Stock Price Prediction in Vietnam using Stacked LSTM

Nguyen Trung Tuan[1[0000-0003-3249-2444]], Thu Hang Nguyen[2[0000-0002-1959-5708]] and Thanh Thi Hien Duong[2[0000-0001-5854-5845]]

[1] National Economics University
[2] Hanoi University of Mining and Geology, Vietnam
tuannt@neu.edu.vn
{nguyenthuhang,duongthihienthanh}@humg.edu.vn

Abstract. The direction of the stock market is always complex, stochastic, and highly volatile. In addition to traditional forecasting models such as linear regression and Automatic Regression Integrated Moving Average (ARIMA) models, analysts are now trying to apply modern deep learning models to predict trends direction of the stock market to achieve more accurate forecasting. In this conducting research, we have investigated and applied the state-of-the-art deep learning sequential model, namely the Stacked Long Short-Term Memory Model (Stacked LSTM) to the prediction of stock prices the next day. The experimental result on three benchmark datasets: stocks of Apple Inc. (AAPL), stocks of An Phat Bioplastic JSC (AAA), and stocks of Bank of Foreign Trade of Vietnam (VCB) has shown the effectiveness of the predictive model. Furthermore, we discovered that the suitable quantity of hidden layers is two, and when we continue to increase the quantity of hidden layers to three or four, the Stacked LSTM model does not improve the predictive power, even though it has a more complex model structure.

Keywords: Stock Price Prediction, Stock Forecasting, Time Series Forecasting, Stacked LSTM.

1 Introduction

Stock prices have been a popular topic in the modern economy. In the stock market, because of high volatility, the stock price prediction has an important impact on the decisions in trading and investing. Hence, stock price forecasting has been an interesting field for researchers, traders, investors, and corporate.

Financial time series with flexible and complex variables are not easy for forecasting. Data of the stock market are numerous and almost nonlinear, so we must develop models which can dissect many hidden layers. Traditional statistical analysis models have been widely used for a long time in economics and finance data analysis, including stock forecasting problems, such as exponential smoothing and Autoregressive Integrated Moving Average (ARIMA) [1]–[3]. Recent studies have shown that deep learning models are capable of exploring hidden and dynamics patterns in the data by learning themselves, so these models can give better forecasting results than statistical [4].

Furthermore, Recurrent neural networks (RNN) are powerful types of neural networks designed to resolve sequence dependence. RNN uses not only the input data, but also the previous outputs for predicting the current output, so it is considered the network for sequential data. RNN association of the nodes makes a directed diagram and internal memory themselves are used to deal with flexible input sequences. The state of each node is time-varying by the real-valued activation function. The learning model in RNN is determined by transition between states so it always has the same input size. Otherwise, the same transaction function which has the same parameters at each step was used in the system [4].

A type of RNN is the Long Short-Term Memory network (LSTM) which has large structures and is used to train data with a successful outcome. An issue RNN faces are having vanishing and exploding gradients. To address these issues, LSTM networks operate "computational gates" which help handle data and keep more needed information. Hence, LSTM models have outperformed with time-series data, compared to other sequences [5]. Furthermore, the price forecasting in the stock market has been implemented with more highly accurate by using LSTM networks. Di Persio and Honchar performed three kinds of RNNs, namely: a basic RNN, an LSTM, and a Gated Recurrent Network (GRU), and included that the accurate results of the other RNNs were not good as the LSTM which was at 72% [6]. Pang et al. suggested two LSTM models to forecast the stock market: one had an embedding layer, and another had an autoencoder. The outcomes of the LSTM with embedding are more quality, with the result of accuracy being 57.2, compared to 56.9% of another [7].

