Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

PREDICTION OF JAKARTA COMPOSITE INDEX USING LEAST SQUARES SUPPORT VECTOR MACHINES APPROACH

[1]ELLEN SUWANDI, [2]DIAZ D. SANTIKA
[1,2]Bina Nusantara University, School of Computer Science
E-mail: [1]suwandiellen@gmail.com , [2]ddsantika@yahoo.com

ABSTRACT

An accurate prediction of stock/index price is highly needed by investors and firms to help them make the right decision regarding to sell, buy, or hold their shares. Recently, a new method within the field of computational intelligence, i.e. Least Squares Support Vector Machines (LSSVM), has been investigated on its performance in predicting stock/index price. LSSVM algorithm itself is introduced as a modified version of SVM algorithm, where the computational complexity of Quadratic Programming (QP) problem in SVM is overcome by solving a set of linear equations instead. In this research, an LSSVM model is built to predict the daily close price of Jakarta Composite Index (JCI). As a comparison, other prediction models are built, namely SVM and ARIMA model which respectively represent the commonly used intelligent and conventional model. Experimental results show that LSSVM model outperforms SVM and ARIMA model in predicting the JCI close price.

Keywords: *Stock Index Prediction, Least Squares Support Vector Machines (LSSVM), Support Vector Machines (SVM), ARIMA, Jakarta Composite Index (JCI)*

1 INTRODUCTION

Prediction of stock index price is seen as one of the most challenging task of time-series prediction. It is often affected by many factors, such as political affairs, company regulations, economic conditions, etc [14],[25]. An accurate prediction of stock index price will, therefore, highly contributes to various parties involved in stock market. With such prediction, investors can anticipate potential investment risks, while speculators and arbitrators can optimize the profits they gained from trading stocks [15],[22].

At the time when computational intelligence has not been recognized by the economic society, predictions were done using traditional forecasting techniques, like ARIMA. One drawback of such models is on the assumption that financial data (e.g. stock prices) is of linear nature, while reality shows otherwise [11]. This problem has lead researchers to propose a computational intelligence based predictor, i.e. Artificial Neural Network (ANN) [20]. However, the use of Empirical Risk Minimization (ERM) in ANN leads to poor generalization, and not to mention, the high computational complexity as well as over fitting issue [27].

A prominent technique is introduced later by Vapnik, i.e. Support Vector Machines (SVM), which adapts Structural Risk Minimization (SRM) instead of ERM [27]. Several studies have proven that this method is more effective than ANN [22],[28]. Nevertheless, the employment of Quadratic Programming (QP) problem has caused computational issue in SVM. Due to that matter, Suykens et al [24] made some modifications to the original formulation of SVM, which is known as Least Squares Support Vector Machines (LSSVM).

The reformulation of SVM by Suykens was done by replacing QP problem with a set of linear equations. By utilizing least square loss function, a set of linear equations is obtained in dual space which results in lower calculation complexity [22]. Empirical studies show that LSSVM is able to perform predictions in various fields of interest with relatively low computational complexity yet better predictive performance compared to other predictive models, including SVM [8],[9],[22],[28].

Many studies have been done to predict of stock/index price in various countries, but only a few done in Indonesian Stock Exchange (IDX). Furthermore, those researches mostly utilize the conventional statistical forecasting techniques that suffer from linearity and therefore are not really helpful for developing a reliable prediction model. To that end, this research proposes the use of an intelligent model based on LSSVM approach to predict the close price of Jakarta Composite Index,

Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

which is a main stock index in IDX. The purpose is to propose a prediction model that is able to accurately predict the JCI price, thus favoring the investors and companies by providing them the knowledge useful for deciding whether to buy, sell, or hold their shares. As a benchmark, the predictive performance of LSSVM model will be compared with that of SVM and ARIMA model, each of which has been one of the most widely used intelligent and conventional models.

Hence, the research question addressed in this research is: How good is the LSSVM-based model at predicting Jakarta Composite Index price in comparison with SVM- and ARIMA-based model?

