Comparative Study of Statistical, Machine Learning, and Deep Learning for Rice Retail Price Forecasting in West Java

I Nyoman Gde Artadana Mahaputra Wardhiana
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
6025231022@student.its.ac.id

Muhammad Zakky Ghufron
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
6025231089@student.its.ac.id

Riyanarto Sarno
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@if.its.ac.id

Agus Tri Haryono
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
7025231020@student.its.ac.id

Shoffi Izza Sabilla
Department of Medical Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
shoffi@its.ac.id

Fadlilatul Taufany
Department of Chemical Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
f_taufany@chem-eng.its.ac.id

Sholiq
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
sholiq@its.ac.id

Kelly Rossa Sungkono
Department of Informatics Engineering
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
kelly@its.ac.id

***Abstract*— This research investigates the effectiveness of machine learning (ML), deep learning (DL), and statistical algorithms for forecasting rice retail prices in West Java, Indonesia. Employing data from 2021 to 2023, this study evaluates Random Forest, LightGBM, XGBoost, LSTM, and SARIMAX models. ML algorithms displayed high error rates with RMSE values exceeding 1100, indicating challenges with complex datasets. In contrast, LSTM models demonstrated robust performance with a substantially lower RMSE of 129.67, but LSTM model's low error rates was achieved by not involving other exogenous variables, focusing solely on historical price data. The SARIMAX model, while less effective in direct forecasting with an RMSE of 799.82, effectively utilized exogenous variables to encapsulate seasonal trends, offering valuable insights into supply chain dynamics affecting rice prices.**

Keywords—LightGBM, LSTM, Random Forest, Rice Price Forecasting, SARIMAX, XGBoost.

I. Introduction

In the intricate landscape of agricultural economics, forecasting prices of staple commodities like rice hold significant implications in decision-making and planning for the government or communities like farmers, traders, and consumers [1], [2]. The province of West Java in Indonesia serves as a significant illustration, notably due to its central involvement in both the production and distribution of rice in Indonesia. According to the Badan Pusat Statistik (BPS), West Java province produced 5.25 million tons of rice in 2023, representing Indonesia's second-largest rice production volume [3]. The rice commodity's supply chain management intricacies further accentuate the necessity for accurate price forecasting. Generally, the farmer and consumer production and distribution process involves various intermediaries, including millers, wholesalers, and retailers. These entities are connected and form a commodity price trajectory [4]. Understanding these complexities underscores the importance of employing robust forecasting models to navigate the terrain of rice markets.

This research explores the efficacy of machine learning (ML), deep learning (DL), and statistical algorithms for forecasting rice retail prices in West Java. With its versatile algorithms capable of extracting patterns from data, machine learning offers a robust forecasting method [5]. This research conducted Random Forest, Light-Gradient Boosting Machine (LightGBM), and Extreme Gradient Boosting (XGBoost) within the ML domain, owing to the capability to handle complex datasets, nonlinear relationships, and high-dimensional feature spaces. Previous research used several machine learning algorithms for product sales forecasting. It produced the decision tree models as the best performance in seasonal data, considering this approach enhances robustness against overfitting, making it an ideal candidate for this dataset research [6]. Deep learning, particularly Long Short-Term Memory (LSTM) networks, emerges as a potent tool in time series forecasting [7]. LSTM's architecture, with its memory cells and gates, enables the model's capability of earlier stages of value remembering and capturing temporal dynamics inherent in historical data. So, the values can be used in the future [8]. This capability is precious in capturing nuanced patterns and irregularities in the agricultural market. Traditional statistical algorithms offer a structured framework for understanding time series dynamics. While Autoregressive Integrated Moving Average (ARIMA) models are commonly used for time series forecasting, this research opts for Seasonal Auto-Regressive Integrated Moving Average with Exogenous Variable (SARIMAX) due to its ability to handle seasonality and trend components more effectively, aligning better with the cyclical nature of rice commodity's data [9].