There are some crucial factors to improve the performance of DNN architecture for prediction. Hiransha et al. showed that large of data is the first factor according to the bigger of the quantity of data, the higher the quality result of the model [4]. Otherwise, Adding or reducing the quantity of hidden layers is also affected to perform of models. The research of Karsoliya confirmed that the model can be had issues in training after the fourth layer. Moreover, the size of hidden notes in each layer can be followed by the thumb regulations that the hidden layer has a quantity of nodes is 2/3 the quantity of nodes in the input layer [8]. Hossain et al. performed two models LSTM and GRU for prediction. First, the features were passing the LSTM to work for forecasting. And then this forecast result was got to the GRU model to perform one more forecasting. The outcome of this model was better than the performance of the LSTM model or GRU model when they worked independently [9]. In recent years, combining models for forecasting have become popular, so the optimization of each model is considered. Assunta et al. predicted stock prices by implementing the multilayer perceptron (MLP) model. It found that in each layer, the optimization of the size of hidden layers and hidden nodes performed differently in each case and must be investigated through trial and error [10].

Many models have been used to predict stock price, but LSTM still has been one of the most common choices for experiments with successful results. Motivated by this trend, in our study, we investigate the Stacked LSTM models by considering the number of hidden layers, hidden nodes in each layer, and size of data aiming to find out the optimized model for forecasting stock price. The models have been verified with three datasets collected from the daily stock prices in the past of three companies, namely:

Apple Inc. (AAPL), An Phat Bioplastic JSC (AAA), and Bank of Foreign Trade of Vietnam (VCB).

The structure of the paper is arranged as follows. In Section 2, we introduce brief presentations of the LSTM and then the Stacked LSTM model - the model that we will optimize architecture for stock price prediction. We then provide some empirical analysis to point out the params settings for the Stacked LSTM model to achieve the best prediction outcomes in the next part - Section 3. Finally, the last section - Section 4 is a conclusion.

2 Methodology

2.1 Long Short-Term Memory

A particular model of RNN is the LSTM network which introduces an internal cell state or memory state and gating mechanisms. It can retain short-term memory while capturing long-range dependencies in data [11]. The LSTM models were applicated in the financial domain [9],[10], sequence learning domain [14] and have achieved a lot of results.

Each cell of LSTM works with gates: Input gate (i_t), output gate (o_t) and forget gate (f_t) (see in Fig.1) [5], [15], [16].

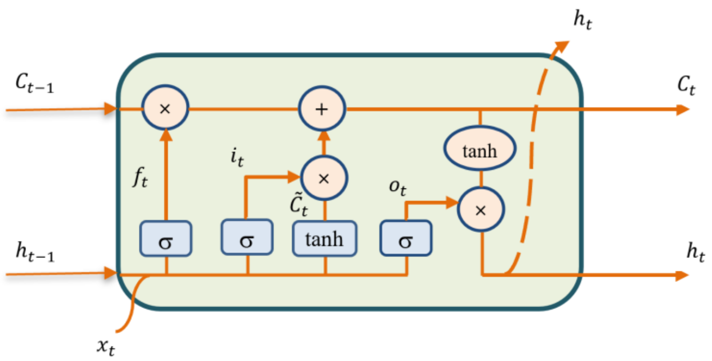


Fig. 1. LSTM cell [15] [13]

The forget gate will get which candidate data from the previous cell state and clear out. It takes the inputs and output of the previous hidden state (h_{t-1}), and input of current state (x_t), and input them through the sigmoid activation function (σ) which each value output is a vector between 0 and 1. The output is 0 means that the information is removed while 1 indicates that the information is kept.

