Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Predicting Jakarta composite index using hybrid of fuzzy time series and support vector regression models

Predicting Jakarta composite index using hybrid of fuzzy time series and support vector regression models

Rian Febrian Umbara, Dede Tarwidi, Erwin Budi Setiawan

Telkom University, Bandung, Indonesia

rianum@telkomuniversity.ac.id, erwinbudisetiawan@telkomuniversity.ac.id, dedetarwidi@telkomuniversity.ac.id

Abstract. The paper discusses the prediction of Jakarta Composite Index (JCI) in Indonesia Stock Exchange. The study is based on JCI historical data for 1286 days to predict the value of JCI one day ahead. This paper proposes predictions done in two stages., The first stage using Fuzzy Time Series (FTS) to predict values of ten technical indicators, and the second stage using Support Vector Regression (SVR) to predict the value of JCI one day ahead, resulting in a hybrid prediction model FTS-SVR. The performance of this combined prediction model is compared with the performance of the single stage prediction model using SVR only. Ten technical indicators are used as input for each model.

1. Introduction

Stock price index is an indicator or a reflection of stock price movements. Index is one of the guidelines for investors to invest capital in the capital market, especially stocks. Stock indices derived from stock prices with high market capitalization can provide an overall picture of the economy and depend on several factors [1].

Predicting stock price index and movement has been considered one of the most challenging applications of time series prediction. While there are many empirical studies dealing with issues in predicting stock price indices, most of the research is done in countries with advanced financial markets. An accurate prediction of stock price index movements is essential for developing an effective market strategy [2]. Thus, investors can minimize the market risk and the opportunity to make a profit with stock index trading [3]. Predicting stock and derivatives prices have been done by many scholars around the world. Patel, Shah, Thakkar, & Kotecha predict the stock price index CNX Nifty and S & P Bombay Stock Exchange (BSE) using SVR-ANN, SVR-RF, SVR-SVR algorithms [1]. While C.P. Lee, W.C. Lin, and C.C Yang [4] used fuzzy time series combined with least square support vector regression and bootstrap method to forecast option prices.

In this study, prediction of Jakarta Composite Index (JCI) is done in two stages. Predictions done in two stages are expected to result in smaller errors than the one-stage prediction. The first stage of the prediction is to calculate ten technical indicators, which are commonly used by analysts to analyze stock price movements, based on JCI data at time t. Next we will use the Fuzzy Time Series method to predict the value of the ten indicators at time t + 1. The predicted value of the ten indicators will then be used to predict the JCI closing value at time t + 1 using the Support Vector Regression method. Support Vector Regression is the most widely used algorithm to predict stock prices and stock price indices proposed by Vladimir N Vapnick in 1996.

2. Construction of Prediction Model

The prediction model to be constructed is a hybrid of Fuzzy Time Series model and Support Vector Regression (FTS-SVR). The prediction process involved 10 technical indicators. Here's an explanation of Fuzzy Time Series, Support Vector Regression, and 10 Technical Indicators..

2.1. Fuzzy Time Series

Fuzzy time series is one of soft computing methods that has been used and applied in time series data analysis. This method was proposed by Song and Chissom. The main purpose of fuzzy time series is to predict time series data that can be widely used in any real time data, including capital market data.

In the last few decades, research on time series has progressed in handling predictions. But in real life, people tend to face many random fuzzy sequences that contain noise or disturbance. Predictions based on traditional time series data do not seem to be able to deal with this problem. But fuzzy mathematics has an advantage in solving this problem. Therefore, Song and Chissom introduced the concept of fuzzy mathematics into time series data and proposed a fuzzy time series concept.

Let $\boldsymbol{U}$ universe of discourse, where $U = \{u_1, u_2, \ldots, u_n\}$. A fuzzy set defined in $\boldsymbol{U}$ can be represented as follow: $A = f_A(u_1)/\, u_1 + f_A(u_2)\,/\, u_2 + \ldots + f_A(u_n)/u_n$, where f_A is a membership function of the fuzzy set A, $f_a: U \rightarrow [0,1]$, and $f_A(u_i)$ represents the membership degree of u_i belongs to fuzzy set A, $f_A(u_i) \in [0,1], and\ 1 \leq\ i\ \leq n$.

