

Boosting Methods For Dengue Incidence Rate Prediction in Bandung District

Fhira Nhita^{a1}, Didit Adytia^{a2}, Aniq Atiqi Rohmawati^{a3}

^aSchool of Computing, Telkom University
Jl. Telekomunikasi, Indonesia

¹fhiranhita@telkomuniversity.ac.id (Corresponding author)

²adytia@telkomuniversity.ac.id

³aniqatiqi@telkomuniversity.ac.id

Abstract

Dengue infections are among the top 10 diseases that cause the most deaths worldwide. Dengue is a severe global threat and problem, especially in tropical countries like Indonesia. The Indonesian Ministry of Health also stated that dengue is as dangerous as COVID-19. One of the preventive actions that can be taken is by controlling vectors (the Aedes aegypti mosquito) where weather factors influence their breeding. In this study, the prediction of the dengue incidence rate is carried out using three boosting methods i.e., Extreme Gradient Boosting, Adaptive Boosting, and Gradient Boosting. The data used are monthly data of the dengue incidence rate and weather data. The case study used is Bandung district, West Java Province, Indonesia. The important issue that is investigated in this study is to find the weather parameters that have the most influence on IR and gradually improve the prediction model through three test scenarios. From the test results, the weather parameter that has the most influence on the next month's IR is temperature. Meanwhile, the best training data length is five years (2016-2020). Finally, the best prediction model achieved by the AdaBoost method with the value of Root Mean Square Error and Correlation Coefficient for testing data (January-December 2021) is 0.55 and 0.95, respectively.

Keywords: *Dengue, Boosting, Extreme Gradient Boosting, Adaptive Boosting, Gradient Boosting, Incidence Rate, Bandung District*

1. Introduction

Dengue infections are among the top 10 diseases that cause the most deaths worldwide [1]. Dengue is a severe global threat and problem, especially in tropical countries like Indonesia [2]. The Indonesian Ministry of Health also stated that dengue is as dangerous as COVID-19 [3]. There is no effective antiviral to treat dengue disease, so an important strategy that can be done is to control the vector (in this case, the Aedes aegypti mosquito). One factor that influences the spread of dengue vectors is the weather [3]–[5]. Several factors in weather influence the increment of dengue cases from other research, including rainfall [6], humidity [7], and temperature [8].

To date, many studies have been carried out the dengue prediction to minimize the spread of dengue disease based on weather parameters using a machine learning approach. In 2019, Harumy et al. used the Neural Network and Regression Method algorithm with an accuracy of 87.16%, involving several regions in Indonesia except for West Java [9]. In 2020, Xu et al. predicted dengue cases in 20 cities in China using dengue incidence data and monthly weather data. The algorithm used is LSTM, BPNN, GAM, SVR, and GBM, with an average RMSE of LSTM, which is 32.02 [10].

Our previous study in 2018 conducted the dengue prediction in the Bandung district using a Support Vector Machine (SVM) and K-Means with 93% accuracy [11]. We took the data from Meteorology Climatology and Geophysics Council with Bandung station as the point due to the unavailability of the weather data in the Bandung district. Furthermore, in this previous study, we

have not analyzed the effect of each weather parameter on IR. Another study was conducted by Salim et al. in 2021 using SVM to predict the dengue outbreak in Malaysia. They found that machine learning has good potential for predicting dengue outbreaks, and they suggest future work using a boosting method [12].

Several studies using the boosting method, including Carjaval et al. in 2018 used several meteorological factors to predict dengue incidence in Manila, Philippines using Random Forest and Gradient Boosting [13]. Meanwhile, Salami et al. used the Random Forest and XGBoost algorithm to predict dengue importation for 21 countries in Europe, with the best value of receiver operating characteristic of 0.94 and sensitivity of 0.88 [14]. In 2020, Puengpreeda et al. predicted the dengue outbreak in Thailand using Random Forest and AdaBoost, with the best MSE value of 9.76 [15]. From these studies, there is still an improvement chance in designing a comprehensive prediction method to obtain better prediction performance. Another critical issue is finding the most influential weather factors according to the conditions of each area.

Therefore, in this study, we used three boosting methods i.e., Extreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), and Gradient Boosting (GB), to predict the dengue incidence rate in the Bandung district. The boosting method was chosen because this method can reduce bias, so it is expected to provide better performance. This study aims to investigate the effect of weather parameters on the dengue incidence rate in Bandung district, find the most influential weather factors, and design a comprehensive methodology to produce the best performance based on the Root Mean Square Error (RMSE) and Correlation Coefficients (CC) values. The results obtained in this study can be used as input for developing an early dengue prediction system in the Bandung district. Also, give the information to the Health Department in Bandung district to make precautions of reducing the dengue incidence rate.

