Journal of Computation Physics and Earth Science
ISSN: 2776-2521 (online)
Volume 5, Number 1, April 2025, Page 21 - 30
https://journal.physan.org/index.php/jocpes/index

Evaluation of the XGBoost Model for Rainfall Prediction and Classification Using BMKG Data and OpenWeather API

M. Saori Isjayan Syah[1], Nardi[2], Agustina Rachmawardani[3]
[1,2,3]Undergraduate Program in Applied of Instrumentation Meteorology, Climatology Geophysics (STMKG)

Article Info

Article history:

Received March 2, 2025
Revised March 7, 2025
Accepted March 8, 2025

Keywords:

Rainfall Prediction,
XGBoost,
Rainfall Classification,
OpenWeather API,
BMKG.

ABSTRACT

Indonesia, situated between two continents and two oceans, experiences significant climate variability, with rainfall patterns shaped by geographical and topographical factors, as well as phenomena like the El Niño Southern Oscillation (ENSO). Accurate rainfall forecasting plays a critical role in disaster mitigation, agricultural planning, and water resource management. This study focuses on developing a rainfall prediction and classification model using the Extreme Gradient Boosting (XGBoost) algorithm. The model leverages historical rainfall data from the Indonesian Meteorological, Climatological, and Geophysical Agency (BMKG) and real-time data from the OpenWeather API. The output includes rainfall trend graphs and classification of rainfall intensity into categories such as light, moderate, or heavy. Model performance is assessed through metrics like accuracy, precision, RMS (Root Mean Square), and RMSE (Root Mean Square Error). This research highlights the integration of historical and real-time data for weather forecasting and demonstrates the application of advanced machine learning techniques like XGBoost to build robust and precise prediction models. The findings are expected to offer practical insights for disaster risk reduction, agricultural strategy planning, and effective water resource management.

Corresponden Author:

M. Saori Isjayan Syah,
Undergraduate Program in Applied of Instrumentation Meteorology, Climatology Geophysics (STMKG)
Tangerang City, Banten, Indonesia
Email: muhammadsaorii@gmail.com

1. INTRODUCTION

Indonesia is situated between the Indian and Pacific Oceans, as well as between two continents—Asia and Australia. This geographical position causes frequent climate variations influenced by multiple factors. Additionally, phenomena like the El Niño Southern Oscillation (ENSO) contribute to the environmental complexity. Rainfall, as a significant component of the climate, is highly intricate. Its characteristics vary across regions, influenced by factors such as geography, topography, and monsoons. The structure and orientation of the islands are yet to be fully accounted for, leading to uneven rainfall distribution across different areas. Moreover, ENSO affects nearly 70% of rainfall variability in Indonesia [1].

Rainfall is a significant and impactful phenomenon within the climate system, directly affecting ecosystems, water resources, management practices, and agriculture. For many people, rainfall serves as the primary source of water, playing a vital role in their daily lives and livelihoods [2][3]. In Indonesia, rainfall is typically measured using two approaches: direct observation with rain gauges and indirect estimation through remote sensing [4]. Developing reliable rainfall forecasting systems remains a significant challenge for researchers across various fields, including weather data analysis, environmental machine learning, operational hydrology, and statistical forecasting methods [5][6].

Rainfall prediction is critically important because it is the variable most strongly correlated with natural disasters such as landslides, floods, mass movements, and avalanches. It involves various techniques such as statistics, data mining, modeling, artificial intelligence, and machine learning. Accurate and real-time rainfall prediction has been a significant challenge for decades due to its nonlinear and stochastic nature. Therefore, having an appropriate approach to predict rainfall allows for preventive and mitigation actions for these naturally occurring phenomena [2][7][8] .

