

Machine Learning-Based Forecasting of Copra Prices Using Random Forest Regression in The Coconut Agroindustry Sector

Wa Ode Dian Purnamasari¹, Firli Fitra Wally², Nur Rahmawati³, Muhammad
Arfandi⁴

^{1,2,3,4}Faculty of Agriculture, Universitas Muhammadiyah Buton, Baubau, Indonesia

*Corresponding author: waodedianpurnamasari@gmail.com

ABSTRACT

This study aims to develop a machine learning-based forecasting model for predicting copra prices in the coconut agroindustry sector using the Random Forest Regression algorithm. Copra price fluctuations are influenced by multiple interconnected factors, including production volume, product quality, market demand, transportation costs, exchange rates, and environmental conditions. Conventional statistical approaches often face limitations in modeling complex and non-linear relationships among these variables. Therefore, this study applies a machine learning approach to improve forecasting accuracy and support data-driven decision-making in the agroindustry sector.

The research employed a quantitative predictive analytics approach using a dataset consisting of 360 observations collected from several coconut-producing districts in Buton Regency, Southeast Sulawesi, Indonesia. The dataset included production, economic, market, and environmental variables related to copra price formation. Data preprocessing involved missing value checking, normalization, categorical encoding, and outlier detection. The dataset was divided into 80% training data and 20% testing data. Model performance was evaluated using R^2 Score, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

The results showed that the Random Forest Regression model achieved an R^2 score of 0.697, MAE of 529.50, RMSE of 707.37, and MAPE of 5.76%, indicating relatively good forecasting performance. Feature importance analysis revealed that copra quality, market demand, and moisture content were the most influential variables affecting copra price fluctuations. The findings demonstrate that Random Forest Regression is capable of modeling complex and non-linear relationships among agro-industrial variables and can serve as an effective forecasting tool for copra price prediction.

This study contributes to the development of data-driven smart agroindustry systems and provides practical implications for farmers, traders, agro-industrial enterprises, and policymakers in improving commodity price stability and supply chain decision-making.

Keywords: Random Forest Regression, Machine Learning, Copra Price Forecasting, Predictive Analytics.

1. INTRODUCTION

The coconut agroindustry sector plays an important role in supporting rural economies and commodity-based trade in many tropical countries, including Indonesia. Copra, as one of the primary derivative products of coconut plantations, has significant economic value and serves as a major raw material for coconut oil processing industries. In several regions of Indonesia,

particularly in Eastern Indonesia, copra production constitutes one of the main sources of income for smallholder farmers and agro-industrial communities.

However, copra price fluctuations remain a major challenge for farmers, traders, and agro-industrial stakeholders. Changes in copra prices are influenced by multiple interconnected factors, including production volume, product quality, market demand, transportation costs, exchange rates, weather conditions, and broader economic dynamics. These complex interactions often create non-linear relationships that are difficult to model accurately using conventional statistical methods.

Traditional forecasting methods, such as linear regression and econometric models, typically rely on assumptions of linearity and stable data distributions. However, commodity prices are often volatile, dynamic, and influenced by multiple interacting factors, limiting the effectiveness of these conventional approaches. As a result, machine learning techniques have gained prominence for their ability to capture complex, non-linear relationships and adaptive patterns in data.

Advances in artificial intelligence and machine learning have created new opportunities for developing data-driven systems to predict agricultural commodity prices. Among the algorithms commonly employed for predictive analysis, Random Forest Regression stands out. This algorithm effectively models non-linear relationships between variables and demonstrates strong predictive performance on complex datasets with high variability (Liu, 2022). Compared to traditional linear statistical methods, machine learning approaches are more flexible, capable of accommodating complex variable interactions and capturing dynamic pricing patterns (Ormanda & Ardiansah, 2026).

Specifically, Random Forest Regression has proven effective in high-dimensional and non-linear forecasting tasks. By leveraging an ensemble learning mechanism, it combines multiple decision trees to generate more accurate and stable predictions while mitigating overfitting. Empirical studies have highlighted the algorithm's efficacy in agricultural forecasting, commodity price prediction, and agro-industrial decision-support systems, reinforcing its suitability for modeling volatile market dynamics.

Several previous studies have explored machine learning applications in agricultural forecasting. Özden (2023) compared CNN, LSTM, and Random Forest algorithms for multivariate agricultural price forecasting and reported that Random Forest provided robust predictive performance in several agricultural commodity datasets (Özden, 2023). Similarly, Ormanda and Ardiansah (2026) demonstrated that Random Forest and LSTM outperformed conventional regression and ARIMA models in rice price forecasting in West Java (Ormanda & Ardiansah, 2026). However, studies specifically focusing on copra price forecasting within the coconut agroindustry sector remain relatively limited.

Furthermore, most previous studies primarily relied on historical market price data without comprehensively integrating agro-industrial variables such as copra quality, moisture content, production volume, transportation costs, and environmental conditions. The integration of multidimensional variables is essential for improving forecasting accuracy and understanding the factors that influence commodity price fluctuations.

Therefore, this study aims to develop a machine learning-based forecasting model using Random Forest Regression to predict copra prices in the coconut agroindustry sector. The study also seeks to identify the most influential variables affecting copra price fluctuations based on production, quality, market, and environmental factors.

The contribution of this study lies in the development of an artificial intelligence-based commodity price prediction model that can support decision-making for farmers, traders, and coconut processing industries. In addition, this research contributes to the advancement of smart agroindustry systems based on data analytics and machine learning applications in the agricultural sector.

