

# Predicting Google's Stock Price with LSTM Model

Tianlei Zhu<sup>1</sup>, Yuexin Liao<sup>1</sup>, Zheng Tao<sup>2\*</sup>

<sup>1</sup>Wenzhou-Kean University, Whenzhou 325000, Zhejiang Province, China

<sup>2</sup>University of Liverpool, Liverpool 999020, Merseyside Province, United Kingdom

\*Corresponding author: Zheng Tao, taozhe@kean.edu

**Copyright:** © 2022 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

**Abstract:** Stock market has a profound impact on the market economy, Hence, the prediction of future movement of stocks is of great significance to investors. Therefore, an efficient prediction system can solve this problem to a great extent. In this paper, we used the stock price of Google Inc. as a prediction object, selected 3810 adjusted closing prices, and used long short-term memory (LSTM) method to predict the future price trend of the stock. We built a three-layer LSTM model and divided the entire data into a test set and a training set according to the ratio of 8 to 2. The final results show that while the LSTM model can predict the stock trend of Google Inc. very well, it cannot predict the specific price accurately.

**Keywords:** Google; Stock prediction; LSTM model; Stock trend

**Online publication:** October 21, 2022

## 1. Introduction

Google is far ahead in global search. It is so eminent that foreign consumers use “Google search” to describe the Internet to retrieve information rather than “online search”, and only the world’s most valuable brands can use their own names in place of product or service’s names. Google also benefits from scale advantages and network effects. The more users use Google’s ecosystem of services, the more the company can learn from its data, and thus its ability to leverage artificial intelligence grows exponentially over time. Network effects provide Google with an additional source of sustained competitive advantage. The more videos are uploaded to YouTube, the more viewers are drawn to the platform. Larger audiences also provide an incentive for creators to keep producing more and better contents. In this case, creators and viewers gravitate to the leading video platform, creating a self-sustaining virtuous cycle for the business. From a fundamental point of view, it’s not an exaggeration to say that Google is one of the most competitive companies in the world, and these mentioned advantages enable the company to continue to create value for investors over the long term. Its remarkable revenue growth, superior profitability and high stock evaluation all lays the solid foundation for stock prediction on Google.

Stock price predictions are theoretically possible. However, it is impossible to pinpoint all of the elements that may influence stock prices and how they will affect stocks. This is because the forecasting model should be able to handle nonlinear problems because the stock prediction is highly nonlinear. Nevertheless, it is suitable to use the recurrent neural network (RNN) to predict since the stock has the characteristics of time series <sup>[1]</sup>. Although the RNN models allows the persistence of information, however, the general RNN model is weak to describe the time series data with long memory. For example, considering the time series is too long, the phenomenon of gradient dissipation and gradient explosion makes RNN training exceedingly difficult. The long short-term memory (LSTM) model proposed by

Hochreiter and Schmid Huber is modified on the basis of the RNN structure, thereby, solving the problem that the RNN model cannot describe the long memory of time series <sup>[2]</sup>.

The rest of the paper is arranged as follows: Section 2 is primarily intended to make a literature review; Section 3 describes the methodology of predicting stocks; Section 4 includes the elucidation of models as well as the analysis of the experimental results; Lastly, the paper ends with concluding remarks and relevant future perspectives in Section 5.