2 RELATED WORKS

The large amount of data produced by stock market has demanded that researchers use the techniques of data mining for this purpose [23]. Many researchers have tried to predict stock trends by using conventional statistical approach, like the use ARIMA model to forecast gold bullion coin selling price by Abdullah [1], electricity price forecasting using ARIMA by Jakaša et al [12], and stock prices prediction with the combination of regression analysis and moving average by Olaniyi et al [21]. However, the performance of such approach is limited by the linearity of the model used.

A more promising alternative is to use the CI-based techniques, as in the studies by Akinwale et al [2], who used ANN to predict stock prices in Nigerian Stock Exchange, and Magaji et al [16], who also achieved the same purpose but using the Naïve Bayes method. Kannan et al [13] proposed an algorithm that is able to predict the direction of stock closing price movement on the following day by combining five stock analysis methods. Kara et al [14] compared the performance on ANN and SVM in predicting stock index movement in Istanbul Stock Exchange, and finally Kumar & Thenmozhi [15] compared SVM and Random Forest in predicting S&P CNX NIFTY Market Index movement.

Several other studies that are specifically more relevant to this research were carried out by Deng & Yeh [8] who used LSSVM to estimate product cost, Gestel et al [9] who reported that LSSVM is a better predictor for financial time series compared to AR and other nonparametric model, and Yu et al [28] who concluded from their experiments that LSSVM outperforms ARIMA, LDA, BPNN, and SVM model in predicting stock price movement. Finally, a more comprehensive work was conducted by Ou & Wang [22] by comparing ten data mining techniques to predict Hang Seng index price movement and experimental results show that SVM and LSSVM model outperforms the eight other models, but specifically, SVM is better than LSSVM for in-sample prediction, whereas for out-sample prediction, LSSVM is better than SVM.

3 LITERATURE REVIEW

Basically, SVM is a machine learning algorithmic technique that aims to generate an optimal separating hyper plane, which is the hyper plane with the largest margin. The original formulation of SVM by Vapnik & Lerner in 1963 and Vapnik & Chervonenkis in 1964 was built to handle linear or separable data [24]. Further formulation for nonlinear case was introduced by Vapnik [27] in 1995 by including the mapping of input data into unlimited high-dimensional feature space, as illustrated by Figure 1. Separating hyper plane will then be constructed in that high-dimensional feature space.

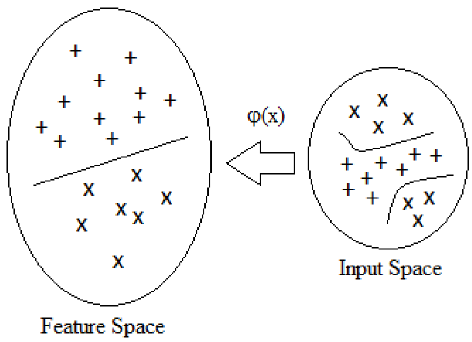

Figure 1: Mapping of Input Space into High-Dimensional Feature Space

With kernel technique, a nonlinear decision function is constructed in the input space, which is equivalent to the linear decision function in the high-dimensional feature space. Several options are available for the kernel function such as linear, polynomial, and Radial Basis Function (RBF) [24].

In SVM, regression problem is formulated as convex quadratic programming (QP) problem. However, this formulation has caused computational complexity [19]. To solve that problem, Suykens et al. proposed a reformulation of SVM, i.e. LSSVM, which employs equality constraints rather than inequality constraints and a least square error term in order to obtain a set of linear equations in the dual space. This formulation has been proven in many studies to have good generalization and consume less computational resource compared to SVM [20],[22],[28].

Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The optimization problem of LSSVM by Suykens et al [24] is expressed as

$$\min_{w,b,e} J_p(w,e) = \frac{1}{2} w \cdot w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 \qquad (3.1)$$

such that

$$y_i = w \cdot \varphi(x_i) + b + e_i, \quad i = 1, \dots, N \qquad (3.2)$$

Decision function of the model in primal space takes the form

$$y(x) = w \cdot \varphi(x) + b \qquad (3.3)$$

where $\varphi(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^{n_h}$ is the mapping into high-dimensional feature space as in the standard SVM.

In the case of nonlinearity, Lagragian form takes place,

$$L(w,b,e;\alpha) = J_p(w,e) - \sum_{i=1}^{N} \alpha_i \{ w \cdot \varphi(x_i) + b + e_i - y_i \} \qquad (3.4)$$

where α_i is the Lagrange multipliers.

The conditions for optimality are

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{N} \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma_{e_i}, \quad i = 1, \dots, N \\ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w \cdot \varphi(x_i) + b + e_i - y_i = 0, \quad i = 1, \dots, N \end{cases} \qquad (3.5)$$

By eliminating w and e, this following linear system is obtained,

$$\begin{bmatrix} 0 & 1_v^T \\ y & \Omega + I/\gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \qquad (3.6)$$

where

$$y = [y_1; \dots; y_N]$$
$$1_v = [1; \dots; 1]$$
$$\alpha = [\alpha_1; \dots; \alpha_N]$$

Kernel tricks can be applied to the matrix Ω,

$$\Omega_{ij} = \varphi(x_i) \cdot \varphi(x_k) = K(x_i, x_j)$$

for $i,j = 1, \dots, N$.

Hence, the LSSVM model generated in the dual space is defined as

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \qquad (3.7)$$

As in the SVM model, several kernel options are available for LSSVM model, such as linear, polynomial, and Radial Basis Function (RBF). In this research, the RBF kernel is used as it tends to give good performance [22]. RBF kernel is expressed as:

$$K(x, x_i) = \exp(-\gamma \|x - x_k\|_2^2 / \sigma^2) \qquad (3.8)$$

To build a prediction model using machine learning based technique, like SVM and LSSVM, one should prepare a data set containing the variable on question as well as other relevant variables. That data set is then split into two, i.e. training set and testing set [11]. Training set is used to build model. The process is called training process, within which the proper parameter values are specified for the model. Predictive performance of the obtained model is then measured by comparing the predicted values with their corresponding actual values in testing set. This process is called testing process. Sometimes, it is important to estimate the goodness of the model built before it is used for testing in order to avoid overfitting. For this purpose, another set, called validation set, is used and the process is called validation process [11].

As opposed to SVM and LSSVM which are based on artificial intelligence, ARIMA is a conventional univariate time series prediction model, where it uses only one series of observation and the model is generated by constructing a linear function of past values of the series and/or previous random shocks [26].

ARIMA model development involves three primary steps [1],[26]. The first step is model identification, which is to determine whether time series is stationary and the possible transformations, i.e. degree of differencing (d) that should be applied to obtain stationarity.

The second is parameter estimation, which is to determine the order of autoregressive (AR) and moving average (MA) of the time series, denoted as p and q. Another set of parameters analogous to these are required in the case of seasonal ARIMA, denoted as P, D, and Q. The general notation of a seasonal ARIMA has the form ARIMA(p,d,q)x(P,D,Q)$_s$ where s is the periodicity [4]. The characteristics of the time series trend as well as its autocorrelation (ACF) and partial autocorrelation functions (PACF) are usually analyzed to identify the optimal values of those parameters [3].

The final step is diagnostic checking to verify the model. A well fitted model is expected to have residuals with characteristics of white noise. Ljung-Box Q statistic is commonly used to test the null hypothesis of the residual normality stating that errors are random or white noise [1]. For a model to be confirmed as fit, the test should give a Q value less than the critical value of chi-square distribution with s degree of freedom, where s is the length of coefficient in the test [7]. Besides Ljung Box Q

Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

statistic, other indicators such as BIC (see Equation (4.2)) are also used in the selection of ARIMA model.

