Journal of Informatics and Data Science (J-IDS)
Vol. 1, No. 1, Juni 2022
ISSN XXXX – XXXX (Online)

Forecating Composite Stock Price Index (CSPI) Using *Long Short Term Memory* (LSTM)

Intan Elprida SIlaban
Prodi Matematika, Jurusan Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan,
Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan, Kab. Deli Serdang, Sumatera Utara
e-mail: intansilaban779@gmail.com

ABSTRAK

The Composite Stock Price Index (CSPI) is an index that displays developments the whole movement of the company's share price in the stock market which refers to the Indonesia Stock Exchange (IDX). Before considering investment, investors can predict the Indonesian stock market is up and down by CSPI analysis. The main objective of this research is to propose forecasting model of CSPI using Long Short Term Memory (LSTM). The performance of LSTM model measured by Root Mean Square Error (RMSE). The results showed that the best LSTM models is model with number of neuron in hidden layer and epoch (iterations) were 10 and 10, respectively. The RMSE values achieved from the LSTM models for testing data is 0,0633. Visually, the prediction graph is almost similar with original data.

Keywords : CSPI, LSTM, Neuron, Hidden Layer, Epoch, RSME

I. INTRODUCTION

The movement of Composite Stock Price Index (CSPI) is an important guide in analyzing the performance of trading of stocks market on the Indonesia Stock Exchange (IDX). The CSPI is also an overview of the Indonesian economy. Before considering investment, investors can predict the Indonesian stock market is up and down by CSPI analysis [1]. According to Suharli [2], it is necessary to forecast the stock price in the future to find out the risks of investments.

Several methods have been proposed in the literature for CSPI forecasting. The method is Fuzzy Time Series Markov Chain [3], ARCH-GARCH models [4], Exponential Smoothing and ARIMA [5], [6] and also Neural Network (NN) [7]. Choosing a specific forecasting technique is based on a compromise between the complexity of the solution, characteristics of the data and the desired prediction accuracy. According to Banarjee [6], ARIMA model cannot be used in case of data fluctuation. ARIMA model focuses on data empirical which is normal, linear, and assumed to be stationary [8].

Artificial Neural Network (ANN) is a system designed based by human brain workings. ANN can predicts stock price movement data very well, and has performs well at predicting based on data pattern [7]. According Ramoz [9], ANN can be classified into several categories. That is Feedforward Network (FFN), Recurrent Neural Network (RNN), Polynomial Networks (PLN) and Modular Networks (MN). It can be applied in forecasting time series data. Long Short-Term Memory (LSTM) is a RNN architecture that was designed by Hochreiter and Schmidhuber to address the vanishing and exploding gradient problems of conventional RNNs. RNNs and LSTMs have been successfully used for speech recognition and forecasting [10].

According to Zheng [11], LSTM applied to electric load forecasting. This reasearch compared LSTM with others method, that is SARIMA, NARX, SVR and NNETAR. The results showed that the LSTM is capable of forecasting complex univariate electric load time series with strong non-stationarity and non-seasonality. Gao [12] also used LSTM for forecasting. The research also compared LSTM with others method, that is Moving Average, Exponential Moving Average, and Support Vector Machine. The performance of models evaluated by RMSE. The results shows that LSTM had good performances with lowest RMSE values.

Diterima: 01-09-2022 Direvisi: 07-09-2022 Disetujui: 15-09-2022

The other research used LSTM is forecasting stock price in Indonesia which has been researched by Adhib [10]. The results shows that LSTM able to predict stock prices with good performance, this is indicated by a low error rate. In the research, concluded that LSTM is able to overcome long-term dependence and able to predict stock prices with accurate results.

In this paper, research do forecasts of Composite Stock Price Index (CSPI) using *Long Short Term Memory* (LSTM). The remainder of the paper is organized as follows. Section II explains LSTM; Section III presents the analysis. Finally, conclusions are indicated in Section IV.

II. METHODE

A recurrent neural network (RNN) is a special case of neural network where the objective is to predict the next step in the sequence of observations with respect to the previous steps observed in the sequence. In fact, the idea behind RNNs is to make use of sequential observations and learn from the earlier stages to forecast future trends. As a result, the earlier stages data need to be remembered when guessing the next steps. In RNNs, the hidden layers act as internal storage for storing the information captured in earlier stages of reading sequential data. RNNs are called “recurrent” because they perform the same task for every element of the sequence, with the characteristic of utilizing information captured earlier to predict future unseen sequential data. The major challenge with a typical generic RNN is that these networks remember only a few earlier steps in the sequence and thus are not suitable to remembering longer sequences of data. This challenging problem is solved using the “memory line” introduced in the Long Short-Term Memory (LSTM) [13] .

LSTM is generally known part of RNN [14]. The LSTM has four gates, which are input gates, new cell state candidate, forget gate and output gate. Input gate takes a new input point from outside and process newly coming data. New cell state candidate takes input from the output of the LSTM cell in the last iteration. Forget gate decides when to forget the output results and thus selects the optimal time lag for the input sequence. Output gate takes all results calculated and generate output. Figure 1 shows the architecture of the LSTM.

