Crude Oil Price Forecasting Using Hybrid Support Vector Machine

Lee Jo Xian¹, Shuhaida Ismail¹, * Aida Mustapha², Mohd Helmy Abd Wahab³, Syed Zulkarnain Syed Idrus⁴

¹Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia.
²Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia.
³Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia.
⁴Faculty of Applied and Human Sciences, Universiti Malaysia Perlis, 01000 Kangar, Perlis, Malaysia.

* Corresponding Author: shuhaida@uthm.edu.my

Abstract. Crude oil price is strongly impacting the world economy. However, it is very fluctuated and difficult for investor to make decision. Hence, forecasting is one of the ways to minimizing risks arise from indecision on future. This paper will apply Support Vector Machine (SVM) and Artificial Neural Network (ANN) and a proposed hybrid model name Empirical mode decomposition-Support Vector Machine (EMD-SVM) forecasting crude oil price. After obtaining the forecasting result, performance evaluation is carry out to show which method can better forecast the crude oil price. The result shows that the performance of cruel oil price forecasting can be significantly increased by using the proposed hybrid EMD-SVM model. Thus proven the hybrid model are out-perform than individual forecasting model.

1. Introduction
West Texas Intermediate crude oil (WTI) is one of the main oil market in the world. The most actively traded world commodity is oil, contributing more that 10% of the total world trade [1]. Moreover, oil has always played a prominent role and fuels almost two-thirds of world energy consumption. This makes oil movement considerably impactful on the world economy especially in January 2004 However due to the volatility of crude oil price time series data, the current forecasting method approaches cannot capture all the factor influencing future crude oil price data. Thus, a method that can better improve the forecast accuracy of crude oil price should be considered. The new forecasting approach that can assist business practitioners and researchers deals with uncertainty future crude oil price. In short, the aim of this study are first to forecast monthly West Texas Intermediate crude oil price using Artificial Neural Network (ANN) and Support Vector Machine (SVM). Second, to generate the hybrid model of Empirical Mode Decomposition (EMD) and SVM in order to improve the forecasting accuracy of crude oil price market and third to evaluate the performance of the individual and hybrid proposed approaches.
2. Materials and Methods

Crude oil price is a standard price officially controlled and agreed upon globally. For this paper study, West Texas Intermediate (WTI) crude oil spot price data from the Energy Information Administration (EIA) website was chosen for experimental purposes. The crude oil price data used in this study is monthly data which covers the period from January 1986 until December 2015. By this mean, there would be a total of 360 observations for this study. The crude oil price are measure in Dollars per Barrel.

2.1. Data Preprocessing

A difference in unit is able to affect the difference in data magnitude consequently have a tendency to impact prediction accuracy in the long run. Thus, the normalization process is important to sort out this problem. In this study, the crude oil price data need to be transformed into a definite appropriate range \([0, 1]\) for prediction set via the normalization Equation (2.1) \[2\].

\[
x_i = \frac{(x - x_{\text{min}})}{(x_{\text{max}} - x_{\text{min}})}
\]  

(2.1)

where \(X_i\) is new \(x\) value, \(x\) is actual value, \(x_{\text{min}}\) is minimum actual value, \(x_{\text{max}}\) is maximum actual value, \(i = 1, \ldots, n\)

2.2. Support Vector Machine (SVM)

Support Vector Machines (SVM) was proposed by Vapnik \[3\]. SVM was initially intended for classification purposes (SVC). Besides classification, SVM also can perform regression and also time series forecasting. This technique is named support vector regression (SVR). The fundamental theory of SVM is structural risk minimization for reducing an upper bound of generalization error instead of an empirical error \[4\]. Generalization error is a measure accuracy degree of an algorithm in predicting outcome values for formerly unseen data. It can be reduced by avoiding overfitting. It is also the difference between error on the training set, expected error and error on the underlying joint probability distribution, empirical error. Besides, SVM models always used linear function to figure out the regression problem. It maps the data into a high-dimensional feature space though a nonlinear mapping \(\phi\) if facing a nonlinear regression in order to make a linear regression in this space.

\[
y(x) = w\phi(x) + b
\]  

(2.2)

The coefficients \(w\) and \(b\) are measured by minimize

\[
R(C) = C \frac{1}{N} \sum_{i=1}^{N} L_{\varepsilon}(d_i, y_i) + \frac{1}{2} ||w||^2 ,
\]  

(2.3)

\[
L_{\varepsilon}(d_i, y_i) = \begin{cases} 
|d - y| - \varepsilon & \text{for } |d - y| \geq \varepsilon \\
0 & \text{otherwise}
\end{cases}
\]  

(2.4)

where \(C\) is set parameters and to assess the trades-off between the empirical risk and the flatness of the model, \(\varepsilon\) is set parameter, \(L_{\varepsilon}(d_i, y_i)\) is \(\varepsilon\)-intensive loss function, \(d_i\) is real stock price during the \(i^{th}\) period, \(C \frac{1}{N} \sum_{i=1}^{N} L_{\varepsilon}(d_i, y_i)\) is the empirical error and \(1 \frac{2}{2} \sum w^2\) is flatness of the function.

