Journal Global Tecnology Computer
Vol 5, No. 3, Agustus 2026, Hal 127-135
ISSN: 2809-6118 (Online - Elektronik)
DOI: 10.47065/jogtc.v5i3.11033
Websitehttps://ejurnal.seminar-id.com/index.php/jogtc

Optimizing Agricultural Commodity Price Forecasts Using an Ensemble Stacking Method Based on Market and Product Characteristics

Abdul Karim[1]*, Kusmanto[2]

[1] Faculty of Science and Technology, Department of Information Technology, Universitas Labuhanbatu, Rantau Prapat, Indonesia
Jl. SM. Raja Aek Tapa No.126 A KM 3.5, Bakaran Batu, Kec. Rantau Selatan, Kab. Labuhanbatu, Sumatera Utara
[2] Faculty of Computer Science, Computer Engineering Program, Universitas Al Washliyah Labuhanbatu, Rantau Prapat, Indonesia
Jl. H. Adam Malik (Lingkar By Pass), Rantau Prapat, Kec. Rantau Selatan, Kab. Labuhanbatu, Sumatera Utara
Email: [1,*]abdulkarim@ulb.ac.id, [2]kusnabara03@gmail.com
(* : coressponding author)

Abstract- Commodity price prediction is a vital aspect of supporting decision-making because price fluctuations can create uncertainty for businesses and stakeholders. This study aims to compare the performance of several machine learning algorithms for commodity price prediction, namely Random Forest, XGBoost, Support Vector Regression (SVR), Gradient Boosting, and Stacking Ensemble. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The experimental results show that Gradient Boosting achieved the best overall performance, with an RMSE of 219.07 and an R^2 of 0.9968, while Random Forest had the lowest MAE of 65.98. The Stacking Ensemble also demonstrated competitive performance, achieving an RMSE of 253.07 and an R^2 of 0.9958. In contrast, SVR produced the lowest performance, with an RMSE of 1,919.65 and an R^2 of 0.7555. Based on these results, Gradient Boosting was selected as the most optimal model because it provided the lowest prediction error and the highest ability to explain data variation. These findings demonstrate that selecting an appropriate machine learning algorithm plays a crucial role in improving commodity price prediction performance.

Keywords: Commodity Price Prediction; Machine Learning; Gradient Boosting; Stacking Ensemble; Random Forest

1. INTRODUCTION

Agricultural commodity prices are a key indicator in the food system and the economy, as price fluctuations influence production decisions, trade, producer income, consumption, and government policy. Commodity price dynamics have become increasingly complex due to rising economic uncertainty, shifts in global demand, supply chain disruptions, changes in input costs, and instability in international markets. Research on global food prices indicates that monetary factors, energy prices, changes in production, demand expectations, and geopolitical risks can contribute to fluctuations in food prices. These conditions make price forecasting crucial not only for trade purposes but also to support risk management and decision-making within the agricultural supply chain[1], [2].

From an economic perspective, commodity price formation can be understood through the interaction between supply and demand within a dynamic market environment. However, the characteristics of agricultural commodities mean that this mechanism does not always result in simple price patterns, as agricultural products vary in quality, variety, production location, market conditions, and the degree of interregional connectivity. Research by Brignoli et al. shows that predicting agricultural commodity prices holds strategic value for supply chain actors, as more accurate price information can be used for marketing decisions, risk management, and investment. At the same time, this research confirms that the use of machine learning offers opportunities to improve the accuracy of out-of-sample predictions for complex commodity prices[3], [4], [5].

Advances in artificial intelligence and machine learning offer new approaches to modeling the nonlinear relationships between commodity characteristics and prices. Unlike statistical models, which generally require specific assumptions about the form of the relationship between variables, machine learning can learn complex patterns from combinations of numerical and categorical variables[6], [7]. Further developments have led to the use of ensemble learning to obtain more robust predictive models. Ensemble learning combines multiple models to leverage their different learning characteristics, so that the weaknesses of one model can be compensated for by another. Research by Ribeiro and Coelho demonstrates that ensemble approaches based on bagging, boosting, and stacking can be used for short-term predictions in agribusiness time series, with ensemble performance that, in some cases, outperforms individual models such as SVR, KNN, and MLP [8]. Recent studies also indicate that the boosting ensemble approach remains a relevant area in agricultural product price forecasting because it is capable of handling volatile and complex price patterns [9], [10].

On a practical level, participants in the agricultural commodity supply chain face challenges such as price uncertainty and price differences across regions and markets. Farmers and traders need price information to help them determine the timing and location of sales, while policymakers need information to understand price dynamics and design more targeted market interventions. Differences in market characteristics, commodities, varieties, and grades mean that the market price of a product can vary even within the same commodity category. Studies on regional food price dynamics show that price patterns can form distinct regional clusters, making it important to consider geographic and market characteristics in price analysis [11], [12].

Another issue is the limitation of predictive models that rely solely on a single algorithm or a few price variables. Such approaches may not fully capture the complex relationships between market characteristics and product

Journal Global Tecnology Computer
Vol 5, No. 3, Agustus 2026, Hal 127-135
ISSN: 2809-6118 (Online - Elektronik)
DOI: 10.47065/jogtc.v5i3.11033
Websitehttps://ejurnal.seminar-id.com/index.php/jogtc

characteristics. For users of price information, models that only produce predicted values without identifying the factors contributing to those predictions also have limitations in supporting decision-making. Machine learning-based research using SHAP demonstrates that interpretability can be used to identify the contributions of various factors to changes in food prices. Therefore, there is a need for predictive models that not only have high accuracy but are also capable of simultaneously utilizing market and product characteristics and can be objectively compared with individual models.

