ICIC Express Letters
Volume **12**, Number **12**, December **2018**

ICIC International ©2018 ISSN 1881-803X
pp. 1277–1283

ALGORITHM AND MODELING OF STOCK PRICES FORECASTING BASED ON LONG SHORT-TERM MEMORY (LSTM)

Adler Haymans Manurung[1], Widodo Budiharto[2] and Harjanto Prabowo[3]

[1]Management Department, BINUS Business School – Doctor of Research in Management
[2]Computer Science Department
School of Computer Science
[3]Computer Science Department, BINUS Graduate Program – Doctor of Computer Science
Bina Nusantara University
Jl. K. H. Syahdan No. 9, DKI Jakarta 11480, Indonesia
Adler.manurung@binus.ac.id; { wbudiharto; harprabowo }@binus.edu

Received May 2018; accepted August 2018

Abstract. *Time series analysis has significance in financial analytics and market forecasting and it can be utilized in any field. For stockbrokers, understanding trends and forecasting supported by software are very important to decision making and reacting to changes in behavioral patterns. This paper proposes an algorithm and model for market forecasting in Indonesian exchange based on the Long Short-Term Memory (LSTM) and compared with ARIMA model. We use data from Bank Central Asia (BCA) from 2013-2018 obtained from Yahoo finance. In our experiments, we predict and simulate the important prices called Open, High, Low and Closing (OHLC) with various parameters. Based on the experiment, the best accurate prediction in LSTM comes from the short term (1 year) with high epoch in training phase rather than using 3 years or 5 years of training data, and our model has better result compared with popular model such as ARIMA. These results should be very useful to be used in stock exchange office.*
Keywords: LSTM, Forecasting, Stock market, Finance, Deep learning, ARIMA

1. **Introduction.** Forecast is a prediction of some future event or events. Analyzing time-oriented data and forecasting future values of a time series are among the most important problems that analysts face in many fields, ranging from finance and economics to managing production operation, to the analysis of political and social policy sessions. Successful time series analysis and forecasting need computer software and good algorithm and model. The forecasting is so important because prediction of future events is a critical input into many types of planning and decision-making processes such as finance, industrial process control risk management [1]. In finance, time series analysis is used for financial forecasting such as stock prices, assets, and commodities. Stock is the most volatile investment with high risk, but with high return to investors if carefully managed in their portfolio. In managing stocks, information on their prices is of utmost importance. Capital markets are markets for buying and selling equity and debt instruments; it also has activities related to public offering and trade of stock and issuance stock of public company.

The government of Indonesia understands that the capital market plays a strategic role in the national development as one of the sources of funding for business and also as investment instruments for Indonesians. Stock exchanges are considered major players in financial sectors of many countries including Indonesia. Stockbrokers, who execute stock trade, use technical, fundamental or time series analysis in trying to predict stock prices, so as to advise clients [2]. The benchmark stock index of Indonesia was affected by

DOI: 10.24507/icicel.12.12.1277

negative market sentiments brought on by domestic factors. There are a lot of complicated financial indicators and the fluctuation of the stock market is highly violent [3].

Time series analysis covers many forecasting methods. Researchers have developed numerous modifications to the basic ARIMA model and found considerable success in these methods. The modifications include clustering time series from ARMA models with clipped data and fuzzy neural network approach. Comparative study conducted by Saad et al. compared methods to predict stock prices using Time Delay, Recurrent and Probabilistic Neural Networks (TDNN, RNN, and PNN, respectively), utilizing conjugate gradient and multistream extended Kalman Filter training for TDNN and RNN [5], but unfortunately the paper does not discuss the result of implementation for stock prices prediction of banking and companies. Bao et al. proposed novel deep learning framework where Wavelet Transforms (WT), Stacked Autoencoders (SAEs) and LSTM are combined for stock price forecasting. High-level denoising features are fed into LSTM to forecast the next day's closing price, but unfortunately the system was very complex for the implementation in real case [6].

