Journal of Theoretical and Applied Information Technology
15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MODIFIED RANDOM FOREST REGRESSION MODEL FOR PREDICTING WHOLESALE RICE PRICES

CHRISTINE DEWI[1], GRUDA SAKTI KRIDA PRASATYA[2], HENOCH JULI CHRISTANTO[3*], SANDRA OCTAVIANI B WIDIARTO[4], GUOWEI DAI[5]

[1,2]Satya Wacana Christian University, Faculty of Information and Technology, Salatiga, Indonesia
[3,4]Atma Jaya Catholic University of Indonesia, Information System Department, Faculty of Engineering, Jakarta, Indonesia
[5]Agricultural Information Institute of CAAS, National Agriculture Science Data Center, Beijing 100081, China

E-mail: [1] christine.dewi@uksw.edu, [2]672019252@student.uksw.edu [3]henoch.christanto@atmajaya.ac.id, [4]sandra.oct@atmajaya.ac.id, [5]dgwstyle@foxmail.com

ABSTRACT

Both in terms of diet and economy, Indonesian people attach great importance to rice as a staple food. In addition, it is very important to monitor rice price fluctuations every month so that overall rice prices remain stable and do not burden the community. Tracking rice price fluctuations helps rice producers, traders, and businesses make informed decisions about when to buy, sell, or store rice. This can optimize their supply chain management and maximize profits. Researchers and analysts can use rice price data to study market trends, identify patterns, and develop predictive models for future price movements. This research purpose to determine the most optimal forecasting model by using the Average Rice Price dataset at the Indonesian Wholesale Trade Level from January 2010 to December 2022. The dataset is obtained from the Central Statistics Agency of Indonesia. Moreover, the best model proposed in this research uses the Random Forest method with hyperparameter tuning using the n estimator parameter of 500. Our proposed method can reduce the MAPE value from 0.0093573 to 0.0089389 and increase the R2 Score value from 0.9916805 to 0.9921578. Moreover, we analyze the performance of our proposed methodology with several other datasets sourced from UCI (University of California Irvine). The experimental outcomes indicate that the suggested model displays superior performance when compared to alternative methods, with a tendency of decreasing MAPE values and increasing R2 values in each experiment for all datasets.

Keywords: *Random Forest, Rice Price Prediction, Machine Learning.*

1. INTRODUCTION

The adequacy of food for a country is very important to realize the development of human resources that are healthy, active, and productive. The primary food commodities consumed by Indonesian people are rice, corn, tubers, and sago. However, the majority of Indonesian people choose rice as a source of fulfilling their daily needs [1]. Rice is an important commodity in the social and economic life of the Indonesian people. For much of the Indonesian population, rice has the status of a staple food because almost the entire population of Indonesia needs rice as its main food ingredient. Apart from being a source of nutrition in the food structure, the provision of rice stocks is very important considering that Indonesia's population is increasing every year [2].

The stability and price level of rice will affect people's accessibility to rice as a food ingredient. Foodstuffs must always be available in sufficient quantities, of high quality and are medically safe for consumption [3]. The availability of staple foods, especially rice, is affected by the amount of rice production in an area. In Indonesia, rice production continues to increase every year [4]. The price of rice has increased every year and the increase in rice prices was caused by issues circulating about the implementation of rice import policies. In addition to the low rice harvest, the increase in rice prices was caused by high demand due to empty stocks in the market [5]. Considering that the price of rice sold by wholesalers will have an impact on the community, it is necessary to forecast rice prices for the next period to maintain the stability of the average price of rice sold by

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

wholesalers[6,7]. Forecasting is a method that can predict a value in the future [8].

Predicting wholesale rice prices can offer several benefits for various stakeholders in the rice industry and policymakers, consumers, and researchers. Some potential benefits of accurate wholesale rice price prediction are as follows: (1) Market Insights and Decision-Making: Accurate price predictions can provide valuable insights into market trends, supply and demand dynamics, and potential price fluctuations. This information is crucial for farmers, traders, processors, and other participants in the rice supply chain to make informed decisions about production, distribution, and pricing strategies. (2) Risk Management: Price prediction models can help stakeholders manage and mitigate risks associated with price volatility. Farmers can plan their planting and harvesting schedules, input purchases, and sales strategies based on anticipated price changes, reducing their exposure to financial losses. (3) Supply Chain Optimization: Knowledge of future price movements can aid in optimizing the entire supply chain, from production to distribution. Efficient supply chain management can result in cost savings, reduced waste, and improved resource allocation. (4) Inventory Management: Rice processors, distributors, and retailers can use price predictions to optimize their inventory management. They can adjust their stock levels based on expected price trends, ensuring they have the right amount of rice available to meet customer demands without holding excessive inventory. (5) Price Stabilization: Governments and policymakers can use price prediction models to implement effective price stabilization measures, such as releasing rice reserves during periods of high prices to stabilize the market and protect consumers from sudden price spikes. (6) Consumer Impact: Accurate price predictions can help consumers plan their food budgets better and make informed purchasing decisions. When consumers have an idea of future price trends, they can adjust their consumption patterns or make bulk purchases when prices are expected to rise. (7) Research and Development: Researchers can use historical price data and predictive models to study the factors influencing rice prices, such as weather conditions, geopolitical events, economic indicators, and technological advancements. This knowledge can lead to better insights into market behavior and potentially guide future policy recommendations. (8) Price Transparency: Price prediction models contribute to increased transparency in the rice market. This transparency can foster trust among market participants and reduce information asymmetry, benefiting all stakeholders. (9) Financial Planning: Investors and financial institutions operating in the agricultural sector can use price predictions to develop investment strategies, assess risk, and make lending decisions. (10) Sustainability and Resource Allocation: With accurate price predictions, farmers can make more informed decisions about resource allocation, such as water and fertilizers, optimizing their operations for both economic and environmental sustainability.

