https://doi.org/10.62836/iaet.v2i1.162

Global Science Publishing

IAET

Innovations in Applied Engineering and Technology

https://ojs.sgsci.org/journals/iaet

Article

Stock Market Analysis and Prediction Using LSTM: A Case Study on Technology Stocks

Zhenglin Li [1,*], Hanyi Yu [2], Jinxin Xu [3], Jihang Liu [4] and Yuhong Mo [5]

[1] *Department of Computer Science and Engineering, Texas A&M University;College Station, TX 77843, USA*

[2] *Department of Computer Science, University of Southern California, Los Angeles, CA 90007, USA; hanyiyu@usc.edu*

[3] *Department of Cox Business School, Southern Methodist University, Dallas, TX 75205, USA; jensenjxx@gmail.com*

[4] *School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104, USA; 550953533@qq.com*

[5] *Deparment of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA; yuhongmo@cmu.edu*

Abstract: This research explores the application of Long Short-Term Memory (LSTM) networks for stock market analysis and prediction, focusing on four major technology stocks: Apple Inc. (AAPL), Google LLC (GOOG), Microsoft Corporation (MSFT), and Amazon. com Inc. (AMZN). Historical stock price data from Yahoo Finance spanning from January 1, 2012, to the present is utilized. The study aims to develop and evaluate an LSTM-based prediction model for forecasting future stock prices. The LSTM model consists of two LSTM layers with 128 and 64 units, respectively, followed by two dense layers. The model is trained using the Adam optimizer and mean squared error (MSE) loss function. Evaluation of the model is done using the root mean squared error (RMSE) metric. The results demonstrate the potential of LSTM models in capturing complex patterns in stock price movements and making reasonably accurate predictions.

Keywords: stock market analysis; stock price prediction; long short-term memory (lstm); machine learning; financial analysis; time series data

1. Introduction

In recent years, the stock market has emerged as a critical component of the global economy, attracting significant attention from investors, traders, and researchers [1]. The ability to accurately analyze and predict stock prices is of paramount importance, as it can lead to substantial financial gains and inform strategic investment decisions [2]. Traditional methods of stock analysis, such as fundamental and technical analysis, have been widely used but often fall short in capturing the complex, non-linear patterns inherent in stock price movements [3].

With the advent of machine learning and artificial intelligence, new avenues have opened up for stock market analysis. Among various machine learning techniques, Long Short-Term Memory (LSTM) networks, a

Received: 19 September 2023; Accepted: 10 November 2023.

* Corresponding: Zhenglin Li (zhenglin_li@tamu.edu)

type of recurrent neural network, have shown great promise in modeling sequential data and capturing temporal dependencies. This research aims to leverage the power of LSTM networks to analyze and predict the stock prices of prominent technology companies, namely Apple (AAPL), Google (GOOG), Microsoft (MSFT), and Amazon (AMZN).

The objectives of this study are twofold: first, to conduct a comprehensive analysis of the historical stock prices of these technology giants, and second, to develop and evaluate an LSTM-based prediction model that can forecast future stock prices with a high degree of accuracy [4,5]. By achieving these objectives, this research seeks to contribute to the growing body of knowledge in the field of financial analysis and provide valuable insights for investors and policymakers.

2. Methodology

2.1. Data Collection and Preprocessing

The study utilizes historical stock price data sourced from Yahoo Finance, accessed through the finance library in Python. The data includes daily adjusted closing prices and trading volumes for four major technology stocks: Apple Inc. (AAPL), Google LLC (GOOG), Microsoft Corporation (MSFT), and Amazon. com, Inc. (AMZN), covering a period from January 1, 2012, to the present.

Data preprocessing involves several steps to prepare the data for analysis and modeling:

Filtering: The dataset is filtered to include only the 'Close' price column, as the focus of the study is on predicting closing stock prices [6].

Scaling: The closing prices are scaled using the MinMaxScaler from the sklearn library to normalize the values between 0 and 1. This step is crucial for the efficient training of the LSTM model [7,8].

Training and Test Set Split: The dataset is divided into training and test sets, with 95% of the data used for training and the remaining 5% for testing. This split ensures that the model can be evaluated on unseen data [9].

Sequence Creation: The training data is transformed into sequences of 60 days each, which serve as the input features (X) for the LSTM model. The corresponding target (Y) for each sequence is the closing price on the 61st day.

Figure 1 and Figure 2 show the visualization of data we use.

