

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 8, No. 3, June, 2024



Selection of Modelling for Forecasting Crude Palm Oil Prices Using Deep Learning (GRU & LSTM)

Gabriella Alexander Tardini¹, Suharjito^{1*}^o

¹ Industrial Engineering Department, BINUS Graduate Program – Master of Industrial Engineering, Bina Nusantara University, Jakarta 11480, Indonesia.

1- Introduction

In 2023, Indonesia emerged as the top global producer of palm oil, contributing 47,000,000 metric tons, or 59% of world production [1]. Palm oil is a vital commodity that enhances the quality of life, bolsters the GDP of nations, and contributes to achieving various sustainable development goals, including eradicating hunger and poverty, providing meaningful employment, and promoting economic growth [2]. The extraction of palm oil from ripe palm fruit involves five key processes, namely sterilization, peeling, crushing, pressing, and clarification [3]. One of the most significant oilseed trees in the world is the oil palm tree, *Elaeis guineensis*. Different phytonutrients, including sterols, vitamin E, squalene, phospholipids, and others, are present in crude palm oil and palm kernel oil. Among all widely used vegetable oils, palm oil contains the greatest concentration of tocopherols and carotenoids. The fleshy fruit pulp (mesocarp), which is used to make palm oil, contains around 29.32% water (in a wet state), 2.03% protein, 68.09% lipids, and 1.11% ash. In contrast, the oil content of palm kernels, which are produced from the inner kernel of the palm fruit, is only around half that of the palm fruit [4].

^{*} **CONTACT**: suharjito@binus.edu

DOI: http://dx.doi.org/10.28991/ESJ-2024-08-03-05

^{© 2024} by the authors. Licensee ESJ, Italy. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC-BY) license (https://creativecommons.org/licenses/by/4.0/).

Cooking oil, biofuel, and cosmetic components are only a few uses for crude palm oil, which is the main product in the palm oil industry [5]. Figure 1 shows the production of crude palm oil in Indonesia in 2022 (tons) based on BPS - Statistics Indonesia (2023). The dynamic nature of palm oil production has a significant impact on prices and downstream usage, particularly the cost comparison between biodiesel and petroleum fuel. As a result, dealers may set inflated or unjust pricing for palm fruit bunches, making it challenging to set different rates for different product attributes. For this industry to have more exact management and planning, reliable forecasting technology is crucial for estimating the dynamics of palm oil output and its pricing [6]. Amin et al. [7] focused on estimating Value at Risk (VaR) for CPO prices using Monte Carlo simulation, offering market players with risk management implications. The integration of the spot market for CPO and the derivatives market is examined in Helbawanti et al. [8], with particular emphasis on the effects of price variations and the derivatives market's influence over the spot market. According to Siregar & Syafarudin [9], who looked into the connection between CPO production volume, CPO pricing, profitability, and stock returns in the plantation industry sector, the impacts on profitability and stock returns were inconsistent and minor.



Figure 1. Production of Crude Palm Oil in Indonesia (2022)

For a wide range of stakeholders, including the palm oil sector, investors, politicians, and consumers, research on crude palm oil (CPO) price predictions is crucial. In an industry with unpredictable pricing, this study is essential for reducing price risk and enabling farmers, dealers, and processors to make well-informed decisions. Additionally, it makes it possible to optimize the supply chain, helping businesses cut costs and improve operational effectiveness and helping investors determine the profitability of their investments. These projections are used by policymakers to create agriculture and trade policies that have an impact on livelihoods, food security, and global commerce. Furthermore, CPO price estimates impact judgments about economic growth and environmental sustainability, as well as market strategies, international trade discussions, and consumer pricing expectations for items made from palm oil. In the end, this research encourages innovation and market response, having a significant influence on commerce, the global economy, and environmental sustainability.

This study will employ deep learning methods, namely long-short-term memory (LSTM) and gated recurrent units (GRU). Which deep learning strategy is most successful at modeling projections for crude palm oil prices using the LSTM and GRU approaches is the question. In the past, GRU and LSTM have been used to forecast stock prices. In the network's hidden layer, LSTM replaces conventional artificial neurons with memory cells and compute units. With the help of these memory cells, the network can efficiently link fresh memories and inputs and utilize the data architecture dynamically, increasing the accuracy of predictions. The fundamental distinction between GRU and LSTM is that the former lacks an output barrier, while the latter does. Because GRU has fewer parameters than LSTM, it is considerably quicker. High prediction accuracy and rapid convergence characterize this modeling. Because the LSTM network has a larger memory capacity to retain and interpret the prior data, it can produce superior results for huge data sets. However, because GRU has fewer parameters than LSTM, it is quicker [10]. Crude palm oil prices have been done using the Single Hidden Layer Feedforward Neural Network (SLFNs) method [11], Multilayer Perceptron (MLP) and LSTM [12], and Support Vector Regression (SVR) [13], but they have not specifically used comparisons between LSTM and GRU. The price of palm oil shows a seasonal pattern because palm oil production has a seasonal pattern every year [14], so the GRU and LSTM methods are used.

The objective of this research is twofold. Firstly, it aims to develop a time series prediction model for crude palm oil prices utilizing the GRU and LSTM methods, with the goal of providing effective price forecasts with a high level of

accuracy. Secondly, the study seeks to ascertain the superior model for predicting crude palm oil prices between the GRU and LSTM methods within the scope of this research. Theoretically, this study helps to improve the accuracy of predictions for the price of crude palm oil. By supporting associated sectors in obtaining more exact inventory management and planning, this research also benefits the users. The rest of the essay is divided into the following sections: an introduction, a review of the literature, a summary of the methods employed, a discussion of the findings, and a conclusion.

2- Literature Review

2-1-Crude Palm Oil

Crude palm oil (CPO) is a byproduct of processed palm oil that has been refined using different techniques, such as centrifugation, vacuum drying, threshing, digestion, and clarifying. Some of these unit processes use heat to handle materials or are made easier by it. The temperatures used to extract the oil from the palm fruit range from 60°C to 140°C, which is significantly higher than the melting point of CPO [15]. There are several uses for palm oil, with the food business accounting for around 80% of its uses, while the other 20% are for non-food purposes. In the food business, refined, bleached, and deodorized (RBD) olein is mostly used to make cooking oils, frying oils, shortening, and margarine. On the other hand, margarine and shortening are mostly made with RBD stearin. RBD palm oil is used in the production of margarine, shortening, cooking oil, and even ice cream. RBD palm oil is simply palm oil that has undergone purification.

In non-dairy creamers and whiteners, palm oil, palm kernel oil, and other fat blends are used as milk fat alternatives. Additionally, the creation of specialty fats, including flexible coating fats, cocoa butter alternatives, and cocoa butter equivalents (CBE), depends heavily on palm oil. Additionally, epoxidized palm oil products (EPOP), polyols, polyurethanes, polyacrylates, and drilling muds are all made from crude palm oil. It functions as a base for drilling mud and as fuel for well-tuned car engines. Additionally, it acts as a cleaner alternative to diesel fuel [3].

2-2-Forecasting

The main objective of forecasting is to offer the most accurate future forecasts possible based on all available data, including historical data and knowledge of upcoming events that can have an impact on the projections. Forecasting is necessary in many circumstances, like deciding whether to predict demand for constructing an extra power plant in the following years, staffing a call center for the upcoming week, and monitoring inventory levels. In scenarios involving investment capital, forecasting may be essential many years in advance, whereas for telecommunications routing, it might only be required for a brief period. Given its potential to affect various sectors within a company, forecasting should hold a crucial role in the decision-making processes of management. Contemporary organizations demand forecasting for different time frames, be it short, medium, or long term, contingent upon the specific application. The process of forecasting typically commences with characterizing the issue, followed by data collection, initial stage analysis, model choice and adaptation, and ultimately, implementing and evaluating the forecasting model [16].

Because palm oil output follows a yearly pattern each year, palm oil prices similarly exhibit seasonal patterns. Every year, from November to February, the output of palm oil begins to decrease because of the rainy season. Production of palm oil will impact the market's supply of palm oil, which will subsequently impact pricing [14].

2-3-Long Short-Term Memory (LSTM)

The LSTM model can preserve and learn long-term input dependencies since it is basically a memory of a recurrent neural network (RNN). These memory extensions provide the model the ability to read, write, and erase data from memory since they increase the data storage capacity. The "gate" cells in these LSTM memories get their name from their capacity to choose whether to store or ignore memory data. The LSTM model records significant input attributes and keeps track of them for a long time. Based on the weight given to the information throughout the training phase, the choice of whether to keep or reject it is made [17].