Today, Artificial Intelligence (AI) is a thriving field with many practical applications and active research topics. Recurrent Neural Networks (RNN) and LSTM [4,12] as part of AI have proved one of the most powerful models for processing sequential data and as state of the art in this topic. Short-term trends, particularly attractive for neural network analysis, can be used profitably in scenarios such as option trading, but only with significant risk. This paper proposes an efficient and simple algorithm and model of stock forecasting based on the LSTM with improvement and innovation in selecting only short-term data for training phase and it is able to give future prediction value and of course it should be very useful for stock prices prediction in Indonesia. The organization of this paper consists of introduction, related works, proposed works and experimental results.

2. **Related Works.**

2.1. **Long short-term memory.**

Sequence prediction problems have been around for a long time especially in financial markets. LSTMs built from the Recurrent Neural Network (RNN). In the figure shown below, a chunk of neural network **A**, looks at some input x_i and outputs a value h_i. A loop allows information to be passed from one step of the network to the next as shown in Figure 1 below.

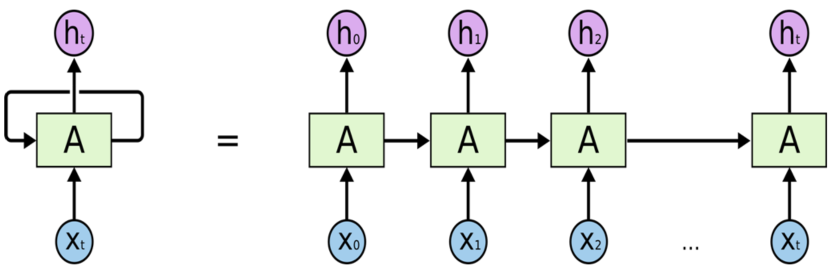


FIGURE 1. The basic model of Recurrent Neural Networks (RNN)

A typical LSTM network is comprised of different memory blocks called cells. There are two states that are being transferred to the next cell: the cell state and the hidden state. The memory blocks are responsible for remembering things and manipulation to this memory is done through three major mechanisms, called gates. LSTMs are particularly well suited to time-series prediction because they can "learn" and "remember" in long-term memory things like market regimes [7].

2.2. **ARIMA model.** ARIMA stands for Autoregressive Integrated Moving Average. ARIMA is also known as Box-Jenkins approach. Auto Regressive (AR) terms refer to the lags of the differenced series, Moving Average (MA) terms refer to the lags of errors and I is the number of difference used to make the time series stationary. In an ARIMA model, the future value of a variable is supposed to be a linear combination of past values and past errors. Assumptions of ARIMA model require that data should be stationary – by stationary it means that the properties of the series do not depend on the time when it is captured. A white noise series and series with cyclic behavior can also be considered as stationary series. We test for stationary using the Augmented Dickey-Fuller unit root test. The stationary of a time series is related to its statistical properties in time [1,9]. An ARIMA(p, d, q) model is for some time series data y_t, where p is the number of autoregressive lags, d is the degree of differencing and q is the number of moving average lags.

3. **Proposed Method.** Considering the complexity of financial time series, combining deep learning with concept of financial market prediction is regarded as one of the most charming topics [13]. Based on that idea, we propose the algorithm for predicting of future values and the RNN model that has LSTM. We use values from the very beginning in the first sliding window to predict the price p in the following window W_{t+1}:

$$W_{t+1} = \left(P_{(t+1)w}, P_{(t+1)w+1}, \ldots, P_{(t+2)w-1}\right) \tag{1}$$

Figure 2 shows our model for stock prices prediction, we try to learn approximation function and our first task is to feed the data into LSTM. Our data is stock price data time series.

