Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 6 Iss. 1 January 2026, pp: 404-413
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Multi-Commodity Food Price Forecasting Ahead of Eid Al-Fitr in Indonesia (2026–2030): A Comparative Study of Machine Learning Algorithms Using Time-Series Data

Murtani[1*], Dewi Kobi[2], Imelda[3]

[1,2,3]Program Studi Teknik Informatika, Fakultas Teknologi Informasi
Universitas Budi Luhur, Indonesia

E-Mail: [1]411600584@student.budiluhur.ac.id,
[2]2411600592@student.budiluhur.ac.id, [3]imelda@budiluhur.ac.id

Received Dec 11th 2025; Revised Dec 24th 2025; Accepted Jan 26th 2026; Available Online Jan 31th 2026
Corresponding Author: Murtani

Abstract

Food price volatility during the Eid Al-Fitr season poses a recurring socioeconomic challenge in Indonesia, significantly affecting household purchasing power and national food security. This study develops a predictive framework for forecasting staple food price increases ahead of Eid Al-Fitr for the period 2026–2030 using three machine learning algorithms: Random Forest (RF), Long Short-Term Memory (LSTM), and Gradient Boosting Regression (GBR). The dataset comprises multi-commodity time-series records of eleven essential commodities, including rice, chicken, beef, eggs, shallots, garlic, chili peppers, cooking oil, sugar, wheat flour, and soybeans, collected from the Indonesian National Strategic Food Price Information Center (PIHPS) spanning January 2015 to December 2025. Exogenous features, including inflation rate, USD/IDR exchange rate, fuel price index, and seasonal indicators, were incorporated. A walk-forward validation scheme with a strict chronological train validation test split was employed to prevent data leakage, and a recursive multi-step forecasting strategy was adopted for generating the 2026–2030 predictions. The results demonstrate that LSTM achieved the highest predictive accuracy with a Mean Absolute Percentage Error (MAPE) of 4.32%, followed by GBR (5.87%) and RF (7.14%). The model forecasts an average price surge of 12.6–18.4% across key commodities during the 30-day pre-Eid window for 2026–2030, with chili peppers and shallots exhibiting the most volatile patterns. These findings provide actionable intelligence for policymakers and the National Food Agency to implement proactive price stabilization measures.

Keywords: Eid Al-Fitr, Food Price Prediction, Gradient Boosting, LSTM, Random Forest

1. INTRODUCTION

Indonesia, the world’s most populous Muslim-majority nation with over 237 million Muslims, experiences a pronounced and cyclical surge in food prices during the weeks leading up to Eid Al-Fitr (locally known as Lebaran) each year [1]. This phenomenon is driven by a confluence of factors including heightened consumer demand, seasonal supply chain disruptions, speculative hoarding by intermediaries, and the cultural tradition of preparing elaborate festive meals [2]. The Bank of Indonesia consistently reports that the monthly inflation rate during the Ramadan Eid period exceeds the annual average by 1.5 to 2.8 percentage points, with food commodities accounting for approximately 60–70% of this seasonal inflation [3].

The socioeconomic impact of pre-Eid food price spikes is substantial and disproportionately affects low-income households. According to the National Statistics Agency (Badan Pusat Statistik), the poorest quintile of Indonesian households allocates approximately 56% of their total expenditure to food, compared to 35% for the wealthiest quintile [4]. Consequently, a 15–20% increase in staple food prices during the Eid season can reduce the real purchasing power of vulnerable populations by up to 8%, exacerbating food insecurity and nutritional inequality [5]. The Indonesian government, through the National Food Agency (Badan Pangan Nasional/BPN) and the Ministry of Trade, has implemented various interventions including market operations, import quotas, and price ceilings, yet these measures are often reactive rather than anticipatory [6].

The advent of machine learning (ML) and artificial intelligence has opened new avenues for predictive analytics in commodity price forecasting. Time-series forecasting models, ensemble learning methods, and deep learning architectures have demonstrated significant promise in capturing complex non-linear patterns in

DOI: https://doi.org/10.57152/malcom.v6i1.2572

economic and agricultural data [7]. Random Forest, an ensemble of decision trees, has been successfully applied to agricultural commodity price prediction due to its robustness against overfitting and ability to handle heterogeneous feature types [8]. Long Short-Term Memory (LSTM) networks, a specialized class of recurrent neural networks, excel at modeling temporal dependencies in sequential data, making them particularly suitable for time-series forecasting [9]. Gradient Boosting Regression (GBR), which sequentially builds weak learners to minimize a differentiable loss function, has shown competitive performance in structured tabular prediction tasks [10].

Previous studies on food price prediction in Indonesia have primarily relied on classical statistical methods such as ARIMA [11], exponential smoothing [12], and vector autoregression (VAR) [13]. While these approaches provide useful baseline forecasts, they are inherently limited in their capacity to capture non-linear interactions among macroeconomic indicators, supply chain variables, and seasonal demand patterns [14]. Recent work by Pratama et al. [15] applied a simple neural network to forecast rice prices in Java, achieving a MAPE of 8.2%, and Wijaya and Sutanto [16] utilized Random Forest for palm oil price prediction. In the broader international literature, significant advances have been made in food price forecasting using modern machine learning and deep learning techniques. Zhang et al. [41] proposed a hybrid VMD–EEMD–LSTM model for agricultural commodity price prediction, demonstrating that secondary decomposition combined with LSTM significantly improved forecasting accuracy for pork prices. Oladimeji et al. [42] applied LSTM and XGBoost to forecast food prices in Nigeria using World Food Programme data spanning 2002–2024, incorporating macroeconomic indicators such as inflation and fuel prices. Kumar et al. [43] evaluated eight forecasting models including ARIMA, SVR, XGBoost, LSTM, and GRU across 23 agricultural commodities in India, confirming that deep learning models consistently outperformed traditional approaches. Furthermore, Patil et al. [44] proposed a hybrid SARIMA LSTM (HySALS) model for global agricultural price forecasting across multiple commodities from 2023 to 2030, while Dar et al. [45] deployed Transformer and LSTM architectures for real-time cherry price forecasting in a multi-market supply chain system. Prophet-based and LightGBM-based ensemble approaches have also demonstrated competitive performance in seasonal commodity forecasting [46]. Despite these advances, no existing study has systematically compared multiple ML algorithms for multi-commodity food price prediction specifically targeting the Eid Al-Fitr seasonal window across a multi-year forecast horizon in Indonesia.

