Electronic Journal of Education, Social Economics and Technology
Vol. 6, No. 1, (2025), pp. 794~803
ISSN 2723-6250 (online)
DOI: https://doi.org/10.33122/ejeset.v6i1.955

SAINTIS publishing

Research Article

Performance Comparison of Manual and Automatic Rain Gauge Using XGBoost and Random Forest Regression

Indri Budiarto*, and Aji Supriyanto

Faculty of Information Technology and Industry, Universitas STIKUBANK Semarang, Jawa Tengah, Indonesia, 50241
*Corresponding Author: indribudiarto0025@mhs.unisbank.ac.id | Phone: +628995839858

ABSTRACT

This study aims to compare the performance of automatic and manual rain gauges in the Central Highlands of Central Java using a machine learning approach based on Extreme Gradient Boosting (XGBoost) and Random Forest Regression (RFR) algorithms. Daily rainfall data were collected from five regencies—Banyumas, Banjarnegara, Wonosobo, Temanggung, and Pemalang—between 2021 and 2024. Preprocessing involved merging data from two types of instruments (Automatic Rain Gauge/AWS and manual ombrometer), correcting anomalies, and standardizing date-time formats. The models were developed using feature engineering techniques, including multi-lag and moving averages, and evaluated using MAE, RMSE, and R-squared (R^2) metrics. The results show that the XGBoost model with automatic data achieved the best performance, with a Mean Absolute Error (MAE) of 17.3632 mm, Root Mean Squared Error (RMSE) of 27.0282 mm, and R^2 of 0.5050. In comparison, the Random Forest model with automatically generated data produced an MAE of 16.6307 mm, an RMSE of 28.5286 mm, and an R^2 of 0.4485. Models with manual data showed lower performance, with R^2 values below 0.30. These findings indicate that automatic measurement data are more stable and effective for building predictive rainfall models using machine learning. This supports the use of automatic instruments as the primary data source in rainfall forecasting and hydrometeorological disaster mitigation systems.

Keywords: Automatic Gauge; Predictive Modeling; Rainfall; Random Forest; XGBoost

1. INTRODUCTION

Rainfall is a crucial climatic parameter essential for various purposes, such as water resource management, agriculture, and hydro-meteorological disaster mitigation. Rainfall distribution in Central Java varies significantly spatially, especially during peak rainy season months. Most of Central Java was in the medium to high category (201–500 mm) last December 2024, but some areas like Wonosobo, Banjarnegara, and western Temanggung even fell into the very high category with rainfall exceeding 500 mm. Due to this high intensity, available data must be complete, accurate, and reliable. The use of automatic rain gauges is considered more efficient and accurate because they can record data continuously without human intervention. However, their utilization in the BMKG Central Java work area is still limited due to the perception that their data quality is lower than manual instruments (Yan & Duan, 2022). Although important for various purposes, accurate rainfall measurement becomes more challenging in areas with complex topography, such as the Central Java Mountains. Variations in rainfall amount, instrument sensitivity, and surrounding environmental conditions are some factors that can cause discrepancies between manual and automatic instrument readings (Shankar et al., 2019).

Rainfall measurements are generally conducted using two types of instruments: manual and automatic. Manual instruments like ombrometers are still widely used because they are simple and accurate under certain conditions. Automatic instruments such as tipping buckets and ultrasonic sensors are better for real-time recording and efficiency but are still often compared to manual instruments, which have long been the standard. Automatic rain gauges have accuracy limitations during high rainfall events, but their precision can be improved through calibration and the implementation of error reduction strategies (Dunkerley, 2025). Manual rain gauges have two main advantages: they are easy to use and highly resistant to technical damage. However, this method relies heavily on the presence and diligence of the observer, so there may be errors in records, and they cannot consistently record rainfall intensity (Droździoł & Absalon, 2023). Consequently, comparative studies are needed to evaluate the performance differences between the two, especially in areas with complex topography like the Central Java Mountains (Segovia-Cardozo et al., 2023) .

To understand the performance differences between manual and automatic rain gauges, this research employs an artificial intelligence-based approach using the Extreme Gradient Boosting (XGBoost) and Random Forest Regression (RFR) methods. XGBoost is effective in handling small datasets, feature selection, and reducing overfitting (No et al., 2024) (Quayesam et al., 2024). Meanwhile, RFR, an ensemble algorithm based on decision trees, can produce more stable models, having achieved 86.55% accuracy in rainfall classification in Indonesia (Rahman et al., 2022). Through the application of these two methods, the research aims to analyze data differences from both types of instruments, identify factors influencing measurement results, and provide recommendations for selecting the most suitable instrument for the geographical conditions of the Central Java Mountains region.

Various studies show that automatic weather stations (AWS) tend to record lower rainfall and higher temperatures compared to manual stations (Lukasová et al., 2024) (Urban & Strug, 2021). To reduce this bias, the monthly regression method has proven to be effective. Also, using machine learning models like XGBoost has been shown to be much better at estimating and predicting rainfall and temperature than traditional methods and other algorithms like Random Forest (RF) and Multivariate Linear Regression (MLR) (Putra et al., 2024) (Liyew & Melese, 2021). These advancements highlight the importance of integrating modern technology into meteorological practices. As the field continues to evolve, leveraging such innovative approaches can enhance the accuracy of climate data, ultimately leading to more reliable weather forecasts and climate modeling. By harnessing the power of big data and advanced analytics, meteorologists can uncover patterns that were previously undetectable. This shift not only improves forecasting capabilities but also aids in better preparedness for extreme weather events. XGBoost also shows high accuracy results in rainfall classification (Yasper et al., 2023) and is capable of handling complex non-linear relationships between meteorological variables and prediction outcomes (Kumar, Kedam, Sharma, Mehta, et al., 2023).