Parameter \(C\) evaluated the trade-off between the empirical risk and the flatness of the model. Variables \(\xi\) and \(\xi^*\) with positive slack represented the distance from the actual values to the relevant boundary values of \(\varepsilon\)-tube. Equation (7) was transformed to the following formation:

Equation (2.5) is transformed to the following constrained formation by minimizing:

\[
R(W, \xi, \xi^*) = \frac{1}{2} w w^T + C \sum_{i=1}^{N} (\xi_i + \xi_i^*)
\]  

(2.5)

where \(R(W, \xi, \xi^*)\) is dual function, \(\xi\) and \(\xi^*\) are positive slack variables, which indicate the distance from the actual values to the corresponding boundary values of \(\varepsilon\)-tube.
subjected to:

\[ w\phi(x_i) + b_i - d_i \leq \epsilon + \xi_i, \]

\[ d_i - w\phi(x_i) - b_i \leq \epsilon + \xi_i^*, \]

where \( \xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \ldots, N. \) \( \quad (2.6) \)

Finally, presenting Lagrange multipliers and maximizing the dual function:

\[ R(\alpha_i - \alpha_i^*) = \sum_{i=1}^{N} d_i (\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) \]

with constraints

\[ \sum_{i=1}^{N} d_i (\alpha_i - \alpha_i^*) = 0, \]

\[ 0 \leq \alpha_i \leq C, \]

\[ 0 \leq \alpha_i^* \leq C, \]

\[ i = 1, 2, \ldots, N. \] \( \quad (2.7) \)

where \( \alpha_i \) and \( \alpha_i^* \) are Lagrangian multipliers. Lagrange multipliers satisfy the equalities, \( \alpha_i \times \alpha_i^* = 0 \)

\[ f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*)K(x_i, x_j) + b \]

\( \quad (2.10) \)

where \( K(x, x_j) \) represented kernel function.

The amount of the kernel is equivalent to the inner product of two vectors \( x_i \) and \( x_j \) in the feature space \( \phi(x_i) \) and \( \phi(x_j) \). Any function that fulfilling Mercer’s condition can be applied as the kernel function. The Gaussian kernel function is apply in this study. The SVM were used to evaluate the nonlinear behavior of the predicting data set because Gaussian kernels aim to present good performance under common efficiency smoothness assumption. The equation of Gaussian kernel function is

\[ K(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \]

\( \quad (2.11) \)

where \( x_i - x_j \) is squared Euclidean distance, \( \sigma \) is a Gaussian parameter.

2.3. Artificial Neural Network (ANN)

Artificial Neural Network is a computational method stimulated by studies of the brain and nervous systems in living organisms. It is also an input-output mathematical model inspired from operation of human brain by implementing the same approach of obtaining knowledge through learning procedure.

Multilayer perceptron means a feedforward network with one or more layers of nodes between the input and output nodes [5]. Basically it made up of three layers. First, the data is presented to the network via input layer. Next, data are processed in the hidden layer and finally reach the output layer where the results produced. A feed forward network is the input nodes in this illustration performing no computation but operates solely to distribute inputs into the network. The information in a feed forward network passes one way from the input layer, through the hidden layer and finally to the output layer. This study focus on three layer feed-forward backpropagation ANN. A three-layer MLP with \( n \) input nodes, \( q \) hidden nodes and one output node can be expressed as

\[ y_i = g \left( \sum_{i=1}^{q} w_i f \left( \sum_{i=1}^{n} w_{ij} x_{i-1} \right) \right), \]

\( \quad (2.12) \)

where \( y_i \) is the output layer, \( x_{i-1} \) is the input of the network, \( w_i \) is the connection weights between the input and hidden layer nodes, \( w_{ij} \) is the connection weights between hidden and output layer nodes, \( g(.) \) and \( f(.) \) are activation functions, \( g(.) \) as the sigmoid function, \( f(.) \) as the linear function.
Linear transfer function: \( f(x) = \text{purelin}(x) = xi \) \hspace{1cm} (2.13)

Log-Sigmoid Transfer function: \( g(x) = \log \text{sig}(x) = \frac{1}{1+\exp(-x)} \) \hspace{1cm} (2.14)

Sigmoid functions are mostly used in financial applications as transfer functions because of its threshold behaviour which describes most of financial and economical series [6]. Backpropagation algorithms were choose as it is the most common applied learning algorithm among all neural network paradigms [7]. The back propagation transmits the errors across the network and permits adaptation of the hidden units. This algorithm helps in minimizing the global error via the steepest descent or gradient approach. The network weights and biases are accustomed by raising the error function’s negative gradient for each iteration.

