

Comparative Performance of SARIMA, Decision Tree, and Hybrid Models for Climate-Informed Rice Production Forecasting

Dewi Arthasari Prescelya¹, Mahmudi², Nur Inayah³

¹⁻³ Universitas Islam Negeri Syarif Hidayatullah Jakarta, Indonesia

✉ dewi1artha1@gmail.com

Received: February 7, 2026 | **Revised:** May 20, 2026 | **Accepted:** May 30, 2026 | **Published:** July 30, 2026

ABSTRACT Rice production is essential to national food security, yet climate-related fluctuations create uncertainty in agricultural planning. This study compared Seasonal Autoregressive Integrated Moving Average (SARIMA), Decision Tree, and hybrid SARIMA--Decision Tree models for forecasting monthly rice production in Bojonegoro, Lamongan, and Ngawi Regencies. The dataset comprised 84 monthly observations for each regency from January 2018 to December 2024, including rice production, rainfall, temperature, and sunshine duration. The procedure included Augmented Dickey--Fuller testing, seasonal decomposition, SARIMA parameter optimization, Decision Tree tuning, and residual correction, in which climate variables and temporal features were used to model SARIMA errors. The dominant predictors were sunshine duration in Bojonegoro and temperature in Lamongan and Ngawi. The hybrid model produced the lowest mean absolute percentage error in all three regencies: 32.23% in Ngawi, 36.76% in Bojonegoro, and 46.18% in Lamongan. Although the hybrid approach consistently improved upon the standalone models, the remaining errors indicate limited absolute forecasting accuracy, particularly in Lamongan. The findings support the residual-based integration of seasonal and non-linear models while highlighting the need for richer, location-specific climate and agricultural data.

Keywords: rice production forecasting, SARIMA, Decision Tree, hybrid model, climate variables

INTRODUCTION

East Java is Indonesia's largest rice-producing province. In 2023, the province produced approximately 9.71 million tonnes of dry unhusked rice and contributed 17.98% of national production (BPS, 2024). Lamongan, Ngawi, and Bojonegoro are among its principal rice-producing regencies. However, monthly production varies substantially because planting and harvesting cycles interact with rainfall, temperature, sunshine duration, irrigation, harvested area, and farm management. Climate variability can therefore complicate production planning and reduce the reliability of forecasts based solely on historical averages.

These data combine strong seasonal dependence with potentially nonlinear climatic effects. SARIMA is appropriate for periodic time-series patterns because it explicitly represents nonseasonal and seasonal autoregressive, differencing, and moving-average components (Divisekara et al., 2021). Its linear structure, however, may not adequately represent threshold effects or interactions among climate variables. Decision Trees can model nonlinear relationships and identify influential predictors, but they do not inherently account for temporal autocorrelation (Demirović and Stuckey, 2021). A residual-based hybrid model can combine these complementary strengths by using SARIMA to represent the temporal structure and a Decision Tree to correct nonlinear errors.

Previous studies support the individual components of this approach. A hybrid SARIMA–Decision Tree model was used for drought forecasting in Tegal and achieved strong classification performance (Yasnita et al., 2020). For rice production in South Sulawesi, SARIMA outperformed an Extreme Learning Machine in terms of MAPE (Jamal et al., 2025). SARIMA has also been applied to monthly rice production in West Nusa Tenggara (Kusuma et al., 2024), whereas transfer-function modeling has demonstrated the value of incorporating harvested area as an explanatory variable (Yulianto and Najib, 2021). In another domain, comparative evidence indicates that seasonal statistical models and Decision Trees offer different advantages: the former represent temporal patterns, whereas the latter provide computationally efficient nonlinear rules (Kyryk and Shatalov, 2024).

Despite this progress, few studies have compared these methods across neighboring rice-producing regions that share a strong seasonal cycle but may respond differently to the same observed climate signals. This study therefore compares SARIMA, Decision Tree, and hybrid SARIMA–Decision Tree models using monthly data from Bojonegoro, Lamongan, and Ngawi. Its objectives are to identify the best-performing model in each regency, examine the relative importance of climate variables, and assess whether nonlinear residual correction improves seasonal forecasts. The study also explicitly considers the limitations associated with the small sample and the use of climate observations from a common station.

RESEARCH METHOD

Data Sources and Research Variables

The study used monthly rice production and climate data from January 2018 to December 2024, yielding 84 observations for each regency. Rice production, measured in tonnes, was obtained from Statistics Indonesia (BPS) of East Java Province, while rainfall, mean temperature, and sunshine duration were obtained from the Meteorology, Climatology, and Geophysics Agency (BMKG). The official data portals were <https://jatim.bps.go.id/> and <https://dataonline.bmkg.go.id/dataonline-home>.

Because complete station-level climate records were unavailable for every regency, a single series of climate observations from one station was matched to all three production series. Consequently, differences in feature importance should not be interpreted as evidence that the regencies experienced different measured climate conditions. Instead, they indicate differences in the modeled sensitivity of each production series to the common climate covariates, potentially reflecting variations in local irrigation, crop varieties, planting calendars, cultivated area, and management practices.

