ISSN 2278-3091
Volume 9, No.4, July – August 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse238942020.pdf
https://doi.org/10.30534/ijatcse/2020/238942020

A Comparative Study on Univariate Time Series based Crude Palm Oil Price Prediction Model using Machine Learning Algorithms

Kanchymalay Kasturi[1], N.Salim[2], Ramesh Krishnan[3], U.R.Hashim[4], M.B. MasAina[5], Indradevi S[6], Mutasem Jarrah[7]

[1]Centre for Advanced Computing Technology (C-ACT), Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia, kasturi@utem.edu.my

[2]School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia naomie@utm.my

[3]Faculty of Business & Management, Universiti Teknologi MARA, KampusBandaraya Melaka,Malaysia, rameshkris@uitm.edu.my

[4]Centre for Advanced Computing Technology (C-ACT),Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia,ummi@utem.edu.my.

[5]Centre for Advanced Computing Technology (C-ACT),Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia,aina@utem.edu.my

[6]Centre for Advanced Computing Technology (C-ACT), Pusat Bahasa dan Pembangunan Insan, Universiti Teknikal Malaysia Melaka (UTeM), indradevi@utem.edu.my

[7]Information Technology Department,Faculty of Computing &Information Technology, King AbdulazizUniversity,Jeddah, Saudi Arabia,mmjarrah@kau.edu.sa

ABSTRACT

Crude palm oil (CPO) price prediction plays an important role in the agricultural economic development. It requires an in-depth knowledge in both economics and agricultural domain. The aim of this paper is to propose a CPO price prediction model to help the plantation organizations in the palm oil sector to effectively anticipate CPO price fluctuations and managing the resources more effectively. The CPO price behavior are non-linear in nature, thus prediction is very difficult. In this paper, a recurrent network, Long Short Term Memory (LSTM) based CPO price prediction system is compared with artificial neural network (ANN) and Holt-Winter method. The findings of this study shows that the LSTM based forecasting model outperformed other models in forecasting the CPO price movement. This study recommends that a LSTM based forecasting could better help the farmer and planters in the agriculture sector in managing the demand of CPO and the operation processes for a better return on investment.

Key words: Crude Palm Oil Price, Long Short Term Memory (LSTM), Machine learning, Time series.

1. INTRODUCTION

Forecasting accuracy in the prediction of commodity prices such as crude palm oil (CPO) price is vital to achieve a positive impact on economic growth of Malaysia. The CPO price fluctuations imposes a tremendous risk to key players in the palm oil industry such as farmers, consumers, traders and producers. The world consumption for CPO is showing an increasing trend, therefore an accurate CPO price prediction model is critical to help in decision-making in situations of uncertainty. As seen in Figure 1, CPO price is found to be highly fluctuated and showing a nonlinear trend.

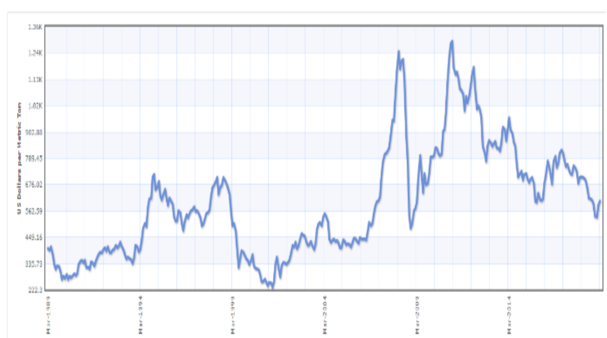

Figure 1: Palm oil monthly price in USD
Source: Oil World (2019) ISTA Mielke GmbH

Most of the time series forecasting for commodity price forecasting models uses traditional statistical methods such as Autoregressive Integrated Moving Average (ARIMA). In this paper, Holt-Winter Exponential Smoothing model, artificial neural networks (ANN) and Long Short Term Memory (LSTM) were utilized for univariate time series based CPO price forecasting for performance comparison.

