 ISSN 2502-3357 (online) | ISSN 2503-0477 (print)

Contents lists available at www.journal.unipdu.ac.id

Register

Journal Page is available to www.journal.unipdu.ac.id/index.php/register

Research article

Enhanced Rice Yield Prediction in Indonesia with Integrated Climate and Agricultural Data Using Decision Tree Regression

Tegar Arifin Prasetyo [a,], Samuel Jefri Siahaan [b], Usman Efendi [c], Mesya Angeliqa Hutagalung [d], Amalia Nur Alifah [e]*
[a,b,d] Department of Information Technology, Faculty of Vocational Studies, Institut Teknologi Del, Sitoluama, Indonesia
[c] Meteorologist in Indonesian Agency for Meteorological, Climatological and Geophysics (BMKG), Semarang, Indonesia
[e] Department of Data Science, Faculty of Informatics, Telkom University, Indonesia
email: [a,] tegar.prasetyo@del.ac.id, [b] if321037@students.del.ac.id, [c] usman.efendi@bmkg.go.id, [d] if321066@students.del.ac.id, [e] amalialifah@telkomuniversity.ac.id*
** Correspondence*

ARTICLE INFO

Article history:
Received January 16th, 2025
Revised March 28th, 2026
Accepted June 21th, 2026
Available online June 28th, 2026

Keywords:
Rice Yield Prediction
Decision Tree Regression
Climate and Agricultural
Hyperparameter Tuning

Please cite this article in IEEE style as:
T. A. Prasetyo, S. J. Siahaan, U. Efendi, M. A. Hutagalung, A.N. Alifah "Enhanced Rice Yield Prediction in Indonesia with Integrated Climate and Agricultural Data Using Decision Tree Regression", Register : Jurnal Ilmiah Teknologi Sistem Informasi, vol. 12, no. 1, pp. 58-70, 2026

ABSTRACT

Rice is central to Indonesia's food security, yet provincial yields are highly sensitive to climatic variability, making reliable forecasting essential for national planning and improving farmer welfare. Most prior Indonesian yield models rely on rainfall and temperature data and omit sunlight exposure duration, which is a limiting factor for photosynthesis in the humid tropics where solar radiation, not temperature, often constrains productivity. This study develops a province-level rice yeild prediction model based on Decision Tree Regression (DTR) that integrates climate data from the Meteorology, Climatology, and Geophysics Agency (BMKG) with agricultural statistics from the Central Statistics Agency (BPS). The dataset comprises data from 34 provinces covering the period from 2018 to 2023 (204 province-year observations), with year, harvested land area, rainfall, and sunlight exposure duration as predictors and rice production as the target variable. The dataset was partitioned into training and testing subsets using an 80:20 ratio. Hyperparameter tuning was performed using k-fold cross-validation, and model performance was performed using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The model attained an average RMSE of 93,046 tons (MAE ≈ 51,824 tons; MAPE ≈ 6.8% on the 2023 hold-out year). A key finding is that absolute RMSE is strongly scale-dependent. When evaluated in relative terms, the highest-producing Java provinces were among the most accurately predicted (relative RMSE ≈ 2.3–2.5%), whereas several small or structurally volatile provinces showed relative errors above 40%. The study contributes to the literature by providing (i) the explicit integration of sunlight exposure into Indonesian rice yield modeling, (ii) a province-disaggregated error analysis that reframes accuracy in scale-independent terms, and (iii) an interpretable decision-support tool for food-policy stakeholders such as Bulog and the Ministry of Agriculture.

1. Introduction

Indonesia, one of the world's largest archipelagos, is renowned for its agrarian economy, where a significant portion of the population works in agriculture [1], with rice being a crucial crop underpinning national food security [2]. Rice is the primary food source for most Indonesians, making Indonesia the world's third-largest producer, following China and India [3]. In 2022, Indonesia's Central Statistics Agency (BPS) recorded an output of approximately 54.75 million tons of dry milled rice, reflecting the country's strong agricultural productivity. However, variability in rice productivity, influenced by factors such as rainfall patterns, sunlight availability, and cultivated areas, emphasizes the need for continuous and targeted efforts to enhance sectoral quality, yield, and sustainability.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 58-70

Enhanced management and reliable forecasting of rice productivity are crucial for ensuring future rice supply, strengthening national food security, and improving the livelihoods of farmers [4].

Variations in global climate patterns, triggered by various climate phenomena, are reshaping rainfall distribution and agricultural cycles in Indonesia, leading to direct effects on rice yields levels [5]. Several previous studies have demonstrated a strong association between rice crop productivity and key climate factors, including rainfall [6], air temperature [7], and sunlight exposure duration [8]. The decision to incorporate sunlight exposure duration is grounded in the distinct agro-climatic conditions of equatorial Indonesia. In temperate agricultural systems, temperature typically acts as the dominant limiting factor for crop growth, governing both the length of the growing season and the rate of development [9]. In the humid tropics, however, ambient temperatures remain consistently high and vary little across seasons, so temperature rarely constrains or differentiates provincial yields. Under these conditions, the primary binding constraint on productivity shifts to the availability of solar radiation. Rice is a C3 crop whose biomass accumulation depends directly on the quantity of photosynthetically active radiation intercepted by the canopy, particularly during the reproductive and grain-filling stages, during which carbohydrate supply determines grain weight [10]. Indonesia's persistent and extensive cloud cover, intensified during the wet monsoon, substantially attenuates incident solar radiation and shortens effective sunshine duration, thereby reducing photosynthetic output even when temperature and water are not limiting. Sunlight exposure duration therefore serves as a more physiologically meaningful predictor of tropical rice productivity than temperature, justifying its explicit inclusion in the proposed model. A study by [11] highlights that extreme rainfall, elevated temperatures, and prolonged sunlight exposure significantly impact rice productivity, indicating that climate change, especially higher temperatures and increased CO_2 levels, reduces rice biomass and yields. Another study shows that rising trends in rainfall and sunlight exposure significantly influence the productivity of rice, with cooler temperatures linked to improved productivity [12]. Additional studies also confirm that climate change, especially the reduction in rainfall, negatively affects rice production [13]. Accordingly, enhancing predictive models for rice productivity by integrating specific climatic and agricultural parameters has become increasingly critical.