Data Preprocessing

Missing climate observations were imputed using the median because it is less sensitive to extreme values than the mean. Isolated, unusually low rice-production values were also replaced through linear interpolation to reduce instability in percentage-based error metrics. The dates were converted into a monthly time index, and each series was divided chronologically into 70% training data and 30% testing data. Although this procedure preserved the temporal order, modifying low production values may affect the resulting accuracy estimates and is therefore treated as a methodological limitation.

SARIMA Model

The seasonal autoregressive integrated moving-average model is denoted by

$$\text{SARIMA}(p, d, q)(P, D, Q)_s, \quad (1)$$

where p , d , and q are the nonseasonal autoregressive, differencing, and moving-average orders; P , D , and Q are their seasonal counterparts; and $s = 12$ represents monthly seasonality. The model can be written as

$$\Phi_P(B^s)\phi_p(B)(1 - B)^d(1 - B^s)^D Z_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t, \quad (2)$$

where B is the backshift operator, Z_t is the observation at time t , and ε_t is the innovation term (Buntara et al., 2023).

Candidate models were identified using the Augmented Dickey–Fuller test, seasonal decomposition, and the autocorrelation and partial autocorrelation functions. A grid search evaluated combinations of nonseasonal and seasonal orders. For each regency, the model with the smallest Akaike information criterion (AIC) was selected, with the Bayesian information criterion (BIC) used as a supporting criterion.

Forecast performance was summarized using root mean squared error, mean absolute error, mean absolute percentage error, and the coefficient of determination. Model-selection criteria were calculated using AIC and BIC:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}, \quad \text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|, \quad (3)$$

$$\text{MAPE} = \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|, \quad R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad (4)$$

$$\text{AIC} = 2k + n \ln(\text{SSE} / n), \quad \text{BIC} = k \ln n + n \ln(\text{SSE} / n), \quad (5)$$

where n is the number of observations and k is the number of estimated parameters (Pangestu et al., 2025).

Decision Tree Model

The Decision Tree regressor recursively partitions the predictor space into increasingly homogeneous regions and assigns a numerical prediction to each terminal node. In this study, the predictors comprised rainfall, temperature, sunshine duration, and temporal lag features, while rice production was the response variable. Hyperparameters were selected through a grid search with five-fold cross-validation, using negative MAPE as the scoring function. The search covered the maximum tree depth, the minimum number of samples required to split a node, the minimum number of samples per leaf, and the number of features considered at each split. Feature importance was calculated from the accumulated reduction in node impurity (Hamid et al., 2025).

Hybrid SARIMA--Decision Tree Model

The hybrid model was estimated in two stages. First, SARIMA generated the baseline forecast and residual

$$e_t = y_t - \hat{y}_t^{\text{SARIMA}}. \quad (6)$$

Second, a Decision Tree was trained to estimate e_t from climate variables, temporal lags, and interaction features. The final forecast was

$$\hat{y}_t^{\text{Hybrid}} = \hat{y}_t^{\text{SARIMA}} + \hat{e}_t^{\text{DT}}. \quad (7)$$

This architecture allows the statistical model to represent linear seasonal dependence, while the tree corrects nonlinear patterns remaining in the residuals. It follows the general residual-correction principle used in hybrid ARIMA–Decision Tree modeling (Fuad et al., 2019).

RESULTS AND DISCUSSION

Figure 1 shows substantial monthly variation in rice production across the three regencies. Recurrent peaks indicate a strong annual harvesting cycle, while differences in peak magnitude suggest that a constant seasonal adjustment alone would be inadequate.

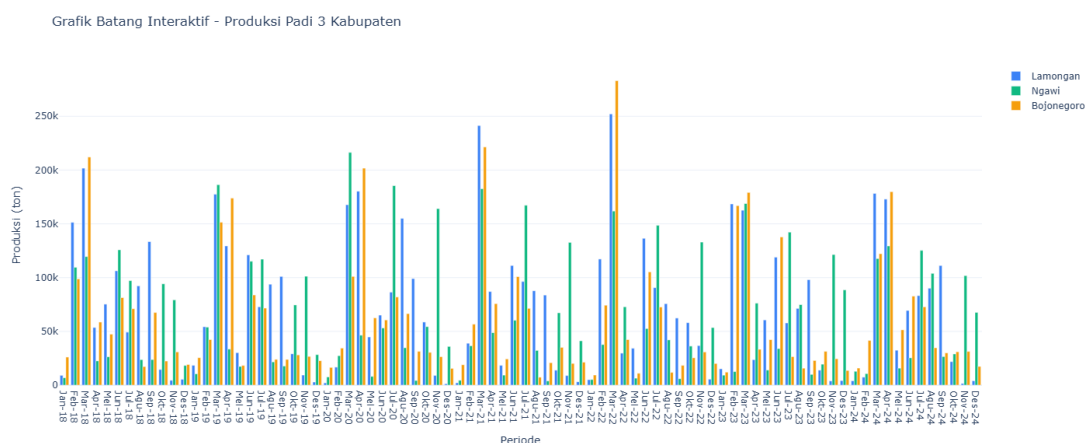


Figure 1. Monthly rice production in Bojonegoro, Lamongan, and Ngawi

Figure 2 presents the common climate series used in the analysis. Rainfall exhibits the greatest short-term variation, temperature remains comparatively stable, and sunshine duration follows a visible seasonal pattern. Because the same station series was used for all locations, the figure represents shared covariate inputs rather than separate local climate histories.

