Available online at http://scik.org
J. Math. Comput. Sci. 11 (2021), No. 6, 8034-8045
https://doi.org/10.28919/jmcs/6680
ISSN: 1927-5307

CRUDE PALM OIL PRICE PREDICTION USING MULTILAYER PERCEPTRON AND LONG SHORT-TERM MEMORY

ICHLASUL AMAL, TARNO*, SUPARTI

Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro,

Jalan Prof. Soedarto S.H. Semarang 50275, Indonesia

Abstract: Crude Palm Oil is a leading commodity from Indonesia. Accurate prediction of Crude Palm Oil prices is very important to ensure future prices and help decision making. Study on crude palm oil prices is needed to anticipate fluctuations. In this study, prediction model was made using Multilayer Perceptron and Long Short-Term Memory. The optimization methods in this study are Stochastic Gradient Descent, Root Mean Square Propagation, and Adaptive Moment Estimation. The best model is selected based on Mean Square Error. Based on the results, Long Short-Term Memory model with Adaptive Moment Estimation optimization method is more optimal than Long Short-Term Memory with Stochastic Gradient Descent and Long Short-Term Memory with Root Mean Square Propagation. The prediction results using Long Short-Term Memory with Adam optimization show that the predicted value is not different from the actual value and Mean Absolute Percentage Error is 2.11%. This model has high forecasting accuracy because Mean Absolute Percentage Error is less than 10%.

Keywords: crude palm oil; prediction; artificial neural network; multilayer perceptron; long short-term memory.

2010 AMS Subject Classification: 62M10, 62M45.

---

*Corresponding author
E-mail address: tarno@lecturer.undip.ac.id
Received August 23, 2021

1. INTRODUCTION

The Agriculture, Forestry and Fisheries sector became the third largest sector in contributing to Indonesia's Gross Domestic Product (GDP) in 2019, with a contribution of 13.26%. Plantation crops contributed greatly to the sector by 25.7%. One of the plantation commodities is palm oil. This plant produces two types of palm oil, namely CPO and CPKO. CPO is more widely used when compared to CPKO (Hariyadi, 2014).

Indonesia is the world's largest producer and consumer of palm oil, producing 44 million tonnes of palm oil and consuming 14.27 million tonnes in 2019. Palm oil prices also fluctuate frequently. The historical price of palm oil from 2010 to October 2020 continues to change significantly. Over the past three years, crude palm oil prices have reached highs and lows. Forecasting of crude palm oil prices is done to predict price fluctuations. Accurate forecasting of Crude Palm Oil prices is very important to ensure future prices and assist decision making. Research on crude palm oil prices is needed to anticipate fluctuations. The uncertainty of crude palm oil prices and the data are non-stationary and non-linear, hence the Artificial Neural Networks (ANN) method is used to forecast crude palm oil prices.

Research by Karia, et al. (2013) concluded that ANN model provides the best results in predicting crude palm oil prices compared to Autoregressive Fractionally Integrated Moving Average (ARFIMA) model and Adaptive Neuro Fuzzy Inference System (ANFIS) model. Based on these results, ANN can be applied to forecasting of crude palm oil prices. ANN has the ability to study non-linear and complex relationships. ANN is useful in time series prediction. One of the ANN methods is Recurrent Neural Networks (RNN), which is designed to recognize patterns as data sequences, can be applied to predictions and forecasts, can also work on text, image, speech, and time series data. RNN has several variants, one of which is Long Short Term Memory (Ciaburro and Venkateswaran, 2017). The design of Long Short-Term Memory can overcome the problem of vanilla RNN, namely vanishing and exploding gradients (Brownlee, 2017). The ANN method as described has advantages over other analytical methods. Therefore, in this study, crude palm oil price predictions were made using ANN. The ANN method applied is Multilayer Perceptron and Long Short-Term Memory.

2. THEORETICAL FRAMEWORK

Forecasting is the art and science of predicting future events. Past data is used for forecasting and projecting into the future with mathematical models (Heizer and Render, 2011). Forecasting is an objective calculation using past data to determine something in the future (Sumayang, 2003).

Forecasting is classified into two categories, quantitative and qualitative methods. Quantitative methods can be applied if there is information from the past, it can be quantified in the form of numerical data and it can be assumed that some aspects of past patterns will continue in the future. Qualitative methods do not require data with the same characteristics as quantitative methods. The input required depends on the specific methods, primarily the result of assessment and accumulated knowledge (Makridakis et al., 1997).