Over the past decade, yield prediction technologies have advanced significantly, especially with the introduction of machine learning (ML) methods such as Decision Tree Regression. The Decision Tree Regression model is widely regarded as an effective approach for yield prediction, capable of managing nonlinear input variables and variable interactions [14], which makes it highly effective for complex datasets with a wide range of features [15]. A study by [16] investigates the application of a combined Decision Tree and AdaBoost algorithm within a machine learning framework to predict crop type and yield based on localized climate variables, achieving strong predictive accuracy and providing targeted crop recommendations tailored to the unique weather conditions of each region. Another previous study assesses the effectiveness of a Decision Tree Regression model for predicting rice production in Indonesia, achieving a Mean Squared Error (MSE) of 0.023467 and an R^2 score of 0.782577 after hyperparameter tuning [17], though the omission of sunlight exposure duration as a variable indicates the potential for further accuracy improvements. Another study proposed a machine learning based prediction system to estimate yields for six crop types in West Africa, using climate, weather, crop yield, and chemical data, with the Decision Tree model achieving the highest accuracy (R^2 of 95.3%), followed by the K-Nearest Neighbor model (R^2 of 93.15%) and logistic regression (R^2 of 89.78%) [18].

A closer examination of these Decision Tree applications reveals a recurring methodological limitation. The Indonesian rice-prediction model of [17], for example, relied on rainfall, temperature, and land-related variables but did not incorporate sunlight exposure duration; its reported R^2 of 0.78 leaves a substantial share of yield variance unexplained, which is plausibly attributable to this omitted radiative factor. Likewise, the crop-yield studies by [15], [18] achieved high accuracy using climate and chemical inputs but were developed for West African savanna conditions and similarly excluded

http://doi.org/10.26594/register.v12i1.5352

sunshine duration, limiting their transferability to Indonesia's cloud-dominated equatorial regime. Across these works, the absence of a direct measure of light availability represents a consistent gap, particularly where it matters most for tropical rice. The present study addresses this gap by integrating BMKG sunlight exposure data with BPS agricultural statistics within a single Decision Tree framework, thereby yielding an enhanced predictive capability that distinguishes the proposed approach from prior models that represented climate primarily through rainfall and temperature.

Our research provides an important contribution to enhancing rice yield prediction models in Indonesia by applying a Decision Tree Regression approach that combines climate data with agricultural data specific to the region. We adopt Decision Tree Regression as the primary model for three reasons. First, its interpretability: its explicit splitting rules can be easily interpreted by extension officers and policy planners, unlike opaque ensemble models. Second, it natively handles nonlinear feature interactions among climate and agricultural variables [14], [15]. Third, it is computationally efficient and reproducible at the provincial scale. This integrated methodology is innovative and remains largely unexplored within the Indonesian context. Leveraging comprehensive datasets, including climate information from the Meteorology, Climatology, and Geophysics Agency (BMKG) and agricultural statistics from the Central Statistics Agency (BPS), this model aims to markedly improve the accuracy and robustness of rice yield predictions. By integrating these high-quality, localized data sources, the model is positioned to serve as a reliable and precise tool for forecasting agricultural yields in the Indonesian context. Although past studies have examined agricultural yield prediction models incorporating climate variables [19], [20], a significant gap persists in developing comprehensive models for rice yield prediction in Indonesia that fully integrate both climate and agricultural factors. In contrast to prior studies, this research considers not only rainfall variability but also incorporates sunlight exposure duration as a critical factor affecting rice productivity. Accordingly, this research aims to enhance the accuracy of rice yield predictions, thereby supporting national food security and advancing farmer welfare through more targeted and sustainable production strategies. In addition, improved prediction accuracy can aid the government in formulating more efficient food policies and securing adequate national rice stocks to manage production fluctuations effectively [21], [22]. A dependable predictive model can further support farmers' well-being by offering guidance on more adaptive and sustainable cultivation practices [23]. Thus, this research aims not only to provide more precise rice yield predictions but also to contribute to shaping more effective climate adaptation strategies and sustainable agricultural resource management for the future.

2. Methods

Figure 1 outlines the process for rice yield prediction in Indonesia using Decision Tree Regression.

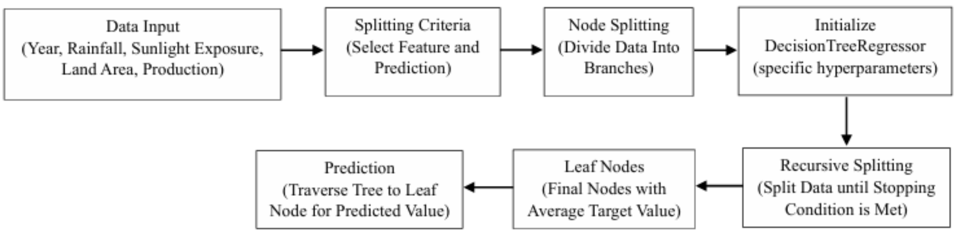

Fig 1. Research Workflow Diagram

Based on Figure 1, the workflow starts with the Data Input stage, which consists of the following variables: year, rainfall, sunlight exposure, land area, and production. The dataset for this research, covering the years from 2018 to 2023, is partitioned into training and testing sets using an 80:20 ratio, with data from 2018 to 2021 designated for training (80%) and data from 2022 to 2023 reserved for testing (20%). Table 1 presents a summary overview of the dataset utilized in this research.