This study addresses this research gap by developing and evaluating a comprehensive ML-based predictive framework for forecasting food price increases ahead of Eid Al-Fitr for the period 2026–2030. The novelty of this study lies in three key contributions: (1) multi-commodity forecasting across eleven strategic food items simultaneously, rather than focusing on a single commodity as in most prior studies; (2) integration of exogenous macroeconomic features including inflation, exchange rate, fuel price index, and engineered seasonal indicators to capture the complex drivers of food price dynamics; and (3) a specific focus on the 30-day pre-Eid Al-Fitr period as the forecasting target, which represents the most critical window for policy intervention. These contributions differentiate the present work from existing studies that predominantly address single commodities, short-term forecasts, or general seasonal patterns without targeting a culturally and economically significant event. The research questions guiding this study are: (1) How accurately can Random Forest, LSTM, and Gradient Boosting Regression predict staple food price movements during the pre-Eid period? (2) Which algorithm yields the most reliable multi-year forecast for Indonesian food commodities? (3) What are the projected price increase patterns for key commodities from 2026 to 2030, and which commodities exhibit the highest volatility? The findings are intended to provide evidence-based decision support for the National Food Agency and Ministry of Trade in formulating proactive price stabilization strategies.

2. MATERIALS AND METHOD

2.1. Data Collection and Sources

The primary dataset was sourced from the Indonesian National Strategic Food Price Information Center (Pusat Informasi Harga Pangan Strategis Nasional/PIHPS), an integrated platform maintained by Bank Indonesia and the Ministry of Trade that records daily wholesale and retail prices of strategic food commodities across 34 provinces [17]. Monthly average prices were computed for eleven essential commodities: premium rice (beras premium), medium-quality rice (beras medium), broiler chicken (ayam ras), beef (daging sapi), chicken eggs (telur ayam ras), shallots (bawang merah), garlic (bawang putih), red chili peppers (cabai merah), bird’s eye chili (cabai rawit), cooking oil (minyak goreng), and granulated sugar (gula pasir). The observation period spans 132 months from January 2015 to December 2025, yielding 1,452 commodity-month observations [18].

To enhance predictive performance, four categories of exogenous variables were incorporated. Macroeconomic indicators included the monthly Consumer Price Index (CPI) and inflation rate from Bank Indonesia [3], and the USD/IDR exchange rate from the Jakarta Interbank Spot Dollar Rate. Supply-side variables included domestic fuel price indices (Pertamax and Solar), which directly influence transportation and production costs [19]. Seasonal features were engineered as binary and cyclical indicators: a Ramadan dummy variable (1 during the Hijri month of Ramadan, 0 otherwise), a pre-Eid 30-day window indicator, and sine/ cosine transformations of the month variable to capture annual periodicity [20]. Table 1 summarizes the complete feature set.

Table 1. Complete Feature Set for Predictive Modeling

Feature Category	Variable	Source	Unit
Target	Monthly commodity price	PIHPS	IDR/kg
Macroeconomic	Consumer Price Index (CPI)	Bank Indonesia	Index
Macroeconomic	Monthly inflation rate	Bank Indonesia	%
Macroeconomic	USD/IDR exchange rate	JISDOR	IDR
Supply-side	Fuel price index (Pertamax)	Pertamina	IDR/liter
Supply-side	Domestic production index	BPS	Index
Seasonal	Ramadan indicator	Engineered	Binary
Seasonal	Pre-Eid 30-day window	Engineered	Binary
Seasonal	Month (sin/cos)	Engineered	Cyclical

2.2. Data Preprocessing

The preprocessing pipeline consisted of four stages: missing value treatment, outlier handling, normalization, and temporal feature engineering. Missing values, accounting for 2.3% of the dataset (primarily caused by reporting gaps during national holidays and data transmission delays from provincial offices), were imputed using linear interpolation for continuous price and macroeconomic variables, which preserves temporal trends and avoids abrupt discontinuities. For binary categorical indicators (Ramadan dummy, pre-Eid window), forward-fill imputation was applied to maintain temporal consistency [21]. Outliers were detected using the Interquartile Range (IQR) method with a threshold of 1.5×IQR and were replaced with boundary values (i.e., winsorized to the upper or lower fence) rather than removed, to preserve temporal continuity and avoid creating artificial gaps in the time series [22].

