World Academy of Science, Engineering and Technology
International Journal of Industrial and Systems Engineering
Vol:17, No:8, 2023

Optimizing Forecasting for Indonesia's Coal and Palm Oil Exports: A Comparative Analysis of ARIMA, ANN, and LSTM Methods

Mochammad Dewo, Sumarsono Sudarto

***Abstract*—**The Exponential Triple Smoothing Algorithm approach nowadays, which is used to anticipate the export value of Indonesia's two major commodities, coal and palm oil, has a Mean Percentage Absolute Error (MAPE) value of 30-50%, which may be considered as a "reasonable" forecasting mistake. Forecasting errors of more than 30% shall have a domino effect on industrial output, as extra production adds to raw material, manufacturing and storage expenses. Whereas, reaching an "excellent" classification with an error value of less than 10% will provide new investors and exporters with confidence in the commercial development of related sectors. Industrial growth will bring out a positive impact on economic development. It can be applied for other commodities if the forecast error is less than 10%. The purpose of this project is to create a forecasting technique that can produce precise forecasting results with an error of less than 10%. This research analyzes forecasting methods such as ARIMA (Autoregressive Integrated Moving Average), ANN (Artificial Neural Network) and LSTM (Long-Short Term Memory). By providing a MAPE of 1%, this study reveals that ANN is the most successful strategy for forecasting coal and palm oil commodities in Indonesia.

***Keywords*—**ANN, Artificial Neural Network, ARIMA, Autoregressive Integrated Moving Average, export value, forecast, LSTM, Long Short Term Memory.

I. Introduction

THE export performance of Indonesia showed a recovery trend due to COVID-19 in 2021. Coal and palm oil play a role in increasing the prices of Indonesia's main commodity. In the past 5 years, Indonesia has held a 20% market share in the global coal trade, making it the second-largest exporter after Australia [1]. This market share positioned coal as one of Indonesia's flagship export commodities, with an average percentage of export value of 49% from 2017 to 2021 [3]. On the other hand, in the global palm oil market share for the past 5 years, Indonesia has held a 50% share of the global demand. The fulfillment of global demand put Indonesia as the market leader, far surpassing competitors from other countries such as Malaysia, the Netherlands, Papua New Guinea and Thailand [2]. From 2017 to 2021, palm oil accounted for an average of 54% of Indonesia's total exports in terms of traded goods quantity [3].

Indonesia's strength in the global supply chain has led to the emergence of new players in commodity trading. As a result, prospective investors and businesses looking to venture into the export business assess the development of a commodity using forecasting or data prediction as a basis for making business decisions. Forecasting for coal and palm oil commodities, using data from the Central Statistics Agent (Badan Pusat Statistik/BPS) [3] and the Exponential Triple Smoothing Algorithm, resulted in a MAPE of 31% for coal and 27% for palm oil.

According to Lewis [4], a MAPE value between 20% and 50% can be classified as a "reasonable" forecasting error, while achieving a value below 10% is considered "very good". Forecasting with an error below 10% assists in decision-making for both the government and businesses.

To achieve an error value below 10%, other forecasting methods such as ARIMA, ANN, and LSTM can be utilized. ARIMA is a forecasting model that attempts to identify historical data patterns and aims to identify the process that generates and influences the historical data patterns, also known as the Data Generating Process. Yee and Samsudin [5] achieved a 9% MAPE in forecasting palm oil prices in Malaysia using the ARIMA method. ANN is inspired by the function of human nervous system, which involves input, processing and output stages. Neural networks can be trained by adjusting various functions that adapt the weight values assigned to the relationships between elements. Pandey [6] forecasted electricity prices in India using the ANN method which resulted in a MAPE of 9%. LSTM is a type of Machine Learning based on the Recurrent Neural Network approach, which is capable of predicting the current state of a machine using large-scale data processing. Manowska [7] forecasted electricity demand in Poland using the LSTM method and obtained a MAPE of 1-3%.

Therefore, to improve the accuracy of forecasting the development of coal and palm oil commodities in Indonesia, we proposed to use three methods, namely: ARIMA, ANN and LSTM. The subsequent step would involve determining which method yields the highest accuracy for both commodities with an error below 10%.