Machine learning can be used for rice price prediction by using historical data to train a model that can accurately forecast future prices. Random Forest is a machine learning method that finds application in tasks such as classification, regression, and other related activities [9]. It belongs to the category of ensemble learning techniques, which merge multiple decision trees to enhance the accuracy of predictions. Additionally, Random Forest is a potent machine learning algorithm that has numerous applications, such as forecasting stock prices, rice prices, and diagnosing diseases [10].

The main contributions of this research can be summarized as follows: (1) Comparing the Mean Absolute Percentage Error (MAPE) values from previous research with different machine learning models, such as SVR, Linear Regression, Decision Tree, GBR, and Random Forest. (2) Establishing the Random Forest method to predict rice prices in this research. Random Forest combines trees by training on the data they have. This method is commonly used because it produces a low percentage of errors, and the accuracy results obtained are quite high [11]. (3) Set hyperparameter tuning to find the optimal value of the Random Forest method. (4) Comparing the error rate and accuracy results of the Random Forest hyperparameter tuning experiments that have been conducted.

The remainder of the paper is organized as follows. An overview of the materials and methods utilized in this research is presented in Section 2. Our study's results and discussion are presented in Section 3, and Section 4 outlines our conclusions and future research goals.

2. MATEIAL AND METHODS

2.1 Forecasting

Forecasting is a methodology employed to estimate future values based on past and present data [12]. But that doesn't mean that after learning this technique we can predict anything accurately, but only learn certain techniques that can be applied to certain situations. In the process of selecting a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

prediction method, it is essential to consider the type of data pattern to test the most suitable method for that particular pattern [13]. Forecasting is used to predict future values or outcomes based on historical data and patterns. It is widely applied across various domains, including economics, finance, business, science, and technology. The goal of forecasting is to make informed decisions and plans by estimating the likely future developments. Besides, forecasting is the process of predicting future values of a variable based on historical data and other relevant factors. It is a critical aspect of decision-making in many industries, including finance, marketing, and operations management [14]. Furthermore, forecasting can be applied to a broad spectrum of applications, including sales forecasting, demand forecasting, and inventory management [15]. Forecasting plays a crucial role in business planning, resource allocation, inventory management, budgeting, and policymaking. While it can provide valuable insights, it's important to note that forecasting is subject to uncertainties and assumptions, and the accuracy of forecasts may vary depending on the complexity of the data and the chosen model.

2.2 Random Forest

Random Forest (RF) is a machine learning technique that improves prediction accuracy through the introduction of randomness. It creates child nodes randomly in its decision tree structure, distinguishing it from traditional decision tree models. This technique is commonly used for both classification and regression tasks. It falls under the category of ensemble learning methods, which involve combining the predictions of multiple individual models to improve overall performance and generalization. The tree-based model works by repeatedly splitting the original dataset into two subgroups using a specific criterion, until a predefined stopping condition is met as seen in Figure 1 [16]. This approach involves constructing a decision tree containing root nodes, internal nodes, and leaf nodes by randomly selecting attributes and data in compliance with relevant guidelines. The topmost node, also known as the root, is situated at the apex of the decision tree [17]. An internal node is a node that splits into branches, having a single input and at least two outputs. The leaf node, also known as the terminal node, is the final node that receives one input and produces no output [18].

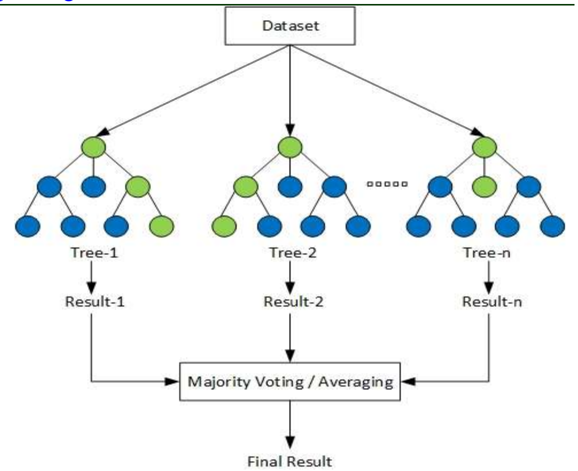


Figure 1: Random Forest Architecture

Predictions in the regression case are taken from the average value of each tree. Formula (1) used for calculating the average value of all tree predictions [19]:

$$\hat{Y}_i = \frac{1}{N_{tree}} \sum_{n=1}^{N_{tree}} \hat{Y}_n \quad (1)$$

Description:

Y^_i = Final prediction results
N_tree = The overall count of trees presents in Random Forest
Y^_n = n-tree prediction results

In Random Forest, the "n_estimators" hyperparameter refers to the number of decision trees that will be used to build the model. The higher the number of estimators, the more accurate the model is likely to be, but this comes at the cost of increased computational time and memory usage [20]. Increasing the number of estimators typically leads to a better fit of the model to the training data, given that each decision tree is trained using a distinct subset of the data and features. However, at a certain point, the benefit of additional estimators may become marginal or even lead to overfitting [21].

