

Comparison of Machine Learning Methods for Monthly Rainfall Forecasting Using k-Nearest Neighbors, Random Forest, and Artificial Neural Networks

Mico Maulana

micomaulana74@gmail.com

Sekolah Tinggi Meteorologi Klimatologi dan Geofisika

Giarno

Sekolah Tinggi Meteorologi Klimatologi dan Geofisika

Research Article

Keywords:

Posted Date: October 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11531449/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Comparison of Machine Learning Methods for Monthly Rainfall Forecasting Using k-Nearest Neighbors, Random Forest, and Artificial Neural Networks

Mico Maulana^{1(*)}, Giarno²

¹State College Of Meteorology Climatology And Geophysics, Tangerang, and Indonesia

² State College Of Meteorology Climatology And Geophysics, Tangerang, and Indonesia

Abstract

Received:

Revised:

Accepted:

Rainfall is an important meteorological parameter across various sectors and exhibits complex nonlinear characteristics, requiring reliable forecasting methods. This study aims to compare the performance of three machine learning algorithms—k-Nearest Neighbors (kNN), Random Forest (RF), and Artificial Neural Networks (ANN)—for monthly rainfall forecasting at the Depati Amir Meteorological Station, Pangkalpinang, using data from 2015 to 2024. The predictor variables included rainfall, mean air temperature, relative humidity, sunshine duration, and rainfall lag features. The modeling process consisted of data preprocessing, feature engineering, hyperparameter optimization using Time Series Cross-Validation, and model evaluation based on Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The results indicate that the Random Forest model achieved the best predictive performance (MAE = 75.73 mm, RMSE = 89.50 mm, R^2 = 0.605), followed by kNN (R^2 = 0.492) and ANN (R^2 = 0.303). These findings demonstrate that Random Forest is more effective than the other two algorithms for monthly rainfall forecasting in the study area.

Keywords:

Rainfall; *Artificial Neural Network*; *K-Nearest Neighbor*; *Random Forest*.

(*) Corresponding Author: micomaulana74@gmail.com

INTRODUCTION

Rainfall is one of the most important climatic variables influencing a wide range of sectors, including the hydrological cycle, agriculture, water resource management, hydrometeorological disaster mitigation, and climate change adaptation. Accurate rainfall information supports informed decision-making by governments, farmers, and water resource managers because it directly affects agricultural productivity, water availability, and the risks of floods, landslides, and droughts. However, rainfall remains one of the most challenging atmospheric variables to forecast due to its highly nonlinear, dynamic nature and its dependence on complex interactions among multiple meteorological and climatological factors (Ejike et al., 2025). Furthermore, improving rainfall forecasting accuracy provides cross-sectoral benefits for telecommunications, agriculture, and environmental management, thereby contributing to the achievement of the Sustainable Development Goals (Raval et al., 2021).

These challenges are particularly pronounced in Indonesia, a tropical archipelagic country located between two continents and two oceans. Rainfall patterns across Indonesia are influenced by the interactions of multiple atmospheric

and oceanic phenomena, including the Asian–Australian monsoon, the El Niño–Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD), the Madden–Julian Oscillation (MJO), local topography, and regional oceanographic conditions (Karnawati, 2026; Ardhasena, 2026). The country's diverse topography generates substantial spatial variability in rainfall, making it difficult for a single forecasting model to accurately represent rainfall patterns across different regions (Aldrian & Susanto, 2003). Moreover, rainfall in tropical regions is predominantly generated by localized convective processes that are more dynamic than the frontal systems characteristic of subtropical climates, resulting in lower predictability (Ejike et al., 2025). Consequently, monthly rainfall forecasting in Indonesia, including the Bangka Belitung Islands served by the Depati Amir Meteorological Station in Pangkalpinang, requires modeling approaches capable of capturing complex nonlinear relationships among meteorological variables.

The limitations of conventional statistical methods in modeling nonlinear relationships have encouraged the adoption of machine learning techniques for rainfall forecasting. Unlike classical statistical approaches, which generally rely on assumptions of linearity and stationarity, machine learning algorithms learn patterns directly from historical observations without requiring strict distributional assumptions, making them well suited for complex meteorological datasets (Pham et al., 2020; Liyew & Melese, 2021). Among the most widely used algorithms, k-Nearest Neighbors (kNN) predicts target values based on the similarity of historical observations, offering a simple yet computationally sensitive approach in high-dimensional feature spaces (Zhang et al., 2017). Random Forest (RF) employs an ensemble of decision trees to improve predictive accuracy while reducing model variance (Appiah-Badu et al., 2022). Meanwhile, Artificial Neural Networks (ANNs) utilize backpropagation-based learning to model highly nonlinear relationships commonly found in meteorological time-series data (Esteves et al., 2019; Kumarasiri & Sonnadara, 2008). Owing to their distinct learning mechanisms, the predictive performance of these algorithms may vary depending on the characteristics of the dataset and the study region.