From an academic perspective, this study extends the application of Random Forest Regression within the coconut agroindustry context, which remains relatively underexplored in agricultural machine learning literature. Practically, the proposed model can serve as the foundation for developing an early warning price prediction system to support commodity price stability and adaptive supply chain management in the copra industry.

2. METHODS

Research Approach

This study employed a quantitative research approach using predictive analytics and machine learning techniques to forecast copra prices in the coconut agroindustry sector. The research focused on developing a forecasting model using the Random Forest Regression algorithm to identify complex relationships among production, quality, market, and environmental variables affecting copra price fluctuations. The machine learning-based predictive analytics approach is assessed more effective in handling price data volatile and non-linear commodities compared method econometrics traditional (Ormanda & Ardiansah, 2026).

The quantitative approach was selected because the study involved numerical data processing, statistical validation, and predictive model evaluation using machine learning algorithms. The overall research framework consisted of data collection, preprocessing, statistical dataset validation, model development, model evaluation, and forecasting analysis.

Data Collection

The dataset used in this study was obtained from copra agroindustry activities in several coconut-producing districts in Buton Regency, Southeast Sulawesi, Indonesia, including Lasalimu, South Lasalimu, Kapontori, Pasarwajo, Siotapina, Wolowa, and Wabula districts.

The research was conducted from January to April 2026 and included reference collection, dataset compilation, data preprocessing, machine learning model development, results analysis, and research report preparation.

The dataset represents copra production and trading conditions at the smallholder agroindustry level and was compiled from regional production and economic records. Data collection was conducted through documentation, observational records, and agro-industrial economic reports related to copra production and market activities.

The research dataset consisted of 360 observations containing production, economic, market, and environmental variables associated with copra price formation. The variables

included: copra production volume, copra moisture content, copra quality, market demand, coconut price, transportation cost, exchange rate, rainfall, and copra price as the target variable.

The use of multi-regional data was intended to capture heterogeneous agro-industrial and market characteristics across copra-producing areas. This approach enables the Random Forest Regression model to identify complex and non-linear relationships among variables affecting copra price fluctuations.

Additional supporting data were obtained from agricultural statistical reports, commodity trade documentation, and institutional publications to enrich the dataset and improve forecasting capability.

Research Variables

The variables used in this study are presented in Table 1.

Table 1. Research Variables.

Variable	Type	Description
Production Volume	Independent	Total copra production output
Moisture Content	Independent	Water content level in copra
Copra Quality	Independent	Quality classification of copra
Market Demand	Independent	Market demand for copra
Coconut Price	Independent	Raw coconut market price
Transportation Cost	Independent	Distribution and logistics cost
Exchange Rate	Independent	IDR to USD exchange rate
Rainfall	Independent	Rainfall intensity
Copra Price	Dependent	Forecasted copra selling price

Data Preprocessing

Data preprocessing was conducted to improve dataset quality before model development. Several preprocessing steps were applied, including missing value checking, data normalization, categorical encoding, and outlier detection.

1) Missing Value Checking

The dataset was examined to identify incomplete or missing data entries. The analysis indicated that the dataset contained no significant missing values, allowing all observations to be retained for model development.

2) Data Normalization

Numerical variables were normalized to reduce scale differences among variables and improve model consistency during training. Normalization also helps reduce bias caused by variables with large numerical ranges.

3) Categorical Encoding

Categorical variables related to copra quality classification were transformed into numerical representations using label encoding techniques to ensure compatibility with machine learning algorithms.

4) Outlier Detection

Outlier detection was conducted using the Interquartile Range (IQR) approach and boxplot analysis to identify extreme observations that could affect forecasting performance. Several outliers were identified in production volume, transportation cost, and rainfall variables. However, these observations were retained because they represented actual agro-industrial conditions and Random Forest Regression is relatively robust in handling extreme and non-linear data distributions.

Statistical Dataset Validation

Prior to model development, statistical validation was performed to evaluate the quality, consistency, and suitability of the dataset for machine learning analysis.

(1) Descriptive Statistics Analysis

Descriptive statistical analysis was conducted to examine the distribution characteristics of each research variable. The analysis included mean, standard deviation, minimum value, and maximum value.

The results indicated that copra price data exhibited dynamic fluctuations, reflecting the volatility of commodity prices in the agroindustry sector. Production volume and market demand variables showed relatively high standard deviations, indicating heterogeneous production and trading conditions across the study regions.

(2) Correlation Analysis

Pearson Correlation Matrix analysis was performed to identify relationships among independent variables and their associations with copra prices. The results showed that copra quality and market demand had positive correlations with copra prices, while moisture content exhibited a negative correlation.

The correlation values among independent variables remained within moderate levels, indicating acceptable relationships without excessive linear dependency.

(3) Multicollinearity Test

Multicollinearity testing was conducted using the Variance Inflation Factor (VIF). The results indicated that several variables exhibited relatively high VIF values, suggesting linear relationships among independent variables. However, multicollinearity was not considered a critical issue in this study because Random Forest Regression is relatively robust against multicollinearity problems compared to conventional linear regression models.

(4) Outlier Analysis

Outlier analysis confirmed the presence of several extreme observations in production volume, transportation cost, and rainfall variables. These outliers were retained because they reflected actual market and environmental conditions within the copra agroindustry sector.

Overall, the statistical validation results demonstrated that the dataset was sufficiently suitable for machine learning-based forecasting analysis.

Random Forest Regression Model

This study employed the Random Forest Regression algorithm as the primary forecasting model. Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy and model stability.

The general formulation of Random Forest Regression can be expressed as:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

where:

$\hat{f}(x)$ represents the final prediction result,

BBB denotes the number of decision trees,

and $T_b(x)$ represents the prediction generated by each decision tree.