Table 1. Summary Overview of the Dataset

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 58-70

Province	Year	Land Area	Production	Rainfall	Sunlight Exposure
Aceh	2018	329515,8	1861567	15,34	5,09
	2019	310012,5	1714438	8,64	5,16
	2020	317869,4	1757313	10,03	4,97
	2021	297058,4	1634640	69,03	4,72
	2022	271750,2	1509456	11,7	4,41
	2023	254318,6	1393474	7,8	89,91
Sumatera Utara	2018	408176,45	2108284,72	5,54	4,03
	2019	413141,24	2078901,59	6,21	3,98
	2020	388591,22	2040500,19	5,9	4,11
	2021	385405	2004142,51	5,26	4,61
	2022	411462,1	2088584	3,86	4,44
	2023	404472,52	2080663,46	6,44	5,31
Sumatera Selatan	2018	581574,61	2994191,84	7,14	5,15
	2019	539316,52	2603396,24	6,49	5,49
	2020	551320,76	2743059,68	8,09	4,65
	2021	496241,65	2552443,19	6,83	4,7
	2022	513378,2	2775069	8,04	4,27
	2023	502162,22	2762059,57	6,28	5,43
…	…	…	…	…	…
Jawa Tengah	2018	1821983,17	10499588,23	32,59	7,01
	2019	1678479,21	9655653,98	4,5	7,47
	2020	1666931,49	9489164,62	6,22	6,48
	2021	1696712,36	9618656,81	8,33	5,98
	2022	1688670	9356445	6,45	5,79
	2023	1640297,54	9061714,85	5,09	7,12
DIY	2018	93956,45	514935,49	5,03	6,09
	2019	111477,36	533477,4	4,82	61,21
	2020	110548,12	523395,95	280,54	5,33
	2021	107506,16	556531,03	6,75	5,31
	2022	110927,2	561699,5	7,95	4,92
	2023	105394,22	532805,26	3,6	6,4
Jawa Timur	2018	1751191,67	10203213,17	6,81	6,99
	2019	1702426,36	9580933,88	5,19	7,51
	2020	1754380,3	9944538,26	7,86	6,61
	2021	1747481,2	9789587,67	7	6,19
	2022	1693211	9526516	35,22	5,66
	2023	1685559,5	9591422,32	60,39	6,99

Table 1 presents a comprehensive overview of agricultural data across various Indonesian provinces from 2018 to 2023, capturing the essential factors influencing rice production. Each row includes key variables, such as land area (in hectares), rice yield, rainfall (in mm), and sunlight exposure duration (in hours per day) for each year and province. Analyzing these factors over time provides insights into how elements like rainfall and sunlight affect rice productivity and land utilization across different provinces.

Before model training, the dataset comprising 204 province-year observations was prepared through three stages: missing-value inspection, outlier screening, and feature scaling. First, the dataset was inspected for missing entries across all predictor and target variables. No missing values were

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 58-70

found, so no imputation was required. Second, each climate variable was screened for outliers using two complementary criteria. As a physical criterion, daily sunlight exposure was bounded by the maximum possible daylight duration in the tropics (approximately 12.5 hours per day); values exceeding this bound are physically impossible and therefore indicate data-entry errors. This screening flagged anomalous records such as Aceh in 2023 (89.91 hours) and the Special Region of Yogyakarta in 2019 (61.21 hours). As a statistical criterion, the interquartile range (IQR) method, when applied to rainfall identified one additional anomaly, namely Yogyakarta in 2020 (280.54 mm). In total, three records were identified as data-entry anomalies. these records were subsequently verified against the original BMKG source and corrected to their true values rather than discarded, thereby ensuring that the panel of 34 provinces over 2018–2023 remained complete at 204 observations. Third, regarding feature normalization, no feature scaling was applied to the primary Decision Tree model or to the tree-based benchmark models (Random Forest and XGBoost), because these algorithms partition the feature space using threshold-based splits and are therefore invariant to monotonic rescaling of the inputs. For the scale-sensitive benchmark models (Support Vector Regression and Linear Regression), features were standardized to have zero mean and unit variance. To prevent data leakage, the scaler was fitted only on the training folds within the cross-validation loop and then applied to the corresponding validation fold.

In the next step, the data containing the four combined variables from the provincial dataset are accessed and fed into the model, where they function as predictor variables. The following step is to determine the splitting criteria, during which the Decision Tree Regression algorithm identifies the most suitable feature and threshold to effectively divide the data at each node [24]. This process aims to reduce prediction errors. The decision tree regression algorithm examines a range of features and threshold values, ultimately choosing the most effective combination to partition the data accurately [25]. Following the selection of the optimal feature and threshold, the process advances to the node-splitting phase. During this phase, the dataset is partitioned into two branches based on the specified splitting criteria, facilitating a structured separation aligned with the model's decision rules. For instance, when rainfall exceeds a specified threshold, the data are assigned to the right branch; if they fall below or match this threshold, they are assigned to the left branch. This approach helps group the data into more homogeneous clusters, thereby enhancing prediction accuracy in later stages.

