A Hybrid ARIMA–LSTM Model Optimized with a Tree-Structured Parzen Estimator for Automatic Hyperparameter Tuning to Forecast Indonesia's Chili and Climate Data

Irwan Sembiring
Universitas Kristen Satya Wacana, Indonesia
irwan@uksw.edu (corresponding author)

Ifan Prihandi
Universitas Kristen Satya Wacana, Indonesia
982022023@student.uksw.edu (corresponding author)

Sutarto Wijono
Universitas Kristen Satya Wacana, Indonesia
sutarto.wijono@uksw.edu

Evi Maria
Universitas Kristen Satya Wacana, Indonesia
evi.maria@uksw.edu

Received: 31 October 2025 | Revised: 28 December 2025 and 23 January 2026 | Accepted: 25 January 2026

 DOI: https://doi.org/10.48084/etasr.15911

ABSTRACT

Red chili peppers are a strategic agricultural commodity in Indonesia, yet their prices and production are highly volatile due to climate variability and supply chain disruptions. Accurate forecasting is therefore essential to support agricultural planning and price stabilization policies. This study proposes an automated forecasting framework based on a hybrid Autoregressive Integrated Moving Average–Long Short-Term Memory (ARIMA–LSTM) computational modeling approach that integrates statistical and deep learning methods while incorporating weather variables, namely rainfall and sunshine duration. The main novelties of this research are the application of Tree-Structured Parzen Estimator (TPE) for efficient automated hyperparameter optimization, improving upon conventional Grid Search (GS), and the construction of a hybrid model capable of capturing both linear and nonlinear temporal patterns. The framework is evaluated using time-series data from North Sumatra, Central Java, and Bali covering the period from March 2021 to December 2023. The results show that the proposed hybrid ARIMA–LSTM model significantly outperforms standalone and benchmark models. For retail price forecasting, the hybrid model achieves a Mean Absolute Percentage Error (MAPE) of 1.45%, compared to 42.74% for ARIMA, 32.41% for Seasonal ARIMA (SARIMA), and 31.73% for Temporal Convolutional Network (TCN), indicating a substantial reduction in forecasting error. These findings confirm the effectiveness of combining TPE-based optimization with hybrid modeling in capturing complex agricultural and climatic dynamics. Overall, the proposed framework provides a scalable and data-driven forecasting tool to support farmers in production planning and assist policymakers in designing more effective price stabilization and agricultural resource management strategies under climate variability.

Keywords-ARIMA; LSTM; TPE; forecasting; automation; time series; hybrid model; computational modeling

I. INTRODUCTION

Red chili peppers are a strategic agricultural commodity in Indonesia, playing a significant role in household consumption and food price stability. Despite their importance, chili prices remain highly volatile due to climate variability, seasonal production patterns, and inter-regional supply chain dependencies. These fluctuations complicate agricultural

planning and market intervention strategies, highlighting the need for reliable forecasting models.

From a time-series perspective, chili price, production, and climate variables exhibit strong temporal dependence. Autoregressive Integrated Moving Average (ARIMA) models are widely used to capture such linear dependencies in historical data and have demonstrated effectiveness in agricultural forecasting applications [1, 2]. In ARIMA, the autoregressive (AR) component models the current observation as a function of past values, whereas the moving average (MA) component captures the influence of past forecast errors. Formally, an AR process of order p is expressed as [3-5]:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \varepsilon_t \quad (1)$$

where Φ_i denote the autoregressive coefficients and ε_t is a white-noise error term. Stationarity is a prerequisite for valid ARIMA modeling; therefore, non-stationary series are transformed through differencing, where the differencing order d represents the number of times the series is differenced to achieve stationarity.

The moving average component complements the autoregressive structure by modeling the current value as a linear combination of present and past error terms. An MA process of order q is defined as:

$$X_t = \varepsilon_t - \Theta_1 e_{t-1} - \Theta_2 e_{t-2} + \cdots - \Theta_q e_{t-q} \quad (2)$$

with Θ_j representing the moving average parameters. Combining the AR and MA components yields the ARMA model, which can be extended to non-stationary series through differencing, resulting in the ARIMA formulation. When the ARMA structure is applied to the differenced series, the ARIMA(p, d, q) model is expressed as:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \varepsilon_t - \Theta_1 e_{t-1} - \Theta_2 e_{t-2} + \cdots - \Theta_q e_{t-q} \quad (3)$$

Although ARIMA models are effective for capturing linear patterns, agricultural and climatic time series often exhibit nonlinear dynamics induced by weather variability and supply-demand interactions. Previous agricultural forecasting studies in Indonesia have demonstrated the effectiveness of ARIMA models for modeling red chili pepper dynamics. For instance, a study in North Sumatra applied ARIMA combined with min–max normalization to forecast retail price, production, rainfall, and sunshine duration [6]. The results showed that the forecasts closely followed the actual observations, achieving very accurate performance for price (ARIMA(1,1,1): MAE Rp687.6107, MAPE 1.34%) and production (ARIMA(6,1,12): MAE 4,796.57 quintals, MAPE 7.07%), as well as very accurate sunshine duration forecasts (ARIMA(1,0,1): MAE 0.2894 h, MAPE 7.33%). Rainfall forecasting also reached fairly accurate performance (ARIMA(6,0,6): MAE 1.4676 mm, MAPE 38.59%). The study reported that min–max normalization helped enhance forecasting accuracy by reducing data variability. However, several limitations were identified, including the potential reliance on manual ARIMA parameter selection, dependence on MAPE as the primary evaluation metric despite its sensitivity to small or zero actual values, the relatively short data span of two years due to data availability, and incomplete meteorological records from the Meteorology, Climatology, and Geophysics Agency (BMKG). These limitations indicate that while ARIMA performs well for linear patterns, more robust and automated approaches are needed to handle nonlinear dynamics and improve generalizability. To address these complexities, recent studies have employed machine learning approaches such as Long Short-Term Memory (LSTM) networks, which are capable of learning nonlinear and long-term dependencies in sequential data.

However, LSTM performance is sensitive to data scaling, necessitating appropriate normalization procedures. In this study, min–max normalization is applied to rescale input data into a fixed range [0,1] [7], defined as [8]:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4)$$

To capture the nonlinear temporal patterns that cannot be modeled by ARIMA, this study employs an LSTM network. LSTM is a variant of recurrent neural networks designed to retain long-term dependencies through a gated memory structure. The information flow within an LSTM cell is regulated by a forget gate, an input gate, and an output gate [9].

At time step t, the LSTM mechanism is mathematically formulated as follows:

$$f_t = \sigma\left(W_f \left[h_{\{t-1\}}, X_t \right] + b_f \right) \quad (5)$$

$$i_t = \sigma\left(W_i \left[h_{\{t-1\}}, X_t \right] + b_i \right) \quad (6)$$

$$o_t = \sigma\left(W_o \left[h_{\{t-1\}}, X_t \right] + b_o \right) \quad (7)$$

$$\widetilde{C_t} = tanh\left(W_{c\left[h_{\{t-1\}}, X_t \right]} + b_c \right) \quad (8)$$

$$C_t = f_t \odot C_{\{t-1\}} + i_t \odot \widetilde{C_t} \quad (9)$$

$$h_t = o_t \odot tanh(C_t) \quad (10)$$

where X_t denotes the input vector, h_t represents the hidden state, C_t is the cell state, and $\widetilde{C_t}$ denotes the candidate cell state. $\sigma(\cdot)$ is the sigmoid activation function, and $\odot$ indicates element-wise multiplication. Through this gating mechanism, LSTM effectively learns nonlinear relationships and long-term dependencies in agricultural and climatic time series. Rather than relying on a single modeling approach, this study integrates ARIMA and LSTM into a hybrid forecasting framework. The final prediction is obtained using a weighted average of both model outputs:

$$\hat{y}_t = w \cdot \hat{L}_t + (1 - w) \cdot \hat{N}_t \quad (11)$$

where $\hat{L}_t$ and $\hat{N}_t$ denote the predicted outputs of the ARIMA and LSTM models, respectively, and $w \in [0,1]$ is the weighting parameter. To evaluate the forecasting performance of the proposed hybrid model, this study employs three widely used error metrics in time-series forecasting: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE) [10]. These metrics provide complementary perspectives on prediction accuracy by measuring absolute deviation, squared deviation, and relative percentage error, respectively.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \quad (12)$$

$$\text{MSE} = \frac{1}{n} \times \sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \quad (13)$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (14)$$

where y_i denotes the actual observed value, $\hat{y}_i$ represents the predicted value, and n is the total number of observations. By combining these three metrics, the evaluation captures both scale-dependent and scale-independent forecasting errors, enabling a balanced assessment of model performance.