The study by Sari et al. [13] developed various optimized machine learning techniques to predict agricultural commodity prices. This study is highly relevant because it demonstrates that machine learning optimization can improve price prediction capabilities. However, that study focused more on evaluating various optimized techniques, whereas this study focuses on developing an ensemble stacking method that integrates multiple base learners by simultaneously utilizing market and product characteristics. Thus, the difference lies in the model combination strategy and the construction of predictor information used.

Brignoli et al. [14] developed a machine learning model to predict grain futures prices and emphasized the importance of out-of-sample evaluation. Their study provides a crucial foundation demonstrating that machine learning can be applied in the context of agricultural economics and price forecasting. However, their research focused on futures prices and the characteristics of the futures market, whereas this study uses agricultural commodity price data that includes both market and product attributes such as region, market, commodity, variety, grade, minimum price, and maximum price. This study thus broadens the focus from market-series-based price prediction to capital-based price prediction using market and product characteristics.

Ribeiro and Coelho [15] demonstrated that a combination of bagging, boosting, and stacking can yield better forecasting performance in several agribusiness cases compared to individual models. The contributions of that study form an important basis for the use of ensemble methods in this research. However, the study by Ribeiro and Coelho focused on short-term forecasting of agribusiness time series, whereas this study emphasizes the integration of categorical information regarding markets and products into capital price regression models. Thus, the proposed study not only tests ensemble methods based on temporal patterns but also explores the contribution of market and product heterogeneity to price forecasting.

Oktoviany et al. [16] used external factors to build a machine learning model for predicting agricultural commodity prices. That study demonstrated the importance of incorporating additional information beyond historical prices when building predictive models. However, that approach focused on price state prediction, whereas this study uses modal price as the target for continuous regression. Furthermore, this study integrates market geographic characteristics and product characteristics as predictor variables and combines several base learners through ensemble stacking. In a recent study, Zhang et al. [17] used a boosting ensemble to predict agricultural product prices and demonstrated that the boosting approach can provide accurate predictions for price data characterized by high volatility and complex patterns. This study reinforces the relevance of ensemble learning in agricultural price forecasting. However, boosting and stacking employ different mechanisms for combining models. Boosting builds models sequentially to correct the errors of previous models, whereas stacking combines the predictions of multiple models through a meta-learner. This study capitalizes on these differences by combining Random Forest, XGBoost, Support Vector Regression, and Gradient Boosting as base learners and then using a meta-learner to generate the final prediction.[18], [19].

Based on these five previous studies, it can be stated that prior research has demonstrated the potential of machine learning, model optimization, external factors, boosting, and stacking in commodity price prediction. However, there remains room for research to develop models that specifically integrate market characteristics and product characteristics as predictors of commodity prices and combine multiple algorithms using a stacking ensemble approach. The research gap in this study lies in the suboptimal integration of market heterogeneity, product characteristics, and a combination of regression models within a single framework for predicting the spot price of agricultural commodities. Therefore, this study proposes the ensemble stacking approach to reduce reliance on a single algorithm and obtain more accurate and robust predictions.

This study aims to develop a price prediction model for agricultural commodities by utilizing market and product characteristics as predictor variables. Market characteristics include information on the region and market location, while product characteristics include the commodity, variety, and grade. The models were built using several machine learning algorithms—namely, Random Forest, XGBoost, Support Vector Regression, and Gradient Boosting—which were then combined using the ensemble stacking approach.

Specifically, this study aims to compare the performance of individual models with the ensemble stacking approach based on MAE, RMSE, MAPE, and R^2, as well as to identify the model that produces the best predictive performance. The study also aims to analyze the contribution of market and product characteristics to the prediction results so that the resulting model is not only accuracy-oriented but also provides relevant information for understanding the formation of agricultural commodity prices.

This study is expected to provide both academic and practical benefits. Academically, the study contributes to the development of ensemble machine learning applications in agricultural commodity price forecasting by integrating market and product information. Practically, the resulting model can serve as the basis for developing information systems or decision support systems for farmers, traders, supply chain actors, and policymakers in analyzing prices and reducing uncertainty in agricultural commodity trading decisions. The interpretability approach can also be used to identify the variables that contribute most significantly to prediction results, making the model's information easier to apply in economic and agribusiness contexts.

Journal Global Tecnology Computer
Vol 5, No. 3, Agustus 2026, Hal 127-135
ISSN: 2809-6118 (Online - Elektronik)
DOI: 10.47065/jogtc.v5i3.11033
Websitehttps://ejurnal.seminar-id.com/index.php/jogtc

2. RESEARCH METHODOLOGY

This study employs a quantitative approach with a machine learning-based experimental design to optimize price predictions for agricultural commodities. The research stages consist of dataset preparation, data understanding, data cleaning, exploratory data analysis (EDA), feature engineering, encoding, data splitting, base learner development, hyperparameter optimization, ensemble stacking, evaluation, and interpretation. Each stage is carried out sequentially to ensure that the data used is of sufficient quality and that the resulting model is capable of providing accurate and interpretable predictions. The overall research workflow is shown in Figure 1.