Reliable climate information plays a vital role in agricultural decision-making, as climate and rainfall data are essential for identifying the most suitable crops for specific agricultural regions. The Oldeman classification is the most suitable system for agricultural zoning in Indonesia. It is commonly applied to assess climate suitability for agriculture because it relies on the consecutive number of wet and dry months over a minimum period of 10 years [9]. The Oldeman classification, similar to the Schmidt-Ferguson system, is based on rainfall intensity, linking wet and dry months to agricultural suitability in specific regions. The classification criteria are as follows:

- Wet month: Average monthly rainfall exceeds 200 mm.
- Humid month: Average monthly rainfall ranges between 100 and 200 mm.
- Dry month: Average monthly rainfall is below 100 mm.

This system is designed based on plant water requirements, particularly for paddy crops, where rainfall above 200 mm is ideal. In contrast, rainfall around 100 mm is more suitable for palawija (secondary) crops. [10]

OpenWeatherMap is an online platform that offers weather information, including real-time data, forecasts, and historical records, to developers of web and mobile applications. It gathers data from various sources such as meteorological services, airport weather stations, radar systems, and other official weather monitoring stations [11]. Weather forecast services, like the OpenWeatherMap API, offer extensive data through an API, including real-time analysis and predictions for upcoming days based on continuously updated forecasts [12].

One clear indicator of technological advancement is the emergence of Machine Learning (ML)[13]. Extreme Gradient Boosting (XGBoost) is an ensemble learning technique. In certain cases, the output of a single machine learning model, such as J48, may be insufficient. While J48 performs well and has been applied to various problems like wildfire prediction and rainfall modeling, its limitations can necessitate more robust approaches [14][15][16]. XGBoost extends this approach using gradient boosting with multi-threading for faster and more efficient computation. It includes sparse-aware processing for handling missing data, supports parallel tree construction, and enables incremental training. These features make XGBoost a powerful tool for classification, regression, and predictive modeling, especially with structured or tabular data [17][18].

Model evaluation is commonly performed using metrics such as Root Mean Square Error (RMSE) to assess the accuracy of estimations. According to research by Adiyo R. (2014), estimation methods such as Maximum Likelihood Robust (MLR) and Weighted Least Squares Mean and Variance adjusted (WLSMV) yield smaller RMSE values, indicating higher accuracy in test score estimation. Additionally, the number of categories and the range of thresholds also affect RMSE values, where a greater number of categories and a balanced range of thresholds tend to result in lower RMSE values, indicating improved model accuracy [19][20].

This study aims to develop a prediction and classification model for rainfall using the Extreme Gradient Boosting (XGBoost) algorithm, leveraging historical rainfall data from BMKG as training data and real-time data from OpenWeather API as input for predictions. The model will output rainfall trend graphs and classify rainfall intensity into categories such as light, moderate, or heavy. The research also seeks to evaluate the model's performance using metrics like accuracy, precision, RMS (Root Mean Square), and RMSE (Root Mean Square Error). The significance of this study lies in its contribution to integrating both historical and real-time data for weather forecasting, applying modern machine learning technologies like XGBoost, and providing a comprehensive model evaluation. Moreover, this research is expected to offer practical benefits for disaster mitigation, agricultural planning, and water resource management by developing a more accurate and reliable weather prediction model

2. RESEARCH METHOD

In this study, we used data from two sources :

Historical Rainfall Data from BMKG

Rainfall data for this study was sourced from the BMKG Maritime Station in Bitung, North Sulawesi, covering the period between January 1, 2019, and January 1, 2024. The dataset includes daily rainfall measurements that adhere to BMKG’s standardized protocols. This data serves as training input for building the prediction model.

Real-Time Data from OpenWeather API

Real-time weather data was retrieved from the OpenWeather API, offering up-to-date information on various weather parameters, including rainfall. This dataset is used to validate the model on data outside the training set, ensuring its predictive capability with real-time input.

Research Approach

Historical rainfall data was obtained from the BMKG Maritime Station in Bitung through official channels. This data covers multiple years and includes daily rainfall measurements. It is essential for training the prediction model as it offers a solid historical framework of weather patterns, helping the model identify long-term trends and anomalies effectively.