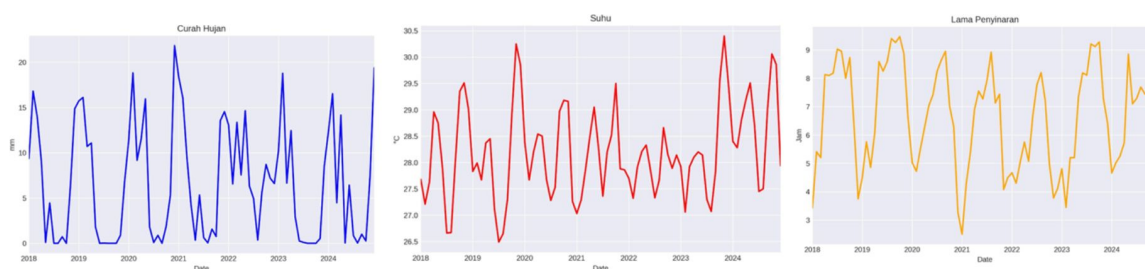


Figure 2. Monthly rainfall, temperature, and sunshine duration

Table 1. Descriptive statistics for rice production and climate variables

Variable	Regency	Mean	SD	Minimum	Maximum
Rice production (tonnes)	Bojonegoro	59,146.2	56,776.9	7,328	283,043
	Lamongan	80,604.4	55,468.3	13,852	252,152
	Ngawi	64,774.9	55,383.8	3,881	216,342
Rainfall (mm)	Bojonegoro	6.67	6.37	0	21.84
	Lamongan	6.67	6.37	0	21.84
	Ngawi	6.67	6.37	0	21.84
Temperature (°C)	Bojonegoro	28.19	0.87	26.49	30.40
	Lamongan	28.19	0.87	26.49	30.40
	Ngawi	28.19	0.87	26.49	30.40
Sunshine duration (hours)	Bojonegoro	6.60	1.81	2.50	9.47
	Lamongan	6.60	1.81	2.50	9.47
	Ngawi	6.60	1.81	2.50	9.47

Lamongan had the highest mean monthly rice production, followed by Ngawi and Bojonegoro. All three production series had large standard deviations relative to their means, confirming substantial temporal variation. The identical climate statistics across the regencies are a direct consequence of using data from the same station and should be considered when interpreting location-specific model results.

Table 2. Pearson correlations between climate variables and rice production

Regency	Rainfall	Temperature	Sunshine duration
Bojonegoro	0.1853	−0.1544	−0.1438
Lamongan	0.0851	−0.3068	−0.0773
Ngawi	−0.0899	−0.1044	−0.0015

The correlations ranged from −0.3068 to 0.1853, indicating weak marginal linear relationships. This finding does not imply that climate is unimportant because lagged, non-linear, and interaction effects may still be present. Instead, it provides an empirical rationale for testing a nonlinear learner alongside SARIMA.

Table 3. Augmented Dickey–Fuller stationarity test results

Regency	ADF statistic	<i>p</i> -value	Conclusion	Differencing
Bojonegoro	−8.38	0.00	Stationary	0
Lamongan	−2.04	0.26	Nonstationary	1
Ngawi	−2.26	0.18	Nonstationary	1

The Bojonegoro series was stationary in levels, whereas the Lamongan and Ngawi series required first-order differencing. These statistical differences describe the observed series but do not, by themselves, identify the agricultural mechanisms responsible for the patterns.

Table 4. Residual variance under additive and multiplicative decomposition

Regency	Additive variance	Multiplicative variance	Selected form
Bojonegoro	1,053,716,111.17	0.20	Multiplicative
Lamongan	1,048,419,149.63	0.20	Multiplicative
Ngawi	484,473,508.72	0.18	Multiplicative

The multiplicative specification produced lower residual variance for all three series. This result indicates that the magnitude of seasonal variation changed with the production level, thereby supporting a logarithmic transformation before seasonal modeling.



Figure 3. Multiplicative decomposition of monthly rice production

Figure 3 confirms the presence of recurrent annual seasonal components in each regency. Lamongan and Bojonegoro also display pronounced irregular peaks, whereas Ngawi shows a more regular seasonal sequence. Nevertheless, the residual panels indicate that decomposition alone does not account for all extreme movements.



Figure 4. ACF and PACF plots for Bojonegoro

For Bojonegoro, the ACF and PACF plots show seasonal signals at approximately multiples of 12 and short-run dependence at early lags. These patterns informed the selection of candidate nonseasonal moving-average and seasonal components.