Here are some steps to construct Fuzzy Time Series model [5]:

1. Define the scope of the universe on the basis of data over a predetermined period of time. Let D_max and D_min be the maximum and minimum values of the data value, then the interval of the universal set is U $= [D_{min}, D_{max}]$.
2. Divide U into several intervals $\{u_1, u_2, \ldots, u_n\}$ with the same length.
3. Define fuzzy sets A_t based on the interval formed and the time series data is fuzzified. The fuzzy set A_t defines the linguistic variables of the time series data..
4. Identify fuzzy relationships using fuzzyfication data. If the time series variable $F(t\text{-}1)$ is fuzzyfied to A_k and $F(t)$ becomes A_m, then A_k has a relationship with A_m.Denote the relation with $A_k \rightarrow A_m$, where A_k is the *current state* of the original value dan A_m is the next *current state*. The entire fuzzy relation will be saved and if there is a repetition of the same fuzzy relation more than once, the same fuzzy relation is stored only once.
5. Insert fuzzy relation into fuzzy relation group. If the fuzzy set relates to more than one set, then another relation is placed on the right of the other set. For example, A_1 relates to A_1, dan A_2, then the resulting fuzzy relation group is $A_1 \rightarrow A_1, A_2$. The results are entered and stored into the ith fuzzy relation group.
6. Defuzzyfy the prediction results with rules:
   a. If there is a one-to-one relationship within the group from A_j, such as $A_j \rightarrow A_k$, and the highest membership degree of A_k is in the interval u_k, then the prediction result from F(t) is the middle value of the interval u_k.
   b. If there is an empty relation, such as A_j than has no relation, or $A_j \rightarrow \emptyset$, and the highest membership degree of lies between the interval ranges u_j, then the prediction result obtained is the middle value of u_j.
   c. If there is a one-to-many relation within the relation group of A_j, such as $A_j \rightarrow A_1, A_2, \ldots, A_n$, and the highest membership degrees lies in the intervals $u_1, u_2, \ldots, u_n$, then the prediction result obtained is the average of the middle values $m_1, m_2, \ldots, m_n$ of the intervals $u_1, u_2, \ldots, u_n$, that is $\frac{m_1 + m_2 + \ldots + m_n}{n}$.

2.2. *Support Vector Regression*

Support Vector Regression (SVR) is a method first introduced by Vladimir N. Vapnik, Harris Drucker, Christopher JC Burges, Linda Kaufman and Alexander J. Smola in 1996. The model introduces Support Vector Classification which relies on a subset of training data. SVR is an application of Upport Vector Machine (SVM) used for regression cases, whose output is a real or continuous number. SVR is a method that can overcome the overfitting, so it will produce a good performance with small error [6]. Overfitting is a condition in which a model does not describe the main relationship between input and output variables but rather describes random error or noise. This condition will result in poor predictions. SVR builds the linear regression function equation as follows

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

Vapnik creates a formula to minimize the function of the linear Ɛ-insensitivity loss function, that is

$$|y - f(x,w)|_\varepsilon = \begin{cases} 0, & jika\ |y - f(x,w)| \leq \varepsilon \\ |y - f(x_i,w)| - \varepsilon, & lainnya \end{cases} \tag{2}$$

Based on the above equation, the linear regression of $f(x,w)$ is estimated by minimizing $||w||^2$ and the sum of linear ε-insensitivity losses. Where the factor $||w||^2$ is called regularization. Minimizing $||w||^2$ will make the function as small as possible, so it can control the function of capacity.

$$R = \frac{1}{2}||w||^2 + c(\textstyle\sum_{i=1}^{m} |y - f(x,w)|_\varepsilon \tag{3}$$

$$R = \frac{1}{2}||w||^2 + c\textstyle\sum_{i=1}^{m}(\xi + \xi^*) \tag{4}$$

$$(w^T x_i + b) - y_i \leq \varepsilon + \xi_i \tag{5}$$

$$y_i - (w^T x_i + b) - y_i \leq \varepsilon + \xi_i \tag{6}$$

$$\xi_i, \xi_i^* \geq 0, i = 1,2,\ldots,m \tag{7}$$

where ξ_i dan ξ_i^* are slack variables, one to exceed the target value by more than ε and other tfor being more than ε below the target. The value of c serves to minimize risk. It controls a trade-off between an approximation error and the norm of weight vector ||w||.