The input gate with another sigmoid layer will choose the value to update in the cell state. A memory cell ($\tilde{C}_t$) is also created which uses tanh activation function on the same inputs, but the output is between -1 and 1. The information from the cell state is dropped with the negative result and is added with the positive value. The result was defined by how much each cell state should be updated, and times the output from tanh and input gate sigmoid activation:

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

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

$$\widetilde{C_t} = tanh(W_C.[h_{t-1}, x_t] + b_C) \tag{3}$$

Finally, a new cell state was obtained by the forget vector (f_t) times the previous cell state (C_{t-1}) and the result is added with the multiplication between input gate and tanh vector:

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \tag{4}$$

The output gate (o_t) (a sigmoid layer) will be used to filter output for the next step, and return to the hidden state (h_t):

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

$$h_t = o_t * tanh(C_t) \tag{6}$$

2.2 Stacked LSTM

The foundational LSTM model includes only one hidden layer. The network was expanded by adding hidden LSTM layers called Stacked LSTM. Each hidden layer of this network includes multiple cell states which are piled on top of each other. Furthermore, we receive one outcome after each input time step through each layer. Therefore, each hidden layer of LSTM can have a sequence output for all input time steps, instead of one output like other models [17].

A Multilayer Perceptron becomes deeper when more hidden layers are added. The learned performances through training process from previous layers were collected by the expanding hidden layer and create a new performance of networks which was developed to high levels of abstraction. Hence, the approach for prediction was improved with more accuracy based on the multi-layer model [18], [19].

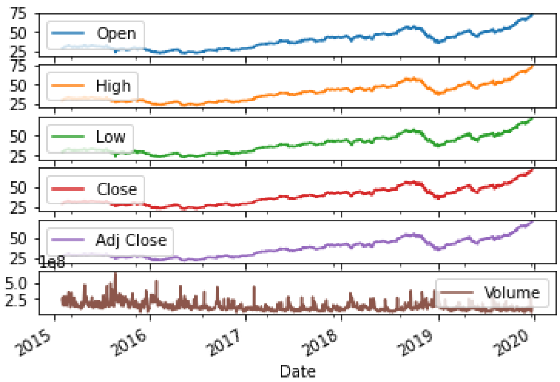
Nowadays, Stacked LSTM is a strong method to handle the complex sequence data for forecasting as of result of outperforming [20], [21].

3 Experiments and Results

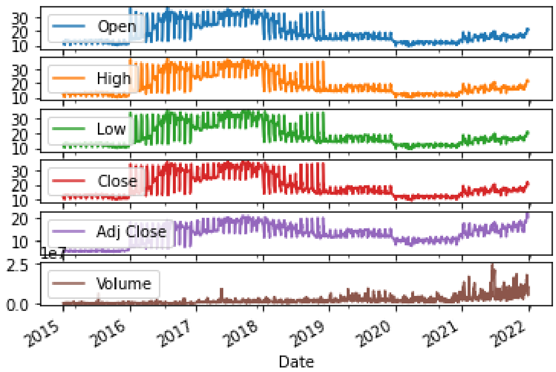
3.1 Datasets

The conducting research was done for three datasets of stock price, namely: AAPL stock was obtained from Yahoo Finance which is used widely in research for forecasting stock price; AAA stock and VCB stock which were collected from Vietstock [22]. The daily history prices of AAA stock and VCB stock are from 5 January 2015 to 31

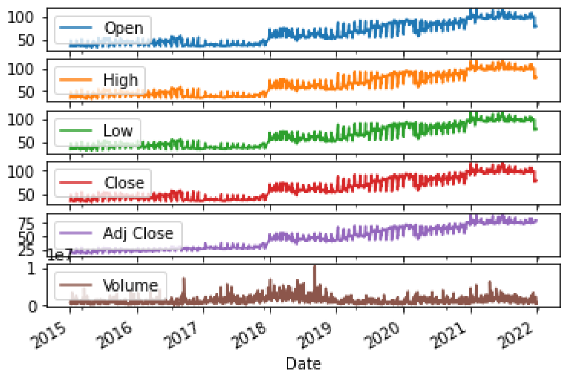
December 2021, including 1744 and 1749 data points, respectively. AAPL stock is between 31 January 2015 to 29 December 2019, with 1236 data points. The attribute categories of these data have six types, namely: Open, High, Low, Close, Adj Close, Volume (see in Fig.2).