2.4. Model input determination
Model input selection is highlighted part for construction of data-driven model such as ANN and SVM as it contribute the elemental information about the system being modelled. There are doubt in many time series forecasting research on how they select appropriate model inputs since plenty of papers does not well describe the input determination methodology used. There are few way for select model input, the easier apply method is using autocorrelation (ACF) and partial autocorrelation (PACF), to identify the appropriate input variables [8]. The second method is ARIMA model proposed by Tang and Fishwick that the number of model inputs is the number of the autoregressive (AR) moving average components in the Box-Jenkins models [9]. Third method is stepwise method for determining the input for ANN models [10]. The method apply in this paper is by using trial and error according to minimum test errors in the ANN and SVM modelling [11]. By reference, sharda and patil proposed model inputs based on 12 inputs for monthly data and four for quarterly data heuristically [12]. Thus, the appropriate lags chosen using a trial-and-error approach \((x−1, x−2,..., x−p, \ where \ p \ is \ 2, 4, ..., 12).\) It gives the number of inputs (I) as 2, 4, 6, 8, 10 and 12.

2.5. Empirical mode decomposition
Huang et al. [13] is the one who first proposed the empirical mode decomposition (EMD) method. EMD is a signal analysis technique that could deal with nonlinear and nonstationary data. The key principle of this method is to decompose the original time series data into a sum of oscillatory functions so-called intrinsic mode function (IMF) by using the Hilbert-Huang transform (HHT). The IMFs kept the behaviour of the initial signal at different time scales. The EMD algorithm is implemented via R software package. IMFs have to fulfill the following two criteria:
(a) At any point, the envelope’s mean value built by the local maxima and minima is zero.
(b) In the data set, the number of zero-crossings and extreme values should be differs at most by one or equal. The following are time series data decompositions process:
1) Detect all the local extrema, comprising local maxima and local minima with \(y(t)\).
2) Select the lower envelope \(y_1(t)\) and upper envelope \(y_u(t)\) of \(y(t)\).
3) Calculate the first mean value \(\mu_1(t)\), that is, \(\mu_1(t) = [y_1(t) + y_u(t)]/2\).
4) Compute the difference between the original time series \(y(t)\) and the mean time series \(\mu_1(t)\). The first IMF \(q_1(t)\) is defined as \(q_1(t) = y(t) - \mu_1(t)\).
5) Assess whether \(q_1(t)\) satisfy the two settings of an IMF property. If not, repeat step 1 to 3 in order to find the first IMF.
6) After obtaining the first IMF, repeat the steps to determine the second IMF until achieve the final time series \(e_s(t)\), residual component that satisfy the termination criteria to stop decomposition process. Sum all the IMF components including the residual component to obtain the original time series \(y(t)\).
\[
y(t) = \sum_{i=1}^{n} q_i(t) + e_s(t)
\] \hspace{1cm} (2.15)

where \(\sum_{i=1}^{n} q_i(t)\) is summation of all the IMF, \(e_s(t)\) is residual component

2.6. Hybrid EMD-SVM
In a time series \( x_t, \ t=1, 2, \ldots, N \), one can predict \( h \)-steps ahead. For example \( X_{t+h}, h=1 \) means make a prediction of 1 step ahead and so on. First, the historical time series \( x_t, \ t=1, 2, \ldots, N \), is decomposed into an IMF component, \( q_i(t), i=1,2,\ldots,n \) and a residual component \( e_y(t) \) using an EMD extraction process. Then, the SVM model is used in modelling each of the IMFs and the residual component, and which predictions are made.

2.7. Evaluation of Performance Forecasts
Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) were used to check the accuracy of prediction. As these performance criteria utilize as good indicators and are most commonly tools of judging the quality of prediction of the proposed model. According to Legates & McCabe [14], there should exist at least one measure of absolute error for each model in performance evaluation. Besides, RMSE is very sensitive to small errors thus it is a good indicator of performance evaluation.

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |(\hat{y}_i - y_i)| 
\]

\[
\text{RMSE} = \left[ \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \right]^{1/2} 
\]

\[
\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| 
\]

where \( y_i \) is the actual values for period \( t \), \( \hat{y}_i \) is the forecast values for period \( t \), \( n \) is the number of observations.