The integration of remote sensing data with machine learning algorithms, particularly XGBoost, also enhances the accuracy of spatiotemporal rainfall predictions (Papacharalampous et al., 2023) (Nicola et al., 2024). Research by (Li et al., 2023) shows that XGBoost is capable of significantly correcting rainfall measurement bias in highland areas, even better than traditional methods such as the Transfer Function Method (TFM). On the other hand, GRU also shows advantages in short-term weather prediction classification based on IoT data, although XGBoost excels in early regression (Darmawan et al., 2023). In general, XGBoost is recommended for applications in the agriculture, transportation, and weather risk mitigation sectors due to its high prediction accuracy and flexibility in handling various types of data (Huber et al., 2022) (Rowe et al., 2022). These attributes make XGBoost a popular choice among researchers and practitioners who require reliable and efficient models for decision-making processes. As the integration of machine learning continues to evolve, it is likely that these methods will further enhance predictive capabilities across multiple industries. This evolution will not only improve existing models but also pave the way for innovative applications that can address complex challenges. By leveraging advanced algorithms like XGBoost, organizations can optimize their operations and achieve better outcomes in an increasingly data-driven world.

By combining the XGBoost and RFR methods, this research is expected to enhance the understanding of rain gauge performance in terms of accuracy, data quality, and measurement stability. The study was conducted in four districts: Banyumas, Banjarnegara, Wonosobo, Temanggung, and Pemalang. The goal is to evaluate data from both manual and automatic instruments and provide technical recommendations appropriate to the geographical characteristics and local needs.

2. RESEARCH METHOD

This study compares the performance of manual and automatic rain gauges using data-based quantitative methods. A literature review on rain measurement instruments and the use of XGBoost and RFR algorithms for meteorological data prediction was conducted before starting the study.

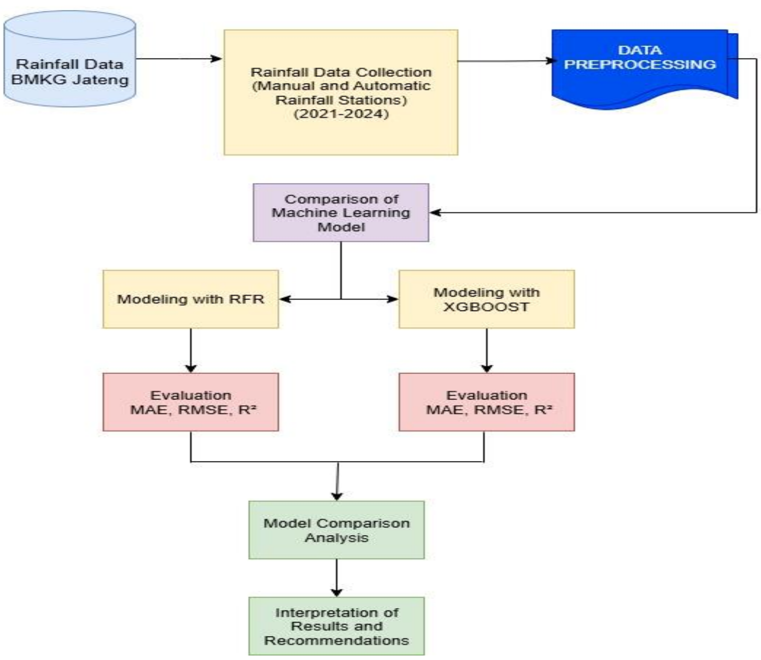

Figure 1. Research Method

2.1 Data Collection

The research locations in this study involve five regencies in the Central Java Mountains, namely Banyumas, Banjarnegara, Wonosobo, Temanggung, and Pemalang. These districts were chosen due to their complex topography and varying levels of rainfall. Data were collected from January 2021 to December 2024 from observation stations owned by the BMKG Central Java Climatology Station and the local Irrigation Office, which recorded daily rainfall using Automatic Rain Gauges (ARG), Automatic Weather Stations (AWS), and manual ombrometers. The diversity of locations and climatic conditions allows for a more comprehensive evaluation of the instrument's performance, both in light and extreme rain conditions, as well as at various elevations. In addition, the selection of these locations also considers the availability of historical data and the continuity of good record-keeping during the observation period.

2.2 Data Preprocessing

Data preprocessing is carried out to ensure that rainfall data from automatic measuring instruments (ARG, AWS) and manual (ombrometer, CSV) are ready to be used in machine learning modeling. The first step is to merge the two datasets, which consist of date columns in datetime format. This is done after eliminating input errors, such as changing the value "8888" to zero. Additionally, the time zone format was adjusted to ensure consistency between the two temporal data sources. Next, the daily datasets are merged, resulting in uniform and clean data ready for analysis using the XGBoost and Random Forest algorithms. This stage is in line with the preprocessing procedures in similar studies, which highlight the importance of transformation, normalization, and handling of missing/anomalous data to improve the accuracy of rainfall prediction using XGBoost and Random Forest (Sangaji & Sutabri, 2025).

2.3 Modeling

The modeling stage employs two primary algorithms: XGBoost and RFR. XGBoost is selected for its efficiency in handling large-scale and complex datasets (Darmawan et al., 2023) (Abdullah & Said, 2024), while RFR is chosen for its ability to produce stable predictive models that are resistant to overfitting (Islam et al., 2023). Model performance is evaluated using three key metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2) (Papacharalampous et al., 2023). These evaluation results are then used to assess the accuracy of each instrument and to provide recommendations for the most appropriate type of rain gauge, taking into account the geographical characteristics of the study area.