For optimization problems it uses langrange theory and Karush-Kuhn-Tucker condition to get the desired regression function weights. SVR can predict the stock price index value at the time of closing by using the kernel assistance and 10 technical indicators as input. The kernel function used is

RadialBasisFunction : $K(x_i, x_j) = exp(-\gamma||x_i \cdot x_j||^2)$ (8)

where γ is constant of radial basis function

2.3. Technical Indicators

Ten technical indicators are some indicators or basis for calculating stock index movement [7]. Here are the indicators

Table 1. Ten technical indicators

No	Indicators	Formula
1.	Simple n-day moving average (SMA)	$\frac{C_t + C_{t-1} + \cdots + C_{t-(n-1)}}{n}$
2.	Weighted moving average (WMA)	$\frac{(n)C_t + (n-1)C_{t-1} + \cdots + (1)C_{t-(n-1)}}{n + (n-1) + \cdots + 1}$
3.	Stochastic K% (STK)	$\frac{C_t - LL_{t-(n-1)}}{HH_{t-(n-1)} - LL_{t-(n-1)}} x100$
4.	Stochastik D% (STD)	$\frac{\sum_{i=0}^{n-1} K_{t-i}}{10} \%$

5.	Relative Strength Index (RSI)	$100 - \frac{100}{1 + (\sum_{i=0}^{n-1} \frac{UP_{t-i}}{n})/(\sum_{i=0}^{n-1} \frac{DW_{t-i}}{n})}$
6.	Moving average convergence divergence (MACD)	$MACD(n)_{t-1} + \frac{2}{n+1} \times (DIFF_t - MACD(n)$
7.	Commodity channel index (CCI)	$\frac{M_t - SM_t}{0.015D_t}$
8.	Larry William's R% (LWR)	$\frac{H_t - C_t}{H_t - L_t} \times 100$
9.	A/D oscillator (AD)	$\frac{H_t - C_{t-1}}{H_t - L_t}$
10.	Momentum (MOM)	$\mathrm{C_t} - \mathrm{C_{t(n-1)}}$

where :
1. C_t = close price of the stock index at time t
2. L_t = lowest price of the stock index at time t
3. H_t = highest price of the stock index at time t
4. LL_t = lowest price of the stock index in time inteval t
5. HH_t = highest price of the stock index in time interval t
6. $DIFF_t$ = $EMA(12)_t - EMA(26)_t$
7. $EMA(k)_t$ = $EMA(k)_{t-1} + \alpha \times (C_t - EMA(k)_{t-1})$
8. α = $\frac{2}{k+1}$
9. k = k-days period for EMA
10. M_t = $\frac{H_t+L_t+C_t}{3}$
11. SM_t = $\frac{(\sum_{i=1}^{n} M_{t-i+1})}{3}$
12. DM_t = $\frac{(\sum_{i=1}^{n} |M_{t-i+1} - SM_t|)}{n}$
13. UP_t = the increase of the stock price index at time t
14. DW_t = the decrease of the stock price index at time t
15. n = number of stocks

2.4. The Proposed Prediction Model
The proposed prediction model is a hybrid model of FTS-SVR, which performs predictions in two stages. The first stage is to predict the value of 10 technical indicators at time t + 1 based on data of 10 technical indicators at time t using Fuzzy Time series. The second stage is to predict the closing value of JCI at time t + 1 based on the prediction of 10 technical indicators generated in the first stage by using Support Vector Regression. The performance of the proposed model will then be compared to its performance with a one-stage prediction model, which predicts the value of JCI at time t + 1 based on 10 technical indicators data at time t by using Support Vector Regression.

3. Prediction Results
Before the prediction model of FTS-SVR was established, JCI closing value data was collected from January 12, 2012 until May 19, 2017. The closing value data was then used to calculate the value of 10

technical indicators per day using the formulas listed in Table 1. Thus we obtained daily value of 10 technical indicators starting from March 1, 2012 to May 19, 2017.
In the first stage, each technical indicator is then divided into 105 intervals to form fuzzy time series models. The formation of Fuzzy Time Series models for each technical indicator is performed using the historical data of each indicator from March 1, 2012 until April 21, 2016. So we get 10 models of Fuzzy Time Series for each indicator. The models obtained are then used to predict the value of 10 technical indicators one day ahead (t + 1) starting from April 22, 2016 to May 19, 2017. The performance of the models is measured by calculating prediction errors, using Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE), with the following formulas

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|A_t-F_t|}{|A_t|}\times 100 \tag{9}$$

$$RMSE = \left(\sum_{i=1}^{n}(A_t-F_t)^2/n\right)^{1/2} \tag{10}$$

where:

A_t= actual value of the stock price index at time t
F_t= predicted value of the stock price index at time t
n = number of data
t = time index

Here are the RMSE and MAPE values from the predictions of each technical indicator using Fuzzy Time Series.