Input, output, forget, and update gates make up the majority of an LSTM cell's architecture (Figure 2). The input gate determines which information the neuron should take in, the output gate builds a new long-term memory, and the update gate changes the memory of the cell. The forget gate determines what should be forgotten based on the data acquired by the preceding memory unit. Together, these four elements produce new long-term, short-term, and output sequences at particular time intervals and receive long-term and short-term input sequences at particular time intervals [18]. Depending on its priority, this gate decides whether or not data may travel through. These gates allow the network to learn what to remember, forget, recall, monitor, and reject. Data that will be processed in the next phase is collected using cell state and hidden state. Therefore, it is possible to safeguard the diminishing gradient.

The Gate equation appears as follows:

Input gate: $i_t = \sigma(W_i h_{t-1} + W_i h_t)$

(1)

Forget gate: $f_t = \sigma(W_f h_{t-1} + W_f h_t)$ (2)Output gate: $o_t = \sigma(W_o h_{t-1} + W_o h_t)$ (3)Intermediate cell state: $\tilde{C} = tan(W_c h_{t-1} + W_c h_t)$ (4)Cell State: $c_t = i_t * \tilde{C} + f_t * c_{t-1}$ (5)New State: $h_t = o_t * tan(c_t)$ (6)

where X_t is an input vector, h_t is an output vector dan W, U, f are the parameter matrix and vector [16].



Figure 2. LSTM Structure [16]

2-4- Gated Recurrent Unit (GRU)

The GRU and LSTM both function similarly, despite the GRU cell having a single hidden state that combines the forget gate and input gate functions into a single update gate. GRU also assembles cells and hidden states into a single state. Because it contains half as many gates as the LSTM, the GRU is a well-liked and smaller variation of the LSTM cell. Figure 3 shows the GRU structure. The update gate and the reset gate are the two GRU gates [18]. The data that has to be transmitted to the output is specified by these updates and reset gates as vectors. How the fresh input is integrated with the old memory will depend on the reset gate. The update gate chooses how much memory was recently saved [19].



It does not possess an output gate, but it includes an update gate denoted as z and a reset gate denoted as r. These gates, represented as vectors, determine the information that should be transmitted to the output. In the equation, X_t represents an input vector, h_t represents an output vector, and W, U, and f denote the parameter matrix and vector [16].

2-5-ARIMA (Autoregressive Integrated Moving Average)

ARIMA, or Autoregressive Integrated Moving Average, is a forecasting algorithm that models the relationship between an observation with a number of delayed observations (autoregression) and an error term (moving average) by using the concept of differencing to make time series data stationary. An approach for simulating the dynamic behavior of a process that generates a time series of consecutive values $Y' = (y_1, ..., y_n)$ of a quantity Y observed in equal intervals {S, ..., nS}, where S is the sampling interval, is called ARIMA (Autoregressive Integrated Moving Average) modeling. The three components of ARIMA—AR (Autoregressive), I (Integrated), and MA (Moving Average)—can be used separately or in combination depending on the situation. As a result, in addition to AR, I (for instance, random walk), MA, ARI, ARMA, IMA, and ARIMA models, there are also white noise models (which have no components). The process is nonstationary when component I is included. A stationary ARMA process is obtained by differentiating a nonstationary ARIMA process. As members of the linear time series model class, ARIMA models frequently presuppose that the vector is jointly normally distributed, sometimes following transformation. Typically, ARIMA modeling is carried out by determining the order in which the AR, I, and MA portions should be identified, estimating the AR and MA parameters, and assessing the validity of the hypotheses. It is frequently challenging to outperform ARIMA models for short-term forecasting in a quality setting when using nonlinear and/or nonnormal models [20]. As to Liu et al. [21], the primary phases involved in ARIMA modeling are as follows:

- 1. Check for stationarity in the time series. a. To achieve stationarity if the series is not stationary, subtract the time series value.
- 2. Choose a model to suit the time series that is now stationary.
- 3. Calculate the chosen model's parameters.
- 4. Assume by means of a model.

2-6-Related Works

Price projections for oil and palm oil have been the subject of prior research. For example, Aisyah et al. [11] conducted research on the application of extreme machine learning to forecast crude palm oil prices using the Single Hidden Layer Feedforward Neural Network (SLFNs) method with datasets from the investing.com website for the period from April 1, 2021, to April 14, 2022. The results showed that the best model had an average MAPE and RMSE of 0.0173 and 0.0308, respectively. When the findings of this research were applied to the following nine days, the average price of crude palm oil was \$6,694. The highest crude palm oil price was obtained on April 27, 2022, at \$6,931, while the lowest price was recorded on April 15, 2022, at \$6,424.

With a dataset received from the Malaysian Palm Oil Board in 2019, Always et al. [22] conducted a study on palm oil utilizing the Discrete Hopfield Neural Network (DHNN2-SAT) and the 2-Satisfiability logic based on the Reverse Analysis (2-SATRA) technique. The core advantages of both HNN and the reverse analysis approach are utilized and combined in this method, which also offers a high degree of flexibility in determining the ideal price trend for palm oil. According to studies, 2SATRA can provide a range of information on pricing fluctuations and can deduce logical principles from these simulations with an acceptable level of accuracy. The top performance had an RMSE of 1.1905 and a MAE of 0.8326.

Additionally, Amal [12] conducted research on projections for the price of crude palm oil utilizing data from the Ministry of Trade of the Republic of Indonesia from January 4, 2010, to December 29, 2020, using the Multilayer Perceptron (MLP) and LSTM algorithms. For MLP, it has a 33,222.80 MSE. Adam is used as the optimization technique, and an LSTM with a time step of 2, 6 hidden neurons, and a 70%–30% split between training and testing data yields the lowest MSE value, 32,224.17. The real number matches the forecast findings exactly, and the MAPE value is under 2.11%. Since the MAPE value is under 10%, the model has a high level of prediction accuracy.

Additionally, Aulia et al. [23] developed the gated recurrent unit approach for predicting palm oil prices using a recurrent neural network (RNN) model. Prices for palm oil are included in the dataset for the period from July 5 to August 2, 2021 (excluding weekends). The selection model with a learning rate of 0.01, a batch size of 100, a hidden state epoch of 512, and a window size of 30 produced the lowest MSE at the 125th iteration in an experiment involving 125 iterations. The accuracy of this model's forecasts or predictions, which has a MAPE score of 4.84%, shows that it has high forecasting skills.

A collection comprising market comments, trade data, commentary data, and news event data (30 September 2016– 30 September 2017) from the Dalian Commodity Exchange, Eastmoney, Xueqiu, and Wind was used by Chai et al. [24] to predict palm oil prices using the Multi-source Heterogeneous Data Analytic (MHDA) technique. A review of palm oil futures data from September 2016 to September 2017 showed how successful the suggested approach is. In comparison to Support Vector Machine (SVM), Random Forest, Multiple Kernel K-Means, and Tensor-Based Learning, this technique has the highest accuracy, reaching 64.15 percent.

Crude oil price prediction was carried out by Wang et al. [25], utilizing the EPPA technique (ensemble probabilistic prediction methodology). The back-propagation neural network (BPNN) method, the long-short-term memory (LSTM) model, the Gaussian process regression (GPR), the least absolute shrinkage and selection operator (LASSO), and the bidirectional long-short-term memory (BI-LSTM) model are some of the techniques that are combined in this method. With no missing data points, the researchers compiled their dataset using daily and weekly European Brent spot pricing (BRE) data in USD per barrel. The US Energy Information Administration provided the statistics, which covered the periods from May 20, 1987, to June 8, 2018, and from May 15, 1987, to June 8, 2018, omitting federal holidays. The prediction interval created by the EPPA method they proposed has a high likelihood of encompassing the actual value because it effectively captures the uncertainty in crude oil prices with a high degree of accuracy.

SARIMA and ARIMA algorithms were used by Lee et al. [26] to predict crude oil prices in Europe and the US. They obtained monthly crude oil price information for the United States and Europe from the Federal Reserve Economic Data (FRED) and the U.S. Energy Information Administration, starting in January 2017 and continuing through September 2021. They utilized observations from January 2017 to March 2021 as the training set and data from April 2021 to September 2021 for testing. The reliability of their forecasts in Europe and the US was evaluated using the performance measures RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), and MAE (Mean Absolute Error). These were the outcomes for SARIMA: For Europe, the corresponding MAPE, MAE, and RMSE values were 0.24, \$12.38, and \$15.51. For the United States, MAPE was 0.30, MAE was \$107, and RMSE was 14.04. It is significant to notice that even though the green ARIMA line wasn't exactly parallel to the genuine line, the time series of data for forecasting natural disasters was still deemed adequate. The graph demonstrated that both the actual trends in the price of crude oil in Europe and the United States closely tracked the prediction line generated by the ARIMA model. In the United States, the outcomes were MAPE of 0.27, MAE of \$11.13, and RMSE of 13.98. The study comes to the conclusion that the ARIMA model is more accurate and effective in forecasting crude oil prices in both Europe and the United States than the SARIMA method. Saadah et al. [13] Utilize information spanning the years 2015 to 2018 from the WTI crude palm oil dataset for conducting a study on support vector regression (SVR) to predict palm oil prices. The assessment of the SVR model, based on accuracy metrics such as MAPE and R-squared, yielded favorable results, affirming the model's reliability with an R-squared value of 98.83% and a MAPE value of 4.022.

