Geoid
Journal of Geodesy and Geomatics

Geoid Vol. 20, No. 1, 2025, 10-21

P-ISSN: 1858-2281; E-ISSN: 2442-3998

# Prediction of Rice Productivity Using the Random Forest Regression Algorithm in Cikaret Subdistrict for the Years 2020-2024

**Muhammad Firdaus*, Firman Hadi, L M Sabri**
Department of Geodetic Engineering, Universitas Diponegoro, Semarang, Indonesia
*Corresponding author: mfirdaus@students.undip.ac.id

Received: 4 January 2025; Revised: 3 March 2025; Accepted: 8 March 2025; Published: 30 April 2025

**Abstract:** The challenges surrounding rice productivity in Indonesia are growing more complex due to factors like climate change, population growth, and limited agricultural land. As the primary food source and main carbohydrate provider, rice is crucial for the majority of Indonesians. This study focuses on predicting rice productivity using the random forest regression algorithm, incorporating predictor variables such as NDVI, NDMI, land area, land surface temperature, rainfall, fertilizer type, and pests. To ensure the accuracy of the model, multicollinearity tests were conducted to check for strong correlations among the independent variables. The tests confirmed the absence of significant linear relationships, allowing all variables to be included in the model. The prediction model was built using time-series data from 2020 to 2023, resulting in 840 samples after eliminating outliers. The optimization process targeted the mtry parameter and the number of decision trees to reduce prediction error. The optimal model, utilizing 7 predictor features and 150 decision trees, achieved a low out of bag (OOB) error and stable mean square error (MSE). Model performance metrics showed a Mean Absolute Error (MAE) of 0.324 tons/hectare, MSE of 0.158 tons/hectare, Root Mean Square Error (RMSE) of 0.398 tons/hectare, and a coefficient of determination ($R^2$) of 0.87. These results demonstrated that the random forest regression algorithm is highly effective in predicting rice productivity, particularly when dealing with complex data involving multiple predictor variables and potential multicollinearity.



Keywords : challanges, forecasting, MAPE, MSE, RMSE

How to cite: Firdaus, M., Hadi, F., & Sabri, L. M. (2025). Prediction of Rice Productivity Using the Random Forest Regression Algorithm in Cikaret Subdistrict for the Years 2020-2024. *Geoid*, *20*(1), 10 - 21.

## Introduction

In Indonesia, food is closely associated with rice, as it serves as the staple food and primary source of carbohydrates for nearly the entire population. Rice also plays a significant role in dietary patterns across most Asian countries and even globally. The availability of rice is crucial for maintaining food security in Indonesia (Suwarno, 2010; Marzani & Juliannisa, 2024). Rice, as the primary crop, significantly impacts food security, with rice availability being a key factor in ensuring a sustainable food supply throughout the year. The sustainability of rice production, in terms of both quality and quantity, is an essential determinant of food availability, which is a basic need for the Indonesian population (Bagio & Athaillah, 2020; Putri, 2023). Population growth greatly influences the balance between rice availability and consumption needs. A region is categorized as having a rice surplus when its availability exceeds consumption needs, whereas it is considered to have a rice deficit if availability falls below consumption requirements.

Rice productivity in Indonesia is of paramount importance because rice is a staple commodity for the majority of the population, and agriculture is a major contributor to the national Gross Domestic Product (GDP). In 2023, the harvested area of rice was estimated to decrease to 10.20 million hectares, leading to a 2.05% decline in production compared to the previous year (BPS, 2024). Factors such as geographic location, climate (rainfall, temperature, wind speed), and solar radiation significantly influence rice production outcomes (Hidayati & Aldrian, 2012; Boamah et al., 2024). Additionally, pests and diseases associated with fluctuations in temperature and humidity also have a substantial impact (Ruminta, 2016).

This study employs six parameters—rainfall, land surface temperature (LST), Normalized Difference Moisture

Indeks (NDMI), land area, fertilizer type, and pests—to analyze rice productivity. Advances in technology, particularly the application of geographic information systems (GIS), have become increasingly important in supporting agricultural decision-making and planning. GIS facilitates optimization of production outcomes by analyzing the relationships between various factors influencing rice productivity (Amrullah et al., 2014; Teng et al., 2023). Factors affecting rice productivity can be integrated using GIS and machine learning techniques such as artificial neural networks (ANN), random forest (RF), and k-nearest neighbor (KNN). Previous studies have demonstrated the effectiveness of machine learning algorithms, such as random forest, in analyzing rice data and predicting yields. For instance, Masdian et al. (2023) utilized random forest to classify imagery and analyze land cover changes, while Nur et al. (2023) employed the method to predict rice yields based on various criteria. RF constructs multiple decision trees and integrates their outputs for classification and regression (Triscowati et al., 2021). Model evaluation involves calculating root mean square error (RMSE) and mean absolute percentage error (MAPE) to assess the accuracy of rice productivity predictions.