(a)



(b)



(c)

Fig. 2. Stock price performance: (a) AAPL stock price, (b) AAA stock price, (c) VCB stock price.

3.2 Parameter settings and evaluation metrics

We use Keras library for Python to implement DNN model and Jupyter Notebook was used to conduct the analysis. To define the LSTM dataset, the data of the model was divided into two sets: the training data was indicated for 80% of the dataset and the test data was set for the rest with 20%. The quantity of layers in the Stacked LSTM model was designed from 2 to 4 hidden layers. A dropout layer was set before the Dense layer to avoid overfitting. The regulation of thumb method was applied so the quantity of nodes in the hidden layers is 2/3 of the size of the input layer [8]. The detail of the quantity of LSTM layers for each case is shown in Table 1.

Table 1. The quantity of layers in the Stacked LSTM model

Quantity of LSTM layers	Nodes of LSTM layers in order
2	(45, 30)
3	(67, 45, 30)
4	(100, 67, 45, 30)

While training data, we used mean squared error (MSE) which is popular and applied the most as a loss function. Besides that, the Stacked LSTM network was optimized by applying the Adam algorithm [15], [21], [23]. Furthermore, with big data, the model was performed with 30 batches, 1000 epochs, and put sliding window size was 7 days.

Finally, the prediction models are evaluated by three used widely metrics in the research community, which are Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) where the lower value the better. Let N be the number of samples, x_t and $\hat{x}_t$ are the ground-truth and prediction results, respectively, the evaluation metrics RMSE, MAE, and MAPE are defined by equations 7, 8, and 9, respectively [20], [24].

$$RMSE\ (x, \hat{x}) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_t - \hat{x}_t)^2} \tag{7}$$

$$MAE(x, \hat{x}) = \frac{1}{N}\sum_{t=1}^{N}|x_t - \hat{x}_t| \tag{8}$$

$$MAPE(x, \hat{x}) = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x_t - \hat{x}_t}{x_t}\right| \tag{9}$$

3.3 Results and discussion

The outcomes of evaluation metrics from Table 2 to Table 4 showed that the case with 2 LSTM layers in all models had the best performance with the lowest error measures. In the prediction models of AAPL stock prices, the RMSE is 0.725, the MAE is 0.503 and MAPE is 0.009. AAA stock prices with RMSE, MAE, and MAPE are 0.285, 0.219, and 0.014, respectively. The results of VCB stock are 1.793, 0.976, and 0.010 for RMSE, MAE, and MAPE, respectively. The prediction graphs for these works are displayed in Fig.3. Otherwise, the value of error measures was increased when we added more the number of LSTM layers to 3 or 4. In the case of both AAPL and VCB stock prices, the error measures of 4 LSTM layers were about double compared with 3 LSTM layers.

Compared with other LSTM works, our performance has an important contribution that we can improve the model by optimizing the parameters which can make the model become the most efficient only with two hidden layers of architecture.

Table 2. The performance of evaluation metrics for AAPL stock predictions

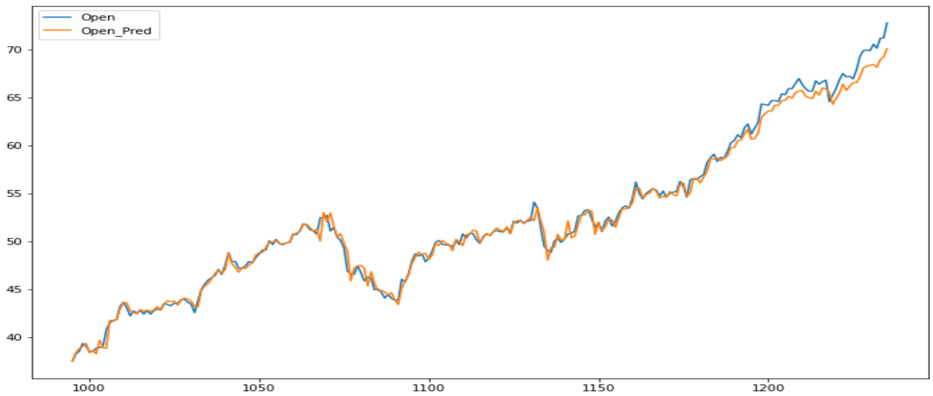
Quantity of LSTM layers	Metrics		
	RMSE	MAE	MAPE
2	**0.725**	**0.503**	**0.009**
3	1.280	0.949	0.017
4	2.695	1.464	0.024