The study tested historical data of CPO prices in time series as predictors. Historical data suggests that complex volatility characteristics of CPO prices are similar to crude oil, such as nonlinearity, uncertainty and dynamics that making the CPO price prediction challenging and the prediction results bear high uncertainties, which subsequently may cause significant reservations for the profits of related investors and social economy steady development. In addition, most of these studies have used statistical methods applying the assumption that time series data are linear and stationery, do not adequately capture the real-time series data which is dynamic, non-linear, and non-stationary for better forecasting accuracy [1],[2]. Traditionally, the time series data that used for statistical methods study have assumptions that they are generated from a linear process [3]. Based on literature, Box-Jenkins model was found to be producing rational results for linear time series data. This model isa widely used technique in the CPO price forecasting [4]. Box-Jenkin's Autoregressive Moving Average (ARMA) model is relatively better in producing short term prediction but incapable of portraying long-run time series data precisely [4]. The Machine learning attempts is to predict the future value automatically by learning the past experience from historical data [5].In a real environment, time series are effectively non-linear, nonparametric, highly noisy, dynamic,complex and chaotic in nature [6]. Traditional shallow neural network methods that consist only a few number of non-linear operations cannot accurately model such complex data [7]. Conventional ANN architectures and algorithms such as Feed Forward Neural Network (FFNN) is not suits the data trends that vary over time[8]. Therefore a temporal prediction has become more challenging. Due to this reason this study proposed LSTM based Deep learning prediction model for CPO price prediction**.**

2. METHODS AND MATERIALS

In this section data and techniques used in this study is discussed.

2.1. Data Collection and Analysis

In this phase, understanding and analysis of the data pattern of the CPO price were performed. Firstly, data collection was carried out. After various phone calls and e-mail exchanges with Malaysian Palm Oil Board (MPOB) officials, data were collected from the MPOB portal in which daily and monthly CPO prices were retrieved and analyzed for the periods between 1984 and 2018.Descriptive statistics analysis was performed on the CPO price data, of which a summary is presented in Table 1. The data distribution, as shown in Table 1, indicates a skewness value 0.999 of the CPO price data, which explains that the data is skewed to the right and presenting positive skewness.

The kurtosis value of the data is 0.3, which indicates that most of the values pertaining to the data were dispersed from the mean value of 511.16. This statistic confirmed that the data were not normally distributed, as shown in Figure 1. This implies that the data did not meet the requirement to perform a linear regression analysis. Therefore, LSTM, a techniques that can handle non-linear data and capable to learn data representation with long term data dependencies is proposed in this study.

Table 1: Summary of Descriptive Statistics for CPO price

Palm oil Price (US Dollars per Metric Ton)	
Mean	511.1611944
Standard Error	12.53984576
Median	447.495
Mode	333.03
Standard Deviation	237.9268446
Sample Variance	56609.18339
Kurtosis	0.352179882
Skewness	0.999986411
Range	1063.48
Minimum	185.07
Maximum	1248.55
Sum	184018.03
Count	360

2.2 Experiment

i. Artificial Neural Network (ANN)

ANN with Feed Forward Neural Network (FFNN) was the most efficient approach for modeling and forecasting the non-linear system[10] and used to solve various prediction challenges in classification[20] and regression areas [19]. McCulloch and Pitts used Mathematical function derived from simulation of the biological neurons to explain the concept of the artificial neuron[12]. ANN includes of a few layers, and each layers consist of a number of neurons. A standard ANN topology consist of an input layer, an output layer and one or more hidden layers in between. As shown in Figure 2, the number of hidden layers and number of neuron in each layer is dependent on the complexity of the problem. From the interaction with the external environment, input layers receive the input features as a vector. The desired output is reflected in output neuron of the model. The development process of ANN model is consist of the following steps. The first step is to find a suitable input feature set. The second step is to establish the number of hidden layers and neurons. The last step is to train and test the network to evaluate the performance. The output of the network can be expressed mathematically as below.

$$Y_t = f_2[\sum_{j=1}^{J} w_j f_1(\sum_{i=1}^{I} w_i x_i)]$$

From the above equation Y_t is denoted as the output of the model. X_i is denoted as input variables whereas w_i and w_j is referring to the weights between neurons of the input and hidden layer and between hidden layer and output respectively. The activation functions for the hidden layer is represented by f_1 and the activation functions output layer is f_2.

