

PAPER • OPEN ACCESS

# Multivariate Time Series Forecasting of Crude Palm Oil Price Using Machine Learning Techniques

To cite this article: Kasturi Kanchymalay *et al* 2017 *IOP Conf. Ser.: Mater. Sci. Eng.* **226** 012117

View the [article online](#) for updates and enhancements.

## Related content

- [Bioactive compounds from palm fatty acid distillate and crude palm oil](#)  
T Estiasih and K Ahmadi
- [The Effect Of Additional Detergent In Crude Palm Oil In The Process Of Separation Stearin](#)  
Vina Rezekyah Hasibuan, Nur aini, Febriyanti et al.
- [Kinetics of Anaerobic Digestion of Palm Oil Mill Effluent \(POME\) in Double-Stage Batch Bioreactor with Recirculation and Fluidization of Microbial Immobilization Media](#)  
L I Ramadhani, S I Damayanti, H Sudibyo et al.

# Multivariate Time Series Forecasting of Crude Palm Oil Price Using Machine Learning Techniques

**Kasturi Kanchymalay<sup>1,2</sup>, N. Salim<sup>2</sup>, Anupong Sukprasert<sup>3</sup>, Ramesh Krishnan<sup>4</sup>, Umami Raba'ah Hashim<sup>1</sup>**

<sup>1</sup>Faculty of Information Technology and Communication, Universiti Teknikal Malaysia Melaka, Malaysia.

<sup>2</sup>Faculty of Computing, Universiti Teknologi Malaysia, Malaysia.

<sup>3</sup>Maharakham Business School, Maharakham University, Thailand.

<sup>4</sup>Faculty of Business & Management, Universiti Teknologi MARA, Malaysia.

Corresponding author: [kasturi@utem.edu.my](mailto:kasturi@utem.edu.my), [umami@utem.edu.my](mailto:umami@utem.edu.my),  
[naomie@utm.my](mailto:naomie@utm.my), [anupong.s@acc.msu.ac.th](mailto:anupong.s@acc.msu.ac.th), [rameshkris@melaka.uitm.edu.my](mailto:rameshkris@melaka.uitm.edu.my)

**Abstract.** The aim of this paper was to study the correlation between crude palm oil (CPO) price, selected vegetable oil prices (such as soybean oil, coconut oil, and olive oil, rapeseed oil and sunflower oil), crude oil and the monthly exchange rate. Comparative analysis was then performed on CPO price forecasting results using the machine learning techniques. Monthly CPO prices, selected vegetable oil prices, crude oil prices and monthly exchange rate data from January 1987 to February 2017 were utilized. Preliminary analysis showed a positive and high correlation between the CPO price and soy bean oil price and also between CPO price and crude oil price. Experiments were conducted using multi-layer perception, support vector regression and Holt Winter exponential smoothing techniques. The results were assessed by using criteria of root mean square error (RMSE), means absolute error (MAE), means absolute percentage error (MAPE) and Direction of accuracy (DA). Among these three techniques, support vector regression (SVR) with Sequential minimal optimization (SMO) algorithm showed relatively better results compared to multi-layer perceptron and Holt Winters exponential smoothing method.