In this study, the choice of RMSE and MAPE as statistical methods for evaluating the predictive accuracy of rice productivity models is driven by their suitability for continuous data, such as rice yield measurements. RMSE provides an absolute measure of prediction error, making it easy to understand how close predicted values are to actual observations, while MAPE expresses the error as a percentage, allowing for a clear understanding of the relative error in predictions. This is particularly useful for comparing model performance across different datasets or scales. Both methods are more robust to outliers compared to traditional metrics like $R^2$, which may not perform well with skewed data or extreme values, common in agricultural research. Additionally, since the study involves regression analysis to predict continuous variables, RMSE and MAPE are more appropriate than classification-based metrics, such as the F1 score, which are typically used for categorical data. These statistical methods offer precise, interpretable, and actionable insights, which are crucial for decision-making in agricultural contexts, such as adjusting farming practices to optimize rice yield.

This study builds on previous research as the foundation for planning further investigations. Earlier research has explored rice productivity using machine learning methods. Masdian et al. (2023) evaluated rice production efficiency in Batang Regency from 2018–2022 using the random forest algorithm. Sentinel-2 imagery data was utilized to monitor land cover changes and rice productivity. The classification results showed high accuracy, with a producer accuracy of 95.556%, user accuracy of 86%, overall accuracy of 91%, and a Kappa value of 0.82. Regression evaluation yielded RMSEs of 1.857 tons/ha for district-level data and 0.498 tons/ha for field data.

Ramadhona et al. (2018) applied the backpropagation neural network method to predict rice productivity, employing min-max normalization and Nguyen-Widrow weight initialization. The model achieved the lowest RMSE of 8.6918 and an average RMSE of 8.2126 with 5-fold cross-validation. Nur et al. (2023) implemented RF to forecast rice yields, with key variables such as land area. The model achieved 95.11% accuracy, a MAPE of 4.884%, an RMSE of 0.250, and an $R^2$ of 0.99. Chang et al. (2023) developed a RF-LUE model to estimate Gross Primary Productivity (GPP) using variables such as temperature and vegetation indices. The model demonstrated $R^2$ values ranging from 0.52 to 0.97, with improved performance on longer temporal scales. Nti et al. (2023) focused on tree-based ensemble learning models to predict crop suitability and productivity, achieving an accuracy of 99.32% and an F1-score of 99.34%. Factors such as rainfall and potassium content were crucial in determining crop selection. Champaneri et al. (2020) in India developed a crop yield prediction system based on machine learning algorithms to assist farmers. The system addressed challenges in yield prediction by utilizing random forest for regression and classification tasks.

This study aims to predict or model rice productivity outcomes. The primary focus of data pre-processing in this research is the processing of parameter data using remote sensing technology via cloud-based processing software, such as Google Earth Engine (GEE). The authors hope that the findings of this study will not only contribute to scientific advancements but also aid stakeholders in decision-making processes. By providing a predictive model for rice productivity, policymakers and field practitioners can use this information for evaluation and strategic planning to enhance rice productivity.

## Data and Method

### *Study Area and Dataset*

The research was conducted in Cikaret Village, Kebonpedes Subdistrict, Sukabumi Regency, West Java (see Figure 1). The study began in May and continued until data processing was completed in August 2024. The research in Cikaret Village, Kebon Pedes District, Sukabumi Regency, West Java, was conducted for several key reasons. Geographically, Cikaret Village is situated at an elevation of 500 to 600 meters above sea level, with an average monthly rainfall of 500 mm and temperatures ranging between 20°C and 27°C. These conditions are highly favorable for agricultural activities, particularly rice cultivation. Out of the village's total area of approximately 200.009 hectares, 101.097 hectares are dedicated to rice fields, making it an important agricultural hub. The potential of Cikaret Village to support food security through sustainable rice production further underscores its importance as a study area. Additionally, understanding the socio-economic dynamics of the community, which largely depends on agriculture, provides valuable insights into the interactions between environmental, social, and economic factors. The limited availability of prior studies on Cikaret Village also offers an opportunity to generate new and significant data. Moreover, as part of Sukabumi Regency, a critical contributor to West Java's rice production, this village serves as a relevant case study to evaluate factors affecting agricultural productivity in the region.