XGBoost Objective Function Formula

$$\mathcal{L}(\emptyset) = \sum_{i-1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{k=1}^{t} \Omega(fk)$$

Where $\mathcal{L}(\emptyset)$ represents the total objective function, which combines the training loss and the regularization term. The term $l(y_i, \hat{y}_i)$ denotes the loss function, such as the squared error commonly used in regression tasks, which measures the difference between the actual value y_i and the predicted value $\hat{y}_i$. The symbol $\hat{y}_i^{(t)}$ indicates the prediction of the model at the t^{th} iteration, reflecting the iterative nature of boosting algorithms. Meanwhile, $\Omega(fk)$ refers to the regularization function, which penalizes model complexity to prevent overfitting and to ensure generalization capability of the model.

RFR Function Formula

$$\hat{f}(x) = \frac{1}{N}\sum_{i=1}^{N} f\,i(x)$$

Where $\hat{f}(x)$ represents the final prediction result of the Random Forest for the input x, N denotes the total number of decision trees in the forest, ad $fi(x)$ is the prediction result produced by the i^{th} individual tree

2.4 Model Evaluation

Model testing is performed to evaluate the performance of each algorithm, using accuracy metrics such as MAE, RMSE, and R^2. In the final step, the results are analyzed and interpreted by comparing all outcomes to draw conclusions regarding the performance differences among the rain gauges. Based on this analysis, relevant recommendations are provided on the optimal use of the instruments within the research area (Tricha & Moussaid, 2024).

Mean Absolute Error

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

Where Y_i denotes the actual value of the i^{th} data point, $\hat{Y}_i$ represents the predicted value of i^{th} data point, and n refers to the total number of data samples.

Root Mean Squared Error

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

Where n represents the total number of data samples, yi is the actual value of the i^{th} data point, and $y\hat{}i$ denotes the predicted value of the i^{th} data point. These variables are commonly used in performance evaluation metrics to measure the accuracy of a predictive model by comparing the predicted values to the actual observed values.

R-Squared

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(\mathrm{y}_i - \mathrm{y\hat{}}_i)^2}{\sum_{i=1}^{n}(\mathrm{y}_i - \mathrm{y\bar{}})^2}$$

Where R^2 evaluates how well the model predicts the actual data compared to the mean of the observed values. Here, yi is the actual value of the i^{th} data point, $y\hat{}i$ is the predicted value of the i^{th} data point, and y^{-} represents the mean of all actual values. An R^2 value of 1 indicates a perfect prediction by the model, while a value of 0 means the model performs no better than a simple average.

2.5 Analysis and Interpretation of Results

The main output of this research is a predictive model that shows the relationship between rainfall data obtained from automatic and manual gauges. This model was developed using the XGBoost and RFR algorithms. Its performance was evaluated using three key metrics MAE, RMSE, and R^2. The reliability of the automatic instruments was assessed in comparison with manual gauges, which have long served as the standard for rainfall measurement. The final outcome of this research is a recommendation for the most suitable type of rain gauge to be used in the study area, based on the results of model analysis and evaluation. This recommendation is expected to improve the accuracy of rainfall data recording and support policies related to water resource management and disaster mitigation in Indonesia.

3. RESULTS AND DISCUSSION

3.1 Preprocessing and Exploration of Rainfall Data

This research begins with a preprocessing stage, where the date columns from both the automatic data (ARG/AWS in Excel format) and the manual data (ombrometer in CSV format) are converted to datetime format. To prevent any negative impact on the model, anomalous values such as "8888" in the manual dataset are replaced with zero. Time zone information is removed from the automatic data, and dates are used as the primary key to merge the two datasets. The result is a clean, standardized dataset ready for modeling using XGBoost and RFR. This stage follows best practices in environmental data preprocessing, as proper data handling significantly improves the accuracy of predictive models (Fadil Danu Rahman et al., 2024).

Table 1. Table dataset

Date	Automatic	Manual
2021-01-02	71.6	61.0
2021-01-03	68.8	22.0
2021-01-04	30.6	13.0
2021-01-05	24.6	29.0
2021-01-06	53.4	48.0

The historical data displayed in the dataset **Table 1** is daily rainfall data from two types of measuring instruments, namely automatic instruments (such as Automatic Rain Gauge/AWS) and manual instruments (ombrometer), which each record the amount of rainfall in millimeters (mm) on the same date. For example, on January 2, 2021, the automatic instrument recorded 71.6 mm while the manual instrument recorded 61.0 mm. This difference indicates a variation in the recording results between the instruments, which could be caused by differences in sensor sensitivity, data collection methods, or the environmental conditions around the instruments. This table shows how both instruments record rainfall at the same time and serves as the basis for regression model analysis. The goal is to determine how well the data from the manual instrument can be compared with the data from the automatic instrument. This has been the reference standard for weather observation systems all along. These different values then become the focus of evaluating the natural regression models XGBoost and Random Forest. Next, the correlation between rainfall data measured by automatic tools (ARG/AWS) and manual tools (ombrometer) can be seen in **Figure 2.**

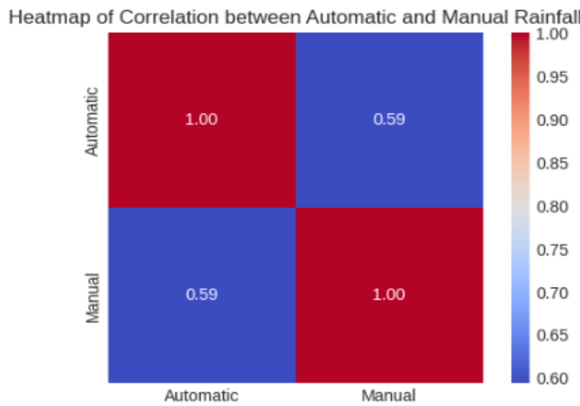

Figure 2. Correlation Between Automatic and Manual Rainfall

Figure 2 shows a Pearson correlation of 0.59 between manual and automatic rainfall measurements, indicating a moderate positive relationship. This means that, although both devices record similar rainfall patterns, there may be value differences caused by various device sensitivity factors. This correlation is sufficient to indicate that both datasets can be compared and used for modeling (Pratiwi & Nurhuda, 2021). To support these findings, the rainfall distribution from each instrument is displayed in the form of a histogram, as shown in **Figure 3**.