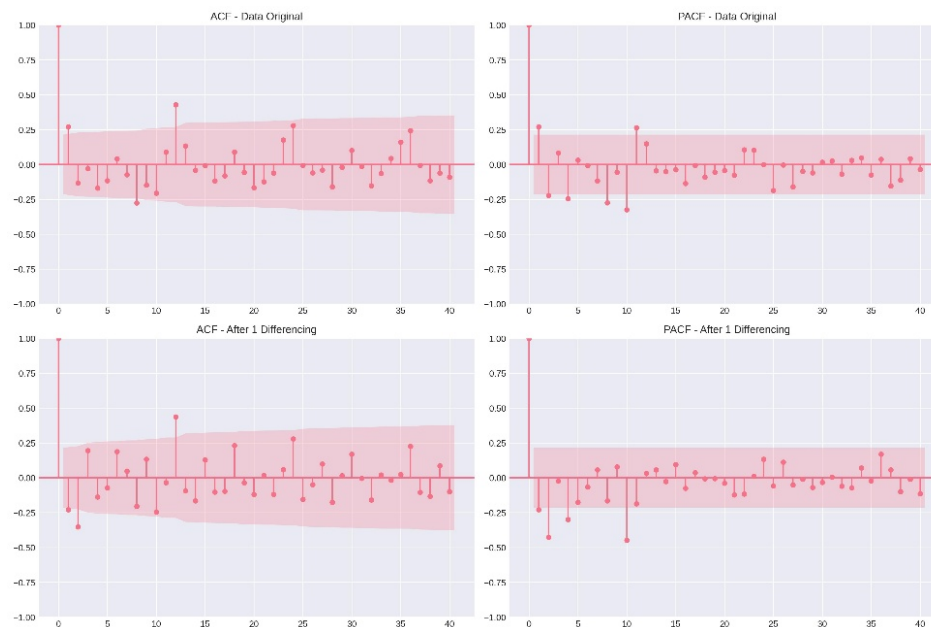


Figure 5. ACF and PACF plots for Lamongan

The Lamongan series required first-order differencing, after which most correlations moved closer to the confidence limits. The remaining early-lag and seasonal dependence supported a model containing both autoregressive and moving-average terms.

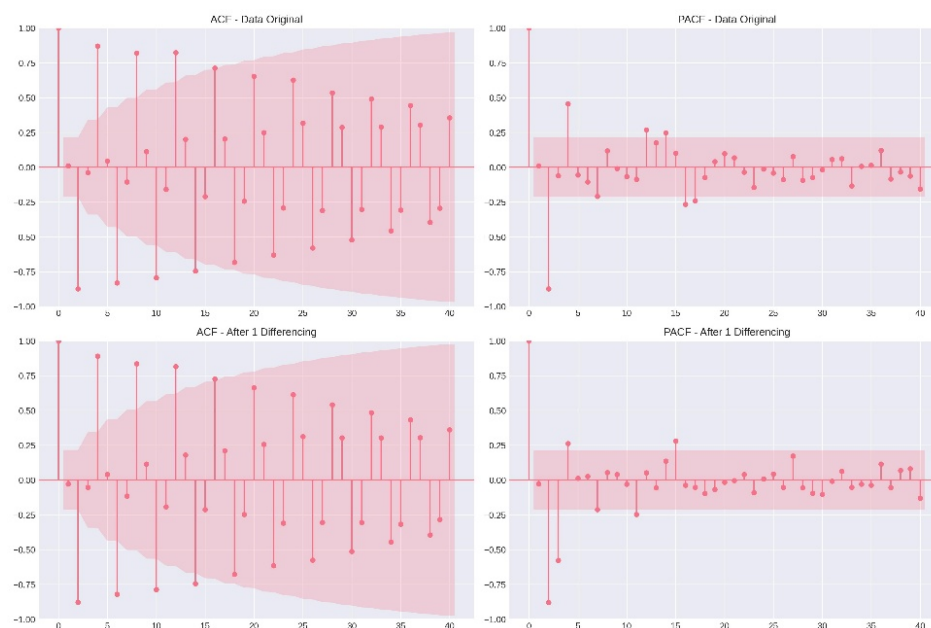


Figure 6. ACF and PACF plots for Ngawi

The Ngawi series also retained a clear seasonal pattern after first-order differencing. Its correlation structure motivated a comparatively richer nonseasonal specification, although the final model was selected using information criteria rather than visual inspection alone.

Table 5. Selected SARIMA parameters

Regency	(p, d, q)	(P, D, Q, s)	AIC	BIC
Bojonegoro	(0, 0, 2)	(1, 1, 1, 12)	58.49	65.66
Lamongan	(2, 1, 1)	(1, 1, 0, 12)	67.87	75.04
Ngawi	(2, 1, 2)	(1, 1, 0, 12)	27.98	36.58

The selected specifications differed across the regencies, reflecting heterogeneous temporal dynamics. Bojonegoro did not require regular differencing, whereas Lamongan and Ngawi did. Information criteria can be compared among candidate models fitted to the same series, but they should not be used to rank model quality across different response series.