Real-time weather data was collected using API requests managed by Python scripts. These scripts were designed to retrieve data at regular intervals, ensuring the continuous collection of the most recent weather parameters, such as rainfall. This real-time data is crucial for model validation, as it assesses the model's ability to make predictions based on actual, current conditions. Moreover, the automated process of data retrieval enhances efficiency and reduces the need for manual intervention.

Preprocessing Data

- Data Collection

  The historical rainfall data was obtained from a cleaned Excel file, resulting in pre-processed data. Real-time weather data was collected using the OpenWeather API for predicting future rainfall.

- Data Preprocessing:

  The historical data was processed by adding additional features such as Year, Month, and Day. Data from the API was converted into a DataFrame, with rainfall values aggregated by date. Rainfall was classified into categories: Dry, Moderate, and Heavy.

- Model Training and Prediction:

  A trained XGBoost model was employed to predict rainfall based on features like Month, Year, and Day.

- Model Evaluation:

  The model's performance was evaluated using metrics such as MSE (Mean Squared Error), RMSE (Root Mean Square Error), precision, and accuracy to assess the quality of its predictions.

- Visualization and Presentation:

  Graphs and tables of the prediction results are presented within a web application to simplify the interpretation of the results.

XGBoost Model Development

- Data Splitting and Model Training

  The dataset was divided into two subsets: a training set and a testing set, with an 80% allocation for training and 20% for testing using train_test_split. The XGBoost model was employed for regression purposes. XGBoost is a boosting model that builds decision trees incrementally, making it especially efficient for handling large datasets and producing more accurate predictions, particularly with data that includes many nonlinear features and complex interactions.

- Hyperparameter Tuning

  GridSearchCV was used for hyperparameter tuning, evaluating different combinations of parameters to identify the best-performing model based on its performance on the training data. The parameters tested include:

  - n_estimators: The number of trees used.
  - max_depth: The maximum depth for each tree.
  - learning_rate: The rate at which the model updates its weights.
  - subsample: The proportion of the data used for training each tree.
  - colsample_bytree: The proportion of features selected for each tree.

Model Evaluation

The performance of the model was assessed using several evaluation metrics, including:

- Root Mean Square Error (RMSE): Measures the overall prediction error.
- RMS (Root Mean Square): Evaluates the level of deviation between predictions and actual data.
- Accuracy: Indicates the percentage of correct predictions for classification.
- Precision: Measures the accuracy of predictions for specific classes (e.g., light or heavy rainfall).

Rainfall Classification Using the Oldeman Method

Rainfall was classified into categories of light, moderate, and heavy using the Oldeman method. The results of this classification were analyzed to identify patterns in the distribution of rainfall over specific time periods.

Visualization and Implementation

- Rainfall Trend Graphs: Trend graphs were created using Python visualization libraries like Matplotlib and Seaborn to display rainfall patterns over the last five years.
- Localhost Application: The trained model was integrated into a Flask-based application, allowing for the visualization of both predictions and classifications of rainfall directly.

Tools and Technologies

This research employs the following tools and technologies:

a. Programming Language: Python 3.x for data processing, model development, and application implementation.

b. Libraries and Frameworks:
   - XGBoost: For developing predictive and classification models.
   - Pandas and NumPy: For data preprocessing.
   - Matplotlib and Seaborn: For data visualization and displaying prediction results.
   - Scikit-learn: For metric evaluation and dataset splitting.
   - Flask: For creating a simple localhost interface.
   - API: OpenWeather API for real-time data retrieval.
   - Hardware: A laptop or server with specifications suitable for intensive computation during model training.