The following stage involves Recursive Splitting, where the data undergo repeated partitioning within each branch formed from earlier node splits. This recursive division persists until a specific stopping condition is satisfied. The stopping criteria may involve constraints such as a maximum tree depth, a minimum sample threshold at each node, or the point at which additional splits fail to produce meaningful reductions in prediction error [26]. Recursive splitting facilitates more granular data partitioning, enabling the model to discern complex patterns embedded within the dataset. Following the completion of the splitting process, the model proceeds to the leaf node stage. Leaf nodes are the terminal points of the decision tree, where no further division is possible. At each leaf node, the decision tree regression algorithm calculates the mean target value (e.g., rice production) from the data present in that node. The values computed at these leaf nodes represent the predictions produced by the decision tree. In essence, each leaf node contains a predicted value that the model applies when processing new data. In the prediction phase, the trained model is utilized to generate forecasts for new data inputs. Each data point is passed through the decision tree, moving through nodes based on predefined features and threshold values, until reaching a leaf node. The value at this leaf node provides the model's prediction for that particular data instance. This approach allows the model to accurately estimate rice production by analyzing essential factors, including harvested area, production volume, rainfall, and sunlight exposure duration.

3. Results and Discussion

http://doi.org/10.26594/register.v12i1.5352

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 58-70

3.1. Result of Benchmark Algorithm

To verify whether the choice of the Decision Tree model is truly appropriate, Table 2 presents the results of five algorithms evaluated on the same dataset and compared fairly using 5-fold cross-validation. In this procedure, the data is divided into five parts in rotation: four parts are used to train the model and the remaining part is used to test it, and this loops until every part has served as the test set. Four metrics are applied simultaneously: RMSE and MAE to assess the magnitude of error in tons, MAPE to measure the error in percentage terms, and R^2 to assess how much ofthe variation in production is explained by the model. The results reveal a consistent pattern. The two tree-based models, Decision Tree and Random Forest, achieved the best performance, with nearly identical error rates and a percentage error of approximately 6 percent. Notably, Random Forest, despite being considerably more complex, does not provide any meaningful improvement in accuracy over a single Decision Tree.

Conversely, Support Vector Regression (SVR) produced the largest error. This may be attributed to the wide variation in production across provinces, ranging from only thousands of tons in smaller provinces to millions of tons in the large provinces of Java. This type of model struggles to handle such extreme differences in scale. Linear Regression, meanwhile, appears strong when judged by its high R^2 value, but its MAPE is substantially higher. This finding highlights an important point: the high R^2 is largely driven by the large provinces, which are easy to predict, even though the model misses badly when predicting the small provinces. This is precisely why a single metric is never sufficient, and why error should be read in relative terms (percentage) rather than as absolute figures alone.

Table 2. Results of Benchmark Algorithm

Model	RMSE	MAE	MAPE	R^2
Linear Regression	206.879	112.557	84,7%	0,993
SVR	849.255	371.275	247,5%	0,887
Decision Tree (based)	175.868	83.742	6,2%	0,995
Random Forest	176.423	77.994	6,7%	0,995
XGBoost	349.639	117.222	17,4%	0,981

The benchmark results in Table 2 provides empirical evidence that overfitting is not a binding problem. Random Forest, an ensemble method specifically designed to reduce the variance and overfitting associated with individual decision trees, produced no meaningful accuracy gain over the tuned single tree (MAPE 6.7% vs 6.2%; identical R^2 of 0.995). Had the single tree been overfitting, the ensemble would have markedly outperformed it. The comparable performance therefore indicates that the regularized Decision Tree captures the underlying signal without excessive variance, while retaining the interpretability that ensembles sacrifice. Consequently, these benchmark resulrs support the selection of the Decision Tree as the primary predictive model: it achieves accuracy ranks among the best, on par with far more complex models, yet it is far simpler and more interpretable, a balance that is exactly what is most needed for a model intended to support food policy.

3.2. Result of Hyperparameter Tuning

In the implementation phase, the Decision Tree Regression model is developed to generate predictions of rice production for each of Indonesia's 34 provinces. Next, 30 hyperparameter tuning iterations are performed for each province. Upon completing these iterations, the researcher selects the predicted values that most closely align with the actual data from 2023. These optimized values are then used to forecast rice production in 2024. Hyperparameter tuning is beneficial for improving the performance of a machine learning model by optimizing hyperparameter values to match the characteristics of the data used [27], [28]. The hyperparameter tuning process starts by selecting the optimal model through cross-validation and refining key parameters. MaxDepth sets the tree's maximum depth, balancing complexity and the risk of overfitting. MinInstancesPerNode specifies the minimum number of data points needed to split a node, controlling node size to avoid excessive fragmentation. MaxBin defines

http://doi.org/10.26594/register.v12i1.5352

the maximum number of bins for categorizing numerical features, influencing the granularity of data splits. Adjusting these parameters ensures that the model is well-suited to the data, enhancing its predictive performance. Table 3 presents the optimal results from the experiments conducted with hyperparameter tuning across the provinces of Indonesia.