2. Research Methods

In this section, we briefly discuss the materials and methods of our study. The stages of research that we carried out in this study are shown in Figure 1. This research methodology included data preparation, measuring the correlation between weather parameters and IR, designing several learning scenarios, and evaluating the performance of each prediction model. The main inputs in this study are IR and weather data. The boosting method is used to predict future IR.

2.1. Dengue cases data

The data used in this study were taken from one area in West Java, Indonesia, namely the Bandung district. West Java province is attractive because it is the province with the largest population in Indonesia. Bandung district was chosen because it is one of the West Java areas with the highest dengue cases. This location has 31 sub-districts and 270 sub-districts. In 2021, the population of Bandung district is 3,633,437 people, with a density of 2,055 people/km². The dengue cases data were obtained from the Bandung district Health Department in the collaboration with School of Computing of Telkom University. The data is the number of cumulative monthly dengue cases from all sub-districts from 2009 until 2021. We used the incidence rate (IR) term, which describes the incidence of dengue cases by 100,000 population as shown in equation (1) [11].

$$IR = \left(\frac{\text{number of dengue case}}{\text{number of population}} \right) \times 100.000 \quad (1)$$

2.2. Weather data

The weather data used in this study is a reanalysis of data from the European Center of Medium-Range Weather Forecasts (ECMWF) provided by ERA5. We retrieved weather data in monthly averages as provided by ERA5 [16]. Siti Aisyah et al. conducted research related to electricity load prediction using weather parameters from ERA5 as input. They found the average trend results similar to data taken from Automatic Weather Station (AWS) [17].

In addition, several studies related to the prediction of dengue incidence in several countries also use weather data taken from ERA5. Cunha et al. conducted an ecological study associated with dengue incidence in Brazil [18]. Also, Lim et al. used ERA5 data, one of which was temperature, to make an inference on dengue epidemics in Singapore [19]. The coordinates of weather data

collection are in Soreang, the capital city of the Bandung district. The description of the location of the study area is shown in Figure 2.

