Vol. 8, No. 4, August 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Residual-Based Hybrid ARIMA–Prophet for Medium Rice Price Forecasting in Kudus City

Aulia Nurhaliza[1], Supriyono[2], Diana Laily Fithri[3]

[1]Information System, Undergraduate Program, Muria Kudus University, Kudus, Indonesia

[2,3]Information System, Faculty of Engineering, Muria Kudus University, Kudus, Indonesia

Received:
February 1, 2026
Revised:
August 6, 2026
Accepted:
August 31, 2026
Published:
September 4, 2026

Corresponding Author:
Author Name*:
Aulia Nurhaliza
Email*:
202253015@std.umk.ac.id

DOI:
10.63158/journalisi.v8i4.1874

Abstract. Rice prices as a primary food commodity in Indonesia often fluctuate, affecting food-price stability and public welfare. This study analyzes and forecasts Medium I and Medium II rice prices in Kudus City using ARIMA, Prophet, Hybrid ARIMA–Prophet, Random Forest, and benchmark models. Daily PIHPS price data from 2021–2025 were divided chronologically into 781 training observations from 2021–2023 and 523 testing observations from 2024–2025. Model performance was evaluated using MAE, RMSE, MAPE, R^2, and directional metrics. Hybrid ARIMA–Prophet achieved MAE of Rp369.68 and MAPE of 2.62% for Medium I, and MAE of Rp467.31 and MAPE of 3.38% for Medium II. However, simpler benchmark models achieved lower price-level errors than the hybrid model. Random Forest showed higher directional Accuracy than Hybrid ARIMA–Prophet, but its low Precision and F1-Score indicate that Accuracy should be interpreted cautiously. The directional analysis also showed a highly imbalanced distribution, with unchanged movements accounting for approximately 94% of test observations. The Diebold–Mariano test, applied only to the regression forecast errors of Hybrid ARIMA–Prophet and Random Forest, showed significant differences for both Medium I and Medium II ($p < 0.001$). Overall, the findings indicate that more complex models do not necessarily outperform simpler benchmarks for the studied rice-price series.

Keywords: Hybrid ARIMA-Prophet, Random Forest, Directional Forecasting, PIHPS, Diebold-Mariano Test.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

1. INTRODUCTION

Rice is a food commodity in Indonesia that plays a vital role in supporting national food security and is a staple food for the majority of the population in Asia [1]. Rice-price stability is important for maintaining purchasing power and economic stability, while uncontrolled price fluctuations can affect inflation and public welfare. Therefore, appropriate monitoring and policy measures are needed to maintain rice-price stability [2]. Rice prices in Indonesia fluctuate due to seasonal changes, the COVID-19 pandemic, and supply-and-demand dynamics [3]. Increased demand and decreased production can lead to shortages and price increases [4]. Macroeconomic factors, including inflation, exchange rates, the BI-7 Day Reverse Repo Rate, and the ENSO index, can also influence rice prices [5]. In addition, climatic conditions such as rainfall may affect rice production and consequently contribute to price fluctuations [6]. These conditions highlight the need for accurate price forecasting to support food-price monitoring and food-security decision-making.

ARIMA (Autoregressive Integrated Moving Average) is a statistical method widely employed for time series analysis and forecasting [7]. It consists of autoregressive (AR), integrated (I), and moving-average (MA) components and can effectively capture linear temporal patterns. However, its ability to represent complex seasonal or nonlinear patterns may be limited. Prophet is a time series forecasting method designed to model trend and seasonal patterns within observed data [8]. It can represent changes in trend and recurring seasonal structures while providing interpretable model components. However, its performance may be limited when data contain highly volatile or irregular short-term fluctuations.

Previous research using ARIMA to forecast prices of four food commodities in Indonesia showed relatively low forecasting errors [9]. Prophet can capture trend and seasonal components without extensive preprocessing, although it may be less effective in capturing short-term autocorrelation patterns [10]. A comparison of Prophet and SARIMAX for food-commodity prices found that Prophet produced a slightly lower MAPE of 3.91% [11]. These characteristics suggest that combining ARIMA and Prophet may provide complementary modeling capabilities. The Hybrid ARIMA-Prophet approach can

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

use ARIMA to model the main linear structure and Prophet to model the remaining residual patterns [12].

Random Forest is a machine learning method based on the ensemble learning approach that combines multiple decision trees to generate predictions [13]. It can be applied to regression problems and provides a nonlinear machine-learning benchmark for continuous price forecasting. The Diebold–Mariano (DM) test is a statistical procedure used to compare the predictive accuracy of two competing forecasting models [14]. Its null hypothesis states that the competing models have equal expected forecast loss.

However, previous research has generally examined individual forecasting models, while the application of residual-based Hybrid ARIMA–Prophet modeling to city-level rice-price forecasting using daily PIHPS data remains limited. In particular, Medium I and Medium II rice prices in Kudus City provide two separate operational PIHPS price series that can be analyzed independently. This study therefore evaluates these two prices series using data from 2021–2025, with 2021–2023 used for training and 2024–2025 for testing.

Kudus City is selected as the study context because city-level rice-price forecasts can provide information relevant to local food-price monitoring. Early identification of potential rice-price increases may support local inflation monitoring, market-operation planning, and the timing of buffer-stock interventions. The findings are intended as context-specific evidence for the studied Kudus City price series and are not assumed to be broadly generalizable to other regions or rice-price categories without further validation. Medium I and Medium II are treated as separate operational price series in the PIHPS dataset rather than as a single combined medium-grade rice category. The study therefore focuses on their price behavior and forecasting performance rather than on the physical or nutritional classification of rice. The regulatory classification is included only to provide context for the terminology used in the PIHPS data [15].

The research gap is the limited application of a residual-based Hybrid ARIMA–Prophet approach to city-level rice-price forecasting using daily PIHPS data, particularly for separately modeled Medium I and Medium II prices in Kudus City. The study also emphasizes both price-level and direction-based evaluation. The practical motivation is to provide evidence that can support local food-price monitoring and food-security early warning, with potential use in market-operation planning and buffer-stock timing.

This study aims to: (1) analyze the characteristics and movement patterns of Medium I and Medium II rice prices in Kudus City; (2) implement ARIMA and Residual modeling Prophet methods for price prediction; (3) develop a Hybrid ARIMA-Prophet model using combination between ARIMA results and Prophet residual modeling results; (4) compare the forecasting performance of the models with machine learning Random Forest; (5) evaluate model performance using both price-level and direction-based metrics, including MAE, RMSE, MAPE, R^2, Accuracy, Precision, Recall, and F1-Score; and (6) determine whether the difference in regression forecast errors between Hybrid ARIMA–Prophet and Random Forest is statistically significant using the Diebold–Mariano test.