Table 2. Error prediction of each technical indicator at time t + 1 with Fuzzy Time Series

No	Prediction of Indicators	MAPE	RMSE
1.	Prediction of SMA	0.8804	88.456
2.	Prediction of WMA	0.7701	74.86
3.	Prediction of STK	86.34	19.18
4.	Prediction of STD	7.5	4.77
5.	Prediction of RSI	311.191	4173.403
6	Prediction of MACD	67.95	5.54
7.	Prediction of CCI	223.76	66.25
8.	Prediction of LWR	157.763	29.49
9	Prediction of AD	384.736	0.53
10	Prediction of MOM	298.1	61.07

In the second stage, the prediction result of 10 technical indicators at t + 1 then becomes the input for SVR process to predict the JCI closing value at t + 1. The predicted indicators values are then divided into two groups, training data from April 22, 2016 until February 13, 2017 (200 business days), and testing data from February 14, 2017 to February 13, 2017 (62 business days). The kernel used in SVR is a radial basis function with parameter values C = 91556.41, ε = 6.0026, and $\gamma = 740.1296$. Prediction errors for training and testing are as follows.

Table 3. Prediction errors of the JCI closing value

No	Process	MAPE	RMSE
1.	Training	0.459	37.047
2.	Testing	2.282	165.393

The following figures show the comparison between the actual JCI closing values and the predicted JCI closing values, for both training and testing data.

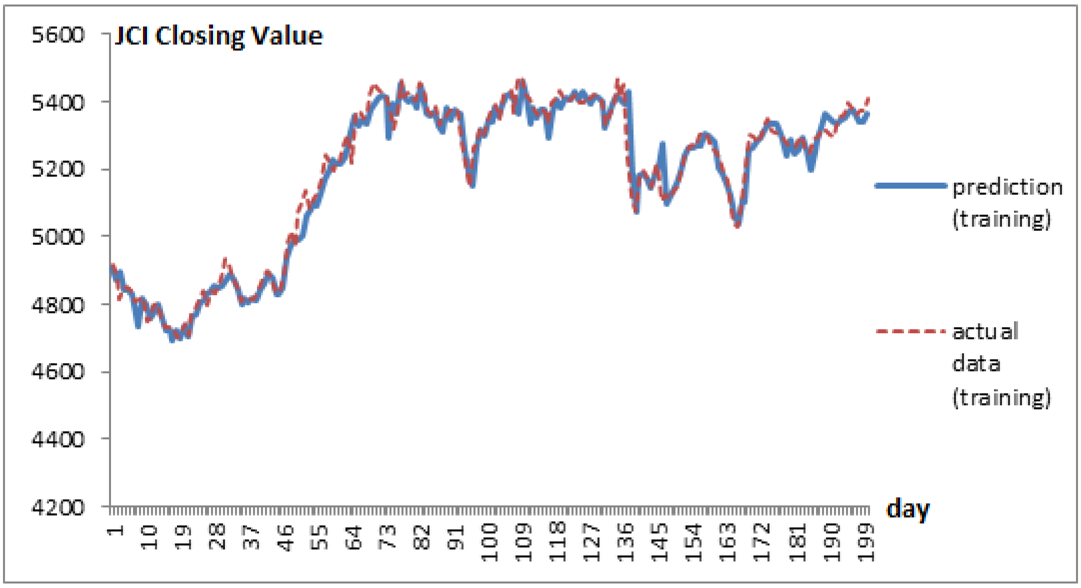

Figure 1. Comparison of FTS-SVR model prediction results with actual data for training data prediction