II. Literature Review

A. ARIMA (Auto Regressive Integrated Moving Average)

ARIMA, developed by Box and Jenkins [8], is a statistical model that combines three components: Autoregressive (AR), Integrated (I) and Moving Average (MA). It is used to analyze and estimate statistical information in a time series. ARIMA

Mochammad Dewo is the Bachelor Student and Dr. Sumarsono Sudarto is the Assistant Professor in Supply Chain Engineering of Industrial Engineering Department, Mercu Buana University, Jakarta, 11650 Indonesia (e-mail: mochdewo@gmail.com, sumarsono@mercubuana.ac.id).

considers a linear combination of past values, past errors, current values and past values of alternative time series. The model aims to explain autocorrelation in the data and can be applied to both stationary and non-stationary statistics. ARIMA is denoted as ARIMA (p, d, q), where p, d and q are parameters. The parameters p and q represent the order of the autoregressive and moving average components, respectively. Parameter d represents the degree of difference needed to make the data stationary. Differentiation is typically used to remove linear and exponential trends in a time series. The value of d determines how many times the data should be differentiated to achieve stationarity state. For prediction, ARIMA sums past period values with constants, which indirectly helps estimate the average changes in predictions over a specific time interval. Various transformation techniques, such as Boxcox and Log transformations, can be used to stabilize the variance. The Auto Correlation Function (ACF) can also be employed to visualize whether the statistics are stationary or not. In the ARIMA model, the forecasting variable is a linear combination of previous observations [9].

B. ANN (Artificial Neural Network)

All ANNs are models inspired by the human brain that use simplified computational units called neurons. These neurons gather input signals, weigh them based on their importance and produce an output. The basic elements of an artificial neuron include input signals, synaptic weights, an aggregator, an activation threshold, an activation potential, an activation function and the final output [10].

In this model, input signals from the external environment are represented by a set of values and synaptic weights to assign importance to each input. The weighted inputs are combined to calculate the activation potential. If the activation potential exceeds a threshold, the neuron produces an output using an activation function. This output can be used as input for other neurons.

The operation of an artificial neuron can be summarized as:

1. Present a set of input values to the neuron.
2. Multiply each input by its corresponding synaptic weight.
3. Combine the weighted inputs to obtain an activation potential.
4. Determine if the activation potential exceeds a threshold to produce an output.
5. Apply an activation function to limit the output value.

This simplified model captures the essence of how artificial neurons process information in neural networks.

C. LSTM

LSTM neural networks are a special type of RNN (Recurrent Neural Network) that can solve the long-term dependency problem in general RNNs. LSTM is suitable for processing and forecasting data with long intervals and delays in time series data. Unlike a single-layer RNN, LSTM neural networks have four loops. In addition to the external RNN loop, they also have an internal LSTM loop [11].

D. Error Evaluation

To assess the forecasting accuracy of the model based on commodity export values, this research chose 4 evaluation methods: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root-Mean-Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

MAE represents the average absolute deviation between predicted and actual values, which effectively reflecting the true prediction error. MSE measures the average squared deviation between predicted and actual values, which can be used to evaluate the magnitude of the deviations. RMSE is the square root of MSE and is used to calculate the deviation between predicted value and actual value. MAPE is a measure of the prediction accuracy of a forecasting method in statistics. It usually expresses the accuracy as a ratio.

E. Literature Review

TABLE I
RELATED JOURNAL

Author	Method	Result
Urrutia et al. [14]	ARIMA & BANN	The most fitted model in forecasting the values of both imports and exports from the year 2018 to 2022 is the Bayesian Artificial Neural Network (BANN)
Rahim et al. [15]	FRBS, ARIMA, ANN, & ARIMA-ANN	The forecast error of the proposed method and hybrid ARIMA-ANN method is lesser compared to other methods, Fuzzy Time Series forecasting with the proposed method provides more proper and simpler method to forecast CPO price.
Liu [16]	LSTM	The coal price forecast in the Bohai Sea region is better with reduced model training errors and improved accuracy
Khalid et al. [17]	ARDL, ARIMA, & ARIMAX	ARIMAX model is the most accurate and the most efficient model as compared to ARDL and ARIMA in forecasting the crude palm oil price.