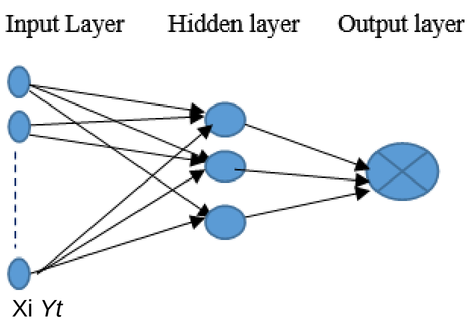

Figure 2: Neural Network [11]

The network we built had combination of one to four inputs and one output. In this architecture, one hidden layer and two hidden neurons was chosen. The architecture is tested with sigmoid in both hidden layer activation function and output layer activation function respectively [13].

There were various methods in splitting the data for training and testing process such as cross-validation, bootstrap and holdout [12]. Data was divided according to holdout method. As for training and test sets, the first 70% of data observations were used as training set and the remaining 30% of data observations were used as the test set.

For the assessment of the model performance, Root Mean Squared Error (RMSE) has been used for model evolution. RMSE measures the square root of average of squares of the prediction errors as shown in equation 1 with Xi is predicted value and X is actual value (Eq.1).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i - X)^2}{n}} \qquad (1)$$

ii. Long Short Term Memory (LSTM)

Long Short Term Memory networks (LSTM) are a special kind of RNN, capable of learning long-term dependencies. They were introduced by [14] Hochreiter & Schmidhuber in 1997, Only a few published papers apply LSTMs to time series forecasting tasks, all of which, to our knowledge, are outside of the CPO price context[18] especially for reading cursive writing[15] and showed best performance using RNN– LSTM.

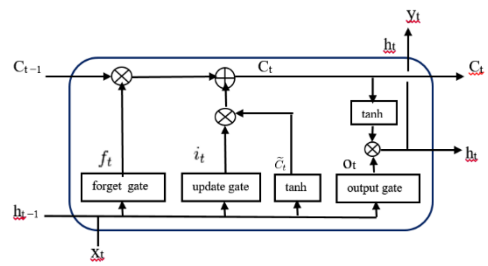

Figure 3: Long Short Term Memory (LSTM) [13]

Most recently, [16] in 2018 uses RNN with LSTM which captures the spatio-temporal dependencies in local rainfall to assess the hydrological impacts of global climate change on regional scale. LSTM model also applied to predict the air pollutants concentration in Indian cities[9].While we could not locate any published papers using LSTMs for multivariate time series price forecasting for palm oil domain, several papers use statistical method and feed forward neural networks for similar study[4][17]. Vanishing or exploding gradients problem in recurrent network is resolved using LSTM for long term learning process. Standard neural network layers replaced by LSTM cell block in a recurrent networks for LSTM networks. These cells have various components called the input gate, the forget gate and the output gate. The graphical representation of the LSTM cell unrolled in time is shown in Figure 3. In Figure 3, x(t) is the cell's input at time t. And h(t) is the cell's output at time t. C(t −1) and C(t) are the LSTM Cell States at time (t−1) and t respectively, which are the key reason for LSTM to learn long term dependency without gradient vanishing risk. Specifically designed structure called Gates control the cell states. There are three gates:

i) Forget gate f(t), conditionally decides what information to throw away from the block. A forget gate is responsible for removing information from the cell state. Using the multiplication of a filter, less important information which are no longer required for LSTM is removed. This is important for performance optimization of the LSTM network.