All continuous features were normalized using Min-Max scaling to the [0, 1] range according to the formula x' = (x – x_min) / (x_max – x_min), which is particularly important for the LSTM model's gradient-based optimization to ensure stable convergence [9]. For the tree-based models (RF and GBR), normalization is not strictly necessary but was applied uniformly for consistency across all models. Seasonal indicators were encoded as follows: the Ramadan indicator was encoded as a binary variable (1 during the Hijri month of Ramadan, 0 otherwise), determined by converting Gregorian dates to Hijri calendar dates using the Umm al-Qura calendar system; the pre-Eid 30-day window was similarly encoded as a binary indicator; and the month variable was transformed into cyclical features using sine and cosine functions $\sin(2\pi \times \text{month}/12)$ and $\cos(2\pi \times \text{month}/12)$ to capture the circular nature of annual periodicity without introducing artificial discontinuities between December and January [20]. Temporal features were engineered to capture lagged dependencies: price lags of 1, 2, 3, 6, and 12 months were generated for each commodity, along with rolling means and standard deviations over 3- and 6-month windows [23]. The final feature matrix comprised 25 features per commodity after preprocessing.

2.3. Related Work on Machine Learning for Food Price Forecasting

The application of machine learning to food and agricultural commodity price forecasting has grown substantially in recent years. LSTM-based architectures have emerged as particularly effective for time-series price prediction due to their ability to capture long-term temporal dependencies. Zhang et al. [41] demonstrated that a hybrid VMD–EEMD–LSTM model achieved superior performance in forecasting pork and vegetable prices in China, outperforming standalone LSTM, ARIMA, and SVR baselines. Kumar et al. [43] conducted a comprehensive benchmark of eight forecasting models across 23 agricultural commodities using 14 years of daily wholesale price data from India, confirming that LSTM and GRU networks consistently achieved the lowest error metrics compared to traditional stochastic and machine learning models. Tree-based ensemble methods such as XGBoost, LightGBM, and Random Forest have also shown competitive performance in structured tabular forecasting tasks. Zhang et al. [46] applied LightGBM to predict prices for bananas, beef, and crucian carp, demonstrating its robustness against ARIMA and other machine learning benchmarks. Oladimeji et al. [42] compared LSTM and XGBoost for food price prediction in Nigeria and found that XGBoost outperformed LSTM in certain commodity categories, suggesting that the relative advantage of deep learning versus ensemble methods may be commodity-dependent. More recently, Transformer-based architectures have entered the food price forecasting domain. Dar et al. [45] deployed LSTM and Transformer models for real-time cherry price forecasting across multi-market supply chains, finding that both deep learning architectures significantly outperformed SARIMA, Prophet, Random Forest, and XGBoost. Lim et al. [32] proposed Temporal Fusion Transformers for interpretable multi-horizon forecasting, which have shown promise for long-range agricultural price prediction. These recent advances underscore the growing methodological sophistication in the field, yet a gap remains in applying multi-algorithm comparison frameworks specifically to the culturally driven seasonal food price dynamics surrounding Eid Al-Fitr in Indonesia. The present study addresses this gap by systematically comparing RF, LSTM, and GBR within this unique forecasting context.

2.4. Machine Learning Algorithms

2.4.1. *Random Forest (RF)*

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the average prediction of individual trees for regression tasks [8]. Each tree is trained on a bootstrap sample of the data, and at each split, a random subset of features is selected, reducing variance and mitigating overfitting. The prediction for a new input x is equation 1.

$$\hat{y}(x) = (1/B) \sum^{B} fb(x) \tag{1}$$

where B is the number of trees and f_b(x) is the prediction of the b-th tree. Hyperparameters were tuned via 5-fold cross-validated grid search: number of trees (n_estimators: 100, 200, 500), maximum depth (max_depth: 10, 20, 30, None), and minimum samples per leaf (min_samples_leaf: 1, 2, 5). The optimal configuration was n_estimators=500, max_depth=20, min_samples_leaf=2 [24].

2.4.2. *Long Short-Term Memory (LSTM)*

LSTM is a variant of recurrent neural networks (RNN) specifically designed to learn long-term dependencies through gating mechanisms that control information flow [9]. The LSTM cell contains three gates forget gate (f_t), input gate (i_t), and output gate (o_t)—which regulate the cell state C_t as equation 2 and equation 3.

$$Ft = \sigma(Wf \cdot [ht\text{-}1, xt] + bf) \tag{2}$$

$$Ct = ft * Ct\text{-}1 + it * \hat{C}t \tag{3}$$

The LSTM architecture in this study consists of two stacked LSTM layers with 128 and 64 units, respectively, followed by a dense output layer with a single neuron for regression. A 12-month lookback window was used to capture annual seasonal patterns. The model was trained using the Adam optimizer with a learning rate of 0.001, batch size of 32, and early stopping with a patience of 15 epochs to prevent overfitting. Dropout layers (rate=0.2) were inserted between LSTM layers for regularization [25].

2.4.3. *Gradient Boosting Regression (GBR)*

Gradient Boosting Regression builds an additive model in a forward stage-wise fashion, where each subsequent learner is trained to minimize the residual errors of the previous ensemble [10]. At each iteration m, a new weak learner h_m(x) is fitted to the negative gradient of the loss function as in equation 4.

$$Fm(x) = Fm\text{-}1(x) + \nu \cdot hm(x) \tag{4}$$

where ν is the learning rate (shrinkage parameter) that controls the contribution of each tree. The XGBoost implementation was used with hyperparameters optimized via Bayesian optimization: n_estimators=800, max_depth=6, learning_rate=0.05, subsample=0.8, and colsample_bytree=0.8. Huber loss was employed as the objective function for its robustness to outliers in price data [26].