Numerous studies have investigated the application of machine learning techniques for rainfall prediction under diverse geographical and climatic conditions. In Southeast Asia, a study conducted in Malaysia reported that Random Forest outperformed five other data mining algorithms in rainfall prediction (Zainudin et al., 2016), whereas research in Singapore identified relative humidity and atmospheric water vapor as the primary predictors of precipitation using Support Vector Machines (Manandhar et al., 2019). In Ghana, a comparative study involving Decision Tree, Multilayer Perceptron, Random Forest, XGBoost, and kNN over a 39-year dataset demonstrated that ensemble methods, particularly Random Forest and XGBoost, consistently outperformed kNN (Appiah-Badu et al., 2022). In Brazil, ANN models achieved seasonal rainfall prediction accuracies ranging from 56% to 78%, highlighting the influence of seasonal variability on model performance (Esteves et al., 2019). Similarly, research conducted in Australia found that an optimized neural network model outperformed several other machine learning algorithms across multiple observation sites, emphasizing the importance of local climatic characteristics in improving predictive accuracy (Raval et al., 2021; Sarasa-Cabezuelo, 2022). Furthermore, comparative studies between

tropical and temperate climates have shown that machine learning models, including Random Forest, generally achieve lower predictive performance in tropical regions due to the greater complexity of rainfall-generating processes (Ejike et al., 2025).

In Indonesia, recent studies have reported similar findings. Random Forest has been shown to outperform Naïve Bayes in rainfall classification based on climate types (Indaryono & Saedudin, 2024) and achieved the highest prediction accuracy (85%) for daily rainfall forecasting in Bandung compared with Logistic Regression and Extreme Gradient Boosting (Hapsari, 2025). Meanwhile, backpropagation-based ANN models have been successfully applied to rainfall prediction in North Sumatra (Manalu, 2016), Balikpapan (Muhammad, 2023), the Ahmad Yani Meteorological Station (Herdhyanti et al., 2022), and monthly rainfall forecasting systems (Sunardi et al., 2020).

Despite these advances, several research gaps remain. First, most previous studies have focused on a single algorithm or compared different sets of algorithms, providing limited insight into the relative performance of kNN, Random Forest, and ANN within a unified forecasting framework. Second, many studies have relied on daily or hourly rainfall data and have been conducted in climatic settings that differ substantially from tropical archipelagic environments such as the Bangka Belitung Islands. Third, model evaluation has primarily emphasized classification metrics, such as accuracy and F1-score, whereas the combined use of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2) for evaluating continuous rainfall forecasts remains relatively limited. Moreover, studies directly comparing kNN, Random Forest, and ANN using observational data from the Indonesian Meteorology, Climatology, and Geophysics Agency (BMKG) for monthly rainfall forecasting at the Depati Amir Meteorological Station are still scarce.

To address these gaps, this study compares the performance of k-Nearest Neighbors, Random Forest, and Artificial Neural Networks within a consistent evaluation framework for monthly rainfall forecasting. The models were developed using BMKG observational data collected between 2015 and 2024, including rainfall, mean air temperature, relative humidity, sunshine duration, and rainfall lag features. Model performance was evaluated using MAE, RMSE, and R^2 to identify the most effective forecasting approach. The novelty of this study lies in the integration of these machine learning algorithms with long-term BMKG observational data from a tropical archipelagic region under a unified evaluation framework, rather than in the development of a new forecasting algorithm.

Accordingly, the objective of this study is to compare the predictive performance of k-Nearest Neighbors, Random Forest, and Artificial Neural Networks for monthly rainfall forecasting. The findings are expected to identify the most suitable machine learning model for improving rainfall forecasting accuracy and to provide practical recommendations for the development of rainfall forecasting systems in tropical archipelagic regions of Indonesia.

METHODS

The research workflow consisted of data validation, data preprocessing, exploratory data analysis (EDA), feature engineering, data partitioning, machine learning model development, model evaluation, and selection of the best-performing model. The overall research framework is illustrated in Figure 1.



Figure 1. The research methodology flowchart.

Data Preprocessing

The preprocessing stage included data structure validation, missing value assessment, duplicate removal, and aggregation of daily observations into monthly data. A monthly completeness check was subsequently performed to ensure the integrity and continuity of the dataset before model development.

The dataset was then divided chronologically into a training set (2015–2022) and a testing set (2023–2024). To preserve the temporal dependency of the

observations during model development and hyperparameter optimization, Time Series Cross-Validation (TSCV) was employed on the training dataset.

Machine Learning Models

k-Nearest Neighbor (kNN)

The k-Nearest Neighbors (kNN) algorithm predicts the target value based on the average of the k nearest observations in the feature space. The similarity between observations is measured using the Euclidean distance:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

where

- $d(x_i, x_j)$ denotes the Euclidean distance between observations
- p is the number of predictor variables,
- x_{ik} represents the value of the k -th predictor for the i -th observation.