At its core, this research is driven by two primary research questions. The first question compares ML, DL, and statistical algorithms to determine which method best forecasts rice prices in West Java using the available dataset with objective methodologies' predictive capabilities, computational efficiency, and interpretability to identify the most effective approach. The second research question focuses on analyzing the supply chain management of rice commodities in West

2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA) | 979-8-3503-5161-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICTIIA61827.2024.10761587

979-8-3503-5161-3/24/$31.00 ©2024 IEEE

Java. It aims to discern the influence of various elements within the supply chain, such as factors from the farmer, industrial, and retail levels on rice prices. All relevant variables related to the supply chain are considered independent, except for the rice price, which is the dependent variable. This comprehensive analysis intends to shed light on the underlying drivers of price fluctuations and provide stakeholders with insights that support informed decision-making and the development of robust predictive models tailored to the specific intricacies of agricultural markets.

II. Related Work

Technological advancements have sparked a notable interest in leveraging several methodologies like statistical methods, machine learning, and deep learning for predicting crop and rice yields. Assessing current yield production research is pivotal for various agricultural aspects such as market planning, insurance, harvest management, and optimizing nutrient strategies while exploring potential crop productivity [10]. Thus, many studies have focused on rice forecasting using various approaches, including machine learning or deep learning. This section discussed several studies related to this scope.

Several studies conducted forecasting using deep learning models [11]. For instance, [12] research focused on using deep learning models to monitor rice production in Indonesia. This research applied a comparison between three deep learning models: Convolutional Neural Network (CNN), Long-Short Term Memory (LSTM), and Recurrent Neural Network (RNN). From the result of the study, the LSTM method was outperformed in rice crop prediction with an R-squared score of 0.79, followed by CNN and RNN.

Instead of using a deep learning model, [13] conducted a regression technique using Support Vector Machine, Linear Regression, and XGBoost. This paper used a dataset from Indonesia rice production over 34 provinces between 2018 and 2022. In this paper, XGBoost outperformed all models with

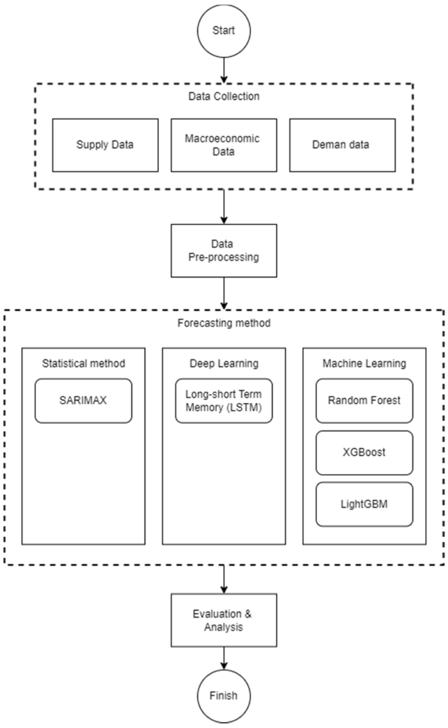

Fig. 1. Proposed method

TABLE I. Data Description

Name	Description	Code
Supply data		
Paddy Harvest Area	Rice paddy field area, percentage (%) of rice harvest area	a
Price of Harvested Dry Unhusked Rice, Farmer-level	Price of harvested dry unhusked rice at the farmer level. (Rp/Kg)	b
Price of Harvested Dry Unhusked Rice, Milling-level	Price of harvested dry unhusked rice at the milling level. (Rp/Kg)	c
Price of Milled Dry Unhusked Rice, Milling-level	Price of milled dry unhusked rice at the milling level. (Rp/Kg)	d
Milled Rice Price	The average of the price of medium and premium milled rice. (Rp/Kg)	e
Rice Production Level	Rice production in west java 2021-2023. (unit tons)	f
Macroeconomic data		
Rice Price	The average rice price of lower quality rice I & II, medium quality rice I & II, super quality rice I & II at traditional market level. (Rp/Kg)	g
Provincial Minimum Wage	Regional Minimum Wage in West Java from 2021-2023 (Represents labor cost)	h
Solar Fuel Price	Solar price from 2021-2023 (Representing distributor entities)	i
Demand data		
Rice Consumption Level	The level of rice consumption in West Java 2021-2023. (unit tons)	j

0.015 in Root Mean Squared Error (RMSE) and a Mean Absolute Error (MAE) score of 0.008. However, that paper did not perform inverse scaling, which resulted in the acquisition of a very small error number, so that the results could be inaccurate. In a similar study, [14] proposed an ensembled voting regression method combining Gradient Boosting Regressor (GBR), XGBoost, and Category Boosting (CatBoost) models. This method achieved the highest R-squared score with less error.