In addition to these error-based evaluation measures, this study applies the Diebold–Mariano (DM) test to statistically assess whether differences in forecasting accuracy between competing models are significant [11, 12]. The DM test evaluates two forecasting models by comparing the difference in their loss functions, defined as [13]:

$$d_t = g(e_{1,t}) - g(e_{2,t}) \quad (15)$$

where $e_{i,t}$ denotes the forecast error of model i at time t, and $g(\cdot)$ represents a loss function such as squared error or absolute error [12]. The DM test statistic is computed as:

$$\text{DM} = \frac{\bar{d}}{\sqrt{\widehat{Var}(\bar{d})}} \quad (16)$$

where $\bar{d}$ is the mean of the loss differential and $\sqrt{\widehat{Var}(\bar{d})}$ is a heteroskedasticity and autocorrelation consistent (HAC) variance estimator. Under the null hypothesis of equal predictive accuracy between the two models, the DM statistic asymptotically follows a standard normal distribution. A statistically significant result indicates that the observed difference in forecasting performance is unlikely to be due to random variation alone.

Based on the hybrid formulation expressed in (11), the accuracy of the final forecast depends not only on the complementary strengths of ARIMA and LSTM in modeling linear and nonlinear patterns, but also on the selection of optimal hyperparameters and appropriate weighting of model outputs. In practice, hyperparameter tuning for both ARIMA and LSTM is often performed manually or through conventional Grid Search (GS) methods, which are computationally expensive, time-consuming, and difficult to scale when applied to multiple time series and regions. Therefore, this study aims to develop an automated hybrid forecasting framework that integrates ARIMA and LSTM models optimized using the Tree-structured Parzen Estimator (TPE). The proposed framework is applied to jointly forecast red chili prices, production volumes, and climate-related variables—namely rainfall and sunshine duration—in three Indonesian provinces: North Sumatra, Central Java, and Bali.

Specifically, this research seeks to address the following research questions:

1. Can TPE-based hyperparameter optimization improve the efficiency and predictive accuracy of ARIMA and LSTM models compared to traditional GS approaches?
2. Does the proposed hybrid ARIMA–LSTM framework consistently outperform standalone ARIMA models across agricultural and climatic time series?
3. To what extent can the hybrid model capture the combined linear and nonlinear dynamics underlying chili market fluctuations influenced by climate variability?

The key contributions of this study are threefold. First, this research implements an automated hyperparameter optimization pipeline based on TPE for both ARIMA and LSTM models, reducing manual intervention while improving computational efficiency. Second, a weighted hybrid ARIMA–LSTM forecasting framework is constructed to integrate linear statistical modeling with nonlinear deep learning capabilities for agricultural and climate-driven time series. Third, an extensive empirical evaluation across multiple provinces and variables is conducted to demonstrate the robustness, scalability, and practical applicability of the proposed approach.

II. METHODOLOGY

This research adopts a quantitative approach utilizing time series analysis, as it is well-suited for examining numerical data collected over time to identify patterns, trends, and relationships among variables. Quantitative methods involve the systematic gathering and statistical analysis of numerical data to explain real-world phenomena and test hypotheses [14]. Time series analysis is particularly appropriate for this study, as it enables the examination of historical data to forecast future developments. This study was implemented using Visual Studio Code as the primary development environment and Python as the main programming language. Several Python libraries were utilized to support different stages of the research. Pandas was used for data preprocessing and manipulation, whereas Matplotlib and Seaborn were employed for data visualization. The Statsmodels library was used to build ARIMA models, and TensorFlow was utilized for developing LSTM models. Model integration through weighted averaging and the computation of error metrics (MAE, MSE, and MAPE) were performed using Scikit-learn. Statistical significance testing was conducted using the DM test, and Optuna was applied to perform TPE-based hyperparameter optimization. Overall, the methodology integrates hybrid ARIMA and LSTM modeling supported by systematic data preprocessing, hyperparameter optimization, and model evaluation. The research stages are illustrated in Figure 1.

A. Data Collection

The data collection stage focuses on obtaining retail price, production, and weather data of red chili peppers across three provinces: North Sumatra, Bali, and Central Java, for the period from March 2021 to December 2023. The study utilizes four datasets in .xlsx format, representing daily retail prices (Rp/kg), monthly production volumes (quintals), daily rainfall intensity (millimeters), and daily sunshine duration (hours). Each dataset consists of a date column followed by three provincial columns corresponding to Bali, Central Java, and North Sumatra.

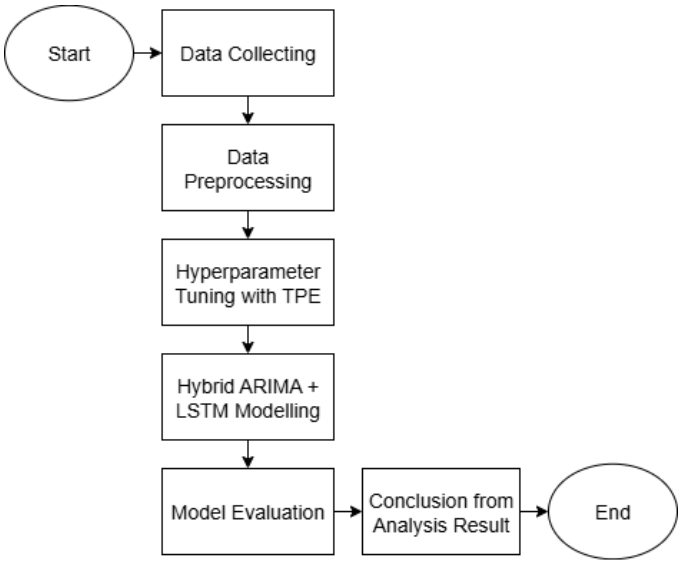

Fig. 1. Overall methodological workflow of the study.

Daily retail price data, which differ by city or regency, are averaged at the provincial level and are sourced from the National Food Agency's public platform, accessible at panelharga.badanpangan.go.id. Monthly production data are obtained from the National Food Agency's Data and Information Center through formal access permission. Weather data, including rainfall and sunshine duration, are collected from the BMKG's Database Center, accessible at data.bmkg.go.id. For each province, data are retrieved from representative weather stations: Silangit and Aek Godang Stations for North Sumatra, I Gusti Ngurah Rai and Bali Climatology Stations for Bali, and Maritim Tanjung Emas Station for Central Java. The daily datasets of price, rainfall, and sunshine duration contain 1,036 rows, whereas the monthly production dataset consists of 34 rows over the same period. Public datasets for retail prices and weather variables are made available through an open repository [15], whereas production data have access restrictions imposed by the data provider.

B. Data Preprocessing

Figure 2 illustrates the preprocessing flow applied in this study. The preprocessing process begins with handling missing values in the time series data. A hybrid interpolation approach is applied to address different missing patterns. Cubic spline interpolation is used to fill gaps located in the middle of the series, as it ensures smooth transitions between surrounding data points. However, linear interpolation is selected for missing values occurring at the beginning or end of the series, such as initial gaps in the price data and edge gaps in the rainfall data, as spline interpolation requires available values on both sides. After resolving missing values, the production data, which are originally recorded on a monthly basis, are disaggregated into daily data to align with the other three variables. Monthly production totals are first converted into average daily values by dividing them by the number of days in each month. To replicate realistic daily variability while preserving the original monthly trend, daily production values are generated using a normal distribution with the daily average as the mean and a standard deviation of 0.1.

Fig. 2. Data preprocessing flow.

After that, the dataset is split into training and testing sets using an 80:20 ratio. The training set consists of the first 80% of the chronologically ordered data, whereas the remaining 20% serves as the testing set for model evaluation. Finally, all variables are normalized using min–max normalization to ensure comparable scales between variables with different units.

C. Hyperparameter Tuning with Tree-Structured Parzen Estimator

Once the preprocessing stage is completed, the next step is hyperparameter tuning using the TPE, as shown in Figure 3. TPE is a Bayesian optimization method that models the probability distribution of hyperparameters and adaptively explores the search space to identify optimal configurations.

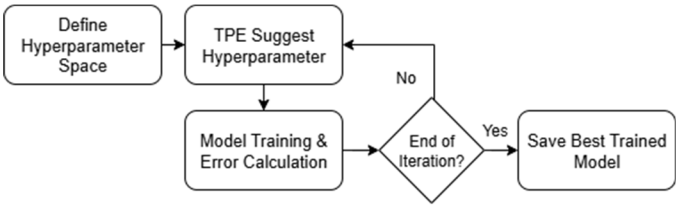

Fig. 3. Hyperparameter tuning process with TPE.

In this study, TPE is applied separately to both ARIMA and LSTM models. For ARIMA, the parameters (p, d, q) are tuned within predefined ranges, whereas for LSTM, hyperparameters such as the number of hidden units, learning rate, batch size, and number of epochs are optimized.