Figure 1. Research Workflow

The dataset contains market and product characteristics, namely State, District, Market, Commodity, Variety, Grade, Min Price, and Max Price, with Modal Price as the target variable. The data was processed by checking for missing values, duplicates, data type mismatches, and outliers. Next, exploratory data analysis (EDA) was conducted to identify the distributions and relationships among variables. Feature engineering was performed by creating relevant derived features, such as Price Range and Mid Price. Categorical variables were then converted to numerical form using appropriate encoding techniques. The entire preprocessing process was applied consistently to avoid data leakage.

The data was split into training, validation, and testing sets. Four algorithms were used as base learners: Random Forest Regression, XGBoost Regression, Support Vector Regression (SVR), and Gradient Boosting Regression. Each base learner was optimized using Grid Search/Randomized Search and cross-validation to obtain the best configuration. Subsequently, the predictions from the base learners were combined using an ensemble stacking method with a meta-learner. Out-of-fold predictions are used during the stacking training phase to prevent data leakage. Model performance is evaluated using MAE, RMSE, MAPE, and R^2. The best model is determined based on the lowest error value and the highest R^2. Next, an interpretation is performed using feature importance or SHAP to identify the contributions of market and product characteristics to price predictions.

2.1 Random Forest Regression

Random Forest Regression is an ensemble method that combines multiple decision trees to generate regression predictions. This method is capable of handling nonlinear relationships, interactions between variables, and data with high variability. In agricultural research, Random Forest has proven to be effective in predicting both yield and price variables because it can utilize many features simultaneously. The use of Random Forest in commodity forecasting also demonstrates that this method can serve as a strong benchmark against boosting algorithms [20].

2.2 XGBoost Regression

XGBoost Regression is a gradient boosting algorithm that builds decision trees incrementally to correct the prediction errors of the previous model. XGBoost is widely used because it can handle nonlinear patterns and complex interactions between features and has a regularization mechanism to reduce overfitting. Recent research shows that XGBoost is capable of delivering high performance in agricultural prediction problems, especially when the variables used have complex relationships with the target [21].

2.3 Support Vector Regression (SVR)

Support Vector Regression is a regression method based on Support Vector Machines that can handle nonlinear relationships through kernel functions. SVR works by finding a regression function whose error falls within a certain tolerance limit. The strength of SVR lies in its ability to generalize to datasets of relatively small to medium size. However, its performance is heavily influenced by kernel selection and parameters such as C, ε, and γ. Recent research also indicates

Journal Global Tecnology Computer
Vol 5, No. 3, Agustus 2026, Hal 127-135
ISSN: 2809-6118 (Online - Elektronik)
DOI: 10.47065/jogtc.v5i3.11033
Websitehttps://ejurnal.seminar-id.com/index.php/jogtc

that SVR can be used for commodity prediction, but its results may be inferior to those of ensemble methods when the data exhibits highly complex patterns [22]

2.4 Gradient Boosting Regression

Gradient Boosting Regression is an ensemble method that builds models incrementally, where each new model is used to correct the errors of the previous model. This approach allows Gradient Boosting to capture nonlinear patterns and complex relationships in the data. Boosting methods are widely used in commodity forecasting because they can achieve high accuracy on data with complex characteristics. Therefore, Gradient Boosting is used in this study to be compared with Random Forest, XGBoost, and SVR in determining the most suitable model for commodity price forecasting [21]

3. RESULTS AND DISCUSSION

3.1 Data Preprocessing and Characteristics

The preprocessing stage was conducted to ensure that the agricultural commodity price data was in a condition suitable for use in the modeling process. The stages included checking for missing values, removing duplicate data, validating price values, transforming data types, and creating new features. Feature engineering produced the “Price Range” variable as the difference between the maximum and minimum prices and the “Mid Price” as the midpoint between those two prices. Categorical variables representing market and product characteristics were then transformed using one-hot encoding.

The processed data is then divided into an 80% training set and a 20% testing set. Feature transformation is performed on the training data and applied to the testing data to prevent data leakage. These steps produce data ready for use in building Random Forest, XGBoost, SVR, Gradient Boosting, and Stacking Ensemble models.

Table 1. Dataset

State	District	Market	Commodity	Variety	Grade	Min Price	Max Price	Base Price
Andhra Pradesh	Chittor	Chittoor	Gur (Jaggery)	No. 2	FAQ	3200.0	3,400.0	3200.0
Andhra Pradesh	Chittor	Chittoor	Mango	Neelam	Medium	700.0	1,500.0	1,200.0
Andhra Pradesh	Chittor	Chittoor	Mango	Totapuri	Medium	1400.0	1,800.0	1,600.0
Andhra Pradesh	Cuddapah	Cuddapah	Groundnut	Local	FAQ	4059.0	7589.0	7,559.0
Andhra Pradesh	Cuddapah	Cuddapah	Turmeric	Bulb	FAQ	4778.0	6,160.0	5,845.0
Andhra Pradesh	Cuddapah	Cuddapah	Turmeric	Finger	FAQ	3012.0	6,619.0	6,175.0
Andhra Pradesh	East Godavari	Peddapuram	Paddy (Dhan) (Common)	1001	FAQ	2040.0	2050.0	2045.0
Andhra Pradesh	East Godavari	Pithapuram	Paddy (Dhan) (Common)	1001	FAQ	2040.0	2060.0	2050.0
Andhra Pradesh	East Godavari	Prattipadu	Paddy (Dhan) (Common)	1001	FAQ	2040.0	2060.0	2050.0
Andhra Pradesh	East Godavari	Sampara	Paddy (Dhan) (Common)	Swarna Masuri (New)	FAQ	2150.0	2200.0	2180.0
Andhra Pradesh	Guntur	Duggirala	Turmeric	Bulb	FAQ	5600.0	6,300.0	6,000.0
Andhra Pradesh	Guntur	Duggirala	Turmeric	Finger	FAQ	5725.0	6,300.0	6,000.0
Andhra Pradesh	Guntur	Tenali	Black Gram (Urd Beans) (Whole)	Black Gram (Whole)	FAQ	8100.0	8300.0	8200.0
Andhra Pradesh	Guntur	Tenali	Lemon	Lemon	FAQ	1000.0	2,500.0	1,500.0
Andhra Pradesh	Kurnool	Alur	Jowar (Sorghum)	Jowar (White)	FAQ	3020.0	3080.0	3050.0
Andhra Pradesh	Kurnool	Kurnool	Groundnut	Local	FAQ	6089.0	7,921.0	7509.0