2. METHODS

2.1. Research Object

The research object consists of medium rice price data in Kudus City obtained from the National Strategic Food Price Information Center (Pusat Informasi Harga Pangan Strategis Nasional) [16]. The dataset contains daily price observations for two commodities, namely Medium I Rice and Medium II Rice, covering the period from January 2021 to December 2025. The data were analyzed as time-series observations to identify temporal patterns and develop rice-price forecasting models.

2.2. Data Analysis Method

The data analysis method follows the flowchart presented in Figure 1, which consists of several stages. The analysis consists of sequential stages, beginning with data collection, followed by data preprocessing and pattern analysis. The resulting data characteristics are then used to determine the appropriate modeling approach. Three modeling approaches are implemented: ARIMA, Prophet, and a Hybrid ARIMA–Prophet model, while Random Forest is used as a machine-learning benchmark. The final stage evaluates and compares the forecasting performance of the models using regression and directional forecasting metrics.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

The methodological framework is designed to address the characteristics of rice-price time series that may contain both linear and nonlinear patterns. ARIMA is employed to capture temporal dependence and linear patterns, while Prophet is applied to model patterns remaining in the ARIMA residuals. The Hybrid ARIMA–Prophet model therefore combines the forecasts generated from the primary ARIMA component and the residual component modeled by Prophet. Random Forest is subsequently used as a nonlinear machine-learning benchmark. The complete research procedure consists of: (1) data collection from PIHPS Nasional, (2) data preprocessing, (3) data pattern analysis, (4) ARIMA modeling, (5) Prophet modeling of ARIMA residuals, (6) development of the Hybrid ARIMA–Prophet model, (7) Random Forest modeling, and (8) model evaluation and comparison.

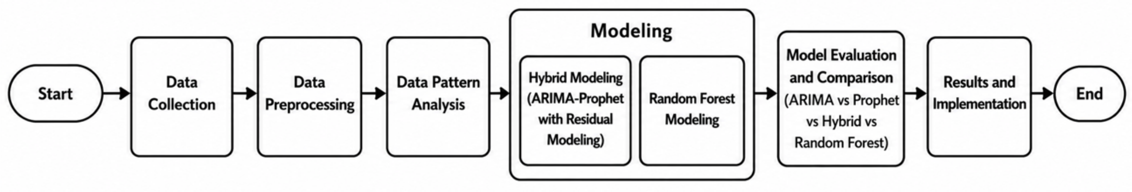

Figure 1. Research Data Analysis Flowchart

2.3. Data Collection

Data collection was conducted using a documentation method a documentation method by retrieving secondary data from the official PIHPS Nasional website [16]. The data were accessed on May 3, 2026, and obtained in Excel format and retained in their original structure before preprocessing. The raw dataset is organized in a wide format, in which each rice commodity is represented as a row and each observation date is represented as a separate column. Two commodities were considered, namely Medium I Rice and Medium II Rice, with observations covering the period from January 2021 to December 2025.

Table 1 presents an example of the raw dataset. As shown in the table, the observation for 1 January 2021 is represented by "–", whereas the observations for 2 and 3 January 2021 are 10,250 for Medium I and 9,750 for Medium II. At the end of the observation period, the recorded prices on 31 December 2025 are 14,150 for Medium I and 13,900 for Medium II. The raw data were subsequently reorganized into a chronological time-series structure for further analysis. Dates for which PIHPS did not provide an observation were not artificially filled or synthesized.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 1. Example Data Structure of Rice Price

No.	Commodity	2021-01-01	2021-01-02	2021-01-03	...	2025-12-31
1	Medium I	-	10,250	10,250	...	14,150
2	Medium II	-	9,750	9,750	...	13,900

2.4. Data Preprocessing

Data preprocessing was conducted before pattern analysis and model development. The preprocessing procedure consisted of missing-value checking, initial visualization, and data transformation assessment.

1) Handling Missing/Absent Observations

The raw dataset was examined to identify missing or absent price observations. In the original PIHPS dataset, an absent observation may be represented by a dash (–), indicating that no price observation was available for the corresponding date. Such observations were not replaced with zero values, interpolated values, or artificially generated prices. Instead, the available PIHPS observations were retained in their original form, while dates without reported observations were excluded from the available observation sequence. This procedure was applied to preserve the original information reported by PIHPS and to avoid introducing artificial price values into the time series.

Table 2. Result After Handling Missing/Absent Data

Date	Medium I	Medium II
2021-01-02	10,250	9.750
2021-01-03	10,250	9,750

2) Data Restructuring

After handling absent observations, the raw data were restructured from a wide format into a chronological time-series format. In the raw dataset, rice commodities were arranged by row, while observation dates were represented as separate columns. The restructuring process converted the date columns into a single date variable and the corresponding price observations into a single price variable. The commodity name was retained as an identifier so that Medium I and Medium II Rice could be analyzed separately. The resulting structure consisted of three main variables: date, commodity,

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

and price. The observations were then sorted chronologically according to their observation dates before subsequent pattern analysis and modeling.

Table 3. Result of Data Restructuring

Date	Commodity	Price
2021-01-02	Medium I	10,250
2021-01-03	Medium I	10,250
2021-01-02	Medium II	9,750
2021-01-03	Medium II	9,750

3) Initial Data Visualization

Initial visualization was performed to obtain an overview of the temporal behavior of the two rice-price series. Line plots were used to observe changes in price over time and to identify possible trends, fluctuations, and recurring seasonal movements. Visualization was also used as an exploratory step before statistical pattern analysis.