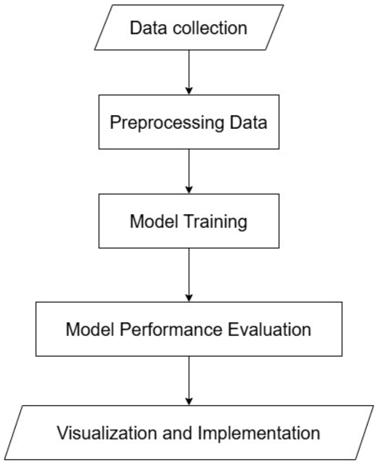

Figure 1. Flowchart

3. RESULT AND DISCUSSION

3.1 Rainfall Prediction Result with the XGBoost Model

The testing results of the XGBoost model demonstrate a good performance in predicting rainfall based on the historical data from BMKG. The model was evaluated using metrics like Mean Square Error (MSE), Root Mean Square Error (RMSE), Precision, and Accuracy. The evaluation results are as follows:

- Mean Square Error (MSE): 112.728 mm²
- Root Mean Square Error (RMSE): 10.6173 mm
- Precision: 0.64
- Accuracy: 0.033

The model's predictions on the historical data show rainfall patterns that are quite close to the actual values, although there are some minor variations in data with high fluctuations. The prediction graph for the historical data is presented in Figure 1.

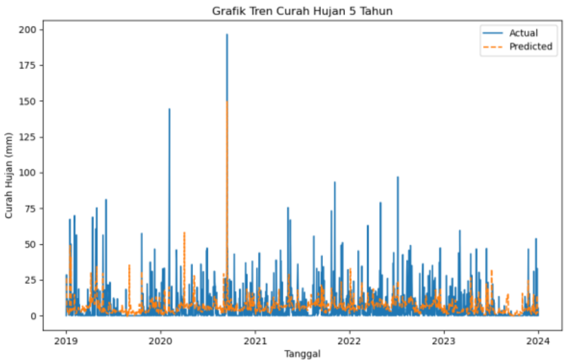

Figure 2. *Rainfall prediction graph based on historical BMKG data*

3.2 Rainfall Classification Based on Historical BMKG Data

Rainfall classification was done using the Oldeman criteria, which divides rainfall into dry, moderate, and heavy categories. Based on the historical data:

- Percentage of dry months: 15.28 %
- Percentage of humid month: 30.56%
- Percentage of wet months: 38.89%

Table 1 shows the annual rainfall classification linked with the number of dry, humid, and wet months for each year in the historical data.

Table 1. Rainfall Classification Based on Historical Data

Year	Total Rainfall (mm)	Classification	Wet Month	Humid month	Dry month
2019	2022.016	Klimat Tropis Basah	4	4	4
2020	2524.105	Klimat Tropis Sangat Basah	6	3	3
2021	2539.676	Klimat Tropis Sangat Basah	7	5	0
2022	2889.014	Klimat Tropis Sangat Basah	8	4	0
2023	1672.199	Klimat Tropis Basah	3	6	3
2024	0.1	Klimat Tropis Kering	0	0	1

3.3 Rainfall Prediction Based on OpenWeather API Data

The predictions of rainfall using real-time data from the OpenWeather API show patterns that match historical trends. These predictions are made for several days ahead, and the results are visualized in Figure 2 and Table 2

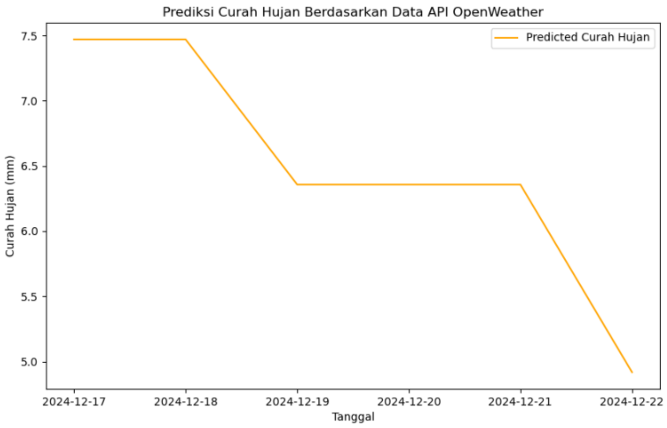

Figure 3. *Rainfall predictions based on OpenWeather API data*

This prediction shows that the model is able to provide fairly accurate rainfall estimates even though it uses data from external sources.