Meanwhile, statistical models can also be used as a forecasting algorithm. [15] proposed comparative study of two hybrid deep learning-statistical models, Neural Networks-Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX) and NNs-Generalized Space-Time Autoregressive Integrated Moving Average with Exogenous Variable (GSTARIMAX). This study's results showed that NNs-GSTARIMAX accuracy was better than NNs-ARIMAX. Therefore, this research proposed comparing several machine learning, deep learning, and statistical methods in rice retail price forecasting in West Java.

III. Methodology

This research compares machine learning, deep learning and statistical algorithms to predict rice retail prices in West Java. This section discussed the research workflow illustrated in Fig. 1.

A. Data Collection

The dataset covers rice commodities in West Java by 2021-2023, consisting of 10 features from different categories: supply-related, macroeconomic factors, and demand-related, which can be seen in Table I. This dataset was taken from BPN (Badan Pangan Nasional), BI PIHPS (Bank Indonesia Pusat Informasi Harga Pangan Strategis Nasional), and BPS (Badan Pusat Statistik). The dataset is a time series of rice commodity developments per day. Rice prices are derived from the traditional market level, while other features are from the producer level.

B. Data Pre-processing

After collecting the dataset, the next step is to carry out pre-processing. A series of pre-processing steps have been conducted to prepare the dataset for analysis. These steps include removing rows containing NaN or empty values; all values in the dataset have been converted to integer data type to suit the analysis needs. Calculating the Pearson correlation helps to understand the relationships between variables and identify patterns that may be hidden in the data. By performing these pre-processing steps, the data used in the analysis is of good quality and ready for further steps. The dataset was split into training and testing sets using an 80:20 ratio for all models, with the training set containing 80% of the data used to train the models and the testing set containing 20% of the data used to evaluate model performance.

This research will use GridSearchCV to search for the ideal machine learning model parameters, Adam optimizer for LSTM models, and grid search to get the lowest Akaike Information Criterion (AIC) for SARIMAX. The dataset was normalized using MinMaxScaler specifically for LSTM models, a step required to maximize model performance. The data were then renormalized during the examination.

C. Forecasting Method

In this research, various ML, DL, and Statistical models are utilized, each selected for its unique capabilities in handling different aspects of time series data. Random Forest, LightGBM, and XGBoost represent ML methods. The Random Forest was an ensemble technique that generates multiple decision trees by randomly selecting subsets of training samples and features to improve predictive accuracy and control over-fitting [16]. It is effective for large datasets and can handle nonlinear data through its ensemble of simple models [17].

The LightGBM was introduced by Ke et al. [18]. It *i*s a gradient-boosting framework that uses tree-based learning algorithms. It is designed for distributed and efficient training, resolving the issue of a large data size and high feature dimension [19]. The XGBoost implements gradient-boosted decision trees designed for speed and performance. The training and predictive results of the preceding decision tree influence the input data for each subsequent decision tree [20]. It is a highly flexible and versatile tool due to its robust handling of various data types and relationships.

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) capable of performing the same operation on all elements in the sequence [21]. The LSTM model was chosen for its ability to handle short-term and long-term data through gating, mitigate the vanishing gradient problem, and efficiently manage variable-length input sequences [22]. This model is particularly well-suited for time series data where long-term context is important.

In the statistical algorithm, SARIMAX (Seasonal Auto Regressive Integrated Moving Average with Exogenous Variable) was selected for its ability to capture complex patterns in time series data and provide accurate forecasts for future values [9]. SARIMAX is an extension of SARIMA for univariate time series forecasting by incorporating exogenous variables, trends, and seasonal effects [23], making it ideal for handling complex seasonal patterns seen in agricultural data like rice prices.

D. Evaluation

The evaluation of this study is assessing the accuracy and efficiency of the models used to predict future rice prices. The primary metrics utilized for this evaluation are Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These metrics help quantify the difference between the predicted values by the models and the actual values observed. MSE focuses on the average of the squares of errors, providing a measure of the quality of an estimator—it is always non-negative, and values closer to zero are good. RMSE, the square root of MSE, gives a clearer idea of the error magnitude between model's output and target value, making interpretation straightforward. MAE measures the average magnitude of the prediction errors, directly providing the average error in the same units as the data. It measures how accurate the predicted values are compared to the observed ones [24].