Vol 5, No. 3, Agustus 2026, Hal 127-135
ISSN: 2809-6118 (Online - Elektronik)
DOI: 10.47065/jogtc.v5i3.11033
Websitehttps://ejurnal.seminar-id.com/index.php/jogtc

3.2 Model Comparison Results

Test results for Random Forest, XGBoost, SVR, Gradient Boosting, and Stacking Ensemble show differences in performance among the algorithms in predicting the capital prices of agricultural commodities. The evaluation used MAE, RMSE, MAPE, and R^2. Lower MAE, RMSE, and MAPE values indicate smaller prediction errors, while a higher R^2 indicates the model's better ability to explain price variations.

Tabel 2. Comparison Results

Model	MAE	RMSE	MAPE (%)	R^2
Random Forest	**65.98**	279.09	451.10	0.9948
XGBoost	99.00	633.27	132.80	0.9734
SVR	222.27	1,919.65	**87.77**	0.7555
Gradient Boosting	74.04	**219.07**	271.94	**0.9968**
Ensemble Stacking	85.65	253.07	292.09	0.9958

Based on the test results, Gradient Boosting demonstrated the best overall performance, with an RMSE of 219.07 and an R^2 of 0.9968. These values indicate that the model was able to produce the lowest prediction error based on RMSE and explain approximately 99.68% of the variation in capital prices in the test data. Meanwhile, Random Forest produced the lowest MAE of 65.98, indicating the smallest mean absolute error. XGBoost produced an R^2 of 0.9734 with an RMSE of 633.27, while SVR produced the highest RMSE of 1,919.65 and the lowest R^2 of 0.7555. These results show that *tree-based ensemble* models are better able to capture nonlinear patterns and interactions between variables in the dataset compared to SVR.

Stacking Ensemble produced an R^2 of 0.9958 and an RMSE of 253.07. Although it performed very well, Stacking Ensemble did not outperform Gradient Boosting. The RMSE of Stacking Ensemble was approximately 15.52% higher than that of Gradient Boosting. These findings suggest that combining multiple *base learners* does not always improve accuracy, especially when one of the individual models is already capable of representing the data patterns very well. MAPE values must be interpreted with caution due to significant differences among models. A high MAPE can be influenced by the presence of small actual values; therefore, this metric is not used as the sole basis for model selection. Based on a combination of MAE, RMSE, and R^2, Gradient Boosting was determined to be the best-performing model in this study.

3.3 Analysis of the Best Model

Based on the test results, Gradient Boosting is the model with the best performance in predicting the capital prices of agricultural commodities. This model yields an MAE of 74.04, an RMSE of 219.07, a MAPE of 271.94%, and an R^2 of 0.9968. The lowest RMSE and highest R^2 compared to other models indicate that Gradient Boosting is capable of producing predictions that most closely approximate the actual values while explaining approximately 99.68% of the price variation in the test data.

The advantages of Gradient Boosting demonstrate that the boosting approach is capable of capturing the nonlinear relationships between market and product characteristics and commodity prices. The stepwise model-building process allows each weak learner to correct the errors of the previous model, enabling more effective learning of complex patterns in the data. This is one of the reasons Gradient Boosting is able to produce a lower RMSE compared to Random Forest, XGBoost, SVR, and Stacking Ensemble. When compared to Random Forest, Gradient Boosting has a slightly higher MAE—74.04 versus 65.98—but produces a lower RMSE—219.07 versus 279.09—as well as a higher R^2—0.9968 versus 0.9948. These results indicate that Random Forest has a smaller average absolute error ( ), but Gradient Boosting is better at controlling large prediction errors. Therefore, Gradient Boosting was selected as the best model based on a combination of evaluation metrics.

These results also show that Ensemble Stacking has not yet yielded a performance improvement. Stacking produced an RMSE of 253.07 and an R^2 of 0.9958, while Gradient Boosting produced an RMSE of 219.07 and an R^2 of 0.9968. Thus, in this research dataset, the use of multiple base learners has not yet been able to outperform Gradient Boosting as a single model. This finding demonstrates that more complex models do not always yield higher accuracy. A model's success is heavily influenced by data characteristics, feature selection, hyperparameter configuration, and the algorithm's ability to capture patterns of relationships among variables. In the context of this study, Gradient Boosting provided the most consistent results and can therefore be established as the primary model for predicting agricultural commodity prices based on market and product characteristics.