Because the kNN algorithm is sensitive to differences in variable scales, all predictor variables were standardized prior to model training to ensure that each feature contributed equally to the distance calculation.

Random Forest

Random Forest (RF) is an ensemble learning algorithm that constructs multiple decision trees using bootstrap sampling and random feature selection. Each tree is trained on a different bootstrap sample of the training data, while a random subset of predictor variables is considered at each split to increase model diversity and reduce overfitting. For regression problems, the final prediction is obtained by averaging the predictions of all individual trees.

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

where

- B is the number of decision trees in the ensemble,
- $T_b(x)$ denotes the prediction produced by the b -th.

In addition to generating predictions, the Random Forest model was used to assess the relative importance of each predictor variable through feature importance analysis, providing insights into the contribution of individual predictors to the model's forecasting performance.

Artificial Neural Network

The Artificial Neural Network (ANN) model employed in this study was a Multilayer Perceptron (MLP). The learning process was performed through forward propagation and backpropagation, enabling the network to iteratively adjust its

weights by minimizing the prediction error. The output of each neuron was computed using an activation function:

$$y = f(\sum_{i=1}^n w_i x_i + b)$$

where

- w_i is the weight associated with the i -th input
- x_i is the i -th input variable;
- b is the bias term; and
- $f(\cdot)$ is the activation function.

The network parameters (weights and biases) were iteratively updated using the backpropagation algorithm until the prediction error converged to a minimum.

Model Evaluation

The predictive performance of the models was evaluated using three statistical metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). These metrics were selected to assess prediction accuracy and the models' ability to explain the variability in the observed rainfall data.

The Mean Absolute Error (MAE) was calculated as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

The Root Mean Square Error (RMSE) was calculated as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

The Coefficient of Determination (R^2) was calculated as:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

where:

- y_i is the observed value;
- $\hat{y}_i$ is the predicted value;
- $\bar{y}$ is the mean of the observed values; and
- n is the total number of observations.

Models with lower MAE and RMSE values and a higher R^2 value were considered to exhibit superior predictive performance and were therefore selected as the best-performing models.

RESULTS & DISCUSSION

Three machine learning algorithms, namely k-Nearest Neighbors (kNN), Random Forest (RF), and Artificial Neural Networks (ANN), were compared for monthly rainfall forecasting. The models were trained using data from 2015 to 2022 and evaluated on an independent test set covering the period from January 2023 to December 2024. Model performance was assessed using three statistical metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). In addition, visual evaluation was conducted through observed-versus-predicted plots, scatter plots, residual time series, and residual distribution plots to examine the prediction accuracy and error characteristics of each model.

Results of Monthly Rainfall Forecasting Using k-Nearest Neighbors (kNN)

The k-Nearest Neighbors (kNN) model achieved its optimal performance with $k = 13$, as determined through the time-series cross-validation optimization process. This optimal value was selected because it provided the best predictive performance on the training data while preserving the temporal structure of the time series. The monthly rainfall predictions generated by the optimized kNN model are presented in Figure 2.