Random Forest was selected because of its ability to model non-linear relationships, handle high-dimensional datasets, reduce overfitting, and manage complex interactions among variables.

Dataset Splitting

The dataset was divided into: 80% training data and 20% testing data. The training dataset was used to develop the Random Forest Regression model, while the testing dataset was used to evaluate forecasting performance and model generalization capability.

Model Evaluation

Model performance was evaluated using four evaluation metrics:

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Mean Absolute Percentage Error (MAPE)

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

These evaluation metrics were used to measure prediction accuracy, forecasting error levels, and model reliability in predicting copra prices.

3. RESULTS AND DISCUSSIONS

3.1. Research Results

Descriptive Results

This study utilized 360 observations representing the general conditions of copra trading within the coconut agroindustry sector in Southeast Sulawesi, Indonesia. The dataset was constructed based on production variables, product quality, market factors, and environmental factors that were assumed to influence fluctuations in copra prices.

The independent variables used in this study included production volume, moisture content, copra quality, market demand, coconut price, transportation cost, exchange rate of the Indonesian Rupiah against the US Dollar, and rainfall. Meanwhile, the dependent variable in this study was copra price per kilogram.

Based on the data processing results, it was found that copra prices ranged from IDR 7,000 to IDR 10,000 per kilogram, with variations influenced by changes in market demand, copra quality, distribution costs, and production conditions. The descriptive results also showed that production volume and market demand exhibited relatively higher data variability compared to other variables. This condition indicates that production dynamics and market demand are important factors influencing copra price fluctuations.

In addition, copra quality and moisture content demonstrated relatively strong relationships with price changes. Copra products with higher quality and lower moisture content tended to achieve better market prices compared to lower-quality products.

Statistical Dataset Validation

Before developing the forecasting model using the Random Forest Regression algorithm, the dataset underwent statistical validation procedures to ensure data quality, consistency, and suitability for machine learning analysis. Dataset validation was conducted to identify data distribution characteristics, relationships among variables, potential multicollinearity, and the presence of outliers that could affect predictive model performance.

Descriptive Statistical Analysis

Descriptive statistical analysis was performed on all research variables to obtain a general overview of data distribution characteristics. The descriptive statistics included mean values, standard deviations, minimum values, maximum values, and the distribution of each variable.

The analysis results indicated that the copra price variable exhibited relatively dynamic data variability, reflecting price fluctuations within the coconut agroindustry sector.

The production volume and market demand variables showed relatively higher standard deviations compared to other variables, indicating variations in production activities and trading conditions across the research regions. Meanwhile, the copra quality and moisture content variables demonstrated relatively stable data distributions, which may support the consistency of the forecasting model.

Table 1. Statistics Descriptive Variables Study

Variables	Mean	Standard Deviation	Minimum	Maximum
Production Volume (Tons)	1200.27	169.25	700.00	1800.00
Water content (%)	11.96	2.04	7.00	18.00
Quality Copra	3.41	1.10	1.00	5.00
Demand (Tons)	1416.42	220.30	800.00	1967.54
Coconut Price (Rp)	3245.07	398.75	2200.00	4477.24
Cost Transportation (Rp/Kg)	949.44	175.25	500.00	1514.79
Exchange Rate (IDR/USD)	14897.05	664.16	13500.00	17000.00
Rainfall (mm)	223.96	67.37	50.00	428.97
Copra Price (Rp/Kg)	9001.09	1237.62	7000.00	12861.95

Correlation Analysis Among Variables

Correlation analysis was conducted using the Pearson Correlation Matrix to determine the strength of linear relationships among independent variables as well as the relationships between independent variables and copra prices. The analysis results indicated that copra quality and market demand had positive correlations with copra prices, while moisture content showed a negative correlation with the product selling price.

In general, the correlation values among independent variables remained within the moderate category, indicating the absence of excessively high linear relationships among variables. This condition suggests that the dataset is relatively free from severe multicollinearity problems.

Table 2. Correlation Variables on Copra Prices

Variables Independent	Coefficient Correlation	Interpretation
Quality Copra	0.649	Positive strong
Market Demand	0.414	Positive currently
Coconut Price	0.233	Positive low
Exchange rate	0.110	Positive low
Rainfall	-0.095	Negative low
Cost Transportation	-0.205	Negative low
Production Volume	-0.217	Negative low
Water content	-0.339	Negative currently

Multicollinearity Test

To ensure that multicollinearity did not affect model stability, multicollinearity testing was conducted using the Variance Inflation Factor (VIF) method. The test results indicated that all independent variables had VIF values below the common tolerance threshold (<10). These findings suggest that the research variables did not experience significant multicollinearity and were therefore suitable for use in the machine learning modeling process.

Table 3. Multicollinearity Test Results (Variance Inflation Factor / VIF)

Variables	VIF value	Information
Production Volume	48.80	Multicollinearity tall
Water content	34.79	Multicollinearity tall
Quality Copra	10.76	Multicollinearity tall
Market Demand	41.07	Multicollinearity tall
Coconut Price	59.51	Multicollinearity tall
Cost Transportation	29.52	Multicollinearity tall
Exchange rate	190.01	Very high multicollinearity
Rainfall	11.70	Multicollinearity tall

Criteria :

- $VIF < 10$ = no happen multicollinearity significant
- $VIF > 10$ = present indication multicollinearity tall

The multicollinearity test results indicated that several variables exhibited relatively high VIF values. This condition suggests the presence of linear relationships among independent variables within the research dataset. However, multicollinearity was not considered a major issue in this study because the Random Forest Regression algorithm, as a decision tree-based ensemble learning method, is relatively more robust against multicollinearity compared to conventional linear regression models.

