Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 12, Number 4 (2016), pp. 3009–3020

http://www.ripublication.com/gjpam.htm

Prediction of Crude Oil Prices using Support Vector Regression (SVR) with grid search – cross validation algorithm

Hasbi Yasin

Department of Statistics,
Universitas Diponegoro
Jl. Prof. Sudarto, SH,
Semarang 50275, Indonesia.

Rezzy Eko Caraka

Indonesia Endowment Fund For Education,
Ali Wardhana Building 2nd Floor,
Lapangan Banteng Timur 10710, Indonesia.

Tarno

Department of Statistics,
Universitas Diponegoro,
Jl. Prof. Sudarto, SH, Semarang 50275, Indonesia.

Abdul Hoyyi

Department of Statistics,
Universitas Diponegoro,
Jl. Prof. Sudarto, SH, Semarang 50275, Indonesia.

Abstract

The aim of this research is forecasting crude oil prices using Support Vector Regression (SVR). Algorithm to determine the optimal parameters in the model using the SVR is a grid search algorithm. This algorithm divides the range of parameters to be optimized into the grid and across all points to get the optimal parameters. In its application the grid search algorithm should be guided by a number of performance metrics, usually measured by cross-validation on the training data. Therefore, it is advisable to try some variations pair hyperplane parameters on SVR. Based on analysis calculation of accuracy and the prediction error using the training data generating R^2 99.10868% while the value of MAPE by 1.789873%. The data testing generates R^2 96.1639% while the value of MAPE by 1.942517%. This indicates to the data of testing using a linear kernel or accuracy of prediction accuracy results are quite large. Best model using the SVR has been formed can be used as a predictive model of crude oil prices. The results obtained showed crude oil prices from period 1 up to 10 experiencing decline.

AMS subject classification:
Keywords: Crude Oil Prices, SVR, Kernel, Grid Search, Cross Validation.

1. Introduction

Final energy consumption in Indonesia for the period 2000–2012 increased by an average of 2.9% per year. The most dominant type of energy is petroleum products which include aviation fuel, avgas, gasoline, kerosene, diesel oil, and fuel oil. These types of fuel consumed mostly by the transport sector. Today, most of the fuel prices are still subsidized. Fuel subsidies in 2013 have reached 199 trillion rupiahs. The government is also still subsidizing electricity for a particular type of users. Total electricity subsidies in 2013 reached 100 trillion rupiahs. The energy subsidy (fuel and electricity) has been increasing steadily. Energy subsidies in 2011 amounted to 195.3 trillion rupiahs and increased to 268 trillion rupiahs in 2013. Total spending on energy subsidies is always greater than the allocated budget and it often causes problems by the end of each fiscal year. Caraka and Yasin (2014) introduced the government has issued a number of policies to reduce petroleum fuel usage. Crude oil price is based on January 2016 data with 22.48 $/barrel (current price) and it assumed to be rising linearly to 40 $/barrel in the end of 2016. Oil production continues to decline while the demand for energy continues to grow which led to the increase in import of crude oil and petroleum products. This was shown by the deficit 3,5 billion Dollar at oil account in the second quarter which increased from 2,1 billion Dollar deficit in the first quarter of 2014 financial year. On the other hand, fuel subsidy is relatively high, due to increased domestic consumption, the increase in international oil prices and the decline in the exchange rate against the dollar and other foreign currencies. It is estimated that fuel subsidies until the end of 2014 will exceed the budget allocation in 2014. Since the publication of the 2015 edition of the WOO in November last year, the most obvious market development has been the oil price collapse. While the average price of the OPEC Reference Basket (ORB) during the first half of 2014 was over $100/barrel, it dropped to less than $60/b in December 2014 and has averaged close to $53/b in the first nine months of 2015. This new oil price environment has had an impact on both demand and supply prospects in the short- and medium-term, and some lasting effects can be expected in the long-term. Crude oil prices are expected to remain low as supply continues to outpace demand in 2016 and more crude oil is placed into storage. EIA estimates that global oil inventories increased by 1.9 million b/d in 2015, marking the second consecutive year of inventory builds. Inventories are forecast to rise by an additional 0.7 million b/d in 2016 before the global oil market becomes relatively balanced in 2017. The first forecasted draw on global oil inventories is expected in the third quarter of 2017, marking the end of 14 consecutive quarters of inventory builds. In the time domain, the long memory is indicated by the fact that the oil prices eventually exhibit strong positive dependence between distant observations. A shock to the series persists for a long time span even though it eventually dissipates. In the frequency domain, the long memory is indicated by the fact that the spectral density becomes unbounded as the frequency approaches zero.