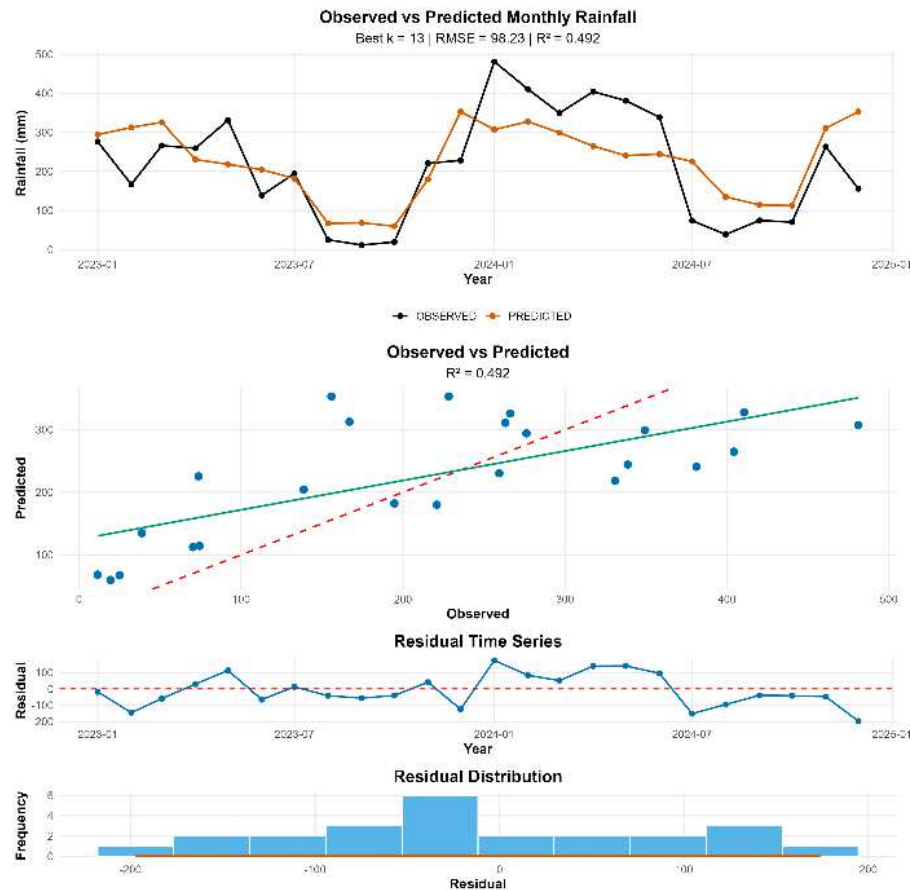


Figure 2. Monthly Rainfall Forecasting Results Using k-Nearest Neighbors (kNN)

As shown in Figure 2, the kNN model was able to capture the overall pattern of monthly rainfall variability during the testing period. However, the model tended to underestimate rainfall during several high-rainfall months, resulting in smoother

predictions than the observed values, indicating an over-smoothing effect. The scatter plot demonstrated a positive relationship between the observed and predicted rainfall values, with an R^2 of 0.4919, indicating that approximately 49.19% of the variability in monthly rainfall was explained by the model.

The model yielded an MAE of 83.5317 mm, indicating a relatively large average absolute difference between the predicted and observed rainfall values. Similarly, the RMSE of 98.2307 mm suggested the presence of relatively large prediction errors, particularly during extreme rainfall events. The residual time series showed that most prediction errors fluctuated around zero, indicating no apparent systematic bias. However, several large residuals were observed during specific periods, reflecting the model's limited ability to accurately capture extreme rainfall events.

Results of Monthly Rainfall Forecasting Using Random Forest

The Random Forest (RF) model was developed using 1,000 decision trees ($n_{tree} = 1000$), with the optimal parameter $mtry = 1$, as determined through the model optimization process. The rainfall predictions generated by the optimized RF model are presented in Figure 3.

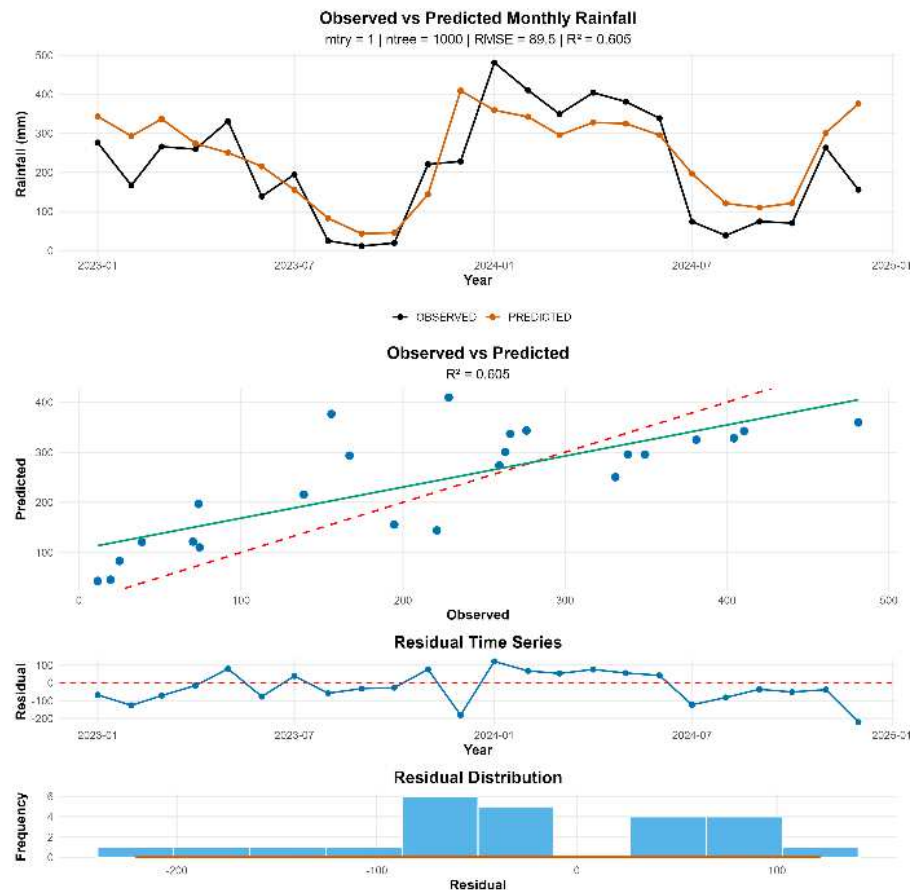


Figure 3. Monthly Rainfall Forecasting Results Using Random Forest

As shown in Figure 3, the Random Forest (RF) model produced predictions that most closely matched the observed rainfall compared with the other models. The model effectively captured rainfall variability during both the wet and dry seasons, resulting in predicted fluctuations that closely followed the observed temporal pattern.