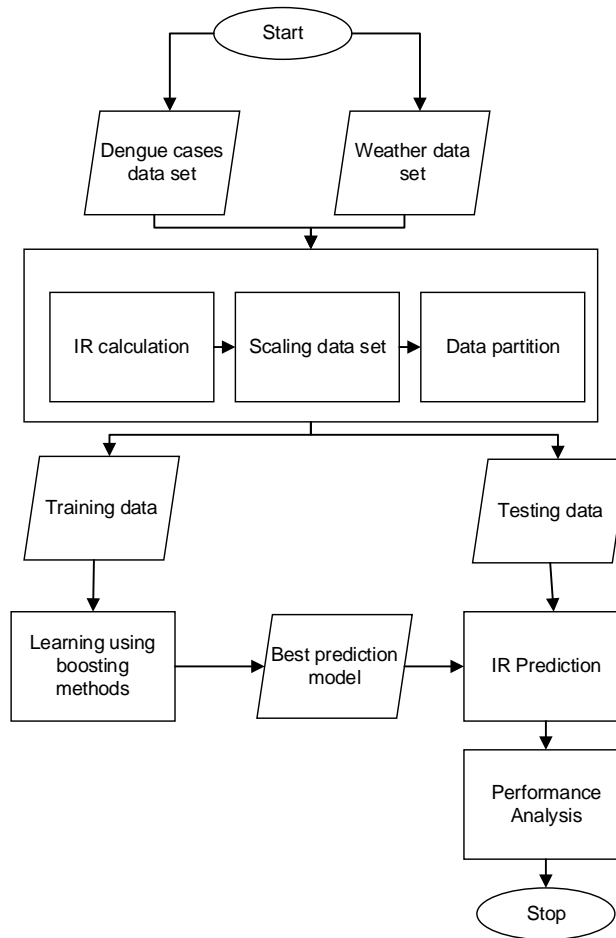


Figure 1. Research methodology for dengue predictions

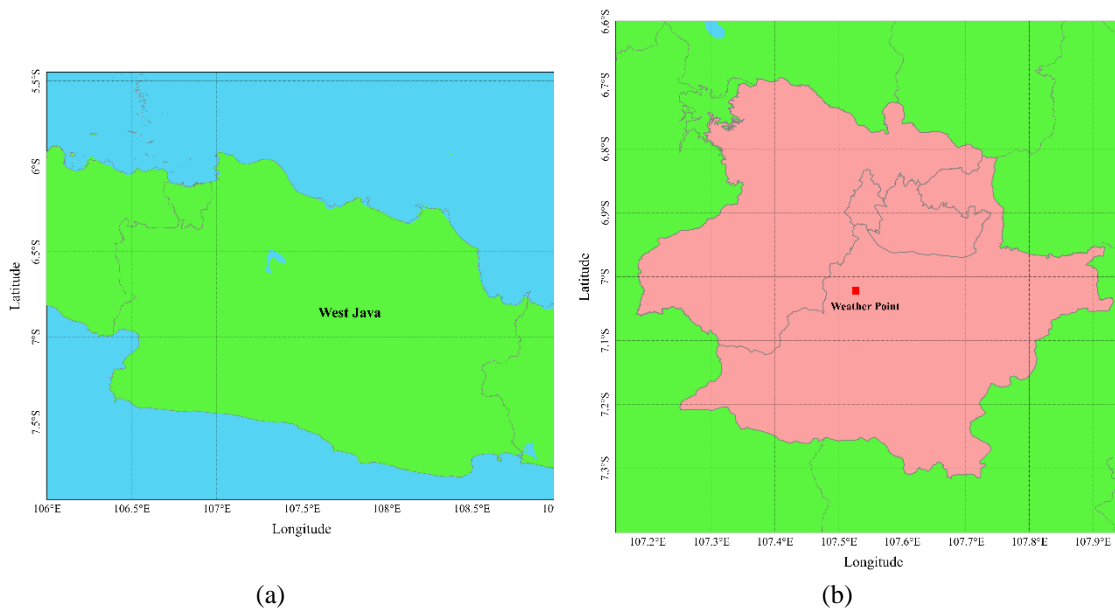


Figure 2. Location of the study area: (a) West Java province, and (b) Bandung district. The red marker denotes the weather point from ERA5

We took seven weather parameters from the ERA5 data set, i.e., 2 meters dew point temperature, 2 meters temperature, surface net thermal radiation-clear sky, surface pressure, mean sea level pressure, relative humidity, and surface net thermal radiation. Detailed information from the weather data is described in Table 1.

Table 1. Weather parameters information

Weather parameter	Abbreviation	Measurement unit
2 meters dewpoint temperature	d2m	Kelvin
2 meters temperature	t2m	Kelvin
surface net thermal radiation-clear sky	strc	Joule Meters**(-2)
surface pressure	sp	Pascals
mean sea level pressure	msl	Pascals
relative humidity	rh	%
surface net thermal radiation	str	Joule Meters**(-2)

2 meters dewpoint temperature represents the temperature to which the air at a height of 2 meters above the Earth's surface must be chilled for saturation to occur. It is a measurement of the air's humidity. It can be used in conjunction with temperature and pressure to calculate relative humidity. Taking into account air conditions, the 2 meters dew point temperature is determined by interpolating between the lowest model level and the Earth's surface. While 2 meters temperature represents the air temperature two meters above the surface of land, sea, or inland water. Taking into account air conditions, 2 meters temperature is determined by interpolating between the lowest model level and the Earth's surface. That parameter is measured in Kelvin (K). Subtract 273.15 from the temperature measured in kelvin to convert it to degrees Celsius (°C).

2.3. Boosting Methods

Boosting is part of the ensemble method that reduces bias to provide better prediction results. In this study, we used three boosting methods i.e., Extreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), and Gradient Boosting (GB).

Adaptive Boosting (AdaBoost) is one of the most popular and broadly used boosting methods [20]. AdaBoost is an ensemble classifier primarily based on a set of rules that mixes more than one vulnerable classifier to provide a sturdy classifier. AdaBoost works by adaptively adjusting the weights of every cycle of the vulnerable classifier of the group. Diversity among weak classifiers allows AdaBoost to provide better results based on the performance of each classifier [21]. The AdaBoost classification has a final equation that can be seen in equation (2) [22],

$$B(x) = \text{sign} \left(\sum_{e=1}^E \alpha_e B_e(x) \right) \quad (2)$$

where E is the train set, B_e stands for the e^{th} weak classifier, and α_e is the corresponding weight coefficient.

Gradient Boosting (GB) is a powerful boosting method that works by developing an ensemble of tree-based models by training each tree sequentially [23], [24]. The most important idea of GB is to construct a predictive version via way of means of acting gradient descent [23]. Below is the gradient boosting method using least-squares approximation as in equation (3) [23], [24],

$$\hat{x}_i = \sum_{n=1}^N k_n(y_i), \quad k_n \in K \quad (3)$$

where n represents the number of trees, k represents the function in the functional space and K represents the set of all possible regression trees.

Extreme Gradient Boosting (XGBoost) is a powerful tree-boosting algorithm that is broadly used by data scientists to improve results [25]. Using the XGBoost method, we can automatically use the CPU's multiple cores for parallel computing, speeding up the calculations [26]. The speed of the model exploration process is helped by this advantage. XGBoost is an enhanced version of GB with better performance and shorter computation time [27]. The objective function calculation of XGBoost is given by equation (4) [26],

$$L = \sum_x l(\hat{a}_i, a_i) + \sum_y \Omega(f_y) \quad (4)$$

where l is the loss function and Ω represents the function used for regularization to prevent overfitting.