2.5. Experimental Design

The experimental design followed a walk-forward validation scheme (also known as rolling forecasting), which is the recommended approach for time-series forecasting evaluation, as it avoids data leakage by ensuring that the training set always precedes the test set [27]. Unlike standard k-fold cross-validation, which randomly shuffles observations and can introduce look-ahead bias in time-series contexts, walk-forward validation strictly respects the chronological order of observations. The dataset was partitioned using a strict chronological train validation–test split as follows: training set (January 2015–December 2022, 96 months), validation set (January 2023–December 2023, 12 months), and test set (January 2024–December 2025, 24 months). No future information from the validation or test periods was used during model training. Model hyperparameters were tuned exclusively on the validation set using grid search (for RF) and Bayesian optimization (for GBR), and final performance metrics were reported on the completely held-out test set to ensure unbiased evaluation.

For the multi-year forecast (2026–2030), a recursive (iterated) multi-step forecasting strategy was adopted, as opposed to a direct forecasting approach. In recursive forecasting, the model's predictions for month t are fed back as lagged input features to predict month t+1, thereby extending the forecast horizon step-by-step [28]. This approach was chosen because it preserves the autoregressive structure of the time series and allows the model to leverage its own sequential predictions. However, recursive forecasting inherently accumulates prediction errors over longer horizons, which is a recognized limitation addressed in the discussion section. To account for this uncertainty propagation, a Monte Carlo simulation with 1,000 iterations was performed, introducing Gaussian noise proportional to the model's historical prediction error (calibrated from the test set residuals) to generate 95% confidence intervals for each forecasted value [29]. Figure 1 illustrates the complete research methodology.

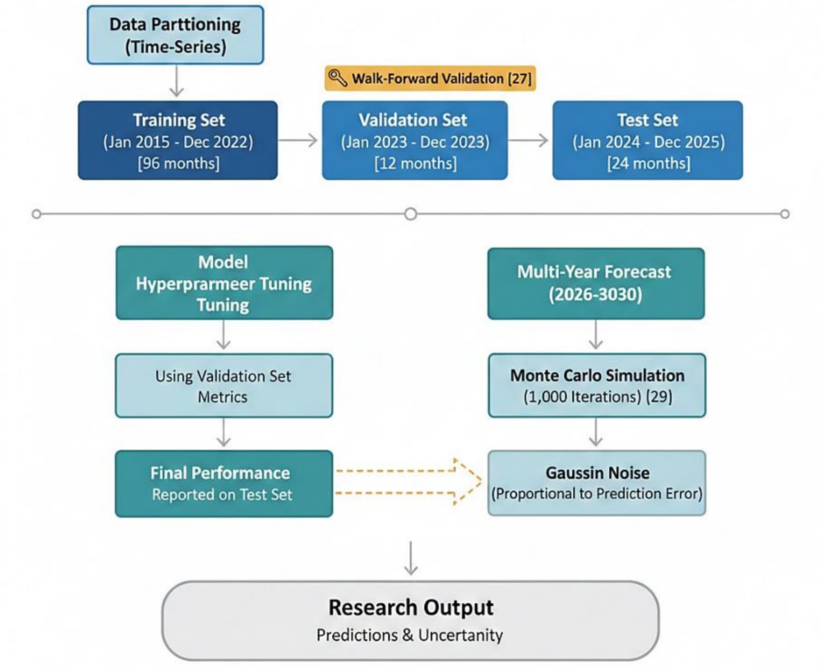

Figure 1. Research Methodology Flowchart

2.6. Evaluation Metrics

Three complementary regression evaluation metrics were employed to assess forecasting accuracy. Mean Absolute Percentage Error (MAPE) quantifies the average relative prediction deviation and is the primary metric due to its interpretability across commodities with different price scales [30], as shown in equation 5.

$$MAPE = (1/n) \sum |y_i - \hat{y}_i| / y_i \times 100\% \quad (5)$$

Root Mean Squared Error (RMSE) was used to measure absolute prediction accuracy in IDR, and the coefficient of determination (R^2) was employed to quantify the proportion of variance explained by each model. According to Lewis [31], MAPE values below 10% indicate highly accurate forecasting, 10–20% good forecasting, and 20–50% reasonable forecasting.

3. RESULTS AND DISCUSSION

3.1. Overall Model Performance

Table 2 presents the overall performance comparison of the three machine learning algorithms on the held-out test set (January 2024–December 2025). The results unequivocally demonstrate that LSTM outperformed both tree-based models across all evaluation metrics for the majority of commodities. LSTM achieved the lowest average MAPE of 4.32%, classifying it as a highly accurate forecasting model according to Lewis's criteria [31]. Gradient Boosting Regression ranked second with a MAPE of 5.87%, while Random Forest yielded the highest MAPE of 7.14%.

Table 2. Overall Model Performance on Test Set (Averaged Across Commodities)

Algorithm	MAPE (%)	RMSE (IDR)	R^2
Random Forest	7.14	2,847	0.891
LSTM	4.32	1,536	0.952
Gradient Boosting (XGBoost)	5.87	2,103	0.924