2. Basic Concept of SVR

Santosa (2007) explained that Support Vector Machines (SVM) is a technique to make predictions, both in the case of classification and regression. SVM with linear classifier has a basic principle which is the case classification linearly separable, but SVM has been developed in order to work on a non-linear problem by incorporating the concept of the kernel in high-dimensional space (Gunn, 1998). By using the concept of ε-insensitive loss function, which was introduced by Vapnik, SVM can be generalized to approach the function or regression Support Vector Regression (SVR) is the application of SVM for regression case. In the case of regression output as real numbers or continuous. SVR is a method that can overcome the overfitting, so it will produce a good performance. Suppose there are l training data $(\boldsymbol{x_i}, y_i)$, $i = 1, \ldots, l$ where $\boldsymbol{x}_i$ an input vector $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_n\} \subseteq \Re^n$ and scalar output $= \{y_i, \ldots y_l\ \} \subseteq \Re$ and l is the number of training data. With the SVR, we want to assign a function $f(x)$ which has the greatest deviation ε from the actual targety_i, for all of the training data. If the value of ε is equal to 0 then obtained a good regression equation.

The main purpose of SVR is to map the input vector into the higher dimensions. Suppose a function below the regression line as optimal hyperplane:

$$f(\boldsymbol{x}) = \boldsymbol{w}^T \varphi(\boldsymbol{x}) + b \tag{1}$$

with

w = weight vector with l dimensional

$\varphi(\boldsymbol{x})$ = the function of mapping x on with l dimensional space

b = bias

In the regression models there are have residual, suppose residual (r) is defined by subtracting the estimated output scalar y to $f(x)$ that is $r = y - f(\boldsymbol{x})$ with

$$E(r) = \begin{matrix} 0 & \text{for } |r| \le \varepsilon \\ |r| - \varepsilon & \text{otherwise} \end{matrix} \tag{2}$$

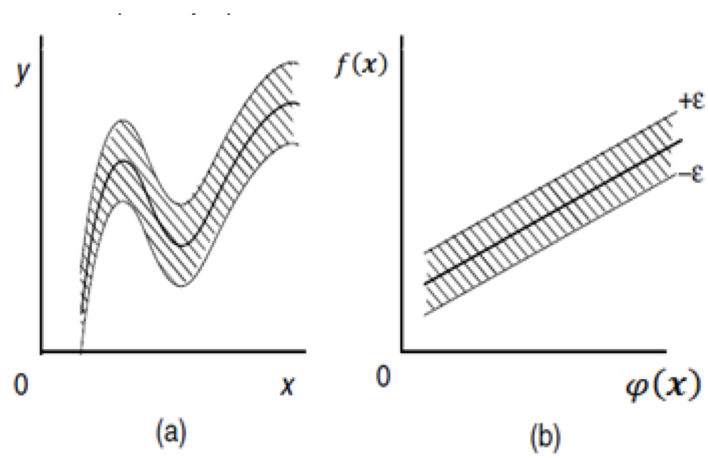

Figure 1: Insensitive Zone (a) Original Input Space and (b) Feature Space.