The tuning process begins by sampling initial candidate configurations from the defined hyperparameter space. Each configuration is trained on the daily sequences using the training set and validated on the corresponding validation data. To ensure consistency with the original temporal resolution of production data, all forecasts and test results are subsequently aggregated back into monthly values before performance evaluation. As a result, while model training and tuning are performed on daily data, the final evaluation metrics are reported on a monthly basis.

By iteratively refining the search distribution, TPE provides a systematic, adaptive, and computationally efficient approach to hyperparameter optimization. This process ensures that both ARIMA and LSTM models are tuned objectively and consistently, laying a robust foundation for subsequent steps in hybrid modeling and performance evaluation.

D. Hybrid ARIMA–LSTM Modeling

Following the hyperparameter tuning stage, the next step is to construct a hybrid ARIMA–LSTM model. This hybridization is designed to integrate the complementary strengths of both models, where ARIMA effectively captures linear and autoregressive patterns, whereas LSTM is capable of

modeling nonlinear relationships and long-term dependencies that cannot be effectively modeled by ARIMA alone.

The hybrid modeling framework is depicted in Figure 4. ARIMA and LSTM are first trained and optimized independently using their respective best hyperparameters. Each model then generates its own forecasts for the target variables. The two sets of forecasts are combined using a weighted average method, where the contribution of ARIMA and LSTM predictions is determined based on their relative performance. Specifically, the weights are calculated using the inverse of the MSE obtained during validation for each model and normalized so that the total weight equals one. This method assigns higher weights to the model with lower forecasting error, so the hybrid framework dynamically prioritizes the contributions of more accurate models for each dataset.

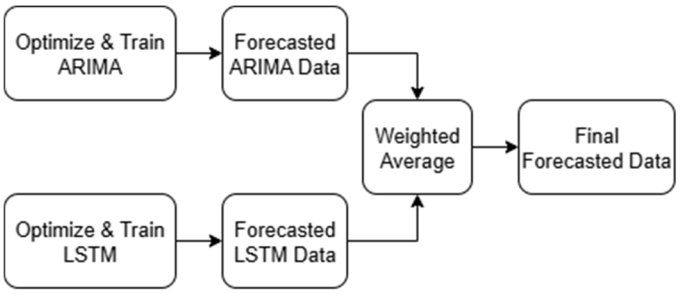

Fig. 4. Hybrid modeling with ARIMA and LSTM.

By combining ARIMA and LSTM forecasts through performance-based weighting, the hybrid ARIMA–LSTM model produces more stable and accurate predictions than either model alone. This approach offers greater robustness in handling the complex dynamics of agricultural and climatic time series.

E. Model Evaluation

The final step in this study is model evaluation, which is conducted to measure the forecasting performance of the hybrid ARIMA–LSTM models. The evaluation framework is divided into three components.

First, the forecasting accuracy of each model is evaluated using three commonly used error metrics: MAE, MSE, and MAPE. These metrics are calculated separately for each province to provide an overview of the overall performance of the ARIMA, LSTM, and proposed ARIMA–LSTM hybrid models. To facilitate a clearer interpretation of forecasting accuracy, this study adopts MAPE accuracy criteria proposed by authors in [16], as shown in Table I.

Second, to determine whether the error reduction achieved by the hybrid model is statistically significant, the DM test is used. The DM test is used to compare the forecasting accuracy between the ARIMA–LSTM hybrid model and the basic ARIMA model for each province. This statistical test ensures that the improvement is not only numerical but also supported by formal significance testing.

Third, the hybrid model is compared with additional models using a price data sample. The models are ARIMA, Seasonal ARIMA (SARIMA), and Temporal Convolutional Network (TCN), using MAE, MSE, and MAPE. This comparison aims to position the ARIMA–LSTM hybrid model relative to classical statistical approaches and modern deep learning, reinforcing its overall effectiveness evaluation.

TABLE I. MAPE CRITERIA ACCORDING TO [16]

No.	MAPE	Interpretation
1	< 10%	Very accurate in prediction
2	10%–20%	Accurate in prediction
3	20%–50%	Fairly accurate in prediction
4	> 50%	Not accurate in prediction

All evaluations were performed on the test dataset, which consisted of the last 20% of chronologically ordered data. Through this evaluation framework, this study ensures a balanced assessment that combines numerical accuracy, statistical significance, and benchmarking comparisons between different modeling approaches.

F. Conclusion from Analytical Results

The analytical results show that the integration of robust preprocessing, TPE-based hyperparameter tuning, and hybrid ARIMA–LSTM modeling improves forecasting accuracy for chili price, production, rainfall, and sunshine duration. ARIMA effectively captures linear patterns, whereas LSTM models nonlinear and long-term dependencies, resulting in lower errors when combined in a hybrid framework.

Weather variables, particularly rainfall and sunshine duration, are found to significantly influence production and price dynamics. However, the effectiveness of the hybrid model is influenced by data quality, especially for highly variable weather data, and by the use of disaggregated production data. In addition, although the model performs well for the studied provinces, its generalization to other regions or commodities may vary due to differences in climate and market conditions.

Overall, the results validate the hybrid ARIMA–LSTM approach as an effective and scalable forecasting framework, while highlighting the need for further validation using more diverse datasets.

III. RESULTS AND DISCUSSION

A. Data Collection

Figure 5 illustrates that the datasets exhibit heterogeneous characteristics across variables and provinces, reflecting differing economic and climatic dynamics. Retail price series show moderate volatility after the initial period, with Bali displaying higher sensitivity to fluctuations compared to Central Java and North Sumatra, which act as major production and distribution hubs. Production data reveal substantial interprovincial disparities in scale, with Central Java and North Sumatra consistently recording higher volumes than Bali, and monthly fluctuations that align with agricultural planting and harvesting cycles. Climatic variables demonstrate contrasting behaviors, as rainfall intensity is highly volatile with frequent

extreme values and strong regional variation, whereas sunshine duration follows smoother and more stable daily patterns. These differences indicate that economic variables are relatively structured, whereas climatic variables, particularly rainfall, are more irregular and prone to abrupt shocks.

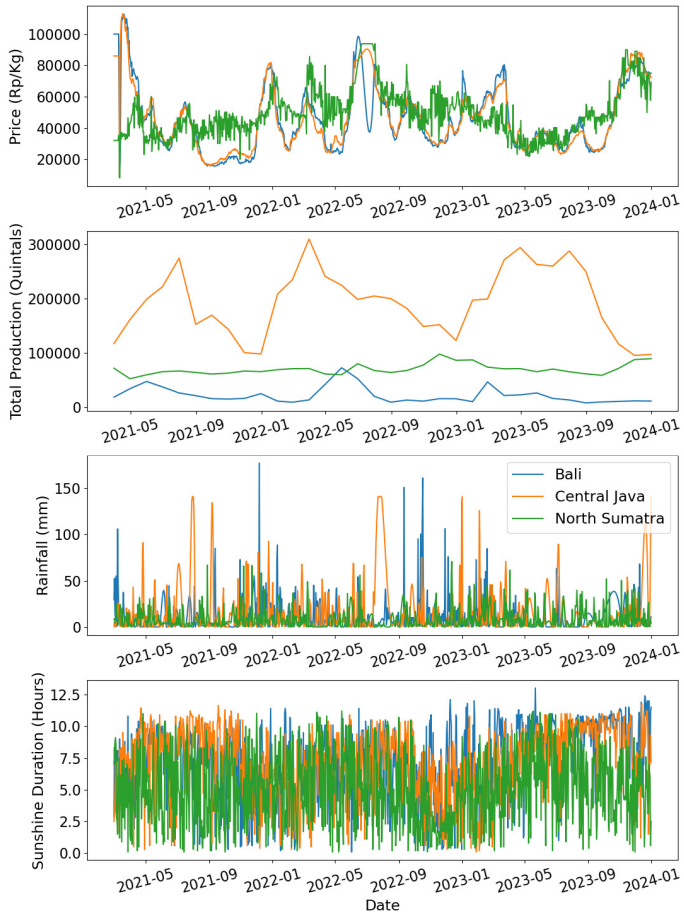

Fig. 5. Temporal patterns of daily retail price, total production, rainfall intensity, and sunshine duration for Bali, Central Java, and North Sumatra from March 2021 to December 2023.