Forget gate takes in two inputs; ht-1 and xt. Hidden state from the previous cell or the output of the previous cell is represented as ht-1.The input at that particular time step is denoted as xt. Multiplication performed on input features together with the weight matrices and added with a bias. Subsequently, the sigmoid function is applied to above value. The sigmoid function output vector with values ranging 0 to 1 is decide to keep or discard the values. If output for a particular value in the cell state is '0', it means that the forget gate wants the cell state to forget that piece of information completely. Else '1' value make the forget gate to keep and remember that whole piece of information.

Keras library used together with phyton and pandas for building the LSTM model for CPO forecasting. The experiment steps are illustrated in Figure 4.

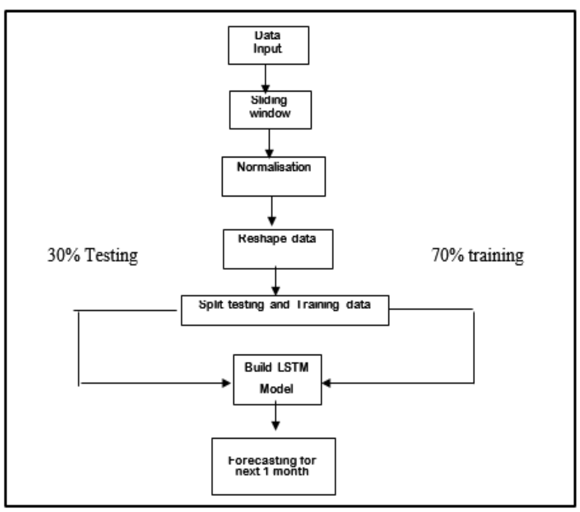

Figure 4: Illustration of data pre-processing before feeding into LSTM Network

3. RESULTS AND DISCUSSION

The result of the CPO price prediction based on univariate time series data of CPO prices is shown in Table 2. Experiments carried out using widely used Holt Winter Exponential Smoothing, a statistical method, ANN and LSTM for machine learning methods. The findings of the experiments is as shown in Table 2.

For CPO price forecasting, LSTM model seems to be the best model to forecast the Malaysian CPO price. The finding of the study shows the LSTM based forecasting model has produced the least error with lowest RMSE and outperformed other methods significantly. Figure 5 shows the predicted and actual CPO price with LSTM. The model that has the lowest error value is said to be the best model for forecasting. Thus, based on the result in Table 2, LSTM model shows a smallest error value as compared to ANN and Holt-Winter models. Table 2 and Figure 5 given below display the outcomes of the experiment. Here, the outcomes of the LSTM exhibited a considerable improvement in precision compared to the other models

Table 2: Result comparison between ANN ,Holt-Winter Method and LSTM

Technique	RMSE
ANN	191.56
Holt-Winter	164.57
LSTM	116.12

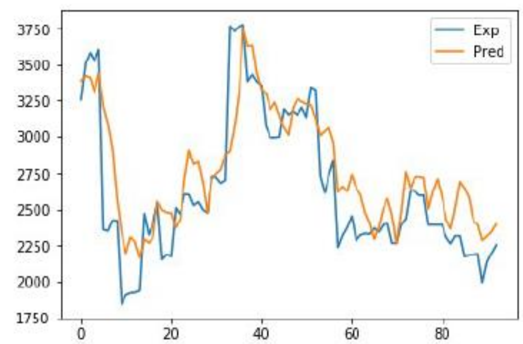

Figure 5: Predicted and the actual CPO price with LSTM

4. CONCLUSION

This study attempted to predict CPO prices by considering historical data of CPO prices as predictors. The result obtained showing the LSTM model demonstrates significant improvement with the least RMSE. LSTM performed better in determining the movement or behavior of crude palm oil price.

5. FUTURE SCOPE

Future experiments should be carried out with more predictors as input features such as other commodities and climate factors. Other machine learning techniques is also proposed for future experiment.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to the Centre for Advanced Computing Technologies (CACT), Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka (UTeM) for supporting this research.