3.4 Evaluation of Stacking Ensemble

The Stacking Ensemble produced an MAE of 85.65, an RMSE of 253.07, a MAPE of 292.09%, and an R^2 of 0.9958. This R^2 value indicates that the Stacking Ensemble still possesses very high predictive capability. However, the test results show that this model has not yet been able to outperform Gradient Boosting.

Table 3. Comparison of Gradient Boosting and Stacking Ensemble

Journal Global Tecnology Computer
Vol 5, No. 3, Agustus 2026, Hal 127-135
ISSN: 2809-6118 (Online - Elektronik)
DOI: 10.47065/jogtc.v5i3.11033
Websitehttps://ejurnal.seminar-id.com/index.php/jogtc

Metrics	Gradient Boosting	Stacking Ensemble	Changes
MAE	74.04	85.65	−15.67%
RMSE	219.07	253.07	−15.52%
R^2	0.9968	0.9958	−0.10%

Based on Table 2, the RMSE of the Stacking Ensemble, at 253.07, is higher than that of Gradient Boosting, at 219.07, representing a performance decline of approximately 15.52%. The MAE also increased from 74.04 to 85.65, while R^2 decreased from 0.9968 to 0.9958. Thus, the complexity of the Stacking Ensemble model has not provided a predictive advantage in this study's configuration. These findings indicate that adding multiple base learners does not always result in more accurate predictions. If individual models are already capable of capturing the main patterns in the data effectively, combining multiple models may offer limited benefits. In this study, Gradient Boosting actually produced more consistent results than the Stacking Ensemble.

3.5 Analysis of Actual vs. Predicted Values and Residuals

The *Actual vs. Predicted* visualization is used to evaluate how closely the predicted results align with the actual values. Points that are closer to the diagonal line y=x indicate a higher level of prediction accuracy.
In Gradient Boosting, an R^2 value of **0.9968** indicates that the predicted points generally align very closely with the actual values. This finding reinforces the numerical evaluation results showing that Gradient Boosting is capable of tracking capital price variations effectively. The deviations still observed in some observations suggest that certain price variations cannot yet be fully explained by the market and product characteristics used.

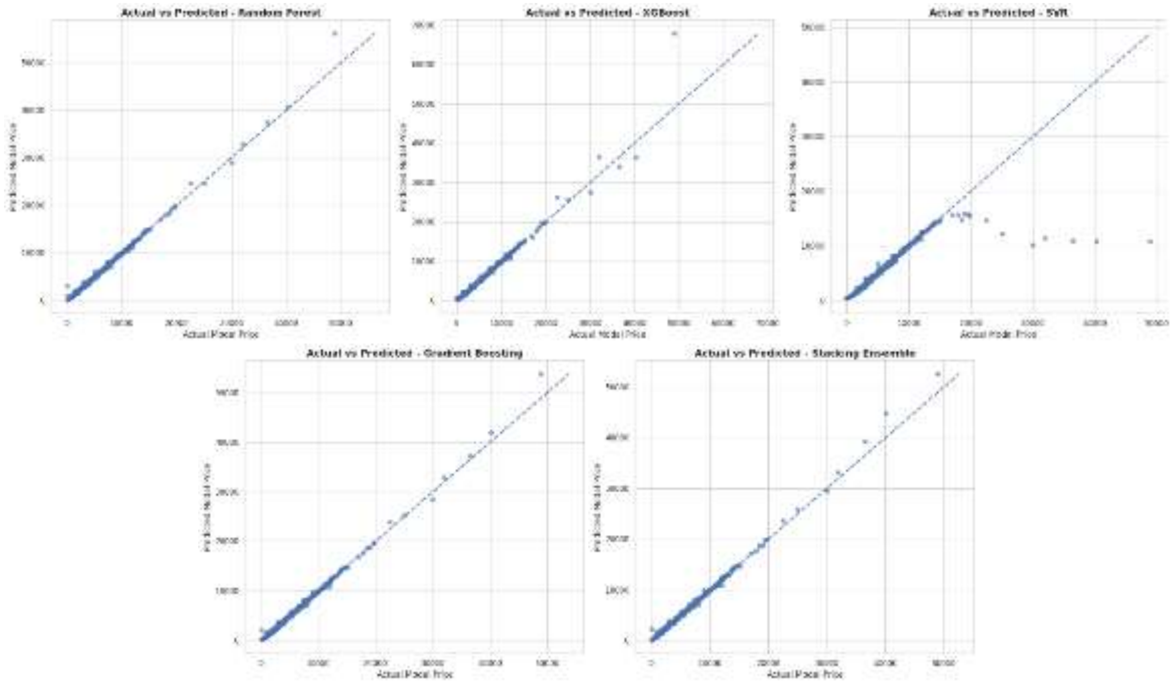

Figure 2. Visualization of Actual vs. Predicted

Figure 2 shows a comparison between the actual price (Actual Modal Price) and the predicted price (Predicted Modal Price) from five models. The dashed line represents the ideal condition, where the predicted value is equal to the actual value. In general, Random Forest, XGBoost, Gradient Boosting, and Stacking Ensemble have data points that are fairly close to the ideal line, especially at lower price values, indicating that these models are capable of providing predictions that are fairly close to the actual values. Gradient Boosting and Stacking Ensemble appear to exhibit a fairly good pattern, as their prediction points follow the diagonal line relatively consistently, even at higher prices. Random Forest also shows fairly good results, although there are a few points that begin to deviate at certain price levels. XGBoost still follows the ideal line pattern, but some deviations are visible at higher prices. Meanwhile, SVR shows the most obvious deviation, especially as the actual price rises, because its predicted values tend to plateau at a certain range and no longer follow the increase in the actual price. Thus, based on this visualization, Gradient Boosting and Stacking Ensemble appear to be the most consistent in tracking actual values, while SVR has limitations in predicting prices in the high-value range.