As shown in Figure 6, missing values were found in all daily datasets except for the production dataset, which is complete across all provinces. The retail price dataset contained 47 missing entries for Bali, 43 for Central Java, and 47 for North Sumatra, mostly at the beginning of the time series. The rainfall dataset had the largest number of gaps, with 504 for Bali, 516 for Central Java, and 295 for North Sumatra, distributed across the beginning, middle, and end of the series. The sunshine duration dataset contained fewer missing entries, with 23 for Bali, 48 for Central Java, and 25 for North Sumatra, mainly concentrated in the middle of the series. In addition to missing values, another challenge was the inconsistency in temporal resolution, since most datasets were daily, whereas the production data were originally available only on a monthly basis.

date	# bali_price	# java_price	# sumatra_price
2021-03-01 00:00:00	*Missing value*	*Missing value*	*Missing value*
2021-03-02 00:00:00	*Missing value*	*Missing value*	*Missing value*
2021-03-03 00:00:00	*Missing value*	*Missing value*	*Missing value*
2021-03-04 00:00:00	*Missing value*	*Missing value*	*Missing value*
2021-03-05 00:00:00	*Missing value*	*Missing value*	*Missing value*
2021-03-06 00:00:00	*Missing value*	*Missing value*	*Missing value*
2021-03-07 00:00:00	*Missing value*	*Missing value*	*Missing value*
2021-03-08 00:00:00	*Missing value*	*Missing value*	*Missing value*
2021-03-09 00:00:00	100000.0	86000.0	32000.0
2021-03-10 00:00:00	100000.0	86000.0	35330.0

(a)

date	# bali_production	# java_production	# sumatra_production
2021-03-01 00:00:00	18176.91	116813.42	71534.7
2021-04-01 00:00:00	33843.14	161138.74	52169.5
2021-05-01 00:00:00	47167.85	198384.44	59343.46
2021-06-01 00:00:00	37062.43	221756.87	65186.19
2021-07-01 00:00:00	25936.06	274201.62	66465.75
2021-08-01 00:00:00	21022.5	152244.91	63776.74
2021-09-01 00:00:00	15519.99	169151.42	60861.62
2021-10-01 00:00:00	14757.32	143167.55	62465.55
2021-11-01 00:00:00	15951.04	100237.91	66340.98
2021-12-01 00:00:00	24770.81	97708.4	65298.66

(b)

date	# bali_rain	# java_rain	# sumatra_rain
2021-03-01 00:00:00	29.5	0.1	*Missing value*
2021-03-02 00:00:00	*Missing value*	5.4	8.5
2021-03-03 00:00:00	15.1	*Missing value*	*Missing value*
2021-03-04 00:00:00	8.1	0.2	12.3
2021-03-05 00:00:00	55.1	1.1	*Missing value*
2021-03-06 00:00:00	0.4	5.4	24.5
2021-03-07 00:00:00	21.0	3.4	13.0
2021-03-08 00:00:00	106.1	8.0	1.5
2021-03-09 00:00:00	52.2	8.0	6.1
2021-03-10 00:00:00	23.7	25.9	6.1

(c)

date	# bali_sun	# java_sun	# sumatra_sun
2021-03-01 00:00:00	2.8	2.5	7.0
2021-03-02 00:00:00	8.3	3.0	8.0
2021-03-03 00:00:00	7.0	4.4	8.8
2021-03-04 00:00:00	6.6	7.0	9.1
2021-03-05 00:00:00	4.4	3.2	8.5
2021-03-06 00:00:00	3.1	6.4	2.8
2021-03-07 00:00:00	6.1	7.2	2.5
2021-03-08 00:00:00	8.1	4.8	1.5
2021-03-09 00:00:00	3.6	4.0	6.1
2021-03-10 00:00:00	6.1	5.5	4.0

(d)

Fig. 6. First 10 examples in all datasets: (a) price data, (b) production data, (c) rainfall intensity data, and (d) sunshine duration data.

B. Data Preprocessing

The first preprocessing step addresses missing values using linear and cubic spline interpolation. As illustrated in Figure 7, which presents a sample illustration based on daily retail price

data from Bali province, missing values in the series occur at the beginning of the dataset and around mid-2021, and the interpolated segments closely follow the prevailing price trajectory without introducing abrupt deviations. The reconstructed curves preserve the magnitude and seasonal rhythm of the surrounding data, indicating that interpolation did not generate artificial volatility or oversmoothed genuine price fluctuations. By allowing the interpolated segments to rejoin the original series smoothly, the preprocessing step stabilizes the data while maintaining essential economic signal structures needed for effective model learning.

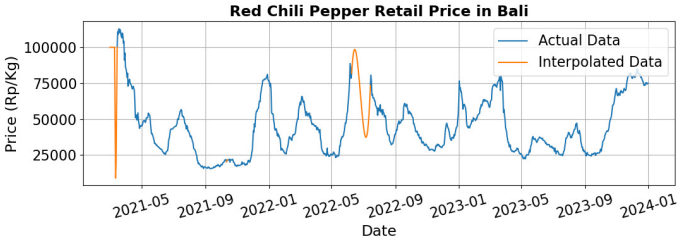

Fig. 7. Example of hybrid interpolation result.

Following the handling of missing values, the production data were disaggregated into a daily frequency to align with the other variables. Figure 8 shows that the disaggregated series preserves relative production dominance across provinces, with Central Java consistently the highest, followed by North Sumatra and Bali. The daily expansion introduces gradual intra-month fluctuations without unrealistic volatility, maintaining proportional relationships with monthly patterns. This indicates that the disaggregation process preserves production hierarchy and seasonal dynamics, providing meaningful daily signals that support adaptive model learning, particularly under climatic variability.

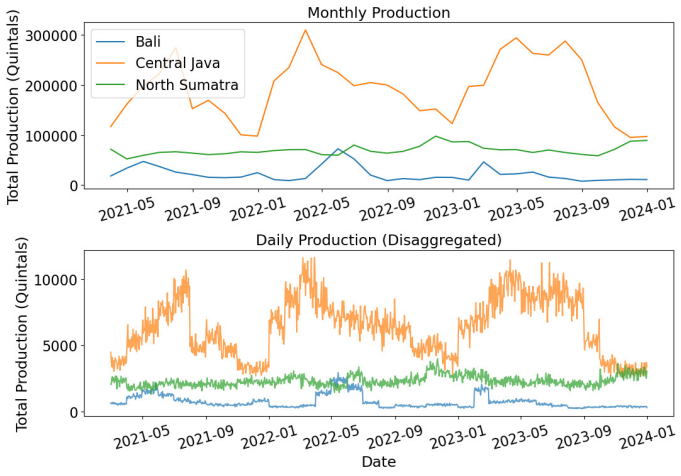

Fig. 8. Disaggregation result on production data.

After interpolation and disaggregation, the datasets were split into training and testing sets using an 80:20 chronological division to preserve temporal dependencies. All variables were represented at a daily frequency, resulting in 822 training observations and 214 testing observations. Min–max normalization was then applied after the data split to ensure comparable scales and to prevent data leakage during model training.

C. *Hyperparameter Tuning with Tree-Structured Parzen Estimator*

The selection of hyperparameter search ranges and the number of optimization trials were guided by theoretical considerations and practical computational constraints. For the ARIMA model, the autoregressive and moving average orders p and q were searched within the range of 0 to 12 to capture short-term and medium-term temporal dependencies, while avoiding excessive model complexity and overfitting, as recommended in recent forecasting studies [17]. The differencing order d was determined separately through stationarity testing and therefore excluded from the optimization space.

For the LSTM model, hyperparameter ranges were defined to balance model capacity and training stability. The number of LSTM units and dense units was limited to 32 and 64, and the number of layers to one or two, as these configurations are widely reported to be effective for medium-scale time series while preventing overparameterization [16, 18]. Dropout rates between 0.2 and 0.5 were applied for regularization, and the learning rate was searched within the range of 0.0001 to 0.001 to ensure stable convergence, consistent with common practices in neural network-based time series forecasting [18].

The number of optimization trials was also selected based on model complexity. ARIMA models are computationally lightweight, allowing the use of 20 TPE trials to adequately explore the parameter space. In contrast, LSTM training is substantially more computationally expensive; therefore, the number of trials was limited to five, as increasing the number of trials yielded diminishing performance gains relative to the added computational cost. With this configuration, the total tuning time for ARIMA with 20 TPE trials and LSTM with five TPE trials remains relatively comparable, ensuring a balanced and fair allocation of computational resources across models.

The following pseudocode summarizes the ARIMA implementation with TPE-based hyperparameter optimization.

```
FOR each column in the 12 daily time
series:
  # Phase 1: Differencing
  d = 0
  WHILE series not stationary:
    difference the series
    d += 1
  Store d
  # Phase 2: TPE-based Hyperparameter
  Tuning
  DEFINE TPE objective():
    p = suggest_int(0, 12)
    q = suggest_int(0, 12)
    Fit ARIMA(p, d, q) on training data
    (daily)
    Predict on test data (daily)
```

```
    Aggregate all data to monthly
    RETURN MAE, MSE, MAPE from monthly
    data
  Run TPE with predefined n_trials = 20
  Store best params and metrics for this
  series
```

The ARIMA pipeline consists of two main phases. First, stationarity is achieved by iteratively differencing the daily series based on the Augmented Dickey–Fuller test until the differencing order d is determined. Second, hyperparameter tuning is conducted using the TPE, which searches for optimal p and q values within the range of 0 to 12 by adaptively sampling promising regions of the parameter space. The number of iterations is fixed at 20 to balance efficiency and exploration. For each trial, the model is trained and evaluated on daily data, with forecasts subsequently aggregated into monthly values. Final performance is assessed using MAE, MSE, and MAPE computed on the aggregated monthly results to ensure consistency with the original production data.