With ε is a small positive value. Suppose the residuals of the linear regression determined output estimate y and $f(\boldsymbol{x})$ with:

$$r = D(\boldsymbol{x}, y) = y - f(\boldsymbol{x}) \tag{3}$$

Based on the equation (3) a good estimate will be obtained when all of the absolute residual values are at an interval of ε.

$$-\varepsilon \leq D(\boldsymbol{x}, y) \leq +\varepsilon \tag{4}$$

Can be written:

$$|D(\boldsymbol{x}, y)| \leq \varepsilon \tag{5}$$

Figure 1 (a) illustrates the original input and output space, if all the training data are among ε it will generate an estimated accordingly. Zones within the range of ε is called ε-insensitive zone. Figure 1(b) illustrates the input - output feature space with new dimensions so that the image becomes more linear throughout the training data are assumed to satisfy the inequality (5). $D(\boldsymbol{x}, y) = \pm\varepsilon$ Support vector is the farthest distance from the hyperplane, then called margin. Maximizing margins will increase the probability of data into a radius of $\pm\varepsilon$. The distance from the hyperplane $D(\boldsymbol{x}, y) = 0$ to the data $(\boldsymbol{x}, y)$ is $||D(\boldsymbol{x}, y)| / | |\boldsymbol{w}^*| |$, where:

$$\boldsymbol{w}^* = (\boldsymbol{1} - \boldsymbol{w}^T)^T \tag{6}$$

It is assumed that the maximum distance the data to the hyperplane is δ, then estimates the ideal would be met with:

$$\frac{|D(\boldsymbol{x}, y)|}{\boldsymbol{w}^*} \leq \delta \tag{7}$$

$$|D(\boldsymbol{x}, y)| \leq \delta\ \boldsymbol{w}^* \tag{8}$$

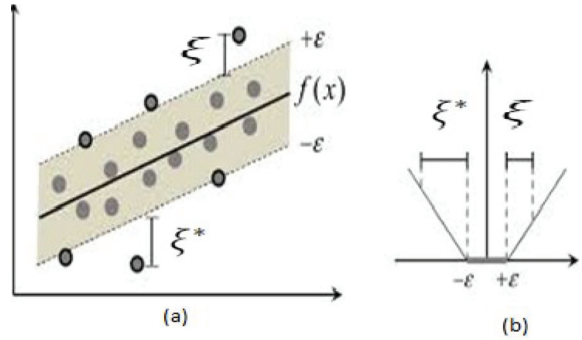

Figure 2: (a) SVR Output, and (b) ε-Insensitive Loss Function.

In inequality (5) and (8) the data that is farthest from the hyperplane filled with $|D(\boldsymbol{x}, y)| = \varepsilon$ then obtained:

$$\delta\ \boldsymbol{w^*} = \varepsilon \tag{9}$$

Therefore, to maximize the margin δ, required $\boldsymbol{w^*}$ minimum. While $\boldsymbol{w}^{*2} = \boldsymbol{w}^2 + 1$ then minimize $\|$ w $\|$ also made the maximum margin. Can be solved with Quadratic Programming:

$$min\frac{1}{2} \parallel \boldsymbol{w} \parallel^2 \tag{10}$$

With

$$y_i - w^T\varphi(\boldsymbol{x_i}) - b \leq \varepsilon \text{ for } i = 1, \ldots, l$$
$$w^T\varphi(\boldsymbol{x_i}) - y_i + b \leq \varepsilon \text{for} \quad i = 1, \ldots, l \tag{11}$$

Factors $||\boldsymbol{w}||^2$ is called regulation. Minimize $||\boldsymbol{w}||^2$ will create a function as flat as possible, so that it can control the capacity of the function. In Equation (**??**) assumed that all points in the range of $f(\boldsymbol{x}) \pm \varepsilon$ is feasible, in the case of infeasibility , where there are some points that may be out of range of $f(\boldsymbol{x}) \pm \varepsilon$ then added slack variables ξ and ξ^* to overcome the problem of restrictions infeasible constraints in the optimization problem.