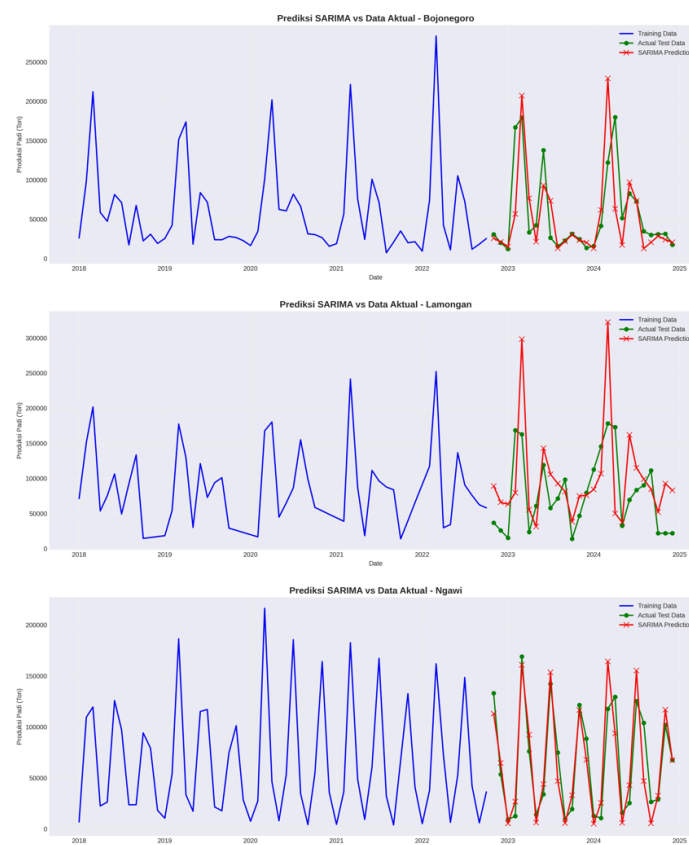


Figure 7. SARIMA predictions and observed rice production

Figure 7 shows that SARIMA reproduced the recurring seasonal cycle but frequently underestimated sharp harvest peaks. The discrepancy was greatest in Lamongan, consistent with its relatively high test MAPE. These remaining peak errors provide a practical target for nonlinear correction.

Table 6. Selected Decision Tree hyperparameters

Regency	Maximum depth	Minimum split	Minimum leaf	Maximum features
Bojonegoro	10	2	1	sqrt
Lamongan	7	2	2	sqrt
Ngawi	7	5	1	sqrt

The selected trees varied in complexity. The Bojonegoro model permitted the greatest tree depth, whereas stronger minimum-sample restrictions regularized the Ngawi model. Given the limited number of monthly observations, these settings should be interpreted cautiously because tree-based models can be sensitive to the particular training period.

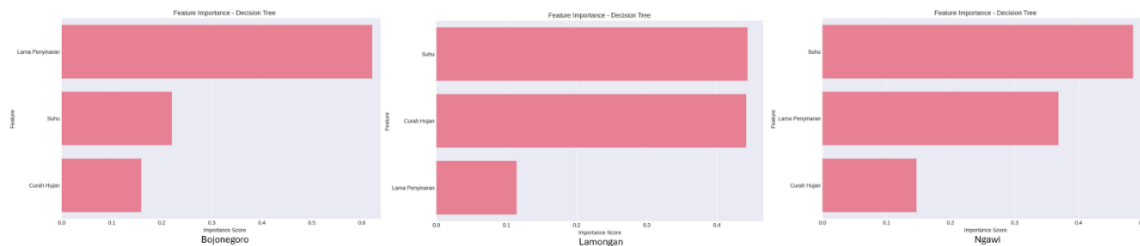


Figure 8. Decision Tree feature importance by regency

Sunshine duration had the highest tree-based feature importance in Bojonegoro, whereas temperature was the most important feature in Lamongan and Ngawi. Feature importance is model-dependent and does not establish causality. Moreover, because the climate inputs were identical across the three regencies, these differences indicate how each production series was partitioned in relation to the common predictors rather than measured differences in local weather. The prominence of temperature is broadly consistent with evidence that warming can reduce the yields of major crops, including rice (Zhao et al., 2017).

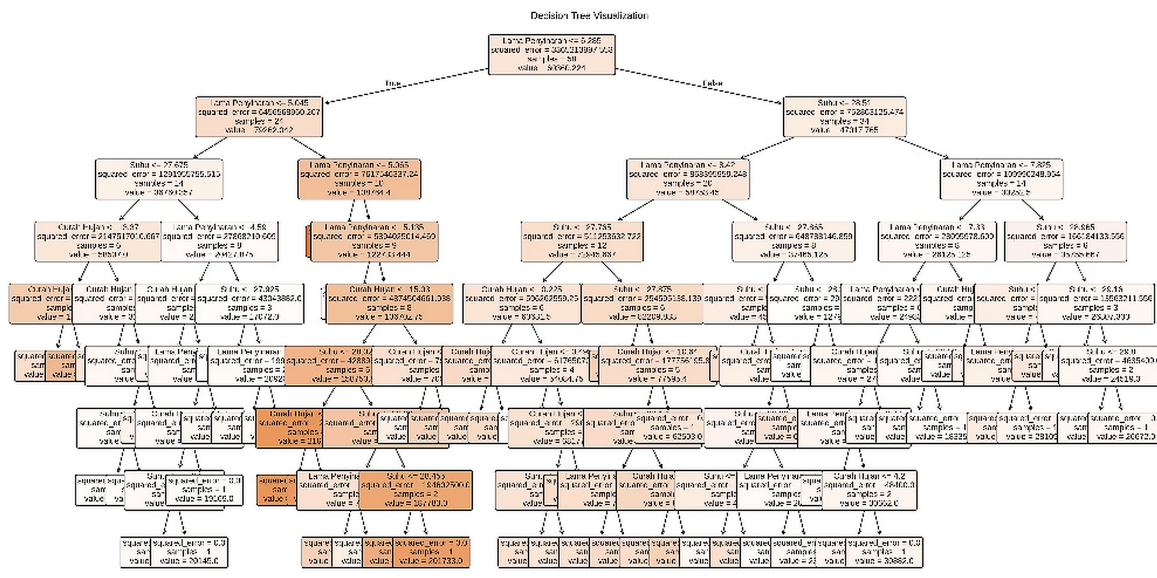


Figure 9. Decision Tree structure for Bojonegoro

The Bojonegoro tree contains numerous terminal nodes, consistent with the selected maximum depth of 10. This complexity allows the model to form flexible local rules, but it also increases the possibility of fitting idiosyncratic variation in the training sample.