The scatter plot yielded an R^2 value of 0.6047, the highest among all evaluated models, indicating that approximately 60.47% of the variability in monthly rainfall was explained by the RF model. Furthermore, the model achieved an MAE of 75.7328 mm and an RMSE of 89.5022 mm, both of which were the lowest among the evaluated methods, indicating superior predictive accuracy.

The residual distribution was also more concentrated around zero than those of the other models, suggesting smaller and more stable prediction errors throughout the testing period. Overall, these results demonstrate that the Random Forest model provided the best predictive performance among the three machine learning algorithms evaluated in this study.

Results of Monthly Rainfall Forecasting Using Artificial Neural Networks (ANN)

The Artificial Neural Network (ANN) model achieved its best performance using a network architecture with five hidden neurons and a decay parameter of 0.0001, as determined through the model optimization process. The monthly rainfall predictions generated by the optimized ANN model are presented in Figure 4.

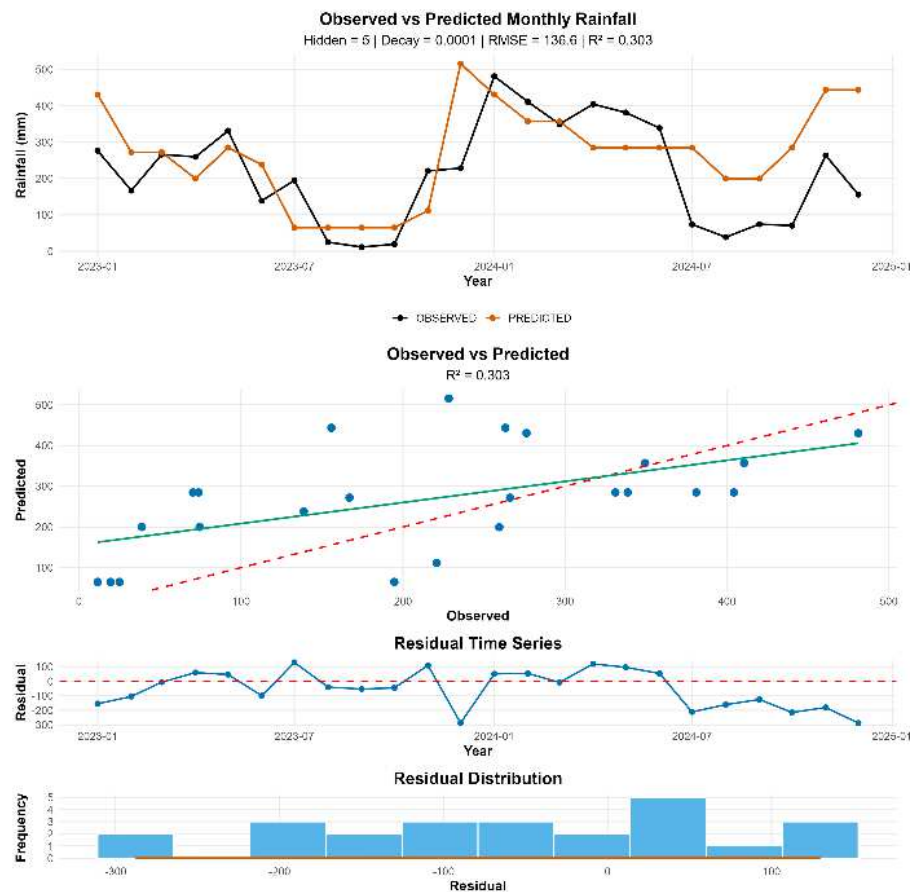


Figure 4. Monthly Rainfall Forecasting Results Using Artificial Neural Networks (ANN)

As shown in Figure 4, the Artificial Neural Network (ANN) model was able to capture the seasonal pattern of monthly rainfall. However, its predictions exhibited larger deviations from the observed values than those of the other two models. The model tended to overestimate rainfall during several high-rainfall events while substantially underestimating rainfall during other periods, resulting in less stable prediction performance.

The scatter plot yielded an R^2 value of 0.3033, indicating that the ANN model explained only 30.33% of the variability in monthly rainfall. In addition, the model produced an MAE of 112.3761 mm and an RMSE of 136.5984 mm, both of which were the highest among the evaluated models, indicating the lowest predictive accuracy.

The residual time series also exhibited greater variability, and the residual distribution was more dispersed than those of the Random Forest and kNN models, indicating larger and less consistent prediction errors throughout the testing period.

Performance Comparison of Machine Learning Models

To identify the most suitable model for monthly rainfall forecasting, the predictive performance of the three machine learning algorithms was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). Lower values of MAE and RMSE indicate smaller prediction errors, whereas a higher R^2 value reflects a greater ability of the model to explain the variability in the observed rainfall data. The performance comparison of the three models is summarized in Table 1.