## 2. Literature review

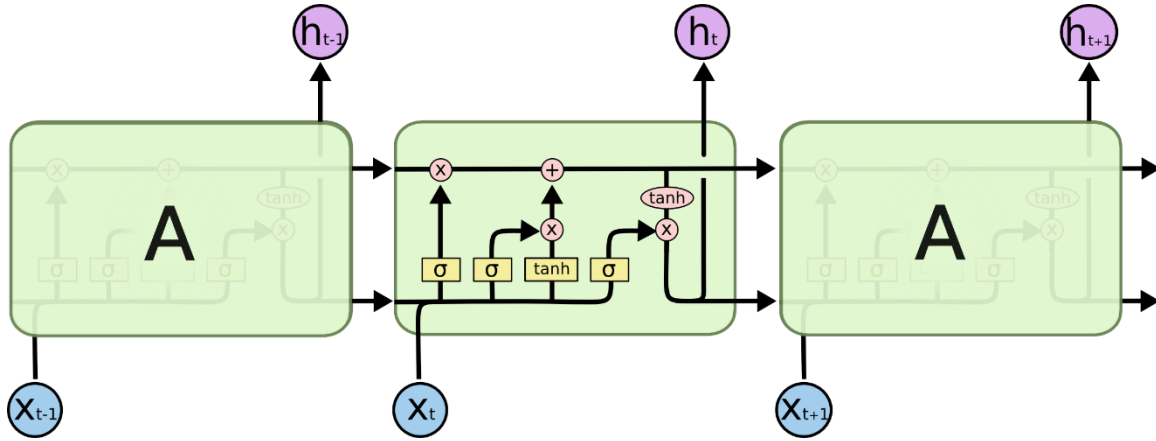
Time series analysis predicts the future values of the series based on data through the analysis of data sets <sup>[3]</sup>. Scholars have previously confirmed the feasibility and effectiveness of time series models in financial markets. The autoregressive–moving-average (ARMA) model and the autoregressive integrated moving average (ARIMA) model based on the ARMA model are important methods for studying time series <sup>[4,5]</sup>, and are adapted for short-term forecasting <sup>[6,7]</sup>. However, there are still some problems with time series models, real-world systems are usually nonlinear and time series data are mostly unstable. Therefore, the combination of artificial neural network model (ANN) and support vector machine (SVM) is the most used method, resulting in better prediction accuracy than traditional methods <sup>[8]</sup>. ANN cannot capture sequence information in the input data required to process sequence data and gradient explosion or disappearance <sup>[9]</sup> while SVM can solve multi-dimensional and nonlinear problems and avoid neural network structure selection and local minimum point problems to a certain extent <sup>[10]</sup>.

The use of machine learning methods to forecast the stock market has been extensively studied in literatures. Google Trends and an improved population-based sine and cosine algorithm (ISCA) have both been used by some researchers to enhance the performance of the optimized artificial neural network (ANN), leading to the conclusion that Google Trends data can aid in more precise stock price direction prediction <sup>[11]</sup>.

This paper adopts the LSTM model, which can pass data to each layer, which ensures the existence of short-term memory while training long-term memory. Moreover, LSTM also solves the vanishing gradient problem encountered by recurrent networks when dealing with long data sequences <sup>[12]</sup>.

## 3. Methodology

Hochreiter and Schmidhuber first suggested the LSTM in 1997, and it later gained enormous popularity, especially for use in solving issues involving time series prediction <sup>[12]</sup>. With the addition of three memory modules: input gate, output gate, and forget gate —LSTM primarily addresses the gradient disappearance issue that is common in classical RNN <sup>[13]</sup>. These three doors and a memory unit together form a memory block (the specific structure of the memory block is shown in **Figure 1**). The upper line inside the square is called the cell state and is used to control the transfer of information to the next moment <sup>[14]</sup>.



**Figure 1.** An LSTM's repeating module with four interconnected layers

In **Figure 1**, we can see that LSTM can add or remove information to the cell state, governed by structures called gates. Gates, which consist of a sigmoid neural network layer followed by a pointwise multiplication operation, are used to selectively let information through. The sigmoid layer outputs a number between 0 and 1, describing how much of each component should pass, and the larger the value, the more it will pass <sup>[15]</sup>.

- (1) The forget gate controls what information can pass through the sigmoid and will pass or partially pass according to the output of the previous moment, to achieve the effect of selective filtering.

$$F_t = \sigma(\omega_f * [h_{t-1}, x_t] + b_f)$$

- (2) An “input gate” layer that determines the values that will be used to update through sigmoid is added to a tanh layer used to generate new candidate values, and the candidate values are obtained to generate new information that needs to be updated (discard unnecessary information, add new information).

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

- (3) The last step is to get an initial output through the sigmoid layer, use tanh to scale the value between -1 and 1, and then multiply the output obtained by the sigmoid pair by pair to get the output of the model.