2.4. Performance Measurement

We used two measurements to evaluate model performance i.e., Root Mean Square Error (RMSE) and Correlation Coefficient (CC). The formula for calculating the RMSE value is explained in equation (5) [28].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (yp_i - yt_i)^2} \quad (5)$$

where n is the number of records, yp_i is the predicted value and yt_i is the target value for each record.

The smaller the RMSE value, the better the IR prediction results because the distance value between the predicted value and the target value is smaller.

While the formula to calculate the CC value is defined in equation (6) [17]. The CC value is in the range of -1 to +1. The greater the CC value, the better the correlation between the observed attributes.

$$CC = \frac{cov(A, B)}{stdev(A) * stdev(B)} \quad (6)$$

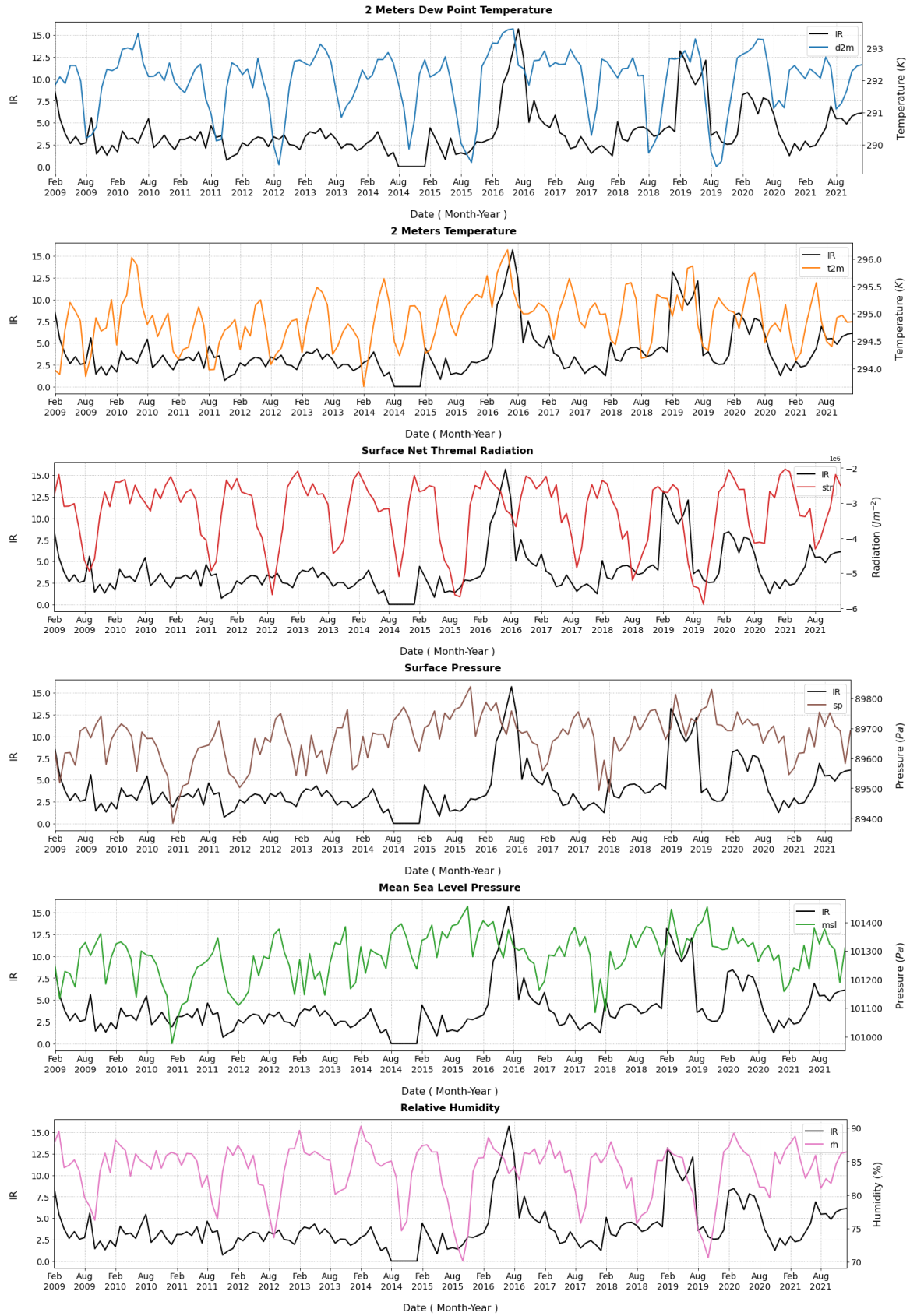
where $cov(A, B)$ is the covariance value between two attributes, namely A and B, $stdev(A)$ and $stdev(B)$ is the standard deviation value of data A and B.