Table 2. Rainfall predictions based on OpenWeather API data.

Date	Rainfall Prediction (mm)
2024-12-17 00:00:00	7.4703521728515625
2024-12-18 00:00:00	7.4703521728515625
2024-12-19 00:00:00	6.357199192047119
2024-12-20 00:00:00	6.357199192047119
2024-12-21 00:00:00	6.357199192047119
2024-12-22 00:00:00	4.918496608734131
2024-12-17 00:00:00	7.4703521728515625

3.4 Rainfall Prediction for the Next 12 Months

The model was also used to predict rainfall for the next 12 months based on synthetic features generated from date data. The prediction graph for this period is shown in Figure 3 and Table 3

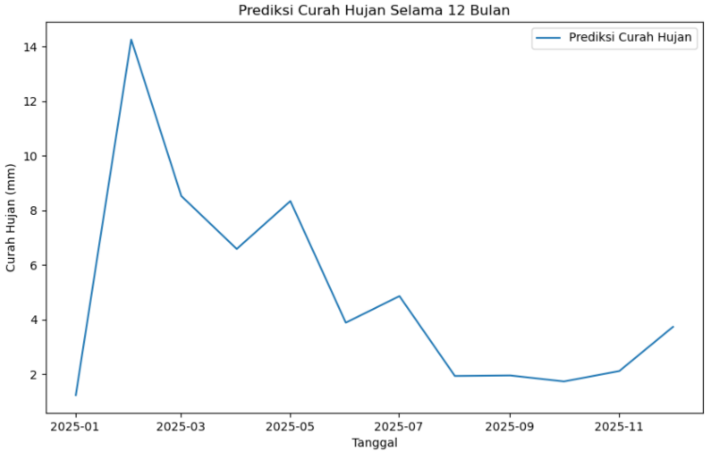

Figure 4. Rainfall predictions for the next 12 months

The prediction results display a seasonal pattern consistent with the characteristics of Indonesia's tropical climate, where rainfall tends to be higher during the rainy season. These predictions can serve as a reference for disaster mitigation and agricultural planning.

Table 3. Rainfall predictions for the next 12 months

Date	Rainfall Prediction (mm)
2025-01-01 08:45:03.968780	1.232616901397705
2025-02-01 08:45:03.968780	14.256916999816895
2025-03-01 08:45:03.968780	8.527671813964844
2025-04-01 08:45:03.968780	6.5872650146484375
2025-05-01 08:45:03.968780	8.341472625732422
2025-06-01 08:45:03.968780	3.886962413787842
2025-07-01 08:45:03.968780	4.860436916351318
2025-08-01 08:45:03.968780	1.9351284503936768
2025-09-01 08:45:03.968780	1.953909993171692
2025-10-01 08:45:03.968780	1.7334764003753662
2025-11-01 08:45:03.968780	2.116471767425537
2025-12-01 08:45:03.968780	3.7357137203216553

The Rainfall Prediction and Classification Dashboard presents evaluation results and predictions using the XGBoost model. On the main page, evaluation metrics are displayed such as MSE of 112.728, RMSE of 10.617, Precision of 0.6409, and Accuracy of 0.0333. These values indicate the model's performance in predicting rainfall based on both historical and real-time data. The website navigation includes a Trends Graph page showing rainfall pattern changes, Oldeman Classification for annual rainfall classification, and API Predictions which present real-time data from OpenWeather API in the form of tables and graphs. Additionally, the 12-Month Prediction page displays rainfall trends for the upcoming year. Visualizations are created using Matplotlib, while data processing is done with Pandas. This dashboard is designed to provide accurate and practical information supporting disaster mitigation, agricultural planning, and water resource management through the integration of machine learning and real-time data technologies.