The following pseudocode summarizes the LSTM implementation with TPE-based hyperparameter optimization.

```
FOR each column in the 12 daily time
series:
  # Phase 1: Windowing with size = 10
  X_train, y_train =
  make_sequence(train_data)
  X_test, y_test =
  make_sequence(test_data)
  # Phase 2: TPE-based Hyperparameter
  Tuning
  DEFINE TPE objective():
    lstm_units = suggest_categorical(32,
    64)
    lstm_layers = suggest_int(1, 2)
    dense_units = suggest_categorical(32,
    64)
    dropout_rate = suggest_float(0.2,0.5)
    lr_rt = suggest_float(1e-4, 1e-3,
    log=True)
    batch_size = suggest_categorical(16,
    32, 64)
    Build LSTM model accordingly
    Train on train data with
    batch_size
    Predict on test data (daily)
    Aggregate all data to monthly
    RETURN MAE, MSE, MAPE from monthly
    data
  Run TPE with predefined n_trials = 5
  Store best params and metrics for this
  series
```

The LSTM pipeline starts with supervised sequence generation using a fixed window size of 10. Hyperparameter tuning is performed using the TPE, with search ranges defined to balance model capacity and training efficiency, including LSTM units between 32 and 64, one to two layers, dense units between 32 and 64, dropout rates from 0.2 to 0.5, learning rates between 1e−4 and 1e−3 on a logarithmic scale, and batch sizes from 16 to 64. The number of optimization trials is set to five to limit computational cost. For each trial, the model is trained and evaluated on daily data, with predictions aggregated to monthly values. Model performance is assessed using MAE, MSE, and MAPE computed on the aggregated results to ensure consistency with the original production data.

Table II compares training efficiency and predictive performance between GS and TPE for ARIMA and LSTM models using completion time and the range of MAPE values across twelve series. TPE substantially reduces training time for both models, particularly for LSTM, which decreases from 15,060.11 s to 819.08 s. For ARIMA, TPE also improves efficiency and narrows the error range by lowering the minimum MAPE from 6.23% to 4.61% and the maximum MAPE from 53.91% to 46.31%. Although GS achieves a lower minimum MAPE for LSTM, TPE reduces the highest MAPE from 30.63% to 25.08%, indicating more stable performance. Overall, TPE provides a more efficient and robust optimization strategy without compromising predictive accuracy.

TABLE II. HYPERPARAMETER TUNING COMPARISON USING GS AND TPE FOR ARIMA AND LSTM MODELS

Model	Completion time (s)	Lowest MAPE (%)	Highest MAPE (%)
GS-ARIMA	3,379.89	6.23	53.91
TPE-ARIMA	1,247.57	4.61	46.31
GS-LSTM	15,060.11	0.69	30.63
TPE-LSTM	819.08	1.54	25.08

These results highlight that TPE not only accelerates the training process but also consistently improves predictive accuracy across both ARIMA and LSTM models. Beyond efficiency, the framework ensures an automated and reproducible pipeline that minimizes human intervention in hyperparameter tuning and maximizes model quality through adaptive exploration of the search space. All experiments were conducted on a Windows 11 local operating system equipped with a Ryzen 7 6800H CPU, a GeForce RTX 3060 Laptop GPU, 2×8 GB DDR5-4800 RAM, and a PCIe Gen 4 SSD.

D. Hybrid ARIMA–LSTM Modeling

Following the forecasting stage, final predictions are generated through a weighted hybridization of ARIMA and LSTM, where model contributions are determined by the inverse of their MSEs. This normalized weighting scheme allows the hybrid framework to adaptively emphasize the more accurate model for each time series. The resulting weight ratios for each variable and province are summarized in the Table III.

Table III presents normalized weight ratios ranging from 0 to 1 for ARIMA and LSTM across the three provinces and four datasets, revealing clear yet complementary contribution patterns between the two models. Although LSTM consistently dominates the weighting for price and production data, with weights exceeding 0.6 across all provinces, this dominance reflects the strong nonlinear characteristics inherent in economic and agricultural variables rather than diminishing the role of ARIMA. In contrast, ARIMA contributes substantially to meteorological variables, particularly rainfall and sunshine

duration in Bali and North Sumatra, indicating its effectiveness in capturing linear, seasonal, and autoregressive structures commonly observed in climatic time series.

TABLE III. NORMALIZED WEIGHT RATIOS FOR EACH MODEL

Data	Province	ARIMA weight	LSTM weight
Price	Bali	0.0174	0.9826
	Central Java	0.0057	0.9943
	North Sumatra	0.0109	0.9891
Production	Bali	0.0590	0.9410
	Central Java	0.0142	0.9858
	North Sumatra	0.1281	0.8719
Rainfall intensity	Bali	0.5817	0.4183
	Central Java	0.2406	0.7594
	North Sumatra	0.5519	0.4481
Sunshine duration	Bali	0.2154	0.7846
	Central Java	0.3606	0.6394
	North Sumatra	0.5472	0.4528

Importantly, even in cases where LSTM receives a higher weight, ARIMA remains a critical component of the hybrid framework by providing a stable linear baseline that captures trend and seasonality, thereby reducing model variance and preventing LSTM from overfitting short-term noise. The weighted averaging mechanism does not aim to balance contributions equally, but rather to allocate weights adaptively based on model performance for each dataset. Consequently, the hybrid model benefits from ARIMA's robustness and interpretability alongside LSTM's capacity to model complex nonlinear dynamics.

E. Model Evaluation

Figure 9 compares ARIMA and the hybrid model using retail price data from Central Java. The hybrid model tracks actual price movements more closely than ARIMA, particularly during sharp increases and sudden drops, whereas ARIMA exhibits lag during abrupt changes. This visual improvement is consistent with the quantitative results in Table IV, where the hybrid model reduces MAPE from 12.21%, corresponding to moderate accuracy, to 1.57%, indicating high predictive accuracy, alongside a substantial reduction in MAE from 4,882.94 to 625.40. This example illustrates the practical benefit of hybridization in capturing nonlinear price dynamics, while serving as an illustrative case rather than conclusive evidence of overall model superiority.

As shown in Table IV, Central Java exhibits the most substantial improvements after hybridization, particularly for economic variables. Price forecasting accuracy improves markedly, with MAPE decreasing from 12.21% to 1.57%, alongside a sharp reduction in MAE from 4,882.94 to 625.40, whereas production forecasting also benefits significantly as MAPE declines from 21.47% to 3.04%. These results indicate that the hybrid ARIMA–LSTM framework effectively captures nonlinear behavior and market-driven volatility in Central Java's price and production data. In contrast, improvements for rainfall intensity are more limited, with MAPE reduced from 46.31% to 21.36%, suggesting that high variability and noise continue to constrain forecasting accuracy. Sunshine duration shows comparable performance between models, with a slight increase in MAPE from 6.70% to 7.33%, indicating that hybridization does not uniformly improve all variables and may introduce marginal trade-offs when ARIMA already provides a strong baseline.

TABLE IV. MODEL PERFORMANCE ON CENTRAL JAVA PROVINCE

Model	Metric	Price data	Production data	Rainfall intensity data	Sunshine duration data
ARIMA	MAE	4,882.94	25,648.29	4.07	0.48
	MSE	86,601,713.95	3,079,479,871.63	56.93	0.57
	MAPE (%)	12.21	21.47	46.31	6.70
Hybrid	MAE	625.40	5,792.52	3.09	0.48
	MSE	513,483.68	46,012,529.33	20.24	0.36
	MAPE (%)	1.57	3.04	21.36	7.33

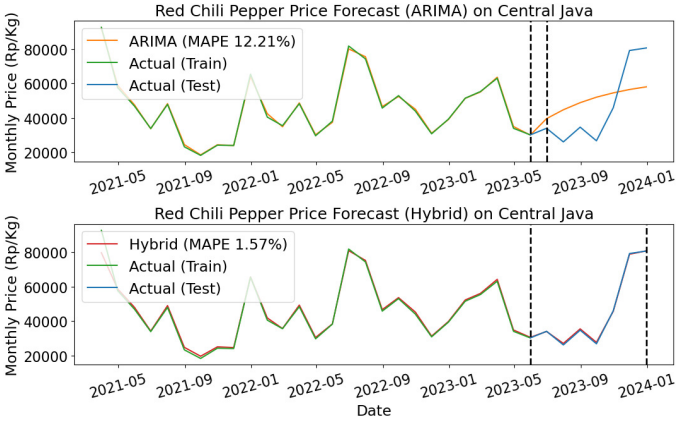

Fig. 9. Sample of hybrid model improvement.