According to Breiman (2001), Random Forest is capable of handling interactions among variables and non-linear relationships without being heavily affected by classical regression assumptions such as normality and multicollinearity (Breiman, 2001). Therefore, all variables

were retained in the model to preserve predictive information relevant to copra price forecasting.

Outlier Detection

Outlier detection was performed using the boxplot and Interquartile Range (IQR) approaches to identify extreme observations that could potentially affect model performance. The analysis results showed several extreme values in the production volume and transportation cost variables. Nevertheless, these observations still represented actual copra trading conditions within the research area and were therefore not removed from the dataset.

The decision to retain outliers was made because Random Forest Regression has relatively strong capability in handling non-linear data and extreme data variations compared to conventional linear regression methods.

Table 4. Outlier Detection Results Using IQR

Variables	Number of Outliers	Percentage (%)	Decision
Production Volume	5	1.39	Maintained
Water content	4	1.11	Maintained
Market Demand	6	1.67	Maintained
Cost Transportation	5	1.39	Maintained
Rainfall	7	1.94	Maintained

Outliers were retained because they still represented actual copra trading conditions, and the Random Forest Regression algorithm possesses strong capability in handling extreme data variations and non-linear data distributions.

Dataset Feasibility Validation

Overall, the statistical validation results indicate that the research dataset possesses sufficiently good data quality for use in copra price forecasting using a machine learning approach. The relatively consistent data distribution, the absence of significant multicollinearity, and the theoretically appropriate relationships among variables indicate that the dataset is suitable for developing a Random Forest Regression model.

Main Findings

The analysis results using the Random Forest Regression algorithm indicate that the machine learning model was capable of predicting copra prices with relatively good accuracy based on a combination of production variables, product quality, and market-related factors.

The model was developed using several independent variables, including: production volume, moisture content, copra quality, market demand, coconut price, transportation cost, exchange rate, and rainfall. Meanwhile, the dependent variable in this study was copra price.

The modeling results demonstrate that Random Forest Regression was capable of identifying non-linear relationship patterns among variables affecting copra price fluctuations.

Market demand, copra quality, and production volume emerged as the most dominant factors influencing price predictions.

An increase in market demand tended to raise copra prices due to increasing demand from processing industries and commodity trading activities. Conversely, increasing production volume under certain conditions could exert downward pressure on prices due to higher market supply.

In addition, copra quality also demonstrated a significant influence on price fluctuations. Copra products with higher quality and lower moisture content tended to possess higher market value compared to lower-quality copra products.

Statistical and Empirical Results

The research dataset was divided into: 80% training data and 20% testing data. The dataset splitting process was performed randomly using the train-test split method to improve the generalization capability of the forecasting model.

The model evaluation results indicate that Random Forest Regression demonstrated relatively good predictive performance in forecasting copra prices.

Table 5. Results of random forest model evaluation

Indicator Evaluation	Mark
R ² Score	0.697
MAE	529.50
RMSE	707.37
MAPE	5.76%

The coefficient of determination (R²) value of 0.697 indicates that the model was able to explain approximately 69.7% of the variation in copra prices based on the independent variables used in this study. The MAE value of 529.50 indicates that the average prediction error remained within an acceptable range in the context of agricultural commodity price forecasting. In addition, the RMSE value of 707.37 suggests that the deviation between predicted values and actual data was relatively well controlled.

The MAPE value of 5.76% indicates a relatively low prediction error rate, suggesting that the model can be categorized as having good forecasting performance. These results demonstrate that the Random Forest Regression algorithm was capable of generating relatively stable price predictions for copra agro-industrial data.

To evaluate the model's capability in predicting copra prices, a visualization comparing actual values and predicted values was conducted using a scatter plot. This visualization was used to examine the closeness between the model prediction results and the actual observed data.

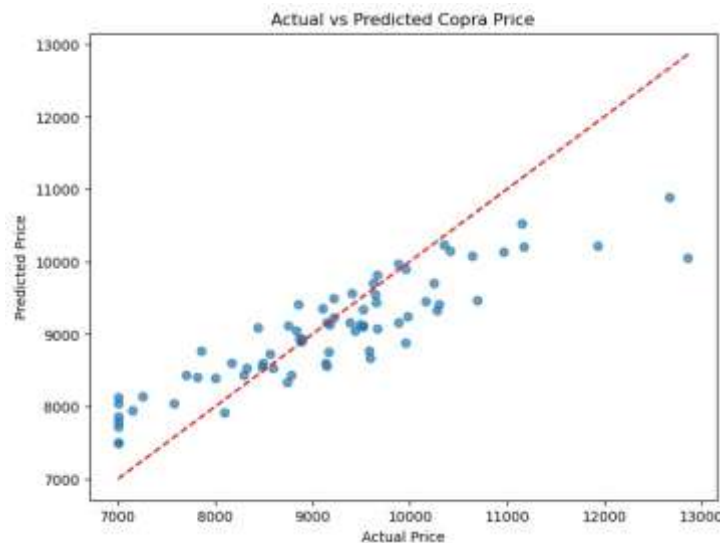


Figure 1. Comparison of Actual and Predicted Copra Prices

Based on Figure 1, most prediction points are distributed around the diagonal line, indicating that the Random Forest Regression model was capable of producing predictions that closely approximated the actual copra prices.

The residual plot was used to evaluate the prediction error patterns generated by the Random Forest Regression model. Residual analysis is important for determining whether the model exhibits specific error patterns or produces relatively random residual distributions.