3. Result and Discussion

In this section, we presented the prediction results of the boosting methods. We calculated the Correlation Coefficient for each weather parameter to the IR and implemented three test scenarios to produce the best performance. The Correlation Coefficient is measured using equation (6) where A and B represent IR and each weather parameter, respectively. The training data used data from 2009 to 2020, while the testing data used data from January until December 2021. The following month's IR prediction is made based on the history of IR data and the weather of the previous month. For example, to predict the IR of February 2021, we used IR and weather in January 2021 as input data.

3.1. Correlation Coefficient between weather parameters and IR

To describe the data trend between IR and each weather parameter used in this study, we plotted the data shown in Figure 3.



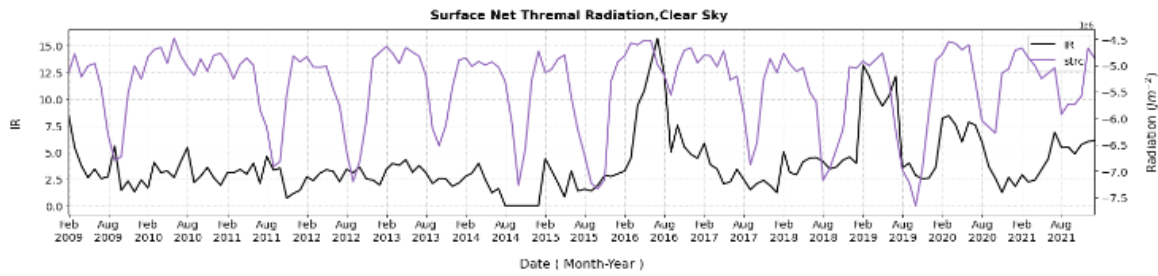


Figure 3. Data plotting between IR and each weather parameter.

The Correlation Coefficient values for each weather parameter with IR are presented in Table 2. The highest Correlation Coefficient value is obtained by 2 meters dewpoint temperature and 2 meters temperature. In contrast, the lowest correlation value is obtained by surface net thermal radiation.

Table 2. Correlation Coefficient between weather parameters and IR

Weather parameter	CC
2 meters dewpoint temperature	0.2916
2 meters temperature	0.2777
surface net thermal radiation-clear sky	0.2255
surface pressure	0.2094
mean sea level pressure	0.1983
relative humidity	0.1867
surface net thermal radiation	0.1462

3.2. Scenario I

In the first scenario, we examined the effect of the length of the training data on the performance of the prediction model for data testing. At this stage, we used all weather parameters as input in the learning process and default parameter settings for each boosting method. The performance of the testing data is presented in Table 3. In this scenario I, we take the best model, which is determined from the highest Correlation Coefficient value. From the four types of training data lengths tested, the best Correlation Coefficient was obtained for the five-year training data length with the highest Correlation Coefficient values being 0.73, 0.94, and 0.67 for XGBoost, AdaBoost, and Gradient Boosting, respectively.

Table 3. Testing performance for the scenario I

Train data length	XGBoost		AdaBoost		Gradient Boosting	
	RMSE	CC	RMSE	CC	RMSE	CC
10 years	1.60	0.43	0.96	0.86	1.58	0.46
8 years	1.58	0.46	0.94	0.87	1.55	0.52
5 years	1.64	0.73	0.91	0.94	1.40	0.67
3 years	1.60	0.51	1.59	0.78	1.64	0.54

3.3. Scenario II

To improve performance in the first scenario, we carried out the second scenario by testing the influence of the weather parameters used as input for the learning process. Weather parameters are entered in stages according to the Correlation Coefficient values generated in Table 2 to see their effect on IR predictions. We determine the best model from the lowest RMSE value in this scenario. The RMSE value is calculated using equation (5) between the predicted value and the target value of IR. Table 4 showed the testing performance results giving the best performance for the d2m parameters with RMSE values are 1.52, 0.67, and 1.06 for XGBoost, AdaBoost, and Gradient Boosting, respectively. These results indicated that 2 meters dewpoint temperature is the weather parameter that has the most influence on future IR predictions.

Table 4. Testing performance for scenario II

Weather parameters	XGBoost		AdaBoost		Gradient Boosting	
	RMSE	CC	RMSE	CC	RMSE	CC
1 (d2m)	1.52	0.70	0.67	0.93	1.06	0.88
2 (d2m, t2m)	1.76	0.66	0.77	0.93	1.58	0.89
3 (d2m, strc, t2m)	1.67	0.74	0.69	0.94	1.46	0.87
4 (d2m, strc, t2m, sp)	1.66	0.69	0.87	0.94	1.39	0.85
5 (d2m, strc, msl, sp, t2m)	1.66	0.68	0.91	0.91	1.37	0.84
6 (d2m, strc, msl, sp, rh, t2m)	1.70	0.71	0.77	0.93	1.47	0.72
All (t2m, d2m, msp, str, strc, sp, msl)	1.64	0.73	0.91	0.94	1.40	0.67