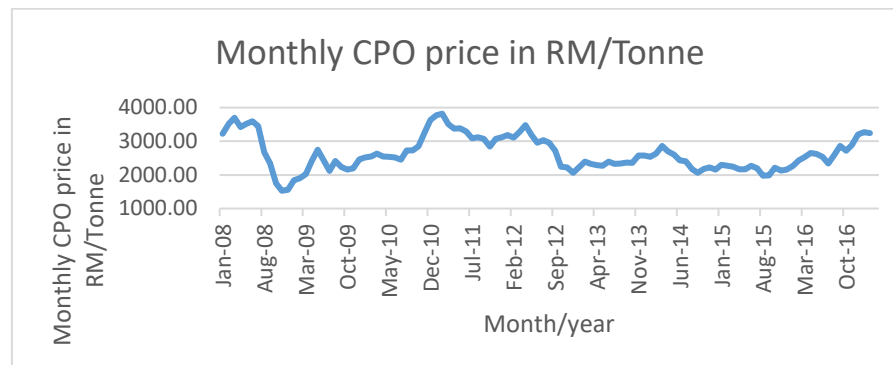
## 1 Introduction

The palm oil industry has grown locally and globally due to Malaysia's position as one of the world's leading palm oil producing country. As one of the top producers and exporters of crude palm oil (CPO) and palm oil products, Malaysia has an important role to sustainably fulfill the growing global need for edible oils and fats. According to Malaysian Palm Oil Board (MPOB), Malaysia is the world's second largest palm oil exporter after Indonesia and oil palm industry forms the economic backbone of Malaysia[1]. As the world's second biggest palm oil producer, Malaysia continues to face new challenges with volatile agricultural commodities prices especially CPO prices. The breadth of the palm oil industry and its significance cannot be understated as its impact ripples through the domestic as well as global economy.

The Malaysian palm oil industry has undoubtedly made significant contribution towards the domestic economy as well as to the development of the world palm oil market[2]. As a fast growing edible oil playing a significant role in Asean economy, it is important to monitor and accurately forecast the price of CPO for the benefit of the Malaysian palm oil industry. Palm oil like any other agricultural commodities is subject to significant price fluctuation as shown in figure 1. The CPO price fluctuation passes on a significant risk to farmers, producers, traders, consumers and others involved



in the production and marketing of CPO. An accurate CPO price forecasting technique is necessary to assist decision-making in risky and uncertain situations.



**Figure 1** Monthly Crude Palm Oil Price Movement.

source: <http://mpoc.org.my> [3]

Typically, CPO price forecasting is handled mostly by economists thus most of the CPO price forecasting research employed statistical methods[4][5][6]. Some of the recent study which employed artificial intelligence techniques in forecasting CPO price have utilized univariate time series analysis for forecasting [7][8] without considering impact of other commodity prices. Thus the main aim of this research is to study the correlation between CPO price and other vegetable oil prices (such as soybean oil, coconut oil, olive oil, rapeseed oil and sunflower oil), crude oil and exchange rate. This research also hopes to forecast the CPO price using other commodity prices employing the machine learning techniques.

### 1.1 Factors Influencing CPO Price

Past research shows that various factors contribute to the prediction of CPO prices. Some of the important factors that have significant relationship with CPO price movements are listed below:

#### i. Price of Soybean

A number of researchers[2],[9],[4],[9],[10][11] suggested that there is a positive relationship between movement of CPO price and soybean oil price. Soybean oil has been a long term competitor of CPO [10]. Past studies[8][11] suggested that there is a short and long run relationship between CPO price and soybean oil price. Our analysis shed some light on how the changes in soybean oil price movement affect the movement of CPO price.

#### ii. Price of Crude Oil

Researchers such as Abdullah in 2010[12], Kantaporn et al[13], Arshad et al.[10] confirmed the relationship between price of crude oil and CPO. Razak et al[14] employed the Engle-Granger co-integration test to demonstrate a significant long term correlation between CPO price and crude oil price which was consistent with the recent work of Appalanaidu [15].

#### iii. Exchange Rate

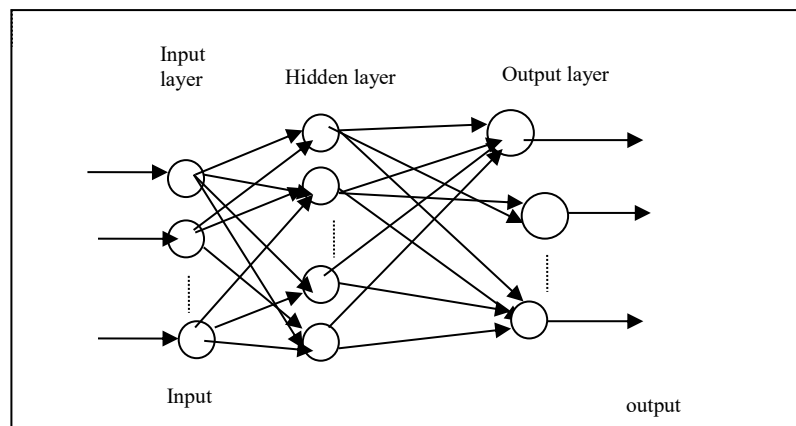
Studies conducted by Kantporn et al [13] and Khin [16] showed a significant relationship between exchange rate and CPO prices as demonstrated by [5] and [2].