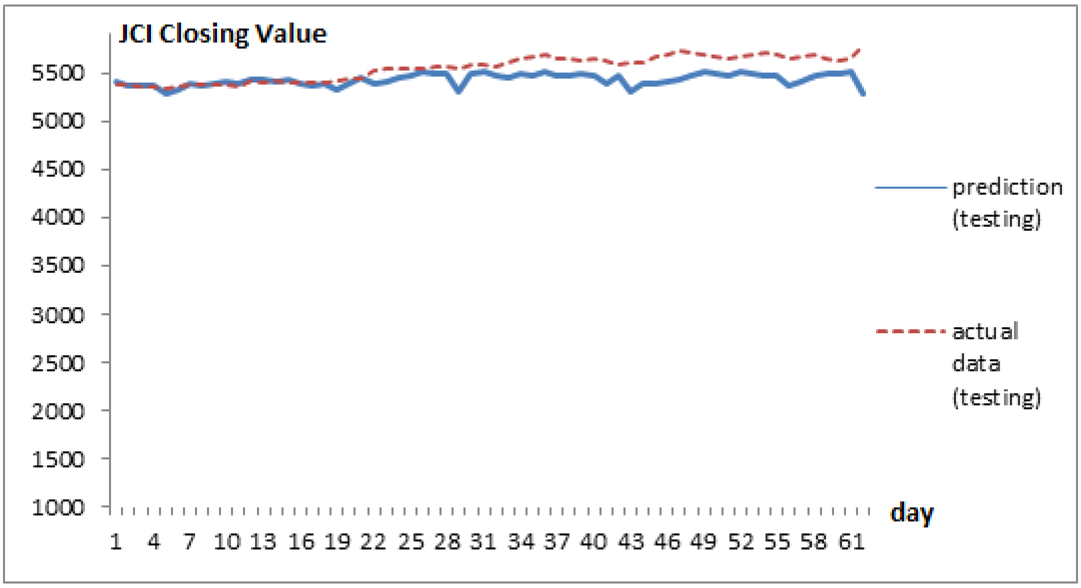

Figure 2. Comparison of FTS-SVR model prediction results with actual data for testing data prediction

JCI closing price prediction results from hybrid model of FTS-SVR are then compared with prediction results from one-stage prediction model using SVR. As the inputs are the actual 10 technical indicators at time t, and the output is the prediction of JCI closing value at t + 1. Training data are the actual values of the ten indicators from April 22, 2016 to February 13, 2017 (200 working days) and testing data are the actual values of the ten indicators from February 14, 2017 to February 13, 2017 (62 working days). Figures 3 and 4 show a comparison of the results of the close price predictions with their actual values for training and testing.

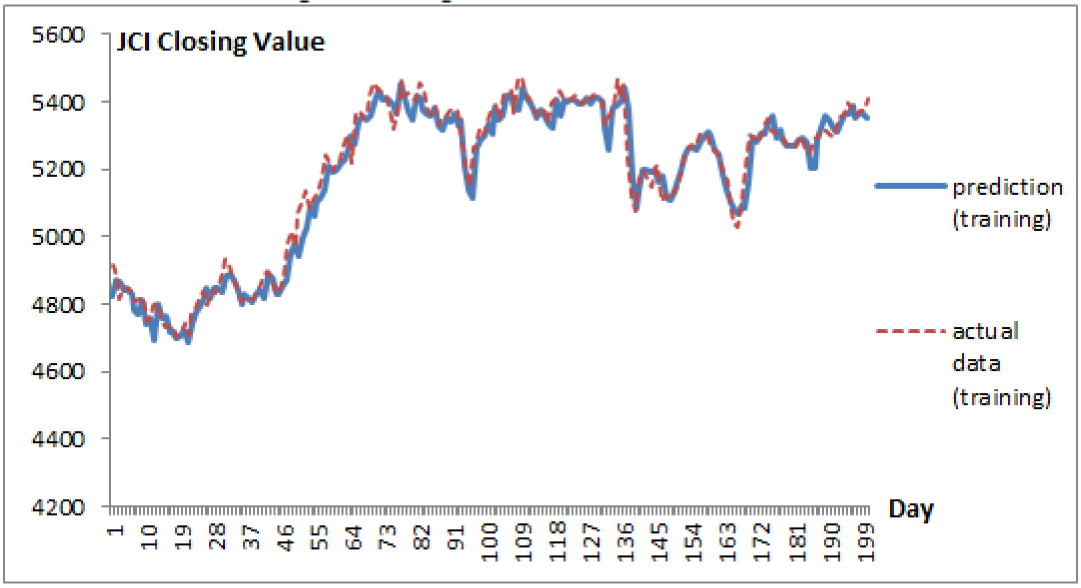

Figure 3. Comparison of SVR model prediction results with actual data for training data prediction