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o)$$

$$h_t = O_t * \tanh(C_t)$$

#### 4. Empirical analysis

This paper used Google stock data for the last 20 years, with data on open, close, high, low, adjusted close and trading volume. The Google stock data is from Yahoo Finance and records 3,810 records from August 19, 2004, to October 4, 2019. In this research, we used the adjusted closing price as the final price of Google Inc.'s daily stock.

##### 4.1. Data preprocessing

We used the data before January 1, 2019, as the training set of the data, and the data after January 1, 2019, as the validation set of the data. After we have dealt with the irrelevant features of the training set, the data is uniformly normalized to reduce the influence of too large dimensional gaps. After confirming the data

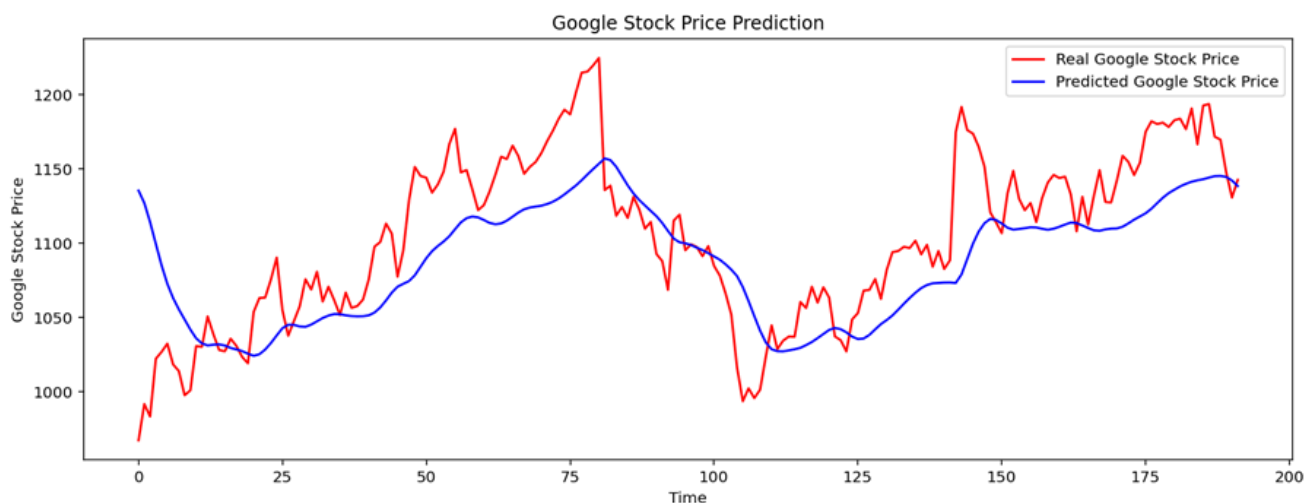
training set, it was then divided with a ratio of 60:1, and then converted to an array with NumPy to complete the data preprocessing.

## 4.2. Modeling

We selected TensorFlow as the deep learning framework for this modeling. When modeling the LSTM model, we set up a 3-layer network. The first layer was the LSTM layer (dimension; 60), and the second layer was the LSTM layer (dimension; 80), the third layer was the LSTM layer (dimension; 120), the dropout layer (dropout=0.2, used to prevent overfitting) was sandwiched between the three LSTM layers, and the fourth layer was the fully connected layer (The neuron number was 1, which was used to predict the future 1 Google stock price), the Adam optimizer was used to estimate the parameters, the learning rate adopted the LR decay method, and the maximum number of iterations was set to 50 times, and the mean square error regression loss was calculated to minimize the loss, until driven to 0.

## 4.3. Visualization

From the **Figure 2**, we can get a line chart with figsize= (14:5) as the ratio, time as the horizontal axis, and price as the vertical axis. The blue line shows the anticipated Google stock price, while the red line shows the actual Google stock price. The price mainly fluctuates around 1000-1200, and the duration is about 200 days.