4 MATERIAL AND METHODS

4.1 Data Collection and Preparation

Data set used in this research comprises four prices of Jakarta Composite Index (i.e. open price, close price, high price, and low price), gold fixing price, and WTI crude oil price, which are gathered from the website of Yahoo Finance (www.finance.yahoo.com), London Bullion Market Association (www.lbma.org.uk), and US Department of Energy (www.eia.gov) respectively. All of those the time series are in daily basis, in the time span between January 8, 2009 and December 11, 2013, which provides 1200 data points. The period is chosen considering the low and stable inflation rate, based on the statistics by BPS-Statistics Indonesia (www.bps.go.id).

4.2 Data Correlation

To determine the correlation of each time series in the data set, Spearman Correlation is utilized since the data set is not normally distributed based on the result of Shapiro-Wilk normality test [18]. The result of Spearman Correlation test is as shown in Table 4.1, from which we can conclude that every time series in the data set are highly correlated to each other because all the significance values are less than 0.01 and the coefficient is close to 1.

Table 4.1: Correlation between Utilized Time Series

		JCI Close	JCI Open	JCI High	JCI Low	Gold A.M.	WTI Crude Oil
JCI Close	Correl. Coef.	1.000					
	Sig. (2-tailed)	0.000					
JCI Open	Correl. Coef.	0.998	1.000				
	Sig. (2-tailed)	0.000	0.000				
JCI High	Correl. Coef.	0.999	0.999	1.000			
	Sig. (2-tailed)	0.000	0.000	0.000			
JCI Low	Correl. Coef.	0.999	0.999	0.999	1.000		
	Sig. (2-tailed)	0.000	0.000	0.000	0.000		
Gold A.M.	Correl. Coef.	0.707	0.707	0.707	0.706	1.000	
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	
WTI Crude Oil	Correl. Coef.	0.772	0.770	0.770	0.773	0.665	1.000
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000

4.3 Data Proportion

Before used for experiment, data is split into three different sets, i.e. training, validation, and testing set. Training set is used to build the prediction model which is then validated using the validation set. Final performance of the model will then be assessed using the testing set. The proportion of each set is tabulated in Table 4.2.

Table 4.2: Distribution of Data for Training, Validation, and Testing Set

Data Set	Period	Number of Observations
Training	8 Jan 2009 - 30 Dec 2011	729
Validation	10 Jan 2012 - 28 Dec 2012	239
Testing	10 Jan 2013 - 11 Dec 2013	222

4.4 Input and Output Variables

During the training process, LSSVM and SVM model will "learn" from some data that are relevant to the data to be predicted. Those data constitute input variables, whilst the data to be predicted constitutes output variable.

Since stock index prediction is a time series prediction, past values in the series have to be considered. Hence, the determination of input variables should involve windowing process, which is selecting the number of samples to be used as successive inputs for the model to predict the subsequent value. For instance, if the window size chosen is three (3), then the prediction output, P(T), is predicted with the input being three successive samples, P(T-1), P(T-2), and P(T-3).

In this research, the window size chosen is five (5), considering that stock trading in the Indonesian Stock Exchange is held five (5) days in a week. In other words, prediction is performed using the data of the previous week. The input variables to be used in this research are then defined as follow:

- Open (T-1) – Open (T-5): JCI open price one to five days in advance.
- Close (T-1) – Close (T-5): JCI close price one to five days in advance.
- High (T-1) – High (T-5): JCI high price one to five days in advance.
- Low (T-1) – Low (T-5): JCI low price one to five days in advance.
- Gold (T): Gold fixing price on the respective day.
- Oil (T): WTI crude oil price on the respective day.

In this research, the one-step prediction approach [6] is applied to the proposed model. For the time series to be predicted, the past values up to

Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the time prior to the prediction point are used as the input and the output is the next upcoming data value, denoted as:

$$S(X_t) = (X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}, \ldots)$$
$$S(X_{t+1}) = (X_t, X_{t-1}, X_{t-2}, X_{t-3}, \ldots)$$
$$\ldots$$

Hence, the output of the proposed model is the JCI close price on the respective day, i.e. Close (T).