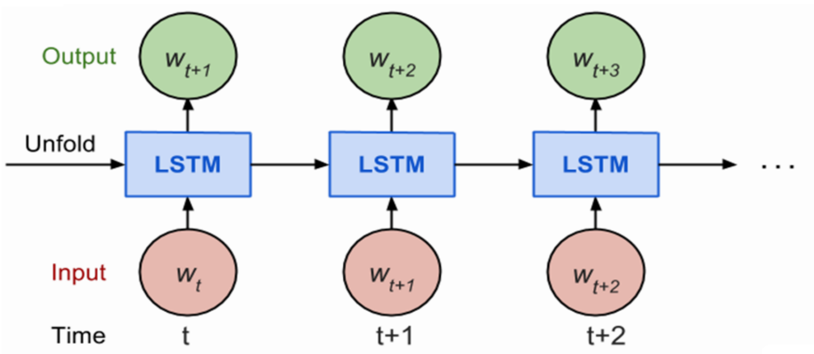


FIGURE 2. The RNN model for stock prices prediction

The efficient algorithm based on Tensorflow and LSTM for prediction of stock prices is shown in algorithm 1.

Algorithm 1. *Predicting stock prices using LSTM*
declare epoch and variables
load data BBCA.JK from yahoo finance
begin
  get and format important columns for processing
  prepare for time series dataset
  #80% training set and 20% testing set
  train-test split
  modeling LSTM
  calculate prediction 1 year, 3 years and 5 years and RMSE
  plotting predictions
  displaying prediction of future values
end

4. **Experimental Results.** We developed LSTM program using Python and Tensorflow [8]. We compare OHLC data with high prices from BCA (BBCA.JK) as shown in Figure 3.

Based on Figure 3, we found that there is no significant error (gap) between original data OHLC average and high price. We use 80% for training data and 20% for testing data, and the result is shown in Figures 4 and 5.

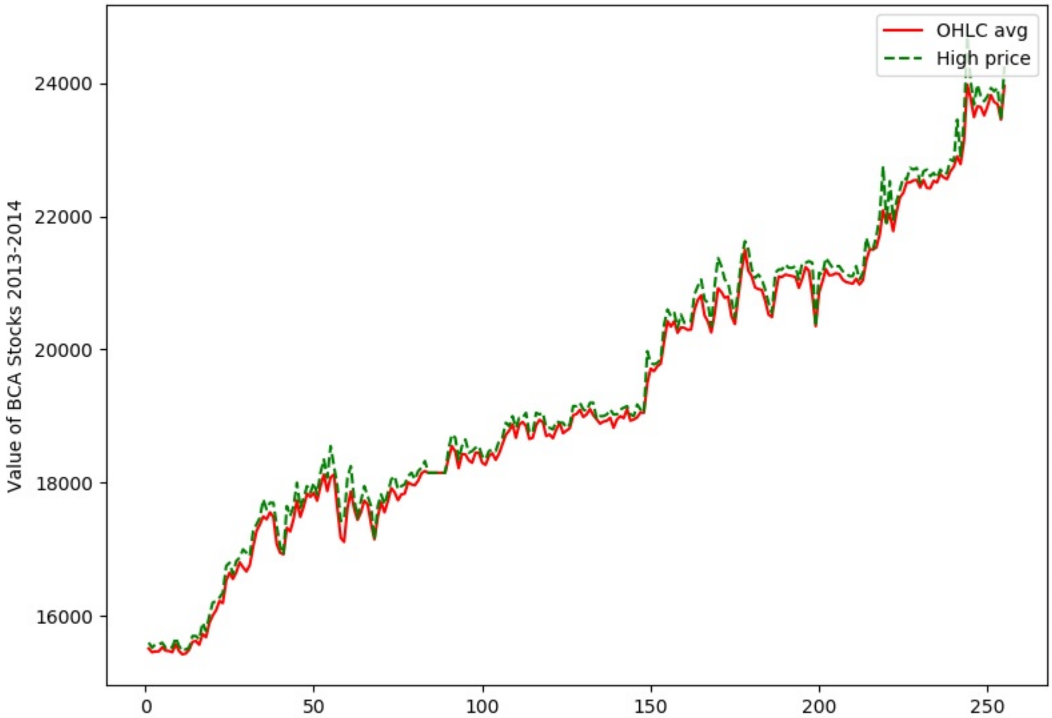


Figure 3. Dataset of OHLC average from stock prices of Bank Central Asia (symbol: BBCA.JK) consists of the data from 20 Feb 2013 until 20 Feb 2014 compared with high price (source: Yahoo finance).

