

# The application of Long Short-Term Memory algorithm in American multinational technology company stock prediction

**Yufei Wang**

Ramona Convent Secondary School, 1701 W. Ramona Rd, Alhambra CA 91803, USA

ywang.23@ramonaconvent.org

**Abstract.** The fast development in American multinational technology companies has attracted both professional and new investors to buy the stocks. However, the price of these companies are unstable and therefore hard to be predicted. The focus of this article is to use AI and deep learning algorithms to find a pattern of the stock price. Long Short-Term Memory Algorithm (LSTM) is the main algorithm used to predict the trend, and other methods including Autoregressive integrated moving average (ARIMA), Seasonal autoregressive integrated moving average (SARIMA), and prophet are also discussed in this piece.

**Keywords:** big data, Long Short-Term Memory, Artificial Intelligence and deep-learning.

## 1. Introduction

Stock, one of the most profitable but also riskiest way to make fortune used to be regarded as random but now are scientifically analyzed and predicted. Long Short Term Memory (LSTM) is widely used in the fields of artificial intelligence and deep learning. This article will mainly focus on forming the basic stock forecasting algorithm in Python with LSTM and constructing a data set for stock of American multinational technology company. With this methodology, people can quickly obtain the precise indicator and result.

## 2. Methodology

In the dataset of American multinational technology company in 2019 [1], the stock information can be imported in the table with multiple variables in one dataset - date, high, low, close, revenue, profit, P/E and growth rate, etc. Date is the period needed to forecast. High and low indicate the highest and lowest prices of the stock that day. Income and profit show how much money was earned on the stock that day, minus the cost price. The price-earnings ratio and growth rate represent the percentage of the stock's gain and the percentage of money added in the market that day.

In the next step, to analyze the data and organize and classify the data, users first apply to the most basic machine learning algorithm, linear regression. Linear regression models generate an equation that determines the relationship between independent and dependent variables. It can be written as:  $y = \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_n X_n$

$x_1, x_2, \dots, x_n$  represent independent variables, and coefficients  $\theta_1, \theta_2, \dots, \theta_n$  represent weights. Due to the large number of the independent variables, features using the time (date) column are extracted and then fitted to a linear regression model since date is the input.

To implement the analysis of the dataset, a program is created to sort the dataset in a certain order, and then generate a separate dataset. Separate datasets can help continue the next step without affecting other required data.

Then, the dataset can be applied in the LSTM algorithm in sequence prediction problems.

LSTM has three gates:

Input gates: Input gates add information to the cell state

forget gate: it removes information that the model no longer needs

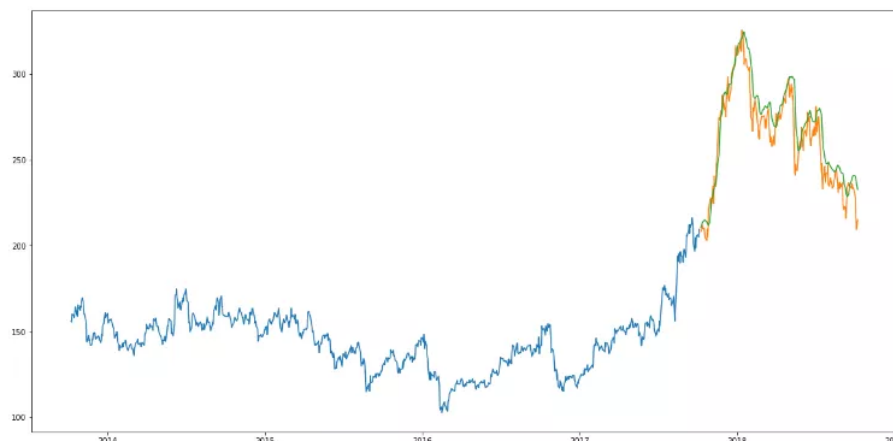
Output gate: The output gate of the LSTM selects the information as the output

In LSTM, the information to omit is determined in the first step with the sigmoid function. It looks at the previous state ( $h_{t-1}$ ) and the current input  $x_t$  and evaluates the function.