Table V compares the forecasting performance of ARIMA and the hybrid model for Bali province across four variables.

TABLE V. MODEL PERFORMANCE ON BALI PROVINCE

Model	Metric	Price data	Production data	Rainfall intensity data	Sunshine duration data
ARIMA	MAE	3,842.31	4,098.62	2.91	0.79
	MSE	83,368,726.27	49,521,344.19	22.85	1.38
	MAPE (%)	8.19	35.99	32.98	10.20
Hybrid	MAE	872.05	1,170.30	3.02	0.52
	MSE	1,656,551.01	3,174,711.08	22.72	0.46
	MAPE (%)	1.59	5.88	26.99	8.08

The hybrid model consistently reduces MAE and MSE, with the most pronounced improvements observed in price and production, where MAPE decreases from 35.99% to 5.88% for production and from 8.19% to 1.59% for price, indicating a substantial enhancement in predictive reliability. These results suggest that hybridization is particularly effective for variables influenced by nonlinear dynamics and abrupt structural changes. In contrast, improvements for rainfall intensity are more moderate, with MAPE declining from 32.98% to 26.99%, reflecting the persistent impact of high volatility and extreme

events. Sunshine duration shows smaller but stable gains, with MAPE reduced from 10.20% to 8.08%, consistent with its smoother temporal structure and the already reasonable baseline performance of ARIMA. These results indicate that hybridization yields variable-specific benefits, delivering strong improvements for economically driven variables while offering more limited gains for meteorological series.

Table VI indicates that ARIMA already provides a competitive baseline in North Sumatra, particularly for production and sunshine duration, resulting in more modest gains after hybridization. Nevertheless, the hybrid model consistently reduces MAE and MSE across all variables. For price forecasting, MAPE decreases from 8.17% to 1.72%, whereas production forecasting improves from 4.61% to 2.65%, indicating refinement rather than structural change. Improvements for rainfall intensity and sunshine duration are smaller, with MAPE declining from 19.73% to 17.34% and from 8.71% to 7.31%, respectively, suggesting that forecasting accuracy remains constrained by dominant linear and seasonal patterns. These results imply that when such structures prevail, ARIMA remains effective, and the hybrid model functions primarily as a complementary enhancement rather than a transformative alternative.

TABLE VI. MODEL PERFORMANCE ON NORTH SUMATRA PROVINCE

Model	Metric	Price data	Production data	Rainfall intensity data	Sunshine duration data
ARIMA	MAE	4,980.62	3,446.57	1.60	0.35
	MSE	138,633,490.74	36,608,043.28	4.79	0.23
	MAPE (%)	8.17	4.61	19.73	8.71
Hybrid	MAE	929.04	1,987.93	1.59	0.31
	MSE	1,611,453.06	6,912,311.71	4.76	0.21
	MAPE (%)	1.72	2.65	17.34	7.31

The DM test was conducted using a one-sided formulation based on MSE on the out-of-horizon forecasts corresponding to the final 20% of the dataset, with a significance level of $\alpha = 0.05$. Under this formulation, the null hypothesis states that the hybrid model does not yield lower forecast errors than the standalone ARIMA model, whereas the alternative hypothesis assumes superior predictive accuracy of the hybrid model. Based on Table VII, the results indicate that statistically significant improvements in favor of the hybrid model are primarily observed for economic variables. In Bali province, significant differences are found for price, production, and sunshine duration, whereas in Central Java, significance is observed for price and production. In North Sumatra, only production data show a statistically significant improvement. These findings suggest that the hybrid model provides meaningful error reductions for variables characterized by market-driven dynamics and structural changes.

In contrast, most rainfall intensity and sunshine duration series do not exhibit statistically significant differences, despite numerical error reductions. This outcome reflects the high variability and stochastic nature of meteorological data, which limits the ability of pairwise forecast comparisons to consistently detect significant improvements. Consequently, the DM test highlights that numerical accuracy gains do not uniformly translate into statistically robust superiority, emphasizing the importance of complementing error metrics with formal significance testing.

TABLE VII. DM TEST RESULTS

Province	Data	DM stat. (MSE)	p-value	Significance ($\alpha = 0.05$)
Bali	Price	-1.9952	0.0465	Yes
	Production	-6.7021	0.0003	Yes
	Rainfall intensity	-1.7326	0.0669	No
	Sunshine duration	-3.2087	0.0092	Yes
Central Java	Price	-3.7174	0.0049	Yes
	Production	-2.5007	0.0232	Yes
	Rainfall intensity	-1.6742	0.0726	No
	Sunshine duration	-1.9072	0.0526	No
North Sumatra	Price	-1.8426	0.0575	No
	Production	-2.1234	0.0390	Yes
	Rainfall intensity	-1.6731	0.0727	No
	Sunshine duration	-0.1563	0.4405	No

On the other hand, Figure 10 compares the MAPE values of ARIMA, SARIMA, TCN, and the hybrid model for retail price forecasting in Central Java. In this comparison, all models have been enhanced with TPE optimization. ARIMA exhibits the highest error, whereas SARIMA, included as a seasonal extension of ARIMA, achieves moderate improvement but remains limited by its linear formulation. TCN, representing a deep learning-based approach, further reduces the error by capturing temporal dependencies more flexibly. The hybrid model attains the lowest MAPE, demonstrating that combining ARIMA's linear structure with LSTM's nonlinear learning capability yields substantially superior accuracy compared to both statistical and standalone deep learning models.

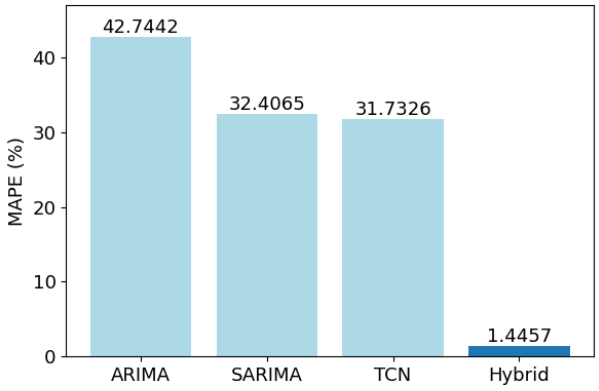

Fig. 10. MAPE comparison of ARIMA, SARIMA, TCN, and the hybrid model for retail price forecasting in Central Java.

Overall, the evaluation shows that although numerical error reductions appear across most variables, statistically robust improvements occur mainly for economic variables such as

price and production. The DM test indicates that gains for meteorological variables are generally not significant, consistent with their high variability and stochastic behavior. While MAPE classifications provide a general sense of predictive accuracy, they do not fully capture model reliability. When considered together, the provincial performance tables, significance testing, and cross-model comparisons demonstrate that the hybrid framework yields meaningful and statistically supported improvements where nonlinear economic dynamics dominate, whereas benefits for climatic variables remain modest and should be interpreted cautiously.

F. Conclusion from Analytical Results

Figure 11 highlights how the hybrid model performs across economic and climatic variables in Bali province, with clear implications for real-world decision making. For retail price and production, the hybrid model forecasts closely track observed values, indicating that the model can capture market-driven dynamics and seasonal production cycles that are critical for supply planning and price stabilization policies. The dominance of the LSTM component for these variables suggests that nonlinear temporal patterns, such as demand shocks and post-harvest adjustments, play a central role in shaping outcomes. However, smoother production forecasts also indicate a limitation in representing abrupt structural changes, implying that while the model is effective for trend-oriented planning, caution is needed when interpreting forecasts during sudden disruptions.

Confidence intervals provide essential insight into forecast reliability under varying levels of uncertainty. Narrower confidence bands for price and sunshine duration reflect lower variance and more stable predictive behavior, supporting their use in short-term planning and operational decision making. In contrast, wider intervals for production and rainfall indicate elevated uncertainty associated with volatile climatic conditions and extreme events, where point forecasts alone may be misleading. The fact that most test observations fall within the confidence bounds suggests that the hybrid model is reasonably well calibrated, even under uncertainty. Overall, these results emphasize that the hybrid framework not only improves point accuracy but also offers valuable uncertainty information, enabling more informed and risk-aware forecasting in agricultural and food system contexts.

Figure 12 presents the hybrid model's monthly forecasts for price, production, rainfall, and sunshine duration in Central Java. The price and production forecasts align closely with the observed series across both training and testing periods, indicating that the model successfully captures the pronounced market fluctuations and large production cycles characteristic of the region. The forecasts remain responsive to major upward and downward movements, reflecting the importance of nonlinear temporal dynamics. However, production forecasts become smoother toward the end of the testing period, suggesting reduced sensitivity to abrupt structural shifts and highlighting a limitation when sudden shocks occur.