3. Results and Discussions

Table 3.1. The Comparison Forecasting Performance between ANN and SVM

| Model | SVM | ANN | EMD-SVM |
|-------|-----|-----|---------|
|       | MAE | MAPE | RMSE    | MAE | MAPE | RMSE    | MAE | MAPE | RMSE    |
| 1     | 10.0641 | 25.5275 | 11.7983 | 8.6763 | 21.7372 | 10.1270 | 2.5257 | 5.3781 | 2.9582 |
| 2     | 9.1604  | 22.9271 | 10.4517 | 9.3463 | 22.7999 | 10.8530 | 2.7086 | 5.8425 | 2.9212 |
| 3     | 8.8563  | 22.1598 | 10.2372 | 8.1417 | 19.7471 | 9.77029 | 2.2777 | 4.9518 | 2.5470 |
| 4     | 9.4557  | 23.4662 | 11.0903 | 8.1157 | 20.4360 | 9.5888 | 2.8649 | 6.2592 | 3.3126 |
| 5     | 10.7757 | 27.3727 | 13.0388 | 11.0076 | 27.2510 | 12.7427 | 2.8396 | 6.3675 | 3.5242 |
| 6     | 7.2411  | 18.4448 | 9.1270  | 11.3096 | 27.5038 | 12.6465 | 2.8108 | 6.2925 | 3.4942 |

Result from Augmented Dickey Fuller test and Anderson Darling test prove that the time series data is nonstationary and nonlinear. By comparing the overall performance of forecasting result, SVM perform better than ANN although ANN is quite competitive with SVM. Hence, SVM is a better forecasting method of crude oil price than ANN. The result of hybrid EMD-SVM is as we expected, its result showed better than individual ANN and even SVM. Thus this result of this research also gain more accuracy in predicting data of crude oil price and can used to replace the old forecasting crude oil price method. This will also directly aid in stabilizing the economy and the inflation rate.

4. Conclusion and Discussion
The cruel oil time series often is highly non-stationary and non-linearity. Therefore, it may have poor prediction performance from applying traditional statistical models. Thus to improve the performance of cruel oil forecasting, a hybrid model based on a combination of EMD and SVM was proposed to predict monthly cruel oil price. The data are transformed into ranged [0,1] before introduced into model. In the proposed model, the EMD extraction process is first used to decompose the series of
stable IMF that are each predicted by separate SVM models. The prediction results from these IMF are combined together to obtain a single SVM model thus obtaining the predicted result. The result shows that the performance of cruel oil price forecasting can be significantly increased by using the proposed hybrid EMD-SVM model. To illustrate the performance of the EMD-SVM model, the indication of performance evaluation such MAPE, MAE and RMSE was apply. The performance of the EMD-SVM are outperform than the individual ANN and SVM methods. In overall proved that hybrid method can increased the forecast accuracy and it is an effective tool as a very promising methodology for complex with highly non-stationarity and non-linearity.

5. References
[1] Verleger P K 1994 Adjusting to volatile energy prices. United State of America: Peterson Institute for International Economics
[2] Ahmed R A and Shabri A B 2014 Daily crude oil price forecasting model using arima, generalized autoregressive conditional heteroscedastic and support vector machines American Journal of Applied Sciences 11 3 pp 425
[3] Vapnik V 1995 The nature of statistical learning theory New York: Springer-Verlag
[4] Vapnik V N and Chervonenkis A J 1974 Theory of pattern recognition Nauka, Moscow
[5] Lippmann R 1987 An introduction to computing with neural nets Assp magazine 4 2 pp 4-22
[6] McNelis P D 2005 Neural networks in financial: Gaining predictive edge in the market New York: Academic Press
[7] Zou H F, Xia G P, Yang F T and Wang H Y 2007 An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting Neurocomputing 70 16 pp 2913- 2923
[8] Aqil M, Kita K and Macalino M 2006 A Preliminary study on the suitability of data driven approach for continuous water laeve modeling Int. J. Comp. Sci 1 pp 246–252
[9] Tang Z and Fishwick P A 1993 Feedforward neural nets as models for time series forecasting ORSA Journal on Computing 5 4 pp 374-385
[10] Refenes A, Burgess A N and Bents Y 2003 Neural networks in financial engineering: A study in methodology IEEE T. Neural Network 8 pp 1223–1267
[11] Behzad M, Asghari K, Eazi M and Palhang M 2009 Generalization performance of support vector machines and neural networks in runoff modeling Expert Syst. Application 36 pp 7624–7629
[12] Sharda R and Patil R B 1992 Connectionist approach to time series prediction: an empirical test J. Intell. Manufuring 3 pp 317–323
[13] Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q and Liu H H 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis A: Mathematical, Physical and Engineering Sciences. London: Royal Society pp 903-995
[14] Legates D R and McCabe G J 1999 Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation Water Resources Research 35 1 pp 233-241