4.5 Model Selection

Model selection in this research is concerned with the selection of optimal parameters for each predictive model, i.e. LSSVM, SVM, and ARIMA model. The parameters required by LSSVM and SVM model are (γ, σ^2) and (C, σ^2, ε), respectively. However, in this research, the SVM model selection is focused on optimizing (C, σ^2), as the more critical parameters [10], while the parameter ε will be assigned with some default value. Meanwhile, for ARIMA model, the parameters to be defined are (p, d, q), which is the non-seasonal part of the model, and $(P, D, Q)_m$,which is the seasonal part of the model.

The optimal combination of parameters for LSSVM and SVM model is chosen using grid search technique. In LSSVM and SVM model selection, the values for each parameter to be tested are taken from some exponentially growing sequences, for example the values tested for LSSVM's γ are $\{2^{-8}, 2^{-1}, \ldots, 2^{20}\}$. The combination chosen is the one that produces the least RMSE.

While for ARIMA model, model selection process will follow the primary three steps, as explained in Section 3, in order to determine the best values for parameter p, d, q, P, D, and Q. Since the time series is generally known to be annual, the value for seasonality parameter, s, can be determined directly, which in this regard it is set with 235 as there are averagely 235 data points in one year period of the time series. During the process of parameter estimation, grid search technique is utilized, which is for selecting the parameters p, q, P, and Q. The sequential values to be tested for p, q, P, and Q are {0, 1, 2, 3, 4}. The best combination is taken from the one that gives the least BIC and Ljung-Box Q statistic value less than its critical value [7].

4.6 Evaluation Method

In this research, evaluation is involved in two stages, i.e. model selection and final evaluation. For model selection, the performance of LSSVM and SVM model is measured in Root Mean Squared Error (RMSE), which represents the mean size of prediction error measured in the same unit as the actual values [5],[17], while the performance of ARIMA model is measured in Bayesian information criterion (BIC), which measures how well a model fit its time series [7].

$$RMSE = \left(\frac{1}{N}\sum_{i=1}^{N}(\tilde{y}_i - y_i)^2\right)^{\frac{1}{2}} \quad (4.1)$$

$$BIC = -2.\ln L + k.\ln(n) \quad (4.2)$$

Meanwhile, final prediction results are evaluated using the indicators of Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). MAE represents the absolute size of the prediction errors, while MAPE calculates the errors as a percentage of the actual values [17].

$$MAE = \left(\frac{1}{N}\sum_{i=1}^{N}|\tilde{y}_i - y_i|\right) \quad (4.3)$$

$$MAPE = \left(\frac{1}{N}\sum_{i=1}^{N}\left|\frac{\tilde{y}_i - y_i}{y_i}\right|\right) \times 100 \quad (4.4)$$

The proposed methodology in this research is summarized in Figure 2.

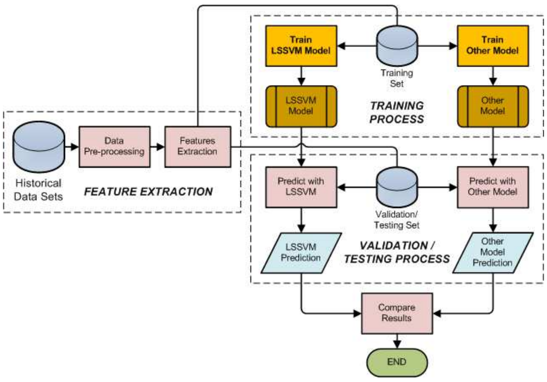

Figure 2: Proposed Methodology

5 EXPERIMENTAL RESULT

The proposed method is brought into experiment by building appropriate predictors based on LSSVM, SVM, and ARIMA approach. The experiment starts with compiling all data from their original sources into one data set, as shown in Table 5.1.