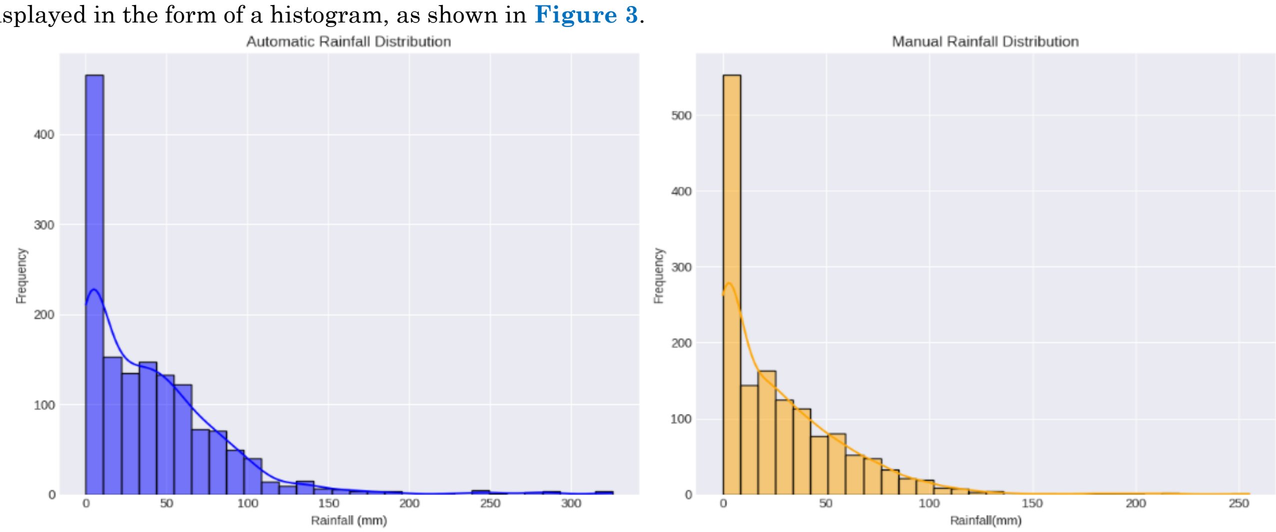

Figure 3. Automatic and manual rainfall distribution

Both distributions exhibit a similar pattern, with most daily rainfall occurring in the lower range (below 20 mm), and frequency decreasing as rainfall intensity increases. The left-skewed nature of the data reflects this trend. While the automatic instrument data displays a slightly wider distribution, the manual instrument data appears to be more concentrated in the lower rainfall range. This discrepancy may be attributed to differences in sensor sensitivity or recording methods between the two instruments. The density estimation curves accompanying each histogram, presented prior to the modeling phase, help clarify these distribution patterns and enhance the understanding of the fundamental characteristics of the data.

3.2 Feature Engineering

At the feature engineering stage, a multi-lag approach is used to identify temporal patterns in daily rainfall data. This technique aims to create input variables based on rainfall values from previous days by adding lag features from one to seven days (Lag_1 to Lag_7) (Kumar, Kedam, Sharma, Khedher, et al., 2023). Since rainfall patterns typically have a short-term memory effect, the addition of these lag features aims to reduce the temporal dependence of daily rainfall. For example, if high rainfall occurs today, it is highly likely that the rainfall values on the following days will also be affected. By using up to seven-day lags, the model can better identify dry periods or consecutive rain patterns. In order to identify short-term trends while reducing extreme daily fluctuations, the 3-day Moving Average (MA_3) and 7-day Moving Average (MA_7) features were added. By smoothing out extreme data fluctuations, the moving average enables the model to capture stable patterns rather than just responding to momentary anomalies (Jamal et al., 2025) (Hasnain, 2025). The feature engineering dataset was then cleaned of null values caused by data shifts and divided into two parts: training data (80%) and testing data (20%). This is done without shuffling, so the time sequence is preserved as a time series modeling.

3.3 Results and Model Evaluation

Modeling was conducted using two machine learning algorithms: Random Forest Regression (RFR) and XGBoost. Hyperparameter tuning was performed using GridSearchCV combined with time series cross-validation. The early stopping technique was applied to XGBoost to prevent overfitting, allowing the model to stop training if its accuracy does not improve after a set number of iterations (Rizky et al., 2024). The model's performance was evaluated using three key metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). The results of the performance evaluation are presented in **Table 2**.