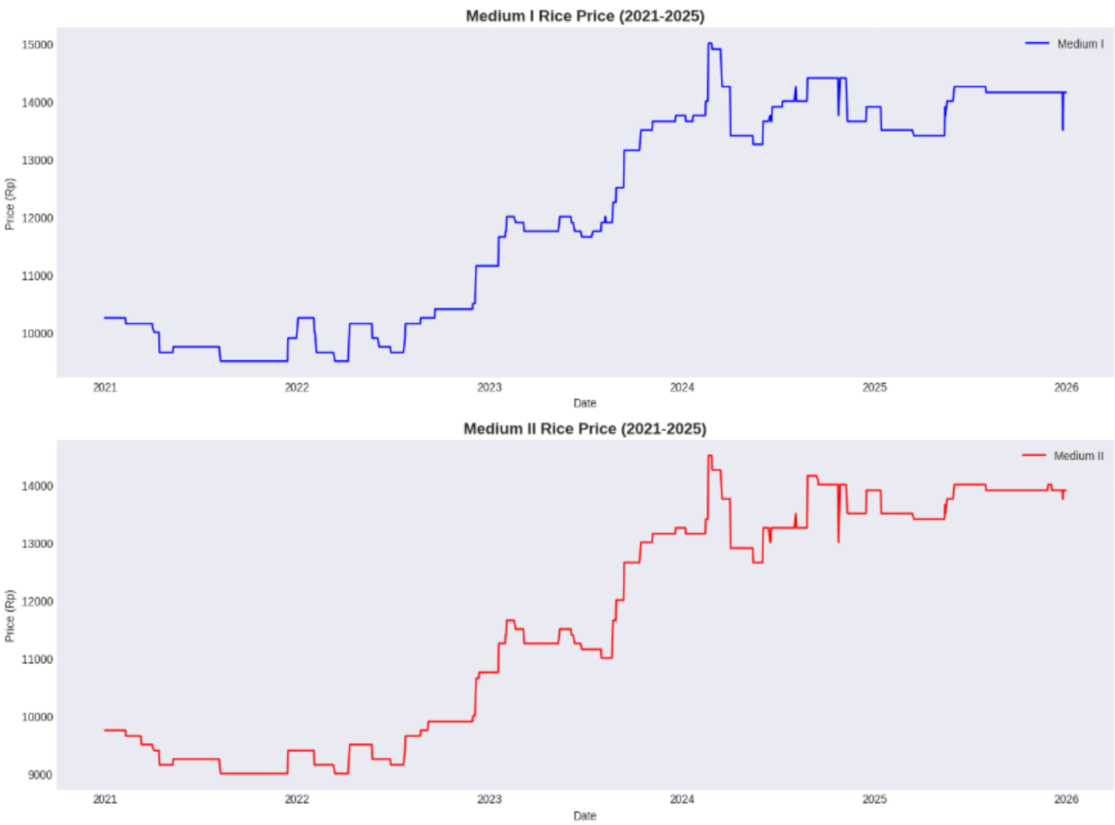

Figure 2. Data Visualization

4) Data Transformation

Data transformation was assessed to determine whether the characteristics of the series required adjustment before statistical modeling. In particular, the series were examined for characteristics such as unstable variance and non-normal behavior. Based on the analysis, the rice-price data were considered sufficiently stable in scale and therefore no additional scale transformation was applied.

2.5. Data Pattern Analysis

Data pattern analysis was conducted to identify the statistical and temporal characteristics of the rice-price series before model estimation. The analysis consisted of trend identification, seasonality identification, and stationarity testing using the Augmented Dickey–Fuller (ADF) test.

1) Trend Identification

Trend identification was performed to determine long-term data movement of the rice-price series. A trend represents the general direction of the series over time, which may be increasing, decreasing, or relatively constant. The temporal sequence of observed prices was visualized and examined to determine whether persistent upward or downward movements were present.

2) Seasonality Identification

Seasonality identification was performed to determine whether recurring patterns occurred at specific time intervals. The analysis considered potential monthly, weekly, and yearly patterns. This step was particularly relevant to the subsequent Prophet configuration, in which yearly and weekly seasonality were enabled.

3) Stationarity Test with ADF (Augmented Dickey-Fuller)

The Augmented Dickey–Fuller (ADF) test was applied to determine whether each rice-price series was stationary and to establish the differencing order required by the ARIMA model. The ADF test examines the null hypothesis that a time series contains a unit root, indicating non-stationarity. For the original level series, Medium I and Medium II Rice produced ADF p-values of 0.996464 and 0.996736, respectively, indicating non-stationarity. First-order differencing was subsequently applied. After differencing, both series became stationary with p-values below 0.001. Therefore, the differencing order for both series was set to (d=1).

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

2.6. ARIMA Modeling

The ARIMA model was specified using three parameters, namely (p), (d), and (q), representing the autoregressive order, differencing order, and moving-average order, respectively [12]. The differencing order was determined from the ADF test, while (p) and (q) were selected using stepwise auto-ARIMA based on the Akaike Information Criterion (AIC). Candidate values of (p) and (q) were evaluated from 0 to 3. For both Medium I and Medium II Rice, the ADF analysis resulted in (d=1). Stepwise auto-ARIMA then selected ARIMA(1,1,0) for both series, with AIC values of 8609.6054 for Medium I and 8535.8270 for Medium II. The selected model was subsequently refitted using the training data and used to generate forecasts for the testing period. The residuals generated by the fitted ARIMA model were retained for the subsequent Prophet residual-modeling stage.

2.6.1. ARIMA Residual Extraction

After ARIMA estimation, residuals were calculated as the difference between the observed training price and the corresponding fitted ARIMA value, as shown in Equation 1.

$$[e\_t=y\_t-y\^t^{ARIMA}] \tag{1}$$

where (y_t) represents the observed price and ($y\^t^{ARIMA}$) represents the fitted ARIMA value. The residuals represent the component of the training series that was not explained by the ARIMA model. The residual series was then used as the target variable for Prophet modeling. Importantly, only training-period residuals were used, while actual residuals from the testing period were not used during model construction.

2.6.2. Prophet Modeling

Prophet was applied to model the residual component remaining after ARIMA estimation. The Prophet formulation decomposes a time series into trend, seasonality, and error components, as shown in Equation 2.

$$[y(t)=g(t)+s(t)+h(t)+\varepsilon(t)] \tag{2}$$

where (g(t)) represents the trend component, (s(t)) represents the seasonal component, (h(t)) represents holiday effects, and (ε(t)) represents the error term. In this study, Prophet was configured with yearly and weekly seasonality enabled, while daily seasonality was disabled. The changepoint prior scale was set to 0.05, the seasonality mode was additive, no holiday effects were supplied, and the prediction interval width was set to 0.80. The same configuration was applied to both rice-price series. Prophet was fitted exclusively to the ARIMA residuals from the training period. The resulting residual forecasts were then used as the additional component of the Hybrid ARIMA–Prophet forecast.

2.6.3. Hybrid Model Development

The Hybrid ARIMA–Prophet model combines the ARIMA forecast with the Prophet forecast of the ARIMA residuals. The hybrid forecast is defined as shown in Equation 3.

$$[Y_t=y\hat{}t^{ARIMA}+e\hat{}t^{Prophet}] \quad (3)$$

where ($y\hat{}t^{ARIMA}$) is the ARIMA forecast and ($e\hat{}t^{Prophet}$) is the forecast of the ARIMA residual produced by Prophet.

The implementation follows three sequential steps. First, ARIMA is fitted to the original rice-price series to capture linear temporal dependence. Second, the ARIMA residuals are calculated from the training observations. Third, Prophet is fitted to these residuals and its forecast is added to the ARIMA forecast. Thus, the hybrid model attempts to capture complementary patterns: ARIMA models the primary linear temporal structure, while Prophet models systematic patterns that remain in the ARIMA residuals.