Figure 2 illustrates that all points outside the margin will be subject to penalties. Furthermore, the above optimization problem can be written as follows:

$$min\frac{1}{2} \parallel \boldsymbol{w} \parallel^2 + \mathtt{C} \int_{i=1}^{\ell} (\xi_i + \xi_i^*) \tag{12}$$

with

$$y_i - \boldsymbol{w}^T\varphi(\boldsymbol{x_i}) - b - \xi_i \leq \varepsilon, \qquad i = 1, \ldots, l \tag{13}$$
$$\boldsymbol{w}^T\varphi(\boldsymbol{x_i}) - y_i + b - \xi_i^* \leq \varepsilon, \qquad i = 1, \ldots, l$$
$$\xi_i, \xi_i^* \geq 0$$

Loss Function is a function that shows the relationship between the residual in how this residual are subject to penalties. Differences loss function will produce different

SVR formula. (Gunn, 1998) Introduced the Loss simplest function is ε-insensitive loss function as an approach to Huber's loss function which allows a series of support vector would be obtained

$$L_\varepsilon(y) = \begin{matrix} 0, \quad \text{for } |f(\boldsymbol{x}) - y| < \varepsilon \\ |f(\boldsymbol{x}) - y| - \varepsilon, \text{ otherwise} \end{matrix} \tag{14}$$

With the constant C > 0 determines the trade-off between flatness of function $f(\boldsymbol{x})$ and the upper limit of deviation is greater than ε is still tolerable. All deviations larger than ε will incur a penalty of$C$. Optimization solutions to equation (12) with a lower limit of inequality (13) Lagrange function is formed as follows:

$$Q\left(\boldsymbol{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*\right) = L$$

$$= \frac{1}{2}\boldsymbol{w}^2 + C\int_{i=1}^{l}\left(\xi_i + \xi_i^*\right) - \int_{i=l}^{l}\alpha_i(\varepsilon + \xi_i - y_i + w^T\varphi(x) + b)$$
$$- \int_{i=1}^{l}\alpha_i^*(\varepsilon + \xi_i^* + y_i - w^T\varphi(x) - b) - \int_{i=1}^{l}(\eta_i\xi_i + \eta_i^*\xi_i^*) \tag{15}$$

L Is called the *Lagrangian*, $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ the Lagrange Multiplier. To obtain an optimal solution, then the partial derivatives of Q to $w,\ b,\ \xi, \xi^*$

$$\alpha_i, \alpha_i^*,\ \eta_i, \eta_i^* \geq 0$$

$$\frac{\partial L}{\partial b} = \int_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0 \tag{16}$$

$$\frac{\partial L}{\partial w} = w - \int_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)\varphi(x_i) = 0 \tag{17}$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0 \tag{18}$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0 \tag{19}$$

From Equation (17) w can be written by:

$$w = \int_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)\varphi(x_i) \tag{20}$$

Then optimal hyperplane can be written:

$$f(x) = \int_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)\varphi^T(x_i)\varphi(x_i) + b \tag{21}$$

Let $\beta_\mathrm{i} = \left(\alpha_i - \alpha_i^*\right)$ then

$$f\left(\boldsymbol{x}\right) = \int_{i=1}^{l} \beta_i \varphi^T(\boldsymbol{x_i})\varphi(\boldsymbol{x_i}) + b$$

$\boldsymbol{w}$ Can be fully described, namely a linear combination of training data patterns. Dual solution to the equation (15) obtained by substituting Equation (18), (19), (20), (21) into the equation (15) by maximizing:

$$Q\left(\alpha, \alpha^*\right) = -\frac{1}{2}\int_{i,j=1}^{l}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)\varphi^T\left(\boldsymbol{x_i}\right)\varphi\left(\boldsymbol{x_j}\right) - \varepsilon\int_{i=1}^{l}\left(\alpha_i + \alpha_i^*\right) + \int_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \quad (22)$$

with

$$\int_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) = 0$$

$$0 \le \alpha_i \le C, \;\; 0 \le \alpha_i^* \le C \quad \text{for } i = 1, 2, 3 \ldots, l$$