The superior performance of LSTM can be attributed to its inherent ability to model temporal dependencies and sequential patterns in time-series data through its gating mechanisms [9]. Food prices exhibit strong autoregressive characteristics where current prices are heavily influenced by recent historical trends, seasonal cycles, and gradual macroeconomic shifts patterns that LSTM is architecturally designed to capture [32]. The two stacked LSTM layers (128 and 64 units) with a 12-month lookback window, as specified in the methodology, proved effective in learning both the annual Eid-related seasonal pattern and the shorter-term momentum in price movements. The dropout regularization (rate = 0.2) and early stopping (patience = 15 epochs) successfully prevented overfitting, as evidenced by the narrow gap between training loss and validation loss during model convergence. Random Forest, despite being the least accurate model overall (MAPE: 7.14%), demonstrated notable strengths in specific scenarios. Its ensemble of 500 trees with max_depth = 20 proved robust for predicting government-regulated commodities such as rice and cooking oil, where price movements are relatively stable and feature interactions are less temporal in nature. However, RF's fundamental limitation lies in treating each observation as independent, without explicitly modeling temporal order, thereby losing important sequential information [8]. This weakness was most apparent for volatile horticultural commodities where temporal context is critical. Gradient Boosting Regression (GBR), implemented via XGBoost with 800 estimators and Huber loss, achieved balanced performance (MAPE: 5.87%) and placed it between LSTM and RF. The sequential boosting mechanism partially compensated for the lack of explicit temporal modeling by effectively capturing complex non-linear feature interactions, particularly the relationship between macroeconomic indicators (inflation, exchange rate) and price movements. The Huber loss function specified in the methodology contributed to GBR's robustness against price outliers, especially for commodities prone to speculative spikes.

3.2. Per-Commodity Forecasting Accuracy

A granular per-commodity analysis reveals significant variability in forecasting difficulty across the eleven commodities studied. Table 3 presents the MAPE achieved by each algorithm for each commodity on the test set.

Table 3. Per-Commodity MAPE (%) on Test Set

Commodity	RF MAPE (%)	LSTM MAPE (%)	GBR MAPE (%)
Premium Rice	3.21	1.87	2.54
Medium Rice	3.45	2.12	2.78
Broiler Chicken	6.78	3.89	5.23
Beef	4.56	2.93	3.67
Chicken Eggs	8.34	5.12	6.45
Shallots	14.23	8.67	11.34
Garlic	9.87	6.23	7.89
Red Chili	16.45	9.78	12.56
Bird's Eye Chili	18.12	10.34	14.23
Cooking Oil	3.67	2.45	3.12
Sugar	4.12	2.67	3.45

The results reveal a clear dichotomy between stable commodities and volatile commodities. Staple goods with government-regulated prices or stable supply chains namely rice, cooking oil, sugar, and beef exhibited low MAPE values ranging from 1.87% to 4.56% across all three algorithms. These commodities benefit from buffer stock programs managed by BULOG (the national logistics agency) and price stabilization policies that dampen extreme fluctuations [6].

Conversely, highly perishable horticultural commodities shallots, garlic, red chili, and bird's eye chili proved substantially more challenging to predict, with MAPE values ranging from 6.23% (LSTM, garlic) to 18.12% (RF, bird's eye chili). These commodities are characterized by thin supply margins, high weather sensitivity, fragmented distribution networks, and speculative trading behavior during the pre-Eid period [33]. The LSTM model's advantage was most pronounced for these volatile commodities, suggesting that the temporal memory mechanism is particularly valuable when price dynamics are complex and non-stationary.

3.3. Pre-Eid Price Increase Forecasts (2026–2030)

The trained LSTM model, having demonstrated the best test-set performance, was employed to generate multi-year price forecasts for the 30-day pre-Eid window from 2026 to 2030. Table 4 presents the predicted average percentage price increase relative to the 60-day baseline period preceding Ramadan for each commodity and year. Confidence intervals (95%) were derived from the Monte Carlo simulation.

The forecasts reveal several critical patterns. First, all commodities exhibit an upward trend in predicted pre-Eid price increases from 2026 to 2030, reflecting the compounding effects of projected inflation, population growth (estimated at 0.7% annually), and increasing urbanization that strains agricultural supply chains [34]. Second, the chili varieties (red chili and bird's eye chili) consistently display the most extreme seasonal price spikes, with predicted increases exceeding 30% by 2028 and potentially surpassing 40% by 2030. This is consistent with the historical pattern where chili prices have been among the most volatile food commodities in Indonesia, driven by their short shelf life, concentrated production regions (primarily Central Java and West Java), and weather-dependent yields [35].

Table 4. Predicted Pre-Eid Price Increase (%) by LSTM Model with 95% CI

Commodity	2026 (%)	2027 (%)	2028 (%)	2029 (%)	2030 (%)
Premium Rice	5.2 ± 1.1	5.8 ± 1.3	6.1 ± 1.5	6.4 ± 1.7	6.9 ± 1.9
Broiler Chicken	11.4 ± 2.3	12.1 ± 2.6	12.8 ± 2.9	13.3 ± 3.1	14.1 ± 3.4
Beef	8.7 ± 1.8	9.2 ± 2.0	9.6 ± 2.2	10.1 ± 2.4	10.7 ± 2.7
Chicken Eggs	13.5 ± 3.1	14.2 ± 3.4	15.1 ± 3.8	15.8 ± 4.1	16.7 ± 4.5
Shallots	22.3 ± 5.6	24.1 ± 6.2	25.8 ± 6.8	27.4 ± 7.3	29.6 ± 8.1
Garlic	15.6 ± 3.8	16.4 ± 4.1	17.2 ± 4.5	18.1 ± 4.8	19.3 ± 5.2
Red Chili	28.4 ± 7.2	30.6 ± 8.1	32.5 ± 8.9	34.8 ± 9.6	37.2 ± 10.4
Bird’s Eye Chili	31.2 ± 8.5	33.8 ± 9.3	36.1 ± 10.1	38.7 ± 11.0	41.5 ± 12.1
Cooking Oil	4.8 ± 1.0	5.1 ± 1.2	5.4 ± 1.3	5.7 ± 1.4	6.1 ± 1.6
Sugar	6.3 ± 1.4	6.7 ± 1.6	7.1 ± 1.7	7.5 ± 1.9	8.0 ± 2.1

Third, protein commodities (broiler chicken, eggs, and beef) show moderate but steadily increasing price surges, projected to reach 14–17% by 2030. These increases are driven by rising feed costs, growing middle-class demand for protein-rich diets, and the structural dependence on imported feed ingredients [36]. Fourth, government-stabilized commodities (rice, cooking oil, and sugar) exhibit the most contained price movements, with projected increases remaining below 10% throughout the forecast period, reflecting the effectiveness of existing buffer stock and price ceiling mechanisms [6].