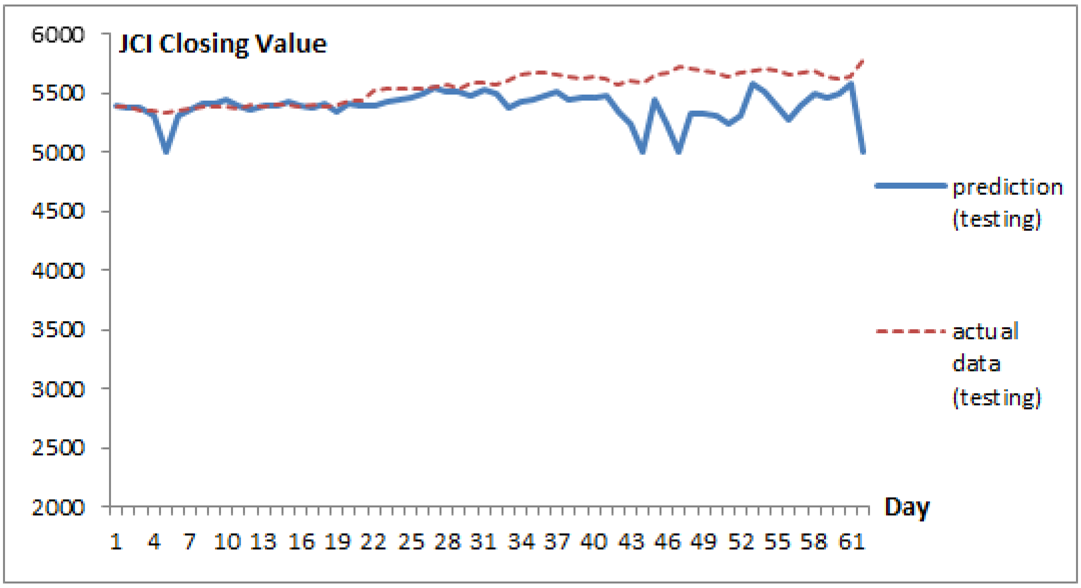

Figure 4. Comparison of SVR model prediction results with actual data for testing data prediction

Table 4 shows the comparison of prediction errors from the FTS-SVR model and the SVR model.

Table 4. Comparison of MAPE and RMSE between predicted values resulted from FTS-SVR model and SVR model

Prediction Model	MAPE for training data prediction	MAPE for testing data prediction	RMSE for training data prediction	RMSE for testing data prediction
FTS-SVR	0.459561173	2.281763222	37.0468724	165.3924739
SVR	0.524795171	3.003133976	41.6945237	242.4911956

It can be seen that the error value (MAPE and RMSE) of predicted JCI closing values resulted from FTS-SVR model is smaller than those of SVR model.

4. Conclusions

The results have shown that the hybrid FTS-SVR model gives a better prediction of JCI closing value than the SVR model. This can be seen from the MAPE and RMSE values of the FTS-SVR models which are smaller than those of the SVR model, both for training data and testing data. For future work, fuzzy time series method can be combined with optimization algorithm, such as genetic algorithm, to get optimal intervals for prediction.

References

[1] Patel, Jigar., Shah, Sahil., Thakkar, Priyank., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. Expert Systems with Applications 42 (2015) 2162-2172.
[2] Leung, M. T., Daouk, H., & Chen, A. S. (2000). Forecasting stock indices: A comparison of classification and level estimation models. International Journal of Forecasting, 16, 173–190.
[3] Manish, K., & Thenmozhi, M. (2005). Forecasting stock index movement: A comparison of support vector machines and random forest. In Proceedings of ninth Indian institute of capital markets conference, Mumbai, India.
[4] Lee, C. P., Lin, W. C., & Yang, C. C. (2014). A strategy for forecasting option prices using fuzzy time series and least square support vector regression with a bootstrap model. Scientia Iranica. Transaction D, Computer Science & Engineering, Electrical, 21(3), 815.
[5] Jens Runi Poulsen. (2009). Fuzzy Time Series Forecasting. Aalborg University Esbjerg.
[6] Parrella. F. (2007). Online support vector regression (Master's thesis). Italy: Department of Information Science, University of Genoa.
[7] Pring, Martin.(2001), Introduction to Technical Analysis. McGraw-Hill. Singapore.