The "n_estimators" hyperparameter is a key parameter in ensemble machine learning algorithms, particularly in methods like Random Forest and Gradient Boosting. It determines the number of base models (trees) to include in the ensemble. Some benefits of tuning the "n_estimators" hyperparameter are as follows: (1) Improved Performance: Increasing the number of estimators often leads to improved predictive performance. (2) Reduced Overfitting: Ensemble methods tend to reduce overfitting by combining the predictions of multiple base models. Increasing the number of estimators can further help mitigate overfitting, especially when individual trees are prone to capturing noise in the data. (3) Stability and Consistency: When we increase the number of estimators, the ensemble's predictions tend to become more stable and consistent. This is because the random fluctuations or noise in individual tree predictions get averaged out as the ensemble size grows. (4) Bias-Variance Tradeoff: Increasing the number of estimators can help strike a balance between bias and variance. Initially, as we add more trees, the bias may decrease while variance increases. However, after a certain point, adding more trees continues to decrease variance without significantly increasing bias. (5) Robustness: A larger number of estimators can make the model more robust to outliers and noisy data. Since each tree focuses on different parts of the feature space, the ensemble becomes less sensitive to individual data points. (6) Feature Importance: Random Forest and Gradient Boosting algorithms can provide feature importance scores based on how often a feature is used across the ensemble. A higher number of estimators can provide more stable and reliable estimates of feature importance. (7) Grid Search Flexibility: Tuning the "n_estimators" hyperparameter provides an additional dimension for hyperparameter grid searches. This allows us to explore different trade-offs between computation time and model performance, giving our research more control over model selection.

The optimal value of the n_estimators hyperparameter depends on the specific dataset and task. In practice, it is common to start with a small number of estimators and gradually increase it until the model's capability to perform on the validation or test set starts to reach a plateau or decrease [22]. Cross-validation can also be used to tune the value of n_estimators and other hyperparameters. Overall, n_estimators is an important hyperparameter in Random Forest that affects the performance and complexity of the model [23]. In this experiment, choosing an appropriate value for this hyperparameter is a key step in building an accurate and efficient Random Forest model.

2.3 Mean Absolute Percentage Error (MAPE)

MAPE is an alternative way to evaluate the performance of regression models, with a straightforward interpretation based on relative errors. Its recommended use is for tasks where being able to detect changes in relative differences is more crucial than detecting changes in absolute values [24]. In the field of forecasting, there is always an element of unpredictability. This means that the predicted value will likely differ from the actual value, and this difference is referred to as the error. While it is impossible to eliminate this error, the goal of forecasting is to reduce it as much as possible. The Formula (2) can be used to perform MAPE calculation [25]:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} |PE_t| \qquad (2)$$

With details, n is a lot of data and t is a period. While the value of 〚PE〛 _t is produced from the Formula (3).

$$PE_t = \left(\frac{X_t - F_t}{X_t}\right) \text{ x } 100\% \quad (3)$$

The Mean Absolute Percentage Error (MAPE) is commonly utilized to determine the mean level of absolute error.

2.4 R-Squared (R2)

The R-squared or R2 is a mathematical indicator that quantifies how well the variation in the dependent variable is clarified by the independent variable in a regression model. It is also referred to as the coefficient of determination and shows how accurately the predicted values align with the actual data points on a regression line. The accuracy of the developed equation is evaluated by measuring the correlation between predicted and observed values [26]. The R2 value ranges between 0 and 1, with 0 indicating that the model is not a good fit for the given data, and 1 indicating that the model fits perfectly with the provided dataset. The R2 is calculated as [27]:

$$R = \sqrt{1 - \frac{\sum_{x=1}^{n} (V_{ot} - V_{op})^2}{\sum_{x=1}^{n} (V_{ot(mean)} - V_{op})^2}} \qquad (4)$$

The Formula (4) above uses n to represent the quantity of data points, while V_ot and V_op refer to the anticipated predictions derived from the regressors and the actual output measurements.

3. METHODOLOGY

3.1 R-Squared (R2)

This section will explain the general design of how this research is conducted. The workflow of this research is outlined in Figure 2, which depicts a series of steps comprising the experiment. The first stage is the collection of the dataset from the website www.bps.go.id. This dataset is the average price of rice at the wholesale level in Indonesia, covering the period from January 2010 to December 2022. This data has three attributes consisting of month, year, and rice prices with a total of 156 data [28]. Second, modeling is performed using the Random Forest method. Random Forest is a supervised machine learning algorithm that repeatedly applies the decision-tree concept to form a forest. This

Journal of Theoretical and Applied Information Technology
15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

algorithm combines predictions based on multiple decision trees [29].

Several experiments had previously been conducted using other methods such as SVR, Linear Regression, Decision Tree, and GBR with the same dataset. From the experiments conducted, it was concluded that the Random Forest method produced the highest accuracy and had the lowest error value. Third, performing hyperparameter tuning of the Random Forest. Tuning involves the process of finding the best hyperparameters for a learning algorithm concerning a particular dataset [30]. A hyperparameter can be the number of trees that should be included in the forest or the number of nodes that each tree can have. This can only be achieved through trial and error methods [31]. To find the optimal value for this hyperparameter tuning, various parameter values are tested. Finally, the results of the hyperparameter tuning are compared to choosing the best parameter based on its accuracy results.

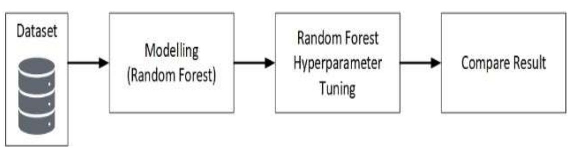


Figure 2: The Research Workflow

3.2 Datasets

This research evaluates the performance of the learning algorithm using the Average Rice Price dataset at the Indonesian Wholesale Trade Level from January 2010 to December 2022. The dataset is obtained from the Central Statistics Agency of Indonesia. The dataset used consists of time range data from January 2010 to December 2022, which includes 156 instances and 3 features. An example of the rice price dataset used is shown in Table 1 and Figure 3.