Thus these three variables will be utilized as important variables in forecasting the CPO price in this study.

### 1.2 Techniques Used in CPO Price Forecasting

Literature search shows that two techniques that were widely used in time series forecasting are statistical methods and computational methods. Most of the price forecasting studies by economists were conducted using statistical methods. Time series forecasting techniques which are popular among researchers are exponential smoothing model, autoregressive models such as ARIMA and MARMA and ARCH /GARCH family models. In an autoregressive integrated moving average model (ARIMA) the future value of a variable is assumed to be a linear function of several past observations and random errors [17]. Arshad et al utilized Box Jenkins univariate ARIMA model to forecast the short-run monthly price of CPO [18]. Khin[15] utilized Vector Error Correction Method (VECM) to analyze the relationship between spot and futures prices in forecasting the CPO price. Even though statistical models were widely used for solving time series forecasting problems, recent trend showed machine learning techniques outperformed the classical statistical methods [8].

Time series forecasting is a difficult task as price movement behaves more like a random walk and varies with time. Recent forecasting trend shows machine learning techniques namely artificial neural network (ANN) and support vector machine (SVM) are being used in a range of different fields such as in business, medicine, energy and science. ANN has been successfully used in forecasting financial data series [19] as statistical methods were inappropriate for strong nonlinear time series such as crude oil price[20]. ANN models has significant advantage over other classes of nonlinear model as they can predict a large class of functions with a high degree of accuracy[[21]. Additionally, no prior assumption of the model form is required in the model building process as the network model is largely determined by the data characteristics [22]. ANN is a bio inspired model that utilize several single processing elements called neurons. A Multi-layer Perceptron (MLP) approach is commonly used in regression problems. It has 3 layers: input layer, output layer and hidden layer as displayed in figure 2. MLP uses back propagation algorithm that calculates the error and then back propagate the error to previous layer. The weight of the network is adjusted by minimizing the error between the target and computed outputs. The weights are continuously revised until minimum error is obtained[23].



**Figure 2.** Architecture of Multi-Layer Perceptron model.

SVM is also widely used in stock market price prediction [24][25]. Support vector for regression (SVR) is built upon statistical learning theory, as a method for solving nonlinear regression estimation problems. Vapnik [26] introduced SVM algorithm for regression and classification problems based on the theory of statistical learning. It is known as SVR to solve the regression prediction problems. SVR differs from ordinary regression methods since it uses structural risk minimization instead minimizing empirical risk used in other learning theory methods like neural networks. Therefore it is expected that this method outperforms other regression methods and is able to have better generalization[27]. According to Vapnik [30], the SVR model is expressed as:

$$f(x) = (a \cdot \phi(x)) + b, \quad (1)$$

where  $a$  is a weight vector,  $b$  is bias, and  $\phi(x)$  is a non-linear function to transform non-linear inputs to a linear mode in a high-dimensional feature space known as a kernel function. SVR with Sequential minimal optimization (SMO) proposed by Smola and Schölkopf [28] for using SVR regression. SMO has the good ability to model regression, prediction with non-linear data[28].

Exponential smoothing is another prediction method that is widely adopted in time series forecasting[29]. The exponential smoothing applied on time series containing noise, trend and seasonality. It was developed by Winters [31]. Three parameters alpha, beta and gamma which define the degree of smoothing are to be applied to noise, trend and seasonal components of time series. At first, a value of alpha is used to dictate the amount of smoothing to apply for noise. A second order of smoothing, so called “double exponential smoothing” using parameter beta is needed in a data set with trend. Finally, a third level of smoothing is introduced to make the process a triple exponential smoothing if a seasonal component is also present in the data set. This third level of smoothing parameter is gamma. Depending upon the nature of the time series one, two or all three of the parameters may be defined in the Holt Winters methodology[29]. A comprehensive performance evaluation of these three methods was examined in this study.

## 2 Experiments

In this section data collection processes and experiment setup is discussed in detail.