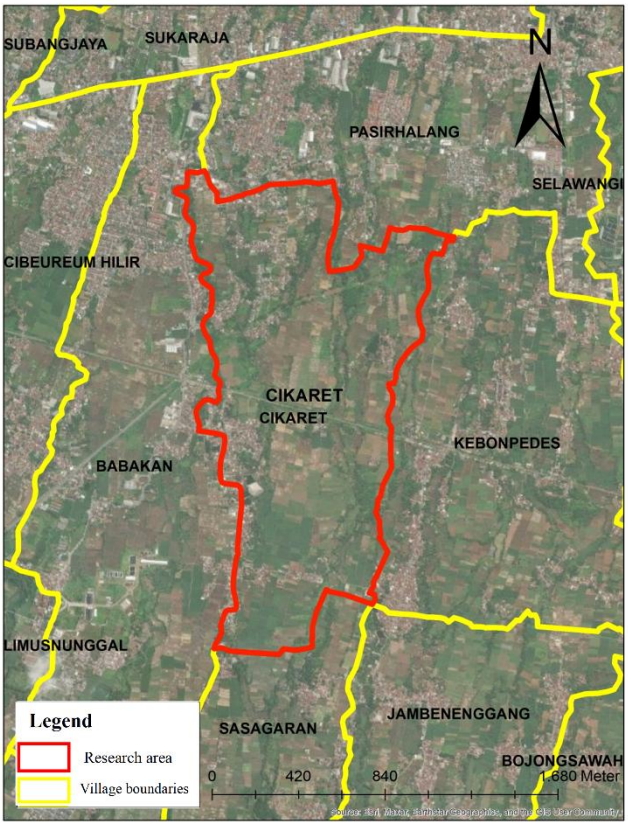


Figure 1. Research Location

The data used in this research consists of primary and secondary data. The primary data refers to the data which collected directly from the field. That includes interview data with farmers in the study area which aims to obtain historical data on rice productivity during the 2020–2023 period. Later on, the data will be used as verification data in the rice productivity prediction model. Meanwhile, the secondary data includes the data which obtained from several digital open sources. The data is presented in Table 1. The data types chosen for this study are essential for understanding the factors affecting rice productivity in Cikaret Village. Rainfall imagery captures spatial data on rainfall patterns, which influence rice growth. Harmonized Landsat-Sentinel

(HLS, L30) and Harmonized Landsat-Sentinel (HLS, S30) provide high-resolution satellite imagery for monitoring land cover and vegetation health. Digital Elevation Model (DEM) data helps assess topography, impacting water drainage, soil erosion, and flood risk. Land parcel shapefile (SHP) data maps agricultural plots, correlating land characteristics with productivity, while administrative boundary shapefile (SHP) defines the study area's limits. Data on the type of fertilizer and pests per parcel is crucial for evaluating their effects on rice yields. Rice productivity data per parcel is the core measure of yield, allowing for a comprehensive analysis of how environmental factors and agricultural practices influence productivity. Combining these datasets offers valuable insights for improving rice farming practices.

Table 1. Secondary Data

| No. | Data | Sources | Years |
|---|---|---|---|
| 1. | Rainfall imagery | CHIRPS | 2020-2024 |
| 2. | HLS (L30) | Google Earth Engine | 2020-2024 |
| 3. | HLS (S30) | Google Earth Engine | 2020-2024 |
| 4. | DEM | DEM SRTM | 2014 |
| 5. | Land parcel SHP | National Land Agency (BPN) | 2019 |
| 6. | Administrative boundary SHP | BPN | 2019 |
| 7. | Type of fertilizer per parcel | Farmer | 2020-2024 |
| 8. | Type of pest per parcel | Farmer | 2020-2024 |
| 9. | Rice productivity data per parcel | Farmer | 2020-2024 |

***Data Processing Stages***

The data processing for rice productivity prediction using satellite imagery and rainfall data involves a series of systematic stages to ensure accurate and reliable modeling. The first stage is the processing of HLS S30 imagery, starting with selecting images with a maximum cloud cover of 20% for the periods from January to May and June to October of the years 2020 to 2023. Cloud masking is then applied to remove clouds and cirrus that could obscure the land surface, ensuring that only relevant data is used. Following this, specific bands from the images are selected to simplify the subsequent processing. The images are clipped to the administrative boundaries of Cikaret Village, focusing the analysis on the relevant area. Indices such as NDVI (Normalized Difference Vegetation Index), NDMI (Normalized Difference Moisture Index), and NDWI (Normalized Difference Water Index) are calculated to visualize vegetation health, moisture levels, and water availability, which are crucial factors influencing rice productivity.

The next stage involves the processing of HLS L30 imagery, which is processed similarly but with an additional focus on surface temperature. Like HLS S30, images are selected based on the same periods and clipped to the research area. However, the primary difference is the calculation of Land Surface Temperature (LST) using the thermal bands of the HLS L30 imagery. Surface temperature is vital for understanding agricultural conditions, as it directly affects plant growth, water stress, and overall productivity. By calculating the average surface temperature from these thermal bands, the study gains insights into temperature variations across the region, which can influence rice productivity predictions.

Next, rainfall data is processed using the CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) dataset, which provides high-resolution rainfall data essential for understanding weather patterns affecting rice growth. This data is then used to calculate the average monthly rainfall values, which will serve as an important predictor in the productivity model. The imagery is reprojected into the WGS 84 coordinate system (EPSG:4326) to ensure compatibility with other datasets, and image downscaling is performed using NDVI and DEM (Digital Elevation Model) Shuttle Radar Topography Mission (SRTM) data as predictors. This downscaling process increases the spatial resolution of the rainfall data, improving the detail and accuracy at the local level.

After processing the satellite imagery and rainfall datasets, the next step is to conduct a multicollinearity test to identify relationships between variables that could affect the model's results. A correlation test is performed

using Past 4 software with Spearman correlation calculations to ensure that the data used in the Random Forest Regression model is independent and reliable. In the final stage, the Random Forest Regression algorithm is applied to predict rice productivity. Data on eight parameters (such as vegetation indices, surface temperature, and rainfall) and actual rice productivity results are collected and split into training and testing datasets. The Random Forest model is then run to predict productivity based on the patterns identified in the training data.