The optimal solution for b using KTT (Karush Kuhn Tucker):

$$\alpha_i\left(\varepsilon + \xi_i - y_i + \boldsymbol{w}^T\varphi\left(\boldsymbol{x_i}\right) + b\right) = 0 \quad (23)$$

$$\alpha_i^*\left(\varepsilon + \xi_i^* - y_i - \boldsymbol{w}^T\varphi\left(\boldsymbol{x_i}\right) - b\right) = 0 \quad (24)$$

$$\eta_i\xi_i = \left(C - \alpha_i\right)\xi_i = 0 \quad (25)$$

$$\eta_i^*\xi_i^* = \left(C - \alpha_i\right)\xi_i^* = 0 \quad (26)$$

$$\text{for } i = 1, 2, 3 \ldots, l$$

From Equation (25) when $0 < \alpha_\mathrm{i} < C$, then $\xi_i = 0$, Equation (26) when $0 < \alpha_i * < C$, then $\xi_i^* = 0$. So to Equation (23) and (24) is given by the equation:

$$\varepsilon - y_i + \boldsymbol{w}^T\varphi(\boldsymbol{x}_i) + b = 0 \quad \text{for } 0 < \alpha_i < C \quad (27)$$

$$\varepsilon + y_i - \boldsymbol{w}^T\varphi(\boldsymbol{x}_i) - b = 0 \quad \text{for } 0 < \alpha_i^* < C \quad (28)$$

This means that the data with the residual$y - f\left(\boldsymbol{x}\right) = +\varepsilon$, α_i filled with $0 < \alpha_i < C$ and data with residual$y - f\left(\boldsymbol{x}\right) = -\varepsilon$, α_i^* filled with$0 < \alpha_i^* < C$. Then:

From Equation (27) obtained estimates of b is

$$b = y_i - \boldsymbol{w}^T\varphi(\boldsymbol{x_i}) - \varepsilon \text{ untuk } 0 < \alpha_i < C \quad (29)$$

From Equation (30) obtained estimates of b is

$$b = y_i - \boldsymbol{w}^{\boldsymbol{T}}\varphi(\boldsymbol{x}_i) + \varepsilon \text{ untuk } 0 < \alpha_i^* < C \quad (30)$$

Table 1: Grid Search – Cross Validation

Cost	Epsilon	Error	Dispersion
0.1	1.00E-01	1.290704	2.212861
1	1.00E-01	1.079881	1.787301
10	1.00E-01	1.069183	1.730493
100	1.00E-01	1.075229	1.732241
0.1	1.00E-02	1.258094	2.247218
1	**1.00E-02**	**1.066267**	**1.84728**
10	1.00E-02	1.072037	1.834004
100	1.00E-02	1.072784	1.833857
0.1	1.00E-03	1.208222	2.181309
1	1.00E-03	1.068912	1.856676
10	1.00E-03	1.066875	1.844631
100	1.00E-03	1.068787	1.846193
0.1	1.00E-04	1.217882	2.189697
1	1.00E-04	1.068931	1.855783
10	1.00E-04	1.067272	1.845259
100	1.00E-04	1.066932	1.844799

3. Simulation

According Santosa (2007) many of the techniques of data mining or machine learning developed with the assumption of linearity, so that the resulting algorithm is limited to cases linear. With a kernel method of data x in the input space is mapped to the feature space with the higher dimensions through φ.

$$\varphi : x \to \varphi(x)$$

Furthermore, the same process as the Linear SVM. The mapping process requires calculating the dot product of two vectors of data on a new feature space. The dot product of two vectors $(\boldsymbol{x}_i)$ and $(\boldsymbol{x}_j)$ is denoted as $\varphi(\boldsymbol{x}_i).\varphi(\boldsymbol{x}_j)$. The dot product value can be calculated without knowing the transformation function φ. Computational technique is called a kernel trick, which calculates the dot product of two vectors in space a new dimension by using both components of these vectors in the space dimension of origin, as follows:

$$\boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{\varphi}(\boldsymbol{x}_i).\boldsymbol{\varphi}(\boldsymbol{x}_j)$$

$\boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is a kernel function that shows a linear mapping in the feature space . Is worth highlighting that $\boldsymbol{K}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ cannot always be expressed explicitly as a combination of α, y and $\boldsymbol{\varphi}(\boldsymbol{x})$, since in many cases $\boldsymbol{\varphi}(\boldsymbol{x})$ is unknown or difficult to quantify. Kernel functions used in this study is the linear kernel with the following functions