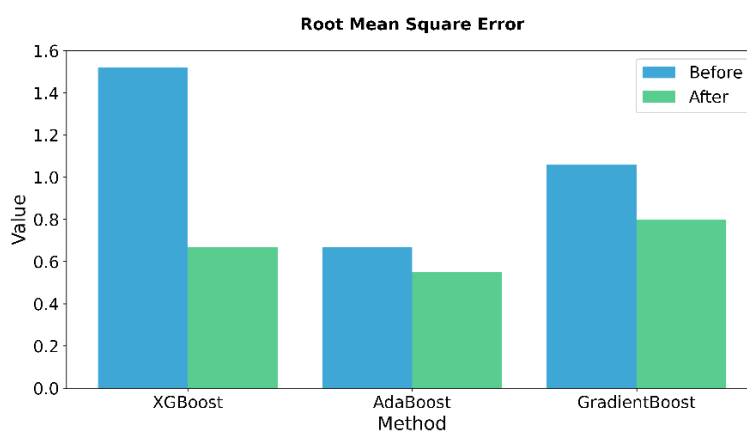
3.4. Scenario III

In the third scenario, we performed the hyperparameter tuning for each boosting method to examine the effect of hyperparameter tuning on RMSE and CC values. In this scenario, the best prediction model obtained in scenarios 1 and 2 is used. The length of the training data is five years, and the best weather parameter is d2m. Table 5 presented the results of data testing performance before and after hyperparameter tuning. These results indicated that hyperparameter tuning significantly affects the RMSE values of all methods. Likewise, the CC value for XGBoost and AdaBoost has increased, while for Gradient Boosting, there has been a slight decrease of 0.01.

Interestingly, the performance of XGBoost after tuning gives a more significant gap between RMSE and CC values than other methods. This indicated that the hyperparameter tuning works very well on the XGBoost method, giving the difference in the RMSE and CC values after tuning that is not far between XGBoost and AdaBoost. In addition, Figure 4 points out the gap between RMSE and CC values before and after hyperparameter tuning for each method is performed. In this last scenario, the best model is produced by the AdaBoost method with an RMSE and CC value are 0.55 and 0.95, respectively.

Table 5. Testing performance for scenario III

Hyperparameter tuning	XGBoost		AdaBoost		Gradient Boosting	
	RMSE	CC	RMSE	CC	RMSE	CC
Before	1.52	0.70	0.67	0.93	1.06	0.88
After	0.67	0.94	0.55	0.95	0.80	0.87



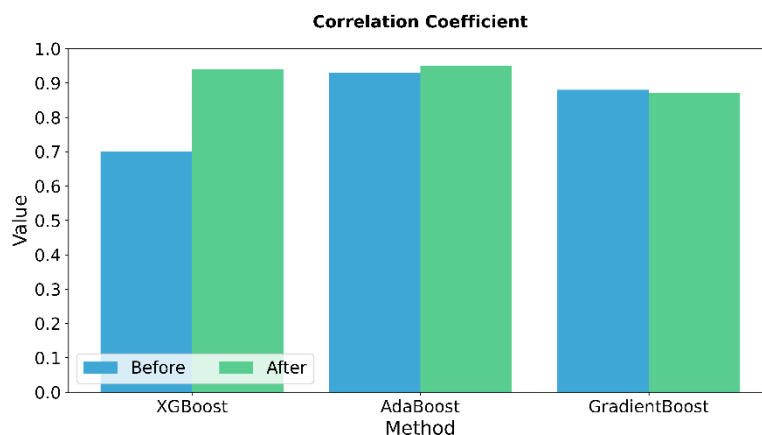


Figure 4. Hyperparameter tuning performances.

3.5. Best prediction model

The three test scenarios discussed in the previous subsection are a comprehensive methodology carried out to obtain better performance in each scenario. The best prediction model is produced by the AdaBoost method with a data training length is five years, the most important of weather parameters is 2 meters dewpoint temperature, and the best method parameters are $n_estimators=20$, $learning_rate=1.5$, $loss='exponential'$. Figure 5 showed the graph between the actual and predicted IR for January-December 2021. The blue color represents the predicted results from AdaBoost, while the black color represents the actual IR. In July 2021, the predicted and actual IR reach the same point, while in other months, there is a difference between the actual and predicted IR. In June, the actual and predicted IR patterns were the same. Both of these values reached their highest peak, which means that the incidence of dengue cases had a peak case in June.