Table 1. Comparison of Machine Learning Model Performance

Model	Optimal Parameters	MAE (mm)	RMSE (mm)	R^2
K-Nearest Neighbor	k = 13	83.5317	98.2307	0.4919
Random Forest	mtry = 1, ntree = 1000	75.7328	89.5022	0.6047
Artificial Neural Network	Hidden = 5, Decay = 0.0001	112.3761	136.5984	0.3033

As presented in Table 1, the Random Forest (RF) model achieved the best predictive performance among the three machine learning algorithms. It produced the lowest MAE (75.7328 mm) and RMSE (89.5022 mm) values, together with the highest R^2 (0.6047). These results indicate that the RF model generated the smallest prediction errors while providing the greatest ability to explain the variability in monthly rainfall.

The k-Nearest Neighbors (kNN) model ranked second, with an MAE of 83.5317 mm, an RMSE of 98.2307 mm, and an R^2 of 0.4919. Although the model successfully captured the seasonal pattern of monthly rainfall, it tended to produce smoother predictions, limiting its ability to accurately represent extreme rainfall variations.

In contrast, the Artificial Neural Network (ANN) model exhibited the lowest predictive performance, with an MAE of 112.3761 mm, an RMSE of 136.5984 mm, and an R^2 of 0.3033. These results indicate that the ANN model produced larger prediction errors and explained a substantially smaller proportion of the variability in the observed rainfall than the other two models.

Comparison of Monthly Rainfall Prediction Patterns

In addition to the statistical evaluation, the three machine learning models were compared based on their ability to reproduce the temporal pattern of monthly rainfall during the testing period. This visual comparison provides further insight into each model's

capability to capture rainfall variability and seasonal fluctuations. The comparison of the predicted and observed rainfall patterns is presented in Figure 5.

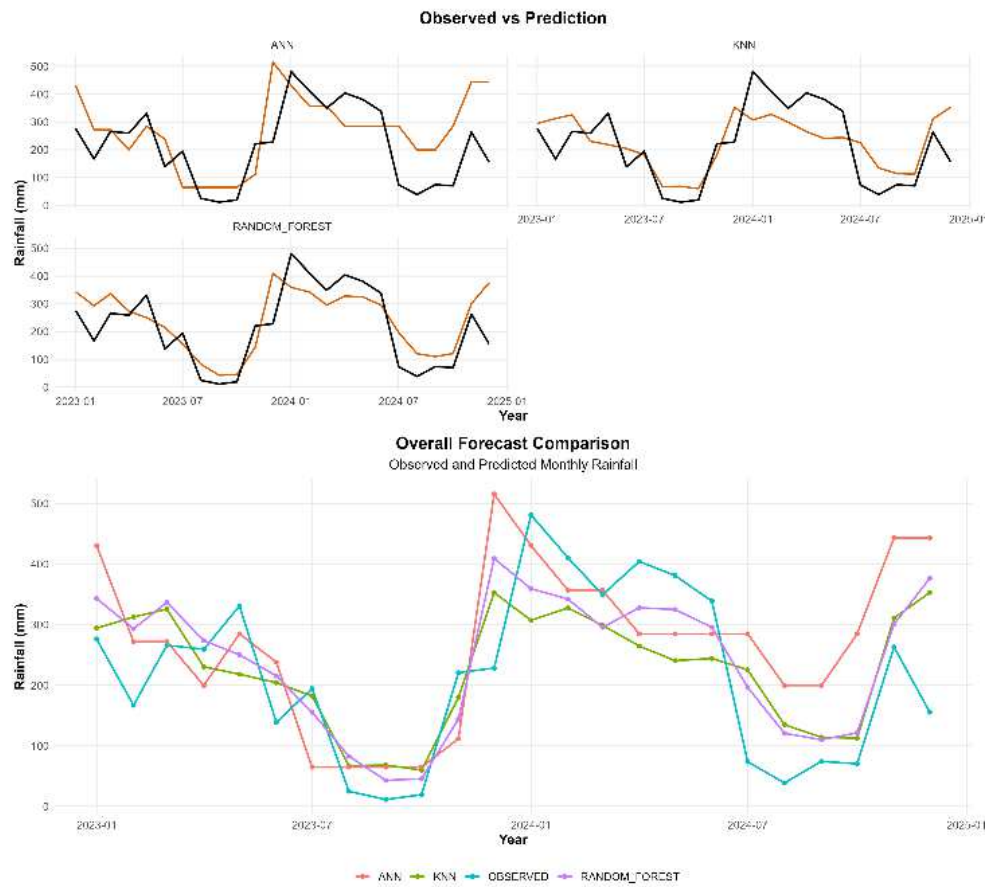


Figure 5. Comparison of Monthly Rainfall Prediction Results from the Three Machine Learning Models

As illustrated in Figure 5, all three machine learning models were able to capture the seasonal pattern of monthly rainfall; however, their agreement with the observed data varied considerably. The Random Forest (RF) model produced predictions that most closely matched the observed rainfall throughout the testing period, demonstrating its superior ability to reproduce the temporal variability of monthly rainfall.