The processing of both HLS S30 and HLS L30 imagery is carried out separately because each type of imagery serves a different purpose in this study. HLS S30 imagery focuses on vegetation indices and moisture, which are critical for monitoring plant health and soil conditions. In contrast, HLS L30 imagery, which contains thermal bands, is used specifically to calculate surface temperature (LST). Surface temperature is a key variable in understanding temperature stress on crops, which directly affects productivity. By using both types of imagery, the study can combine information on vegetation health, moisture levels, and surface temperature, leading to a more comprehensive and accurate prediction of rice productivity.

## Results and Discussion

### *Multicollinearity Test*

The multicollinearity test aims to identify the degree of correlation between independent variables. High correlation among independent variables can reduce the accuracy of the model in predicting outcomes. In this study, the multicollinearity test was conducted on seven independent variables that serve as predictors in the development of a rice productivity prediction model using machine learning algorithms. These seven variables are shown in Figure 2.

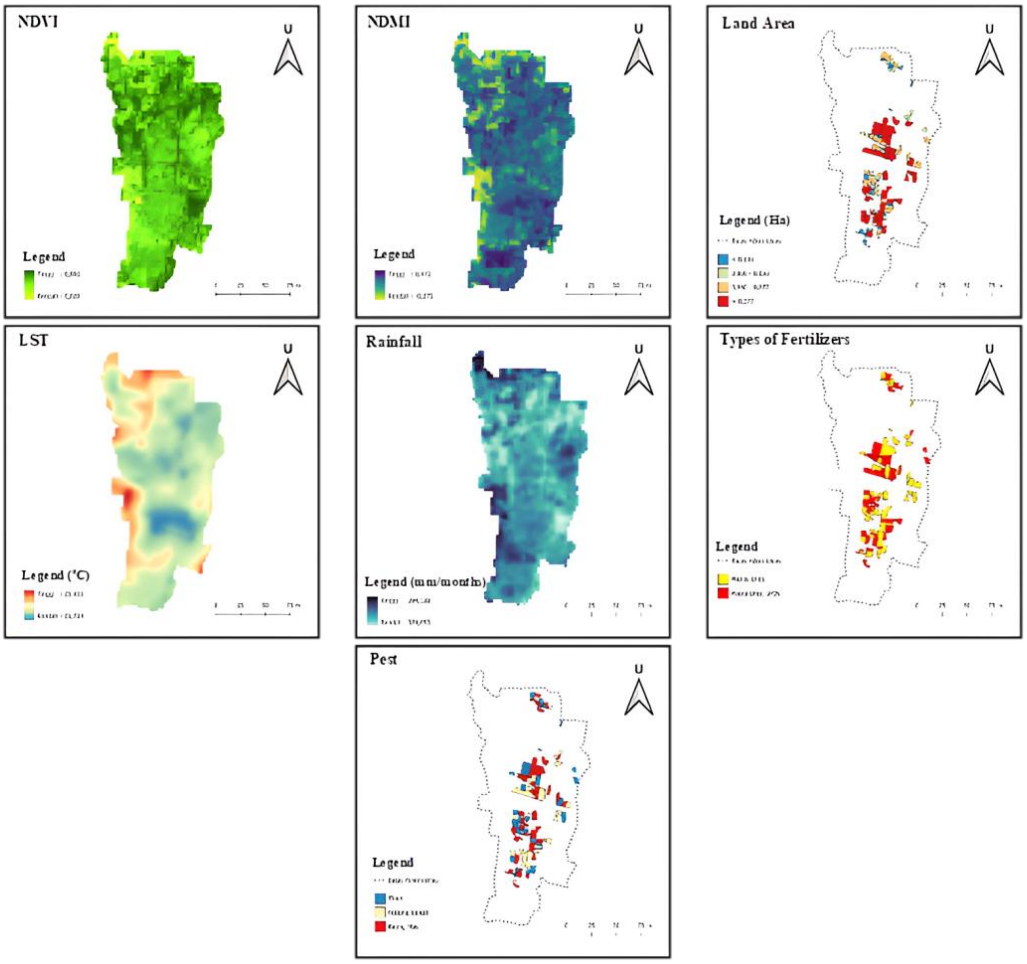


Figure 2. Independent variables in the study

The NDVI, NDMI, land area, LST, and rainfall variables are numerical data, while fertilizer type and pests are categorical data. The multicollinearity test used Spearman's correlation to measure the relationship between variables without requiring additional preprocessing. In conducting a multicollinearity test using the entire dataset used to build the rice productivity prediction model. The results of the multicollinearity test (see Figure 3) revealed that most variables have low correlations, with correlation values below 0.20, indicating no significant linear relationship between the following variables: rainfall with land area, NDMI, NDVI, fertilizer type, and pests; LST with land area, NDMI, fertilizer type, and pests; NDMI with land area, fertilizer type, and pests; NDVI with land area, fertilizer type, and pests; and fertilizer type with land area and pests. Variables with low correlation include NDVI and LST, while moderate linear relationships are observed between LST and rainfall, as well as NDVI and NDMI. Strong correlations are indicated by values between 0.70 and 0.90, while values above 0.90 show a very strong relationship. No strong or very strong linear relationships were found between the variables, allowing all variables to be used in the prediction model. This test is crucial to ensure the reliability of the prediction model, preventing potentially inaccurate models due to multicollinearity. If the model cannot understand the true relationships between rice productivity outcomes and the independent predictor variables, it may fail to reflect the actual dynamics on the ground (Ariani et al., 2020).