III. RESEARCH METHOD

In this study, quantitative research methodology was used to enhance the accuracy level of forecasting coal and palm oil commodities using statistical mathematical models. The study also evaluates the comparison of various methods used in the research. The required data are obtained from Trademap.org [1], [2], developed by the International Trade Center (ITC). The data needed to support the research include the export data of coal and palm oil commodities in terms of trade value, which forms the basis for calculations according to the research title.

Secondary data used in the study are sourced from the Central Statistics Agent (Badan Pusat Statistik/BPS) [3]. The secondary data taken for this research consisted of Indonesia's international trade data, particularly the export data of coal and palm oil commodities in terms of trade value. The data collection method was carried out by collecting data export data on exports of coal and palm oil commodities in terms of trade value throughout 2002 to 2021 from the ITC. These data are processed using software such as Minitab and Google Colab Research to perform statistical calculations and train artificial intelligence models for forecasting purposes.

The presented flowchart encapsulates a systematic approach to enhance the accuracy of commodity export value forecasting for Indonesia's pivotal coal and palm oil sectors.

Problem Identification: The process commences by identifying the core issue: the current forecasting methods yield error rates that can significantly impact industrial output and economics, prompting the quest for improved accuracy.

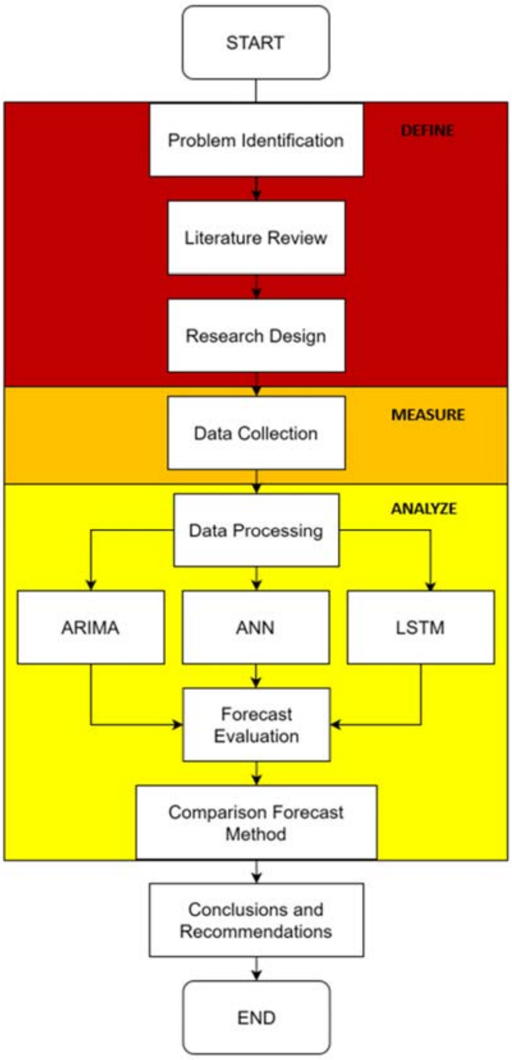

Fig. 1 Research Methodology Flowchart

Literature Review: An exploration of prior research methodologies provides insights into various forecasting approaches, such as ARIMA, ANN, and LSTM. This knowledge forms a solid foundation for the subsequent steps.

Research Design: The design phase outlines the roadmap for the study, highlighting the significance of achieving an error rate below 10%

Data Collection: Trusted sources like Trademap.org and Central Statistics Agent (BPS) contribute essential export data of coal and palm oil, which serves as the bedrock for accurate predictions.

Data Processing: Data preprocessing comes next, involving transformation and differentiation to ensure data are ready for meaningful analysis, an integral step in ensuring reliable results.

ARIMA, ANN, LSTM: The flowchart then branches into the heart of the research, where three distinct methodologies—ARIMA, ANN, and LSTM—are employed to forecast the trends of coal and palm oil commodities.