### 2.1 Data and Method

Data collection and pre-processing of data are important tasks in data analysis. In this study nine time series data were used. The monthly closing price of all variables such as CPO price, sunflower oil price, olive oil price, rapeseed oil price, coconut oil price, peanut oil price, soybean oil price and West Texas Intermediate (WTI) crude oil spot price were retrieved from <http://www.indexmundi.com>. Exchange rates of US dollar to Malaysian Ringgit was retrieved from <http://www.tradingeconomics.com>. The frequency of data used is the monthly price from January 1987 to February 2017.

First, pre-processing of data was carried out on the collected data for missing values. Data deletion was performed to handle the missing data. Statistical analysis was carried out to discover the underlying patterns and trends between the variables involved. A correlation analysis was then conducted to investigate the nature of the relationship between CPO price movement and other vegetable oil prices such as soybean oil, coconut oil, olive oil, rapeseed oil and sunflower oil. The highly correlated vegetable oils prices are chosen as the predictors to predict the CPO price. Next, the relationship between CPO price and crude oil price and the relationship between CPO price and exchange rate is also examined. The correlation between CPO price and other commodities prices was analysed using Pearson correlation test. The Pearson correlation test between CPO price and soybean oil price showed the highest  $r$  value (Pearson correlation coefficient) which is 0.941741044 followed by rapeseed oil 0.915504766, coconut oil 0.869002225, peanut oil 0.826758802 and sunflower oil 0.77520504. However, olive oil has the lowest value of  $r$  which is 0.154381116. As shown in table 1 coconut oil, rapeseed oil, sunflower oil and soybean oil has short run relationship with CPO using Pearson correlation analysis whereas olive oil has no short run relationship with CPO price. In line with past researches[13], [30], this study exhibits that CPO price has positive and high correlation with soy bean oil prices with a positive correlation coefficient value of 0.941741044. This results showed that an increase in the price of soy bean oil price will also result an increase in the price of the CPO. However, olive oil has the lowest value of  $r$  which is 0.154381116. This indicates olive oil price has insignificant correlation with CPO price. As a conclusion, we can say that coconut oil price, rapeseed oil price and sunflower oil price also has short run relationship with CPO whereas olive oil has no short run relationship with CPO price thus olive is excluded in the forecasting of CPO price.

**Table 1.** Correlation results between CPO price and other vegetable oil prices.

CPO price	Pearson Correlation Coefficient (r value for monthly data )
Soybean oil price	0.941741044
Sunflower oil price	0.775205041
Rapeseed oil price	0.915504766
Coconut oil price	0.869002225
Olive oil price	0.154381116
Peanut oil price	0.826758802

Crude oil price and exchange rate were also considered in this study to predict the CPO price as research were carried out in the past[13]to test this relationship . In analyzing the relationship between CPO price and exchange rate, a negative correlation coefficient was recorded (-0.3094641) as shown in table 2. This results showed that an increase in the exchange rate causes a decrease in the price of CPO [31]. Whereas the relationship between crude oil price and CPO price showed a positive correlation as r value of 0.80252 is obtained as demonstrated by [14].

**Table 2.** Correlation between CPO price, crude oil price and exchange rate.

CPO price	Pearson Correlation Coefficient (r value for monthly data )
Crude oil price	0.80252
Exchange rate	-0.3094641

Based on the above correlation analysis results, soy bean oil price, coconut oil price, rapeseed oil price, sunflower oil price and peanut oil prices together with crude oil price and exchange rate are used as main predictors of CPO price in this study.

Weka, the data mining tool was employed to apply different algorithms on multivariate time series data to forecast the monthly CPO price for short term. First, ANN based Multi-layer perceptron regressor (MLP) with activation function Approximate Sigmoid is applied to train the regression model. Number of hidden neuron is the main parameter and 2 hidden units was used. Next, Support vector regression (SVR) with Sequential minimal optimization (SMO) function for regression was used for CPO price forecasting. Lastly The Holt Winters exponential smoothing algorithm was applied to forecast the monthly CPO price. A CPO forecast price of five month ahead of last data was derived from the above three forecasting techniques.

### 3 Results and Discussion

Performance of the three forecasting techniques of multivariate time series was evaluated. Forecasting evaluation is done by comparing the evaluation metrics of MSE, RMSE ,MAE and MAPE [32] and the best model is determined by lowest error value of the evaluation metrics.