IV. Result and Discussion

A. Dataset Analysis

The dataset used in this study consists of 678 rows and 10 columns, Fig. 2 displayed multiple and complex time series, each representing different variables within the rice commodity's supply chain management in West Java. Each

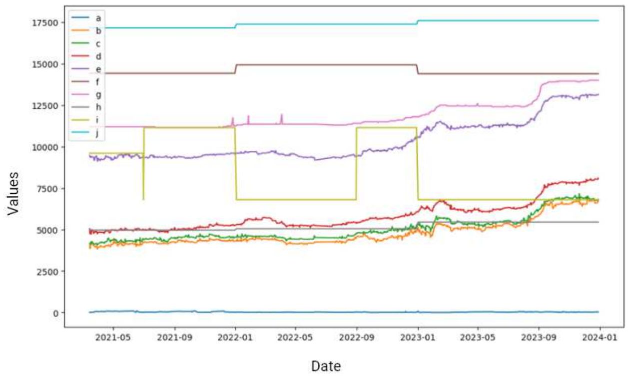

Fig. 2. Dataset visualization

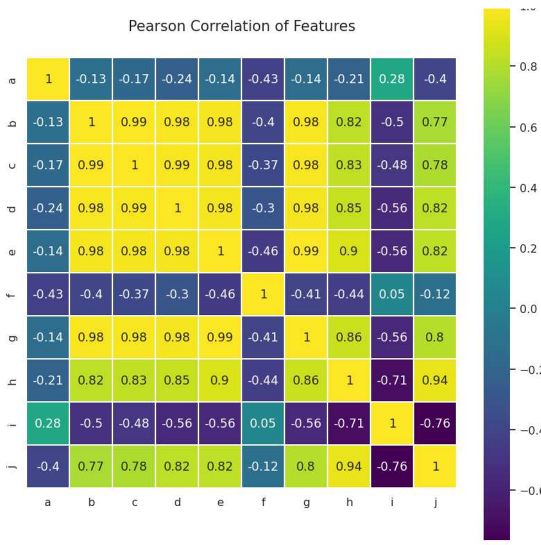

Fig. 3. Pearson correlation

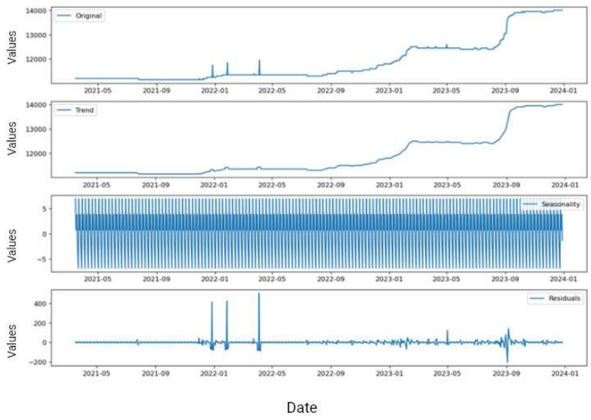

Fig. 4. Trend, seasonality, and residuals

colored line represent variables from Table I, illustrating trends from May 2021 to January 2024. Notably, some variables shown here are based on assumptions (the flat lines), which are common when actual data are hard to obtain or estimate. These include estimated rice consumption levels or projected production outputs.

The Pearson correlation matrix displays the linear relationships between various variables related to rice commodity's supply chain management in West Java. The matrix in Fig. 3 highlights strong positive correlations among variables directly involved in rice pricing and production stages, such as price of harvested dry grain, price of milled dry grain, and milled rice price, indicating a tight linkage where changes in the others closely mirror changes in one variable. Notably, economic factors like provincial minimum wage and solar fuel price show varying degrees of negative correlations with some variables, suggesting that lower fuel prices might be associated with higher rice prices and wages. The correlation of time-related variables (Year and Month) with others indicates potential seasonal or annual patterns affecting these metrics.

Characteristics of the dataset is shown on Fig. 4, displays decomposition with four graphs: original, trend, seasonality, and residual components of rice price. The "Original" graph shows the raw data, which exhibits an increasing trend over time with some notable spikes. The "Trend" graph emphasizes a smooth, gradual increase, suggesting a consistent upward movement in the data across the observed period. The "Seasonality" graph reveals a very regular, predictable pattern that repeats frequently, indicating a strong seasonal influence. Finally, the "Residuals" graph, which is mostly stable with a few spikes, shows the deviations from the modeled trend and seasonality, capturing the unpredicted or random fluctuations in the data.