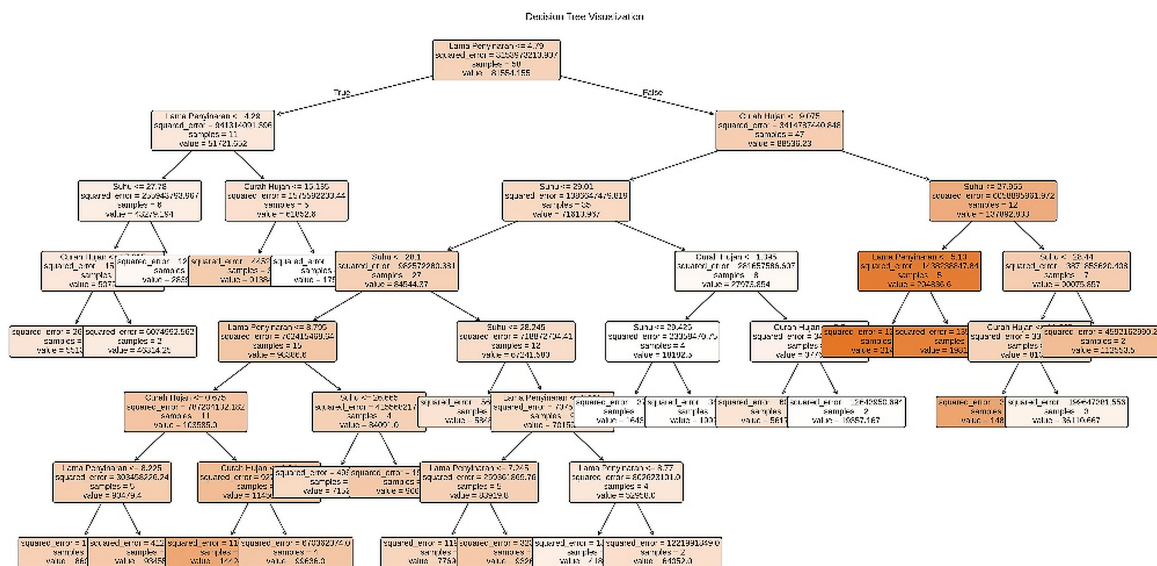


Figure 10. Decision Tree structure for Lamongan

The Lamongan tree is shallower and requires a minimum of two observations per leaf. This structure provides some regularization, although the standalone tree still lacks an explicit mechanism for preserving seasonal continuity.

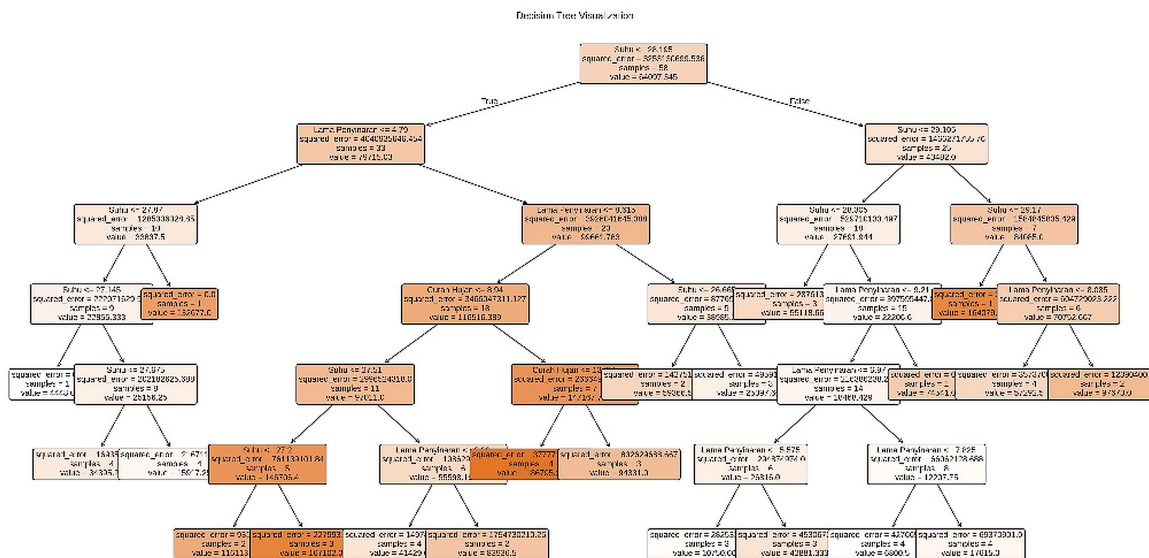


Figure 11. Decision Tree structure for Ngawi

The Ngawi tree applies the strongest restriction on split size. Even with this constraint, its standalone predictions generalized poorly, indicating that the climate and lag features alone did not adequately represent the seasonal structure of the test period.