Residual analysis is used to identify patterns of model error through the difference between actual and predicted values: A good model is expected to produce residuals that are scattered around zero without forming any specific pattern. Based on the residual visualization, an evaluation is conducted to determine whether the prediction errors tend to be systematic or random.

Journal Global Tecnology Computer
Vol 5, No. 3, Agustus 2026, Hal 127-135
ISSN: 2809-6118 (Online - Elektronik)
DOI: 10.47065/jogtc.v5i3.11033
Websitehttps://ejurnal.seminar-id.com/index.php/jogtc

If the Gradient Boosting residuals are scattered relatively randomly around the zero line, the results indicate that the model has been able to capture the main patterns in the data. Conversely, if there is a specific pattern, this may indicate the presence of information not yet represented in the model, such as seasonal factors, supply and demand, distribution, or external market conditions.

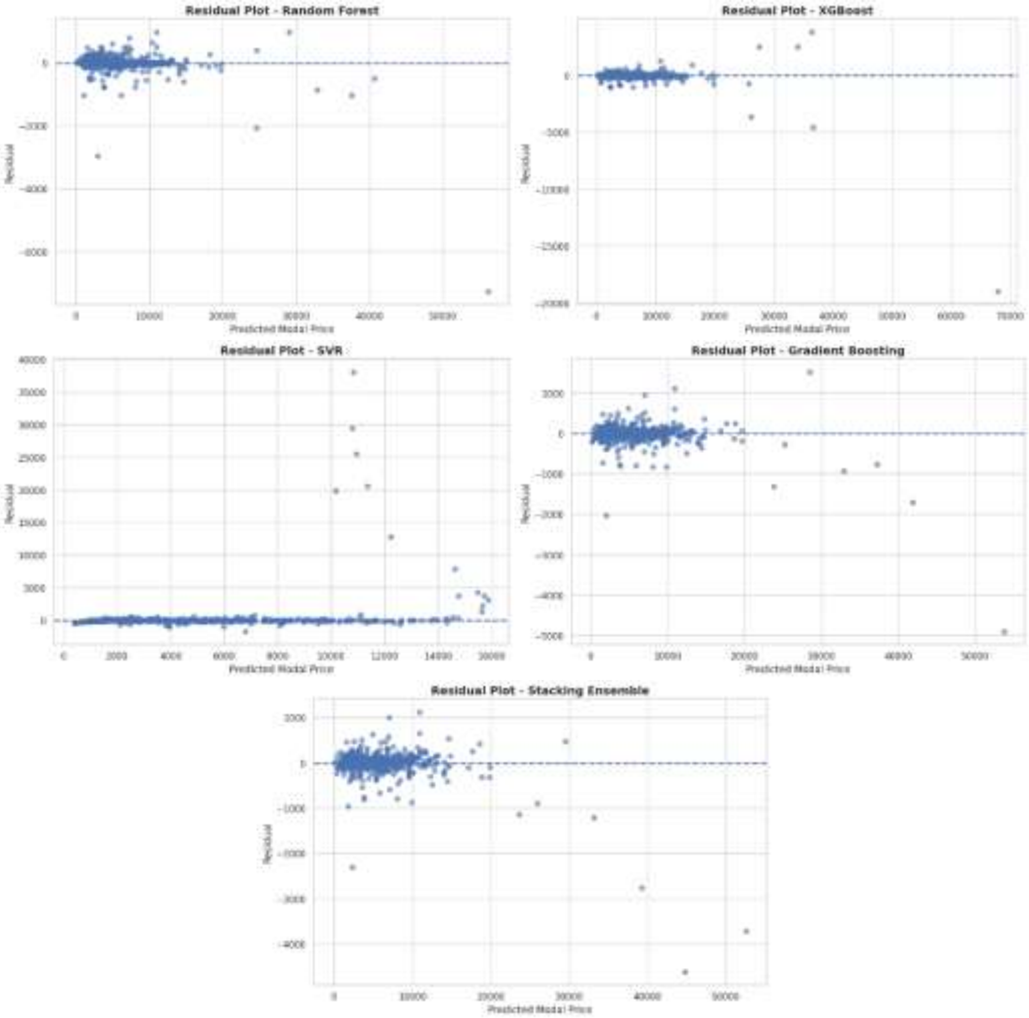

Figure 3. Residuals

Based on the residual plot, it can be seen that most of the residuals for all five models are clustered around the zero line, indicating that the prediction results are generally quite close to the actual values. However, there are still some points with relatively large residuals or outliers. In Random Forest and XGBoost, there are several quite extreme negative outliers, while SVR has several very high positive outliers. Gradient Boosting exhibits a relatively more controlled residual pattern, while Stacking Ensemble has a residual distribution that is quite close to the zero line and relatively fewer outliers. Overall, this graph shows that Stacking Ensemble and Gradient Boosting have more stable residual patterns compared to the other models, although determining the best model still requires examining evaluation results based on MAE, RMSE, and R^2 values.