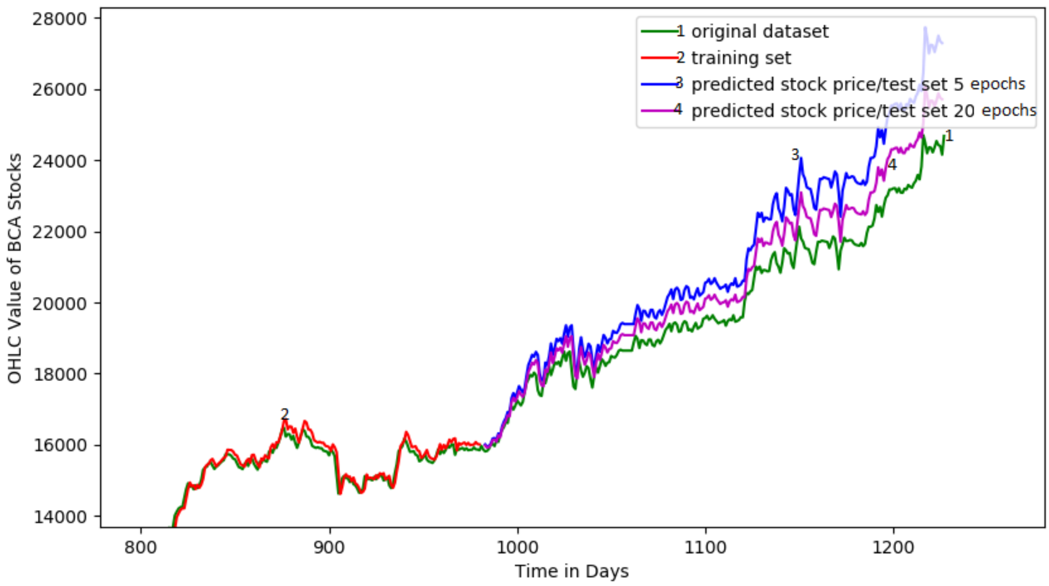
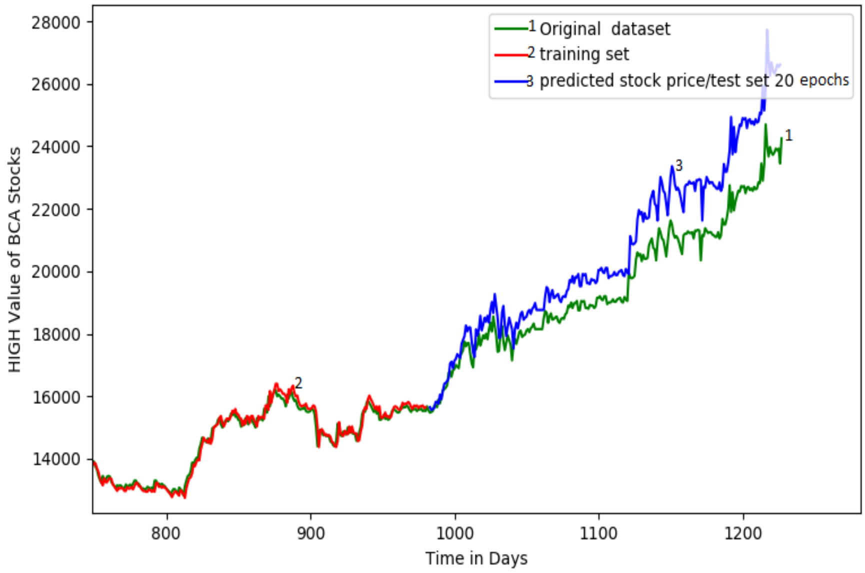
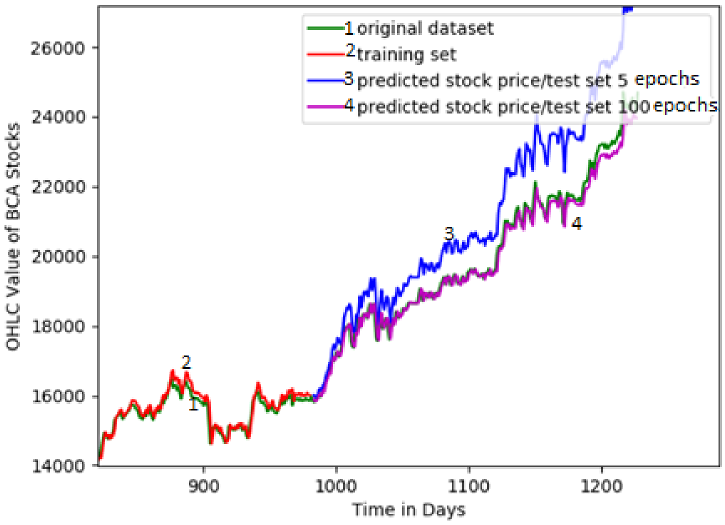