Table 3. The performance of evaluation metrics for AAA stock predictions

Quantity of LSTM layers	Metrics		
	RMSE	MAE	MAPE
2	**0.285**	**0.219**	**0.014**
3	0.380	0.278	0.018
4	0.352	0.261	0.016

Table 4. The performance of evaluation metrics for VCB stock predictions

Quantity of LSTM layers	Metrics		
	RMSE	MAE	MAPE
2	**1.793**	**0.976**	**0.010**
3	1.994	1.112	0.012
4	4.063	3.078	0.031



(a)

(b)

(c)

Fig. 3. Performance of Stock price prediction for 2 LSTM layers: (a) AAPL stock price, (b) AAA stock price, (c) VCB stock price

4 Conclusion

In conducting research on stock price prediction, we recommend a neural network based on optimizing Stacked LSTM architecture. One of the approaches to improve performance is considering specific factors which are the quantity of data, the ratio of dataset divided into training and test set, the quantity of LSTM layers, and the nodes of each LSTM layer. Our implementation showed that we can achieve better outcomes after only 2 LSTM layers. The model has a lot of data that need to be trained so this performance helps to save the training time and complexity. Furthermore, our models were tested on three stock prices datasets of Apple Inc. (AAPL), An Phat Bioplastic JSC (AAA), and Bank of Foreign Trade of Vietnam (VCB), so it will help the traders and investors have more useful information to make good decisions and get more profit in Vietnam stock market. In future works, we will extend more crucial features that affect the stock price and could achieve real-time stock market forecasting in Viet Nam.

Acknowledgements This work was funded by Hanoi University of Mining and Geology under grant number 65/QD-MDC.

References

[1] M. Khashei and Z. Hajirahimi, "A comparative study of series arima/mlp hybrid models for stock price forecasting," *Communications in Statistics - Simulation and Computation*, vol. 48, no. 9, pp. 2625–2640, Oct. 2019, doi: 10.1080/03610918.2018.1458138.

[2] M. Kumar and M. Thenmozhi, "Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models," *IJBAAF*, vol. 5, no. 3, p. 284, 2014, doi: 10.1504/IJBAAF.2014.064307.

[3] S. A. Wadi, M. Almasarweh, and A. A. Alsaraireh, "Predicting Closed Price Time Series Data Using ARIMA Model," *MAS*, vol. 12, no. 11, p. 181, Oct. 2018, doi: 10.5539/mas.v12n11p181.

[4] H. M, G. E.A., V. K. Menon, and S. K.P., "NSE Stock Market Prediction Using Deep-Learning Models," *Procedia Computer Science*, vol. 132, pp. 1351–1362, 2018, doi: 10.1016/j.procs.2018.05.050.

[5] C. Olah, "Understanding lstm networks." [Online]. Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/, 2015.

[6] L. Di Persio and O. Honchar, "Analysis of recurrent neural networks for short-term energy load forecasting," Thessaloniki, Greece, 2017, p. 190006. doi: 10.1063/1.5012469.

[7] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, "An innovative neural network approach for stock market prediction," *J Supercomput*, vol. 76, no. 3, pp. 2098–2118, Mar. 2020, doi: 10.1007/s11227-017-2228-y.