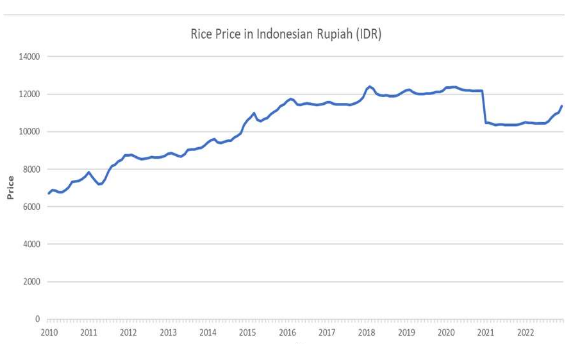


Figure 3: Rice Price Chart from January 2010 to December 2022

Table 1: Rice Price Dataset in Indonesian Rupiah (IDR)

Month	Year	Price
January	2010	6702
February	2010	6888
March	2010	6854
April	2010	6761
May	2010	6772
June	2010	6873
July	2010	7026
August	2010	7318
September	2010	7351
October	2010	7391
November	2010	7457
December	2010	7617
…..	…..	…..
…..	…..	…..
January	2022	10496
February	2022	10471
March	2022	10463
April	2022	10455
May	2022	10448
June	2022	10448
July	2022	10449
August	2022	10551
September	2022	10772
October	2022	10947
November	2022	11012
December	2022	11363

4. RESULT AND DISCUSSION

4.1 Experiment Result

At this stage, the results of the conducted experiment are explained. This research evaluates the performance of the learning algorithm using a dataset of rice prices. The first experiment involved comparing the error rates of previous research with several other methods. The data used is the price of rice from January 2010 to July 2020. The model used in the previous research by Anandyani was Arima Box Jenkins with a MAPE value of 0.87 [8].

Figure 4 shows the error rate results of the rice price dataset using various types of machine learning methods. The results of the experiment presented in Figure 4 used several machine learning methods such as Support Vector Regression (SVR), Linear Regression, Decision Tree, Gradient Boosting Regression (GBR), and Random Forest. Based on the results of the experiment, the smallest error value was obtained using the Random Forest method with a MAPE value of 0.0083385. In

Journal of Theoretical and Applied Information Technology
15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

conclusion, it can be inferred that the Random Forest method outperforms other methods, based on the obtained results of the experiment and previous studies using the same dataset.

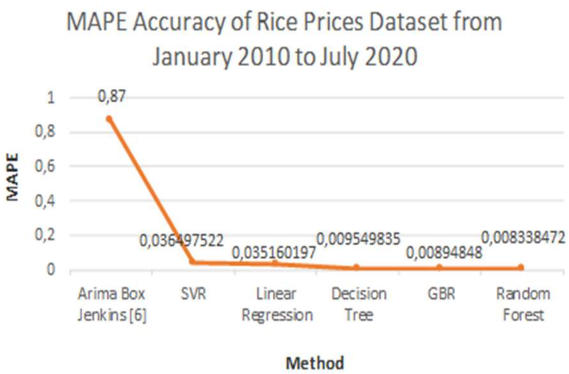


Figure 4: MAPE result using Rice Price Dataset from Januari 2010 to July 2020

The next experiment is to add new data using a rice price dataset with a time range from January 2010 to December 2022. The methods that will be compared are the same as the previous experiment, namely Support Vector Regression (SVR), Linear Regression, Decision Tree, Gradient Boosting Regression (GBR), and Random Forest. The values that will be compared are the accuracy values using R2 Score and the error values using MAPE. The accuracy results and error rates are displayed in Table 2.

Table 2: Accuracy and MAPE results using Rice Price Dataset from January 2010 to December 2022.

Method	MAPE	R2 Score
Linear Regression	0.0771416	0.6292070
SVR	0.0671858	0.4620580
Gradient Boosting Regression	0.0117876	0.9912672
Decision Tree	0.0097670	0.9907611
Random Forest	**0.0093573**	**0.9916805**

From the results of the experiment presented in Table 2, this can be inferred that Random Forest is the optimal method with the smallest MAPE value of 0.0093573 and the highest R2 Score value of 0.9916805. Therefore, it was decided in this research to create a model using the Random Forest method with a dataset of rice prices from January 2010 to December 2022. After being established using the Random Forest method, the next experiment is the adjustment of Random Forest's hyperparameter values to improve accuracy results and reduce error rates. The hyperparameter to be adjusted is the n_estimator parameter. Moreover, n_estimator is the number of trees with increasing gradients [32]. This experiment will attempt to change the value of the n_estimator parameter, and the results will be displayed in Table 3.

Table 3: Evaluation of hyperparameter tuning using data from January 2010 to December 2022.

Method	MAPE	R2 Score
Random Forest	0.0093573	0.9916805
Random Forest (n_estimator = 200)	0.0090941	0.9920341
Random Forest (n_estimator = 1000)	0.0089954	0.9919881
Random Forest (n_estimator = 500)	**0.0089389**	**0.9921578**

The experiment presented in Table 3 is a hyperparameter tuning experiment for Random Forest by trying several values. The first row represents the Random Forest method without the n_estimator hyperparameter. Next, hyperparameter tuning was attempted by inputting n_estimator values of 200, 1000, and 500. From the results displayed in Table 3, a Random Forest with a n_estimator value of 500 can minimize the MAPE value from 0.0093573 to 0.0089389 and increase the R2 Score value from 0.9916805 to 0.9921578. Therefore, it can be concluded that a Random Forest with a n_estimator value of 500 is the best method when compared to the other experiments. In another research, several values of n_estimator were also used, such as 50, 100, 200, and 500. The best model obtained also used n_estimator 500, resulting in an accuracy score of 0.9840 [33].