Preprocessing Data

Preprocessing data is required by machine learning algorithms to run properly. The machine learning process doesn't work properly when the input numeric attributes are scaled differently. There are two common ways for attributes to be the same scale, using min-max scaling/normalization and standardization. Min-max scaling runs by shifting the value and re-scaling in the range 0 to 1. Min-max scaling is done by subtracting the minimum value and dividing by maximum minus minimum as an equation 1 (Géron, 2017).

$$X_{scaled} = \frac{X - min(X)}{max(X) - min(X)} \tag{1}$$

Multilayer Perceptron

ANN is a collection of algorithms inspired by the biological brain. ANN consists of simple connected units called neurons and functions to receive, process, and transmit signals to other neurons. The elements in ANN are quite simple. The complexity and strength of this system comes from the interaction between the elements (De Marchi and Mitchell, 2019).

Multilayer Perceptron consists of one input layer, one or more hidden layers, and one output layer. Every layer except the output layer includes biased neurons and is fully connected to the next layer. The training algorithm on the Multilayer Perceptron uses backpropagation (Géron, 2017). The backpropagation training algorithm consists of feedforward to get prediction results and backpropagation of error to get corrected weight and bias values and used to update weights and biases (Fausett, 1994). Multilayer perceptron architecture which consists of one input layer with q neurons, one hidden layer with p neurons and one output layer with one neuron can be written as equation 2.

$$y_o = \psi_o \left\{ w_{co} + \sum_{n=1}^{p} w_{no} \varphi_n \left(w_{cn} + \sum_{i=1}^{q} w_{in} x_i \right) \right\} \tag{2}$$

where w_{in}, w_{no} are weights for the unit relationship in the input layer-hidden layer and weights for the unit relationships in the hidden layer-output layer, then w_{cn}, w_{co} are weights for the constant unit relationships in the input layer-hidden layer and weights for constant unit relationships in the hidden layer-output layer. φ_n, ψ_o are activation functions in the hidden layer and output layer.

Long Short-Term Memory

ANN has several variants based on the number of hidden layers and data flows. One of them is RNN, where the connections between neurons form a cycle. RNN uses internal memory for processing and displays connections between hidden layers that are propagated over time to learn sequences (Ciaburro and Venkateswaran, 2017). RNN is good to use when it is important to know the sequence. The advantage of RNN is that there is no limitation on the input length and without affecting the size of the model and taking into account the sequential history of the network (De Marchi and Mitchell, 2019). RNN that has too many layers can cause it to be untrainable, known as vanishing gradient. Networks that have vanishing gradient cannot converge to good results (Rungta, 2018).

Hochreiter and Schmidhuber (1997) introduced Long Short-Term Memory (LSTM) as a solution to the vanishing and exploding gradients contained in the RNN. Gers et al. (2000) proposed the addition of a forget gate to determine the information to be removed from the cell state. The main idea of LSTM is cell state, information can only be written or deleted explicitly. Cell state can only be changed by certain gates. These gates consist of logistic sigmoid functions and element-wise multiplication. LSTM consists of three gates, namely forget gate, input gate, and output gate (Vasilev et al., 2019). The equation in the LSTM can be written as follows:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{3}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{4}$$

$$c_t' = tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{5}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot c_t' \tag{6}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{7}$$

$$h_t = o_t \odot tanh(c_t) \tag{8}$$

where f_t, i_t, o_t are forget gate, input gate, and output gate then c_t', c_t, h_t are candidate cell state, cell memory state, and output. Cells in the LSTM can be shown in Fig. 1.

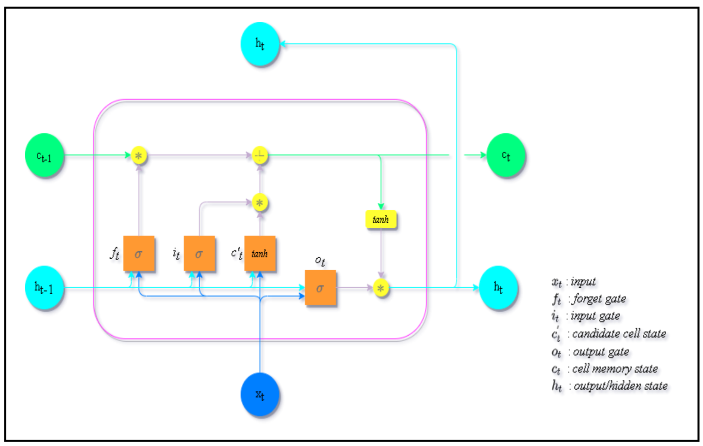

Fig. 1. LSTM Cell

Optimization Methods

Loss function is calculated after making predictions and to improve accuracy an optimizer is used. The optimizer is used to adjust the weight of the model during the training. The optimizer forms the model into the most accurate by adjusting the weights based on the loss function value. Gradient descent is one of the popular optimization algorithms. There are three variants of gradient descent. Batch gradient descent calculates the gradient for the entire data set to perform a single update. In contrast, stochastic gradient descent (SGD) performs parameter updates for each training instance x_i and label y_i (Ruder, 2016). Mini-batch gradient descent takes the idea of the two previous variants by accumulating errors for every n samples and performing one update (Vasilev et al., 2019).