Figure 4. Detailed result of prediction based on the 80% training set and 20% testing set. It shows the significant accuracy from 5 epochs to 20 epochs.



(a)



(b)

Figure 5. Result of prediction of high price using 20 epochs (a) and OHLC price using 5 and 100 epochs (b). Prediction using 100 epochs is better than 5 epochs.

We compare result of the experiment by varying epoch and historical data between 1-5 years as shown in Table 1. It shows the best prediction using 1 year data with the best accuracy 94.59% at 100 epoch. Epoch is one of the best methods to compare various data for forecasting. For analyzing the efficiency of the system we used the Root Mean Square Error (RMSE).

We simulated the ARIMA model [10] to get return equation and accuracy as shown in Figure 6, we call arima function and we got coefficient values and RMSE.

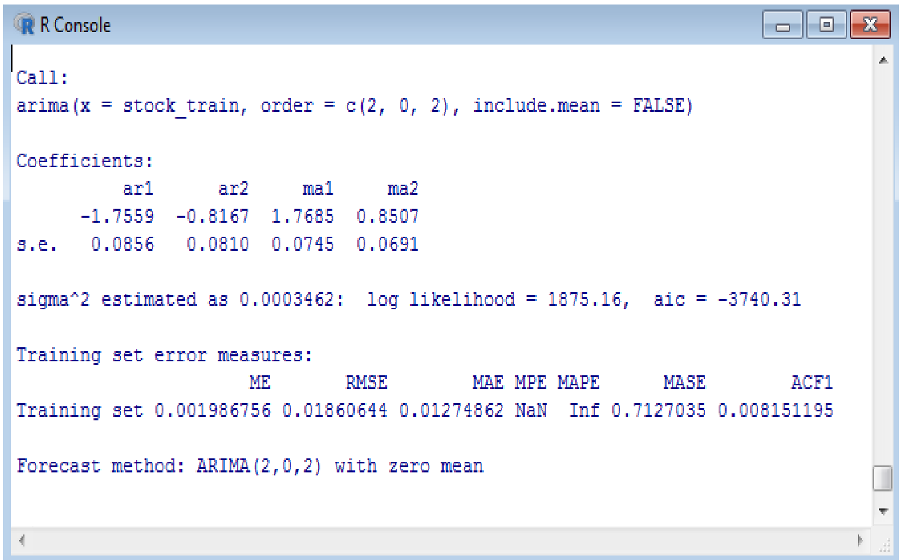
From the coefficients obtained, the return equation can be written as:

$$Y_t = -1.7559 * Y_{(t-1)} - 0.8167 * Y_{(t-2)} + 1.7685\varepsilon_{(t-1)} + 0.8507\varepsilon_{(t-2)}$$