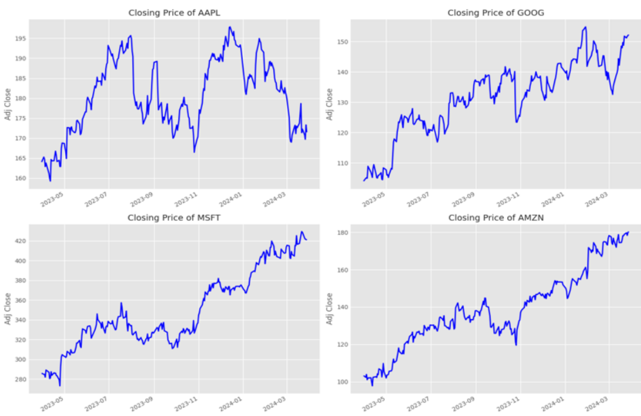

Figure 1. Visualization of Historical Stock Prices for AAPL, GOOG, MSFT, and AMZN (2012-Present).

Figure 2. Preprocessed Data for LSTM Model Training: Scaled and Sequenced Closing Prices.

2.2. LSTM Model Architecture.

The study employs a Long Short-Term Memory (LSTM) neural network model for stock price prediction. The LSTM model is chosen for its ability to capture temporal dependencies and handle time-series data effectively [10]. Figure 3 shows the visualization of the architecture. The model architecture comprises the following layers.

Input Layer: The input layer is configured to accept sequences of 60 days' worth of scaled closing prices.

LSTM Layers: Two LSTM layers are used, with the first layer having 128 units and returning sequences to feed into the subsequent layer. The second LSTM layer has 64 units and does not return sequences.

Dense Layers: Following the LSTM layers, two dense layers with 25 and 1 unit(s) respectively are added. The final dense layer outputs the predicted closing price.

Compilation: The model is compiled using the Adam optimizer and mean squared error (MSE) as the loss function.

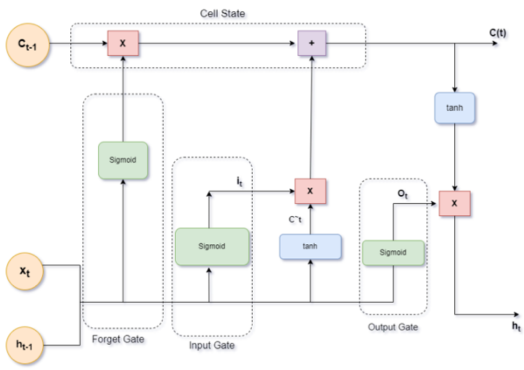

Figure 3. Architecture Visualization of LSTM Neural Network Model for Stock Price Prediction.

2.3. LSTM cell.

The whole computation can be defined by a series of equations as follows:

$$i_t = \sigma\left(W^i H + b^i\right) \quad (1)$$

$$f_t = \sigma\left(W^f H + b^f\right) \quad (2)$$

$$o_t = \sigma\left(W^o H + b^o\right) \quad (3)$$

$$c_t = \tanh\left(W^c H + b^c\right) \quad (4)$$

$$m_t = f_t \odot m_{t-1} + i_t \quad (5)$$

$$h_t = \tanh\left(o_t \odot m_t\right) \quad (6)$$

2.4. Model Training and Evaluation

The LSTM model is trained on the prepared training dataset with a batch size of 1 and 1 epoch. The training process involves adjusting the model's weights to minimize the loss function, thereby improving the accuracy of predictions [11].

Model evaluation is conducted on the test dataset, where the model's predictions are compared against the actual closing prices. The root mean squared error (RMSE) is calculated to quantify the model's prediction accuracy [12]. Lower RMSE values indicate better model performance, where n is the number of observations, y_i is the actual value of the i-th observation, and $\widehat{y}_i$ is the predicted value of the i-th observation.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2} \quad (7)$$

3. Results and Discussion

In this study, we applied a Long Short-Term Memory (LSTM) model to predict the closing stock prices of four major technology companies: Apple (AAPL), Google (GOOG), Microsoft (MSFT), and Amazon (AMZN). The data was sourced from Yahoo Finance and spanned from January 1, 2012, to the current date. The dataset was split into training (95%) and testing (5%) sets, with the LSTM model trained on the former and evaluated on the latter.

The LSTM model consisted of two LSTM layers with 128 and 64 units, respectively, followed by two dense layers with 25 and 1 units. The model was compiled using the Adam optimizer and mean squared error loss function. It was trained for 1 epoch with a batch size of 1.

The model's performance was evaluated using the root mean square error (RMSE) metric is achieved by 18.89. The RMSE for the test set was calculated to assess the accuracy of the predictions. The results showed that the LSTM model was able to capture the trends in the stock prices of the analyzed companies with a reasonable degree of accuracy. Figure 4 shows the visualization.