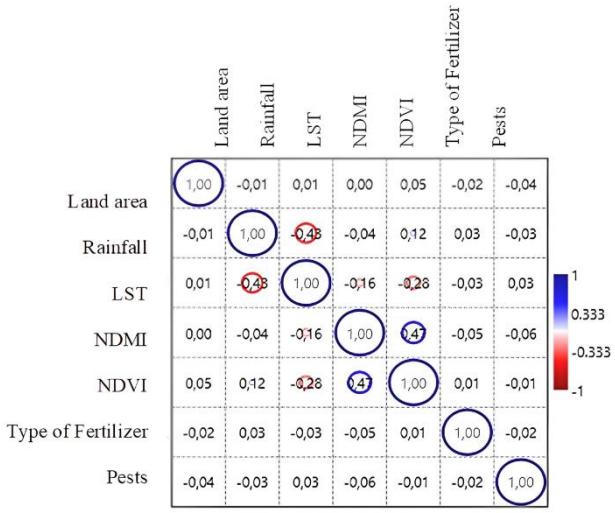


Figure 3. Results of Multicollinearity Test

### *Optimization of the Random Forest Regression Model*

The rice productivity prediction model uses the random forest regression algorithm with seven predictor variables. The sample data consists of 107 parcels from the period 2020 to 2023, totaling 856 data points. After removing outliers, 840 data points were used, split into 70% for training data (590 data points) and 30% for testing data (250 data points). The training data is used to build the model, while the testing data is used to evaluate its accuracy. The distribution of the random forest regression parcel samples is presented in Figure 4.

Optimization of parameters in the prediction model using the random forest regression algorithm is performed to minimize prediction errors. Two important parameters are considered: mtry (the number of random features considered at each split) and ntree (the number of decision trees). The mtry parameter refers to the number of features (predictor variables) randomly selected from the total predictor variables to be considered for splitting at each node in each decision tree (López et al., 2022). Overfitting can be avoided by adjusting the mtry parameter, which also helps control the complexity of the model (Firmansyach et al., 2023). In addition, the Out of Bag Error (OOB) calculated with the mtry parameter is used to evaluate the performance of the prediction model (Lee et al., 2020). The analysis results indicate that mtry set to 7 yields the lowest OOB error

(0.00189), compared to the highest value (0.00279) at mtry 2. Therefore, the optimal mtry parameter for this model is 7. The Out of Bag Error value is shown in Figure 5.