Table 3. Results of Hyperparameter Tuning

Province	maxDepth	minInstances PerNode	maxBins	RMSE	Best Parameter	Prediction in 2023	Actual in 2023
Aceh	[20, 25, 30]	[20, 30, 40]	[30, 40, 50]	196952,42	20,20,30	1619511,07	1393474,11
Sumatera Utara	[7, 12, 17]	[12, 22, 32]	[12, 17, 22]	74925,11	7,12,12	2079067,63	2080663,46
Sumatera Selatan	[1, 5, 10]	[1, 5, 10]	[2, 5, 10]	416272,13	1,1,2	2577919,72	762059,57
Sumatera Barat	[8, 13, 18]	[13, 23, 33]	[13, 18, 23]	114509,10	8,13,13	1431718,44	1457502,44
Bengkulu	[1, 5, 10]	[1, 5, 10]	[2, 5, 10]	12446,88	1,1,2	276363,65	277310,01
Riau	[1, 5, 10]	[1, 5, 10]	[2, 5, 10]	15365,93	1,1,2	215508,04	209190,02
Jambi	[1, 5, 10]	[1, 5, 10]	[2, 5, 10]	4311,01	1,1,2	293838,24	274557,09
Lampung	[27, 30, 30]	[27, 37, 47]	[37, 47, 57]	326530,33	27,27,37	2609030,52	2728780,6
Bangka Belitung	[29, 30, 30]	[29, 39, 49]	[39, 49, 59]	10470,83	29,29,39	60242,33	65500,85
Kalimantan Barat	[8, 13, 18]	[13, 23, 33]	[13, 18, 23]	91774,16	8,13,13	721561,91	688413,14
Kalimantan Timur	[1, 5, 10]	[1, 5, 10]	[2, 5, 10]	5252,66	1,5,2	244677,96	215290,58
Kalimantan Selatan	[1, 5, 10]	[1, 5, 10]	[2, 5, 10]	31450,62	1,1,2	984862,93	835282,46
Kalimantan Tengah	[2, 6, 11]	[2, 6, 11]	[2, 6, 11]	68456,64	2,2,2	412375,44	334732,63
Kalimantan Utara	[1, 5, 10]	[1, 5, 10]	[2, 5, 10]	8862,29	1,1,2	29967,31	24347,22
Banten	[3, 7, 12]	[3, 7, 12]	[3, 7, 12]	213192,5	3,3,3	1683695,85	1678765,59
DKI Jakarta	[24, 29, 30]	[24, 34, 44]	[34, 44, 54]	1292,81	24,24,34	3304,39	2803,24
Jawa Barat	[1, 5, 10]	[1, 5, 10]	[2, 5, 10]	485102,28	1,5,2	9162256,47	9095938,03
Jawa Tengah	[23, 28, 30]	[23, 33, 43]	[33, 43, 53]	646023,55	23,23,33	9422804,81	9061714,85
Jawa Timur	[21, 26, 30]	[21, 31, 41]	[31, 41, 51]	460083,47	21,21,31	9632345,85	9591422,32
Bali	[2, 6, 11]	[2, 6, 11]	[2, 6, 11]	54924,52	2,2,2	673835,33	668612,13
NTT	[23, 28, 30]	[23, 33, 43]	[33, 43, 53]	143766,85	23,23,33	756169,03	757505,4
NTB	[28, 30, 30]	[28, 38, 48]	[38, 48, 58]	94669,26	28,28,28	1444281,22	1546819,76
Sulawesi Barat	[29, 30, 30]	[29, 39, 49]	[39, 49, 59]	40273,73	29,29,39	309231,01	294026,68
Sulawesi Tengah	[9, 14, 19]	[14, 24, 34]	[14, 19, 24]	114835,01	9,14,14	812143,65	812948,49
Sulawesi Utara	[1, 5, 10]	[1, 5, 10]	[2, 5, 10]	60878,18	1,1,2	240882,12	230832,14

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 58-70

Province	maxDepth	minInstances PerNode	maxBins	RMSE	Best Paramet er	Prediction in 2023	Actual in 2023
Sulawesi Tenggara	[28, 30, 30]	[28, 38, 48]	[38, 48, 58]	18937,38	28,28,28	512513,69	482371,05
Sulawesi Selatan	[23, 28, 30]	[23, 33, 43]	[33, 43, 53]	688186,11	23,23,33	4881315,97	4943096,36
Maluku Utara	[28, 30, 30]	[28, 38, 48]	[38, 48, 58]	16302,05	28,28,28	30160,82	28168,81
Maluku	[7, 12, 17]	[12, 22, 32]	[12, 17, 22]	16084,46	7,12,12	100434,37	83065,17
Papua Barat	[20, 25, 30]	[20, 30, 40]	[30, 40, 50]	3433,27	20,20,30	25089,73	23808,11
Papua	[6, 11, 16]	[11, 21, 31]	[11, 16, 21]	51244,76	6,11,11	200670,91	200115,34

3.3. Results of Predicted Rice Production

This section analyzes the predicted rice prediction for 2024 , with prior testing conducted using 2023 data. Figure 2 provides a side-by-side comparison of actual production data from 2023 and predicted production for 2024. The close alignment between the two suggests that the model's predictions closely reflect the trends observed in the previous year. This comparison is presented for all 34 provinces, with Figure 2 also indicating which provinces are expected to experience increases or decreases in rice production.