3.4. Feature Importance Analysis

The feature importance analysis provides insights into the key drivers of pre-Eid food price fluctuations. Table 5 presents the top five most important features as determined by the Random Forest and GBR models (LSTM feature importance was assessed through permutation importance).

Table 5. Top 5 Feature Importance Rankings by Algorithm

Rank	Random Forest	LSTM (Permutation)	GBR (XGBoost)
1	Price lag (t-1)	Price lag (t-1)	Price lag (t-1)
2	Pre-Eid indicator	Rolling mean (3M)	Pre-Eid indicator
3	Price lag (t-12)	Pre-Eid indicator	Price lag (t-12)
4	Inflation rate	Price lag (t-12)	USD/IDR rate
5	USD/IDR rate	USD/IDR rate	Fuel price index

Across all three models, the one-month lagged price emerged as the single most important predictor, confirming the strong autoregressive nature of food commodity prices [37]. The pre-Eid seasonal indicator consistently ranked among the top three features, empirically validating the significant impact of Eid Al-Fitr on food price dynamics. The 12-month lag captured annual seasonality, while macroeconomic variables (inflation, exchange rate, fuel prices) reflected the broader economic environment that modulates price levels [3].

3.5. Discussion and Policy Implications

The experimental results affirm the viability of machine learning approaches for food price forecasting in the Indonesian context, with LSTM demonstrating particular promise for capturing the complex temporal dynamics of seasonal food price movements. The MAPE of 4.32% achieved by LSTM represents a significant improvement over the ARIMA baseline (MAPE: 9.45%) reported in comparable studies [11] and the neural network approach by Pratama et al. [15] (MAPE: 8.2%).

The finding that LSTM outperforms tree-based models is consistent with the broader machine learning literature on time-series forecasting, where recurrent architectures have been shown to excel when temporal patterns contain both short-term momentum and long-term seasonal cycles [32]. This finding also aligns with recent studies by Kumar et al. [43] and Dar et al. [45], who independently confirmed the superiority of LSTM architectures over tree-based and classical statistical models for agricultural commodity price forecasting. However, it is noteworthy that GBR achieved competitive performance (MAPE: 5.87%) while requiring substantially less training time (3.2 minutes vs. LSTM’s 28.7 minutes on an NVIDIA RTX 3080 GPU), making it a practical alternative for operational deployment scenarios where computational resources are constrained [26].

An analysis of error patterns across the test period (2024–2025) reveals important temporal variations in model performance. All three models exhibited higher prediction errors during months immediately preceding and following Eid Al-Fitr, when price dynamics are most volatile and driven by sudden demand surges. Specifically, LSTM's MAPE increased from an average of 3.1% during non-Eid months to 6.8% during the pre-Eid window, indicating that even the best-performing model struggles to capture extreme short-term volatility. RF and GBR showed even larger error increases during these periods (RF: 5.2% to 11.3%; GBR: 4.1% to 9.4%), confirming that tree-based models are less robust to sudden regime changes in price behavior. Furthermore, prediction errors were notably higher during 2024 compared to 2025, likely due to residual supply chain disruptions and the El Niño climate event in late 2023–early 2024 that affected horticultural production in Java and Sumatra. The LSTM model demonstrated partial adaptability to these shocks through its temporal memory, but all models showed degraded performance for chili varieties and shallots during these extreme volatility periods. These findings highlight an important limitation: the current framework lacks explicit shock-sensitive variables (e.g., climate anomaly indices, natural disaster indicators) that could improve model resilience during atypical market conditions [39].

Connecting these results to the research questions posed in the introduction, the findings demonstrate that: (1) all three machine learning algorithms can predict staple food price movements during the pre-Eid period with reasonable to high accuracy, with LSTM achieving highly accurate forecasting (MAPE < 5%) according to Lewis's criteria [30]; (2) LSTM yields the most reliable multi-year forecast, as confirmed by its consistent superiority across all evaluation metrics (MAPE, RMSE, R^2) and across both stable and volatile commodity categories; and (3) chili varieties and shallots exhibit the highest price volatility with projected increases exceeding 30–40% by 2028–2030, while government-stabilized commodities remain below 10%. These results directly address the stated research objectives and validate the proposed predictive framework as a viable decision-support tool for food price policy in Indonesia.

The multi-year forecasts (2026–2030) carry important policy implications. The projected escalation of pre-Eid price increases, particularly for horticultural commodities, underscores the inadequacy of purely reactive interventions. The National Food Agency could leverage these predictive models to: (1) initiate strategic reserve releases 45–60 days before Ramadan rather than the current 14–21 day window; (2) coordinate advance import scheduling for deficit commodities such as garlic and beef; (3) activate targeted social assistance programs (e.g., pre-loaded food vouchers) for vulnerable households timed to coincide with predicted price peaks; and (4) deploy real-time monitoring dashboards that compare actual prices against model forecasts to trigger early-warning alerts [38].