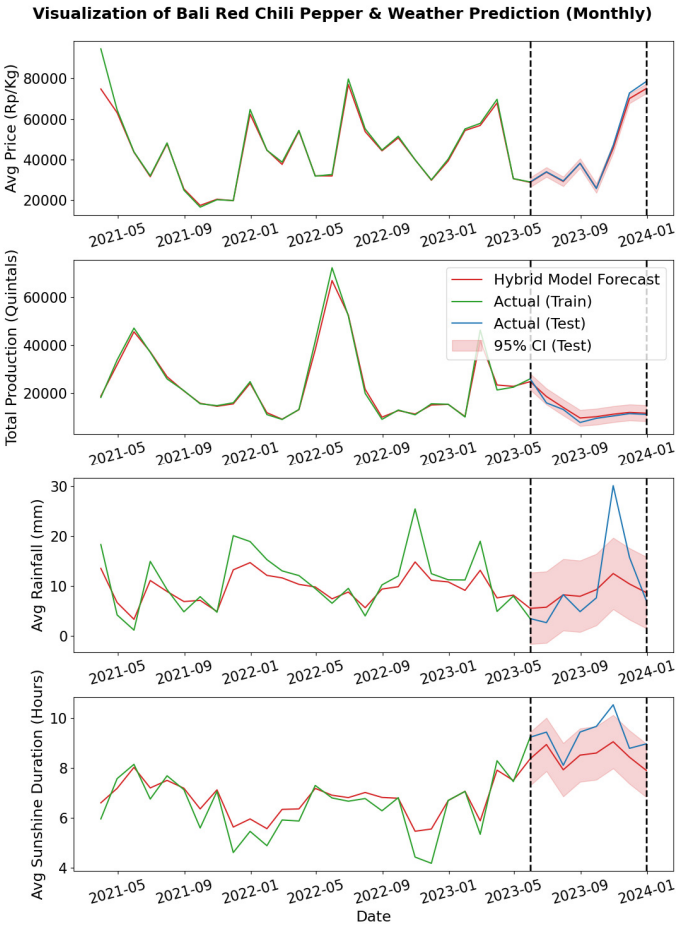

Fig. 11. Actual data and hybrid model forecast graph on Bali province.

The confidence intervals further contextualize forecast reliability. Price and sunshine duration display relatively narrow intervals, indicating lower uncertainty and more stable temporal behavior suitable for short-term planning. Production and rainfall exhibit wider intervals, particularly near the forecast horizon, signaling increased uncertainty under volatile market and climatic conditions. This widening illustrates that although the hybrid model captures central tendencies effectively, variability intensifies when structural changes or extreme events become more prominent. Overall, the Central Java results reinforce the importance of interpreting hybrid forecasts jointly through point predictions and confidence intervals, offering a balanced perspective on accuracy and uncertainty that complements the earlier error metrics and statistical testing discussed in this section.

Figure 13 presents the hybrid model forecasts for monthly price, production, rainfall, and sunshine duration in North Sumatra. Compared to the other provinces, the observed and forecasted series indicate a stronger dominance of linear and seasonal patterns, particularly for production and sunshine duration. For retail price, the hybrid model follows the overall upward and downward movements well during both training and testing periods, although the trajectory appears smoother during the transition into the test horizon, suggesting a more conservative response to rapid market changes. Production forecasts similarly capture the general trend but show limited

sensitivity to short-term fluctuations, indicating that the hybrid model in North Sumatra functions more as a refinement of the existing ARIMA structure rather than a fundamental shift in dynamics.

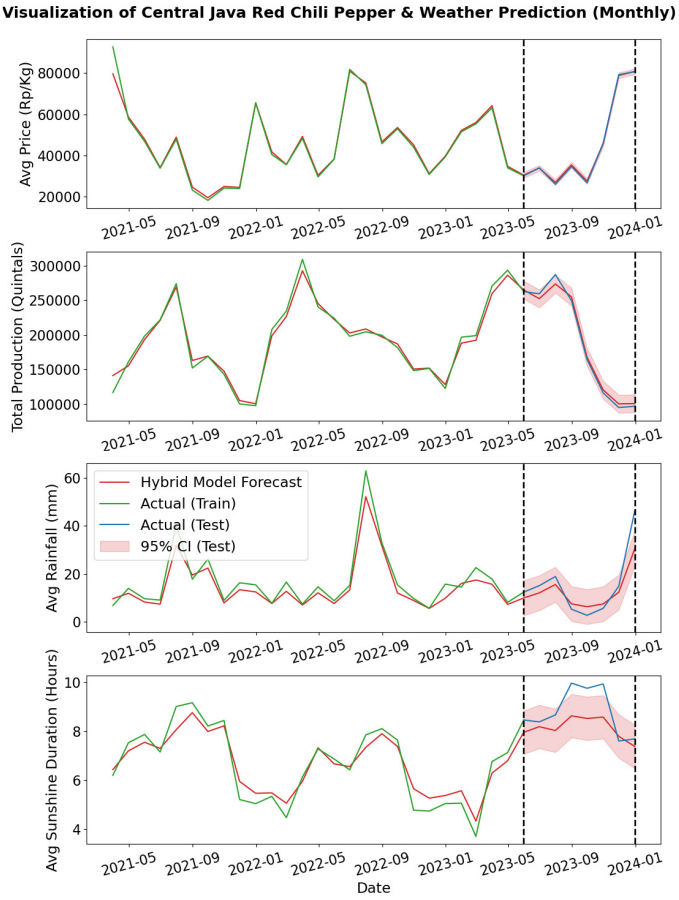

Fig. 12. Actual data and hybrid model forecast graph on Central Java province.

The confidence intervals provide important insight into uncertainty behavior across variables. Price and production exhibit moderately narrow confidence bands, reflecting relatively stable variability and supporting their use for medium-term planning. In contrast, the third graph shows clearer discrepancies between actual and predicted values, particularly during the test period. While the predicted rainfall series successfully preserves the overall seasonal pattern, it consistently smooths several sharp peaks and sudden drops observed in the actual data, especially during months with extreme precipitation. This indicates that the hybrid model tends to underestimate short-term rainfall volatility and extreme events, leading to noticeable deviations between observed and forecasted values despite alignment in general trend. Sunshine duration maintains comparatively tighter bounds consistent with its smoother temporal pattern.

The majority of observed test values remain within the confidence intervals, suggesting that the hybrid model remains well calibrated despite limited gains in responsiveness. Overall, the North Sumatra results emphasize that when linear and seasonal structures dominate the data, the hybrid framework offers incremental improvements and reliable uncertainty quantification, reinforcing the need to interpret forecasts jointly through point estimates and confidence intervals rather than accuracy metrics alone.

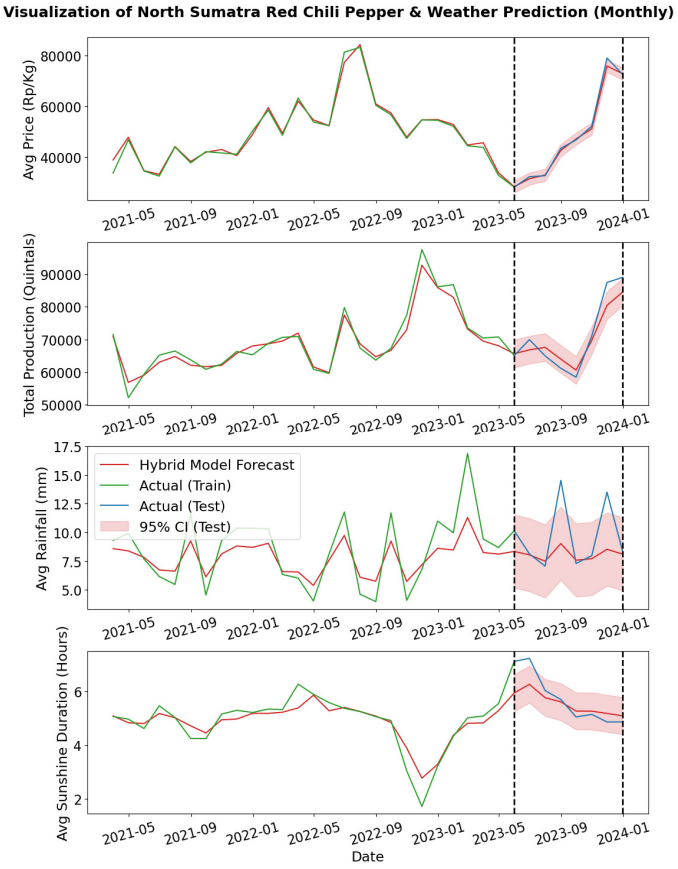

Fig. 13. Actual data and hybrid model forecast graph on North Sumatra province.