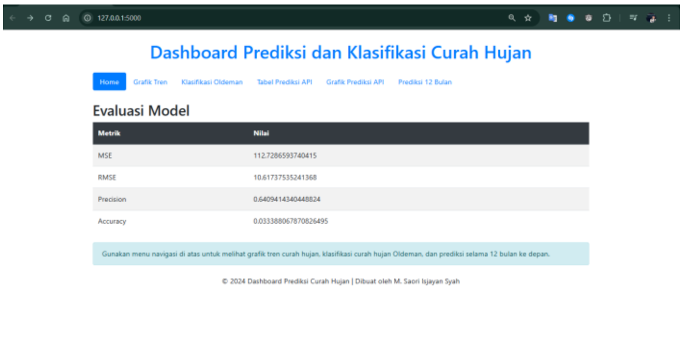

Figure 5. Model Visualization

The evaluation results show that the XGBoost model delivers strong performance in predicting rainfall using historical data. With an RMSE and accuracy, the model effectively reduces prediction errors. These results reflect the model's ability to produce predictions that closely align with actual data, demonstrating XGBoost's effectiveness in managing complex and non-linear rainfall datasets.

Several key factors that contributed to the success of the XGBoost model in this study include:

a. Capability to Process Tabular Data: XGBoost is specifically designed to handle structured tabular data, such as the BMKG dataset, making it ideal for modeling weather parameters like rainfall.
b. Optimized Gradient Boosting: XGBoost utilizes a boosting technique that iteratively corrects errors from previous iterations to progressively enhance model performance and improve prediction accuracy.
c. Automatic Handling of Missing Values: XGBoost can effectively manage datasets with missing values, a common issue in weather data. Enhanced Feature Engineering: Incorporating additional features such as year, month, and day into the model helps uncover seasonal trends and patterns in rainfall data

The XGBoost model's prediction results show a rainfall trend that aligns with the seasonal patterns observed in the study area (Bitung, North Sulawesi). This is evident in Figures 1 and 3, where rainfall increases during the rainy season and decreases during the dry season. The model effectively captured these seasonal variations, supported by the historical data spanning the 2019–2024 period used for training. Nevertheless, there are slight deviations between the predicted and actual values, especially during periods of significant fluctuation. These discrepancies may be influenced by several factors:

a) ENSO Variability (El Niño–Southern Oscillation): The ENSO phenomenon contributes to approximately 70% of rainfall variability in Indonesia. Consequently, in years with El Niño or La Niña events, prediction accuracy may decline.
b) Topography and Local Factors: The XGBoost model does not explicitly account for local factors such as topography and elevation, which can influence rainfall distribution in specific areas.
c) Input Data Limitations: Data obtained from the OpenWeather API may have different spatial and temporal resolutions compared to BMKG data, which can affect the quality of the predictions

Rainfall classification based on Oldeman’s criteria provides additional insight into the climatic conditions of the study area. For example:

- Years with a dominance of dry months indicate potential drought conditions that can impact the agricultural sector and water resources.
- Years with a dominance of wet months indicate potential risks of flooding and landslides.

The classification results presented in Table 1 show significant annual variations in the number of wet, humid, and dry months. This information can be used for agricultural planning and disaster mitigation. Additionally, these classification trends can help in determining more climate-adaptive planting patterns.

Integrating real-time data from the OpenWeather API highlights the flexibility of the XGBoost model for short-term predictions. The prediction results shown in Figure 3.2 demonstrate that the model can provide reasonably accurate rainfall estimates even with data from external sources. However, it’s important to consider the following challenges:

1) Data Quality: The data from the OpenWeather API depends on global modeling and spatial interpolation, which may be less precise compared to direct observations from BMKG stations. Temporal Resolution
2) Discrepancy: Real-time data from the API may have a higher temporal resolution (e.g., hourly), while the historical BMKG data used is in daily or monthly resolution.
3) Infrastructure Limitations: Utilizing API data requires a stable internet connection, which may be a challenge in some remote locations. Despite these challenges, the prediction results can provide early information on rainfall conditions in a region, which is valuable for decision-making in disaster risk mitigation and daily activities.