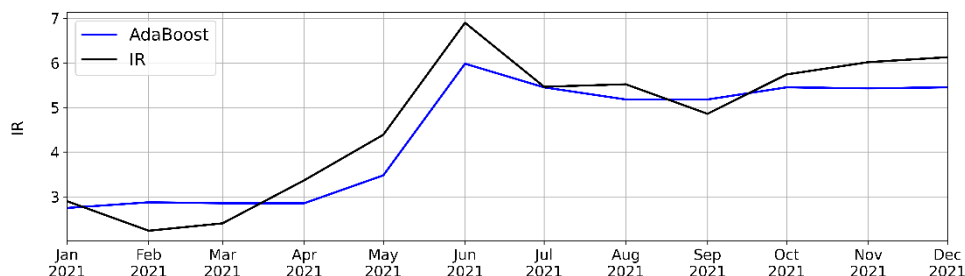


Figure 5. Prediction results for data testing (2021)

4. Conclusion

This study implemented three boosting methods for predicting the dengue incidence rate (IR) in Bandung district, West Java, Indonesia. The data used are monthly data of IR and weather data. Three test scenarios were conducted to find the best predictive model. In the first scenario, the best predictive model is obtained when using a five-year training data length. In the second scenario, we found the most influential weather parameter on IR, which is the temperature (2 meters dewpoint temperature). Meanwhile, in the third scenario, the hyperparameter tuning for each method significantly affects the RMSE and Correlation Coefficient values. The best prediction model was generated by the AdaBoost method with an RMSE and Correlation Coefficient value are 0.55 and 0.95, respectively.

For future work, several issues can be investigated further. First, determine the several weather data points to obtain a more representative weather point with a higher correlation to IR. Second, it is possible to observe the effect of lookback data not only from the previous month to predict the next month of IR. Third, apply the other machine learning methods, such as Random Forest to improve the performance of the prediction model.

References

- [1] P. Sิริyasatien, S. Chadsuthi, K. Jampachaisri, and K. Kesorn, "Dengue epidemics prediction: A survey of the state-of-the-art based on data science processes," *IEEE Access*, vol. 6, pp. 53757–53795, 2018, doi: 10.1109/ACCESS.2018.2871241.
- [2] S. Choudhary, V. Gaurav, T. Sharma, V. V, and P. K R, "Forecasting Dengue and Studying its Plausible Pandemy using Machine Learning," *SSRN Electronic Journal.*, May 2019, doi: 10.2139/SSRN.3507320.
- [3] S. Tiffany, D. Sarwinda, B. D. Handari, and G. F. Hertono, "The comparison between extreme learning machine and artificial neural network-back propagation for predicting the dengue incidences number in DKI Jakarta," *Journal of Physics: Conference Series*, vol. 1821, no. 1, p. 012025, Mar. 2021, doi: 10.1088/1742-6596/1821/1/012025.
- [4] A. M. Najar, M. I. Irawan, and D. Adzkiya, "Extreme Learning Machine Method for Dengue Hemorrhagic Fever Outbreak Risk Level Prediction," *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, Nov. 2018, doi: 10.1109/ICSCEE.2018.8538409.
- [5] W. Anggraeni *et al.*, "Modified Regression Approach for Predicting Number of Dengue Fever Incidents in Malang Indonesia," *Procedia Computer Science.*, vol. 124, pp. 142–150, Jan. 2017, doi: 10.1016/J.PROCS.2017.12.140.
- [6] J. Cheng *et al.*, "Extreme weather conditions and dengue outbreak in Guangdong, China: Spatial heterogeneity based on climate variability," *Environmental Research*, vol. 196, p. 110900, May 2021, doi: 10.1016/J.ENVRES.2021.110900.
- [7] M. Mamenun, Y. Koesmaryono, R. Hidayati, A. Sopaheluwakan, and B. D. Dasanto, "Kemajuan Penelitian Pemodelan Prediksi Demam Berdarah Dengue menggunakan Faktor Iklim di Indonesia : A Systematic Literature Review," *Buletin Penelitian Kesehatan*, vol. 49, no. 4, pp. 231–246, Dec. 2021, doi: 10.22435/BPK.V49I4.4762.
- [8] V. J. Jayaraj, R. Avoi, N. Gopalakrishnan, D. B. Raja, and Y. Umasa, "Developing a dengue prediction model based on climate in Tawau, Malaysia," *Acta Tropica*, vol. 197, Sep. 2019, doi: 10.1016/J.ACTATROPICA.2019.105055.
- [9] T. H. F. Harumy, H. Y. Chan, and G. C. Sodhy, "Prediction for Dengue Fever in Indonesia Using Neural Network and Regression Method," *Journal of Physics: Conference Series*, vol. 1566, no. 1, p. 012019, Jun. 2020, doi: 10.1088/1742-6596/1566/1/012019.
- [10] J. Xu *et al.*, "Forecast of dengue cases in 20 chinese cities based on the deep learning method," *International Journal of Environmental Research and Public Health*, vol. 17, no. 2, Jan. 2020, doi: 10.3390/IJERPH17020453.
- [11] M. M. Muzakki and F. Nhita, "The spreading prediction of Dengue Hemorrhagic Fever (DHF) in Bandung regency using K-means clustering and support vector machine algorithm," *2018 6th International Conference on Information and Communication Technology (ICoICT)*, pp. 453–458, Nov. 2018, doi: 10.1109/ICOICT.2018.8528782.
- [12] N. A. M. Salim *et al.*, "Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques," *Scientific Reports 2021*, vol. 11, no. 1, pp. 1–9, Jan. 2021, doi: 10.1038/s41598-020-79193-2.
- [13] T. M. Carvajal, K. M. Viacrusis, L. F. T. Hernandez, H. T. Ho, D. M. Amalin, and K. Watanabe, "Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines," *BMC Infectious Diseases*, vol. 18, no. 1, pp. 1–15, Apr. 2018, doi: 10.1186/S12879-018-3066-0/FIGURES/3.
- [14] D. Salami, A. Sousa, M. Do, and R. Oliveira Martins, "Predicting Dengue Importation Into Europe, Using Machine Learning and Model-agnostic Methods," *Scientific Reports*, doi: 10.1038/s41598-020-66650-1.
- [15] A. Puengpreeda, S. Yhusumrarn, and S. Sirikulvadhana, "Weekly Forecasting Model for Dengue Hemorrhagic Fever Outbreak in Thailand," *Engineering Journal*, vol. 24, no. 3, pp. 71–87, May 2020, doi: 10.4186/ej.2020.24.3.71.
- [16] H. Hersbach *et al.*, "The ERA5 global reanalysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020, doi: 10.1002/QJ.3803.
- [17] S. Aisyah, A. A. Simaremare, D. Adytia, I. A. Aditya, and A. Alamsyah, "Exploratory Weather Data Analysis for Electricity Load Forecasting Using SVM and GRNN, Case Study in Bali, Indonesia," *Energies*, vol. 15, no. 10, pp. 1–17, 2022, Accessed: Sep. 07, 2022. [Online]. Available: <https://ideas.repec.org/a/gam/jeners/v15y2022i10p3566-d814588.html>.