B. Machine Learning Results

As this research focuses on forecasting rice prices, the dependent variable is the rice price itself, while other features are considered independent variables. All available features were utilized to discern which variables have the most significant impact on the fluctuations in rice prices. The evaluation process involved splitting the data into training and testing sets using an 80:20 ratio for all ML models. Hyperparameter tuning was conducted on all machine learning models using GridSearchCV to obtain the optimal score with optimum parameters. Table III shows the hyperparameter used in each ML model.

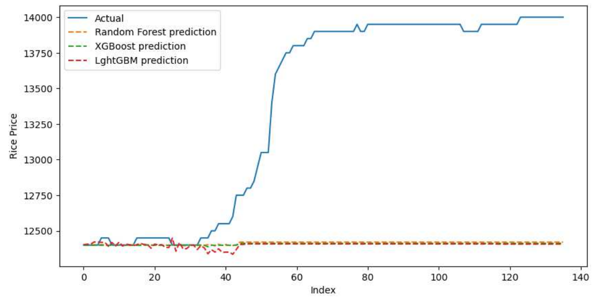

Fig. 5. Machine Learning prediction results

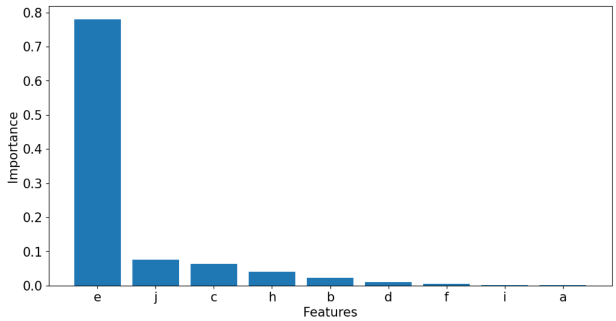

Fig. 6. Feature importance results

Based on the model's results in Table II, it is noted that these machine learning models showed relatively poor performance across all the evaluation metrics used (MSE, RMSE, and MAE). This observation suggests that despite its robustness in handling large and complex datasets typically, based on Fig. 5, Random Forest, XGBoost, and LightGBM did not adapt well to the specific nuances of this dataset. This underperformance could be attributed to several factors inherent to the model and the nature of the data.

The poor results led to further investigation by removing highly correlated features to reduce redundancy and potential noise. However, this approach only led to a slight reduction in error rates, indicating that the fundamental challenges with the model's fit to this dataset were not solely due to feature redundancy.

Despite the poor performance, Fig. 6 showed that the feature importance measures reported by the Random Forest model reflect similar relationships to those observed in the Pearson correlation matrix. It suggests that the model recognizes the same influential factors but perhaps struggles due to its non-temporal nature and inability to capture

TABLE III. COMPARATIVE RESULTS

Models	***Evaluatioin metrics***		
	MSE	***RMSE***	***MAE***
SARIMAX	639714.51	799.82	650.35
RF	1384875.01	1176.80	958.96
XGBoost	1389609.34	1178.81	961.11
GBM	1406810.24	1186.09	971.82
LSTM	**16816.55**	**129.67**	**103.94**

TABLE II. MACHINE LEARNING PARAMETERS

ML models	***Parameters***
RF	max_depth: 10, min_samples_split: 2, n_estimator: 50
XGBoost	learning_rate: 0.05, max_depth: 3, n_estimators: 100
LightGBM	learning_rate: 0.2, max_depth: 3, n_estimators: 100