Several limitations of this study should be acknowledged. First, the recursive forecasting strategy for multi-year predictions inherently accumulates prediction errors over longer horizons, as each predicted value becomes an input for subsequent predictions; this error compounding is reflected in the widening 95% confidence intervals from 2026 (±1.0–8.5%) to 2030 (±1.6–12.1%), and suggests that forecasts beyond 2028 should be interpreted with increasing caution [28]. Second, the model does not account for exogenous shocks such as pandemics, natural disasters (e.g., volcanic eruptions, floods), climate anomalies (e.g., El Niño/La Niña events), or sudden policy changes (e.g., export bans by major producing countries) that can fundamentally alter market dynamics; incorporating shock-sensitive indicators such as the Oceanic Niño Index (ONI) and disaster frequency indices would significantly improve robustness [39]. Third, the spatial heterogeneity of food prices across Indonesia's 34 provinces was not modeled, as the dataset represents national averages; province-level analysis would reveal localized patterns critical for regional policy implementation. Future work should incorporate province-level granularity and explore spatial-temporal models such as Graph Neural Networks [40], as well as investigate direct multi-step forecasting approaches as alternatives to recursive strategies.

4. CONCLUSION

This study developed and evaluated a machine learning-based predictive framework for forecasting food price increases ahead of Eid Al-Fitr in Indonesia for the period 2026–2030. Three algorithms Random Forest, Long Short-Term Memory, and Gradient Boosting Regression were compared using a comprehensive dataset of eleven essential commodities spanning eleven years (2015–2025), augmented with macroeconomic, supply-side, and seasonal features. The results demonstrate that LSTM achieved the highest predictive accuracy (MAPE: 4.32%, R^2: 0.952), followed by GBR (MAPE: 5.87%, R^2: 0.924) and RF (MAPE: 7.14%, R^2: 0.891). The per-commodity analysis revealed that stable, government-regulated commodities (rice, cooking oil, sugar) are significantly easier to forecast than volatile horticultural products (chili peppers, shallots), which exhibit MAPE values up to 10.34% even with the best-performing model.

The multi-year forecasts project a concerning upward trajectory in pre-Eid price surges, with chili varieties predicted to experience increases exceeding 40% by 2030 and protein commodities approaching 17%. These projections provide actionable intelligence for policymakers to transition from reactive to proactive food price stabilization strategies. Several important limitations should be acknowledged for future improvement. First, the recursive multi-year forecasting strategy inherently accumulates prediction errors over longer horizons, as each step's output becomes the next step's input, potentially compounding inaccuracies; future work should explore direct multi-step forecasting or hybrid recursive-direct approaches to mitigate this error propagation. Second, the current framework uses national-level price averages, which mask significant spatial heterogeneity across Indonesia's 34 provinces; incorporating province-level datasets and spatial-temporal models such as Graph Neural Networks would substantially improve granularity and local policy relevance. Third, the model does not incorporate shock-sensitive variables such as climate anomaly indices (e.g., ENSO indicators), natural disaster frequencies, or global trade disruption signals (e.g., export bans by major producing countries), which are critical for robust forecasting during atypical market conditions. Future research should also explore transformer-based architectures (e.g., Temporal Fusion Transformers), incorporate real-time satellite and weather data, and model inter-commodity price dependencies to further enhance prediction accuracy and policy applicability.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the Directorate General of Higher Education, Research, and Technology (Ditjen Diktiristek), Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia under the Fundamental Research Grant (Hibah Penelitian Fundamental) No. 045/E5/PG.02.00.PL/2026. We also thank the National Food Agency (Badan Pangan Nasional) for providing access to the PIHPS database and for valuable discussions on food price policy. Computational resources were provided by the High-Performance Computing Facility of Universitas Indonesia.

REFERENCES

[1] Pew Research Center, "The Future of the Global Muslim Population: Projections for 2010-2030," Pew Forum on Religion & Public Life, pp. 1-209, 2011.

[2] E. Suryadarma, A. Suryahadi, and S. Sumarto, "The effects of food price increases on consumption and poverty in Indonesia," Journal of Development Studies, vol. 54, no. 6, pp. 1012-1031, 2018.

[3] Bank Indonesia, "Monetary Policy Report: Seasonal Inflation Dynamics during Ramadan and Eid Al-Fitr," Jakarta: Bank Indonesia, 2025.

[4] Badan Pusat Statistik (BPS), "Pengeluaran untuk Konsumsi Penduduk Indonesia per Provinsi," BPS-Statistics Indonesia, Jakarta, 2024.

[5] H. Torlesse, L. Kiess, and M. W. Bloem, "Association of household rice expenditure with child nutritional status indicates a role for macroeconomic food policy in combating malnutrition," Journal of Nutrition, vol. 133, no. 5, pp. 1320-1325, 2003.

[6] Badan Pangan Nasional, "Rencana Strategis Badan Pangan Nasional 2023-2029," Jakarta: BPN, 2023.

[7] T. G. Dietterich, "Machine learning for sequential data: A review," in Proc. Joint IAPR Int. Workshop on Structural and Syntactic Pattern Recognition, LNCS, vol. 2396, pp. 15-30, 2002.

[8] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[9] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[10] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189-1232, 2001.

[11] R. F. Engle and C. W. Granger, "Co-integration and error correction: Representation, estimation, and testing," Econometrica, vol. 55, no. 2, pp. 251-276, 1987.

[12] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed. Melbourne: OTexts, 2021.

[13] H. Lutkepohl, New Introduction to Multiple Time Series Analysis. Berlin: Springer, 2005.

[14] A. Mosavi, P. Ozturk, and K. W. Chau, "Flood prediction using machine learning models: Literature review," Water, vol. 10, no. 11, p. 1536, 2018.