Table 2. Evaluation of automatic and manual rainfall prediction models

Model	MAE	RMSE	R^2
Random Forest (Otomatis	16.6307	28.5286	0.4485
XGBoost (Otomatis)	17.3632	27.0282	0.5050
Random Forest (Manual)	16.5438	24.9370	0.2534
XGBoost (Manual)	16.5542	24.5059	0.2790

3.3.1 Evaluation of the RFR Model

The results of the RFR model training, as shown in **Table 2**, indicate that when using automatic data as input, the model yields an MAE of 16.63 mm, an RMSE of 28.53 mm, and an R^2 of 0.4485. This suggests that the model can explain approximately 44.85% of the data variation, demonstrating reasonably good predictive performance. In contrast, when using manual data, although the MAE (16.54 mm) and RMSE (24.94 mm) are slightly lower, the R^2 value decreases to 0.2534. This suggests that the manual data is less consistent in capturing patterns, possibly due to irregular recording practices or human error. Overall, automatic data provides more stable and accurate results for building predictive models, making it more suitable for the development of machine learning-based rainfall prediction systems. A visualization of the RFR model's prediction performance on historical data and future projections is presented in **Figure 4** and **Figure 5**.

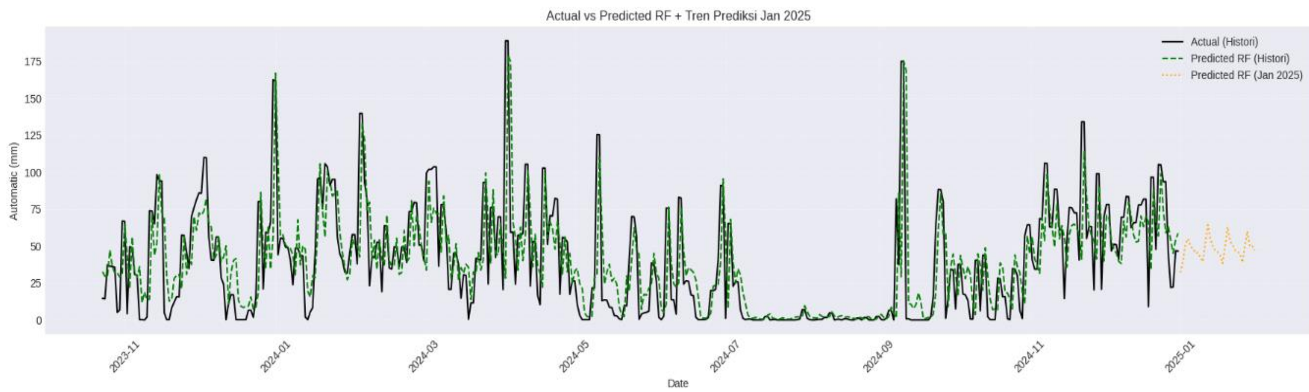

Figure 4. Visualization of RFR Predictions Against Historical Data on Automatic Rainfall

Figure 4 shows a comparison between the predictions of the Random Forest (RF) model for the historical period and the January 2025 projection with actual rainfall data from automatic instruments (ARG/AWS). The black line shows the actual data, while the dashed green line shows the historical RF predictions, which generally follow the seasonal rainfall pattern quite well. On the right side of the graph, the dashed orange line shows the RF projection for January 2025, with a stable pattern and rainfall of 30–70 mm per day.

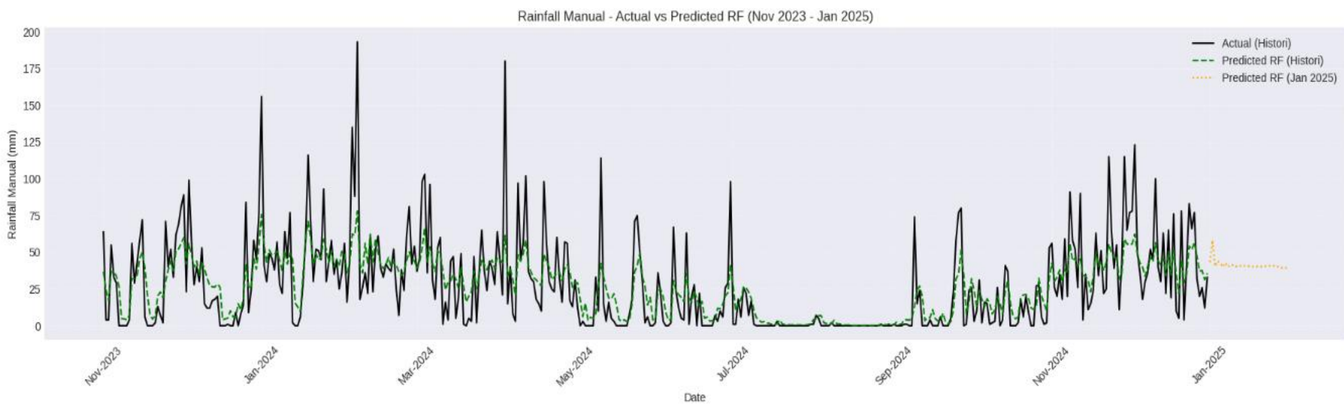

Figure 5. Visualization of RFR Predictions Against Historical Data on Manual Rainfall

Conversely, the RFR prediction results using manual data as the target are shown in Figure 6. Although the overall trend can still be improved, the predictions for extreme peaks such as those in February and May 2024 tend to differ significantly. The prediction curve appears smoother, indicating that the model tends to reduce the large variations typically observed in manual recordings. This is consistent with the evaluation metrics, which show a lower R^2 value when manual data is used. Both statistically and visually, the results demonstrate that automatic data provides more consistent and accurate prediction outcomes compared to manual data. The RFR model is capable of capturing seasonal rainfall patterns well; however, the best performance is achieved when the data source exhibits the stability and temporal consistency characteristic of automatic instruments. Therefore, automatic data is supported as a more reliable foundation for developing a machine learning-based rainfall prediction system in the Central Java Central Mountains region. Overall, it can be concluded that using data from automatic instruments enables the RFR model to generate more stable and realistic predictions compared to using data from manual instruments. This finding reinforces the importance of utilizing automated data in machine learning-based weather prediction systems, especially in regions with complex climate variability such as the Central Java Central Mountains.