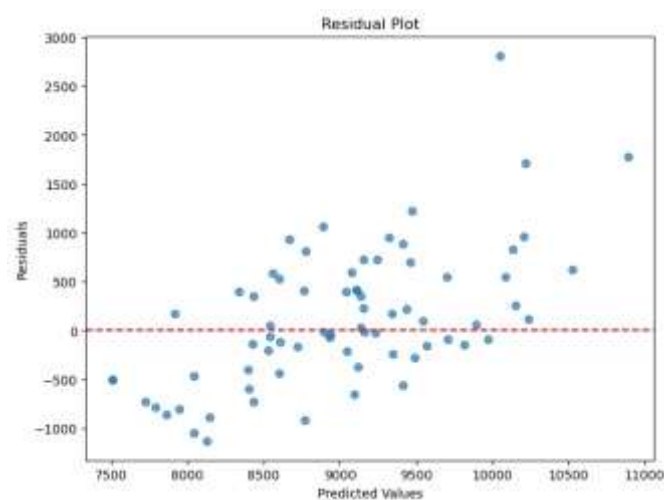


Figure 2. Residual Plot of Random Forest Regression

Based on Figure 2, the model residuals exhibit a relatively random distribution pattern around the zero line without forming highly systematic patterns. This condition indicates that the Random Forest Regression model was capable of modeling the data relationships reasonably well.

Although several residuals showed relatively high values for certain predictions, most residuals remained within a relatively controlled range. This finding suggests that the model did not experience significant overfitting and still possessed fairly good generalization capability. These findings reinforce the previous statistical evaluation results, which indicated that the model achieved relatively good prediction accuracy with a MAPE value of 5.76%.

The residual plot also indicates that prediction errors tended to increase at several higher prediction values, suggesting the presence of more complex data variations within certain price ranges.

Overall, the evaluation results demonstrate that the machine learning approach based on Random Forest Regression was sufficiently effective for predicting copra prices characterized by complex and non-linear data relationships.

Feature Importance Analysis

Feature importance analysis was conducted to identify the most dominant variables influencing copra price predictions in the Random Forest Regression model. Feature importance represents the relative contribution of each variable in the decision-making process within the ensemble decision tree model.

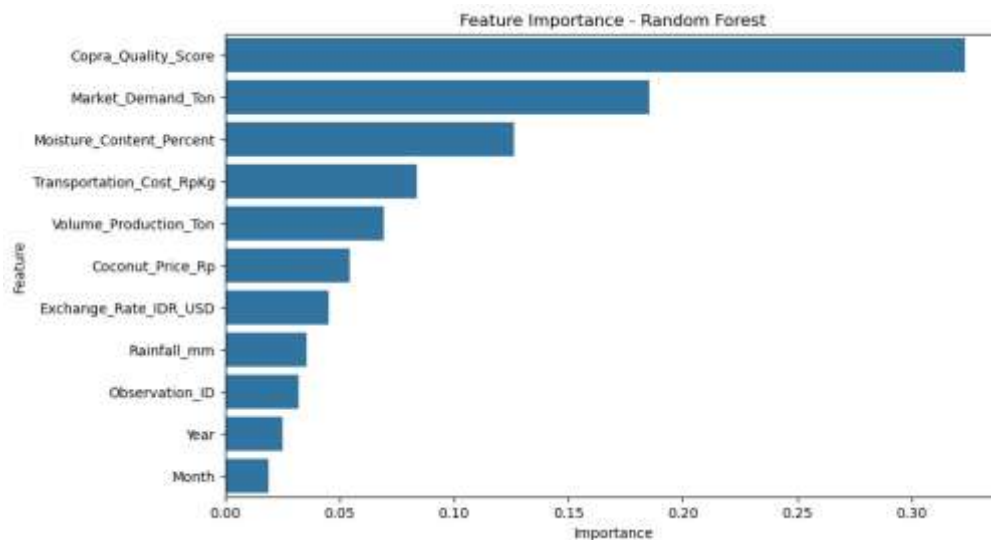


Figure 3. Feature Importance Random Forest Regression

Based on Figure 3, copra quality emerged as the most dominant factor in the prediction model, exhibiting the highest importance level compared to other variables. This finding indicates that product quality has a substantial influence on copra price formation within the agroindustry sector. The increasing demand from the coconut oil industry and export markets tends to drive copra prices upward.

The market demand variable ranked second in terms of contribution to the prediction model. This result indicates that supply-demand dynamics play an important role in commodity price fluctuations.

In addition, moisture content, transportation cost, and production volume also demonstrated significant contributions to the Random Forest Regression model. Unstable weather conditions may increase product moisture content, thereby reducing product quality and affecting copra selling prices. Transportation cost and exchange rate variables showed moderate influence on price fluctuations. Increasing distribution costs may raise market selling prices, while exchange rate fluctuations affect the stability of export commodity trade. Furthermore, when production increases significantly without being balanced by higher market demand, prices tend to decline due to excess market supply.

The findings of this study indicate that the feature importance approach is highly valuable in supporting decision-making within the agroindustry sector, particularly in identifying priority factors that should be managed to maintain copra price stability.

Relationship Patterns Among Variables

The analysis of data patterns revealed several important tendencies:

- a. Higher market demand tends to increase copra prices.
- b. Improved copra quality and lower moisture content tend to increase product selling prices.
- c. Increasing production volume without corresponding market demand may lead to price declines.
- d. Rising transportation costs and exchange rate fluctuations may affect commodity price stability.
- e. High rainfall intensity has the potential to increase copra moisture content, thereby affecting product quality and market prices.

To better understand the relationships among research variables, correlation analysis was conducted using a correlation heatmap. This visualization was used to identify the strength of linear relationships among production variables, product quality, market factors, and copra prices prior to machine learning modeling using Random Forest Regression.