We then examine the ARMA model using parameters in R language:
# Summary of the ARIMA model using the determined (p, d, q) parameters
fit = arima(x = stock_train, order = c(2, 0, 2), include.mean = FALSE)
summary(fit)

Table 1. Result of experiment with various historical data and epochs

<table>
<tr><th rowspan="3">Historical data</th><th colspan="3">OHLC Value</th><th colspan="3">High Value</th></tr>
<tr><th colspan="6">epochs</th></tr>
<tr><th>5</th><th>20</th><th>50</th><th>20</th><th>50</th><th>100</th></tr>
<tr><td></td><td colspan="6">Test RMSE value</td></tr>
<tr><td rowspan="3">1 year</td><td>537.07</td><td>335.33</td><td>257.42</td><td>565.18</td><td>394.76</td><td>205.65</td></tr>
<tr><td colspan="2">Prediction value (next day)</td><td>25545.20</td><td>25341.33</td><td>25343.35</td><td>25344.05</td></tr>
<tr><td colspan="2">Accuracy</td><td>93.83%</td><td>94.57%</td><td>94.58%</td><td>94.59%</td></tr>
<tr><td rowspan="3">3 years</td><td>1023.43</td><td>1022.43</td><td>1023.45</td><td>929.80</td><td>193.30</td><td>193.31</td></tr>
<tr><td colspan="2">Prediction value (next day)</td><td>30619.65</td><td>28690.60</td><td>27094.52</td><td>27092.09</td></tr>
<tr><td colspan="2">Accuracy</td><td>78.28%</td><td>83.54%</td><td>88.46%</td><td>88.47%</td></tr>
<tr><td rowspan="3">5 years</td><td>1403.34</td><td>423.55</td><td>324.23</td><td>1294.42</td><td>228.16</td><td>223.82</td></tr>
<tr><td colspan="2">Prediction value (next day)</td><td>32492.25</td><td>29248.09</td><td>27213.09</td><td>29964.86</td></tr>
<tr><td colspan="2">Accuracy</td><td>73.77%</td><td>81.95%</td><td>88.07%</td><td>79.99%</td></tr>
</table>



```
R Console
Call:
arima(x = stock_train, order = c(2, 0, 2), include.mean = FALSE)

Coefficients:
          ar1      ar2     ma1     ma2
      -1.7559  -0.8167  1.7685  0.8507
s.e.   0.0856   0.0810  0.0745  0.0691

sigma^2 estimated as 0.0003462:  log likelihood = 1875.16,  aic = -3740.31

Training set error measures:
                      ME       RMSE        MAE MPE MAPE      MASE        ACF1
Training set 0.001986756 0.01860644 0.01274862 NaN  Inf 0.7127035 0.008151195

Forecast method: ARIMA(2,0,2) with zero mean
```

Figure 6. The return equation using R for ARIMA model

ARIMA model is only able to get 56% in accuracy and the actual (black line) vs forecasted (red line) are shown in Figure 7.

5. **Conclusions.** This paper develops a model and program for stock prices prediction using data from Yahoo finance. Good prediction systems for stock prices help traders, investors, and analysts by providing supportive information like the future direction of the stock market. Our experiment successfully predicts the stock prices using LSTM compared with ARIMA model. We found that for LSTM, we should use short term historical data for the best accuracy. Historical data using 1 year is the best compared with 3 years and 5 years data. Deep learning technology is expanding the options available to data scientists to solve interesting problems with high accuracy. LSTM is also superior in short term data until 94% compared with ARIMA model that only has 56% in accuracy. Studies suggest that additional factors should be taken into account on top of the basic or unmodified model such as the online news information related to the stock and haircut estimation value, and it will be proposed for future work.