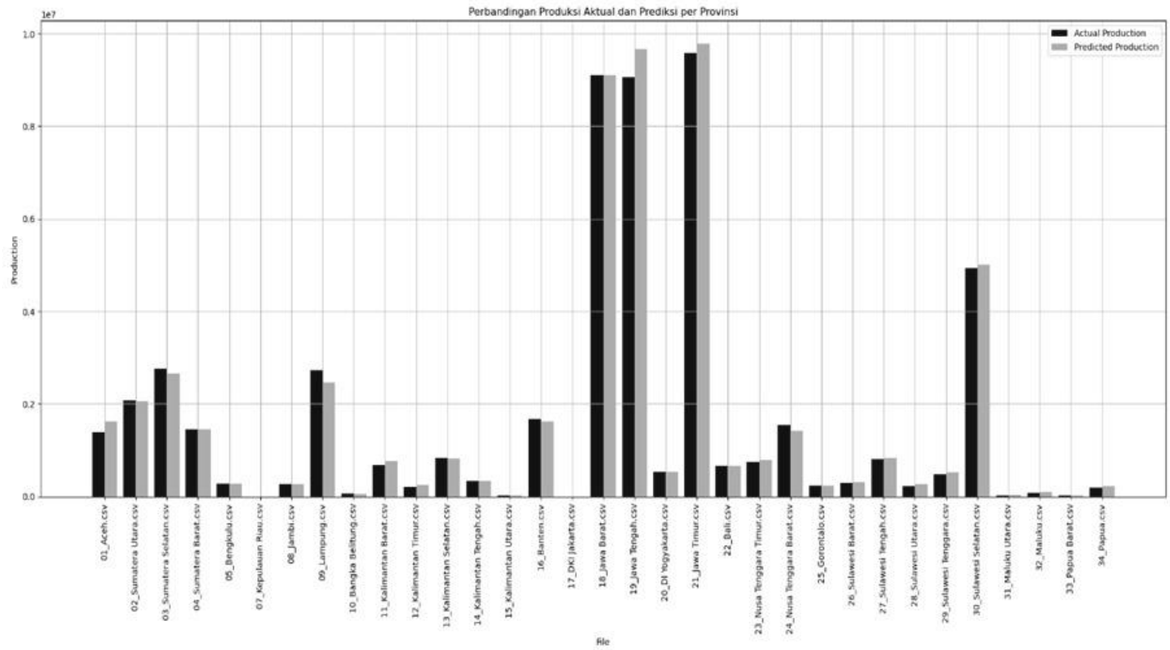

Fig 2. Graph of Predicted Rice Production Results for 2024

Table 4 provides the Root Mean Squared Error (RMSE) values for rice production prediction across Indonesia's 34 provinces. The findings indicate that the Riau Islands Province recorded the lowest RMSE value at 525.50, while the average RMSE across all provinces was 93,046.25. Comparing these RMSE values with the manually computed standard deviation of the actual 2023 data confirms that the 2024 prediction errors fall within an acceptable range, underscoring the model's accuracy.

http://doi.org/10.26594/register.v12i1.5352

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 58-70

Table 4. RMSE Results for 2024 Predictions

Province	RMSE	Province	RMSE	Province	RMSE	Province	RMSE
Aceh	186175,48	Jambi	36797,90	Kalimantan Utara	10336,86	Bali	47637,95
Sumatera Utara	21989,50	Lampung	203367,10	Banten	117880,67	Nusa Tenggara Timur	55905,68
Sumatera Selatan	245226,44	Bangka Belitung	10789,47	DKI Jakarta	1197,90	Nusa Tenggara Barat	35397,54
Sumatera Barat	110973,03	Kalimantan Barat	24574,52	Jawa Barat	208725,03	Gorontalo	6319,13
Bengkulu	5900,88	Kalimantan Timur	30467,28	Jawa Tengah	230004,38	Sulawesi Barat	29080,63
Riau	36222,83	Kalimantan Selatan	265060,93	DI Yogyakarta	33062,89	Sulawesi Tengah	72420,73
Kepulauan Riau	525,50	Kalimantan Tengah	149754,06	Jawa Timur	216321,68	Sulawesi Utara	38732,16
Sulawesi Tenggara	29091,95	Sulawesi Selatan	664026,44	Maluku Utara	9631,42	Maluku	6473,10
Papua Barat	2374,35	Papua	21127,37				
Average RMSE: 93046,25824							

3.4. Error Analysis

While Table 4 reports the absolute RMSE for each province, absolute error is strongly scale-dependent and therefore difficult to compare across provinces whose production differs by several orders of magnitude. To enable a fair comparison, Table 5 reports the relative RMSE, defined as the absolute RMSE divided by the province's 2023 production level. As shown in Table 5, this reframing reverses the apparent ranking: Sulawesi Selatan, which has the largest absolute RMSE (664,026 tons), achieves a low relative error of only 13.4%, whereas DKI Jakarta, with one of the smallest absolute RMSE values (1,198 tons), records a relative error of 42.7%. A clear structural pattern emerges. The high-volume producers of Java, Jawa Timur (2.26%), Jawa Barat (2.29%), and Jawa Tengah (2.54%), are among the most accurately predicted provinces in relative terms, because their large and stable production base provides a strong, consistent signal for the model to learn. In contrast, the largest relative errors occur in small or structurally atypical provinces such as Kalimantan Tengah (44.7%), DKI Jakarta (42.7%), and

http://doi.org/10.26594/register.v12i1.5352

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 58-70

Kalimantan Utara (42.5%), where production is low, volatile, or driven by non-climatic factors (e.g., the predominantly urban character of DKI Jakarta), meaning that even a modest absolute deviation translates into a large percentage error.

Table 5. RMSE Results for 2024 Predictions

Province	RMSE	Rel. RMSE (%)	Province	RMSE	Rel. RMSE (%)
Sumatera Utara	21.989,50	1,06	DI Yogyakarta	33.062,89	6,21
Bengkulu	5.900,88	2,13	Banten	117.880,67	7,02
Jawa Timur	216.321,68	2,26	Bali	47.637,95	7,12
Nusa Tenggara Barat	35.397,54	2,29	Nusa Tenggara Timur	55.905,68	7,38
Jawa Barat	208.725,03	2,29	Lampung	203.367,10	7,45
Jawa Tengah	230.004,38	2,54	Sumatera Barat	110.973,03	7,61
Kalimantan Barat	24.574,52	3,57	Maluku	6.473,10	7,79
Sulawesi Tenggara	29.091,95	6,03	Sumatera Selatan	245.226,44	8,88
Sulawesi Tengah	72.420,73	8,91	Bangka Belitung	10.789,47	16,47
Sulawesi Barat	29.080,63	9,89	Sulawesi Utara	38.732,16	16,78
Papua Barat	2.374,35	9,97	Riau	36.222,83	17,32
Papua	21.127,37	10,56	Kalimantan Selatan	265.060,93	31,73
Aceh	186.175,48	13,36	Maluku Utara	9.631,42	34,19
Jambi	36.797,90	13,40	Kalimantan Utara	10.336,86	42,46
Sulawesi Selatan	664.026,44	13,43	DKI Jakarta	1.197,90	42,73
Kalimantan Timur	30.467,28	14,15	Kalimantan Tengah	149.754,06	44,74