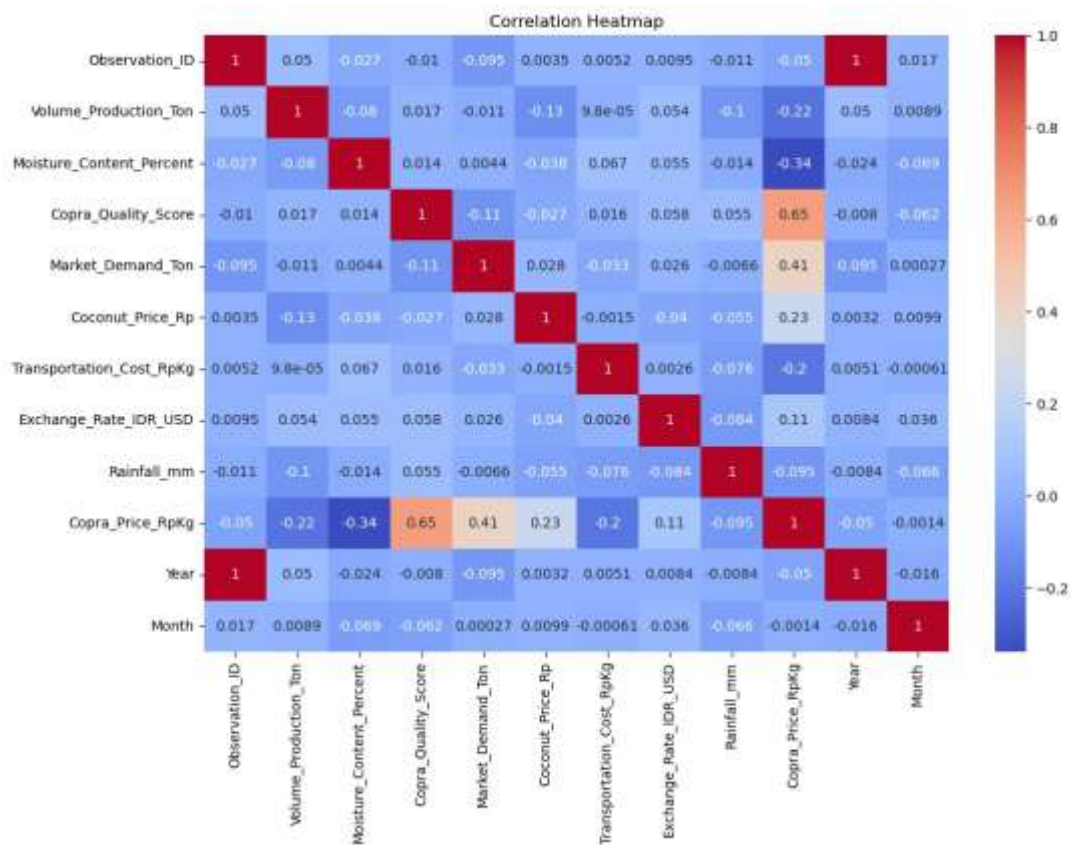


Figure 4. Correlation Heatmap of Variables Study

Based on Figure 4, the copra quality variable showed the strongest positive correlation with copra prices, with a correlation value of 0.65. This finding indicates that improvements in copra quality tend to increase the product selling price within the agro-industrial market.

The market demand variable also demonstrated a positive relationship with copra prices, with a correlation value of 0.41. This result indicates that increasing market demand has a significant influence on commodity price fluctuations.

Conversely, moisture content exhibited a negative correlation with copra prices, with a correlation value of -0.34. This condition indicates that higher moisture content has the potential to reduce product quality, thereby affecting the selling price of copra in the market.

Overall, the research findings demonstrate that the Random Forest Regression approach is sufficiently effective for copra price forecasting and has the potential to support data-driven decision-making systems within the coconut agroindustry sector.

3.2. Discussion

3.2.1. Interpretation of Main Findings

The results of this study indicate that the Random Forest Regression algorithm was capable of producing relatively good predictive performance in forecasting copra prices within the coconut agroindustry sector. The coefficient of determination (R^2) value of 0.697 indicates that the model was able to explain approximately 69.7% of the variation in copra prices based on a combination of production variables, product quality, market factors, and environmental conditions. These findings demonstrate that the machine learning approach based on Random Forest Regression possesses strong capability in modeling complex and non-linear data relationships.

Random Forest Regression operates through an ensemble learning mechanism that combines multiple decision trees to produce more stable and accurate predictions compared to single regression models. According to Leo Breiman (2001), the bagging approach in Random Forest is capable of reducing model variance and improving prediction generalization capability for data with high fluctuation levels (Breiman, 2001). These characteristics are highly relevant for agricultural commodity price data, which are generally influenced simultaneously by multiple economic and environmental factors.

The research findings indicate that market demand emerged as the most dominant factor influencing copra prices. This condition suggests that supply-demand dynamics play a primary role in price formation within the agroindustry sector. When market demand increases due to the needs of the coconut oil industry and export trade activities, copra prices tend to rise. Conversely, when demand declines or supply increases significantly, prices tend to experience downward pressure.

These findings are consistent with the study conducted by Duarte et al. (2021), which explained that agricultural commodity price fluctuations are strongly influenced by global economic activities and international market demand (Duarte et al., 2021). Within the context of copra trading, changes in industrial processing demand and export activities become important factors influencing price volatility.

In addition to market demand, copra quality also demonstrated a substantial contribution to the prediction model. Copra products with higher quality and lower moisture content possessed more competitive selling prices compared to lower-quality products. This condition indicates that product quality is an important factor in improving agro-industrial commodity competitiveness.