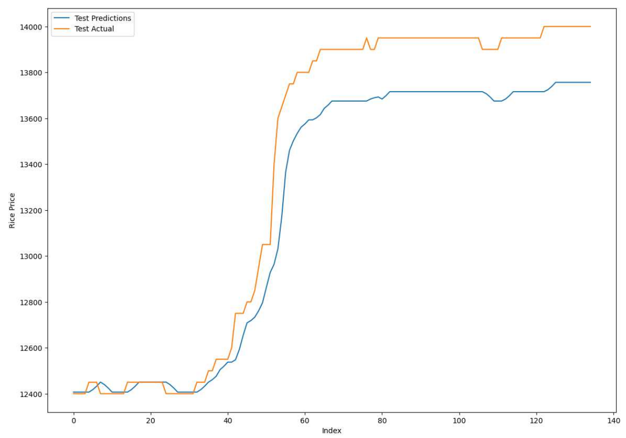

Fig. 7. LSTM prediction result

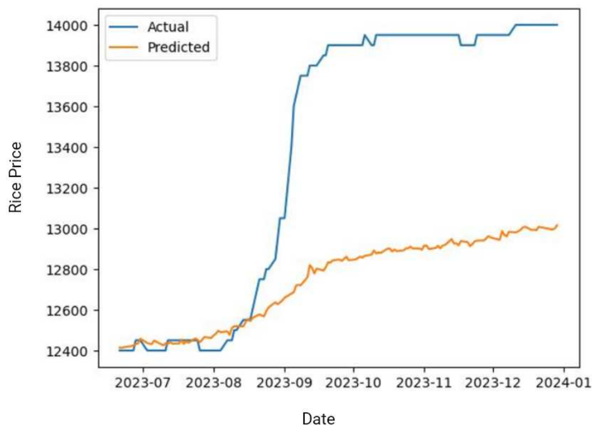

Fig. 8. SARIMAX prediction result

complex, nonlinear interactions effectively. This congruence would mean that while the model did not predict well, it still recognized the key variables impacting rice prices. However, Random Forest is not a sequential model like SARIMAX or LSTM. It cannot forecast future time steps based on past data.

The study found that machine learning models like Random Forest, LightGBM, and XGBoost faced challenges in forecasting rice prices due to high error metrics (MSE, RMSE, and MAE). Thus, ML models cannot answer Research Question 1. While ML models showed poor performance in forecasting prices directly, their ability to analyze feature importance and their robustness in handling complex datasets may allow them to provide insights into supply chain influences on prices. Therefore, ML can contribute to answering Research Question 2.

C. *LSTM Results*

A multivariate LSTM model, designed to incorporate multiple input variables, was employed to predict rice prices. However, the performance of the multivariate LSTM model was found to be suboptimal. LSTM struggles with multivariate time-series data due to this dataset's complexity and high dimensionality, which contains intricate, nonlinear interdependencies among variables that are difficult to model.

Hence, historical data on rice prices only was used to predict rice prices. In this stage, the normalized data was split into 80% training and 20% testing data. This model was trained using an Adam optimizer, with a learning rate 0.0001, relu activation, and 100 epochs. Fig. 7 shows the visualization of the prediction results using the LSTM method. Based on Table II, this deep learning model carried out good predictions and resulted in a better error rate than machine learning models.

The LSTM model showcased its strength in understanding long-term data dependencies, achieving significantly lower MAE scores compared to other models. This suggests that LSTM can answer Research Question 1 effectively. Although LSTM excels at capturing temporal patterns in price data, its ability to directly analyze complex interdependencies in supply chain factors might not be as robust without modifications or integrations with other models. It might not effectively answer Research Question 2 alone.

D. *SARIMAX Results*

After employing a grid search to optimize model parameters, the best SARIMAX configuration identified is SARIMAX (0, 1, 1) x (0, 1, 1, 12), achieving the lowest AIC value of 5528.86.

With test data of 20%, the SARIMAX model performs better than the machine learning model but still worse than the deep learning model regarding the errors' magnitude and consistency. The MSE, RMSE, and MAE values are in Table II. The actual values show a sharp rise around November 2023, which then levels off, while the predicted values increase gradually but fail to capture this abrupt spike. That indicates that while the SARIMAX model tracks the overall trend, it does not accurately predict sudden changes in the data, suggesting that the model does not fit the data, with reasonable errors in the context of rice price forecasting. Fig. 8 shows the visualization of the predicted data.

This model also performed reasonably well, although not as effectively as LSTM, in forecasting rice prices, capturing complex seasonal patterns. Hence, SARIMAX can answer Research Question 1. With its capacity to incorporate exogenous variables, SARIMAX can analyze how different supply chain elements influence rice prices, making it suitable to answer Research Question 2.