The next step is to perform an experiment on the Random Forest model with n_estimator 500 using several different datasets. The purpose of this experiment is to guarantee that the model created using Random Forest with n_estimator 500 can produce good accuracy even when using several different datasets. Table 4 describes the dataset descriptions that we use in our experiment.

Table 4: Dataset Descriptions

No	Dataset	Instance	Feature	Year
1	Rice Price Dataset [26]	156	3	2022
2	Power Consumption of Tetouan City Dataset [32]	52417	9	2021

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3	Combined Cycle Power Plant Dataset [33,34]	9568	4	201 4

Table 5: Evaluation of hyperparameter tuning using Power Consumption of Tetouan City Dataset [34].

Method	MAPE	R2 Score
Random Forest	0.0249722	0.9908976
Random Forest (n_estimator = 200)	0.0246895	0.9911778
Random Forest (n_estimator = 1000)	0.0246157	0.9911920
Random Forest (n_estimator = 500)	**0.0245714**	**0.9912360**

Table 5 shows the results of the hyperparameter tuning evaluation for the Tetouan City Electricity Consumption dataset. The dataset consists of 52417 instances and 9 features. It can be deduced from the results presented that the proposed model using Random Forest with n_estimator 500 produces the best MAPE and R2 Score values compared to other models. Further, Random Forest with n_estimator 500 can minimize the initial MAPE value of 0.0249722 to 0.0245714 and increase the R2 Score value from 0.9908976 to 0.9912360.

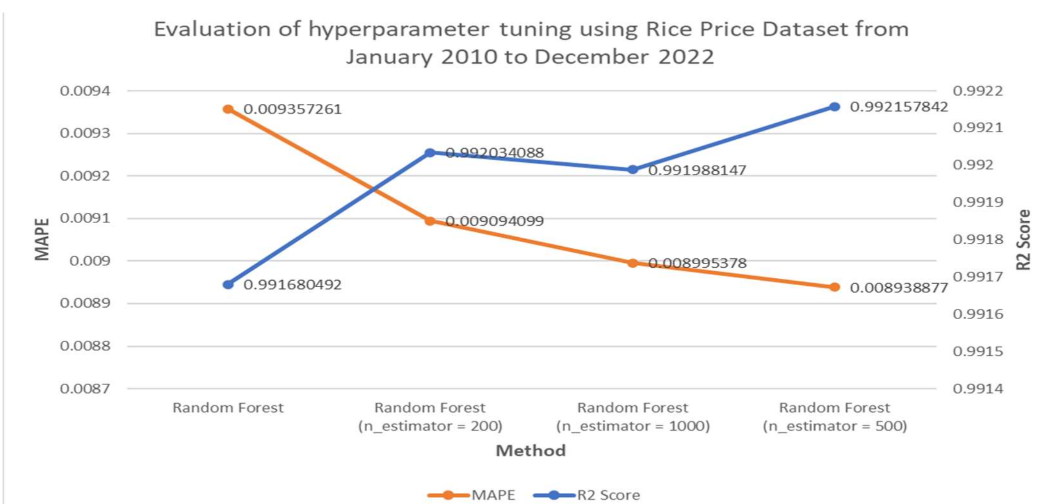
Table 6: Evaluation of hyperparameter tuning using Combined Cycle Power Plant Dataset [35,36].

Method	MAPE	R2 Score
Random Forest	0.0051784	0.9648999
Random Forest (n_estimator = 200)	0.0051621	0.9652738
Random Forest (n_estimator = 1000)	0.0051442	0.9654174
Random Forest (n_estimator = 500)	**0.0051439**	**0.9653407**

The evaluation results for hyperparameter tuning on the Combined Cycle Power Plant dataset are presented in Table 6. This dataset comprises 9568 instances and includes 4 distinct features. Upon a thorough examination of the outcomes, it becomes evident that the Random Forest model, with an n_estimator value of 500, consistently outperforms other models. Specifically, the Random Forest model with n_estimator set at 500 successfully reduces the initial MAPE value from 0.0051784 to 0.0051439, simultaneously elevating the R2 Score from 0.9648999 to 0.9653407.

In the broader context, a higher MAPE value implies a suboptimal performance, while a lower MAPE value is considered more favorable. Conversely, a higher R2 Score is generally preferable over a lower one. An R2 value of 1 signifies that the regression prediction fits the data exceptionally well, as noted by reference [37]. The findings of this research unequivocally demonstrate a consistent downward trend in MAPE values and a corresponding upward trend in R2 values across all datasets for each experiment conducted.