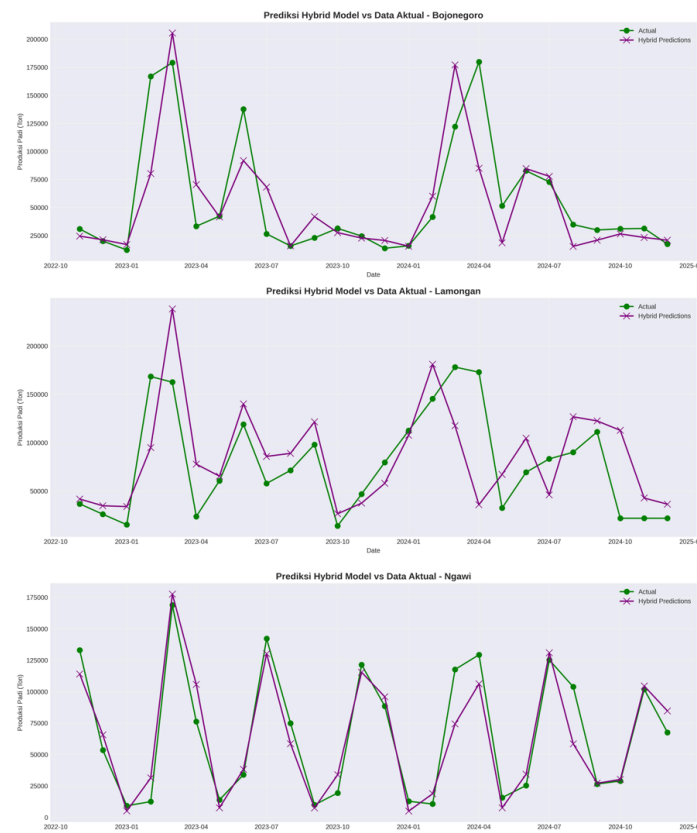


Figure 12. Hybrid-model predictions and observed rice production

Figure 12 shows that adding Decision Tree residual correction generally brought the predictions closer to the observed values. The improvement was most visible during periods in which SARIMA produced systematic peak errors. Nevertheless, several large discrepancies remained, especially in Lamongan, demonstrating that relative improvement did not translate into high absolute accuracy.

Table 7. Test-set MAPE for the three forecasting models

Regency	Model	MAPE (%)
Bojonegoro	SARIMA	40.09
	Decision Tree	56.02
	Hybrid SARIMA–DT	36.76
Lamongan	SARIMA	59.46
	Decision Tree	52.58
	Hybrid SARIMA–DT	46.18
Ngawi	SARIMA	41.18
	Decision Tree	196.33
	Hybrid SARIMA–DT	32.23

The hybrid model achieved the lowest MAPE in all three regencies. Compared with SARIMA, the hybrid model reduced MAPE by 3.33 percentage points in Bojonegoro, 13.28 points in Lamongan, and 8.95 points in Ngawi. The standalone Decision Tree performed particularly poorly in Ngawi, reinforcing the importance of explicitly modeling seasonality. However, the hybrid MAPE values, which ranged from 32.23% to 46.18%, remain substantial. Therefore, the results identify the hybrid model as the best of the three tested alternatives, but not as a fully accurate operational forecasting system.

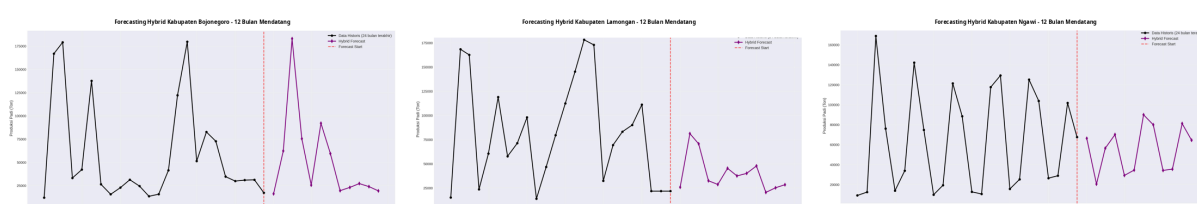


Figure 13. Twelve-month forecasts from the compared models

The projections in Figure 13 preserve the seasonal patterns identified in the historical series. Bojonegoro shows an early peak followed by a decline, Lamongan displays moderate variation, and Ngawi retains the most regular fluctuations. Because the analysis does not specify how future climate covariates were generated, these values should be interpreted as model-based scenarios rather than definitive operational forecasts.

This study has several limitations. The sample contains only 84 observations for each regency, climate observations from the same station were used for all regencies, important agricultural predictors such as harvested area and irrigation were omitted, and low production values were interpolated. In addition, five-fold cross-validation of a flexible tree using a short time series may produce unstable hyperparameters. These limitations may have contributed to the high test errors and may restrict the generalizability of the inferred climate sensitivities.

CONCLUSION AND RECOMMENDATIONS

The hybrid SARIMA–Decision Tree model produced the lowest test MAPE in Bojonegoro, Lamongan, and Ngawi. SARIMA successfully captured recurring temporal and seasonal patterns, while the Decision Tree residual-correction stage accounted for part of the remaining nonlinear variation associated with climate and lagged features. The hybrid MAPE values were 36.76% for Bojonegoro, 46.18% for Lamongan, and 32.23% for Ngawi. Thus, hybridization improved comparative performance, but the remaining errors are too large to support a claim of consistently high forecasting accuracy.