The LSTM algorithm is used in Sen [27] to forecast the price of crude palm oil from 2011 to 2020. The price of Bitcoin, the S&P index, the price of gold, the Shanghai Stock Exchange index, the German Dollar Bund Yield index, inventory levels, and inventory estimates were some of the variables that the researchers utilized. Other variables were the price of Bitcoin, the S&P index, the price of gold, the Shanghai Stock Exchange index, the German Dollar Bund Yield index, and inventory levels. As a result, the estimated inventory was able to forecast real inventory outcomes between 2011 and 2020 with an adjusted R-squared value of 0.343. The daily volatility of crude oil prices was not higher than that of Bitcoin, the Shanghai Stock Exchange, German 10-year bond yields, or US 10-year bond rates, which is significant. They predicted inventory levels based on how the Granger variance of the S&P 500 affected the daily variation in crude oil prices. In addition, after a few days, Granger variations in the dollar index and the price of gold caused changes in the price of crude oil. Their prediction model, consisting of one LSTM layer and one dense layer, successfully predicted crude oil price variations while taking into account expected inventory, the previous day's variation in crude oil price, the S&P 500, the dollar index, and gold as independent variables. After testing several hyperparameter combinations, they discovered that the Adam optimizer with a one-day lookback period, 68 iterations, and a tan activation function generated the best accuracy, with an outstanding R-squared value of 80.18%. During the COVID-19-related crude oil crisis, this method was successful in anticipating sizable changes in oil prices. After hyperparameter optimization, it was discovered that the Nadam optimizer, in combination with an eight-day lookback period, 68 iterations, and a tan activation function, produced the highest accuracy, with an Rsquared value of 80.74%. This technique also showed promise in predicting substantial fluctuations in oil prices during the COVID-19 crisis.

In order to predict the price of crude palm oil, Suryani [28] utilized data they collected from websites run by the Ministry of Energy and Mineral Resources (ESDM) using the Multiple Linear Regression approach. The study's data source included monthly pricing information for 55 various types of crude oil during a four-year period (2018–2022). With SLC = 9%, Attaka = 45%, Duri = 126%, Belida = 33%, Banyu = 150%, and SC = 50%, the model generated calculates MAPE with an extremely high degree of accuracy. The MAPE computation findings are used to estimate Indonesian crude oil prices, and the linear regression equation that does so demonstrates the high caliber of the model created by SLC crude oil. In comparison to Duri and Banyu crude oils, Attaka, Belida, and SC crude oils fared better.

2-7-State of the Art

The benefits and drawbacks of the various approaches were contrasted, particularly in research that simply made price predictions for palm oil. The SLFNs technique has the benefit of learning considerably more quickly and performing greater generalization than artificial neural networks, but it also has the drawback of not optimizing the bias and initial number of neurons [11]. The 2-SATRA method, which applies 2-SAT logic to the Discrete Hopfield Neural Network (DHNN2-SAT), has the benefit of being able to extract logical rules with a respectable degree of accuracy and providing a variety of insights on price movements. However, 2SATRA has a limited ability to accurately predict palm oil prices at particular periods. This drawback results from the nature of the neurons in 2SATRA, which are only capable of detecting broad trends (upward or downward) in commodity prices. The Multilayer Perceptron (MLP) approach, on the other hand, provides a unique set of benefits. This approach works well with large amounts of input data, offers quick predictions after training, and maintains constant accuracy even with smaller datasets. It may be efficiently applied to difficult, non-linear situations. The extent of its flaws is still unknown, though. Calculations are difficult and time-consuming since each independent variable influences the dependent variable. Additionally, the effectiveness of the training procedure affects the model's performance [29].

The LSTM technique provides excellent predicted accuracy, strong interpretation, and the ability to adjust to nonlinear data behavior. It can also maintain memory and forget states that take into account past information. However, there are disadvantages to LSTM, including challenging data gathering, intricate modeling, and time-consuming corrections. SVR has the benefit of swiftly constructing representative trees and providing users with a visual representation of the model's cuts, but it also has the disadvantage of being less accurate than other predictive analytic techniques when working with highly variable, sparse, or discontinuous data [30]. This study will be innovative in its ability to forecast the price of crude palm oil by integrating LSTM and GRU approaches. While GRU has a simpler structure but converges more slowly, LSTM enables powerful modeling but is constrained by poor data and interpretation. The modeling presented in the Liu et al. [10] study on forecasting stock prices using the LSTM and GRU approaches demonstrated high prediction accuracy and a rapid rate of convergence.

3- Methods

3-1-Research Framework

The main idea emerges from the complexity of the dynamics of the palm oil market [6], influenced by several external factors such as climate change, trade policies, and global economic factors. Previous research indicates that traditional methods often prove ineffective in capturing these complex patterns. Therefore, we chose Deep Learning methods due to their ability to handle time series data and address intricate temporal relationships.

In the literature review, this research details various forecasting methods previously employed and highlights their weaknesses in handling the fluctuating dynamics of the market. Literature review results demonstrate that Deep Learning methods, particularly GRU and LSTM, have shown good performance in price forecasting across various fields, including finance and economics. Hence, the selection of GRU and LSTM is based not only on their popularity but also on their proven ability to handle complex price forecasting challenges. The data used in this study is sourced from reliable outlets and covers a significant time series. Data preprocessing is a crucial initial stage, involving data cleaning and normalization to ensure the accuracy and reliability of the model. The GRU and LSTM model structures are discussed in detail, including the number of layers, neurons, and activation functions chosen to maximize forecasting performance.

Implementation and experiments are conducted by dividing the data into training and testing subsets. Model parameters are adjusted and optimized to achieve optimal results. Experiment results are evaluated using established metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-Squared value, and Mean Absolute Percentage Error (MAPE) to assess forecasting accuracy.

Additionally, comparisons are made with other forecasting methods to demonstrate the superiority of GRU and LSTM in the context of crude palm oil price forecasting. The main conclusions and practical implications of this research are outlined, providing a significant contribution to enhancing the understanding and forecasting capabilities of crude palm oil prices, as well as laying the foundation for further research in the development of more sophisticated forecasting models.

3-2-Research Stages

This research was conducted to produce a low-error price prediction model for crude palm oil. The modeling in this study uses the deep learning method with the LSTM and GRU models. The framework in this study consists of several stages, such as the planning stage, the implementation stage, and the testing stage (Figure 4). These stages are further divided into several components. Each of these components plays an important role, so they must be done sequentially before getting the desired final model.



Figure 4. Research Framework

The earliest step of this research is the planning phase. The construction of the best model and understanding of the GRU and LSTM techniques for predicting the price of crude palm oil are involved in this stage's presentation. Utilizing the previously discussed ideas, a literature study is the following stage. To achieve this, locate and compile all of the research publications on crude palm oil price forecasts that have been published in the last five years. Opportunities that had not previously been explored in research on crude palm oil utilizing the combined LSTM and GRU approaches were recognized from the findings of the literature analysis that was undertaken.

The phase of implementation comes next. This stage, which contains a number of procedural components, is the foundation of the study. Finding a dataset on the cost of crude palm oil from August 1, 2018 to August 31, 2023 is the first stage. Before utilizing the model, the crude palm oil price data is preprocessed after being collected to ensure accuracy. To get the best outcomes, deep learning (LSTM and GRU) applications come next. The outcomes of the modeling approach are then evaluated and examined.

The project's last step is the model testing phase. After the model has been created, it is tested using data on the price of crude palm oil. If the test results indicate that a certain model has the lowest mean absolute error (MAE), root mean squared error (RMSE), R value, and mean absolute percentage error (MAPE) of all the models evaluated, that model will be picked and put up against other models in comparison. The goals of this research have been met if it outperforms earlier investigations.

3-3-Data Collection

The data set for crude palm oil utilized in this study was obtained from https://bappebti.go.id/harga_komoditi_bursa (Indonesian Ministry of Trade). A data set with 3,653 data points may be found in Table 1. The time range covered is from August 1, 2018, to August 31, 2023. There are four datasets used: Medan, Rotterdam SPOT (date and delivery in the same month), Rotterdam Forward 1 Month (date and delivery one month forward), and Rotterdam Forward 2 Months (date and delivery two months forward), each processed separately in different files. The Medan dataset consists of 1,064 entries, Rotterdam SPOT has 617 entries, and Rotterdam Forward 1 Month and 2 Months have 1,022 and 950 entries, respectively. Each portion of the data, training, and testing has properties like Date, Contract, Trade Type, Location, Price, and Unit (see Table 2). Date is the cost of daily crude palm oil. Contract is the selected commodity. A trade type is a type of commodity trading. Location is the CPO price in the chosen location. Price is the price of crude palm oil that was opened in one day. The unit is the unit of measurement used for CPO. The data is divided into two categories: training and testing.