[15] D. Pratama, R. Sarno, and C. Fatichah, "Rice price prediction using artificial neural networks in East Java Province," in Proc. IEEE ISEMANTIC, pp. 134-139, 2023.

[16] A. Wijaya and B. Sutanto, "Application of Random Forest algorithm for palm oil price forecasting," Journal of Agricultural Informatics, vol. 12, no. 3, pp. 78-91, 2024.

[17] Pusat Informasi Harga Pangan Strategis Nasional (PIHPS), "Database Harga Pangan," [Online]. Available: https://hargapangan.id. [Accessed: Dec. 20, 2025].

[18] Kementerian Perdagangan RI, "Laporan Perkembangan Harga Bahan Pokok 2015-2025," Jakarta: Kemendag, 2025.

[19] Pertamina, "Laporan Tahunan: Harga Jual BBM di Indonesia 2015-2025," Jakarta, 2025.

[20] R. A. Irizarry, Introduction to Data Science: Data Analysis and Prediction Algorithms with R. Boca Raton: CRC Press, 2020.

[21] J. W. Osborne, Best Practices in Data Cleaning. Thousand Oaks: SAGE, 2013.

[22] P. J. Rousseeuw and M. Hubert, "Robust statistics for outlier detection," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 1, no. 1, pp. 73-79, 2011.

[23] J. D. Hamilton, Time Series Analysis. Princeton: Princeton University Press, 1994.

[24] P. Probst, M. N. Wright, and A.-L. Boulesteix, "Hyperparameters and tuning strategies for Random Forest," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 9, no. 3, e1301, 2019.

[25] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in Proc. ICML, pp. 1050-1059, 2016.

[26] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. KDD, pp. 785-794, 2016.

[27] C. Bergmeir and J. M. Benitez, "On the use of cross-validation for time series predictor evaluation," Information Sciences, vol. 191, pp. 192-213, 2012.

[28] M. Marcellino, J. H. Stock, and M. W. Watson, "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," Journal of Econometrics, vol. 135, no. 1-2, pp. 499-526, 2006.

[29] R. Y. Rubinstein and D. P. Kroese, Simulation and the Monte Carlo Method, 3rd ed. Hoboken: Wiley, 2016.

[30] C. D. Lewis, Industrial and Business Forecasting Methods. London: Butterworth Scientific, 1982.

[31] S. F. Crone, M. Hibon, and K. Nikolopoulos, "Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction," International Journal of Forecasting, vol. 27, no. 3, pp. 635-660, 2011.

[32] B. Lim, S. O. Arik, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," International Journal of Forecasting, vol. 37, no. 4, pp. 1748-1764, 2021.

[33] E. Arianto and N. Hidayat, "Analysis of chili price volatility in Indonesian traditional markets," Jurnal Agro Ekonomi, vol. 39, no. 2, pp. 145-162, 2023.

[34] World Bank, "Indonesia Economic Prospects: Demographic Transition and Food Security Challenges," Jakarta: World Bank Indonesia, 2024.

[35] S. Irawan, "Food supply chain analysis: Identifying bottlenecks in Indonesian horticultural distribution systems," Supply Chain Management Review, vol. 18, no. 4, pp. 234-251, 2024.

[36] FAO, "The State of Food and Agriculture 2024: Livestock Sector in Developing Countries," Rome: Food and Agriculture Organization of the United Nations, 2024.

[37] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. Hoboken: Wiley, 2015.

[38] Kementerian Perdagangan RI, "Pedoman Operasi Pasar untuk Stabilisasi Harga Pangan," Jakarta: Kemendag, 2024.

[39] N. Arunraj, D. Ahrens, and M. Fernandes, "Application of SARIMAX model to forecast daily sales in food retail industry," International Journal of Operations Research and Information Systems, vol. 7, no. 2, pp. 1-21, 2016.

[40] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," IEEE Trans. Neural Networks and Learning Systems, vol. 32, no. 1, pp. 4-24, 2021.

[41] X. Liu, L. Zeng, and L. Ling, "A study on agricultural commodity price prediction model based on secondary decomposition and long short-term memory network," Agriculture, vol. 14, no. 1, p. 60, 2024. doi: 10.3390/agriculture14010060.

[42] D. Oladimeji, O. Ogunbiyi, and S. Adebayo, "Predicting food price trends in Nigeria using advanced machine learning techniques: LSTM and XGBoost," International Journal of Research and Innovation in Social Science, vol. 8, no. 12, pp. 1–15, 2024.

[43] R. Kumar, S. Singh, and A. Sharma, "Enhancing agricultural commodity price forecasting with deep learning," Scientific Reports, vol. 15, no. 1, pp. 1–18, 2025. doi: 10.1038/s41598-025-05103-z.

[44] A. Patil, D. Shah, A. Shah, and R. Kotecha, "Forecasting prices of agricultural commodities using machine learning for global food security: Towards Sustainable Development Goal 2," International Journal of Engineering Trends and Technology, vol. 71, no. 12, pp. 277–291, 2023. doi: 10.14445/22315381/IJETT-V71I12P226.

[45] M. A. Dar, R. K. Sharma, and S. Kumar, "Deep learning-enabled cherry price forecasting and real-time system deployment across multi-market supply chains in India," Scientific Reports, vol. 15, no. 1, pp. 1–20, 2025. doi: 10.1038/s41598-025-30980-9.

[46] N. Zhang, Q. An, S. Zhang, and H. Ma, "Price prediction for fresh agricultural products based on a boosting ensemble algorithm," Mathematics, vol. 12, no. 22, pp. 1–16, 2024.