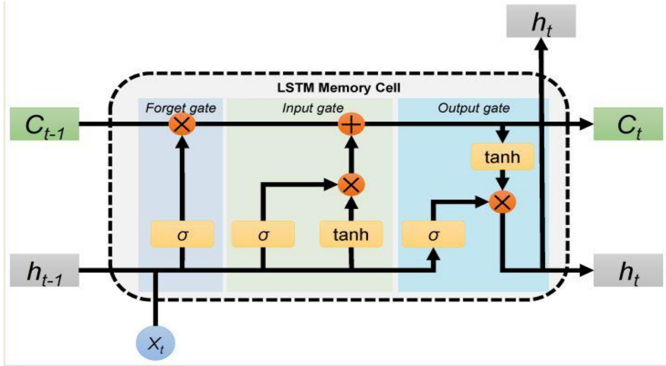

Fig. 1 : The architecture of LSTM

The equation is described as follows [11] :

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (1)$$

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \quad (2)$$

$$g_t = tanh(W_c.[h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * g_t \quad (4)$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * tanh\ tanh\ (C_t) \quad (6)$$

σ and tan *h* are applied which represent the specific, elementwise applied activation functions of the LSTM [14], where σ stand for the standard sigmoid function. The sigmoid layer outputs numbers between zero and one, describing how much of each component should be let through. A value of zero means “let nothing through,” while a value of one means “let everything through” i, f, o and c denote the mentioned inner-cell gates, respectively the input gate, forget gate, output gate, and cell activation vectors. c need to be equal to the hidden vector h. The W terms again denote weight matrices. The input gate can determine how incoming vectors x_t alter the state of the memory cell. The output gate can allow the memory cell to have an effect on the outputs. Finally, the forget gate allows the cell to remember or forget its previous state.

The equations of the sigmoid tanh function [10] are shown :

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (7)$$

$$tanh\ tanh\ (x) = 2\sigma(2x) - 1 \quad (8)$$

Diterima: 01-09-2022 Direvisi: 07-09-2022 Disetujui: 15-09-2022

Journal of Informatics and Data Science (J-IDS)
Vol. 1, No. 1, Juni 2022
ISSN XXXX – XXXX (Online)

III. ANALYSIS

The dataset used in this paper is daily CSPI dataset from Nov. 2014 to Oct. 2020, provided by www.finance.yahoo.com. The dataset shows in figure 2.

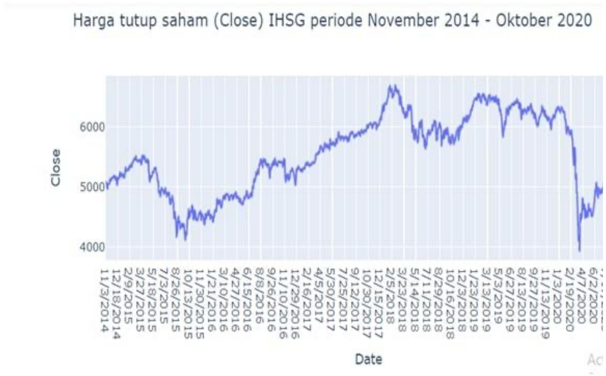

Fig. 2 : The histogram of daily CSPI dataset from Nov. 2014 to Oct.2020

Before the data is divided, the data is normalized using min-max-scaling. In this experiment, data is divided into two parts. The training datasets with 80% of the observations is used to train model, the remaining 20% is used to test the model prediction accuracy [15].

All the data we used make up a time series . We use the window method in the LSTM for regression. The size of the window is a parameter which we set as 7. The form of the training set is like $\{x_{t-6}, x_{t-5}, \dots, x_t \rightarrow x_{t+1}\}$ [15]. The LSTM architecture used in this study consists of 4 hidden layers with each layer containing dropout 0,2 and 1 output layer with 1 node. The number of neurons in the hidden layer will be trained by several numbers, which are respectively 10,20,30,40 and 50. The numbers the number of neurons in the hidden layer will be trained by several numbers. The number will be trained with the number of epoch which are 10, 20, 30 and 40. From such combinations obtain RMSE value. The training combinations with RMSE value are shown in the table 1.

Table 1 : The training combinations with RMSE value

Test	Number of neurons in hidden layer	Number of epoch	RMSE
1	10	10	0,3021
2	10	20	0,3088
3	10	30	0,3108
4	10	40	0,3114
5	20	10	0,3145
6	20	20	0,3153
7	20	30	0,3110
8	20	40	0,3144
9	30	10	0,3152
10	30	20	0,3128
11	30	30	0,3149
12	30	40	0,3175
13	40	10	0,3161
14	40	20	0,3175
15	40	30	0,3166
16	40	40	0,3110
17	50	10	0,3165
18	50	20	0,3135
19	50	30	0,3178
20	50	40	0,3101

Diterima: 01-09-2022 Direvisi: 07-09-2022 Disetujui: 15-09-2022

From the table it can be seen that the model with the number of neurons in hidden layer 10 and number of epoch 10 obtained the smallest RMSE. then this model is used in testing. For information, the RMSE value is calculated through the formula in the eq. 9.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \quad (y - \tilde{y})^2} \quad \textbf{(9)}$$

with n is numbers of dataset, y is predicted data value and $\tilde{y}$ is actual data value.