Location	Time Interval	Amount of Data
Medan	1 August 2018-31 August 2023	1.064
Rotterdam (SPOT)	1 August 2018-31 August 2023	617
Rotterdam (Forward 1 Month)	1 August 2018-31 August 2023	1.022
Rotterdam (Forward 2 Months)	1 August 2018-31 August 2023	950
Total		3.653

Table 1. Data Collection of Crude Palm Oil Prices

Table 2. Example of Crude Palm Oil Price Data Set

Date	Contract	Trade Type	Location	Price	Unit
02/01/19	CPO	Jan19	Rotterdam	507,5	US\$/MT
02/01/19	CPO	Feb19	Rotterdam	525	US\$/MT
02/01/19	CPO	Mar19	Rotterdam	540	US\$/MT
02/01/19	CPO	SPOT	Medan	6311,84	IDR/Kg
09/01/19	CPO	Jan19	Rotterdam	530	US\$/MT
09/01/19	CPO	Feb19	Rotterdam	540	US\$/MT
09/01/19	CPO	Mar19	Rotterdam	560	US\$/MT

3-4-Data Preprocessing

The pre-processing stage involves the data separation process. The data separation process involves separating the data based on the prices of CPO. The data for CPO prices in Medan and Rotterdam is separated. The Rotterdam data is further divided into SPOT data (date and trade type in the same month), forward 1 month (date and trade type for one-month futures), and forward 2 months (date and trade type for two-month futures). The "trade type" column listed with two different months is not used, as it would introduce data bias. The separation of Rotterdam data is done to ensure that each date has only one price, eliminating any instances of different prices for the same date. Columns such as contract, trade type, and location can be omitted during processing since the data processing for each dataset will be carried out in separate documents. The dependent variable used in this research is the price data. Table 3 illustrates an example of the Rotterdam SPOT dataset after pre-processing, where each date contains only one price data entry.

Table 3. The pre-processed Rotterdam SPOT dataset

Date	Price	Unit
02/01/19	507,5	US\$/MT
02/01/19	525	US\$/MT
02/01/19	540	US\$/MT
02/01/19	6311,84	IDR/Kg
09/01/19	530	US\$/MT
09/01/19	540	US\$/MT
09/01/19	560	US\$/MT

The data undergoes a normalization process. One of the most well-known methods for standardizing information is the Min-Max Scaler technique. The most extreme value for each component is changed to a 1, the base estimation of that element is changed to a 0, and every other value is altered to a decimal between 0 and 1 [31]. The purpose of data normalization is to avoid the influence of dominant patterns in price behavior that have higher values over smaller price features. Using the same range of values ensures adherence to the actual patterns of price behavior that generally prevail or occur in the market [23]. The Min-Max normalization method is applied to ensure that prices follow actual price patterns. The use of the Min-Max normalization method guarantees that all price values will have exactly the same scale and a significant reduction [32]. This rule applies to all crude palm oil price data obtained.

$$X_{Scaled} = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} \tag{11}$$

where x_{max} is the highest value for X Feature and x_{min} is its minimum value [30].

3-5-Prediction of CPO Using the GRU and LSTM Methods

Data preparation is carried out to separate inaccurate data and standardize all data by making sure that they are all the same size in order to decrease the possibility of mistakes throughout the training phase (Figure 5). The dataset will then

be split into training data and testing data groups. The tools used for running the LSTM and GRU methods are from Google Colab. Regarding the hyperparameters under consideration, we define a set of parameters as follows: learning rate = $\{0.001, 0.01, and 0.1\}$, nodes = $\{64, 216, and 512\}$, and optimizer types such as $\{Adam, SGD, and RMSProp\}$. The most optimal parameters used are: (a) learning rate: (0.001); (b) batch size: 100; (c) nodes: (512); (d) Optimizer: Adam; (e) and epochs: 50. Based on [12, 23]), the best learning rate is 0.001 with the Adam optimizer. The best model uses a batch size of 100 [23]. Table 4 shows the hyperparameters of the LSTM, GRU, and Bi-LSTM models for comparison. Additionally, the comparison methods used include ARIMA (1, 1, 5) according to Wei & Samsudin [14], LSTM, Simple Recurrent Neural Network (RNN), and ARIMA (2,2,2) according to Ofuoku & Ngniatedema [34], as well as Support Vector Regression (SVR) according to Saadah et al. [13]. The data processing for ARIMA is done using Orange software.



Figure 5. Data Processing Architecture Crude Palm Oil Price Prediction with LSTM and GRU

Table 4. Hyperparameters of LSTM, GRU, and B1-LSTM modelli	ad Bi-LSTM modelling
--	----------------------

Learning Rate	0.0010
Batch Size	100
Optimizer	Adam
Epochs	50
Units	512 (LSTM & GRU); 1024 (Bi-LSTM)

3-6-Evaluation and Analysis

The mean absolute error (MAE), root mean squared error (RMSE), R value, and mean absolute percentage error (MAPE) were used to calculate errors in this study. The four evaluation standards are described as follows:

1. Mean Absolute Error (MAE)

The mean absolute error (MAE) between actual and anticipated values is referred to. If p stands for the original power and \hat{p} for the estimated power, then MAE may be mathematically represented as [35]:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |p - \hat{p}|$$
(12)

2. Root Mean Square Error (RMSE)

A formula for computing polynomials is RMSE. This phrase refers to an error index that is utilised to evaluate different forecasting models. What is being measured is the squared difference between actual observations and forecasts [28].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_i - A_i)^2}$$
(13)

where n is the sample size, P is the projected value, A is the actual value [36].

3. R value

The relationship between the actual and expected values is expressed in terms of the correlation coefficient's value, or R-value. If R is 1, there is a strong connection; if R is 0, there is a chance connection [37].

$$R = \frac{n(\Sigma PA) - (\Sigma P)(\Sigma A)}{\sqrt{(n(\Sigma P^2 - (\Sigma P)^2 [n \Sigma A^2 - (\Sigma A)^2])}}$$
(14)

4. Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100\%}{N} \sum_{i=1}^{n} \left| \frac{y_{i'} - y_{i}}{y_{i}} \right|$$
(15)

where y_i' denotes the price's expected value, y_i denotes the price's actual value, and N denotes the number of predicted samples [37].

Many machine learning models are now in use. The model evaluation is determined by measuring the mean absolute error (MAE), root mean squared error (RMSE), R value, and mean absolute percentage error (MAPE) using a different set of data from the training data set. This study's testing phase, which assesses how well the predictions held true, is the last stage. The next stage is to conduct testing by displaying the outcomes of plotting a sample of test data to compare the outcomes of the projected values with the actual data values. The optimum model is chosen from a variety of available models once each machine learning model has through the data training procedure. We will compare the findings of the GRU and LSTM data analysis methods with those of the Bi-LSTM method.

When comparing forecasting methods applied to a time series or multiple time series with the same units, MAE is popular because it is easy to understand and calculate. Forecasting methods that minimize MAE will result in median forecasts, while minimizing RMSE will result in mean forecasts. MAPE has the advantage of being unitless, making it often used for comparing forecasting performance between datasets [16]. The R-value indicates the inconsistency of dependent parameters explained by independent parameters [35].

4- Result

4-1-Pre-processing and Data Characteristics

The statistics for crude palm oil prices are separated using the trading locations of Medan and Rotterdam. The Rotterdam dataset is additionally pre-processed into SPOT, Forward 1 Month, and Forward 2 Months segments. The same month, day, and type of trade are indicated by SPOT data. The data for Forward 1 Month is represented using a month trade type that is moved from the date. The data is displayed with a two-month trade type that has been pushed from the date in the more likely forward 1-month dataset, Forward 2 Months. Figures 6 to 9 display the price distribution of crude palm oil in Medan and Rotterdam (SPOT, Forward 1 Month, and Forward 2 Months), respectively. From 2018 to 2023, the price of crude palm oil in Medan will typically vary between Rp 5,000 and Rp 10,000 per kg. Rotterdam's price (SPOT, forward 1 month, and forward 2 months) is often between US\$1.200 per MT.



Figure 6. Price distribution of Crude Palm Oil in Medan (IDR/Kg)



Figure 7. Price distribution of Crude Palm Oil in Rotterdam Forward 1 Month (US\$/MT)



Figure 8. Price distribution of Crude Palm Oil in Rotterdam SPOT (US\$/MT)



Figure 9. Price distribution of Crude Palm Oil in Rotterdam Forward 2 Month (US\$/MT)

Table 5 presents statistical data for each dataset. Each dataset provides statistical measures, offering an overview of the characteristics of their respective markets. The "Medan" dataset indicates a mean price of 13,957.12, with a median value of 14,635.88. The range from 0 to 25,385.91 shows a significant price variation. The variance, at 33,364,316.52, and the standard deviation, at 5,776.19, further emphasize price dispersion, providing insights into potential fluctuations.