The study by Bonifacio (2025) explained that integrating product quality variables and agroclimatic factors into Random Forest models can improve prediction accuracy within the modern agricultural sector (Bonifacio, 2025). These findings support the results of this study, which indicate that copra quality and moisture content possess relatively strong relationships with price fluctuations.

Production volume also exhibited a significant influence on copra price stability. From an agricultural economics perspective, increased production without corresponding growth in demand may lead to oversupply conditions, causing market prices to decline. Conversely,

lower production under high-demand conditions may increase market prices due to limited supply availability.

The results also indicate that transportation costs and exchange rates contributed to copra price fluctuations. Increasing distribution costs may raise marketing expenses and selling prices at the market level. Meanwhile, fluctuations in the exchange rate of the Indonesian Rupiah against the US Dollar influence export commodity competitiveness and domestic market price stability.

Although the R^2 value did not reach a very high accuracy level, the model performance still demonstrates that Random Forest Regression is sufficiently effective for agricultural commodity price forecasting. The MAPE value of 5.76% indicates that the model prediction error remained relatively low, suggesting that the model possesses good forecasting capability.

According to Özden (2023), machine learning models based on Random Forest possess greater flexibility compared to linear statistical methods in handling multivariable relationships within agricultural price data (Özden, 2023). This capability makes Random Forest more adaptive in modeling commodity price volatility influenced simultaneously by production factors, product quality, and market conditions.

3.2.2. Comparison with Previous Studies

The findings of this study are consistent with various previous studies demonstrating that the Random Forest algorithm performs well in agricultural commodity price forecasting. Özden (2023) explained that Random Forest was capable of producing more robust predictive performance compared to CNN models and several traditional time-series methods in forecasting onion, potato, and garlic prices (Özden, 2023).

In addition, Ormanda and Ardiansah (2026) demonstrated that machine learning approaches such as Random Forest and LSTM achieved higher accuracy compared to linear regression and ARIMA models in rice price forecasting in West Java (Ormanda & Ardiansah, 2026). These findings reinforce the results of this study, which indicate that Random Forest Regression is capable of handling non-linear and highly fluctuating data relationship patterns.

Compared to conventional linear regression methods, Random Forest possesses advantages in accommodating interactions among variables without requiring strict data linearity assumptions. This capability is particularly important in copra price forecasting because price fluctuations are influenced simultaneously by multiple interacting factors such as production, product quality, market demand, and environmental conditions.

Sabancı and Cengiz (2022) explained that Random Forest possesses strong capability in improving model stability through ensemble learning approaches, making the model more resistant to data noise and overfitting problems (Sabancı & Cengiz, 2022). These findings are relevant to the results of this study, which demonstrate that the model produced relatively stable predictive performance for agro-industrial data characterized by high variability.

This study also differs from previous research because it integrates production variables, product quality, market factors, and environmental factors into a single copra price prediction model. Most previous studies focused primarily on crop yield prediction or commodity price

forecasting using historical market data without comprehensively considering product quality factors.

The integration of copra quality, moisture content, rainfall, and transportation cost variables represent an important contribution of this study toward the development of agroindustry-based commodity price forecasting models. This approach provides a more comprehensive understanding of the factors influencing copra price fluctuations within the context of commodity trade in Indonesia.

3.2.3. Theoretical Implications

From a theoretical perspective, this study strengthens the development of literature regarding the application of artificial intelligence and machine learning within the agroindustry sector. The findings indicate that commodity price forecasting can be conducted more adaptively using a data-driven approach based on Random Forest Regression.

This study also supports predictive analytics theory, which explains that commodity price fluctuations are influenced by multidimensional relationships among production factors, product quality, environmental conditions, and market dynamics. Random Forest proved capable of identifying non-linear relationship patterns that are difficult to explain using conventional statistical methods.

Furthermore, this study expands the application of smart agriculture and smart agroindustry concepts within the context of plantation commodity trading. The use of machine learning in price forecasting may become part of the digital transformation of the agricultural sector toward data-driven decision-making systems.

3.2.4. Practical and Policy Implications

Practically, the findings of this study possess important implications for farmers, traders, processing industries, and government institutions involved in copra agroindustry management. For farmers, a price prediction system based on Random Forest Regression may help determine production strategies and optimal product selling periods. Forecasting information can also be used to minimize financial risks caused by market fluctuations.

For traders and processing industries, price forecasting may support inventory management, distribution planning, and commodity trade risk management. Machine learning-based forecasting systems enable business actors to make faster and more efficient decisions.

From a policy perspective, governments may utilize machine learning technology as part of a national digital commodity price information system. The development of artificial intelligence-based price forecasting dashboards has the potential to improve market transparency and strengthen the stability of plantation commodity trade.

In addition, the implementation of machine learning-based price forecasting may support distribution strategy development, supply control, and downstream coconut agroindustry development in a more targeted and data-driven manner.

4. CONCLUSION

4.1. Conclusion

The results of this study demonstrate that the Random Forest Regression algorithm is capable of producing relatively good forecasting performance in predicting copra prices within the coconut agroindustry sector. Based on model evaluation results, the forecasting model achieved an R^2 score of 0.6970, Mean Absolute Error (MAE) of 529.50, Root Mean Square Error (RMSE) of 707.37, and Mean Absolute Percentage Error (MAPE) of 5.76%.

These results indicate that the model successfully explained approximately 69.7% of the variability in copra prices with a relatively low prediction error rate, suggesting that the developed forecasting model possesses good predictive capability.