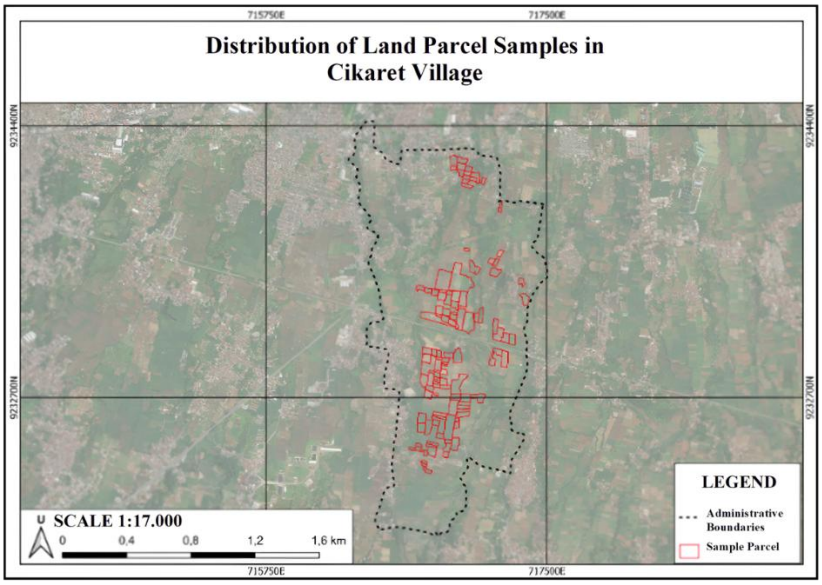


Figure 4. Distribution of Random Forest Regression Parcel Samples

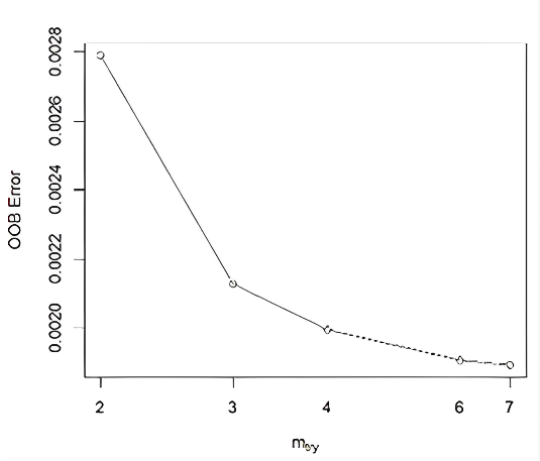


Figure 5. Out of Bag Error Values

The prediction error rate in the model can be monitored through the Mean Square Error (MSE) values for each tree in the ensemble, as shown in Figure 6. A very low error rate indicates that the prediction model is highly effective. At the beginning of the graph, the error rate is relatively high due to the limited number of trees (ntree), meaning the random forest model lacks sufficient complexity to effectively model the data. As the number of trees increases, the error rate decreases significantly because the ensemble effect of multiple trees reduces prediction variability and improves accuracy. After approximately 100 trees, the graph shows a diminishing reduction in error rate, indicating the model has reached a stabilization point where adding more trees does not yield significant accuracy improvements. Adding trees up to 500 demonstrates diminishing returns in reducing the error rate. Table 3 shows that the lowest MSE is found at the 150th tree, while the highest $R^2$ value occurs at the 291st tree. This indicates that beyond the stabilization point, adding more trees becomes inefficient in terms of accuracy and computational time. This finding aligns with Oshiro et al. (2012), who stated that the optimal number of trees does not always correlate with a larger number of trees.

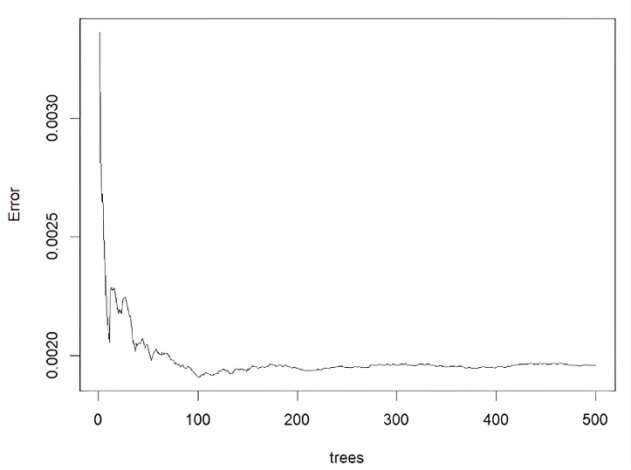


Figure 6. Overall Model Error Rate

***Optimization of Model Performance***

Table 3 represents the statistical results before and after optimization, including mean squared residual (MSE) and explained variance percentage. The explained variance percentage indicates how well the model accounts for the variability in the dependent variable based on the independent variables (Agustian et al., 2019). As seen, an explained variance percentage of 81.32%, meaning the model can explain 81.32% of the variability in the response data, while the remaining 18.68% cannot be explained and may be attributed to random factors or unmodeled variables. The model's performance is also improved, as evidenced by the reduction in MSR. MSR measures the average squared differences between the model's predicted values and the actual values of the response variable. Figure 7 shows that the MSR value is lower than in Figure 10, indicating that the model's predictions are closer to the actual values. The decrease in MSR and the increase in explained variance percentage demonstrate that the model exhibits excellent performance and accuracy in predicting rice productivity outcomes.

Table 3. Performance of the model before and after the optimization

| Metric | Before Optimization | After Optimization |
|---|---|---|
| Number of Trees (ntree) | 500 | 500 |
| Number of Variables Tried at Each Split (mtry) | 2 | 7 |
| Mean of Squared Residuals | 0.002687342 | 0.00195958 |
| % Variance Explained | 74.39% | 81.32% |

***Variable Contribution***

In the random forest regression model, the contribution of each variable is assessed by measuring its importance, which evaluates how much each feature contributes to the prediction outcomes. Two primary metrics are used to determine variable importance: Percentage Increase in Mean Squared Error (%IncMSE) and Increase in Node Purity (IncNodePurity). %IncMSE measures variable importance by randomly permuting the values of a single feature within the dataset and observing its impact on the model's Mean Squared Error (MSE). MSE is the average squared difference between the model's predicted values and the actual values. A variable with high %IncMSE indicates that it is highly important, as permuting its values causes a significant increase in prediction error. Conversely, a low %IncMSE suggests that the variable is less important. Figure 7 shows that the land area variable has the highest %IncMSE value, followed by the rainfall variable. This

indicates that land area and rainfall have the most significant impact on prediction error, thereby strongly influencing the model's performance.

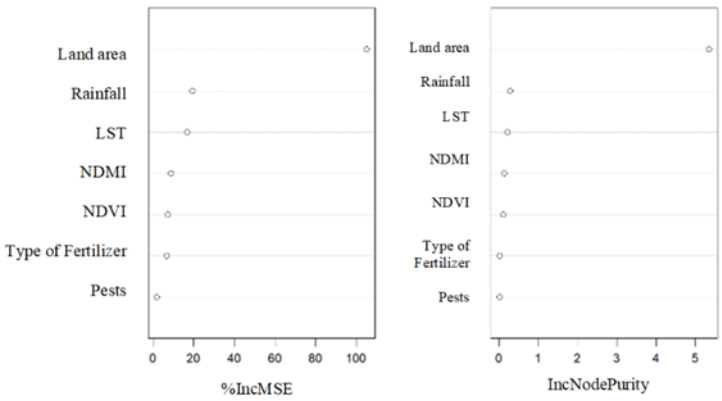


Figure 7. Variable Importance in Random Forest

In this study, variable importance was measured using two metrics: %IncMSE and IncNodePurity. %IncMSE evaluates the importance of a variable based on the increase in prediction error when the feature values are permuted, while IncNodePurity measures how frequently a feature is used to increase the purity of nodes in decision trees. A high IncNodePurity value indicates that the variable is frequently used for node splitting and contributes significantly to improving node purity, thus being considered highly important. Conversely, a low value suggests that the feature is less informative. The land area variable has the highest value for both metrics. Overall, the most important variables according to both metrics are land area, rainfall, and LST, which significantly contribute to predicting rice productivity. The importance of variables in the random forest model is presented in Table 4.