Datasets	Mean	Median	Range	Variance	Standard Deviation
Medan	13.957,12	14.635,88	25.385,91	33.364.316,52	5.776,19
Rotterdam SPOT	976,04	990,00	1.545,0	123.570,37	351,53
Rotterdam Forward 1 Month	948,22	977,50	1.535.00	118.124,11	343.69
Rotterdam Forward 2 Months	926,72	970,00	1.447,5	105.401,64	324,66

Similarly, the "Rotterdam SPOT" dataset shows an average price of 976.04, with a median of 990.00. A narrower range, i.e., 1,545.0, indicates less variation compared to the "Medan" dataset. Its variance (123,570.37) and standard deviation (351.53) emphasize the extent of price dispersion in this market. The "Rotterdam Forward 1 Month" and "Rotterdam Forward 2 Months" datasets exhibit similar patterns. Despite having slightly lower means (948.22 and 926.72, respectively) and lower medians (977.50 and 970.00, respectively), both datasets show relatively narrow ranges (1,535.00 and 1,447.5, respectively). Their corresponding variances (118,124.11 and 105,401.64) and standard deviations (343.69 and 324.66) highlight the extent of price dispersion for these forward contracts.

Overall, the statistical analysis of these datasets provides a comprehensive understanding of price dynamics in the examined locations. Mean and median offer insights into central tendencies, while range, variance, and standard deviation explain the extent of price variation. These findings provide valuable information for stakeholders seeking to understand and navigate the complexities of these diverse markets.

4-2-Results of Training and Testing Model

After being separated, the data is normalized using the Min-Max Scaler method. Then, the data is split into training and testing in the ratio 70:30. The data distribution for the train and test data for the price of crude palm oil in Medan and Rotterdam (SPOT, forward 1 Month, and Forward 2 Months) is shown in Figures 10 to 13. The data is processed using the LSTM and GRU algorithms once the training and testing sets of data have been obtained, with Bi-LSTM being utilized for comparison. Similar to the stock price data in Liu et al. [10], the dataset of crude palm oil prices is time series data, thus processed using LSTM and GRU methods.



Figure 10. Data distribution for the train and test for Crude Palm Oil in Medan



Figure 11. Data distribution for the train and test for Crude Palm Oil in Rotterdam Forward 1 Month



Figure 12. Data distribution for the train and test for Crude Palm Oil in Rotterdam SPOT



Figure 13. Data distribution for the train and test for Crude Palm Oil in Rotterdam Forward 2 Month

The optimizer used to make the model more optimal is Adam, with a learning rate of 0.001. The batch size used is 100, and the number of epochs is set to 50. Ghane et al. [38] performed the best modeling of the Recurrent Neural Network (RNN) for the elastoplasticity of woven composites with a learning rate of 0.001 and the Adam optimizer, using 512 units for both LSTM and GRU. Weigelt et al. [39] demonstrated that a batch size of 100 results in the best modeling for intelligent systems.

Data from Medan and Rotterdam are analyzed using the GRU technique, producing a model architecture with two GRU layers and two dense layers. 512 GRU units make up the first layer, which is set to produce sequential outputs (return_sequences=True) and take shape input (xTrain.shape[1], 1). It acts as the model's input layer. The second layer, a 512-unit GRU layer, however, does not provide sequential outputs (return_sequences=false). As the first dense layer in the model, the third layer, which has 25 units, is dense. The last output layer of the model is the fourth layer, which is another dense layer with 1 unit. The last output layer of the model is the fourth layer, which is another dense layer with 1 unit. The model has 2,379,827 total parameters, which will be modified during training using the crude palm oil price dataset.

The analysis of data from Medan and Rotterdam using the Bi-LSTM approach produces a model architecture with two Bi-LSTM layers and two dense layers. A Bi-LSTM layer with 512 units makes up the top layer. Its setup allows input of shape (xTrain.shape[1], 1]) and needs sequential outputs (return_sequences=True). This acts as the model's input layer. The second layer, on the other hand, is a second 512-unit Bi-LSTM layer, but it does not provide sequential outputs (return_sequences=false). As the first dense layer in the model, the third layer, which has 25 units, is dense. The last output layer of the model is the fourth layer, which is another dense layer with 1 unit.

The data is processed using the LSTM and GRU algorithms once the training and testing sets of data have been obtained, with Bi-LSTM being utilized for comparison. The comparison of actual data and anticipated crude palm oil prices in Medan, processed using the LSTM, GRU, and Bi-LSTM algorithms, is shown in Figures 14 to 16. Those figures show that the real graph (blue line) is closely related to the predicted graph (orange line). To learn more about the evaluation, we need to calculate the error using MAE, MAPE, RMSE, and R-squared.



Figure 14. Comparison of the LSTM method's anticipated and real pricing for palm oil in Medan (IDR/Kg)



Figure 15. Comparison of the GRU method's anticipated and real pricing for palm oil in Medan (IDR/Kg)



Figure 16. Comparison of the Bi-LSTM method's anticipated and real pricing for palm oil in Medan (IDR/Kg)

The palm oil dataset in Rotterdam underwent comparable procedures. The outcomes of the Rotterdam SPOT dataset processing are displayed in Figures 17 to 19. The actual (blue) and estimated (orange) graphs are closely related, yielding findings that are somewhat consistent with the previous dataset.



Figure 17. Rotterdam's SPOT actual and anticipated prices for palm oil are compared using the LSTM technique (US\$/MT)



Figure 18. Rotterdam's SPOT palm oil price comparison between actual and forecasted prices using the GRU technique (US\$/MT)



Figure 19. Rotterdam's SPOT actual and forecasted prices of palm oil were compared using the Bi-LSTM approach (US\$/MT)

Figures 20 to 22 show the results of Rotterdam Forward 1 Month Data Processing. The real and projected plots are close to one another, just as in the prior Rotterdam dataset. Additionally, for the Rotterdam Forward 2 Month data processing shown in Figures 23 to 25. There is a much closer graph of real and predicted data than in the dataset before.



Figure 20. Rotterdam's Forward 1 Month actual and anticipated prices for palm oil are compared using the LSTM technique (US\$/MT)



Figure 21. Rotterdam's Forward 1 Month actual and anticipated prices for palm oil are compared using the GRU technique (US\$/MT)



Figure 22. Rotterdam's Forward 1 Month actual and anticipated prices for palm oil are compared using the BiLSTM technique (US\$/MT)



Figure 23. Rotterdam's Forward 2 Month actual and anticipated prices for palm oil are compared using the LSTM technique (US\$/MT)



Figure 24. Rotterdam's Forward 2 Month actual and anticipated prices for palm oil are compared using the GRU technique (US\$/MT)



Figure 25. Rotterdam's Forward 2 Month actual and anticipated prices for palm oil are compared using the BiLSTM Model Performance Comparison and Discussion technique (US\$/MT)

4-3-Model Performance Comparison and Discussion

To further study the Medan data evaluation, we need to calculate errors using MAE, MAPE, RMSE, and R-squared (Table 6). MAPE below 10% indicates high accuracy; 11–20% indicates good forecasting; 21–50% indicates reasonable forecasting; and over 51% indicates inaccurate forecasting [7]. Table 5 presents the evaluation of the Medan model. With the hyperparameters used by the tester, GRU still has the lowest MAPE compared to LSTM and Bi-LSTM. However, when compared to similar literature, the lowest MAPE for the Medan data is achieved with the LSTM method version [34], followed by the GRU method (3.0547%) and LSTM (3.9412%) versions by the author. MAPE reaches 36.6571% and 4.1200% when we apply the ARIMA (2,2,2) approach and RNN by Ofuoku & Ngniatedema [34] to the Medan data. The ARIMA-type approach (1, 1, 5) produces a MAPE of 35.9599% when applied to estimate CPO prices [14]. To predict CPO in Indonesia with the lowest MAPE, Saadah et al. [13] studied the support vector regression approach. The SVR MAPE result is 47.3958%. In short, LSTM and GRU approaches remain the best ways to anticipate CPO prices for the Medan dataset.