Journal of Theoretical and Applied Information Technology
20th May 2014. Vol. 63 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5.1 Samples of Original Data Set

Date	JCI High Price (IDR)	JCI Low Price (IDR)	JCI Open Price (IDR)	JCI Close Price (IDR)	Gold Fixing (USD)	WTI Crude Oil (USD)
9-Dec-13	4,227.23	4,192.47	4,210.20	4,214.34	1,228.50	97.10
10-Dec-13	4,275.95	4,220.91	4,224.45	4,275.68	1,245.75	98.32
11-Dec-13	4,282.10	4,235.42	4,273.94	4,271.74	1,255.25	97.25

The next step is to extract features from the data set for use by LSSVM and SVM model. The resulting feature set is as shown in Table 5.2. Features extracted include all JCI prices one to five days in advance, gold fixing price, and WTI crude oil price on the respective day, as the input variables, as well as JCI close price on the respective day, as the output variable for the prediction model.

Table 5.2: Samples of Extracted Features

Date	Input							Output
	High (T-5)	Low (T-5)	...	Open (T-1)	Close (T-1)	Gold (T)	Oil (T)	Close (T)
9-Dec-13	4331.59	4265.60	...	4202.26	4180.79	1228.50	97.10	4214.34
10-Dec-13	4320.06	4269.35	...	4210.20	4214.34	1245.75	98.32	4275.68
11-Dec-13	4267.60	4241.30	...	4224.45	4275.68	1255.25	97.25	4271.74

Once features have been extracted, models are ready to be constructed. The process starts with searching the optimal parameter values for LSSVM, SVM, and ARIMA model. Using grid search techniques, the best hyper parameters obtained for the LSSVM model is ($\gamma = 2^{20}$, $\sigma^2 = 2^{19}$) that generates the smallest RMSE for prediction on validation set, that is 36.0492, whereas the best parameter values for the SVM model is ($C = 2^{19}$, $\sigma = 2^{17}$), with the smallest RMSE for validation is 34.2324.

Meanwhile for ARIMA model, from the model identification step, it is found that the first non-seasonal (d=1) and seasonal (D=1) difference has been enough to bring the series to stationarity. Then, from various statistics obtained by fitting possible models to the series, it is obtained that the best fitted model for the time series is ARIMA(0,1,0)x(0,1,1)$_{235}$ that produces the smallest BIC, i.e. 7594.601. The Q value of 26.1875 that is lower than the critical value of 31.4104 indicates the acceptance of null hypothesis that errors for the model is white noise.

By this point, each obtained model has been ready to be used to make final prediction against testing set to measure the final performance. Figure 3 shows the predictions made by LSSVM, SVM, and ARIMA model, respectively, in comparison with the actual price.

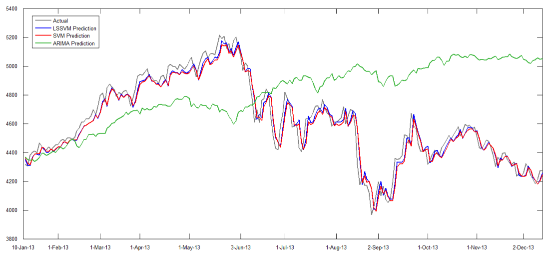

Figure 3: Predicted and Actual JCI Close Price

The predictive performance of LSSVM, SVM, and ARIMA model measured in RMSE, MAE, and MAPE is tabulated in Table 5.3.

Table 5.3: Performance of LSSVM, SVM, and ARIMA Model in Predicting JCI Close Price

Indicator	LSSVM	SVM	ARIMA
RMSE	67.7453	71.5178	452.0385
MAE	51.9116	55.0293	376.7195
MAPE	1.1294%	1.1935%	8.3793%

The experimental results imply that LSSVM model is the best predictor for JCI price compared to SVM and ARIMA model. The prediction indicators shown in Table 5.3 verify the theories in literatures and previous works that LSSVM model is able to make prediction with better accuracy compared to SVM, which is due to the utilization of least square loss function to obtain a set of linear equations in dual space that produces better generalization performance and lower computational cost in LSSVM.