REFERENCES

1. A.A. Karia and Bujang, I. **Progress accuracy of CPO price prediction: Evidence from ARMA family and Artificial Neural Network approach,** *International Research Journal of Finance and Economics.* ISSN 1450-2887, 2011.
2. Cheng, C., Sa-Ngasoongsong, A., Beyca, O., Le, T., Yang, H., Kong, Z. and Bukkapatnam, S. T**ime series forecasting for nonlinear and non-stationary processes: A review and comparative study**, *IIE Transactions*. 47(10), 1053-1071,2015.https://doi.org/10.1109/ICPRIM E.2013.6496450
3. Kumar, D. A. and Murugan, S.**Performance analysis of Indian stock market index using neural network time series model,** *IEEE 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*. 72-78, 2013.
4. A. A. Karia .**Forecasting on crude palm oil prices using Artificial Intelligence approaches,** *Am. J. Oper. Res.*, vol. 03, no. 02, pp. 259–267, 2013.
5. Kavitha ,Charan and Jannu. **Supervised machine learning approaches,** *A Computer Science Engineering,* Vol 5, Is 2 E-ISSN: 2348-2273, P-ISSN : 2454-1222, 2018.
6. Si, Y.-W. and Yin, J. **OBST-based segmentation approach to financial time series**,*Engineering Applications of Artificial Intelligence*. 26 (10), 2581–2596, 2013.
7. Längkvist, M., Karlsson, L. and Loutfi, A. **A review of unsupervised feature learning Survey,** *International Journal of Electrical Electronics & and deep learning for time-series modeling. Pattern Recognition Letters*. 42(1), 11-24. ISSN0167-8655,http://dx.doi.org/10.1016/j.patrec. 2014.01.008, 2014.
8. Weigend, A. S. **Time series prediction: forecasting the future and understanding the past,** Routledge. 2018.
9. Keerthana R, and Varghese S. Fo**recasting of the air pollution based on meteorological data and air pollutants using Deep Learning: A novel review,** *International Journal of Advanced Trends in Computer Science and Engineering*, 9(1), January – February, 2020.
10. Al-Janabi A.A, Alhendawi K.A, and Alsewari A.A. **An intelligent neural model for assessing Web Systems performance,***International Journal of Advanced Trends in Computer Science and Engin*eering, 9(2), March-April, 2020.
11. Negnevitsky Artificial **intelligence: A guide to Intelligent Systems (3rd Edition)**, Pearson Education. 2005.
12. I. H. Witten and E. Frank.**Data mining: Practical machine learning tools and techniques**: Morgan Kaufmann, 2005.
13. ChristopherOlah, http://colah.github.io /posts /2015-08-Understanding-LSTMs, 2015
14. S. Hochreiter and J. Urgen Schmidhuber. **Long Short Term Memory**, *Neural Computing*, vol.9 vol.9, no. 8, pp. 1735–1780, 1997.
15. A. Graves and J. Schmidhuber. **Offline handwriting recognition with multi-dimensional recurrent neural networks**, *NIPS 2008*, Vancouver, Canada, pp. 545-552, 2008.
16. Misra, S., Sarkar, S. and Mitra,P. **Statistical downscaling of precipitation using long Theorem.** ApplClimatoly 134, 1179–1196. https://doi.org/10.1007/s00704-017-2307-2.,2018.
17. G. P. Yean, Gan, Peck Yean and Li, Zhi Dong. **Econometric study on Malaysia's palm oil position in the world market to 2035**, *Renewable and Sustainable Energy Reviews*, Elsevier, vol.39, pages 740-747. 2014.
18. Kanchymalay, K., Salim, N., and Krishnan, R. **Time series Based Crude palm oil price forecasting with weather elements using LSTM Network**, *International journal of Engineering and Advanced Technology* , Vol 9-1,Oct 2019.
19. R.Suganyaand Sasi Kumar A. **Stock price prediction using tech news based soft computing approach,** *International Journal of Advanced Trends in Computer Science and Engineering*, 9(2), March - April 2020.
20. P. Santhi, N. Deeban, N. Jeya punitha, B. Muthu kumaran, and R. Ravi kumar. **Prediction of Diabetes using Neural Networks,** *International Journal of Advanced Trends in Computer Science and Engineering*, 9(2), March-April 2020.