2.6.4. Random Forest Modeling

Random Forest was implemented as a nonlinear benchmark using lagged prices, rolling statistics, and calendar variables. The feature vector consisted of lag 1, lag 7, and lag 30 price variables; 7-day and 30-day rolling means; 7-day and 30-day rolling standard deviations; and calendar variables consisting of month, day, day of week, and day of year. The current target price was excluded from the feature vector to prevent data leakage. The model was configured with 100 trees, unlimited maximum depth, a minimum of two samples required for splitting, one sample as the minimum leaf size, random state 42, and parallel processing using (n_jobs=-1). Random Forest forecasts were generated recursively throughout the testing period. After each prediction, the generated forecast value was used when constructing subsequent lagged and rolling features. Actual prices from the testing period were not fed back into the model, thereby preventing test-set information leakage.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 4. Model Configuration for Medium I and Medium II

Model	Configuration	Medium I	Medium II
ARIMA	ARIMA order (p,d,q)	(1,1,0)	(1,1,0)
	AIC	8609.6054	8535.8270
Prophet	Yearly seasonality	True	True
	Weekly seasonality	True	True
	Daily seasonality	False	False
	Changepoint prior scale	0.05	0.05
	Seasonality mode	Additive	Additive
	Holidays	None	None
	Interval width	0.80	0.80
	Lag feature	Lag 1, 7, 30	Lag 1, 7, 30
	Rolling windows	7-days, 30-days	7-days, 30-days
	Rolling statistic	Mean, standard deviation	Mean, standard deviation
	Calendar features	Month, day, day of week, day of year	Month, day, day of week, day of year
	Number of trees	100	100
	Max depth	None	None
	Minimum samples split	2	2
	Minimum samples leaf	1	1
	Random state	42	42
	n_jobs	-1	-1

2.6.5. Model Evaluation

Model performance was evaluated using two perspectives: price-level forecasting accuracy and directional forecasting performance. Price-level forecasting was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). Directional forecasting was evaluated using Accuracy, macro-Precision, macro-Recall, macro-F1,

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

balanced accuracy, specificity, and Matthews correlation coefficient (MCC). Confusion matrices were also generated to examine class-specific prediction errors. Let (y_t) denote the actual rice price at observation (t), $(\hat{y}_t)$ denote the predicted price, and (n) denote the total number of test observations. Model performance was evaluated using the following metrics:

1) Mean Absolute Error (MAE)

MAE calculates the average absolute difference between predicted and actual values, treating all errors equally regardless of direction. It is intuitive and robust to outliers, with lower values signifying that predictions are, on average, closer to the true observations, as shown in Equation 4.

$$MAE = (1/n) \Sigma |y_t - \hat{y}_t| \quad (4)$$

2) Root Mean Square Error (RMSE)

RMSE computes the square root of the average squared differences between predictions and actuals, heavily penalizing large deviations. It is more sensitive to significant errors than MAE, making it useful when large mistakes are particularly undesirable, and lower values indicate better overall accuracy, as shown in Equation 5.

$$RMSE = \sqrt{[(1/n) \Sigma (y_t - \hat{y}_t)^2]} \quad (5)$$

3) Mean Absolute Percentage Error (MAPE)

MAPE expresses prediction accuracy as a percentage by averaging the absolute percent errors relative to actual values. It provides a scale-independent interpretation, but can be skewed or undefined when actual values are near zero; smaller percentages reflect more precise forecasts, as shown in Equation 6.

$$MAPE = (100/n) \Sigma |(y_t - \hat{y}_t)/y_t| \quad (6)$$

4) Coefficient of Determination (R^2)

R^2 indicates the proportion of total variation in the dependent variable that is explained by the independent variables in the model. Ranging from 0 to 1, values closer to 1 suggest

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

a strong fit, though it does not account for overfitting and should be interpreted alongside other metrics, as shown in Equation 7.

$$R^2 = 1 - [\Sigma(y\_t - \hat{y}\_t)^2 / \Sigma(y\_t - \bar{y})^2] \quad (7)$$

5) Classification Metrics

Directional forecasting was evaluated using three classes: upward, downward, and unchanged. An upward movement was defined as a price increase relative to the previous observation, a downward movement as a price decrease, and an unchanged movement as no price change. Accuracy, macro-Precision, macro-Recall, macro-F1, balanced accuracy, specificity, and Matthews correlation coefficient (MCC) were used to evaluate directional performance. Confusion matrices were also generated to examine class-specific prediction errors, as shown in Equation 8 to 11.

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \quad (8)$$

$$\text{Precision} = TP / (TP+FP) \quad (9)$$

$$\text{Recall} = TP / (TP+FN) \quad (10)$$

$$F1 = 2(\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall}) \quad (11)$$

6) Diebold-Mariano Test

The Diebold–Mariano (DM) test was applied to statistically compare the squared forecast errors of the Hybrid ARIMA–Prophet model and Random Forest. The test was conducted separately for Medium I and Medium II Rice using the test-period forecast errors. The resulting statistics were DM = −10.736 (($p<0.001$)) for Medium I and DM = −14.649 (($p<0.001$)) for Medium II, indicating a statistically significant difference in forecast errors between the two models.

3. RESULTS AND DISCUSSION

3.1. Data Pattern Analysis

Figure 3 shows trend and seasonality identification, Medium I rice price data shows a fluctuating movement pattern with a long-term upward trend. The Augmented Dickey-Fuller (ADF) test results showed that the initial data was not stationary (p-value = 0.858), so first-order differencing was performed. After differencing, the data became stationary with p-value = 0.0. Seasonality analysis revealed recurring annual and weekly patterns in the data, indicating that rice prices are influenced by seasonal factors such as harvest seasons and demand cycles.

Similar patterns were observed in Medium II rice data. The initial ADF test p-value was 0.879, and after differencing, it became 0.0000000000001024. Trend analysis showed a general upward movement with periodic fluctuations. Seasonality analysis also revealed recurring patterns, confirming the presence of seasonal components in both commodities shown on Figure 3.