IV. CONCLUSION AND FUTURE WORK

This study highlights two main novelties: the implementation of Tree-Structured Parzen Estimator (TPE)-based hyperparameter tuning and the construction of a hybrid Autoregressive Integrated Moving Average–Long Short-Term Memory (ARIMA–LSTM) model that integrates the strengths of both statistical and deep learning approaches. The results confirm that the hybrid ARIMA–LSTM consistently outperforms standalone ARIMA, producing estimates classified as very accurate for price, production, and sunshine duration variables, while being fairly accurate to accurate for rainfall variables.

The benchmark comparison further strengthens these findings. For retail price forecasting, the hybrid ARIMA–LSTM achieves a Mean Absolute Percentage Error (MAPE) of 1.45%, which is substantially lower than ARIMA (42.74%), Seasonal ARIMA (SARIMA) (32.41%), and Temporal Convolutional Network (TCN) (31.73%). This result clearly shows that the hybrid ARIMA–LSTM approach provides

superior predictive performance by effectively combining linear pattern modeling and nonlinear representation learning.

In conclusion, this research demonstrates that the automation of hybrid ARIMA–LSTM model construction, supported by robust preprocessing techniques such as hybrid interpolation, disaggregation, chronological data splitting, and min–max normalization, provides a reliable framework for forecasting agricultural time series data. The integration of ARIMA for modeling linear dependencies and LSTM for capturing nonlinear and long-term patterns, combined through weighted averaging, ensures improved stability and accuracy compared to standalone models. The optimization of hyperparameters using TPE further enhances model performance by systematically exploring the parameter space for both ARIMA and LSTM. Analytical results show that the hybrid approach consistently reduces forecasting errors across multiple variables, validating its technical robustness for use under climate variability.

However, the effectiveness of hybrid models is highly dependent on the quality of available data, especially for weather variables that exhibit high variability. In addition, although the hybrid approach shows high performance for the provinces studied, its generalization to other regions or commodities may vary due to differences in climate behavior, market dynamics, and production systems.

Future research can be directed toward several improvements. First, it is recommended to extend the dataset with longer time horizons, additional factors, or Internet of Things (IoT)-based real-time data to improve forecasting reliability and adaptability to climate variability. Second, applying this framework to other commodities and regions could provide broader benefits for agricultural planning, food availability, and market stability. Finally, since the hybrid results show that LSTM contributes a greater weight in the forecasting process, future studies may focus more on deep learning approaches, such as Gated Recurrent Unit (GRU) or Transformer-based models, to further enhance predictive accuracy and capture complex nonlinear patterns.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support provided by the Doctoral Program in Computer Science (Prodi Doktor Ilmu Komputer), Faculty of Information Technology, Universitas Kristen Satya Wacana (UKSW) Salatiga. The authors also appreciate the valuable contributions and collaboration of all co-authors in developing and refining this research.

DATA AVAILABILITY

The datasets used in this study include retail prices and production figures for red chili peppers obtained from the National Food Agency (BPN), as well as weather data sourced from the Meteorology, Climatology, and Geophysics Agency (BMKG). Public datasets for retail prices and weather variables are available at [15], whereas production data are subject to access restrictions imposed by the data provider.

DECLARATION OF GENERATIVE AI USE

During the preparation of this work, the authors used ChatGPT in order to translate and paraphrase the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

[1] W. A. Zakaria, V. P. Tambunan, and Y. Indriani, "Predicting the cassava economy in Lampung province: An ARIMA-based forecast of supply, demand, and price," *Edelweiss Applied Science and Technology*, vol. 9, no. 4, pp. 2905–2922, Apr. 2025, https://doi.org/10.55214/25768484.v9i4.6698.

[2] S. B. Bhusanar, S. S. Meena, and A. Beniwal, "Forecasting of Arrival and Price of Groundnut in Rajasthan by ARIMA Model for Livelihood of Farmers," *Asian Journal of Agricultural Extension, Economics & Sociology*, vol. 41, no. 9, pp. 684–690, Aug. 2023, https://doi.org/10.9734/ajaees/2023/v41i92092.

[3] M. Omar, F. Hassan, S. Shahin, and M. shahat, "The usage of the autoregressive integrated moving average (ARIMA) model for forecasting milk production in Egypt (2022-2025)," *Open Veterinary Journal*, vol. 14, no. 1, pp. 256–265, Feb. 2024, https://doi.org/10.5455/OVJ.2024.v14.i1.22.

[4] R. Faridi, S. Ahmmed, and S. Ahmmed, "Forecasting Inflation Rate with Algorithmic Arima Modeling," *Indian Journal of Economics and Financial Issues*, vol. 5, no. 2, pp. 143–157, Dec. 2024.

[5] A. Sutany, "Forecasting Weekly Stock of PT. Astra International Tbk (ASII) using ARIMA Model," *Proceeding of The Symposium on Data Science*, vol. 5, pp. 23–32, Aug. 2025.

[6] I. Prihandi, S. Wijono, I. Sembiring, and E. Maria, "Implementation of ARIMA with Min-Max Normalization for predicting the Price and Production Quantity of Red Chili Peppers in North Sumatra Province considering Rainfall and Sunlight Duration Factors," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 21876–21887, Apr. 2025, https://doi.org/10.48084/etasr.9875.

[7] D. Protić *et al.*, "Numerical Feature Selection and Hyperbolic Tangent Feature Scaling in Machine Learning-Based Detection of Anomalies in the Computer Network Behavior," *Electronics*, vol. 12, no. 19, Oct. 2023, Art. no. 4158, https://doi.org/10.3390/electronics12194158.

[8] A. K. Rad *et al.*, "Machine Learning for Determining Interactions between Air Pollutants and Environmental Parameters in Three Cities of Iran," *Sustainability*, vol. 14, no. 13, June 2022, Art. no. 8027, https://doi.org/10.3390/su14138027.

[9] X. Wen and W. Li, "Time Series Prediction Based on LSTM-Attention-LSTM Model," *IEEE Access*, vol. 11, pp. 48322–48331, 2023, https://doi.org/10.1109/ACCESS.2023.3276628.

[10] I. Naouadir, O. E. Ogri, J. E. Mekkaoui, M. Benslimane, and A. Hjouji, "Celestial Object Detection in Astronomical Images Using MSE and Jacobi Moments," *Statistics, Optimization & Information Computing*, vol. 12, no. 3, pp. 660–671, Feb. 2024, https://doi.org/10.19139/soic-2310-5070-1959.

[11] Q. Deng, S. Wang, and J. Lyu, "Comparative effectiveness analysis of univariate time-series forecasting models for disease mortality rates in the global burden of disease database: a case study of global hypertensive heart disease among women of childbearing age," *Frontiers in Public Health*, vol. 13, Nov. 2025, Art. no. 1681569, https://doi.org/10.3389/fpubh.2025.1681569.

[12] A. M. Khan and M. Osińska, "Comparing forecasting accuracy of selected grey and time series models based on energy consumption in Brazil and India," *Expert Systems with Applications*, vol. 212, Feb. 2023, Art. no. 118840, https://doi.org/10.1016/j.eswa.2022.118840.

[13] J. Zhou, H. Li, and W. Zhong, "A modified Diebold–Mariano test for equal forecast accuracy with clustered dependence," *Economics Letters*, vol. 207, Oct. 2021, Art. no. 110029, https://doi.org/10.1016/j.econlet.2021.110029.

[14] D. F. McCaffrey, "Volume 14: Quantitative Research and Educational Measurement," in *International Encyclopedia of Education*, 4th ed., R. J. Tierney, F. Rizvi, and K. Ercikan, Eds. Amsterdam, Netherlands: Elsevier, 2023, pp. xix–xxiv, https://doi.org/10.1016/B978-0-12-818630-5.02014-5.

[15] I. Prihandi, "Public Dataset for Chili Price and Weather Variables in Indonesia (2021–2023)." Zenodo, Jan. 23, 2026, https://doi.org/10.5281/zenodo.18344258.

[16] C. D. Lewis, *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. Oxford, UK: Butterworth Scientific, 1982.

[17] U. Singh, S. Tamrakar, K. Saurabh, R. Vyas, and O. P. Vyas, "Hyperparameter Tuning for LSTM and ARIMA Time Series Model: A Comparative Study," in *2023 IEEE 4th Annual Flagship India Council International Subsections Conference*, Mysore, India, 2023, pp. 1–6, https://doi.org/10.1109/INDISCON58499.2023.10270325.

[18] A. E. Ouardi, B. Er-Raha, and K. Tatane, "LSTM Adaptive Hyperparameter Tuning for Financial Time Series Forecasting Using Custom Gradient-Based Methods: A Comparative Study with Bayesian Optimization," *Journal of Theoretical and Applied Information Technology*, vol. 102, no. 24, pp. 8913–8936, Dec. 2024.