4. Conclusion

This study developed a province-level model for forecasting rice production across Indonesia's 34 provinces using Decision Tree Regression, integrating climate data (rainfall and sunlight exposure) from BMKG with agricultural statistics (harvested area and production) from BPS covering the period of 2018–2023. The model achieved an average RMSE of 93,046 tons. A benchmark against four alternative algorithms confirmed that the tuned Decision Tree achieved accuracy comparable to more complex ensembles such as Random Forest while remaining substantially more interpretable. Additionally, feature importance analysis identified harvested land area as the dominant predictor, with sunlight exposure and rainfall providing secondary refinement, a hierarchy consistent with the agronomic determinants of yield. A scale-independent error analysis further showed that the model is most reliable for the high-volume producers of Java (relative RMSE 2.3–2.5%), while larger relative errors are primarily observed in small or structurally volatile provinces. Beyond its predictive accuracy, the model offers practical operational value for food-policy stakeholders such as the National Logistics Agency (Bulog) and the Ministry of Agriculture. Delivered through an accessible web interface that displays annual production trends by province, the province-level forecasts can inform the timing and geographic allocation of rice procurement, the positioning of buffer stocks, and inter-provincial distribution planning before anticipated shortfalls, enabling a more pre-emptive rather than reactive response to supply risks. Importantly, the error analysis itself functions as a decision aid: the relative-error map distinguishes provinces whose forecasts can be relied upon directly, particularly the major

http://doi.org/10.26594/register.v12i1.5352

producing regions that dominate national supply, from provinces where the model is comparatively less certain. For the latter group, characterized by high relative error and structurally volatile production, the appropriate response is not to act on the point forecast alone but to prioritize these provinces for intensified ground-level monitoring, more frequent data updates, and larger precautionary buffers. Consequently, identifying where the model is least confident helps target surveillance and contingency resources, directly supporting the mitigation of localized food crises.

This study is nonetheless subject to certain temporal limitation. The six-year window (2018–2023), while sufficient to establish province-level predictive relationships, may not fully capture long-term macro-climatic cycles. In particular, the El Niño–Southern Oscillation (ENSO), through its El Niño and La Niña phases, exerts a strong and irregular influence on rainfall and drought across the archipelago, with documented effects on planting calendars and rice yields. The relationships learned by the model may therefore be partly conditioned on the specific ENSO states prevailing during the study period. Future research should therefore focus on several directions. Extending the temporal coverage to span multiple complete ENSO cycles while incorporating explicit climate indices such as the Oceanic Niño Index as additional predictors, would improve the model's robustness under extreme conditions. Expanding the model from one-year-ahead to multi-year forecasting would enhance its value for long-term agricultural planning, while further refinement aimed at reducing error in the structurally volatile provinces would improve precision across all regions. Such advances would further strengthen the model's role as a reliable, interpretable, and practical decision-support tool for national food-security planning.

Acknowledgment

We thank the Information Technology Department, ITDel and BMKG, Indonesia for their support. We are deeply grateful for the funding provided by LPPM Insitut Teknologi Del. The authors acknowledge the use of AI-based tools for language improvement and grammar checking. The AI system did not create, alter, or add to the article's substantive content; rather, it was used solely to enhance readability, coherence, and clarity.

References

[1] D. Chrisendo, H. Siregar, and M. Qaim, "Oil palm and structural transformation of agriculture in Indonesia," *Agricultural Economics*, vol. 52, no. 5, pp. 849–862, Sep. 2021, doi: 10.1111/AGEC.12658.

[2] F. Rozi *et al.*, "Indonesian market demand patterns for food commodity sources of carbohydrates in facing the global food crisis," *Heliyon*, vol. 9, no. 6, p. e16809, Jun. 2023, doi: 10.1016/J.HELIYON.2023.E16809.

[3] T. Sitaresmi *et al.*, "Advances in the development of rice varieties with better nutritional quality in Indonesia," *J. Agric. Food Res.*, vol. 12, p. 100602, Jun. 2023, doi: 10.1016/J.JAFR.2023.100602.

[4] K. E. Giller *et al.*, "The future of farming: Who will produce our food?," *Food Secur.*, vol. 13, no. 5, pp. 1073–1099, Oct. 2021, doi: 10.1007/S12571-021-01184-6/FIGURES/8.

[5] E. Grigorieva, A. Livenets, and E. Stelmakh, "Adaptation of Agriculture to Climate Change: A Scoping Review," *Climate 2023, Vol. 11, Page 202*, vol. 11, no. 10, p. 202, Oct. 2023, doi: 10.3390/CLI11100202.

[6] M. Joseph, S. Moonsammy, H. Davis, D. Warner, A. Adams, and T. D. Timothy Oyedotun, "Modelling climate variabilities and global rice production: A panel regression and time series analysis," *Heliyon*, vol. 9, no. 4, p. e15480, Apr. 2023, doi: 10.1016/J.HELIYON.2023.E15480.

[7] S. Saud *et al.*, "Comprehensive Impacts of Climate Change on Rice Production and Adaptive Strategies in China," *Front. Microbiol.*, vol. 13, p. 926059, Jun. 2022, doi: 10.3389/FMICB.2022.926059/BIBTEX.

[8] M. El Khayat, D. A. Halwani, L. Hneiny, I. Alameddine, M. A. Haidar, and R. R. Habib, "Impacts of Climate Change and Heat Stress on Farmworkers' Health: A Scoping Review," *Front. Public Health*, vol. 10, p. 782811, Feb. 2022, doi: 10.3389/FPUBH.2022.782811/FULL.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 58-70
[9] M. Soni, A. Y. Sheshukov, and J. Aguilar, "The critical role of temperature in determining optimal planting schedule for cotton: A review," *Agric. For. Meteorol.*, vol. 373, p. 110741, Oct. 2025, doi: 10.1016/J.AGRFORMET.2025.110741.