V. Conclusion

The research conducted provides a detailed comparison of various forecasting models, machine learning (ML), Long Short-Term Memory (LSTM), and Seasonal Auto Regressive Integrated Moving Average with Exogenous Variable (SARIMAX) in their efficacy to predict rice prices in West Java and analyze the supply chain management impacts on these prices. Results indicate that LSTM models capture long-term dependencies in price data, making them highly effective for direct price forecasting. However, they fall short of directly analyzing the impact of supply chain factors. In contrast, despite their lower performance in forecasting prices, ML models can still offer insights into supply chain factors due to their ability to handle complex datasets and analyze feature importance. Although SARIMAX can incorporate exogenous variables, which is beneficial for understanding the impacts on rice prices, it does not match LSTM's precision in price prediction, but this model addresses both research questions.

Future research should explore the integration of additional exogenous variables, such as weather patterns, political changes, and economic indicators that significantly impact agricultural outputs and market prices. Experimenting with hybrid models that combine the strengths of various forecasting approaches, like merging LSTM's temporal accuracy with SARIMAX's seasonal adjustability, could yield more robust predictions.

ACKNOWLEDGMENT

This research was funded by Institut Teknologi Sepuluh Nopember (ITS) under Penelitian Dana Departemen Program, Penelitian Keilmuan, Penelitian Flagship, Penelitian Kolaborasi Pusat and under the Project Scheme of the Publication Writing and Intellectual Property Rights (IPR) Incentive (Penulisan Publikasi dan Hak Kekayaan Intelektual/PPHKI) Program 2024.

REFERENCES

[1] B. Jokonowo, J. Claes, R. Sarno, and S. Rochimah, “Process Mining in Supply Chains: A Systematic Literature Review,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, p. 4626, May 2018, doi: 10.11591/ijece.v8i6.pp4626-4636.

[2] V. C. Sugiarto, R. Sarno, and D. Sunaryono, “Sales forecasting using Holt-Winters in Enterprise Resource Planning at sales and distribution module,” in *2016 International Conference on Information & Communication Technology and Systems (ICTS)*, 2016, pp. 8–13. doi: 10.1109/ICTS.2016.7910264.

[3] “Badan Pusat Statistik Provinsi Jawa Barat.” Accessed: May 12, 2024. [Online]. Available: https://jabar.bps.go.id/indicator/53/713/1/produksi-beras-menurut-kabupaten-kota.html

[4] P. A. Wibowo Putro, E. K. Purwaningsih, D. I. Sensuse, R. R. Suryono, and Kautsarina, “Model and implementation of rice supply chain management: A literature review,” *Procedia Comput Sci*, vol. 197, pp. 453–460, 2022, doi: https://doi.org/10.1016/j.procs.2021.12.161.

[5] M. M. Taye, “Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions,” *Computers*, vol. 12, no. 5, 2023, doi: 10.3390/computers12050091.

[6] R. S. Mallik, R. Abhiram, S. R. Reddy, and R. M. Jagadish, “A Comprehensive Survey on Sales Forecasting Models Using Machine Learning Algorithms,” in *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, 2022, pp. 1–6. doi: 10.1109/ICERECT56837.2022.10060168.

[7] A. Haryono, R. Sarno, and K. Sungkono, “Stock price forecasting in Indonesia stock exchange using deep learning: a comparative study,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, pp. 861–869, May 2024, doi: 10.11591/ijece.v14i1.pp861-869.

[8] A. Kumar Dubey, A. Kumar, V. García-Díaz, A. Kumar Sharma, and K. Kanhaiya, “Study and analysis of SARIMA and LSTM in forecasting time series data,” *Sustainable Energy Technologies and Assessments*, vol. 47, p. 101474, 2021, doi: https://doi.org/10.1016/j.seta.2021.101474.

[9] S. Mulla, C. B. Pande, and S. K. Singh, “Times Series Forecasting of Monthly Rainfall using Seasonal Auto Regressive Integrated Moving Average with EXogenous Variables (SARIMAX) Model,” *Water Resources Management*, vol. 38, no. 6, pp. 1825–1846, 2024, doi: 10.1007/s11269-024-03756-5.

[10] K. Gavahi, P. Abbaszadeh, and H. Moradkhani, “DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting,” *Expert Syst Appl*, vol. 184, p. 115511, 2021, doi: https://doi.org/10.1016/j.eswa.2021.115511.

[11] A. T. Haryono, R. Sarno, and K. R. Sungkono, “Transformer-Gated Recurrent Unit Method for Predicting Stock Price Based on News Sentiments and Technical Indicators,” *IEEE Access*, vol. 11, pp. 77132–77146, 2023, doi: 10.1109/ACCESS.2023.3298445.