Table 6. MAE, R	RMSE, R-S	Squared, and	I MAPE of	Crude Palm	Oil Price in Medan
-----------------	-----------	--------------	-----------	------------	--------------------

Method	MAE	RMSE	R-squared	MAPE (%)
• LSTM	667.6389	896.7875	0.9425	3.9412
• LSTM [34]	429.8600	568.2900	0.9100	2.7800
• GRU	531.0581	718.0000	0.9631	3.0547
• Bi-LSTM	878.4119	1,108.9177	0.9120	5.2493
• ARIMA (1,1,5) [14]	4,700.3681	5,817.1785	-0.011	35.9599
• ARIMA (2,2,2) [14]	4,898.0502	5922.1584	-0.0476	36.6571
• Simple RNN [11]	615.5000	808.4800	0.8100	4.1200
• SVR [13]	4,848.7671	5,807.5774	-0.0109	47.3958

The evaluation of each model is presented in the table below, making it suitable for use in forecasting this data (Table 7). The LSTM, GRU, and Bi-LSTM methods have MAPE below 10%, specifically 4.2732%, 3.0493%, and 4.5689%, which, according to Amin et al. [7], indicates high accuracy. With the hyperparameters used by the author, the GRU method still leads compared to LSTM and Bi-LSTM. The proposed tester's method is also compared with other literature. The lowest MAPE is achieved by the ARIMA (1,1,5) method version by Wei & Samsudin [14], which is 2.4500%, followed by ARIMA (2,2,2) and LSTM versions by Ofuoku & Ngniatedema [34] at 2.5126% and 2.8500%, respectively. The RNN method by Ofuoku & Ngniatedema [34] and SVR by Saadahet al. [13] are also used for processing the Rotterdam (SPOT) dataset. The MAPE of these RNN and SVR methods reaches 3.8100% and 34.9876%, respectively. The Medan and Rotterdam SPOT datasets show clear differences. The Medan dataset receives higher errors when processed with the ARIMA method, while the Rotterdam SPOT dataset receives smaller errors. Compared to GRU, the ARIMA (1,1,5) and ARIMA (2,2,2) algorithms provide lower errors, making them suitable for forecasting this data as well.

Method	MAE	RMSE	R-squared	MAPE (%)
•LSTM	52.1434	74.6622	0.9431	4.2732
•LSTM [34]	436.2500	578.0000	0.9000	2.8500
●GRU	37.5861	57.7891	0.9659	3.0493
•Bi-LSTM	55.4324	77.5116	0.9387	4.5689
•ARIMA (1,1,5) [14]	14.0376	43.5947	0.9846	2.4500
•ARIMA (2,2,2) [14]	14.4271	45.4282	0.9833	2.5126
•Simple RNN [11]	38.5600	50.8500	0.3900	3.8100
•SVR [13]	279.4392	343.3406	0.0460	34.9876

Table 7. MAE, RMSE, R-Squared, and MAPE of Crude Palm Oil Price in Rotterdam (SPOT)

In Table 8, model evaluation is presented for the Rotterdam Forward 1 Month dataset. With the hyperparameters used by the author, GRU still has the lowest MAPE (2.4221%) among LSTM, GRU, and Bi-LSTM, all followed by LSTM (3.3865%) and Bi-LSTM (3.4881%). The results are also compared with those from other literature, where ARIMA (1,1,5) [14] has the lowest MAPE at 2.2928%, followed by ARIMA (2,2,2) version [34] at 2.3601%, and the GRU method used by the author. When tested with the LSTM method and RNN version [34], it results in MAPEs of 2.7900% and 2.5300%, respectively. Meanwhile, testing with SVR version [13] yields a MAPE of 34.8624%. Based on the testing, this dataset is suitable for use with the GRU and ARIMA methods.

Table 8. MAE, RMSE, R-Squared, and MAPE of Crude Palm Oil Price in Rotterdam (Forward 1 Month)

Method	MAE	RMSE	R-squared	MAPE (%)
• LSTM	40.6554	67.4243	0.9467	3.3865
• LSTM [34]	426.2100	568.0700	0.9100	2.7900
• GRU	29.1776	53.2126	0.9668	2.4221
• Bi-LSTM	40.7634	65.3825	0.9499	3.4881
• ARIMA (1,1,5) [14]	11.7408	45.2919	0.9826	2.2928
• ARIMA (2,2,2) [14]	11.7696	46.7696	0.9815	2.3601
• Simple RNN [11]	25.1100	33.5600	0.6900	2.5300
• SVR [13]	271.3113	336.6703	0.0404	34.8624

The evaluation was conducted to observe each method for the Rotterdam Forward 2 Months dataset. The evaluation results are presented in Table 9. This dataset yields the best Mean Absolute Percentage Error (MAPE) for GRU, which is 2.1659%, followed by LSTM (2.7411%) and Bi-LSTM (3.4405%) when tested with the hyperparameters used by the examiner. The testing results are also compared with other literature. The lowest MAPE is achieved by the LSTM method version [34], which is 1.7500%, followed by ARIMA (1,1,5) [14, 34] at 1.9623%, and ARIMA (2,2,2) version [34] at 2.0069%. Testing the dataset with RNN version [34] and SVR version [13] methods resulted in consecutive MAPE values of 2.4800% and 33.3810%, respectively. In addition to GRU and LSTM, this dataset also allows for low error rates with the ARIMA method.

Method	MAE	RMSE	R-squared	MAPE (%)
•LSTM	32.8323	45.5708	0.9708	2.7411
•LSTM [34]	17.6600	23.8200	0.8500	1.7500
• GRU	26.1072	38.6758	0.9790	2.1659
•Bi-LSTM	40.4446	54.7845	0.9578	3.4405
•ARIMA (1,1,5) [14]	10.7053	33.8785	0.9891	1.9623
• ARIMA (2,2,2) [14]	10.5755	35.2762	0.9882	2.0069
• Simple RNN [11]	24.9500	31.6200	0.7400	2.4800
•SVR [13]	255.8624	318.9873	0.0346	33.3810

Table 9. MAE, RMSE, R-Squared, and MAPE of Crude Palm Oil Price in Rotterdam (Forward 2 Months)

This journal collectively indicates that the Gated Recurrent Unit (GRU) method performs better than the Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM) methods for price forecasting, based on the Mean Absolute Percentage Error (MAPE) metric [40]. Hamayel & Owda [41] focused on predicting cryptocurrency prices and concluded that GRU outperforms LSTM and Bi-LSTM in predicting the prices of Bitcoin, Litecoin, and Ethereum. Busari & Lim [42] compared AdaBoost-LSTM and AdaBoost-GRU models for crude oil price prediction and found that both models outperformed benchmarking models, with AdaBoost-GRU being superior. Ubrani & Motwani [43] examined LSTM and GRU models for forecasting the Market Clearing Price (MCP) in the electricity market and determined that both models perform well, with LSTM slightly outperforming GRU. In summary, the paper collectively suggests that GRU is the best method for price forecasting, followed by LSTM and Bi-LSTM, when comparing the MAPE metric.

The paper collectively indicates that the relative performance of time series forecasting methods such as ARIMA, LSTM, and GRU can depend on several factors and vary from dataset to dataset. While ARIMA may perform better in some specific cases based on Mean Absolute Percentage Error (MAPE), LSTM and GRU can still outperform ARIMA in many other cases for various reasons. First, LSTM and GRU, as part of the deep learning family, have a higher capacity to capture complex non-linear relationships in the data and adapt to various patterns, trends, and seasonality that ARIMA may struggle with, according to Siami-Namini et al. [17]. Second, LSTM and GRU are highly effective when dealing with large and complex datasets, especially those with high-dimensional inputs, as neural networks can effectively leverage this data [44]. Third, LSTM and GRU are designed to handle sequential data and can capture longterm dependencies and remember past information, which is crucial for time series forecasting. On the other hand, ARIMA, as a linear model, may not effectively capture complex sequential patterns [17]. This is supported by Wu et al. [45], where the ARIMA model shows significant advantages in short-term predictions but has more weaknesses in the long term, making it not the best choice. Atique et al. [46] explain that the simplicity of the ARIMA model lies in its applicability to stationary time series data. Therefore, for seasonal and non-stationary time series data, a transformation to stationarity is needed before applying the ARIMA model. This simplicity arises from the assumption of a linear correlation between past and present time series values, which is a major drawback of the ARIMA model, although it can model various types of time series data. From the explanations above, it can be concluded that if forecasting with data tends to be non-stationary, long-term, and seasonal, deep learning methods like GRU and LSTM are recommended over ARIMA.

4-4-Managerial Implication

The application of Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) models to forecast crude palm oil (CPO) prices has benefits for the industry. The dynamic expectation of future prices will enhance decision-making at various levels [47]. The need to predict the price of crude palm oil is crucial for improving the implementation of business intelligence [48]. Accurately forecasting CPO prices using these models provides managers with valuable insights into future market conditions, facilitating well-informed decision-making regarding inventory management, production planning, and overall business strategy. Additionally, the proposed LSTM-based prediction method is a useful and reliable machine learning technique that can provide valuable information to businesses, industries, and government institutions [34]. The utilization of advanced forecasting models enables industry managers to identify and assess potential risks associated with CPO price fluctuations, leading to the development of effective risk mitigation strategies. Supply chain optimization is another crucial consideration, as accurate forecasts align production levels with anticipated demand, resulting in more efficient inventory management and overall supply chain performance. Reliable CPO price predictions contribute to better budget and financial planning, allowing for more efficient resource allocation and realistic financial target setting. Timely and accurate forecasts enable industry players to position themselves strategically in the market, adjusting marketing strategies and pricing based on anticipated price trends, thereby gaining a competitive advantage.