Forecast Evaluation: The evaluation phase critically assesses the accuracy of forecasts using key metrics like MAE, MSE, RMSE, and MAPE, determining the effectiveness of each methodology.

Comparison Forecast Method: The comparative analysis stage consolidates the results of ARIMA, ANN, and LSTM forecasts, revealing the standout method in terms of precision: the ANN approach.

IV. Finding

The ITC serves as a vital platform to facilitate international trade by providing valuable information and intelligence to support trading processes. Developed by the United Nations Conference on Trade and Development (UNCTAD) and the World Trade Organization (WTO), the ITC has established a Trade Support Institutions (TSIs) that focus on trade promotion, sectoral performance and trade development strategies. Through its strategic market research website, trademap.org, the ITC offers comprehensive statistical information on international trade flows, assisting in measuring national and sectoral trade competitiveness and identifying priority products and markets for trade development. This article explores the data collected from trademap.org for coal and palm oil commodities, providing insights into their market trends over the years.

A. ARIMA Analysis of Coal and Palm Oil

The coal data were analyzed using ARIMA to forecast future trends. The data transformed and differentiated to achieve stationary state. After model identification, the selected ARIMA model was used for forecasting. The forecasted values for 2018, 2019, 2020 and 2021 are as follows:

- 2018: 20,628,877
- 2019: 22,163,219
- 2020: 23,015,827
- 2021: 23,489,607

The palm oil data underwent transformations and differentiation to achieve stationarity. Following model identification, six ARIMA models were tested and assessed. The chosen ARIMA model was then utilized for making forecasts. The forecasted values for 2018, 2019, 2020 and 2021 are as follows:

- 2018: 21,361,961
- 2019: 22,433,930
- 2020: 23,505,899
- 2021: 24,577,868

The forecasting data of ARIMA Coal and Palm Oil in this method do not yield good values, in the error evaluation such as MAE with values 8.755.815 and 5.195.029, MSE with values 104.871.368.434.617 and 31.250.112.761.241, RMSE with value 10.240.672 and 55.900.180, and MAPE, which produces values of 49% and 31%. These values are larger than the initial method that underlies the research, which is 31% for coal and 27% for palm oil. Unfavorable forecasting results are due to less than optimal data for forecasting, where 20 data points are used, 16 data points are processed as initial data for forecasting and four data points are used for testing. The processed data using Minitab show white noise in the ACF and PACF tests on the differentiated data. As consequence of this white noise, the data are overfitted following [12], which assumes the values of AR (p) and MA (q) as 1 in Monte Carlo overfitting tests. As a result, 6 models are used for testing to obtain the best model for forecasting using the ARIMA method.

B. ANN and LSTM Analysis

1. Import Libraries: Import necessary modules for data processing, modeling and visualization.
2. Create and Save Dataset: Create two data frames, 'coal_df' and 'palm_df', then save them as CSV files for easy data loading.
3. Load Dataset: Load data from "coal_data.csv" and "palm_ data.csv" into data frames. Apply Min-Max Scaling using MinMaxScaler from scikit-learn. Split the data into training and testing datasets ('coal_train', 'coal_test', 'palm_ train' and 'palm_test').
4. Create LSTM Model: Define the 'create_lstm_model' function to build an LSTM model using Keras. The model consists of two LSTM layers with 50 units each. The first layer has 'return_sequences=True' to pass the output sequentially to the second layer. An additional layer serves as the output layer. Compile the LSTM model using the Adam Optimizer and mean squared error loss. Create separate LSTM models for coal ('coal_lstm_model') and palm oil ('palm_lstm_model'). Train the models with the training data reshaped to match the LSTM model dimensions, using 100 epochs and a batch size of 16.
5. Create ANN Model: Define the 'create_ann_model' function to build an ANN model using Keras. The model has two layers with 50 and 100 units, respectively. The input dimension is set to 1. ReLU activation is used for both layers, and an output layer is added. Compile the ANN model using the Adam Optimizer and mean squared error loss. Create separate ANN models for coal ('coal_ann_ model') and palm oil ('palm_ann_model'). Train the models with the training data reshaped to match the ANN model dimensions, using 100 epochs and a batch size of 16.
6. Prediction from Trained Models: Use the trained models to predict the test data. For LSTM models, reshape the test data before prediction. For ANN models, use the test data as is. Apply the inverse transform to obtain the actual results. Reshape the predicted data back to (-1, 1) and flatten it to a one-dimensional array.
7. Evaluation of Trained Models: Calculate error metrics (mean squared error and mean absolute error) using 'mean_ squared_error' and 'mean_absolute_error' from scikit-learn.