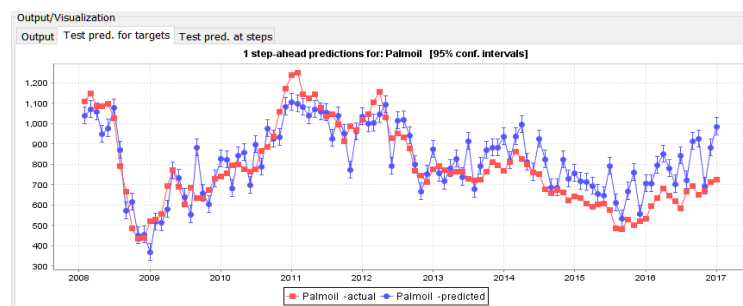
Direction accuracy (DA) is the number of times the movement of the predicted values matches the movement of the actual values, expressed as a percentage of the number of values predicted. The result for each algorithm is shown in table 3

**Table 3.** Comparison of the performance of forecasting techniques

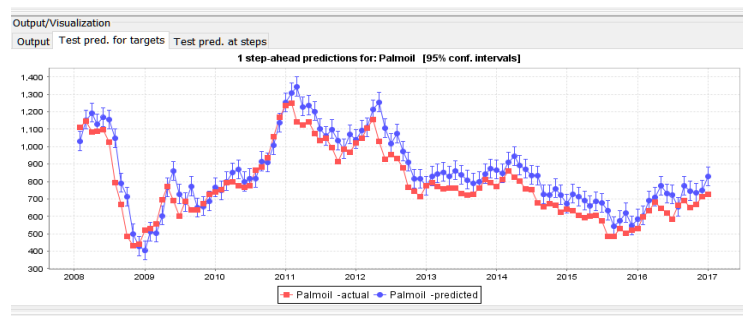
Forecasting Techniques Evaluation Metrics	MLP	SMO	Holt Winter
MAE	157.5563	59.15	252.431
DA	71.02	62.616	56.0748
MAPE	19.8991	7.8073	30.8689

RMSE	179.0278	76.4853	396.1274
MSE	32050.94	5849.9946	156916.92

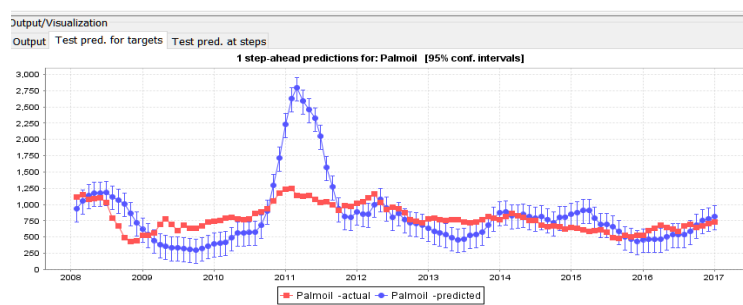
Table 3.0 shows the performance of the three forecasting method. MLP and SMO achieved better predictive accuracy than Holt Winter technique. MLP had MAPE of 19.89% and SMO had MAPE of 7.8% which are considerably lower than MAPE for Holt Winter technique of 30.86%. RMSE value for SMO was also lower compared to MLP and Holt Winter techniques. Analysis of DA shows MLP having the highest accuracy compared to SMO even though MAE and MSE of MLP is higher than SMO. In this study we found that SMO performed considerably better than MLP and Holt Winter techniques. SMO gives a promising result compared to MLP and Holt Winter. The actual and predicted CPO price trend can be viewed for each methods as shown in the graphs in figures 3, 4 and 5. The visual observation of the graphs shows that multivariate time series forecasting of CPO price has performed better using support vector regression.



**Figure 3.** Multilayer Perceptron



**Figure 4.** Support Vector Machine SMO



**Figure 5.** Holt Winter Exponential Smoothing.



#### 4 Conclusions

Support vector regression, multi-layer perceptron and Holt Winter exponential smoothing were utilized in this study to forecast the CPO price using multivariate time series. The prediction results exhibits that the support vector regression had higher predicted accuracy compared to multi-layer perceptron and Holt Winter exponential smoothing methods. In this study nine attributes were chosen and the results of this analysis showed the strength of support vector regression in forecasting multivariate time series of CPO price. In future, more relevant attributes could be included to improve forecasting of the CPO price. Feature selection method also can be added in future studies in order to improve accuracy of CPO price forecasting.