[12] V. J. Halim, A. Ryanlie, K. Djong, K. E. Setiawan, and M. Fikri Hasani, “Rice Yield Production Forecasting using Deep Learning Models,” in *2023 International Conference on Networking, Electrical Engineering, Computer Science, and Technology (IConNECT)*, 2023, pp. 54–59. doi: 10.1109/IConNECT56593.2023.10327193.

[13] E. Yoshino, F. I. Kurniadi, and B. Juarto, “Forecasting Rice Production in Indonesia using Regression Techniques: A Comparative Analysis of Support Vector Machine, Linear Regression, and XGBoost Regression,” in *2023 10th International Conference on ICT for Smart Society (ICISS)*, 2023, pp. 1–5. doi: 10.1109/ICISS59129.2023.10291271.

[14] A. Islam, R. A. Ifty, M. A. H. Saim, J. A. Mahin, Md. F. Nizamee, K. E. Delowar., “Crop Yield Prediction: A Fusion of IoT and Machine Learning for Precision Agriculture,” in *2023 26th International Conference on Computer and Information Technology (ICCIT)*, 2023, pp. 1–6. doi: 10.1109/ICCIT60459.2023.10441217.

[15] H. Primageza, R. A. Vinarti, R. Tyasnurita, E. Riksakomara, and A. Muklason, “Comparison of NNs-ARIMAX and NNs-GSTARIMAX on Rice Price Forecasting in Indonesia,” in *2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2021, pp. 1–8. doi: 10.1109/ICACSIS53237.2021.9631332.

[16] D. B. Ansori, J. Slamet, M. Z. Ghufron, M. A. R. Putra, and T. Ahmad, “Android Malware Classification Using Gain Ratio and Ensembled Machine Learning,” *International Journal of Safety and Security Engineering*, vol. 14, no. 1, pp. 259–266, Feb. 2024, doi: 10.18280/ijsse.140126.

[17] A. Antoniadis, S. Lambert-Lacroix, and J.-M. Poggi, “Random forests for global sensitivity analysis: A selective review,” *Reliab Eng Syst Saf*, vol. 206, p. 107312, 2021, doi: https://doi.org/10.1016/j.ress.2020.107312.

[18] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T. Liu, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” in *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

[19] X. Guo, X. Gui, H. Xiong, X. Hu, Y. Li, H. Cui, Y. Qiu, C. Ma, “Critical role of climate factors for groundwater potential mapping in arid regions: Insights from random forest, XGBoost, and LightGBM algorithms,” *J Hydrol (Amst)*, vol. 621, p. 129599, 2023, doi: https://doi.org/10.1016/j.jhydrol.2023.129599.

[20] M. Ma, G. Zhao, B. He, Q. Li, H. Dong, S. Wang, and Z. Wang, “XGBoost-based method for flash flood risk assessment,” *J Hydrol (Amst)*, vol. 598, p. 126382, 2021, doi: https://doi.org/10.1016/j.jhydrol.2021.126382.

[21] B.-S. Kwon, R.-J. Park, and K.-B. Song, “Short-Term Load Forecasting Based on Deep Neural Networks Using LSTM Layer,” *Journal of Electrical Engineering & Technology*, vol. 15, no. 4, pp. 1501–1509, 2020, doi: 10.1007/s42835-020-00424-7.

[22] L. Muflikhah, I. Cholissodin, N. Widodo, F. E. Hermanto, T. L. Wargasetia, H. Ratnawati, and R. Sarno, “Single nucleotide polymorphism based on hypertension potential risk prediction using LSTM with Adam optimizer,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, p. 1126, May 2024, doi: 10.11591/ijeecs.v33.i2.pp1126-1139.

[23] S. S. Kumar, A. Kumar, S. Agarwal, M. Syafrullah, and K. Adiyarta, “Forecasting indoor temperature for smart buildings with ARIMA, SARIMAX, and LSTM: A fusion approach,” in *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2022, pp. 186–192. doi: 10.23919/EECSI56542.2022.9946498.

[24] A. Ibrahem Ahmed Osman, A. Najah Ahmed, M. F. Chow, Y. Feng Huang, and A. El-Shafie, “Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia,” *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 1545–1556, 2021, doi: https://doi.org/10.1016/j.asej.2020.11.011.