In the ANN method, the same variety of data is used, with 16 data points to train the machine learning model and 4 data points to test the trained model. Python is used as a tool for machine learning with Keras and Tensorflow to optimize the data training process. Machine learning was performed using 2 layers with 50 and 100 layers respectively, along with 100 epochs and a batch size of 16, gathered using the Adam Optimizer, resulting in an optimal data training process. The data that are trained from each layer is passed directly to the next layer without being input again and end with the output layer. This method provides the best results among the three methods, both for coal and palm oil commodities, with MAE with a value of 152.826 and 73.460, MSE with a value of 47.120.895.824 and 221.585.339.480, RMSE with a value 217.073 and 146.920, and MAPE values of 0.77% and 0.27%, respectively. These results indicate that the ANN method through machine learning is the best method for forecasting coal and palm oil commodities.

The LSTM method is performed concurrently with the ANN data processing, so the process is similar between the ANN and LSTM methods. However, for the LSTM method, it produces optimal forecasting results with MAE values of 2.203.326 and 3.177.152, MSE with values 8.347.257.137.608 and 15.126.134.951.321, RMSE with values 2.889.162 and 3.907.574, and MAPE values of 9% and 15% for coal and palm oil, respectively. The LSTM method is not as optimal as the ANN method in machine learning because the initial data in the input layer is reused in the subsequent input process. Repetition in this process causes the LSTM method to be less optimal than ANN.

C. Comparison of Forecast Methods

TABLE II
COMPARISON FORECAST METHODS (ARIMA, ANN, & LSTM)

Method	Commodity	MAE	MSE	RMSE	MAPE
ARIMA	Coal	8.755.815	104.871.368.434.617	10.240.672	49%
ARIMA	Palm Oil	5.195.029	31.250.112.761.241	55.900.180	31%
ANN	Coal	152.826	47.120.895.824	217.073	**0,77%**
ANN	Palm Oil	73.460	221.585.339.480	146.920	**0,27%**
LSTM	Coal	2.203.326	8.347.257.137.608	2.889.162	9%
LSTM	Palm Oil	3.177.152	15.126.134.951.321	3.907.574	15%

The three employed methods have yielded diverse outcomes in the forecasting process. Simplifying the comparison, we employ the MAPE, a metric expressing errors in percentage terms. Significantly, the MAPE values for the ARIMA method are 49% (coal) and 39% (palm oil); for the ANN method, they are 0.77% (coal) and 0.27% (palm oil); and for the LSTM method, they are 9% (coal) and 15% (palm oil).

Within the framework of this study, the aim is to achieve an error rate below 10%. Among the three methodologies, only ANN and LSTM fulfill this criterion. Specifically, ANN demonstrates error rates below 10% for both commodities. Meanwhile, LSTM attains this threshold for coal; however, it records a 15% error for palm oil. Nonetheless, this palm oil approximation is in close proximity to the targeted error threshold.

In conclusion, for enhancing the accuracy of coal and palm oil commodity data forecasting with errors below 10%, the most suitable approach is the application of the ANN.

V. CONCLUSION

Based on the research findings, it can be concluded that the most optimal forecasting method for coal and palm oil commodities is the ANN method, which has the lowest error value. The ARIMA method produces suboptimal results due to the limited amount of data used (20 data points), while the optimal data processing requires a larger dataset of 36 data points [13]. Utilizing machine learning as a data processor, even a small amount of data can be optimized to create accurate forecasts.