- [18] M. da C. M. Cunha *et al.*, "Disentangling associations between vegetation greenness and dengue in a Latin American city: Findings and challenges," *Landscape and Urban Planning*, vol. 216, p. 104255, Dec. 2021, doi: 10.1016/J.LANDURBPLAN.2021.104255.
- [19] J. T. Lim, B. S. Dickens, S. Haoyang, N. L. Ching, and A. R. Cook, "Inference on dengue epidemics with Bayesian regime switching models," *PLOS Computational Biology*, vol. 16, no. 5, p. e1007839, May 2020, doi: 10.1371/JOURNAL.PCBI.1007839.
- [20] H. Lu, H. Gao, M. Ye, and X. Wang, "A Hybrid Ensemble Algorithm Combining AdaBoost and Genetic Algorithm for Cancer Classification With Gene Expression Data," *IEEE/ACM Transaction on Computational Biology and Bioinformatics*, 2019.
- [21] I. Kurniawan, M. Rosalinda, and N. Ikhsan, "Implementation of Ensemble Methods on QSAR Study of NS3 Inhibitor Activity as Anti-dengue Agent," *SAR and QSAR Environmental Research*, vol. 31, no. 6, pp. 477–492, 2020.
- [22] J. Wang and S. Tang, "Time series classification based on arima and AdaBoost," *MATEC Web of Conferences*, vol. 309, p. 03024, 2020, doi: 10.1051/MATECONF/202030903024.
- [23] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," *Advanced Neural Information Processing Systems*, vol. 2018-December, pp. 6638–6648, Jun. 2017, Accessed: Dec. 31, 2021. [Online]. Available: <https://arxiv.org/abs/1706.09516v5>.
- [24] L. Liu, M. Ji, and M. Buchroithner, "Combining Partial Least Squares and the Gradient-Boosting Method for Soil Property Retrieval Using Visible Near-Infrared Shortwave Infrared Spectra," *Remote Sensing 2017, Vol. 9, Page 1299*, vol. 9, no. 12, p. 1299, Dec. 2017, doi: 10.3390/RS9121299.
- [25] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [26] W. Li, Y. Yin, X. Quan, and H. Zhang, "Gene expression value prediction based on XGBoost algorithm," *Frontier in Genetics*, vol. 10, p. 1077, 2019.
- [27] R. Dhia'a Abdu-Aljabar and O. A. Awad, "A Comparative analysis study of lung cancer detection and relapse prediction using XGBoost classifier," *IOP Conference Series: Materials Science and Engineering*, vol. 1076, no. 1, p. 012048, Feb. 2021, doi: 10.1088/1757-899X/1076/1/012048.
- [28] A. W. Ramadhan, D. Adytia, D. Saepudin, S. Husrin, and A. Adiwijaya, "Forecasting of Sea Level Time Series using RNN and LSTM Case Study in Sunda Strait," *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, vol. 12, no. 3, pp. 130–140, 2021.