3.3.2 Evaluation of the XGBoost Model

The XGBoost model training results, as shown in Table 2, produced an MAE of 17.3632 mm, an RMSE of 27.0282 mm, and an R^2 of 0.5050 when using automated data as input features. This indicates that although the average prediction error is slightly higher compared to the Random Forest model, the higher R^2 value around 0.51 suggests that the XGBoost model is better at capturing data variation and understanding the dynamic patterns of rainfall. With an R^2 value above 0.5, the model successfully explains more than 50% of the variation in the actual data, making it statistically the most robust model among those tested. When trained using manual data, the XGBoost model achieved an MAE of 16.5542 mm, an RMSE of 24.5059 mm, and an R^2 of 0.2790. Although the MAE and RMSE values are nearly identical to those of the Random Forest model, the increase in R^2 (approximately 2.56% higher) indicates that XGBoost offers better regularization and weighting capabilities. The performance of the XGBoost model in relation to historical data and future projections is visualized in **Figure 6** and **Figure 7**.

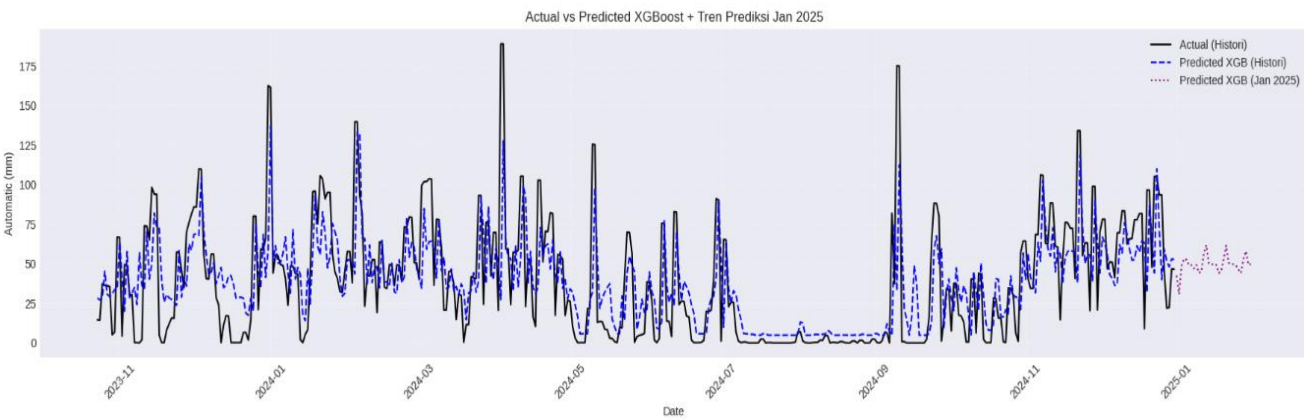

Figure 6. Visualization of XGBoost Predictions Against Historical Data on Automatic Rainfall

Figure 6 presents a comparison between the historical prediction results of the XGBoost model, its projected prediction for January 2025, and the actual rainfall data. The vertical axis represents daily rainfall intensity in millimeters (mm), while the horizontal axis covers the period from November 2023 to the end of January 2025. The black line illustrates actual data recorded by the Automatic Rain Gauge (ARG/AWS), showing a fluctuating pattern due to seasonal dynamics. The dashed blue line depicts the XGBoost model's predictions during the historical period, which closely follow the observed rainfall trends and peaks, particularly during January, March, and late 2024. Meanwhile, the dashed purple line on the right side of the graph represents the model's projections for January 2025. This projection displays a relatively stable and realistic pattern, with rainfall intensities ranging from moderate to high (35–60 mm/day), aligning with early rainy season trends from the previous year. The model's ability to capture seasonal dynamics and project short-term trends accurately is evident in this visualization. It further supports the quantitative evaluation, where the XGBoost model achieved the highest R^2 value (0.5050) among all tested models, confirming its suitability for data-driven rainfall prediction systems in regions with high climate variability, such as the Central Java Central Mountains.

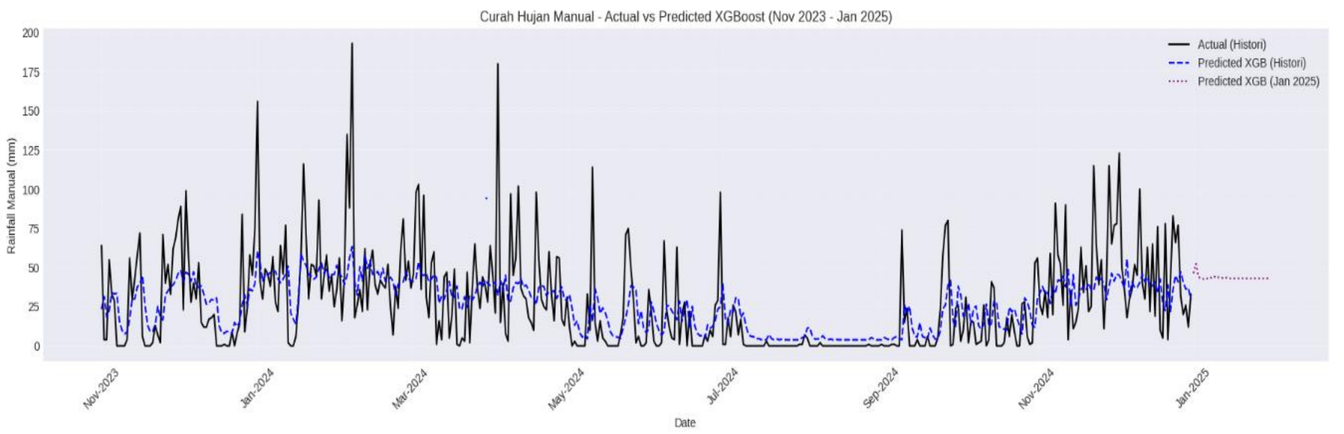

Figure 7. Visualization of XGBoost Predictions Against Historical Data on Manual Rainfall