The k-Nearest Neighbors (kNN) model also captured the overall rainfall pattern but generated smoother prediction curves, resulting in an underrepresentation of several peak rainfall events. In contrast, the Artificial Neural Network (ANN) model exhibited larger deviations from the observed values, particularly during periods characterized by abrupt changes in rainfall, indicating a lower capability to capture rapid rainfall fluctuations.

The results of this study indicate that the Random Forest (RF) model provided the best predictive performance for monthly rainfall forecasting at the study site. This superiority is reflected by the lowest RMSE (89.5022 mm) and MAE (75.7328 mm) values, together with the highest R^2 (0.6047) among the evaluated models. These findings suggest that the RF model explained approximately 60% of the variability in monthly rainfall while producing smaller prediction errors than the other machine learning algorithms.

The superior performance of the RF model can be attributed to its ensemble learning framework, which combines multiple decision trees through the bootstrap aggregating (bagging) technique. This approach enables the model to capture complex nonlinear

relationships between predictor variables, including air temperature, relative humidity, sunshine duration, and antecedent monthly rainfall, and the rainfall observed in the subsequent month. Furthermore, averaging the predictions from a large number of decision trees reduces model variance, resulting in more robust and stable predictions while minimizing the risk of overfitting.

The k-Nearest Neighbors (kNN) model achieved the second-best performance, with an R^2 value of 0.4919. Although the model successfully reproduced the overall seasonal pattern of monthly rainfall, it exhibited limited capability in predicting extreme rainfall events. This limitation arises because the kNN algorithm estimates predictions based on the average values of the nearest neighboring observations, producing smoother prediction curves. Consequently, several peak rainfall events were underestimated, contributing to relatively higher RMSE values than those obtained by the RF model.

In contrast, the Artificial Neural Network (ANN) model demonstrated the lowest predictive performance among the evaluated algorithms. One possible explanation is the relatively limited size of the training dataset, which consisted of approximately 96 monthly observations. Such a dataset may not provide sufficient information for the neural network to learn the underlying nonlinear relationships effectively. In general, ANN models require substantially larger datasets to optimize network parameters and achieve good generalization performance. Consequently, the limited training data likely contributed to the larger prediction errors observed in this study.

Overall, the findings demonstrate that the Random Forest algorithm is the most suitable machine learning approach for monthly rainfall forecasting in the study area. By achieving the lowest MAE and RMSE values together with the highest R^2 , the RF model provided the best balance between predictive accuracy and model stability. These results further suggest that ensemble-based machine learning approaches are more effective than instance-based and neural network-based methods for monthly rainfall forecasting when only a relatively limited number of observations are available.

CONCLUSION

This study compared the performance of three machine learning algorithms, namely k-Nearest Neighbors (kNN), Random Forest (RF), and Artificial Neural Networks (ANN), for monthly rainfall forecasting. The results demonstrated that the Random Forest model outperformed the other two algorithms, achieving the lowest prediction errors (MAE = 75.7328 mm and RMSE = 89.5022 mm) and the highest coefficient of determination (R^2 = 0.6047). These findings indicate that the ensemble learning approach employed by Random Forest is more effective in capturing the nonlinear relationships among meteorological variables in the dataset used in this study.

Overall, the results suggest that Random Forest is a reliable method for monthly rainfall forecasting in the study area. Future research may further improve forecasting performance by incorporating additional predictor variables, extending the length of the training dataset, or comparing Random Forest with more advanced machine learning and deep learning algorithms.

REFERENCES

- Aldrian, E., & Susanto, R. D. (2003). Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature. *International Journal of Climatology*, 23(12), 1435–1452. <https://doi.org/10.1002/joc.950>