Table 2: Values of w and b

Variabels	w	b
Yt_1	1.15081	−0.0063
Yt_2	−0.1568	

$\boldsymbol{\varphi}(\boldsymbol{x}) = \boldsymbol{K}(x, x') = \boldsymbol{x}^{\boldsymbol{T}}\boldsymbol{x}$. Yao (2014) introduced one algorithm to determine the optimal parameters in the model using the SVR is a grid search algorithm. This algorithm divides the range of parameters to be optimized into the grid and across all points to get the optimal parameters. In its application the grid search algorithm should be guided by a number of performance metrics, usually measured by cross-validation on the training data. Yasin (2014) explained it is advisable to try some variations pair hyperplane parameters on SVR. SVR adopts the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization. The optimal parameters of SVR can be use Grid Search Algorithm method. Concept of this method is using cross validation (CV). Based on this simulation using data of crude oil prices we can see from Tabel 1 the value of C and epsilon values obtained best by looking at the smallest error value of Grid Search – cross validation process. Cross-Validation is used in the LOO or Leave One Out in order to obtain the best parameter values are $C = 1$ and epsilon = 0.01.

From Table 2 it can be seen that the value of beta and w will be used in the search for support vector, support vector generated will be used to predict the equation SVR formed. Calculation of accuracy and the prediction error using the training data generating R^2 is 0.9910868 or 99.10868% while the value of MAPE by 1.789873%. The data testing generates R^2 is 0.9616390 or 96.1639% while the value of MAPE by 1.942517%. This indicates to the data of testing using the mapping function has a linear kernel or accuracy of prediction accuracy results are quite large, to the value of the error in the prediction process is relatively small.

Based on Figure 3 and Figure 4 we can see that the graph data in training and testing compared with the prediction vs actual have the same pattern, the results of the predictive value of crude oil prices almost entirely below it can be said that the model overfitting SVR can solve problems that usually occur in the data of time series. From the best model using the SVR has been formed can be used as a predictive model of crude oil prices, we wanted to know the price of 10 periods crude oil prices after the last data is taken . The results obtained showed crude oil prices from period 1 up to 10 experiencing decline. These results can describe the state of crude oil prices are unstable.

4. Conclusion

The oil price is partly determined by actual supply and demand, and partly by expectation. Demand for energy is closely related to economic activity. If producers think the price

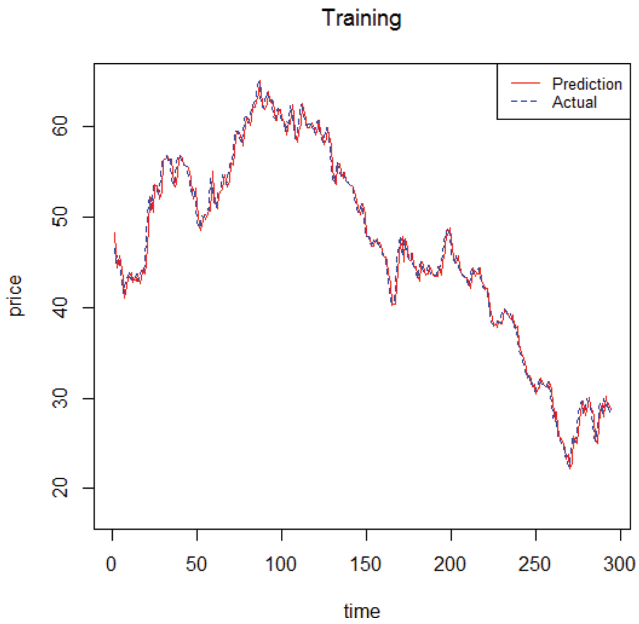

Figure 3: Training Phase of Crude Oil Prices.