#### Acknowledgments

The authors would like to acknowledge the Ministry of Higher Education Malaysia (MOHE), Universiti Teknikal Malaysia Melaka (UTeM) and Universiti Teknologi Malaysia (UTM) for the financial support and research funding of this study.

#### References

- [1] P. Y. Gan and Z. D. Li, "Econometric study on Malaysias palm oil position in the world market to 2035," *Renew. Sustain. Energy Rev.*, vol. 39, pp. 740–747, 2014.
- [2] B. A. Talib and Z. Darawi, "An Economic Analysis of the Malaysian Palm Oil Market," *Oil Palm Ind. Econ. J.*, vol. 2, no. 1, pp. 19–27, 2002.
- [3] MPOC, Monthly Palm Oil Trade Statistics, *Malaysian Palm Oil Counc.*, <http://www.mpoc.org.my>, Accessed March 2017.
- [4] C. Kochaphum, S. H. Gheewala, and S. Vinitnantharat, "Does biodiesel demand affect palm oil prices in Thailand?," *Energy Sustain. Dev.*, vol. 17, no. 6, pp. 658–670, 2013.
- [5] M. N. Shamsudin and F. M. Arshad, "Short Term Forecasting of Malaysian Crude Palm Oil Prices," in <http://www.econ.upm.edu.my/-fatimah/pipoc-.html>, 2000, pp. 1–12.
- [6] S. Mohammadi, F. M. Arshad, B. K. Bala, and A. Ibragimov, "System Dynamics Analysis of the Determinants of the Malaysian Palm Oil Price," *Am. J. Appl. Sci.* 2015, 12 355.362 DOI 10.3844/ajassp.2015.355.362, pp. 1–8, 2015.
- [7] A. A. Karia and I. Bujang, "Progress Accuracy of CPO Price Prediction : Evidence from ARMA Family and Artificial Neural Network Approach," *Int. Res. J. Financ. Econ. ISSN 1450-2887 Issue 64 © EuroJournals Publ. Inc. 2011*, vol. 64, no. 64, pp. 66–79, 2011.
- [8] A. A. Karia, "Forecasting on Crude Palm Oil Prices Using Artificial Intelligence Approaches," *Am. J. Oper. Res.*, vol. 3, no. 2, pp. 259–267, 2013.
- [9] B. N. Zainal, "A Study on the Factors Affecting Crude Palm Oil (CPO) Price in Malaysia," in Available at SSRN: <https://ssrn.com/abstract=2279006> or <http://dx.doi.org/10.2139/ssrn.2279006>, 2010, pp. 1–10.
- [10] F. M. Arshad, A. Awad, and A. Hameed, "Review of Economics & Finance Crude Oil , Palm Oil Stock and Prices : How They Link 1," *Rev. Econ. Financ. - Toronto Better Adv. Press. ISSN 1923-7529, ZDB-ID 26371911. - 2013, 3, p. 48-57*, pp. 48–57, 2013.
- [11] C. Kantaporn, S. Sriboonchitta, S. Rahman, and A. Wiboonpongse, "Predicting Malaysian palm oil price using Extreme Value Theory," *Int. J. Agric. Manag. Vol. 2 Issue 2,2013 Int. Farm Manag. Assoc. Inst. Agric. Manag. ISSN 2047-3710.*, vol. 2, no. 2, pp. 91–99, 2013.
- [12] S. N. Abdullah, X. Zeng, C. Science, and O. Road, "Machine Learning Approach for Crude Oil Price Prediction with Artificial Neural Networks-Quantitative ( ANN-Q ) Model," in *The 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, 2010, pp. 1-8. doi:*



10.1109/IJCNN.2010.5596602, 2010, vol. 44, no. 0.