REFERENCES

[1] ITC (2023b). Trade Statistics for International Business Development. Trade Map - List of importers for the selected product in 2022 (Coal; briquettes, ovoids and similar solid fuels manufactured from coal). https://www.trademap.org/Country_SelProduct.aspx?nvpm=1%7C%7C%7C%7C%7C270120%7C%7C%7C6%7C1%7C1%7C1%7C1%7C1%7C2%7C1%7C1%7C1

[2] ITC (2023a). Trade Statistics for International Business Development. Trade Map - List of exporters for the selected product (Palm oil and its fractions, whether or not refined (excl. chemically modified)). https://www.trademap.org/Country_SelProduct_TS.aspx?nvpm=1%7C%7C42%7C%7C%7C1511%7C%7C%7C4%7C1%7C1%7C2%7C2%7C1%7C3%7C1%7C1%7C1

[3] Badan Pusat Statistik. (2017-2021). Statistik Perdagangan Luar Negeri Indonesia Ekspor (2017-2021), Jilid I. Badan Pusat Statistik. https://www.bps.go.id/publication/2022/07/06/d3580f9e1b55a44b265d5ad8/statistik-perdagangan-luar-negeri-indonesia-ekspor-2021-jilid-i.html

[4] Lewis, C. D. (1982). Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. Butterworth Scientific.

[5] Yee, K. W., & Samsudin, H. B. (2021). Comparison between Artificial Neural Network and ARIMA Model in Forecasting Palm Oil Price in Malaysia. International Journal of Scientific Engineering and Science, 5(11), 12–15.

[6] Pandey, A. K., Sahay, K. B., Tripathi, M. M., Chandra, D. (2014). Short-term load forecasting of UPPCL using ANN.

[7] Manowska, A., & Bluszcz, A. (2022). Forecasting crude oil consumption in Poland based on LSTM recurrent neural network. Energies, 15(13), 4885. https://doi.org/10.3390/en15134885

[8] G.E.P. Box, G. Jenkins, “Time Series Analysis, Forecasting and Control”, Holden-Day,San Francisco, CA, 1970.

[9] Jain, G., & Mallick, B. (2017). A study of time series models Arima and ETS. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2898968

[10] da Silva, I. N., Spatti, D. H., Flauzino, R. A., Liboni, L. H. B., & Reis Alves, S. F. dos. (2017). Artificial Neuron. In Artificial Neural Network A Practical Course (pp. 11–13). essay, Springer International Publishing Switzerland.

[11] Huang, B., Tong, L., & Zuo, Y. (2022). Short-term Power Generation Load Forecasting based on LSTM Neural Network. Journal of Physics: Conference Series, 2247(1), 012033. https://doi.org/10.1088/1742-6596/2247/1/012033

[12] Liu, Yamei, "Overfitting and forecasting: linear versus non-linear time series models " (2000). Retrospective Theses and Dissertations. 13915. https://lib.dr.iastate.edu/rtd/13915

[13] Heizer, J., & Render, B. (2009). Operations management (9th ed.). Pearson Prentice Hall.

[14] Urrutia, J. D., Abdul, A. M., & Atienza, J. B. (2019). Forecasting Philippines imports and exports using Bayesian Artificial Neural Network and autoregressive integrated moving average. AIP Conference Proceedings. https://doi.org/10.1063/1.5139185

[15] Rahim, N. F., Othman, M., & Sokkalingam, R. (2018). A comparative review on various method of forecasting crude palm oil prices. Journal of Physics: Conference Series, 1123, 012043. https://doi.org/10.1088/1742-6596/1123/1/012043

[16] Liu, X. (2021). Research on the forecast of coal price based on LSTM with improved Adam Optimizer. Journal of Physics: Conference Series, 1941(1), 012069. https://doi.org/10.1088/1742-6596/1941/1/012069

[17] Khalid, N., Hamidi, H. N. A., Thinagar, S., & Marwan, N. F. (2018). Crude palm oil price forecasting in Malaysia: An econometric approach. Jurnal Ekonomi Malaysia, 52(3). https://doi.org/10.17576/jem-2018-5203-19.