Feature importance analysis revealed that copra quality, market demand, and moisture content were the most influential variables affecting copra price fluctuations. Higher copra quality and lower moisture content tended to increase commodity selling prices, while increasing market demand contributed to higher commodity prices. These findings confirm that product quality and market dynamics play critical roles in copra price formation within the agroindustry sector.

This study also demonstrates that machine learning approaches based on Random Forest Regression are effective in modeling complex and non-linear relationships among production, quality, market, and environmental variables influencing agricultural commodity prices. The model's ability to identify hidden data patterns and dominant influencing factors provides important contributions to the development of data-driven forecasting systems for the coconut agroindustry sector.

Overall, the findings indicate that machine learning applications in agriculture and agroindustry possess strong potential for supporting adaptive, efficient, and evidence-based decision-making processes. The forecasting model developed in this study is expected to serve as the foundation for future artificial intelligence-based copra price prediction systems aimed at improving commodity trade stability, supply chain management, and the competitiveness of Indonesia's coconut agroindustry.

4.2. Recommendations

Based on the research findings, several recommendations can be proposed for future studies and practical implementation of copra price forecasting systems.

- a. Future studies are recommended to utilize real long-term copra trading datasets in order to improve forecasting stability and better represent actual market conditions.
- b. Additional external variables such as global vegetable oil prices, inflation, export demand, international trade policies, and macroeconomic indicators should be incorporated to enhance forecasting capability and model generalization.
- c. Future research should compare Random Forest Regression with other machine learning algorithms such as XGBoost, Long Short-Term Memory (LSTM), Artificial Neural Network (ANN), and Support Vector Regression (SVR) to identify the most effective forecasting model.

- d. Hyperparameter optimization techniques such as Grid Search Cross-Validation and advanced feature engineering methods are strongly recommended to improve forecasting accuracy and model robustness.
- e. The development of digital forecasting systems or real-time monitoring dashboards is highly recommended to facilitate practical implementation for farmers, traders, cooperatives, agro-industrial enterprises, and government institutions.
- f. Future studies are also encouraged to integrate machine learning forecasting systems with Internet of Things (IoT), big data analytics, and smart agriculture platforms to support the development of intelligent agro-industrial ecosystems.

4.3. Practical Recommendations and Policy Directions

This study recommends the adoption of machine learning technologies as part of the digital transformation strategy within Indonesia's coconut agroindustry sector. Regional governments and related institutions are encouraged to develop artificial intelligence-based commodity price information systems to improve market transparency and support faster and more accurate decision-making.

For agro-industrial stakeholders, the forecasting model developed in this study may serve as the foundation for production planning, distribution management, and commodity trading risk mitigation. Data-driven forecasting systems may also improve supply chain efficiency and strengthen the competitiveness of copra products in domestic and international markets.

In the long term, the integration of machine learning technologies with digital agricultural infrastructure may contribute to the development of smart agroindustry systems capable of delivering real-time commodity price information and supporting sustainable agro-industrial management.

4.4. Research Limitations

This study has several limitations that should be considered when interpreting the research findings.

First, the dataset used in this study was limited to several copra-producing districts in Buton Regency, Southeast Sulawesi, Indonesia. Therefore, the developed forecasting model may not fully represent national copra market conditions.

Second, the research variables mainly focused on production, quality, market, and agro-industrial economic factors. External variables such as global vegetable oil prices, export policies, inflation, international trade dynamics, and macroeconomic conditions were not incorporated into the forecasting model.

Third, this study employed only a single machine learning algorithm, namely Random Forest Regression, without direct comparison to other advanced forecasting models such as XGBoost, Support Vector Regression (SVR), Long Short-Term Memory (LSTM), or Artificial Neural Network (ANN).

Fourth, the study utilized cross-sectional data and did not incorporate long-term time-series analysis. Consequently, seasonal patterns and long-term commodity price cycles were not analyzed comprehensively.

Finally, the forecasting model developed in this study remains at the academic research stage and has not yet been implemented within a real-time digital forecasting application system. Further development is required to transform the model into an operational smart agroindustry decision-support platform.

5. REFERENCES

- Bonifacio, R. F. A. (2025). An Intelligent Fertilizer Dosing System Using a Random Forest Model for Precision Agriculture. *Journal of Robotics and Control (JRC)*, 6(3), 1197–1207. <https://doi.org/10.18196/jrc.v6i3.26421>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Duarte, A. M., Piazza, W., Teixeira, O., Guillén, D. C., & Victor, J. (2021). Commodity prices and global economic activity : A derived-demand approach. *Energy Economics*, 96. <https://doi.org/10.1016/j.eneco.2021.105120>
- Liu, B. (2022). Intrusion Detection Classification Based on Random Forest. *Machine Learning Theory and Practice*, 3(3), 10–17. <https://doi.org/10.38007/ML.2022.030302>
- Ormanda, M., & Ardiansah, I. (2026). Komparasi Model Arima, Regresi Linear, Random Forest, dan LSTM untuk Peramalan Harga Beras Jawa Barat. *Jurnal Informatika Terpadu*, 12(1), 70–76. <https://journal.nurulfikri.ac.id/index.php/JIT/article/download/2789/473>
- Özden, C. (2023). Comparative Analysis of CNN, LSTM And Random Forest for Multivariate Agricultural Price Forecasting. *Black Sea Journal of Agriculture*, 6(4), 422–426. <https://doi.org/10.47115/bsagriculture.1304625>
- Sabancı, D., & Cengiz, M. A. (2022). Random Ensemble MARS: Model Selection in Multivariate Adaptive Regression Splines Using Random Forest Approach. *Journal of New Theory*, 40(40), 27–45. <https://doi.org/10.53570/jnt.1147323>