By comparing the conventional statistical ARIMA model with the intelligent LSSVM model (and also SVM model), it is found that intelligent models constantly outperforms the conventional models. There are two reasons for this. First, ARIMA model is a linear model and therefore it fails to capture the nonlinear patterns in stock market, which on the other hand are successfully captured by LSSVM and SVM model. Second, ARIMA model use only one stock price and other correlated factors in the time series are not taken into account, whereas in LSSVM and SVM model, all those factors are considered during the prediction.

6 CONCLUSIONS

Experiments done in this research aim to prove that LSSVM model is able to predict the Jakarta Composite Index price with high accuracy. By comparison, LSSVM model is the best predictor for

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

JCI price as it outperforms SVM and ARIMA model the JCI close price. In this research, performance of each model in predicting the JCI close price is calculated in RMSE, MAE, and MAPE, in which LSSVM model generates the lowest RMSE, MAE, and MAPE.

In summary, as compared to the coequal intelligent model, SVM, LSSVM has better predictive performance as a result of the adaptation of least square loss function, and when compared with the conventional ARIMA model, likewise, LSSVM model appears to give better predictive performance due to the linearity and the use of only one single factor in ARIMA. Therefore, LSSVM approach is highly recommended for predicting Jakarta Stock Index price.

7 FURTHER WORKS

Further works will include experiments on using other approaches to predict JCI price, enhancing LSSVM model with some parameter optimization algorithms, using other kernel methods for LSSVM, and varying window size during feature extraction.

REFERENCES:

[1] Abdullah, L. (2012). ARIMA Model for Gold Bullion Coin Selling Prices Forecasting. *International Journal of Advances in Applied Sciences (IJAAS)*, 1(4), 153-158.

[2] Akinwale, A. T., Arogundade, O. T., & Adekoya, A. F. (2009). Translated Nigeria Stock Market Prices Using Artificial Neural Network for Effective Prediction. *Journal of Theoretical and Applied Information Technology (JATIT)*, 36-43.

[3] Andreoni, A., & Postorino, M. N. (2006). A Multivariate ARIMA Model to Forecast Air Transport Demand. *European Transport Conference*. Strasburgo.

[4] Ayinde, K., & Abdulwahab, H. (2013). Modeling and Forecasting Nigerian Crude Oil Exportation: Seasonal Autoregressive Integrated Moving Average Approach. *International Journal of Science and Research (IJSR)*, 2(12), 245-249.

[5] Challagulla, V., Bastani, F., Yen, I.-L., & Paul, R. (2005). Empirical Assessment of Machine Learning based Software Defect Prediction 54echniques. *10th IEEE International Workshop on Object-Oriented Real-Time Dependable Systems*, (pp. 263-270).

[6] Cheng, C.-H., & Wei, L.-Y. (2010). One Step-ahead ANFIS Time Series Model for Forecasting Electricity Load. *Optimization and Engineering*, 11(2), 303-317.

[7] Clement, E. (2013). Using Normalized BIC to Improve Box-Jenkins Model Building. *Journal of Modern Mathematics and Statistics*, 7(1), 1-7.

[8] Deng, S.-G., & Yeh, T.-H. (2011). Using Least Squares Support Vector Machines to the Product Cost Estimation. *Journal of C.C.I.T.*, 40(1), 1-16.

[9] Gestel, T. V., Suykens, J. A., Baestaens, D.-E., Lambrechts, A., Lanckriet, G., Vandaele, B., et al. (2001). Financial Time Series Prediction Using Least Squares Support Vector Machines Within the Evidence Framework. *IEEE Transactions on Neural Networks*, 12(4), 809-821.

[10] Hsu, C.-w., Chang, C.-c., & Lin, C.-j. (2010). A Practical Guide to Support Vector Classification. CiteSeerX.