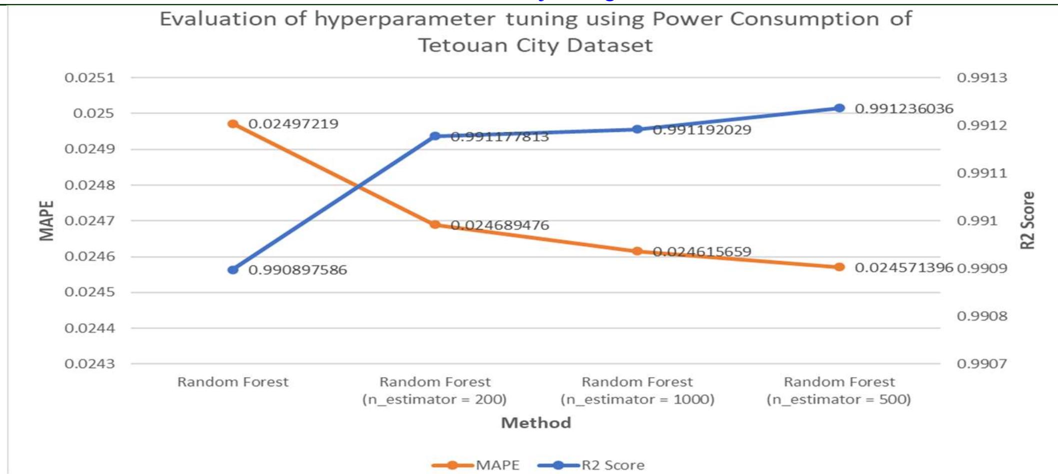
Based on the comprehensive evaluation results displayed in Figure 5, it is evident that the proposed model employing the Random Forest method with an n_estimator value of 500 consistently exhibits superior performance when compared to alternative methods across all datasets. The Random Forest method, with n_estimator set at 500, consistently excels in minimizing the MAPE value and maximizing the R2 Score value, thus confirming its superiority.



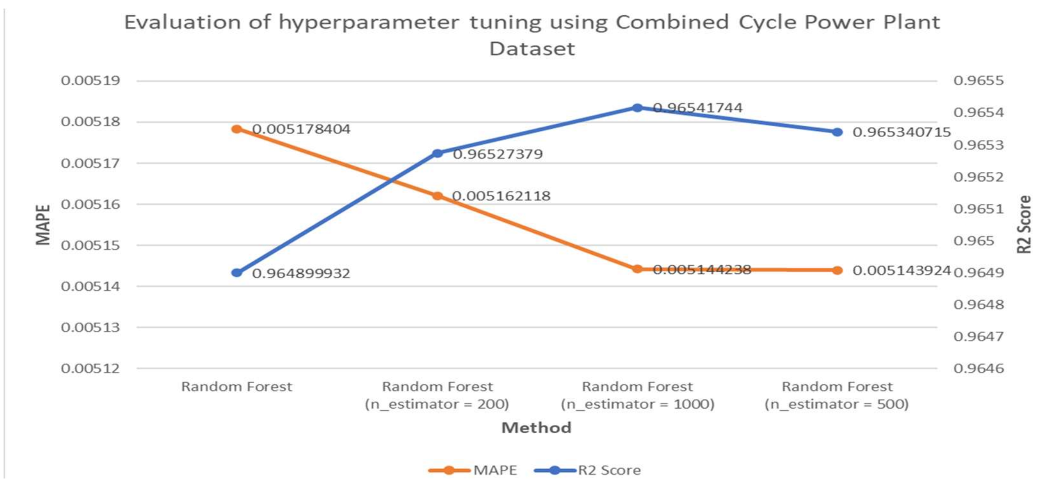
(a)

Journal of Theoretical and Applied Information Technology
15th December 2023. Vol.101. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(b)



(b)

Figure 5: Evaluation of hyperparameter tuning with (a) Rice Price Dataset, (b) Power Consumption of Tetouan Dataset, and (c) Combined Cycle Power Plant Dataset.

5. CONCLUSION

The purpose of this research is to try to improve the error rate from previous research by exploring some other methods. From the results of the experiments conducted, it can be inferred that the Random Forest method outperforms other methods based on the performance results. After determining to use the Random Forest method, the next experiment is to perform Random Forest hyperparameter tuning to improve accuracy results and reduce error rates. The hyperparameter used in this research is the n_estimator parameter. Based on the evaluation results of the hyperparameter tuning presented in Table 3, Table 5, and Table 6, the proposed model using Random Forest with n_estimators 500 was able to reduce the MAPE value and maximize the R2 Score value. For example, in Table 3 Random Forest with n_estimator 500 can reduce the MAPE value from 0.0093573 to 0.0089389 and increase the R2 Score value from 0.9916805 to 0.9921578. Model tuning is very important to perform because it can optimize the parameters to generate the best accuracy. The results of the simulation experiment prove the feasibility and accuracy of the proposed algorithm. While the current study focuses on the Random Forest method, future research could benefit from an expanded comparison with a more diverse set of machine learning models, including deep learning approaches and ensemble methods. This would contribute to a comprehensive understanding of the most suitable techniques for rice price prediction

REFERENCES:

[1] D. Dwipajati, E. Widajati, A.F. Ainaya, R.D. "Novanda, Potential of Indonesian Community Food Sources which are Rich in Fiber as an Alternative Staple Food for Type 2 Diabetics: A Scoping Review", *Open Access Macedonian Journal of Medical Sciences*, Vol. 10, 2022, pp. 47-53.

[2] A.D. Guritno, N.E. Kristanti, M.R. Tanuputri, "Collaborative Strategy for the Supply Chain of Rice: A Case Study on Demak and Sukoharjo Regency, Central Java, Indonesia", *AgriTECH*, Vol. 41, No. 1, 2021, pp. 1-7. https://doi.org/10.22146/agritech.48929.

[3] A. Shahzad, S. Ullah, A.A. Dar, M.F. Sardar, T. Mehmood, M.A. Tufail, A. Shakoor, M. Haris, "Nexus on climate change: agriculture and possible solution to cope future climate change stresses", *Environmental Science and Pollution Research*, Vol, 28 January 2021, 14211–14232. https://doi.org/10.1007/s11356-021-12649-8.