[10] C. Vishwakarma *et al.*, "Bioengineering of Canopy Photosynthesis in Rice for Securing Global Food Security: A Critical Review," *Agronomy 2023, Vol. 13, Page 489*, vol. 13, no. 2, p. 489, Feb. 2023, doi: 10.3390/AGRONOMY13020489.

[11] L. H. Phuoc *et al.*, "Rice Growth and Yield Responses to Climate Variabilities and Scenarios," *Trends in Sciences*, vol. 20, no. 2, p. 6390, Dec. 2022, doi: 10.48048/tis.2023.6390.

[12] S. Abbas, J. E. Nichol, and M. S. Wong, "Trends in vegetation productivity related to climate change in China's Pearl River Delta," *PLoS One*, vol. 16, no. 2, p. e0245467, Feb. 2021, doi: 10.1371/JOURNAL.PONE.0245467.

[13] P. A. Beding, F. Palobo, B. M. W. Tiro, R. H. S. Lestari, and M. K. Rumbarar, "Climate change impact on rice productivity in the rainfed Merauke District, Papua," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 648, no. 1, p. 012104, Feb. 2021, doi: 10.1088/1755-1315/648/1/012104.

[14] A. Durap, "A comparative analysis of machine learning algorithms for predicting wave runup," *Anthropocene Coasts*, vol. 6, no. 1, pp. 1–26, Dec. 2023, doi: 10.1007/S44218-023-00033-7/FIGURES/10.

[15] L. S. Cedric *et al.*, "Crops yield prediction based on machine learning models: Case of West African countries," *Smart Agricultural Technology*, vol. 2, p. 100049, Dec. 2022, doi: 10.1016/J.ATECH.2022.100049.

[16] M. Keerthana, K. J. M. Meghana, S. Pravallika, and M. Kavitha, "An ensemble algorithm for crop yield prediction," *Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021*, pp. 963–970, Feb. 2021, doi: 10.1109/ICICV50876.2021.9388479.

[17] Erlin, A. Yunianta, L. A. Wulandhari, Y. Desnelita, N. Nasution, and Junadhi, "Enhancing Rice Production Prediction in Indonesia Using Advanced Machine Learning Models," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3478738.

[18] L. S. Cedric *et al.*, "Crops yield prediction based on machine learning models: Case of West African countries," *Smart Agricultural Technology*, vol. 2, p. 100049, Dec. 2022, doi: 10.1016/J.ATECH.2022.100049.

[19] R. Aworka *et al.*, "Agricultural decision system based on advanced machine learning models for yield prediction: Case of East African countries," *Smart Agricultural Technology*, vol. 2, p. 100048, Dec. 2022, doi: 10.1016/J.ATECH.2022.100048.

[20] T. Hu *et al.*, "Climate change impacts on crop yields: A review of empirical findings, statistical crop models, and machine learning methods," *Environmental Modelling & Software*, vol. 179, p. 106119, Aug. 2024, doi: 10.1016/J.ENVSOFT.2024.106119.

[21] S. K. C. Kyire, J. K. M. Kuwornu, R. K. Bannor, E. K. Apiors, and E. Martey, "Perceived risk and risk management strategies under irrigated rice farming: Evidence from Tono and Vea irrigation schemes-Northern Ghana," *J. Agric. Food Res.*, vol. 12, p. 100593, Jun. 2023, doi: 10.1016/J.JAFR.2023.100593.

[22] D. Dorairaj and N. T. Govender, "Rice and paddy industry in Malaysia: governance and policies, research trends, technology adoption and resilience," *Front. Sustain. Food Syst.*, vol. 7, p. 1093605, Jun. 2023, doi: 10.3389/FSUFS.2023.1093605/BIBTEX.

[23] A. A. Mana, A. Allouhi, A. Hamrani, S. Rahman, I. el Jamaoui, and K. Jayachandran, "Sustainable AI-based production agriculture: Exploring AI applications and implications in agricultural practices," *Smart Agricultural Technology*, vol. 7, p. 100416, Mar. 2024, doi: 10.1016/J.ATECH.2024.100416.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
12 (1) 2026 58-70

[24] O. A. Montesinos López, A. Montesinos López, and J. Crossa, "Random Forest for Genomic Prediction," *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, pp. 633–681, 2022, doi: 10.1007/978-3-030-89010-0_15.

[25] A. Durap, "A comparative analysis of machine learning algorithms for predicting wave runup," *Anthropocene Coasts*, vol. 6, no. 1, pp. 1–26, Dec. 2023, doi: 10.1007/S44218-023-00033-7/FIGURES/10.

[26] W. Kanyongo and A. E. Ezugwu, "Feature selection and importance of predictors of non-communicable diseases medication adherence from machine learning research perspectives," *Inform. Med. Unlocked*, vol. 38, p. 101232, Jan. 2023, doi: 10.1016/J.IMU.2023.101232.

[27] T. A. Prasetyo, T. L. Gaol, N. F. Sipahutar, T. Siahaan, T. E. Manik, and R. Chandra, "Refining tomato disease recognition: hyperparameter tuning on ResNet-101 architecture for precise leaf-based classification," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 2, pp. 1204–1213, May 2024, doi: 10.11591/IJEECS.V34.I2.PP1204-1213.

[28] T. A. Prasetyo, V. L. Desrony, H. F. Panjaitan, R. Sianipar, and Y. Pratama, "Corn Plant Disease Classification Based on Leaf using Residual Networks-9 Architecture," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 3, pp. 2908–2920, Jun. 2023, doi: 10.11591/IJECE.V13I3.PP2908-2920.