[11] Huang, W., Lai, K. K., Nakamori, Y., & Wang, S. (2004). Forecasting Foreign Exchange Rates with Artificial Neural Networks: A Review. *International Journal of Information Technology & Decision Making*, 3(1), 145–165.

[12] Jakaša, T., Andročec, I., & Sprčić, P. (2011). Electricity Price Forecasting – ARIMA Model Approach. *2011 8th International Conference on the European Energy Market (EEM)* (pp. 222-225). Zagreb, Croatia: IEEE.

[13] Kannan, K. S., Sekar, P. S., Sathik, M. M., & Arumugam, P. (2010). Financial Stock Market Forecast using Data Mining Techniques. *Proceedings of International MultiConference of Engineers and Computer Scientists*, I. Hong Kong.

[14] Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, 38(5), 5311–5319.

[15] Kumar, M., & Thenmozhi, M. (2005). Forecasting stock index movement: A comparison of support vector machines and random forest. *Proceedings of ninth Indian institute of capital markets conference*. Mumbai, India: Social Science Electronic Publishing, Inc.

[16] Magaji, A. S., Waziri, V. O., Isah, A., & R., A. K. (2012). A Generic Prediction of Nigeria Stock Exchange Market, Implementation

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Using Naive Bayes Model. *International Journal of Advanced Computer Technology*, 10-14.

[17] Moghaddasi, R., & Badr, B. R. (2008). An Econometric Model for Wheat Price Forecasting in Iran. *International Conference on Applied Economics*, 4, pp. 671-678. New Delhi.

[18] Morgan, G. A., Griego, O. V., & Gloeckner, G. W. (2001). SPSS for Windows: An Introduction to Use and Interpretation in Research. New Jersey: Lawrence Erlbaum Associates, Inc.

[19] Mustaffa, Z., & Yusof, Y. (2011). Optimizing LSSVM Using ABC For Non-Volatile Financial Prediction. *Australian Journal of Basic and Applied Sciences*, 5(11), 549-556.

[20] Mustaffa, Z., & Yusof, Y. (2012). A Hybridization of Enhanced Artificial Bee Colony-Least Squares Support Vector Machines for Price Forecasting. *Journal of Computer Science*, 8(10), 1680-1690.

[21] Olaniyi, S. A., Adewole, K. S., & Jimoh, R. G. (2011). Stock Trend Prediction Using Regression Analysis – A Data Mining Approach. *ARPN Journal of Systems and Software*, 1(4), 154-157.

[22] Ou, P., & Wang, H. (2009). Prediction of Stock Market Index Movement by Ten Data Mining Techniques. *Modern Applied Science*, 3(12), 28-42.

[23] Prasanna, S., & Ezhilmaran, D. (2013). An analysis on Stock Market Prediction using Data Mining Techniques. *International Journal of Computer Science & Engineering Technology (IJCSET)*, 4(2), 49-51.

[24] Suykens, J. A., Gestel, T. V., Brabanter, J. D., Moor, B. D., & Vandewalle, J. (2002). Least Squares Support Vector Machines. K. U. Leuven, Beilgium: World Scientific Publishing Co. Pte. Ltd.

[25] Tan, T. Z., Quek, C., & Ng, G. S. (2007). Biological brain-inspired genetic complementary learning for stock market and bank failure prediction. *Computational Intelligence*, 23(2), 236–261.

[26] Tsitsika, E., Maravelias, C. D., & Haralabous, J. (2007). Modeling and Forecasting Pelagic Fish Production Using Univariate and Multivariate ARIMA Models. *Fisheries Science*, 73, 979-988.

[27] Vapnik, V. N. (2000). The Nature of Statistical Learning Theory (2nd ed.). New York, USA: Springer-Verlag New York, Inc.

[28] Yu, L., Chen, H., Wang, S., & Lai, K. K. (2009). Evolving Least Squares Support Vector Machines for Stock Market Trend Mining. Evolutionary Computation, *IEEE Transactions on*, 13(1), 87 - 102.