Another optimization algorithm is Root Mean Square Propagation (RMSProp). RMSProp works by maintaining the moving average of the squared gradient for each weight. The recommended γ value is 0.9 and learning rate (η) is 0.001 (Tieleman and Hinton, 2014). The next optimization method is Adaptive Moment Estimation (Adam). This method calculates the adaptive learning speed for each parameter (Ruder, 2016). Adam combines ideas from Momentum optimization and RMSProp (Géron, 2017).

Model Performance Measures

One of the important parts of ANN is evaluating its performance. The metric that is often used in regression problems is Mean Square Error (MSE). MSE is defined as the difference of the mean square of the expected value and the predicted result. MSE equation can be written as follows:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{n} \tag{9}$$

where y_i is the actual value and $\widehat{y_i}$ is the predicted value.

The next metric is Mean Absolute Percentage Error (MAPE). MAPE calculates the mean of the absolute percentage error of predictions. The prediction results are better if the MAPE value is getting smaller (Swamidass, 2000).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right| \times 100 \tag{10}$$

Model with MAPE value of 10% or less is model with high forecasting accuracy, then model with MAPE value of more than 10% and less than or equal to 20% has a good forecasting rate. MAPE value in the interval of more than 20% and less than or equal to 50% then has a reasonable forecasting rate, and if MAPE is more than 50% then it has an inaccurate forecasting rate (Lewis, 1982).

3. RESEARCH METHOD

Data in this study is daily price of crude palm oil in Rupiah per kilogram from January 4, 2010 to December 29, 2020 obtained from the Ministry of Trade of the Republic of Indonesia, then converted into supervised and splitted into training data and testing data. There are three ways of splitting the data. The first is 70% training-30% testing, then 75% training-25% testing, and 80%

training-20% testing. The data is processed using Google Collaboratory with the methods are Multilayer Perceptron and Long Short-Term Memory. The steps of data analysis are as follows:

1. Load research data.
2. Conduct exploratory data analysis (EDA).
3. Change the form of the data into a supervised.
4. Split the data into training data and testing data.
5. Normalize the data into a range of 0 to 1.
6. Make model architecture for Multilayer Perceptron and Long Short-Term Memory.
7. Search for the best hyperparameters for each model.
8. Conduct training on each model that has been made using the obtained hyperparameters.
9. Select the best model based on the MSE value in the training data.
10. Make predictions on the testing data using the best model.
11. De-normalize the predicted data.
12. Measurement of model performance on data testing using MAPE

4. MAIN RESULTS

Crude palm oil price data is converted into supervised and splitted into training data and testing data. The proportion of training and testing is 70%-30%, 75%-25%, and 80%-20%. Then normalize using min-max scaling, with the scaler derived from the training data. The minimum value is 5764 and the maximum value is 10981. Normalization calculation for the first three data:

$$X_1' = \frac{X_1 - min_x}{max_x - min_x} = \frac{7537 - 5764}{10981 - 5764} = 0.33985049$$

$$X_2' = \frac{X_2 - min_x}{max_x - min_x} = \frac{7725 - 5764}{10981 - 5764} = 0.37588652$$

$$X_3' = \frac{X_3 - min_x}{max_x - min_x} = \frac{7724 - 5764}{10981 - 5764} = 0.37569484$$

Multilayer Perceptron

The first method to predict crude palm oil prices is Multilayer Perceptron (MLP). The input for MLP uses significant lag from the Partial Autocorrelation Function (PACF). PACF of crude palm oil price data can be shown in Fig. 2.