Furthermore, testing is carried out to obtain predictive results. the model used is a model that has been previously selected. The RMSE value for testing is 0,0633. The test results obtained are shown in table 2.

Table 2 : The test results with model numbers of hidden layer 10 and numbers of epoch 10

No	Actual data (min-max-scaling)	Predicted data (min-max-scaling)	Actual data	Predicted data
1	0,8275	0,8518	6269,62	6182,62
2	0,8506	0,8241	6205,35	6205,35
3	0,8519	0,8257	6209,68	6209,68
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
283	0,4384	0,4048	5144,05	5051,42
284	0,4327	0,4049	5128,23	5051,97

IV. CONCLUSION

The conclusions from the research of Forecating Composite Stock Price Index (CSPI) using *Long Short Term Memory* (LSTM) as follows :

1. The optimal LSTM model for forecasting CSPI is a model with the number of neurons in hidden layer 10 and the number of epoch 10 with RMSE training and RMSE testing, respectively 0.3021 and 0.0633.
2. Forecasting results generated from the LSTM model with the number of neurons on hidden layer 10 and the number of epochs 10 for the next 284 days, are between the initial price range from 6182.62 to the last data is 5051.97. Visually, the prediction graph is almost like the real data.

For future researchers who will continue and develop this research suggests to adding training variations, both in the number of hidden layer neurons as well as variations in the number of epochs and research can be carried out, thats the comparison of prediction results LSTM with other forecasting models.

REFERENCE

[1] Ekayana, Sangkasari Paranita., S. D. J. S., (2017): Pergerakan Indeks Harga Saham Gabungan dan Faktor-Faktor yang Mempengaruhinya, Jurnal Studi Manajemen Organisasi, 2(14).

[2] Suharli, Michell., d. M. O., (2005): Memprediksi Tingkat Pengembalian Investasi pada Equity Securities melalui Rasio Profitabilitas, Likuiditas dan Hutang pada Perusahaan Publik di Jakarta, Simposium Nasional Akuntansi, 8, 288– 296.

[3] Aristyani, Y., d. E. S., (2015): Peramalan Indeks Harga Saham Gabungan (IHSG) dengan Metode Fuzzy Time Series Markov Chain, Jurnal MIPA, 38(2), 186– 196.

[4] Syafitri, E., (2013): Peramalan Indeks Harga Saham Gabungan di Bursa Efek Jakarta dengan Model ARCH-GARCH, Skripsi FMIPA Universitas Sumatera Utara, .

[5] Warsini (2011): Perbandingan Metode Pemulusan (Smoothing) Eksponensial dan ARIMA (Box-Jenkins) sebagai Metode Peramalan Indeks Harga Saham Gabungan, Skripsi FMIPA Universitas Sumatera Utara, .

[6] Banarjee

[7] Halimi, Imam., d. W. A. K., (2018): Prediksi Indeks Harga Saham Gabungan (IHSG) Menggunakan Algoritma Neural Network, Jurnal Edukasi dan Penelitian Informatika, 4(1).

[8] Wartati, Desy., d. N. A., (2016): Aplikasi Jaringan Syaraf Tiruan dan Particle Swarm Optimation untuk Peramalan Indeks Harga Saham Bursa Efek Indnesia, Jurnal TeknoSains, 6(1), 22–30.

[9] Ramoz., d. M., (2012): A Review of Artificial Neural Networks : How Well Do They Perform in Forecasting Time Series, Analitika : Revista de analisis estadiscito, 6(2), 7–15.

[10] Adhib, Arfan., d. L. E., (2019): Prediksi Harga Saham di Indonesia Menggunakan Algoritma Long Short Term Memory, Seminar Nasional Teknologi Informasi dan Komunikasi, 3(1).

[11] Zheng, Jian., C. X. Z. Z. X. L., (2017): Electric Load Forecasting in Smart Grid Using Long Short Term Memory based Recurrent Neural Networks, 51st Annual Conference on Information Sciences and Systems (CISS), 1–6.

[12] Gao, Tingwei., Y. C. d. Y. L., (2018): Applying Long Short Term Memory Neural Networks for Predicting Stock Closing Price, 2017 8th IEEE International Conference on Software Engineering and Service Scince (ICSESS), 3–6.

Diterima: 01-09-2022 Direvisi: 07-09-2022 Disetujui: 15-09-2022

[13] Fazano, Fikri El Huda., d. R. Y., (2020): Implementasi Long Short-Term Memory Pada Harga Saham Perusahaan Perkebunan Di Indonesia, Jurnal UJMC, 6(1), 9–18.

Diterima: 01-09-2022 Direvisi: 07-09-2022 Disetujui: 15-09-2022