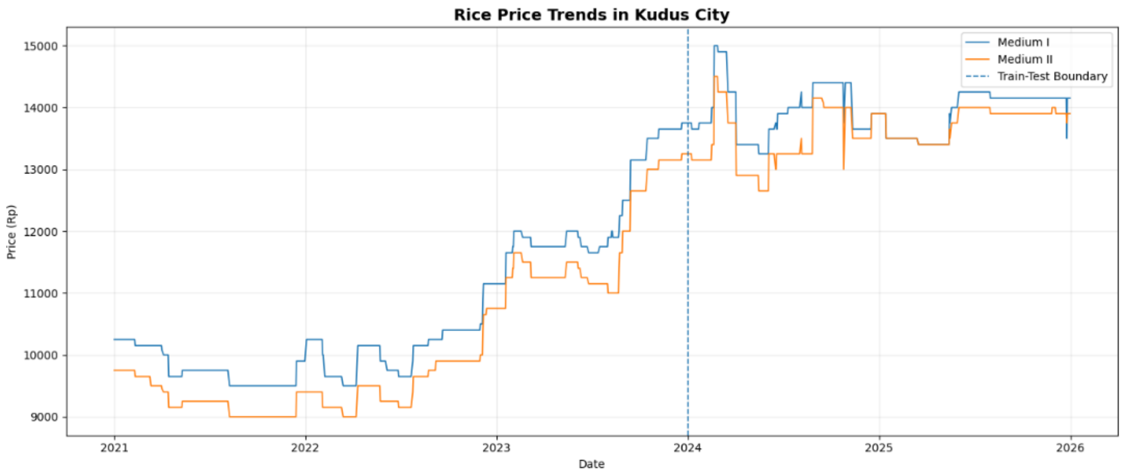

Figure 3. Medium I and Medium II Rice Price Trends

3.2. Model Performance Comparison

Table 5 presents the regression and directional forecasting performance of the evaluated models for Medium I and Medium II rice prices. The results show that the simpler baseline models generally achieved lower price-level forecasting errors than the more complex models. For Medium I, the Naive, Seasonal Naive, and ARIMA-only models produced an MAE of Rp 367.50, RMSE of Rp 433.15, MAPE of 2.61%, and R^2 of −0.216. Hybrid ARIMA–Prophet produced a slightly higher MAE of Rp 369.68, RMSE of Rp 446.91, MAPE of 2.62%, and R^2 of −0.295. Random Forest produced larger errors, with an MAE of Rp 516.21, RMSE of Rp 643.03, MAPE of 3.64%, and R^2 of −1.680.

A similar pattern was observed for Medium II. The Naive, Seasonal Naive, and ARIMA-only models achieved an MAE of Rp 452.92, RMSE of Rp 542.86, MAPE of 3.28%, and R^2 of

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

−0.850. Hybrid ARIMA–Prophet resulted in an MAE of Rp 467.31, RMSE of Rp 554.18, MAPE of 3.38%, and R^2 of −0.928. Random Forest produced an MAE of Rp 682.28, RMSE of Rp 771.60, MAPE of 4.94%, and R^2 of −2.738, while Prophet-only produced an MAE of Rp 679.90, RMSE of Rp 827.21, MAPE of 4.96%, and R^2 of −3.297.

Table 5. Comparison of Regression Metrics for Hybrid ARIMA-Prophet and Random Forest

Commodity	Model	MAE (Rp)	RMSE (Rp)	MAPE (%)	R^2
Medium I	Naïve	367.49	433.15	2.60	-0.21
Medium I	Seasonal Naïve (7-days)	367.49	433.15	2.60	-0.21
Medium I	ARIMA-only	367.49	433.15	2.60	-0.21
Medium I	Prophet-only	1298.02	1586.61	9..33	-15.31
Medium I	Hybrid ARIMA-Prophet	369.67	446.90	2.61	-0.29
Medium I	Random Forest	516.21	643.02	3.63	-1.68
Medium II	Naïve	452.91	542.86	3.27	-0.85
Medium II	Seasonal Naïve (7-days)	452.91	542.86	3.27	-0.85
Medium II	ARIMA-only	452.91	542.86	3.27	-0.85
Medium II	Prophet-only	679.90	827.20	4.96	-329
Medium II	Hybrid ARIMA-Prophet	467.31	554.18	3.38	-0.92
Medium II	Random Forest	682.27	771.59	4.94	-2.73

In addition to regression evaluation, this study also conducted classification evaluation to measure both models' ability to predict price movement direction (up or down). Table 2 presents the classification evaluation results. Table 6 shows a clear difference between price-level and direction-based evaluation. For directional forecasting, Random Forest achieved the highest accuracy among the evaluated models, with 89.46% for Medium I and 90.42% for Medium II. However, these accuracy values should be interpreted cautiously because Precision and F1-Score remained very low. For Medium I, Random Forest achieved a Precision of 2.38% and an F1-Score of 3.51%, while for Medium II, Precision was 4.88% and F1-Score was 7.41%. These results indicate that the high overall accuracy did not translate into reliable identification of individual price movements.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Hybrid ARIMA–Prophet showed substantially lower directional accuracy, with 44.06% for Medium I and 45.02% for Medium II. Its Precision and F1-Score were also low, at 2.41% and 4.58% for Medium I and 2.10% and 4.01% for Medium II, respectively. Therefore, neither model provides strong evidence of reliable directional classification based on the reported classification metrics. Random Forest has higher Accuracy than Hybrid ARIMA-Prophet, but its Precision and F1-score remain very low. Therefore, the Accuracy advantage should not be interpreted as evidence that Random Forest reliably identifies upward movements or is suitable for buy/sell signals.

The low Precision values indicate that positive-direction predictions were frequently incorrect. Accuracy can therefore be misleading when upward, downward, and unchanged movements are unevenly distributed. The revised analysis should report class counts, the treatment of unchanged movements, confusion matrices, balanced accuracy, specificity, MCC, macro-F1, and simple directional baselines before drawing conclusions about directional usefulness. Overall, the directional results do not provide sufficient evidence that either model can reliably predict individual price movements. In particular, the high Accuracy of Random Forest must be interpreted alongside its low Precision, F1-score, and the class distribution.

Table 6. Comparison of Classification for Hybrid ARIMA-Prophet and Random Forest

Commodity	Model	Accuracy (%)	Precision	Recall	F1-Score
Medium I	Naïve	0.97	0.00	0.00	0.00
Medium I	Seasonal Naïve (7-days)	0.97	0.00	0.00	0.00
Medium I	ARIMA-only	0.97	0.00	0.00	0.00
Medium I	Prophet-only	0.37	0.02	0.60	0.05
Medium I	Hybrid ARIMA-Prophet	0.44	0.02	0.46	0.04
Medium I	Random Forest	0.89	0.02	0.06	0.03
Medium II	Naïve	0.97	0.00	0.00	0.00
Medium II	Seasonal Naïve (7-days)	0.97	0.00	0.00	0.00
Medium II	ARIMA-only	0.97	0.00	0.00	0.00
Medium II	Prophet-only	0.44	0.01	0.38	0.03
Medium II	Hybrid ARIMA-Prophet	0.45	0.02	0.46	0.04
Medium II	Random Forest	0.90	0.04	0.15	0.07

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.3. Prediction Visualization

Figure 4 presents a visualization comparing Hybrid ARIMA-Prophet and Random Forest predictions against actual data for both commodities. Based on the visualization, Hybrid ARIMA-Prophet follows the observed price level more closely than Random Forest, consistent with the regression metrics. Random Forest shows larger deviations during substantial price changes, consistent with its negative R^2 values. The final figures should use explicit dates, commodity labels, price units, and model labels.