The findings of this research have several important implications, including:

- Disaster Mitigation: Accurate rainfall predictions enable governments and communities to take proactive measures against disasters such as floods, landslides, and droughts.
- Agricultural Planning: The annual rainfall classification information can help farmers determine more effective planting schedules based on seasonal patterns of wet and dry periods.
- Water Resource Management: Rainfall trend predictions assist in managing reservoirs, irrigation, and the supply of clean water during dry seasons.

- Real-Time Technology Use: Integrating real-time data from the OpenWeather API highlights the potential of Internet of Things (IoT)-based technologies for monitoring weather conditions in real time

While the research results demonstrate good performance, there are some limitations that need to be considered:

- Climate Variability: The model does not explicitly account for global climate factors like ENSO in its input features.
- Data Resolution: Historical BMKG data is used at a daily resolution, whereas the real-time API predictions have an hourly resolution, which may impact the prediction outcomes.
- Local Factors: Local topography has not been fully integrated into the model, yet it influences rainfall patterns in hilly or mountainous areas.

Based on the findings of this research, several recommendations for future development include:

- Integrating ENSO Factors: Incorporating ENSO indices (El Niño/La Niña) as input variables to improve prediction accuracy during extreme conditions.
- Higher Spatial Resolution: Combining data from multiple BMKG stations within the same region to enhance the spatial resolution of predictions.
- Satellite Data Use: Integrating satellite observation data such as TRMM (Tropical Rainfall Measuring Mission) to complement historical data.
- Model Optimization: Evaluating other models like LSTM (Long Short-Term Memory) or combinations of XGBoost and deep learning to further enhance prediction performance.

4. CONCLUSION

This study successfully developed a rainfall prediction and classification model utilizing the Extreme Gradient Boosting (XGBoost) algorithm, incorporating historical data from BMKG and real-time data from the OpenWeather API. The model's performance evaluation yielded a Mean Squared Error (MSE) of 112.728, Root Mean Squared Error (RMSE) of 10.617, Precision of 0.6409, and Accuracy of 0.0333, reflecting moderate predictive capability. The results are presented through a Flask-based dashboard that showcases model evaluations, rainfall trend visualizations, Oldeman-based rainfall classification, and forecasts for the next 12 months. By integrating historical and real-time data, the system enables dynamic and up-to-date predictions, offering practical applications for disaster management, agricultural planning, and water resource optimization. This research highlights the potential of machine learning technologies in enhancing rainfall prediction accuracy, supporting informed decision-making across multiple sectors.

REFERENCE

[1] M. A. Priatna and E. C. Djamal, "Precipitation prediction using recurrent neural networks and long short-term memory," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 5, pp. 2525–2532, Oct. 2020, doi: 10.12928/TELKOMNIKA.V18I5.14887.

[2] S. Poornima and M. Pushpalatha, "Prediction of rainfall using intensified LSTM based recurrent Neural Network with Weighted Linear Units," *Atmosphere (Basel)*, vol. 10, no. 11, Nov. 2019, doi: 10.3390/atmos10110668.

[3] N. K. A. Appiah-Badu, Y. M. Missah, L. K. Amekudzi, N. Ussiph, T. Frimpong, and E. Ahene, "Rainfall Prediction Using Machine Learning Algorithms for the Various Ecological Zones of Ghana," *IEEE Access*, vol. 10, pp. 5069–5082, 2022, doi: 10.1109/ACCESS.2021.3139312.

[4] M. Putra, M. S. Rosid, and D. Handoko, "High-Resolution Rainfall Estimation Using Ensemble Learning Techniques and Multisensor Data Integration," *Sensors*, vol. 24, no. 15, Aug. 2024, doi: 10.3390/s24155030.

[5] E. G. Dada, H. J. Yakubu, and D. O. Oyewola, "Artificial Neural Network Models for Rainfall Prediction," *European Journal of Electrical Engineering and Computer Science*, vol. 5, no. 2, pp. 30–35, Apr. 2021, doi: 10.24018/ejece.2021.5.2.313.