The XGBoost prediction results using manual rainfall data, as shown in Figure 8, demonstrate that although the prediction curve still shows a correlated pattern with the actual data, extreme variations in the actual data remain difficult to fully capture. The model tends to produce smoother curves and dampen significant spikes, particularly in February and May 2024. This supports the finding that inconsistencies in the manual data continue to limit the model's performance, even though XGBoost performs better than Random Forest when applied to manual data. Overall, the evaluation results indicate that the XGBoost model offers the best balance between accuracy and the ability to capture seasonal patterns, especially when using automated data. Its highest R^2 value among all tested models confirms that XGBoost is the most suitable model for rainfall prediction based on automatic data in the Central Java Central Mountains region. This conclusion is further supported by the visualizations, which show that the model effectively follows major rainfall trends and maintains stability in future projections. These findings highlight the greater stability and reliability of automatic instrument data for developing machine learning-based rainfall prediction models using the XGBoost algorithm.

4. CONCLUSION

The results of this study confirm that the XGBoost and Random Forest Regression (RFR) algorithms are effective for building daily rainfall prediction models, particularly when using data from automatic instruments such as Automatic Rain Gauges (ARG) and Automatic Weather Stations (AWS). The XGBoost model demonstrated the best performance, with an MAE of 17.3632 mm, an RMSE of 27.0282 mm, and an R^2 of 0.5050, indicating that the model could accurately explain more than 50% of the variation in actual rainfall data. Meanwhile, the RFR model using automatic data achieved an MAE of 16.6307 mm, an RMSE of 28.5286 mm, and an R^2 of 0.4485. Although slightly lower than XGBoost in terms of determination, it still delivered stable results. In contrast, when trained using data from manual instruments, both models performed noticeably worse. The RFR model achieved only an R^2 of 0.2534, and XGBoost 0.2790, highlighting the greater susceptibility of manual data to recording inconsistencies. Based on this evaluation, it can be concluded that combining automated data with the XGBoost algorithm is the most suitable approach for developing a reliable rainfall prediction system, especially in regions with complex topography such as the Central Java Mountains. This approach also contributes significantly to the advancement of more accurate and responsive weather monitoring and disaster mitigation systems.

REFERENCES

Abdullah, M., & Said, S. (2024). Performance Evaluation of Machine Learning Regression Models for Rainfall Prediction. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, *December*. https://doi.org/10.1007/s40996-024-01691-4

Darmawan, H., Yuliana, M., & Hadi, M. Z. S. (2023). GRU and XGBoost Performance with Hyperparameter Tuning Using GridSearchCV and Bayesian Optimization on an IoT-Based Weather Prediction System. *International Journal on Advanced Science, Engineering and Information Technology*, *13*(3), 848–859. https://doi.org/10.18517/ijaseit.13.3.18377

Droździoł, R., & Absalon, D. (2023). Evaluation of Selected Amateur Rain Gauges with Hellmann Rain Gauge Measurements. *Climate*, *11*(5). https://doi.org/10.3390/cli11050107

Dunkerley, D. (2025). *Limited Time Resolution of Event Data Loggers Can Bias Intensity Measurements from Tipping-Bucket Rain Gauges Limited Time Resolution of Event Data Loggers Can Bias Intensity Measurements from Tipping-Bucket Rain Gauges*. 0–11. https://doi.org/10.20944/preprints202407.0322.v2

Fadil Danu Rahman, Mulki, M. I. Z., & Taryana, A. (2024). Clustering Dan Klasifikasi Data Cuaca Cilacap Dengan Menggunakan Metode K-Means Dan Random Forest. *Jurnal SINTA: Sistem Informasi Dan Teknologi Komputasi*, *1*(2), 90–97. https://doi.org/10.61124/sinta.v1i2.15

Hasnain, K. N. (2025). *Integrating Machine Learning for Real - Time Energy Load Forecasting in US Smart Grids : A Multi - Model Comparative Approach Abstract :* 1–19.

Huber, F., Yushchenko, A., Stratmann, B., & Steinhage, V. (2022). Extreme Gradient Boosting for yield estimation compared with Deep Learning approaches. *Computers and Electronics in Agriculture*, *202*. https://doi.org/10.1016/j.compag.2022.107346

Islam, K. I., Elias, E., Carroll, K. C., & Brown, C. (2023). Exploring Random Forest Machine Learning and Remote Sensing Data for Streamflow Prediction: An Alternative Approach to a Process-Based Hydrologic Modeling in a Snowmelt-Driven Watershed. *Remote Sensing*, *15*(16). https://doi.org/10.3390/rs15163999

Jamal, N. A., Mohamed Noor, N., Jafri, I. A. M., & Wibowo, G. J. (2025). *Assessment of Moving Average (MA) Method for Rainfall Prediction in Yogyakarta, Indonesia*. 5. https://doi.org/10.3390/eesp2025033005

Kumar, V., Kedam, N., Sharma, K. V., Khedher, K. M., & Alluqmani, A. E. (2023). A Comparison of Machine Learning Models for Predicting Rainfall in Urban Metropolitan Cities. *Sustainability (Switzerland)*, *15*(18). https://doi.org/10.3390/su151813724