[8] S. Karsoliya, "Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture," *International Journal of Engineering Trends and Technology*, p. 4, 2012.

[9] M. A. Hossain, R. Karim, R. Thulasiram, N. D. B. Bruce, and Y. Wang, "Hybrid Deep Learning Model for Stock Price Prediction," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, Bangalore, India, Nov. 2018, pp. 1837–1844. doi: 10.1109/SSCI.2018.8628641.

[10] Universiti Malaysia Terengganu, H. Salleh, and A. M. P. Atrick Vincent, "AN INVESTIGATION INTO THE PERFORMANCE OF THE MULTILAYER PERCEPTRON ARCHITECTURE OF DEEP LEARNING IN FORECASTING STOCK

PRICES," *UMT JUR*, vol. 3, no. 2, pp. 61–68, Apr. 2021, doi: 10.46754/umtjur.2021.04.006.
[11] S. Hochreiter, "The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions," *Int. J. Unc. Fuzz. Knowl. Based Syst.*, vol. 06, no. 02, pp. 107–116, Apr. 1998, doi: 10.1142/S0218488598000094.
[12] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, Oct. 2018, doi: 10.1016/j.ejor.2017.11.054.
[13] J. B. Heaton, N. G. Polson, and J. H. Witte, "Deep learning for finance: deep portfolios: J. B. HEATON, N. G. POLSON AND J. H. WITTE," *Appl. Stochastic Models Bus. Ind.*, vol. 33, no. 1, pp. 3–12, Jan. 2017, doi: 10.1002/asmb.2209.
[14] M.-T. Luong, I. Sutskever, Q. V. Le, O. Vinyals, and W. Zaremba, "Addressing the Rare Word Problem in Neural Machine Translation," *arXiv:1410.8206 [cs]*, May 2015, Accessed: Mar. 01, 2022. [Online]. Available: http://arxiv.org/abs/1410.8206
[15] S. Aryal, D. Nadarajah, P. L. Rupasinghe, C. Jayawardena, and D. Kasthurirathna, "Comparative Analysis of Deep Learning Models for Multi-Step Prediction of Financial Time Series," *Journal of Computer Science*, vol. 16, no. 10, pp. 1401–1416, Oct. 2020, doi: 10.3844/jcssp.2020.1401.1416.
[16] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[17] Jason Brownlee, "Stacked Long Short-Term Memory Networks (machinelearningmastery.com)." Aug. 18, 2017. [Online]. Available: https://machinelearningmastery.com/stacked-long-short-term-memory-networks/
[18] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to Construct Deep Recurrent Neural Networks," *arXiv:1312.6026 [cs, stat]*, Apr. 2014, Accessed: Feb. 24, 2022. [Online]. Available: http://arxiv.org/abs/1312.6026
[19] A. Graves, N. Jaitly, and A. Mohamed, "Hybrid speech recognition with Deep Bidirectional LSTM," in *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, Olomouc, Czech Republic, Dec. 2013, pp. 273–278. doi: 10.1109/ASRU.2013.6707742.
[20] Allison Koenecke, "Applying Deep Neural Networks to Financial Time Series Forecasting." Institute for Computational & Mathematical Engineering, Stanford, California, USA, 2020. [Online]. Available: https://web.stanford.edu/~koenecke/files/Deep_Learning_for_Time_Series_Tutorial.pdf
[21] M. Al Ridhawi, "Stock Market Prediction Through Sentiment Analysis of Social-Media and Financial Stock Data Using Machine Learning," Oct. 2021, doi: 10.20381/RUOR-27045.
[22] "Vietstock." Vietstock. [Online]. Available: https://vietstock.vn/
[23] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Jan. 2017, Accessed: Mar. 02, 2022. [Online]. Available: http://arxiv.org/abs/1412.6980
[24] Pietro Terna, *A Deep Learning Model to Forecast Financial Time-Series*. 2015.