Table 4. Variable Importance in Random Forest

| Variables | %IncMSE | IncNodePurity |
|---|---|---|
| Land Area | 150.366135223 | 5.345914290 |
| Rainfall | 27.050013119 | 0.220501891 |
| LST | 23.643045798 | 0.273914511 |
| NDMI | 4.771512642 | 0.117579357 |
| NDVI | 12.298924944 | 0.104248476 |
| Fertilizer Type | 10.169299692 | 0.009779692 |
| Pests | 10.808511633 | 0.012214961 |

***Model Accuracy***

The accuracy of the rice productivity prediction model was tested using a test dataset comprising 250 data points, which were not included in the model development. Accuracy was evaluated by calculating RMSE, MAPE, and $R^2$, comparing the actual values with the predictions from the test data. Details of this comparison are presented in Table 5. The Random Forest model with 500 trees shows an RMSE of 0.037 and an $R^2$ of 0.852, indicating a significant effect of the predictor variables on the response variable. The $R^2$ value approaching 1 indicates a high coefficient strength. Additionally, the MAPE of 0.38% suggests the model's excellent forecasting ability, as a low MAPE (<10%) indicates high prediction accuracy, reaching 99.62%.

Table 5. Performance Evaluation

| RMSE | $R^2$ | MAPE | Accuracy |
|---|---|---|---|
| 0,037 | 0,852 | 0,38 % | 99,62 % |

The model's accuracy was tested again using rice productivity data from January to May 2024. The predicted results were very close to the actual values, demonstrating that the model still meets the criteria for good

regression. As shown in Table 6, the predicted results remain close to the actual values, indicating that the prediction model still meets the criteria for good regression.

Table 6. Comparison of Predicted Results with Actual Data for 2024

| No. | Actual Data (ton) | Prediction Data (ton) |
|---|---|---|
| 1 | 0.420 | 0.573 |
| 2 | 0.928 | 0.866 |
| 3 | 0.332 | 0.273 |
| 4 | 0.272 | 0.273 |
| 5 | 0.412 | 0.343 |
| 6 | 1.378 | 1.163 |
| 7 | 0.758 | 0.638 |
| 8 | 1.152 | 0.930 |
| 9 | 0.697 | 0.621 |
| 10 | 1.096 | 0.947 |
| 11 | 0.592 | 0.543 |

After re-evaluating the accuracy, the prediction model shows an RMSE of 0.114 and an $R^2$ of 0.87, while the MAPE is 5.28%, resulting in a model accuracy of 94.72% (Table 7). Although there is a decrease in the model's accuracy, this is due to the limited test data, with only 11 data points for the year 2024. However, the model still maintains high accuracy above 90%, at 94.72%.

Table 7. Performance Evaluation for 2024 Data

| RMSE | $R^2$ | MAPE | Accurate |
|---|---|---|---|
| 0,114 | 0,87 | 5,28 % | 94,72 % |

## Conclusions

Based on this study, several conclusions and recommendations can be drawn. First, the rice productivity prediction model using the random forest regression algorithm with historical data from 2020 to 2024 demonstrated highly accurate results. By utilizing 500 trees and splitting the dataset into 70% training data and 30% testing data, the model achieved an accuracy of 99.62% with an RMSE value of 0.037, MAPE of 0.38%, and $R^2$ of 0.852. Although there was a decrease in accuracy to 94.72% with the latest 2024 data, the RMSE, MAPE, and $R^2$ values still indicated excellent performance. Second, this modeling involved seven variables, with land area, rainfall, and LST contributing the most to the prediction results. In contrast, the NDMI variable had the least influence due to its variation in correlation with the dependent variable. For future research, several suggestions are proposed: Incorporate additional variables related to factors affecting rice productivity, such as water availability, to enhance the comprehensiveness of predictions and employ other machine learning algorithms, such as Support Vector Machine, Artificial Neural Network, or K-Nearest Neighbors, to identify the most effective algorithm for predicting rice productivity outcomes.

## References

Agustian, I., Saputra, H. E., & Imanda, A. (2019). Pengaruh Sistem Informasi Manajamen Terhadap Peningkatan Kualitas Pelayanan Di Pt. Jasaraharja Putra Cabang Bengkulu. *Profesional: Jurnal Komunikasi Dan Administrasi Publik*, *6*(1), 42–60. https://doi.org/10.37676/professional.v6i1.837

Amrullah., S. D. (2014). Peningkatan Produktivitas Tanaman Padi (Oryza sativa L.) melalui Pemberian Nano Silika. *Pangan*, *23*(1-2), 17-32.