Kumar, V., Kedam, N., Sharma, K. V., Mehta, D. J., & Caloiero, T. (2023). Advanced Machine Learning Techniques to Improve Hydrological Prediction: A Comparative Analysis of Streamflow Prediction Models. *Water (Switzerland)*, *15*(14). https://doi.org/10.3390/w15142572

Li, H., Zhang, Y., Lei, H., & Hao, X. (2023). Machine Learning-Based Bias Correction of Precipitation Measurements at High Altitude. *Remote Sensing*, *15*(8). https://doi.org/10.3390/rs15082180

Liyew, C. M., & Melese, H. A. (2021). Machine learning techniques to predict daily rainfall amount. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00545-4

Lukasová, V., Varšová, S., Onderka, M., Bilčík, D., Buchholcerová, A., & Nejedlík, P. (2024). Continuity of long-term climate data series after the transition from manual to automatic weather station. *Contributions to Geophysics and Geodesy*, *54*(3), 251–266. https://doi.org/10.31577/congeo.2024.54.3.2

Nicola, D. S., D, E., E, P., & M, B. (2024). Comparative analysis of conventional and machine learning techniques for rainfall threshold evaluation under complex geological conditions. *Landslides*, *21*(12), 2893–2911. https://doi.org/10.1007/s10346-024-02336-3

No, V., Syahreza, A., Ningrum, N. K., & Syahrazy, M. A. (2024). *Edumatic : Jurnal Pendidikan Informatika Perbandingan Kinerja Model Prediksi Cuaca : Random Forest , Support Vector Regression , dan XGBoost*. *8*(2), 526–534. https://doi.org/10.29408/edumatic.v8i2.27640

Papacharalampous, G., Tyralis, H., Doulamis, A., & Doulamis, N. (2023). Comparison of Tree-Based Ensemble Algorithms for Merging Satellite and Earth-Observed Precipitation Data at the Daily Time Scale. *Hydrology*, *10*(2). https://doi.org/10.3390/hydrology10020050

Pratiwi, A., & Nurhuda, W. (2021). Analisis Perbandingan Data Pengukuran Curah Hujan Metode Otomatis (Automatic Rain Gauge) dan Metode Observasi. *Prosiding Seminar Nasional Fisika*, 116–128. https://fisika.fmipa.unesa.ac.id/proceedings/index.php/snf/article/view/162%0Ahttps://fisika.fmipa.unesa.ac.id/proceedings/index.php/snf/article/download/162/156

Putra, M., Rosid, M. S., & Handoko, D. (2024). A Review of Rainfall Estimation in Indonesia: Data Sources, Techniques, and Methods. *Signals*, *5*(3), 542–561. https://doi.org/10.3390/signals5030030

Quayesam, D., Akubire, J., & Darkwah, O. (2024). *A Comparison of Machine Learning Algorithms for Predicting Sea Surface Temperature in the Great Barrier Reef Region*. http://arxiv.org/abs/2411.15202

Rahman, A. U., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., Khan, M. A., & Mosavi, A. (2022). Rainfall Prediction System Using Machine Learning Fusion for Smart Cities. *Sensors*, *22*(9). https://doi.org/10.3390/s22093504

Rizky, M. H., Faisal, M. R., Budiman, I., Kartini, D., & Abadi, F. (2024). Effect of Hyperparameter Tuning Using Random Search on Tree-Based Classification Algorithm for Software Defect Prediction. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *18*(1), 95. https://doi.org/10.22146/ijccs.90437

Rowe, F., Mahony, M., & Tao, S. (2022). *Assessing Machine Learning Algorithms for Near-Real Time Bus Ridership Prediction During Extreme Weather*. 1–35. http://arxiv.org/abs/2204.09792

Sangaji, D., & Sutabri, T. (2025). Analisis XGBoost dan Random Forest untuk Prediksi Curah Hujan dalam Mendukung Mitigasi Karhutla. *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence)*, *5*(1), 13–18. https://doi.org/10.55382/jurnalpustakaai.v5i1.905

Segovia-Cardozo, D. A., Bernal-Basurco, C., & Rodríguez-Sinobas, L. (2023). Tipping Bucket Rain Gauges in Hydrological Research: Summary on Measurement Uncertainties, Calibration, and Error Reduction Strategies. *Sensors*, *23*(12). https://doi.org/10.3390/s23125385

Shankar, B., Soni, A., Chandrasekar, H., Raghavan, S., & Shrivastava, M. (2019). First Observations on the Trap-Induced Avalanche Instability and Safe Operating Area Concerns in AlGaN/GaN HEMTs. *IEEE Transactions on Electron Devices*, *66*(8), 3433–3440. https://doi.org/10.1109/TED.2019.2919491

Tricha, A., & Moussaid, L. (2024). *Evaluating machine learning models for precipitation prediction in Casablanca City*. *June*, 1325–1332. https://doi.org/10.11591/ijeecs.v35.i2.pp1325-1332

Urban, G., & Strug, K. (2021). Evaluation of precipitation measurements obtained from different types of rain gauges. *Meteorologische Zeitschrift*, *30*(5), 445–463. https://doi.org/10.1127/metz/2021/1084

Yan, X., & Duan, G. (2022). The Real-Time Prediction of Product Quality Based on the Equipment Parameters in a Smart Factory. *Processes*, *10*(5). https://doi.org/10.3390/pr10050967

Yasper, A., Handoko, D., Putra, M., Aliwarga, H. K., & Rosid, M. S. R. (2023). Hyperparameters Optimization in XGBoost Model for Rainfall Estimation: A Case Study in Pontianak City. *Jurnal Penelitian Pendidikan IPA*, *9*(9), 7113–7121. https://doi.org/10.29303/jppipa.v9i9.3890