[4] S. Slameto, E. Jannah, M. Meidaliyantisyah, W. Wibawa, "Analysis of rice farming in tidal swampland, Tulang Bawang Regency, Lampung Province", *International Conference On Agriculture and Applied Science*, 2021. https://doi.org/10.25181/icoaas.v1i1.2061.

[5] F.N. RASHID, "IS INCREASING INPUT USE FOR RICE PRODUCTION A PROFITABLE PROPOSITION IN TANZANIA?", *Review of Agricultural and Applied Economics*, Vol. 23, No. 2. June 2020. pp. 54-63. https://doi.org/10.15414/raae.2020.23.02.54-63.

[6] H. J. Christanto, S. A. Sutresno, A. Denny and C. Dewi, "USABILITY ANALYSIS OF HUMAN COMPUTER INTERACTION IN GOOGLE CLASSROOM AND MICROSOFT TEAMS", *Journal of Theoretical and Applied Information Technology*, Vol. 101, No. 16. August 2023. pp. 6425-6436.

[7] H. J. Christanto, S. A. Sutresno, V.S. Simi, C. Dewi and G. Dai, "ANALYSIS OF GAME THEORY IN MARKETING STRATEGIES OF TIKTOK AND INSTAGRAM", *Journal of Theoretical and Applied Information Technology*, Vol. 101, No. 22. November 2023. pp. 7100-7109.

[8] A.R. Anandyani, D.K.A. Astutik, N. Bariroh, A. Indrasetianigsih, "Prediction of the Average Price of Rice Sold by Wholesalers Using the ARIMA Box-Jenkins Method", *Teknosains Media Informasi Sains Dan Teknologi*, Vol. 15, No. 2, August 2021. pp. 151-160. https://doi.org/10.24252/teknosains.v15i2.17721.

[9] S. Mehtab, J. Sen, "A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing", *SSRN*,2020. https://doi.org/10.2139/ssrn.3502624.

[10] A. Ghimire, S. Thapa, A.K. Jha, S. Adhikari, A. Kumar, "Accelerating Business Growth with Big Data and Artificial Intelligence", *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC),* Institute of Electrical and Electronics Engineers Inc., Palladam, India, 2020, pp. 441–448.https://doi.org/10.1109/I-SMAC49090.2020.9243318.

[11] M.H.L. Louk, B.A. Tama, "Tree-Based Classifier Ensembles for PE Malware Analysis: A Performance Revisit", *Algorithms*, Vol.15, No. 9, September 2022. pp. 332, https://doi.org/10.3390/a15090332.

[12] M. Rizki, A. Wenda, F.D. Pahlevi, M.I.H. Umam, M.L. Hamzah, S. Sutoyo, "Comparison of Four Time Series Forecasting Methods for Coal Material Supplies: Case Study of a Power Plant in Indonesia", *2021 International Congress of Advanced Technology and Engineering (ICOTEN),* Institute of Electrical and Electronics Engineers Inc., Yemen, 2021. https://doi.org/10.1109/ICOTEN52080.2021.9493522.

[13] M. Hosoe, M. Kuwano, T. Moriyama, "A method for extracting travel patterns using data polishing", *Journal of Big Data*, Vol.8, No. 13, 2021, pp. 1-22. https://doi.org/10.1186/s40537-020-00402-w.

[14] A. Thakkar, K. Chaudhari, "Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions", *Information Fusion*, Vol. 65, January 2021, pp. 95–107. https://doi.org/10.1016/j.inffus.2020.08.019.

[15] S. Ren, H.L. Chan, T. Siqin, "Demand forecasting in retail operations for fashionable products: methods, practices, and real case study", *Annals of Operations Research*, Vol. 291, August 2020, pp. 761–777. https://doi.org/10.1007/s10479-019-03148-8.

[16] M. Schonlau, R.Y. Zou, "The random forest algorithm for statistical learning", *The Stata Journal*, Vol. 20, No. 1, 2020, pp. 3–29. https://doi.org/10.1177/1536867X20909688.

[17] I.H. Sarker, A. Colman, J. Han, A.I. Khan, Y.B. Abushark, K. Salah, "BehavDT: A Behavioral Decision Tree Learning to Build User-Centric

Context-Aware Predictive Model", *Mobile Networks and Applications*, Vol. 25, 2020, pp. 1151–1161. https://doi.org/10.1007/s11036-019-01443-z.

[18] A. Adebowale, S. a Idowu, A.A. A, Comparative Study of Selected Data Mining Algorithms Used For Intrusion Detection, *International Journal of Soft Computing and Engineering (IJSCE)*, Vol. 3, No. 3, July 2013, pp. 237-241.

[19] E.S. Lestari, I. Astuti, F. Ilmu, T. Informasi, U. Gunadarma, J.M. Raya, J. Barat, "Penerapan Random Forest Regression Untuk Memprediksi Harga Jual Rumah Dan Cosine Similarity Untuk Rekomendasi Rumah Pada Provinsi Jawa Barat", *Jurnal Ilmiah FIFO*, Vol. 14, No. 2, November 2022, pp 131–146. http://dx.doi.org/10.22441/fifo.2022.v14i2.003

[20] M.J. Jun, "A comparison of a gradient boosting decision tree, random forests, and artificial neural networks to model urban land use changes: the case of the Seoul metropolitan area", *International Journal of Geographical Information Science,* Vol. 35, No. 11, 2021, pp. 2149–2167. https://doi.org/10.1080/13658816.2021.1887490.