1) Hybrid ARIMA-Prophet produces a prediction curve very close to actual data, with relatively small deviations. The prediction line almost coincides with the actual line, confirming the very low MAE and MAPE values.
2) Random Forest produces rougher and more deviating predictions, especially during periods with significant price changes. This is reflected in the negative R^2 values it produced.

Residual analysis should be interpreted together with formal diagnostics rather than visual amplitude alone. If ARIMA adequacy is claimed, the final manuscript should include actual residual plots, Ljung–Box results, and ACF/PACF plots used for ARIMA identification.

1) Hybrid ARIMA-Prophet residuals are scattered around zero with small amplitude (average ±Rp 200), indicating that this model has low bias and errors are randomly distributed.
2) Random Forest residuals have larger amplitude (average ±Rp 1,000) and show certain patterns, indicating that this model still has systematic bias in its predictions.

Figure 5 showed that the test data were highly imbalanced toward upward and downward price movements. Medium I contained 15 upward, 14 downward, and 494 unchanged observations, while Medium II contained 13 upward, 17 downward, and 493 unchanged observations. The confusion matrices in Figure 5 show that both models had difficulty correctly identifying the three directional classes. Although Hybrid ARIMA–Prophet and Random Forest correctly identified several downward movements, their ability to identify upward movements was limited, and both models showed poor classification of unchanged movements. Therefore, overall Accuracy should not be interpreted as evidence of reliable directional forecasting, and class-specific metrics are necessary for evaluating the models' directional performance.

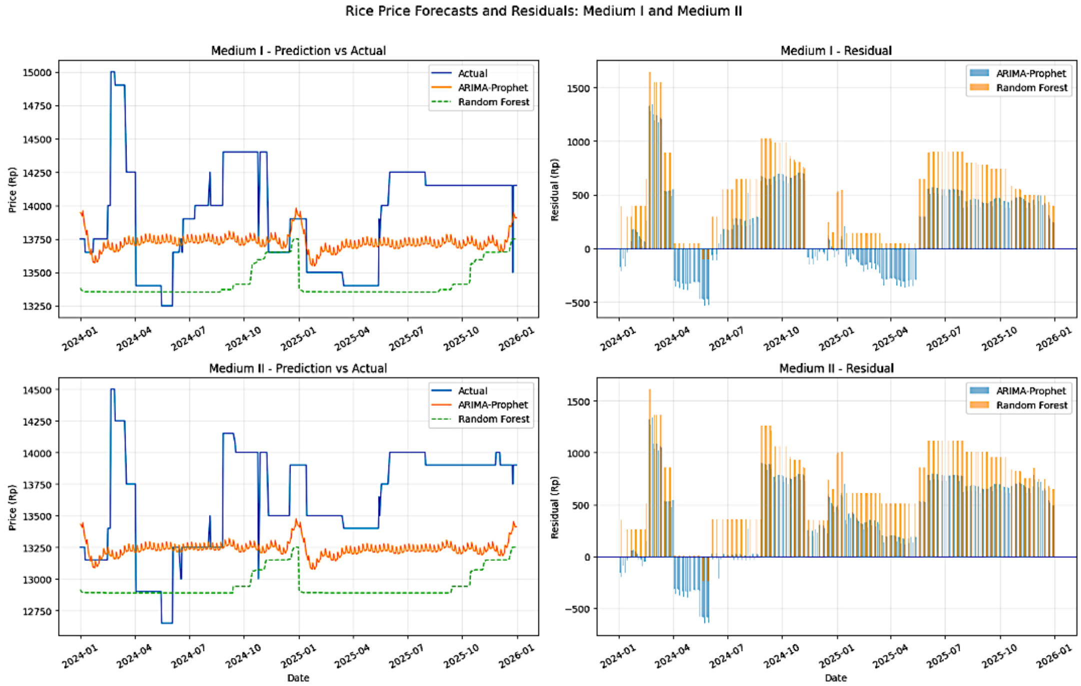

Figure 4. Comparison of Hybrid ARIMA-Prophet and Random Forest Predictions

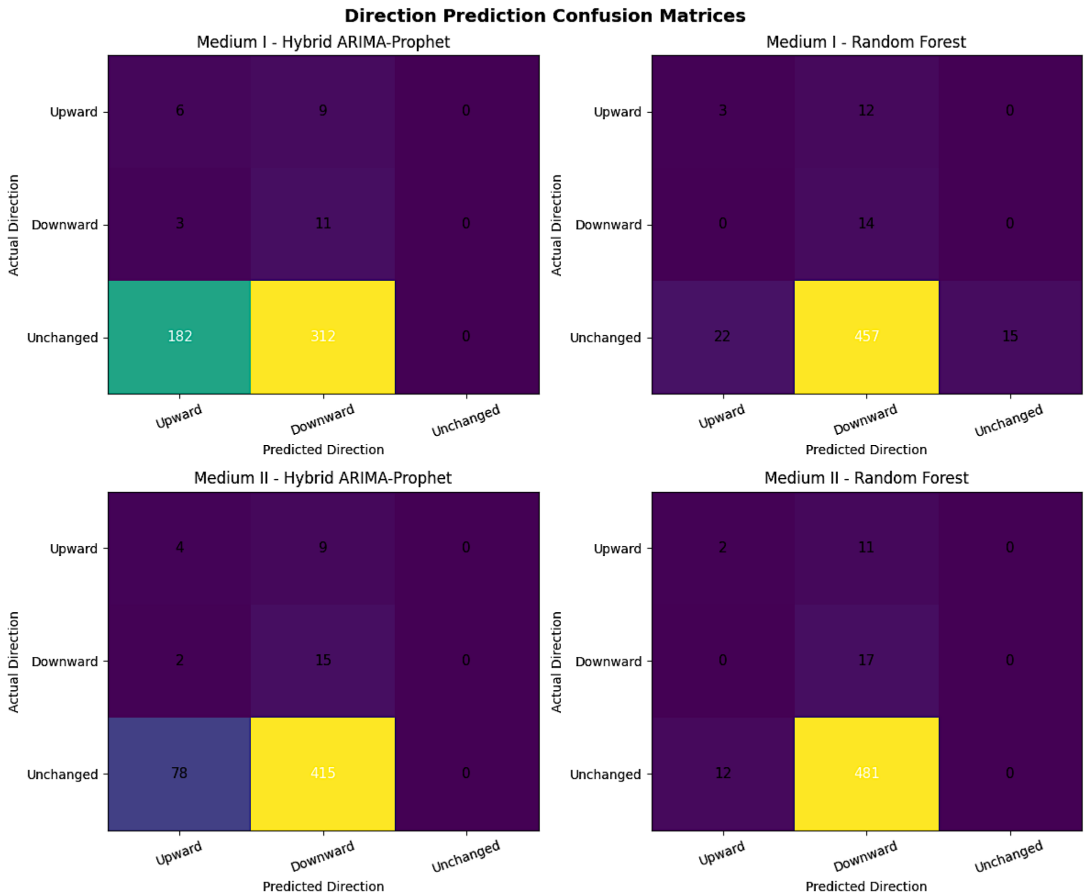

Figure 5. Confusion Matrices

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.4. Diebold-Mariano Test Results