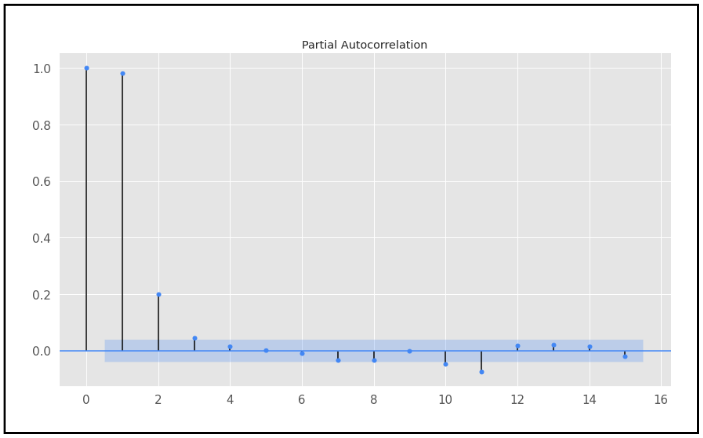

Fig. 2. PACF Plot of Crude Palm Oil Price

Fig. 2 shows the PACF of crude palm oil price data. Based on the PACF plot, the significant lags are lag 1, lag 2, lag 3, lag 10, and lag 11. In order to avoid many lag combinations, the lag used as input is a combination of lag 1, lag 2, lag 3, and lag 11. Using one hidden layer, with the choice of neurons in the hidden layer is 3, 4, 5, 6, 7, and 8. The activation function in the hidden layer is ReLU. The batch size is 64, with 200 epochs. The optimization methods are SGD, RMSProp, and Adam. Computed with MLP method on all combinations of hyperparameters. Then the best hyperparameters are selected from each optimization method. The best results from each optimization are shown in Table 1.

Table 1. MLP Results

Input Lag	Hidden Neurons	Data Split	Optimization	MSE
1 and 3	5	75%-25%	SGD	33391.34
1 and 2	8	70%-30%	RMSProp	33195.77
1 and 2	6	75%-25%	Adam	33222.80

Combination of input lag, hidden neurons, data splitting, and optimization method that have small MSE is lag 1 and 2, hidden neurons 8, the proportion of training and testing data is 70%-30%, and the optimization method is RMSProp has MSE value of 33195.77, then lag 1 and 2, hidden neurons 6, proportion of training and testing data is 75%-25%, and the optimization method is Adam has

MSE value of 33222.80.

Both have MSE values that are not much different. Besides based on the MSE value, it also considers the number of parameters. Lag 1 and 2, hidden neurons 6, the proportion of training and testing data is 75%-25%, and the optimization method is Adam has weight and bias parameters of 2×6+6×1+6+1=25, less than lag 1 and 2, hidden neurons 8, the proportion of training and testing data is 70%-30%, and the optimization method is RMSProp which has weight and bias parameters of 2×8+8×1+8+1=33, thus lag 1 and 2, hidden neurons 6, the proportion of training and testing data is 75%-25%, and the optimization method is Adam simpler than lag 1 and 2, hidden neurons 8, the proportion of training and testing data is 70%-30%, and the optimization method is RMSProp. Finally the best MLP is lag 1 and 2, hidden neurons 6, the proportion of training and testing data is 75%-25%, and the optimization method is Adam.

Long Short-Term Memory

The second method is LSTM. The time steps is from 1 to 10. Using one hidden layer, with the choice of neurons in the hidden layer is 3, 4, 5, 6, 7, and 8. The activation function in the hidden layer is ReLU. The batch size is 32, with 200 epochs. The optimization methods are SGD, RMSProp, and Adam. Computed with LSTM method on all combinations of hyperparameters. Then the best hyperparameters are selected from each optimization method. The best results from each optimization are shown in Table 2.

Table 2. LSTM Results

Time steps	Hidden Neurons	Data Split	Optimization	MSE
3	5	70%-30%	SGD	33539.59
2	4	70%-30%	RMSProp	33019.59
2	6	70%-30%	Adam	32224.17

Table 2 shows the best LSTM results from each optimization method. Combination of input lag, hidden neurons, data splitting, and optimization method that have small MSE is time steps 2, hidden neurons 6, the proportion of training and testing data is 70%-30%, and the optimization method is Adam has MSE value of 32224.17. Finally, the best LSTM model is time steps 2, hidden neurons 6, the proportion of training and testing data is 70%-30%, and the optimization method is Adam.

Model Selection

The best MLP and LSTM results from Table 1 and Table 2 are then shown in Table 3. The smallest MSE value is 32224.17 which is obtained when using LSTM, with time steps 2, hidden neurons 6, the proportion of training and testing data is 70%-30%, and the optimization method is Adam.

Table 3. Best Results

Method	Input Lag	Time steps	Optimization	Hidden Neurons	Data Split	MSE
MLP	1 dan 2	-	Adam	6	75%-25%	33222.80
LSTM	-	2	Adam	6	70%-30%	32224.17

The best model is model using the LSTM, with time steps is 2, hidden neurons 6, the proportion of training and testing data is 70%-30%, and the optimization method is Adam. The model is used to predict crude palm oil prices. Predictions were made on testing data, from March 15, 2017 to December 29, 2020. The actual data and the predicted data are shown in Fig. 3.