**Figure 2.** Google stock price prediction

## 5. Conclusion

In this paper, we used the LSTM recurrent neural networks to extract feature value and analyze the stock data. The LSTM deep learning model we used this time, which combines the attention mechanism with depth and uses the gradient descent method to achieve a faster speed approximation, had better performance than the previous ARIMA, ANN and SVM models, and based on the algorithm, it solved the problems of easily falling into local extreme values and slow convergence speed. In general, the overall trend of image prediction is basically consistent with the actual trend, and it is also suitable for predicting long-term trends. Although there will be delayed prediction due to the time difference and small changes that cannot be noticed in a short period of time, the upward or downward trend progressively starts to match as time goes on, and the coincidence degree in the later stage is greatly enhanced.

While the accuracy rate was not very satisfactory, we found that it can still be improved, especially if the correct threshold was set to effectively exclude very low or very high yield sequences. This is useful

when selecting stocks for analysis. Additionally, we have taken into account fewer learning characteristics and have not taken into account the subjective impact of political policies, business trade conditions, or even social climate, which will be the direction we quantify in the future. Instead, the image fit can be improved by adding more learned features and optimizing the neural network weight matrix. Finally, based on our comparison of different neural networks and optimization algorithms, better models should be designed to improve prediction accuracy in the future.

### **Disclosure statement**

The author declares no conflict of interest.

### **Author contributions**

Z.T. conceived the idea of the study. Y.L. wrote the introduction and integrated various parts of the paper. T.Z. analyzed the data and wrote the rest of the paper.

### **References**

- [1] Li S, Li W, Cook C, et al., 2018, Independently Recurrent Neural Network (IndRNN): Building a Longer and Deeper RNN. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5457-5466
- [2] Staudemeyer RC, Morris ER, 2019, Understanding LSTM: A Tutorial into Long Short-term Memory Recurrent Neural Networks.
- [3] Mondal P, Shit L, Goswami S, 2014, Study of Effectiveness of Time Series Modeling (ARIMA) in Forecasting Stock Prices. *International Journal of Computer Science, Engineering and Applications*, 4(2): 13.
- [4] Chen S, Lan X, Hu Y, et al., 2014, The Time Series Forecasting: From the Aspect of Network.
- [5] Yuan Q, Shi Y, Hu G, et al., (eds) 2020, Computer Applications in Industry and Engineering. *Proceedings of 32nd International Conference on Computer Applications in Industry and Engineering*.
- [6] Ariyo AA, Adewumi AO, Ayo CK, 2016, Stock Price Prediction Using the ARIMA Model. *Proceedings of 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, 106–112.
- [7] Zhang G, Patuwo BE, Hu MY, Forecasting with Artificial Neural Networks: The State of the Art. *International Journal of Forecasting*, 14(1): 35–62.
- [8] Vui CS, Soon GK, On CK, et al., 2013, A Review of Stock Market Prediction with Artificial Neural Network (ANN). *Proceedings of IEEE International Conference on Control System, Computing and Engineering*, 477-482, <https://www.doi.org/10.1109/ICCSCE.2013.6720012>
- [9] Liang X, Ge Z, Sun L, et al., 2019, LSTM with Wavelet Transform-Based Data Preprocessing for Stock Price Prediction. *Mathematical Problems in Engineering*.
- [10] Sands TM, Tayal D, Morris ME, et al., 2015, Robust Stock Value Prediction Using Support Vector Machines with Particle Swarm Optimization. *Proceedings of IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 3327-3331.
- [11] Hu H, Tang L, Zhang S, et al., 2018, Predicting the Direction of Stock Markets Using Optimized Neural Networks with Google Trends. *Neurocomputing*, 285: 188-195.
- [12] Hochreiter S, Schmidhuber J, 1997, Long Short-term Memory. *Neural Comput.*, 9(8): 1735–1780.

<http://dx.doi.org/10.1162/neco.1997.9.8.1735>

- [13] Ma Q, 2020, Comparison of ARIMA, ANN and LSTM for Stock Price Prediction. E3S Web of Conferences, 218: 01026.
- [14] Lu W, Li J, Li Y, et al., 2020, A CNN-LSTM-based Model to Forecast Stock Prices. Complexity, 2020: 6622927
- [15] Ghosh A, Bose S, Maji G, et al., 2019, Stock Price Prediction Using LSTM on Indian Share Market. Proceedings of 32nd International Conference on Computer Applications in Industry and Engineering, 101–110.

**Publisher's note**

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.