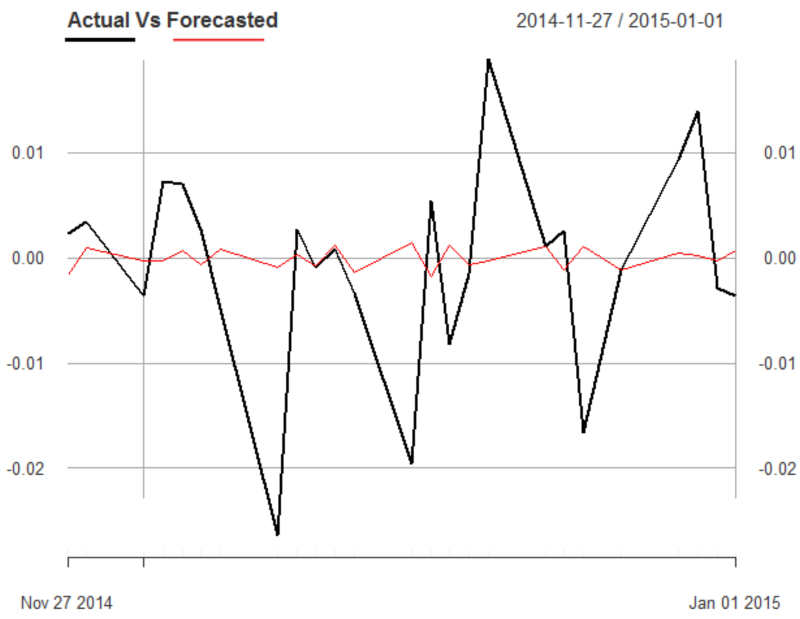


Figure 7. Actual vs forecasted using ARIMA model

Acknowledgment. This work is fully supported by Bina Nusantara University.

REFERENCES

[1] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting (Springer Text in Statistics)*, Springer International Publishers, Switzerland, 2016.

[2] *Introduction to Capital Market and Problem in Indonesia*, http://www.indonesiaforeigninvestment law.com/capital-market/indonesia-invesment-capital-market-in-indonesia/.

[3] *Analysis and Forecast of Indonesia's Jakarta Composite Index (IHSG)*, https://www.indonesia-investments.com/id/finance/financial-columns/analysis-and-forecast-of-indonesias-jakarta-composit e-index-ihsg/item1278.

[4] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning (Adaptive Computation and Machine Learning Series)*, MIT Publisher, Cambridge, MA, 2016.

[5] E. W. Saad, D. V. Prokhorov and D. C. Wunsch, Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks, *IEEE Trans. Neural Networks*, vol.9, pp.1456-1470, 1998.

[6] W. Bao, J. Yue and Y. Rao, A deep learning framework for financial time series using stacked autoencoders and long-short term memory, *PLoS ONE*, vol.12, no.7, pp.1-24, 2017.

[7] F. A. Gers, J. Schmidhuber and F. Cummins, Learning to forget: Continual prediction with LSTM, *Neural Computation*, vol.12, no.10, pp.2451-2471, 2000.

[8] *https://github.com/NourozR/Stock-Price-Prediction-LSTM.*

[9] P.-F. Pai and C.-S. Lin, A hybrid ARIMA and support vector machines model in stock price forecasting, *International Journal of Management Science*, vol.33, no.3, pp.497-505, 2005.

[10] *Stock Prediction Using ARIMA Model*, https://www.r-bloggers.com/forecasting-stock-returns-using-arima-model/.

[11] *Tutorial on Stock Prediction Using RNN*, https://lilianweng.github.io/lil-log/2017/07/08/predict-stock-prices-using-RNN-part-1.html.

[12] K. Chen, Y. Zhou and F. Dai, A LSTM-based method for stock returns prediction: A case study of China stock market, *IEEE International Conference on Big Data*, 2015.

[13] R. C. Cavalcante, R. C. Brasileiro, V. L. F. Souza, J. P. Nobrega and A. L. I. Oliveira, Computational intelligence and financial markets: A survey and future directions, *Expert Systems with Applications*, vol.55, pp.194-211, 2016.