The second layer has two functions. The first is the sigmoid function and the second is the tanh function. The sigmoid function decides which values to let through (0 or 1). The tanh function assigns weights to the passed values, determining their level of importance, from -1 to 1.

The final step is to run a sigmoid layer to determine which parts of the cell state make it output. Then setting a value between -1 and 1 via the tanh function can be multiplied by the output of the sigmoid gate.

### 3. Results



**Figure 1.** A figure name.

In the figure above, people can compare and contrast the results of the linear and LSTM algorithm. Since LSTM is known for its accurate prediction with numerous uncertain inputs [2], LSTM is obviously more in line with the actual trend. The LSTM model can adjust various parameters to make stock predictions more accurate. And because linear regression is too fixed, it will lead to some parameter changes that are very different from the parameters.

After comparing the predicted results, it can be found that at least the accuracy of the prediction is more than 50%. However, the LSTM prediction is generally lower than the actual. It might be the influence of randomness of market.

### 4. Discussion

With the analysis of stocks, it is not difficult to find that there is a big difference between linear regression and LSTM in the prediction of stocks. The reason for these differences is the randomness of the market including globalization and development in technology. Optimization of the parameters and development in better algorithm are necessary to better prediction [3].

Certain bias in LSTM algorithm has been discussed and improved by combining with other algorithm or adding more restrictions [4]. However, for beginners in the stock market, this model is a reliable source to obtain the result.

After fitting the model, the predicted trend can lead to the result. Then users can perform some analysis through the decision tree model and other methods, so that they can clearly see whether there is a problem with the data through the decision tree. Other forecasting techniques like Autoregressive integrated moving average (ARIMA)[5], Seasonal ARIMA (SARIMA)[6], and Prophet[7] are also useful in forecasting.

## 5. Conclusion

Time series forecasting is still developing, especially after the epidemic which greatly influence the global economy. And stock analysis is an extension of this field. When people use machine learning to analyze stocks, they can find that stocks have strong randomness and variability. Even so, making some simple predictions using methods like LSTM would at least give some advice when it comes to buying American multinational technology company stocks, especially useful for people who do not have a clue but want to make a purchase. The accuracy is high enough for people to make rational decision and the difference can be adjusted with further studies.

## References

- [1] Google Stock Data | Kaggle
- [2] Kumar G, Singh UP, Jain S. An adaptive particle swarm optimization-based hybrid long short-term memory model for stock price time series forecasting. *Soft comput.* 2022;26(22):12115-12135. doi: 10.1007/s00500-022-07451-8. Epub 2022 Aug 26. PMID: 36043118; PMCID: PMC9415266.
- [3] Huang Y, Gao Y, Gan Y, Ye M. A new financial data forecasting model using genetic algorithm and long short-term memory network. *Neurocomputing.* 2021;425:207–218.
- [4] Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions. *Eur J Oper Res.* 2018;270(2):654–669.
- [5] Alwee R, Shamsuddin SM, Sallehuddin R. Hybrid support vector regression and autoregressive integrated moving average models improved by particle swarm optimization for property crime rates forecasting with economic indicators. *ScientificWorldJournal.* 2013 May 23;2013:951475. doi: 10.1155/2013/951475. PMID: 23766729; PMCID: PMC3677664.
- [6] Suradhaniwar S, Kar S, Durbha SS, Jagarlapudi A. Time Series Forecasting of Univariate Agrometeorological Data: A Comparative Performance Evaluation via One-Step and Multi-Step Ahead Forecasting Strategies. *Sensors (Basel).* 2021 Apr 1;21(7):2430. doi: 10.3390/s21072430. PMID: 33916026; PMCID: PMC8037998.
- [7] Crash Diagnosis and Price Rebound Prediction in NYSE Composite Index Based on Visibility Graph and Time-Evolving Stock Correlation Network