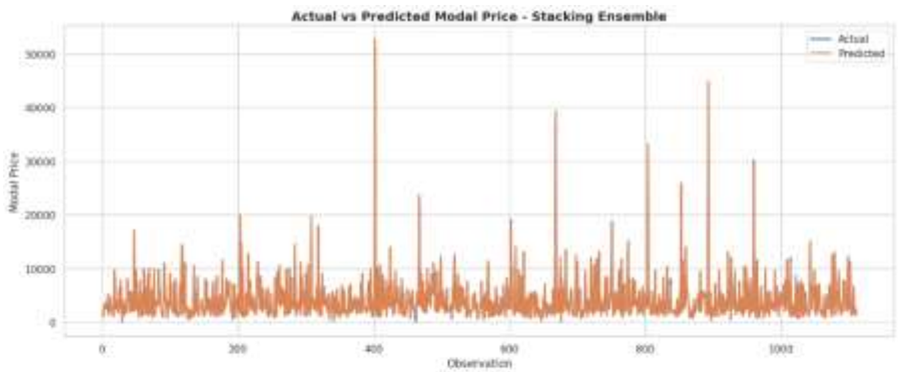

Figure 4. Actual vs. Predicted Based on Data Order

Vol 5, No. 3, Agustus 2026, Hal 127-135
ISSN: 2809-6118 (Online - Elektronik)
DOI: 10.47065/jogtc.v5i3.11033
Websitehttps://ejurnal.seminar-id.com/index.php/jogtc

Figure 4 shows a comparison between the actual capital price (Actual) and the predicted capital price (Predicted) from the Stacking Ensemble model for each observation. In general, the actual and predicted lines appear quite close for most observations, particularly in the price range of approximately 2,000–10,000, indicating that the model is able to follow the actual price pattern quite well. However, for some observations, there are significant spikes in the predicted values—for example, around observations 400, 660, 800, 850, and 900—where the predicted values are much higher than the actual values. This indicates that the model still experiences prediction errors for certain data points with specific price characteristics, particularly at extreme values. Despite these deviations, the overall prediction pattern still follows actual price changes and does not show significant discrepancies in the majority of observations. Thus, the Stacking Ensemble can be considered reasonably capable of predicting stock prices, but attention must still be paid to the presence of some outliers that cause predictions to be excessively high at certain observations.

4. CONCLUSION

Based on the research results, Gradient Boosting is the best model for predicting commodity prices compared to Random Forest, XGBoost, SVR, and Stacking Ensemble. The Gradient Boosting model produced the lowest RMSE of 219.07 and the highest R^2 of 0.9968, indicating the model's ability to explain approximately 99.68% of the data's variation. Although Random Forest has the lowest MAE value of 65.98, overall, Gradient Boosting provides the most optimal performance based on a combination of key evaluation metrics. Stacking Ensemble also demonstrates excellent performance with an R^2 of 0.9958 and an RMSE of 253.07, but it has not yet been able to surpass Gradient Boosting. Meanwhile, SVR showed the lowest performance with an RMSE of 1,919.65 and an R^2 of 0.7555. Thus, Gradient Boosting can be considered the most effective model for prediction on this research dataset.

REFERENCES

[1] K. Kurumatani, "Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method," *SN Applied Sciences 2020 2:8*, vol. 2, no. 8, pp. 1434–, Jul. 2020, doi: 10.1007/S42452-020-03225-9.

[2] E. Rahmani, M. Khatami, and E. Stephens, "Using Probabilistic Machine Learning Methods to Improve Beef Cattle Price Modeling and Promote Beef Production Efficiency and Sustainability in Canada," *Sustainability 2024, Vol. 16, p. 1789*, Vol. 16, No. 5, p. 1789, Feb. 2024, doi: 10.3390/SU16051789.

[3] N. Zhang, Q. An, S. Zhang, and H. Ma, "Price Prediction for Fresh Agricultural Products Based on a Boosting Ensemble Algorithm," *Mathematics 2025, Vol. 13, p. 71*, vol. 13, no. 1, p. 71, Dec. 2024, doi: 10.3390/MATH13010071.

[4] E. Rahmani, M. Khatami, and E. Stephens, "Using Probabilistic Machine Learning Methods to Improve Beef Cattle Price Modeling and Promote Beef Production Efficiency and Sustainability in Canada," *Sustainability 2024, Vol. 16, p. 1789*, Vol. 16, No. 5, p. 1789, Feb. 2024, doi: 10.3390/SU16051789.

[5] C. Liu, L. Zhou, L. Höschle, and X. Yu, "Food price dynamics and regional clusters: machine learning analysis of egg prices in China," **China Agricultural Economic Review**, vol. 15, no. 2, pp. 416–432, May 2023, doi: 10.1108/CAER-01-2022-0003.

[6] R. Gupta and C. Pierdzioch, "Multi-Task Forecasting of the Realized Volatilities of Agricultural Commodity Prices," *Mathematics 2024, Vol. 12, p. 2952*, vol. 12, no. 18, p. 2952, Sep. 2024, doi: 10.3390/MATH12182952.