Future research should use climate observations from stations representative of each regency and incorporate harvested area, irrigation, humidity, soil conditions, crop varieties, fertilizer use, and planting-calendar information. Longer monthly series, robust treatment of extreme values, and time-series-aware cross-validation are also recommended. Comparative testing with SARIMAX, Random Forest, gradient boosting, and recurrent or attention-based models could determine whether richer exogenous structures produce more stable forecasts. Before the model is used for policy or operational planning, its performance should be validated prospectively using future production observations.

Artificial Intelligence Statement: The authors declare that no generative artificial intelligence or AI-assisted technologies were used in conducting the research or preparing the submitted manuscript

Author Contribution Statement: DAP: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Visualization. M: Validation, Writing – review & editing, Supervision. NI: Validation, Resources, Data curation, Supervision, Project administration.

Funding Statement: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest Statement: The authors declare no conflict of interest

REFERENCES

- Badan Pusat Statistik Provinsi Jawa Timur. 2024. *Indikator Pertanian Provinsi Jawa Timur 2023*. Badan Pusat Statistik Provinsi Jawa Timur. <https://jatim.bps.go.id/id/publication/2024/11/22/d5055617ff89f10adc29dc93/indikator-pertanian-provinsi-jawa-timur-2023.html>.
- Buntara, M. S., Napitupulu, H., & Gusriani, N. 2023. Pemrograman Python untuk peramalan data deret waktu menggunakan metode Seasonal Autoregressive Integrated Moving Average (SARIMA). *In Search*, 22(2), 354–362. <https://doi.org/10.37278/insearch.v22i2.774>.
- Demirović, E., & Stuckey, P. J. 2021. Optimal decision trees for nonlinear metrics. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5), 3733–3741. <https://doi.org/10.1609/aaai.v35i5.16490>.
- Divisekara, R. W., Jayasinghe, G. J. M. S. R., & Kumari, K. W. S. N. 2021. Forecasting the red lentils commodity market price using SARIMA models. *SN Business & Economics*, 1(1), Article 20. <https://doi.org/10.1007/s43546-020-00020-x>.
- Fuad, I. I. M., Rosli, L. H., & Husni, H. 2019. Application of hybrid ARIMA-Decision Tree algorithm in reservoir pressure surveillance. *Abu Dhabi International Petroleum Exhibition and Conference*. <https://doi.org/10.2118/197762-MS>.
- Hamid, I. Y., Yousif, S. M. A., Aljaloud, S., Abaker, A. O. I., Elemam, H. Z. S., Alruwaitee, K. A., & Anja, M. I. A. 2025. Machine learning techniques in a hybrid forecasting model for oil prices combining ARIMA. *Applied Mathematics & Information Sciences*, 19(6), 1241–1252. <https://doi.org/10.18576/amis/190601>.
- Jamal, R., Baso, A. M. A., Febriyanti, A., Sahriman, S., Siswanto, S., Yunita, A. I., Angriany, A. M. N., Rahim, R., & Fadil, M. 2025. Perbandingan metode Seasonal ARIMA dan Extreme Learning Machine dalam prediksi produksi padi di Sulawesi Selatan. *ESTIMASI: Journal of Statistics and Its Application*, 6(2), 264–278. <https://doi.org/10.20956/ejsa.v6i2.45821>.
- Kusuma, W., Setiawan, R. N. S., & Widiyanti, N. M. N. Z. 2024. Model Seasonal Autoregressive Integrated Moving Average (SARIMA) untuk memprediksi jumlah produksi beras di Provinsi Nusa Tenggara Barat. *Agroteksos*, 34(3), 830–839. <https://doi.org/10.29303/agroteksos.v34i3.1214>.
- Kyryk, V., & Shatalov, Y. 2024. SARIMA(X) and Decision Tree models comparison for load forecasting in electrical grids. *Modern Engineering and Innovative Technologies*, 1(33-01), 12–17. <https://doi.org/10.30890/2567-5273.2024-33-00-008>.

- Pangestu, B. M., Kamarudin, K., & Windarsyah, W. 2025. Penerapan model Seasonal Autoregressive Integrated Moving Average untuk prediksi hasil panen jamur tiram berbasis data historis. *Digital Transformation Technology*, 5(1), 133--142. <https://doi.org/10.47709/digitech.v5i1.5903>.
- Yasnita, Sutoyo, E., & Musnansyah, A. 2020. A hybrid of Seasonal Autoregressive Integrated Moving Average (SARIMA) and Decision Tree for drought forecasting. *Proceedings of the 2020 International Conference on Engineering and Information Technology for Sustainable Industry*. <https://doi.org/10.1145/3429789.3429870>.
- Yulianto, S., & Najib, A. J. 2021. Perbandingan metode SARIMA dan metode transfer pada produksi padi di Kabupaten Kendal. *Seminar Nasional Official Statistics*, 2020(1), 582--590. <https://doi.org/10.34123/semnasoffstat.v2020i1.433>.
- Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., Huang, M., Yao, Y., Bassu, S., Ciais, P., Durand, J.-L., Elliott, J., Ewert, F., Janssens, I. A., Li, T., Lin, E., Liu, Q., Martre, P., Müller, C., et al. 2017. Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences*, 114(35), 9326--9331. <https://doi.org/10.1073/pnas.1701762114>.