To test whether the performance difference between the two models is statistically significant, the Diebold-Mariano test was conducted. Table 4 presents the test results. Based on Table 7, The Diebold–Mariano (DM) test was applied only to the regression forecast errors of Hybrid ARIMA–Prophet and Random Forest. For Medium I, the DM statistic was −10.736 with a p-value below 0.001, while for Medium II, the DM statistic was −14.649 with a p-value below 0.001. These results indicate a statistically significant difference in regression forecast errors between Hybrid ARIMA–Prophet and Random Forest for both rice-price categories. The lower MAE, RMSE, and MAPE of Hybrid ARIMA–Prophet relative to Random Forest are consistent with this result. However, the DM test compares only these two models and does not establish that Hybrid ARIMA–Prophet is superior to the simpler baseline models. The previous explanation based on approximately 500 testing observations should be removed because the test set contains 731 observations. A nonsignificant DM result should instead be investigated through the paired daily loss differential, loss function, and autocorrelation correction.

Table 7. Diebold-Mariano Test Results

Commodity	DM Statistic	p-value	Conclusion
Medium I	-10.73	0.0	Significant
Medium II	-14.64	0.0	Significant

3.5. Summary of Complete Model Comparison

Table 8 summarizes the currently reported metrics for the two models. ARIMA-only, Prophet-only, naive, and seasonal-naive baselines requested by the reviewer are not present in the current manuscript and should be added after evaluation on the same test period.

Table 8. Complete Comparison of Evaluation Metrics

Metric	Medium I (Hybrid)	Medium I (RF)	Medium II (Hybrid)	Medium II (RF)
MAE	369.67	516.21	467.31	682.27
RMSE	446.90	643.02	554.18	771.59
MAPE (%)	2.61	3.63	3.38	4.94
R^2	-0.29	-1.68	-0.92	-2.73

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Metric	Medium I (Hybrid)	Medium I (RF)	Medium II (Hybrid)	Medium II (RF)
Accuracy (%)	0.44	0.89	0.45	0.90
Precision	0.02	0.02	0.02	0.04
Recall	0.46	0.06	0.46	0.15
F1-Score	0.04	0.03	0.04	0.07

3.6. Discussion

The main strength of this study is the comparison of forecasting approaches with different modeling characteristics, allowing the performance differences to be interpreted from both statistical and machine-learning perspectives. The use of a chronological train–test split preserves the temporal order of the rice-price series and reduces the risk of using future information during model development. In addition, evaluating both price-level and directional performance is important because a model may produce relatively accurate numerical forecasts while still having limited ability to correctly identify individual price movements.

The results indicate that model complexity does not necessarily lead to better forecasting performance for the studied rice-price series. Although Hybrid ARIMA–Prophet produced relatively low MAPE values of 2.62% for Medium I and 3.38% for Medium II, the Naive, Seasonal Naive, and ARIMA-only models achieved lower MAE, RMSE, and MAPE. One possible explanation is that the rice-price series contains strong temporal persistence, meaning that recent or seasonal price information may already provide substantial information for short-term forecasting. Under such conditions, adding a residual forecasting component may not provide sufficient additional information to compensate for the increased model complexity. This result is consistent with Amirudin [12], who reported strong performance from Hybrid ARIMA–Prophet, but differs in terms of the relative advantage of the hybrid model. The difference may be related to variations in dataset characteristics, observation periods, and forecasting settings.

The competitive performance of ARIMA also suggests that the temporal dependence in the rice-price series can be captured effectively through its autoregressive and moving-average structure. This may explain why ARIMA remained competitive with the more

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

complex hybrid approach. The finding is consistent with Dianco & Novita [9], who reported that ARIMA could provide relatively low forecasting errors for food commodities, including rice. Therefore, the present results suggest that model selection should be determined by the characteristics of the target series rather than by model complexity alone.

The different performance between Hybrid ARIMA–Prophet and Random Forest can also be interpreted from their modeling mechanisms. The Hybrid ARIMA–Prophet approach explicitly models the temporal structure of the series through ARIMA and then models the remaining residual component using Prophet. In contrast, Random Forest depends on engineered predictors such as lagged prices, rolling statistics, and calendar variables. If these predictors do not contain sufficient additional information about future price movements, the nonlinear learning capability of Random Forest may not translate into better forecasting accuracy. This may explain why Hybrid ARIMA–Prophet achieved lower regression errors than Random Forest in the tested comparison. The result also differs from Sentosa [6], where SVR outperformed ARIMAX-GARCH, reinforcing that forecasting performance depends on the data structure, predictors, and experimental design rather than on a particular model being universally superior.

For directional forecasting, Random Forest obtained higher Accuracy than Hybrid ARIMA–Prophet, but its low Precision and F1-score indicate that the higher Accuracy should be interpreted cautiously. This may occur when the directional classes are imbalanced, particularly when unchanged or dominant movement classes occur more frequently than other classes. Consequently, a model can obtain relatively high overall Accuracy while performing poorly when identifying specific upward or downward movements. This explains why the regression results and directional results do not necessarily identify the same model as the best performer. Furthermore, because Prophet was used to model the ARIMA residual component, the Hybrid ARIMA–Prophet model should not automatically be interpreted as a nonlinear forecasting approach. Additional residual diagnostics would be required to determine whether the hybrid structure successfully captures patterns that remain unexplained by ARIMA.

The observed forecasting performance should be interpreted within the characteristics of the dataset and experimental design. The use of only two rice-price categories from

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Kudus City limits the extent to which the observed model behavior can be generalized to other commodities or locations. In addition, the models rely primarily on historical price information. The absence of exogenous variables such as weather, policy interventions, supply conditions, and market factors may limit the models' ability to explain price changes caused by external shocks. The absence of walk-forward or rolling-origin validation also limits the assessment of how consistently the models would perform under changing market conditions. These limitations may partly explain why the models showed differences in performance across price-level and directional evaluation. The findings suggest that simpler forecasting approaches should not be disregarded when forecasting rice prices. The lower errors obtained by Naive, Seasonal Naive, and ARIMA indicate that recent and seasonal historical information may already provide useful short-term predictive information for the studied series. At the same time, the limited directional performance indicates that accurate price-level forecasting does not necessarily imply reliable prediction of individual price movements. Therefore, the models are more appropriately used as supplementary information for short-term price monitoring and planning rather than as standalone trading or buy/sell decision systems.