- Alzahrani, A. J., Ayyub, N., & Parchami, A. (2020). A comparative analysis of machine learning models for rainfall prediction using Orange. *International Journal of Engineering Research and Technology*, 9(10), 893–898.
- Appiah-Badu, N. K. A., Missah, Y. M., Amekudzi, L. K., Ussiph, N., Frimpong, T., & Ahene, E. (2022). Rainfall prediction using machine learning algorithms for the various ecological zones of Ghana. *IEEE Access*, 10, 5069–5082. <https://doi.org/10.1109/ACCESS.2021.3139312>
- Ardhasena. (2026, 10 Juni). *BMKG: Faktor geografis tingkatkan kompleksitas prediksi musim di Indonesia*. Media Indonesia. <https://mediaindonesia.com/humaniora/899123/bmkg-faktor-geografis-tingkatkan-kompleksitas-prediksi-musim-di-indonesia>
- Badan Meteorologi, Klimatologi, dan Geofisika. (2024). *Climate outlook 2024*. <https://www.bmkg.go.id/iklim/buletin-iklim/climate-outlook-2024>
- Badan Meteorologi, Klimatologi, dan Geofisika. (t.t.). *Prediksi musim*. Diakses 11 Juli 2026, dari <https://www.bmkg.go.id/iklim/prediksi-musim>
- Chen, S. T., Yu, P. S., & Tang, Y. H. (2010). Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of Hydrology*, 385(1–4), 13–22. <https://doi.org/10.1016/j.jhydrol.2010.01.021>
- Ejike, O., Ndzi, D., & Shakir, M. Z. (2025). Comparative study of machine learning-based rainfall prediction in tropical and temperate climates. *Climate*, 13(8), 167. <https://doi.org/10.3390/cli13080167>
- Esteves, J. T., de Souza Rolim, G., & Ferraudo, A. S. (2019). Rainfall prediction methodology with binary multilayer perceptron neural networks. *Climate Dynamics*, 52, 2319–2331. <https://doi.org/10.1007/s00382-018-4252-x>
- Hapsari, I. (2025). Evaluasi model prediksi curah hujan berbasis machine learning di Kota Bandung. *Jurnal Nasional Teknologi dan Sistem Informasi*, 11(2), 136–143. <https://doi.org/10.25077/TEKNOSI.v11i2.2025.136-143>
- Herdhyanti, A., Muflikhah, L., & Cholissodin, I. (2022). Prediksi curah hujan dengan empat parameter menggunakan backpropagation (Studi kasus: Stasiun Meteorologi Ahmad Yani). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(12).
- Indaryono, N. A. P., & Saedudin, Rd. R. (2024). Analisa perbandingan algoritma random forest dan naïve bayes untuk klasifikasi curah hujan berdasarkan iklim di Indonesia. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 9(1), 158–167. <https://doi.org/10.29100/jupi.v9i1.4421>
- Karnawati, D. (2026, 5 Juni). *Cuaca Indonesia dipengaruhi kompleksitas geografis dan oseanografi*. RRI.co.id. <https://rri.co.id/nasional/2470521/dwikorita-cuaca-indonesia-dipengaruhi-kompleksitas-geografis-dan-oseanografi>
- Kumarasiri, A. D., & Sonnadara, U. J. (2008). Performance of an artificial neural network on forecasting the daily occurrence and annual depth of rainfall at a tropical site. *Hydrological Processes*, 22(18), 3535–3542. <https://doi.org/10.1002/hyp.6964>
- Liyew, C. M., & Melese, H. A. (2021). Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, 8, 153. <https://doi.org/10.1186/s40537-021-00545-4>
- Manalu, M. T. (2016). Jaringan syaraf tiruan untuk memprediksi curah hujan Sumatera Utara dengan metode back propagation (Studi kasus: BMKG Medan). *Jurnal Riset Komputer*, 3(1), 35–40.
- Manandhar, S., Dev, S., Lee, Y. H., Meng, Y. S., & Winkler, S. (2019). A data-driven approach for accurate rainfall prediction. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11), 9323–9330. <https://doi.org/10.1109/TGRS.2019.2926110>

- Muhammad, F. A. P. (2023). *Prediksi curah hujan menggunakan artificial neural network backpropagation di Kota Balikpapan* [Skripsi sarjana, Institut Teknologi Kalimantan]. Repository ITK.
- Pham, B. T., Le, L. M., Le, T.-T., Bui, K.-T. T., Le, V. M., Ly, H.-B., & Prakash, I. (2020). Development of advanced artificial intelligence models for daily rainfall prediction. *Atmospheric Research*, 237, 104845. <https://doi.org/10.1016/j.atmosres.2020.104845>
- Pham, Q. B., Yang, T. C., Kuo, C. M., Tseng, H. W., & Yu, P. S. (2019). Combining random forest and least square support vector regression for improving extreme rainfall downscaling. *Water*, 11(3), 451. <https://doi.org/10.3390/w11030451>
- Raval, M., Sivashanmugam, P., Pham, V., Gohel, H., Kaushik, A., & Wan, Y. (2021). Automated predictive analytics tool for rainfall forecasting. *Scientific Reports*, 11, 17704. <https://doi.org/10.1038/s41598-021-95735-8>
- Sarasa-Cabezuelo, A. (2022). Prediction of rainfall in Australia using machine learning. *Information*, 13(4), 163. <https://doi.org/10.3390/info13040163>
- Sunardi, S., Yudhana, A., & Muflih, G. (2020). Sistem prediksi curah hujan bulanan menggunakan jaringan saraf tiruan backpropagation. *Jurnal Sistem Informasi Bisnis*, 10(2).
- Zainudin, S., Jasim, D. S., & Bakar, A. A. (2016). Comparative analysis of data mining techniques for Malaysian rainfall prediction. *International Journal on Advanced Science, Engineering and Information Technology*, 6(6), 1148–1153. <https://doi.org/10.18517/ijaseit.6.6.1487>
- Zhang, S., Li, X., Zong, M., Zhu, X., & Wang, R. (2017). Efficient kNN classification with different numbers of nearest neighbors. *IEEE Transactions on Neural Networks and Learning Systems*, 29(5), 1774–1785. <https://doi.org/10.1109/TNNLS.2017.2673241>