LSTM-based predictions can assist palm oil growers worldwide in decision-making regarding the management and operational processes of palm oil plantations [49]. The implementation of GRU and LSTM models can enhance operational efficiency by aligning production schedules with expected market demand, ultimately reducing waste and optimizing resource utilization. Anticipating market trends through accurate forecasts also contributes to customer satisfaction by ensuring product availability and avoiding stockouts. Continuous monitoring facilitated by these models allows for quick adaptation to changes in market conditions, helping the industry remain agile and responsive to new trends. Supported by accurate forecasts, managers can make better investment decisions related to infrastructure, technology, and capacity expansion, ensuring alignment with market needs and future demand projections. The implementation of sophisticated forecasting models may require cross-functional collaboration within the organization, necessitating cooperation between data scientists, analysts, and industry experts for the successful development, implementation, and interpretation of model results. It is crucial for industry managers to be aware that while GRU and LSTM models offer strong forecasting capabilities, a holistic approach involving data quality assurance, model validation, and continuous monitoring is required for successful implementation. Additionally, effective communication and collaboration between different departments are essential to reaping all the managerial benefits of these forecasting tools.

5- Conclusion

Crude palm oil is a commodity that contributes to many sectors, making it important to determine the best prediction method to accurately estimate its price. This research employs LSTM, GRU, and Bi-LSTM methods as comparative techniques. The forecasting technology used is Google Colab. Data will be collected from August 1, 2018, to August 31, 2023, to predict future prices. The data is then divided into training and testing sets. Subsequently, the training and testing sets are created from the data. The dataset is categorized according to the trading centers of Medan and Rotterdam, collected from the Indonesian Ministry of Trade. The Rotterdam data is divided into SPOT, Forward 1 Month, and Forward 2 Months. Each method produces four layers (two layers for the LSTM/GRU/Bi-LSTM methods and two dense layers). Each of the two layers in these methods has 512 units, while the two dense layers have 25 and 1 unit. The optimizer used for each method is Adam.

The predicted data is then examined for accuracy. The forecast results show MAPE values below 10%, representing highly accurate predictions. The best hyperparameters are a learning rate of 0.001, the Adam optimizer, 50 epochs, and 512 units using the GRU method. However, ARIMA may have lower errors than GRU and LSTM, but it also depends on the complexity, dimensions, and nature of the data. LSTM and GRU, with their deep learning capabilities, are often strong contenders, especially in scenarios where data exhibits complex and non-linear patterns and when dealing with large, high-dimensional datasets. However, ARIMA may still have its place in cases where the data follows more linear patterns or when the dataset is relatively simple. Overall, it can be concluded that LSTM and GRU methods can be used to forecast crude palm oil prices, but both need to be tailored to the characteristics of the data.

Suggestions for future research include optimizing parameters with specific algorithms. Additionally, future research can modify the forecasting structure to be more efficient and optimal. Further development in subsequent studies can also involve the design of a decision-making visualization system for predicting results using modern and more accurate tools.

6- Declarations

6-1-Author Contributions

Conceptualization, G.A. and S.S.; methodology, G.A.; software, G.A.; validation, G.A. and S.S.; formal analysis, G.A.; investigation, G.A.; resources, G.A.; data curation, G.A.; writing—original draft preparation, G.A.; writing—review and editing, G.A.; visualization, G.A.; supervision, S.S.; project administration, G.A.; funding acquisition, S.S. All authors have read and agreed to the published version of the manuscript.

6-2-Data Availability Statement

Publicly available datasets were analyzed in this study. This data can be found here: [https://bappebti.go.id/harga_komoditi_bursa].

6-3-Funding

This research receive grant from the Directorate General of Higher Education Ministry of Education, Culture, Research, and Technology together with the Education Fund Management Institute with contract number 179/E5/PG.02.00.PL/2023.

6-4-Acknowledgements

The authors would like to express their gratitude to BINUS University for all their support.

6-5-Institutional Review Board Statement

Not applicable.

6-6-Informed Consent Statement

Not applicable.

6-7- Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

- [1] FAS. (2024). Production-Palm Oil. U. S. Department of Agriculture, Foreign Agriculture Service, Washington, United States. Available online: https://fas.usda.gov/data/production/commodity/4243000 (accessed on May 2024).
- [2] Ayompe, L. M., Schaafsma, M., & Egoh, B. N. (2021). Towards sustainable palm oil production: The positive and negative impacts on ecosystem services and human wellbeing. Journal of Cleaner Production, 278, 123914. doi:10.1016/j.jclepro.2020.123914.
- [3] Achaw, O.-W., & Danso-Boateng, E. (2021). Chemical and Process Industries. Springer International Publishing, Cham, Switzerland. doi:10.1007/978-3-030-79139-1.
- [4] Tanwar, B., & Goyal, A. (2021). Oilseeds: Health Attributes and Food Applications. Springer, Singapore. doi:10.1007/978-981-15-4194-0.
- [5] Al-Sabaeei, A. M., Al-Fakih, A., Noura, S., Yaghoubi, E., Alaloul, W., Al-Mansob, R. A., Imran Khan, M., & Aliyu Yaro, N. S. (2022). Utilization of palm oil and its by-products in bio-asphalt and bio-concrete mixtures: A review. Construction and Building Materials, 337, 127552. doi:10.1016/j.conbuildmat.2022.127552.
- [6] Lim, C. H., Lim, S., How, B. S., Ng, W. P. Q., Ngan, S. L., Leong, W. D., & Lam, H. L. (2021). A review of industry 4.0 revolution potential in a sustainable and renewable palm oil industry: HAZOP approach. Renewable and Sustainable Energy Reviews, 135(February 2020), 110223. doi:10.1016/j.rser.2020.110223.
- [7] Amin, F. A. M., Sukeri, N. A. S. M., Hasbullah, N., & Jamaludin, N. (2019). Forecasting Value at Risk for Malaysian Palm Oil using Monte Carlo Simulation. International Journal of Academic Research in Business and Social Sciences, 9(9), 1253–1260. doi:10.6007/ijarbss/v9-i9/6418.
- [8] Helbawanti, O., Permata Bunda, C. A., & Faqihuddin, F. (2022). Integrasi Pasar Berjangka Komoditi Crude Palm Oil (CPO) Indonesia Di Masa Pandemi Covid-19. Jurnal Agrica, 15(1), 61–74. doi:10.31289/agrica.v15i1.5401.
- [9] Siregar, K., & Syafarudin, A. (2020). Analysis of Crude Palm Oil (CPO) Production Volume and Price on Profitability and Its Impact on Stock Returns. International Journal of Engineering Technologies and Management Research, 6(7), 87–100. doi:10.29121/ijetmr.v6.i7.2019.419.
- [10] Liu, Y., Wang, Z., & Zheng, B. (2019). Application of Regularized GRU-LSTM Model in Stock Price Prediction. 2019 IEEE 5th International Conference on Computer and Communications (ICCC), Chengdu, China. doi:10.1109/iccc47050.2019.9064035.
- [11] Aisyah, S., Ulinnuha, N., & Hamid, A. (2023). Application of Extreme Learning Machine in Forecasting Crude Palm Oil Prices. KUBIK: Jurnal Publikasi Ilmiah Matematika, 7(2), 97–105. doi:10.15575/kubik.v7i2.20460. (In Indonesian).
- [12] Amal, I. (2021). Crude palm oil price prediction using multilayer perceptron and long short-term memory. Journal of Mathematical and Computational Science. doi:10.28919/jmcs/6680.
- [13] Saadah, S., Fakhira, Z.Z., & Hasna, H.Z. (2021). Support Vector Machine (SVM) To Predict Crude Oil Palm in Indonesia and Exchange Rate of EUR/USD. Journal of Computer Science and Informatics Engineering (J-Cosine), 5(1), 85–92. doi:10.29303/jcosine.v5i1.403.
- [14] Wei, K.Y., & Banu Samsudin, H. (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.
- [15] Shehu, U. E., Chow, T. Q., Hafid, H. S., Mokhtar, M. N., Baharuddin, A. S., & Nawi, N. M. (2019). Kinetics of thermal hydrolysis of crude palm oil with mass and heat transfer in a closed system. Food and Bioproducts Processing, 118, 187–197. doi:10.1016/j.fbp.2019.09.009.
- [16] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts, Melbourne, Australia.
- [17] Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, United States. doi:10.1109/bigdata47090.2019.9005997.