The Diebold–Mariano results indicate statistically significant differences in forecast errors between Hybrid ARIMA–Prophet and Random Forest for both Medium I and Medium II. The significant differences suggest that the two models generated systematically different forecast-error behavior during the evaluation period. The result is consistent with the lower MAE, RMSE, and MAPE obtained by Hybrid ARIMA–Prophet relative to Random Forest, suggesting that the difference favored the hybrid model in this specific comparison. However, the statistical significance does not explain why the hybrid model was better than all other models, nor does it establish universal superiority. In particular, the lower errors obtained by the simpler benchmark models indicate that the advantage of Hybrid ARIMA–Prophet is dependent on which models are included in the comparison.

4. CONCLUSION

This study evaluated rice-price forecasting for Medium I and Medium II Rice in Kudus City using ARIMA, Prophet, Hybrid ARIMA–Prophet, Random Forest, and benchmark models. The results indicate that Hybrid ARIMA–Prophet is suitable for the studied price series

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

and outperformed Random Forest in regression forecasting errors for both rice categories; however, the simpler Naive, Seasonal Naive, and ARIMA-only models achieved lower MAE, RMSE, and MAPE, indicating that the hybrid model is not universally superior. For directional forecasting, Random Forest achieved higher Accuracy than Hybrid ARIMA–Prophet, but its low Precision and F1-score indicate limited reliability for individual price-movement prediction and trading or buy/sell decision support. The Diebold–Mariano test showed statistically significant differences between Hybrid ARIMA–Prophet and Random Forest forecast errors, but this result applies only to the comparison between these two models. The study is limited by its focus on one city and two rice-price categories, the absence of exogenous variables, the lack of walk-forward or rolling-origin validation, and the specific model configurations and benchmarks used. Future research should incorporate exogenous variables, evaluate additional machine-learning and deep-learning approaches such as XGBoost, SVR, and LSTM, conduct multiregion validation, apply walk-forward validation, and address class imbalance in directional forecasting to obtain more robust and generalizable results.

REFERENCES

[1] V. Arinal and M. Azhari, “Penerapan Regresi Linear Untuk Prediksi Harga Beras Di Indonesia,” *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 341–6, Sep. 2023, Accessed: Aug. 07, 2026, doi: 10.55338/saintek.v5i1.1417

[2] N. A. Muzakir and M. Z. Yahya, “Analisis Perbandingan Model Double Exponential Smoothing dan ARIMA untuk Prediksi Harga Beras di Indonesia,” *VARIANSI: Journal of Statistics and Its application on Teaching and Research*, vol. 7, no. 01, pp. 7–20, Apr. 2025, doi: 10.35580/variansiunm349.

[3] N. V. Sijabat and Suharjito, “Unhulled rice prices forecast at farmer level for development of Indonesian sustainable local production,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1488, no. 012114, pp. 1–15, Apr. 2025, doi: 10.1088/1755-1315/1488/1/012114.

[4] L. Diawati and A. S. Rasyid, “Mitigating retail rice price volatility for sustainable supply chains: an optimization and regression-based approach,” *F1000Res.*, vol. 14, no. 311, pp. 1–35, Mar. 2025, doi: 10.12688/f1000research.161723.1.

[5] M. F. Al Anshari, "Forecasting Rice Price Volatility in Indonesia: Integrating ENSO Indicators with Machine Learning Approaches," *International Journal of Sustainable Development and Planning*, vol. 20, no. 11, pp. 4643–4652, Nov. 2025, doi: 10.18280/ijsdp.201109.

[6] Y. D. Sentosa, L. Cahyadi, and F. V. Ferdinand, "Analysis Of Rainfall Effects And Comparison Of Arimax-Garch And Support Vector Regression Methods Based On Rice Price Forecasting In Seven Cities On Java Island," *ICIC Express Letters, Part B: Applications*, vol. 15, no. 12, pp. 1285–1293, Dec. 2024, doi: 10.24507/icicelb.15.12.1285.

[7] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks," *Future Internet*, vol. 15, no. 8, pp. 1–31, Aug. 2023, doi: 10.3390/fi15080255.

[8] H. Mahajan, "Comparative Study of Arima, Lstm and Prophet Models for Time Series Forecasting: A Comprehensive Review," *Preprints (Basel)*, pp. 1–15, Jan. 2026, doi: 10.20944/preprints202601.1377.v1.

[9] R. F. Dianco and M. Novita, "Real time prediction of four main food commodities in Indonesia and the mapping based on autoregressive integrated moving average model," *J. Phys. Conf. Ser.*, vol. 1725, no. 012023, pp. 1–9, Jan. 2021, doi: 10.1088/1742-6596/1725/1/012023.

[10] A. A. Mahira and M. Herlina, "Prediksi Harga Bawang Merah di Provinsi Jawa Barat Menggunakan Model Neural Prophet," *Bandung Conference Series: Statistics*, vol. 5, no. 2, pp. 333–342, Aug. 2025, doi: 10.29313/bcss.v5i2.19681.

[11] A. Nurramadhan, D. Priyanto, and N. Sulistianingsih, "Perbandingan Metode Prophet dan SARIMAX untuk Peramalan Harga Pangan di Lombok Barat," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 4, pp. 9439–9446, Jan. 2026, doi: 10.31004/riggs.v4i4.4964.

[12] A. Amirudin, S. A. Prasetyo, and G. Triyono, "Implementasi Hybrid Arima-Prophet Untuk Prediksi Harga Beras Nasional," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 4, pp. 3837–3850, Dec. 2025, doi: 10.29100/jipi.v10i4.8689.

[13] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, Dec. 2024, doi: 10.58496/BJML/2024/007.

[14] A. Grant, M. Mrazik, and S. Satchell, "Evaluating Forecasts at Multiple Horizons: An Extension of the Diebold–Mariano Approach," *J. Forecast.*, vol. 45, no. 5, pp. 2379–2392, Mar. 2026, doi: 10.1002/for.70150.

[15] *Peraturan Badan Pangan 2 Tahun 2023 tentang Mutu dan Label Beras*. Indonesia, 2023, p. 12. [Online]. Available: www.peraturan.go.id

[16] PIHPS Nasional, "Tabel Harga Berdasarkan Komoditas," PIHPS Nasional. Accessed: May 06, 2026.