[7] X. Han, T. Yuan, D. Wang, Z. Zhao, and B. Gong, "How to Understand High Global Food Prices? Using SHAP to Interpret Machine Learning Algorithms," *PLoS One*, vol. 18, no. 8, p. e0290120, Aug. 2023, doi: 10.1371/JOURNAL.PONE.0290120.

[8] A. Thaker, L. H. Chan, and D. Sonner, "Forecasting Agricultural Commodity Futures Prices Using Convolutional Neural Networks: An Application to Wheat Futures," *Journal of Risk and Financial Management 2024, Vol. 17, p. 143*, vol. 17, no. 4, p. 143, Apr. 2024, doi: 10.3390/JRFM17040143.

[9] C. Sun, M. Pei, B. Cao, S. Chang, and H. Si, "A Study on an Agricultural Commodity Price Prediction Model Based on Secondary Decomposition and a Long Short-Term Memory Network," *Agriculture 2024, Vol. 14, p. 60*, Vol. 14, No. 1, p. 60, Dec. 2023, doi: 10.3390/AGRICULTURE14010060.

[10] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 13–17, 2016, pp. 785–794, Aug. 2016, doi: 10.1145/2939672.2939785;CSUBTYPE:STRING:CONFERENCE.

[11] C. Liu, L. Zhou, L. Höschle, and X. Yu, "Food price dynamics and regional clusters: a machine learning analysis of egg prices in China," **China Agricultural Economic Review**, vol. 15, no. 2, pp. 416–432, Feb. 2023, doi: 10.1108/CAER-01-2022-0003.

[12] A. Thaker, L. H. Chan, and D. Sonner, "Forecasting Agricultural Commodity Futures Prices Using Convolutional Neural Networks: An Application to Wheat Futures," *Journal of Risk and Financial Management 2024, Vol. 17, p. 143*, Vol. 17, No. 4, p. 143, Apr. 2024, doi: 10.3390/JRFM17040143.

[13] M. Sari, S. Duran, H. Kutlu, B. Guloglu, and Z. Atik, "Various optimized machine learning techniques to predict agricultural commodity prices," *Neural Computing and Applications 2024 36:19*, vol. 36, no. 19, pp. 11439–11459, Apr. 2024, doi: 10.1007/S00521-024-09679-X.

[14] P. L. Brignoli, A. Varacca, C. Gardebroek, and P. Sckokai, "Machine learning to predict grain futures prices," *Agricultural Economics (United Kingdom)*, vol. 55, no. 3, pp. 479–497, May 2024, doi: 10.1111/AGEC.12828;WGROUP:STRING:PUBLICATION.

[15] M. H. D. M. Ribeiro and L. dos Santos Coelho, "An ensemble approach based on bagging, boosting, and stacking for short-term prediction in agribusiness time series," *Appl. Soft Comput.*, vol. 86, p. 105837, Jan. 2020, doi: 10.1016/J.ASOC.2019.105837.

[16] P. Oktoviany, R. Knobloch, and R. Korn, "A machine learning-based price trend prediction model for agricultural commodities using external factors," *Decisions in Economics and Finance 2021 44:2*, vol. 44, no. 2, pp. 1063–1085, Nov. 2021, doi: 10.1007/S10203-021-00354-7.

[17] N. Zhang, Q. An, S. Zhang, and H. Ma, "Price Prediction for Fresh Agricultural Products Based on a Boosting Ensemble Algorithm," *Mathematics 2025, Vol. 13, p. 71*, vol. 13, no. 1, p. 71, Dec. 2024, doi: 10.3390/MATH13010071.

[18] Z. Budiarso, H. Listiyono, and A. Karim, "Optimizing LSTM with Grid Search and Regularization Techniques to Enhance Accuracy in Human Activity Recognition," *Journal of Applied Data Sciences*, vol. 5, no. 4, pp. 2002–2014, Nov. 2024, doi: 10.47738/jads.v5i4.433.

[19] Retno Andani Sundari, Poningsih, and A. Karim, "Enhancing Lung Cancer Detection: Optimizing CNN Architectures through Hyperparameter Tuning," https://jurnal.iaii.or.id/index.php/RESTI/article/view/6357/1127. Accessed: Dec. 19, 2025. [Online]. Available: https://jurnal.iaii.or.id/index.php/RESTI/article/view/6357/1127

[20] A. Theofilou, S. A. Nastis, A. Michailidis, T. Bournaris, and K. Mattas, "Predicting Prices of Staple Crops Using Machine Learning: A Systematic Review of Studies on Wheat, Corn, and Rice," *Sustainability 2025, Vol. 17, p. 5456*, vol. 17, no. 12, p. 5456, Jun. 2025, doi: 10.3390/SU17125456.

[21] J. C. M. Sánchez, H. G. A. Mesa, A. T. Espinosa, S. R. Castilla, and F. G. Lamont, "Improving wheat yield prediction through variable selection using Support Vector Regression, Random Forest, and Extreme Gradient Boosting," *Smart Agricultural Technology*, vol. 10, p. 100791, Mar. 2025, doi: 10.1016/J.ATECH.2025.100791.

[22] R. L. Manogna, V. Dharmaji, and S. Sarang, "Enhancing agricultural commodity price forecasting with deep learning," *Scientific Reports 2025 15:1*, vol. 15, no. 1, pp. 20903–, Jul. 2025, doi: 10.1038/s41598-025-05103-z.