[6] M. Mohammed, R. Kolapalli, N. Golla, and S. S. Maturi, "Prediction Of Rainfall Using Machine Learning Techniques," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 9, p. 1, 2020, [Online]. Available: www.ijstr.org

[7] A. Sharma and G. Nijhawan, "Rainfall Prediction Using Neural Network," *International Journal of Computer Science Trends and Technology*, vol. 3, 2015, [Online]. Available: www.ijcstjournal.org

[8] Q. Zou, Y. Liu, and X. Linge, "A survey on rainfall forecasting using artificial neural network'," 2019.

[9] I. S. Harahap, I. Z. Matondang, Suryanto, E. K. Indah, and I. Fitri, "Mapping Climate Classification of Oldeman in Agricultural Resources Management in South Tapanuli District," *IOP Conf Ser Mater Sci Eng*, vol. 1156, no. 1, p. 012002, Jun. 2021, doi: 10.1088/1757-899x/1156/1/012002.

[10] B. Ahmada *et al.*, "Climate-based crop pattern determination using Standard Precipitation Index (SPI) and the Oldeman classification in Sangiran Site," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Mar. 2020. doi: 10.1088/1755-1315/451/1/012019.

[11] C. Dewi and R.-C. Chen, "Integrating Real-Time Weather Forecasts Data Using OpenWeatherMap and Twitter," 2019. [Online]. Available: http://ejournal.uksw.edu/ijiteb

[12] A. Musah *et al.*, "An Evaluation of the OpenWeatherMap API versus INMET Using Weather Data from Two Brazilian Cities: Recife and Campina Grande," *Data (Basel)*, vol. 7, no. 8, Aug. 2022, doi: 10.3390/data7080106.

[13] C. Sugiarto, F. Debora Abigael, Y. Faiz Athallah, A. Hari Saputra, and S. Tinggi Meteorologi Klimatologi dan Geofisika, "UTILIZATION OF HUBERREGRESSOR MACHINE LEARNING MODEL TO PREDICT CARBON MONOXIDE CONCENTRATION IN SURABAYA CITY," *JCEBT*, vol. 8, no. 1, 2024, [Online]. Available: http://ojs.uma.ac.id/index.php/jcebt

[14] M. T. Anwar, H. D. Purnomo, S. Y. J. Prasetyo, and D. K. Hartomo, "Decision Tree Learning Approach To Wildfire Modeling on Peat and Non-Peat Land in Riau Province," 2018.

[15] B. Mahesh, "Machine Learning Algorithms - A Review," *International Journal of Science and Research (IJSR)*, vol. 9, no. 1, pp. 381–386, Jan. 2020, doi: 10.21275/art20203995.

[16] J. Kang, H. Wang, F. Yuan, Z. Wang, J. Huang, and T. Qiu, "Prediction of precipitation based on recurrent neural networks in jingdezhen, jiangxi province, China," *Atmosphere (Basel)*, vol. 11, no. 3, Mar. 2020, doi: 10.3390/atmos11030246.

[17] N. Uzir, S. Raman, S. Banerjee, and R. S. Nishant Uzir Sunil R, "Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets," *International Journal of Control Theory and Applications*, vol. 9, 2016, [Online]. Available: https://www.researchgate.net/publication/318132203

[18] Aditya Gumilar, Sri Suryani Prasetiyowati, and Yuliant Sibaroni, "Performance Analysis of Hybrid Machine Learning Methods on Imbalanced Data (Rainfall Classification)," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 3, pp. 481–490, Jul. 2022, doi: 10.29207/resti.v6i3.4142.

[19] R. Adityo, "PENGARUH JUMLAH KATEGORI, RENTANG THRESHOLD DAN METODE ESTIMASI TERHADAP AKURASI SKOR TES PADA BEBERAPA MODEL POLITOMI," 2014.

[20] H. Zheng, J. Yuan, and L. Chen, "Short-Term Load Forecasting Using EMD-LSTM neural networks with a xgboost algorithm for feature importance evaluation," *Energies (Basel)*, vol. 10, no. 8, Aug. 2017, doi: 10.3390/en10081168.