[21] Z. Tian, J. Xiao, H. Feng, Y. Wei, "Credit Risk Assessment based on Gradient Boosting Decision Tree", *Procedia Computer Science*, Vol. 174, July 2020, pp. 150–160. https://doi.org/10.1016/j.procs.2020.06.070.

[22] P. Probst, M.N. Wright, A.L. Boulesteix, "Hyperparameters and tuning strategies for random forest", *WIREs Data Mining and Knowledge Discovery*, Vol. 9, No. 3, 2019. https://doi.org/10.1002/widm.1301.

[23] P. Contreras, J. Orellana-Alvear, P. Muñoz, J. Bendix, R. Célleri, "Influence of Random Forest Hyperparameterization on Short-Term Runoff Forecasting in an Andean Mountain Catchment", *Atmosphere,* Vol. 12, No. 2, February 2021, pp. 238. https://doi.org/10.3390/atmos12020238.

[24] D. Chicco, M.J. Warrens, G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation", *PeerJ Computer Science*, Vol. 7, 2021, pp. 1–24. https://doi.org/10.7717/PEERJ-CS.623.

[25] F. Liantoni, A. Agusti, "Forecasting Bitcoin using Double Exponential Smoothing Method Based on Mean Absolute Percentage Error", *International Journal on Informatics Visualization*, Vol. 4, No. 2, 2020, pp. 91–95. https://doi.org/10.30630/joiv.4.2.335.

[26] A. Cominotte, A.F.A. Fernandes, J.R.R. Dorea, G.J.M. Rosa, M.M. Ladeira, E.H.C.B. van Cleef, G.L. Pereira, W.A. Baldassini, O.R. Machado Neto, "Automated computer vision system to predict body weight and average daily gain in beef cattle during growing and finishing phases", *Livestock Science*, Vol. 232, 2020. https://doi.org/10.1016/j.livsci.2019.103904.

[27] A. Afzal, S. Alshahrani, A. Alrobaian, A. Buradi, S.A. Khan, "Power Plant Energy Predictions Based on Thermal Factors Using Ridge and Support Vector Regressor Algorithms", *Energies,* Vol. 14, No. 21, 2021. https://doi.org/10.3390/en14217254.

[28] BPS - Statistics Indonesia, The Average Price of Rice at the Wholesale Trade Level in Indonesia, (2022). https://www.bps.go.id/indicator/20/295/2/rata-rata-harga-beras-di-tingkat-perdagangan-besar-grosir-indonesia.html (accessed February 17, 2023).

[29] D. Ganatra, "Ensemble Methods to Improve Accuracy of a Classifier", *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 9, No. 3, 2020, pp. 3434-3439. https://doi.org/10.30534/ijatcse/2020/145932020.

[30] L. Amusa, D. North, T. Zewotir, "Optimal hyperparameter tuning of random forests for estimating causal treatment effects", *Songklanakarin J. Sci. Technol*, Vol. 43, No. 4, 2021, pp. 1004-1009.

[31] C.G. Siji George, B. Sumathi, "Grid Search Tuning of Hyperparameters in Random Forest Classifier for Customer Feedback Sentiment Prediction", *International Journal of Advanced Computer Science and Applications(IJACSA),* Vol. 11, No. 9, 2020, pp. 173–178. https://doi.org/10.14569/IJACSA.2020.0110920

[32] S.H. Encinas, J. Cruz, N.J. Beltran, F. Pineda, C. Romero, J.F.C. Acero, W.M. Machaca, "Modeling And Prediction Of A Multivariate Photovoltaic System, Using The Multiparametric Regression Model With Shrinkage Regularization And Extreme Gradient Boosting", *LACCEI International Multi-Conference for Engineering, Education, and Technology*, Argentina, July 2021. https://doi.org/10.18687/LACCEI2021.1.1.557.

[33] I. Agustina, Y. Mulyani, T. Septiana, M. Mardiana, "ANALISIS PENGEMBANGAN MODEL PREDIKSI KESUKSESAN KICKSTARTER MENGGUNAKAN ALGORITMA BACKPROPAGATION DAN

RANDOM FOREST", *JITET (Jurnal Informatika dan Teknik Elektro Terapan).* Vol. 10, No. 3, 2022, pp. 173–182. https://doi.org/10.23960/jitet

[34] A. Salam, A. El Hibaoui, "Comparison of Machine Learning Algorithms for the Power Consumption Prediction : - Case Study of Tetouan city -,", *2018 6th International Renewable and Sustainable Energy Conference (IRSEC*), Institute of Electrical and Electronics Engineers Inc., Morocco, 2018. pp. 1-5 https://doi.org/10.1109/IRSEC.2018.8703007.

[35] P. Tüfekci, "Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods", *International Journal of Electrical Power & Energy Systems*, Vol. 60, 2014, pp. 126–140. https://doi.org/10.1016/j.ijepes.2014.02.027.

[36] H. Kaya, P. Tüfekci, S.F. Gürgen, "Local and Global Learning Methods for Predicting Power of a Combined Gas & Steam Turbine", *International Conference on Emerging Trends in Computer and Electronics Engineering (ICETCEE'2012)*, Dubai, UAE, 2012, pp. 13–18.

[37] C. Dewi, R.C. Chen, "Random forest and support vector machine on features selection for regression analysis", *International Journal of Innovative Computing, Information and Control,* Vol. 15, No. 6, 2019, pp. 2027–2037. https://doi.org/10.24507/ijicic.15.06.2027.