Ariani, D., Prasetyo, Y., & Sasmito, B. (2020). Estimasi Tingkat Produktivitas Padi Berdasarkan Algoritma NDVI, EVI Dan SAVI Menggunakan Citra Sentinel-2 Multitemporal (Studi Kasus: Kabupaten Pekalongan, Jawa Tengah). *Jurnal Geodesi Undip*, *9*(1), 207–216.

Bagio., & Athaillah, T. (2020). Pembukuan Usaha Tani Padi di Desa Leuhan Kecamatan Johan Pahlawan Kabupaten Aceh Barat. *Jurnal Abdimas Bina Bangsa*, *1*(1), 80-86. https://doi.org/10.46306/jabb.v1i1.13

Boamah, P. O., Onumah, J., Apam, B., Salifu, T., Abunkudugu, A. A., & Alabil, S. A. (2024). Climate variability impact on crop evapotranspiration in the upper East region of Ghana. *Environmental Challenges*, *14*, 100828. https://doi.org/10.1016/j.envc.2023.100828

BPS. (2024). *Luas Panen Padi dan Produksi Padi di Indonesia 2023.* Badan Pusat Statistik.

Champaneri, M., Chachpara, D., Chandvidkar, C., & Rathod, M. (2020). Crop Yield Prediction Using Machine Learning. *International Journal of Science and Research*, *9*(4), 1-5.

Chang, X., Xing, Y., Gong, W., Yang, C., Guo, Z., Wang, D., . . . Yang, S. (2023). Evaluating gross primary productivity over 9 ChinaFlux sites based on random forest regression models, remote sensing, and eddy covariance data. *Science of the Total Environment 875*(162601).

Firmansyach, W. A., Hayati, U., & Arie Wijaya, Y. (2023). Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(1), 262–269. https://doi.org/10.36040/jati.v7i1.6329

Google, E. E. (2004). *Earth Engine Data Catalog*. Dipetik July 24, 2024, dari https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED#bands

Hidayati, D., & Aldrian, E. (2012). *Perubahan iklim: upaya peningkatan pengetahuan dan adaptasi petani dan nelayan melalui radio.* Bogor: Sarana Komunikasi Utama.

López, M. O. A., López, M. A., & Crossa, J. (2022). *Multivariate statistical machine learning methods for genomic prediction*. Mexico: Springer.

Marzani, Y., & Juliannisa, I. A. (2024). Assessment Ketersediaan Beras Pada 34 Provinsi Di Indonesia. *Jurnal Ekonomi Pertanian dan Agribisnis (JEPA)*, *8*(2), 700–711.

Masdian, A. R., Bashit, N., & Hadi, F. (2023). Analisis Produktivitas Padi Menggunakan Algoritma Machine Learning Random Forest Di Kabupaten Batang Tahun 2018 - 2022. *Elipsoida: Jurnal Geodesi dan Geomatika*, *06*(01), 43-51.

Nti, I. K., Zaman, A., Nyarko-Boateng, O., Adekoya, A. F., & Keyeremeh, F. (2023). A predictive analytics model for crop suitability and productivity with tree-based ensemble learning. *Decision Analytics Journal*, *8*(August), 1-11.

Nur, N., Wajid, F., Sulfayanti, & Wildayani. (2023). Implementasi Algoritma Random Forest Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga. *Jurnal Komputer Terapan, 9(1)*, 58-64.

Oshiro, T.M., Perez, P.S., Baranauskas, J.A. (2012). How Many Trees in a Random Forest?. In: Perner, P. (eds) Machine Learning and Data Mining in Pattern Recognition. MLDM 2012. *Lecture Notes in Computer Science*, vol 7376. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31537-4_13

Putri, F. A. (2023). Optimalisasi Produksi Padi Menuju Ketahanan Pangan di Jawa Tengah. *Seminar Nasional Official Statistics*, *2023*(1), 827–838. https://doi.org/10.34123/semnasoffstat.v2023i1.1888

Ramadhona, G., Setiawan, B. D., & Bachtiar, F. A. (2018). Prediksi Produktivitas Padi Menggunakan Jaringan Syaraf Tiruan Backpropagation. J*urnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(12), 6048–6057.

Ruminta. (2016). Analisis Penurunan Produksi Tanaman Padi Akibat Perubahan Iklim di Kabupaten Bandung Jawa Barat. *Jurnal* Kultivasi, 15(1), 37-45.

Suwarno. (2010). Meningkatkan Produksi Padi Menuju Ketahanan Pangan yang Lestari. *Pangan*, *19*(3), 234-243.

Teng, Z., Chen, Y., Meng, S., Duan, M., Zhang, J., & Ye, N. (2023). Environmental Stimuli: A Major Challenge during Grain Filling in Cereals. *International Journal of Molecular Sciences*, *24*(3). https://doi.org/10.3390/ijms24032255

Triscowati, D. W., Buana, W. P., & Marsuhandi, A. H. (2021). Pemetaan Potensi Lahan Jagung Menggunakan Citra Satelit Dan Random Forest Pada Cloud computing Google Earth Engine. *Seminar Nasional Official Statistics, Daring: Politeknik Statistika STIS*, 1001-1011