- [18] ArunKumar, K. E., Kalaga, D. V., Mohan Sai Kumar, C., Kawaji, M., & Brenza, T. M. (2022). Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. Alexandria Engineering Journal, 61(10), 7585–7603. doi:10.1016/j.aej.2022.01.011.
- [19] Yamak, P. T., Yujian, L., & Gadosey, P. K. (2019). A comparison between ARIMA, LSTM, and GRU for time series forecasting. ACM International Conference Proceeding Series, 49–55. doi:10.1145/3377713.3377722.
- [20] Luceño, A., & Peña, D. (2007). Autoregressive Integrated Moving Average (ARIMA) Modeling. Encyclopedia of Statistics in Quality and Reliability, 2. doi:10.1002/9780470061572.eqr276.
- [21] Liu, T., Liu, S., & Shi, L. (2020). ARIMA Modelling and Forecasting. Springer Briefs in Statistics, 61–85, Springer, Singapore. doi:10.1007/978-981-15-0321-4_4.
- [22] Alway, A., Zamri, N. E., Kasihmuddin, M. S. M., Asyraf Mansor, M., & Sathasivam, S. (2020). Palm oil trend analysis via logic mining with discrete hopfield neural network. Pertanika Journal of Science and Technology, 28(3), 967–981.
- [23] Aulia, A. A., Elhanafi, A. M., Dafitri, H., Aulia, A., Elhanafi, A. M., & Dafitri, H. (2021, October). Implementation of the Gated Recurrent Unit Algorithm in Predicting Palm Oil Prices Using the Recurrent Neural Network (RNN) Model. Prosiding SNASTIKOM: Seminar Nasional Teknologi Informasi & Komunikasi Paper, Indonesia.
- [24] Chai, L., Xu, H., Luo, Z., & Li, S. (2020). A multi-source heterogeneous data analytic method for future price fluctuation prediction. Neurocomputing, 418, 11–20. doi:10.1016/j.neucom.2020.07.073.
- [25] Wang, J., Niu, T., Du, P., & Yang, W. (2020). Ensemble probabilistic prediction approach for modeling uncertainty in crude oil price. Applied Soft Computing Journal, 95. doi:10.1016/j.asoc.2020.106509.
- [26] Lee, J. Y., Nguyen, T. T., Nguyen, H. G., & Lee, J. Y. (2022). Towards Predictive Crude Oil Purchase: A Case Study in the USA and Europe. Energies, 15(11). doi:10.3390/en15114003.
- [27] Sen, A. (2021). An Analysis of Crude Oil Prices in the Last Decade (2011-2020): With Deep Learning Approach. SSRN Electronic Journal. doi:10.2139/ssrn.3906517.
- [28] Suryani, D., Fadhilla, M. & Labellapansa, A. (2022). Indonesian Crude Oil Price (ICP) Prediction Using Multiple Linear Regression Algorithm. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 6(6), 1057–1063. doi:10.29207/resti.v6i6.4590.
- [29] Çolakoğlu, N., & Akkaya, B. (2019). Comparison of multi-class classification algorithms on early diagnosis of heart diseases. y-BIS 2019 Conference Book: Recent Advances in Data Science and Business Analytics, 25-28 September, 2019, Istanbul, Turkey.
- [30] Solano Meza, J. K., Orjuela Yepes, D., Rodrigo-Ilarri, J., & Cassiraga, E. (2019). Predictive analysis of urban waste generation for the city of Bogotá, Colombia, through the implementation of decision trees-based machine learning, support vector machines and artificial neural networks. Heliyon, 5(11), 2810. doi:10.1016/j.heliyon.2019.e02810.
- [31] Rahman, M., & Raju, V. (2020). Patients' perceptions on Factors Determining of Total Quality Management (TQM) Implementation into Bangladeshi Public Hospitals Using Technology: An Empirical Study. PalArch's Journal of Archaeology of Egypt/Egyptology, 17(7), 6258-6272.
- [32] Lawi, A., Mesra, H., & Amir, S. (2022). Implementation of Long Short-Term Memory and Gated Recurrent Units on grouped time-series data to predict stock prices accurately. Journal of Big Data, 9(1), 89. doi:10.1186/s40537-022-00597-0.
- [33] Andrea, I. P., Sitepu, B., & Simatupang, P. (2018). Analysis the Influence of Total Quality Management (TQM) and Joint Commission International (JCI) Toward Patient Safety Program A Case Study: Omni Hospital Group. Master Thesis, Swiss German University, Tangerang, Indonesia.
- [34] Ofuoku, M., & Ngniatedema, T. (2022). Predicting the Price of Crude Palm Oil. International Journal of Strategic Decision Sciences, 13(1), 1–15. doi:10.4018/ijsds.305830.
- [35] Shahid, F., Zameer, A., & Muneeb, M. (2021). A novel genetic LSTM model for wind power forecast. Energy, 223, 120069. doi:10.1016/j.energy.2021.120069.
- [36] Wang, Y., Liu, P., Zhu, K., Liu, L., Zhang, Y., & Xu, G. (2022). A Garlic-Price-Prediction Approach Based on Combined LSTM and GARCH-Family Model. Applied Sciences (Switzerland), 12(22), 11366. doi:10.3390/app122211366.
- [37] Namasudra, S., Dhamodharavadhani, S., & Rathipriya, R. (2023). Nonlinear Neural Network Based Forecasting Model for Predicting COVID-19 Cases. Neural Processing Letters, 55(1), 171–191. doi:10.1007/s11063-021-10495-w.
- [38] Ghane, E., Fagerström, M., & Mirkhalaf, M. (2023). Recurrent neural networks and transfer learning for elasto-plasticity in woven composites. arXiv preprint arXiv:2311.13434. doi:10.48550/arXiv.2311.13434.
- [39] Weigelt, S., Steurer, V., Hey, T., & Tichy, W. F. (2020). Roger that! Learning How Laypersons Teach New Functions to Intelligent Systems. 2020 IEEE 14th International Conference on Semantic Computing (ICSC). doi:10.1109/icsc.2020.00020.

- [40] Setiawan, K. E., Elwirehardja, G. N., & Pardamean, B. (2023). Indoor Climate Prediction Using Attention-Based Sequence-to-Sequence Neural Network. Civil Engineering Journal, 9(5), 1105-1120. doi:10.28991/CEJ-2023-09-05-06.
- [41] Hamayel, M. J., & Owda, A. Y. (2021). A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms. AI (Switzerland), 2(4), 477–496. doi:10.3390/ai2040030.
- [42] Busari, G. A., & Lim, D. H. (2021). Crude oil price prediction: A comparison between AdaBoost-LSTM and AdaBoost-GRU for improving forecasting performance. Computers & Chemical Engineering, 155, 107513. doi:10.1016/j.compchemeng.2021.107513.
- [43] Ubrani, A., & Motwani, S. (2019). LSTM- and GRU-based time series models for market clearing price forecasting of Indian deregulated electricity markets. Advances in Intelligent Systems and Computing, 898, 693–700. doi:10.1007/978-981-13-3393-4_70.
- [44] Iaousse, M., Jouilil, Y., Bouincha, M., & Mentagui, D. (2023). A Comparative Simulation Study of Classical and Machine Learning Techniques for Forecasting Time Series Data. International Journal of Online and Biomedical Engineering, 19(8), 56– 65. doi:10.3991/ijoe.v19i08.39853.
- [45] Wu, H., Chen, S., & Ding, Y. (2023). Comparison of ARIMA and LSTM for Stock Price Prediction. Financial Engineering and Risk Management, 6(1), 1–7. doi:10.23977/ferm.2023.060101.
- [46] Atique, S., Noureen, S., Roy, V., Subburaj, V., Bayne, S., & MacFie, J. (2019). Forecasting of total daily solar energy generation using ARIMA: A case study. 2019 IEEE 9th Annual Computing and Communication Workshop and Conference, CCWC 2019, 114–119. doi:10.1109/CCWC.2019.8666481.
- [47] Moitra, N., Raj, P., Saxena, S., & Kumar, R. (2020). Crude oil prediction using LSTM. International Journal of Innovative Science and Research Technology, 5(2), 900-905.
- [48] Al-Khowarizmi, Nasution, I. R., Lubis, M., & Lubis, A. R. (2020). The effect of a secos in crude palm oil forecasting to improve business intelligence. Bulletin of Electrical Engineering and Informatics, 9(4), 1604–1611. doi:10.11591/eei.v9i4.2388.
- [49] Kanchymalay, K., Salim, N., & Krishnan, R. (2019). Time series based crude palm oil price forecasting model with weather elements using LSTM network. International Journal of Engineering and Advanced Technology, 9(1), 3188–3192. doi:10.35940/ijeat.A9994.109119.