- [13] K. Chuangchid, A. Wiboonpongse, S. Sriboonchitta, and C. Chaiboonsri, "Factors Affecting Palm Oil Price Based on Extremes Value Approach," *Int. J. Mark. Stud.*, vol. 4, no. 6, 2012.
- [14] A. R. Hadi, M. H. Yahya, A. H. Shaari, M. H. Huridi, and U. K. Lumpur, "Investigating relationship between crude palm oil and crude oil prices – cointegration approach," *2nd Int. Conferfen Bus. Econ. Res. (2nd ICBER 2011) prooceeding*, no. 76, pp. 1554–1565, 2011.
- [15] S. D. Applanaidu, F. M. Arshad, M. N. Shamsudin, and A. Hameed, "An Econometric Analysis of the Link between Biodiesel Demand and Malaysian Palm Oil Market," *Int. J. Bus. Manag.*, vol. 6, no. 2, 2011.
- [16] A. A. Khin, Zainal Abidin Mohamed;, and C. Agamudai;, "Price Forecasting Methodology of the Malaysian Palm Oil Market," *Int. J. Appl. Econ. Financ.*, ISSN 1991-0886, vol. 7, no. 1, pp. 23–36, 2013.
- [17] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.
- [18] F. M. Arshad and R. Ghaffar, "Crude Palm Oil Price Forecasting : Box-Jenkins Approach," *Pertanika* 9(3), 359 - 367, vol. 9, no. 3, pp. 359–367, 1986.
- [19] B. Oancea and S. C. Ciucu, "Time series forecasting using neural networks," in *Proceedings of the "Challenges of the Knowledge Society" Conference*, pp. 1402-1408; eprint arXiv:1401.1333, 2013, pp. 1402–1408.
- [20] A. Hedayati, M. Hedayati, and M. Esfandyari, "Stock market index prediction using artificial neural network," *J. Econ. Financ. Adm. Sci.*, vol. 21, no. 41, pp. 89–93, 2016.
- [21] I. Koprinska, M. Rana, and V. G. Agelidis, "Knowledge-Based Systems Correlation and instance based feature selection for electricity load forecasting," *Knowledge-Based Syst.*, vol. 82, pp. 29–40, 2015.
- [22] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting," *Appl. Soft Comput. J.*, vol. 11, no. 2, pp. 2664–2675, 2011.
- [23] W.-C. Yeh, "New Parameter-Free Simplified Swarm Optimization for Artificial Neural Network Training and Its Application in the Prediction of Time Series," *IEEE Trans. NEURAL NETWORKS Learn. Syst.*, vol. 24, no. 4, pp. 661–665, 2013.
- [24] A. Khashman and N. I. Nwulu, "Support Vector Machines versus Back Propagation Algorithm for Oil Price Prediction," *Liu D., Zhang H., Polycarpou M., Alippi C., He H. Adv. Neural Networks – ISNN 2011. ISNN 2011. Lect. Notes Comput. Sci. vol 6677. Springer, Berlin, Heidelb.*, pp. 530–538.
- [25] H. Ince and T. B. Trafalis, "Short term forecasting with support vector machines and application to stock price prediction," *Int. J. Gen. Syst.*, vol. 37, no. 6, pp. 677–687, 2008.
- [26] V. N. Vapnik, "An overview of statistical learning theory.," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–99, 1999.
- [27] M. S. Baawain and A. S. Al-Serihi, "Systematic approach for the prediction of ground-level air pollution (around an industrial port) using an artificial neural network," *Aerosol Air Qual. Res.*, vol. 14, no. 1, pp. 124–134, 2014.
- [28] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat. Comput.*, vol. 14, pp. 199–222, 2004.
- [29] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting , Second Edition Springer Texts in Statistics*. 2002.

- [30] J. Othman, “Cointegration Between Palm Oil Price and Soybean Oil Price : A Study on Market Integration Mohammad Haji Alias,” *J. Ekon. Malaysia* 32(1998) 39-50, vol. 32, pp. 39–50, 1998.
- [31] T. Mielke and O. World, “Global Oil Supply , Demand and Price Outlook With Special Emphasis on Palm Oil,” in *Price Outlook Conference (POTS) in Iran on Feb 6, 2017*, 2017.
- [32] G. Zhang, B. E. Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks : The state of the art,” *Int. J. Forecast.* 14 35–62, vol. 14, pp. 35–62, 1998.