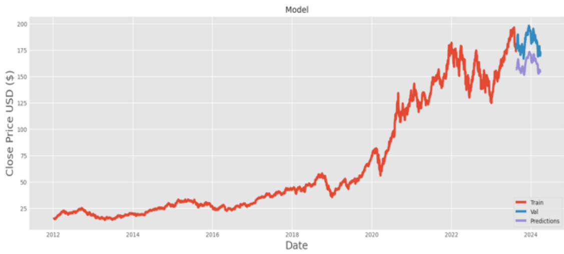

Figure 4. Comparison between Predicted and Actual Closing Stock Prices with LSTM Model.

The results of this study indicate that LSTM models are a promising tool for stock market analysis and prediction. The model was able to predict the closing stock prices of major technology companies with a satisfactory level of accuracy. This finding is consistent with previous research that has highlighted the effectiveness of LSTM models in capturing the temporal dependencies in time-series data, such as stock prices.

However, there are several limitations to this study that should be addressed in future research. Firstly, the model was trained on a relatively small dataset, which may limit its ability to generalize to different time periods or market conditions. Additionally, the model was evaluated using only one metric (RMSE), and further analysis using additional metrics could provide a more comprehensive assessment of its performance.

Moreover, the stock market is influenced by a wide range of factors, including economic indicators, company news, and market sentiment. Incorporating these factors into the model could potentially improve its predictive accuracy. Future research could also explore the use of more complex LSTM architectures or hybrid models that combine LSTM with other machine learning techniques.

4. Conclusion

In this study, we explored the application of Long Short-Term Memory (LSTM) networks for stock market analysis and prediction, focusing on four major technology stocks: Apple Inc. (AAPL), Google LLC (GOOG), Microsoft Corporation (MSFT), and Amazon. com Inc. (AMZN). Our research demonstrated the potential of LSTM models in capturing complex patterns and trends in stock price movements, leading to reasonably accurate predictions of future prices.

The findings of this study hold significant implications for both academia and the finance industry. For researchers and data scientists, the successful application of LSTM models in stock market prediction offers a promising avenue for further exploration and refinement of machine learning techniques in financial analysis. For investors and traders, the insights gained from our model could aid in making more informed decisions, potentially leading to improved investment strategies and portfolio management.

In conclusion, our study underscores the potential of LSTM networks in stock market analysis and prediction, offering valuable insights for both researchers and practitioners in the field of finance. As machine learning continues to evolve, its application in financial markets holds the promise of unlocking new opportunities for understanding and predicting stock price movements.

Funding

Not applicable.

Author Contributions

Conceptualization, Z.L.; writ-ing—original draft preparation and writing—review and editing, Z.L., H.Y., J.X., J.L. and Y.M. All of the authors read and agreed to the published the final manuscript.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Conflicts of Interest

The authors declare no conflict of interest.

References

1 Liu S, Liao G, Ding Y. Stock Transaction Prediction Modeling and Analysis Based on LSTM. In Proceedings of the 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, China, 18–22 August 2018.

2 Borovkova S, Tsiamas I. An Ensemble of LSTM Neural Networks for High - Frequency Stock Market Classification. *Journal of Forecasting* 2019; **38**(**6**): 600–619.

3 Baek Y, Kim HY. ModAugNet: A New Forecasting Framework for Stock Market Index Value With an Overfitting Prevention LSTM Module and a Prediction LSTM Module. *Expert Systems with Applications* 2018; **113**: 457–480.

4 Eapen J, Bein D, Verma A. Novel Deep Learning Model With CNN and Bi-Directional LSTM for Improved Stock Market Index Prediction. In Proceedings of the 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 7–9 January 2019.

5 Selvin S, Vinayakumar R, Gopalakrishnan EA, Menon VK, Soman KP. Stock Price Prediction Using LSTM, RNN and CNN-Sliding Window Model. In Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 13–16 September 2017.

6 Nelson DM, Pereira AC, de Oliveira RA. Stock Market' S Price Movement Prediction With LSTM Neural Networks. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017.

7 Sherstinsky A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Physica D: Nonlinear Phenomena* 2020; **404**: 132306.

8 Nabipour M, Nayyeri P, Jabani H, Mosavi A, Salwana E, Shahab S. Deep Learning for Stock Market Prediction. *Entropy* 2020; **22**(**8**): 840.

9 Bhandari HN, Rimal B, Pokhrel NR, Rimal R, Dahal KR, Khatri RKC. Predicting Stock Market Index Using LSTM. *Machine Learning with Applications* 2020; **9**: 100320.

10 Althelaya KA, El-Alfy E-SM, Mohammed S. Evaluation of Bidirectional LSTM for Short-and Long-Term Stock Market Prediction. In Proceedings of the 2018 9th International Conference on Information and

Communication Systems (ICICS), Irbid, Jordan, 3–5 April 2018.

11 Moghar A, Hamiche M. Stock Market Prediction Using LSTM Recurrent Neural Network. *Procedia Computer Science* 2020; **170**: 1168–1173.

12 Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J. LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*2017; **28**(**10**): 2222–2232.