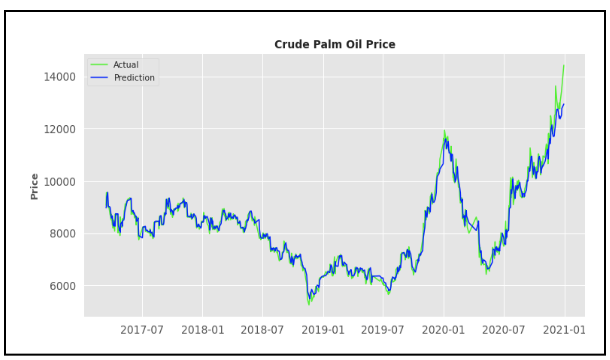

Fig. 3. Crude Palm Oil Price Prediction

Fig. 3 shows plot of the actual data and the predicted data of crude palm oil prices. The green plots are actual data, while the blue plots are predictive data. Based on Fig. 3, it can be seen that the model can produce the appropriate output, because the prediction results are similar to the actual data. Model performance is measured using MAPE. Based on the prediction results, MAPE value is 2.11%. Model has high forecasting accuracy because MAPE is less than 10%.

5. CONCLUSION

According to the results and discussion, it can be concluded that the best time series model for predicting AALI returns for the period of 3 October 2012 to 1 October 2019 is ARIMA(0,0,1)-GARCH (Based on the results, it can be concluded that the best model is using the LSTM method, time steps 2, hidden neurons 6, and the optimization method is Adam. The prediction results are not different from the actual value, with MAPE value is less than 10% so that it has high forecasting accuracy.

ACKNOWLEDGEMENT

This research supported by Laboratory of Statistics Department, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] J. Brownlee. Long Short-term Memory Networks with Python: Develop Sequence Prediction Models with Deep Learning. Jason Brownlee, (2017).

[2] G. Ciaburro, B. Venkateswaran. Neural Networks with R: Smart Models using CNN, RNN, Deep Learning, and Artificial Intelligence Principles. Birmingham: Packt Publishing, (2017).

[3] L. De Marchi, L. Mitchell. Hands-On Neural Networks: Learn How to Build and Train Your First Neural Network Model using Python. Birmingham: Packt Publishing Ltd., (2019).

[4] L. Fausett. Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. Englewood Cliffs, New Jersey: Prentice-Hall Inc., (1994).

[5] A. Géron. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc., (2017).

[6] F.A. Gers, J. Schmidhuber, F. Cummins. Learning to Forget: Continual Prediction with LSTM. Neural Comput. 12 (2000), 2451–2471.

[7] P. Hariyadi. Mengenal Minyak Sawit dengan Beberapa Karakter Unggulnya. Jakarta: GAPKI, (2014).

[8] J. Heizer, B. Render. Operations Management. Pearson Higher Ed., (2011).

[9] A.A. Karia, I. Bujang & I. Ahmad. Forecasting on Crude Palm Oil Prices Using Artificial Intelligence Approaches. Amer. J. Oper. Res. 3 (2013), 259-267.

[10] S. Hochreiter, J. Schmidhuber. Long Short-Term Memory. Neural Comput. 9 (1997), 1735–1780.

[11] C.D. Lewis. Industrial and Business Forecasting Methods A Practical Guide to Exponential Smoothing and Curve Fitting. London: Butterworth-Heinemann, (1982).

[12] S. Makridakis, S.C. Wheelwright, R.J. Hyndman. Forecasting: Methods and Applications. John Wiley & Sons, (1997).

[13] S. Ruder. An Overview of Gradient Descent Optimization Algorithms. arXiv preprint arXiv:1609.04747, (2016).

[14] K. Rungta. TensorFlow in 1 Day: Make Your Own Neural Network. Amazon Digital Services LLC - Kdp Print Us, (2018).

[15] L. Sumayang. Dasar-Dasar Manajemen Produksi dan Operasi. Jakarta: Salemba Empat, (2003).

[16] P.M. Swamidass. Encyclopedia of Production and Manufacturing Management. Springer US, (2000).

[17] T. Tieleman, & G. Hinton. RMSprop Gradient Optimization, (2014). https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf. Accessed on December 4, 2020.

[18] I. Vasilev, D. Slater, G. Spacagna, P. Roelants, & V. Zocca. Python Deep Learning: Exploring Deep Learning Techniques and Neural Network Architectures with PyTorch, Keras, and Tensor Flow. Birmingham: Packt Publishing Ltd, (2019).