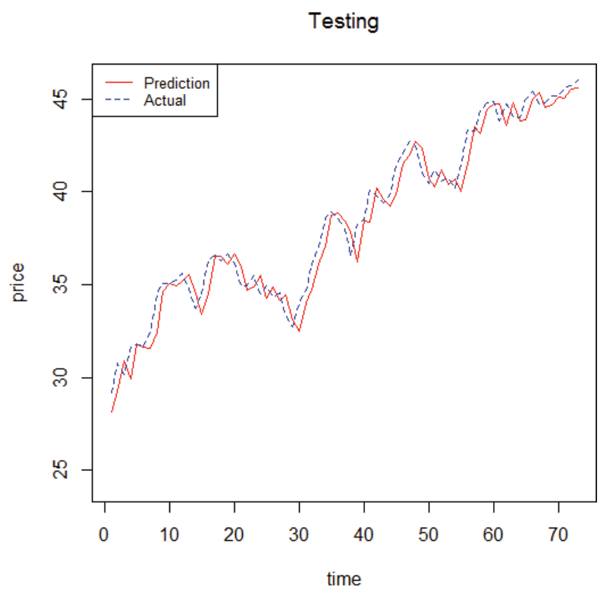

Figure 4: Testing Phase of Crude Oil Prices.

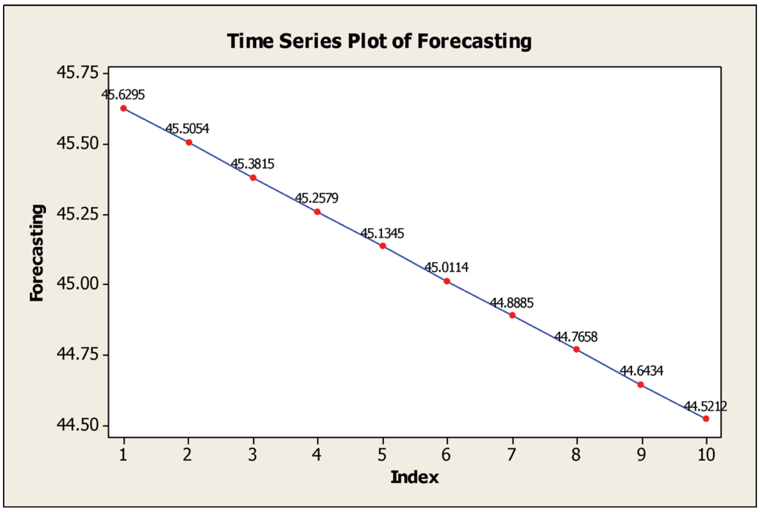

Figure 5: Grafik of Forecasting Crude Oil Prices Using SVR.

is staying high, they invest, which after a lag boosts supply. Similarly, low prices lead to an investment drought. Crude oil price is based on 2016 data with 44 $/barrel. Improvement in the economy will encourage petroleum fuel utilization, especially in the transport sector as its main user. This has to be supported by an adequate increase in crude oil supply. SVR can be used as an alternative method to forecasting crude oil prices and grid search as search algorithm to determine the optimal parameters in the model of SVR.

Acknowledgement

This research is supported by Directorate of Research and Community Services, The Ministry of Research, Technology, and Higher Education.

References

[1] Abe, S. 2005. Support Vector Machine for Pattern Classification. Springer-Verlag. London Limited.

[2] Caraka, R.E., Yasin, H. 2014. Prediksi Produksi Gas Bumi dengan General Regression Neural Network (GRNN), in Proc. National Statistics Conference (SNS IV). ISSN: 2087–2590, No 270–277.

[3] Gunn, S. 1998. Support Vector Machines for Classification and Regression. Technical Report, ISIS.

[4] Hamel, L. 2009. Knowledge Discovery With Support Vector Machine. United State Of America: A John Wiley & Sons, INC., Publication.

[5] Santosa, B. 2007. Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu.

[6] Scholkopf B., and Smola A. 2003. A Tutorial on Support Vector Regression.

[7] Yao, Y, et al. 2014. An Improved Grid Search Algorithm and Its Application in PCA and SVM Based Face Recognition, Journal of Computational Information Sistems. Vol. 10, No. 3:1219–1229.

[8] Yasin. H., Prahutama.A., and Utami, T.W. 2014. Prediksi Harga Saham Menggunakan Support Vector Regresion dengan Algoritmaa Grid Search, Statistika. Universitas Diponegoro. Media Statistika. Vol 7, No 1:29–35