



(RESEARCH ARTICLE)



Stock market prediction using LSTM

Yasmin Akter Bipasha ^{1, 2, *}

¹ College of Business, Westcliff University, Irvine, CA 92614, USA.

² Bangladesh University of Professionals, Mirpur Cantonment, Dhaka-1216, Bangladesh.

International Journal of Science and Research Archive, 2024, 12(02), 3146-3153

Publication history: Received on 10 July 2024; revised on 22 August 2024; accepted on 28 August 2024

Article DOI: <https://doi.org/10.30574/ijrsra.2024.12.2.1542>

Abstract

The stock market prediction patterns are seen as an important activity and it is more effective. Hence, stock prices will lead to lucrative profits from sound taking decisions. Because of the stagnant and noisy data, stock market-related forecasts are a major challenge for investors. Therefore, forecasting the stock market is a major challenge for investors to use their money to make more profit. This presents a stock market prediction based on Long Short-Term Memory (LSTM), designed to handle time series data and long-term dependencies. Historical stock price data for major technology companies, including Apple, Amazon, Google, and Microsoft, is collected from Yahoo Finance. The dataset contains key attributes such as open price, close price, high price, low price, and trading volume. These features are used to analyze stock behavior and predict future price movements. LSTM-based models can capture temporal relationships in stock market data more effectively than traditional statistical methods.

Keywords: Stock market prediction; Long Short-Term Memory (LSTM); Time series analysis; Financial forecasting; Yahoo Finance data

1. Introduction

A correct prediction of stocks can lead to huge profits for the seller and the broker. Frequently, it is brought out that prediction is chaotic rather than random, which means it can be predicted by carefully analyzing the history of respective stock market. Machine learning is an efficient way to represent such processes. It predicts a market value close to the tangible value, thereby increasing the accuracy. Introduction of machine learning to the area of stock prediction has appealed to many researches because of its efficient and accurate measurements [1] [2]. The vital part of machine learning is the dataset used. The dataset should be as concrete as possible because a little change in the data can perpetuate massive changes in the outcome. In this project, supervised machine learning is employed on a dataset obtained from Yahoo Finance. This dataset comprises of following five variables: open, close, low, high and volume. Open, close, low and high are different bid prices for the stock at separate times with nearly direct names. The volume is the number of shares that passed from one owner to another during the time period. The model is then tested on the test data.

Investors use two main methods, fundamental analysis and technical analysis, to make decisions in the stock market with the aim of earning higher returns and reducing risk. Research on stock market prediction has become an important topic for investors, analysts, and researchers. Studying stock market movements and price behavior is difficult because financial markets contain noise and uncertainty, as explained in studies on market efficiency and behavioral finance [3]. Stock prices change due to many factors, including company earnings reports, economic conditions, and market news.

Stock market indices are created based on market capitalization and represent the performance of a group of selected stocks [4]. Because financial markets change over time, predicting stock prices with high accuracy remains a challenging

* Corresponding author: Yasmin Akter Bipasha

task. As a result, researchers and analysts have worked on understanding stock market behavior using different statistical and computational methods. Traditional approaches such as autoregressive integrated moving average (ARIMA) models and clustering techniques have been used for stock market prediction.

Machine learning algorithms have become important tools for studying complex financial data and making predictions. Similar machine learning models are also used in other fields such as healthcare systems and Internet of Medical Things applications, which shows the flexibility of these techniques across different domains [5]. Stock market prediction has remained an active research area because it offers the possibility of financial gain; however, stock price data often follows patterns like a random walk, which makes prediction difficult [23]. Studies show that machine learning methods can handle large datasets and identify patterns that are not clear using traditional techniques [6], [21].

New learning methods such as federated learning focus on data privacy and shared model training across different locations, and they offer future opportunities for stock market prediction where data security is important [20, 25]. By using machine learning techniques, investors and traders can improve decision making, find price trends, and control financial risk, which can support better performance in the stock market. ML is used in healthcare rapidly [24, 26, 27, 28].

Time Series data is a series of data points indexed in time order. Time series data is everywhere, so manipulating them is important for any data analyst or data scientist. In this work, stock market data, particularly some technology stocks (Apple, Amazon, Google, and Microsoft) will be discussed. Future stock prices will be predicted through a Long Short Term Memory (LSTM) method. The following questions will be answered:

- What was the change in price of stock over time?
- What was the daily return of the stock on average?
- What was the moving average of the various stocks?
- How can we predict future stock behavior? (Predicting the closing price stock price of APPLE inc using LSTM)

2. Related Work

When it comes to stock price prediction, there are essentially two techniques. For projecting the price of a stock, we need to look at its historical performance, its debt, the percentages of shares held by the company's promoters, its business competitors, and one of the most essential factors is the company's management. Although a model has a high probability of properly predicting the price of a stock, variables such as government intervention in the business sector might cause price fluctuations. For example, if a corporation has a monopoly over the business in its industry, the company will be a profitable investment regardless of the market situation. However, if the government makes a few modifications to the sector's operations, the company's stock price may see significant variations. Most studies have explored a relationship between the volume of shares traded and the price of the share in attempt to anticipate the price of the share. M. F. M. Osborne investigated the link between the price of a share and the total number of shares traded at the moment (Volume of share) in 1959. Then he built a model in which he attempted to show that there is a useful relationship between the absolute value of price change between two points in time. Some scholars have attempted to forecast market attitudes by examining the number of shares traded in the market at a certain time [8]. D. Godfrey, Clive W. J. Granger, and Oskar Morgenstern discovered a link between the number of shares traded and the difference between the opening and closing price of the stock on a given day in 1964 [9]. In 1970, Peter K. Clark discovered that the square of change in price and the volume of shares traded in the market have a positive relationship [10]. Then, for a few months, I. G. Morgan conducted an experiment in which he monitored data linked to the change in price of the share for four consecutive days and discovered that there is a relationship between price and volume [11]. After that, Anice Moulton Wester field discovered a similar beneficial link with approximately 300 stocks [12]. Based on data from several hourly market indexes, P Jain and GH Joh discovered a favorable link between price and volume [13]. Guo Liang and Zhou Weixing conducted an infinitesimal level price volume relationship factual investigation among a variety of Chinese stock market equities utilizing high frequency data last year. Then they noticed a link between the difference in share prices and the number of shares exchanged that day. Authors of [17]-[19], have used the machine learning technique to solve the real-life unsolved issue. Then have also used a non-linear convex function curve to represent the link between price and volume [14]-[16]. When the number of shares traded exceeds the average, it is regarded as a negative indication because the price of a share falls when there is an abundance of shares available, whereas the price of a share rises when there are fewer shares available to meet market demand. In simple business logic, if individuals are convinced that the price of a product will rise in the future, they will not sell it; but, if they believe that the price of anything will fall, they will rush to sell it as soon as possible to avoid selling it at a lower price than the present price.

However, this might cause issues. For example, if a firm requires funding, it can sell a large portion of its stock to get funds and then reinvest in a business that would generate greater profit. As a result, this strategy of estimating market mood based on volume to forecast future prices is not always reliable.

3. Results

The first step is to get the data and load it to memory. stock data is collected from the Yahoo Finance website. Yahoo Finance is a rich resource of financial market data and tools to find compelling investments. To get the data from Yahoo Finance, yfinance library is used which offers a threaded and Pythonic way to download market data from Yahoo.

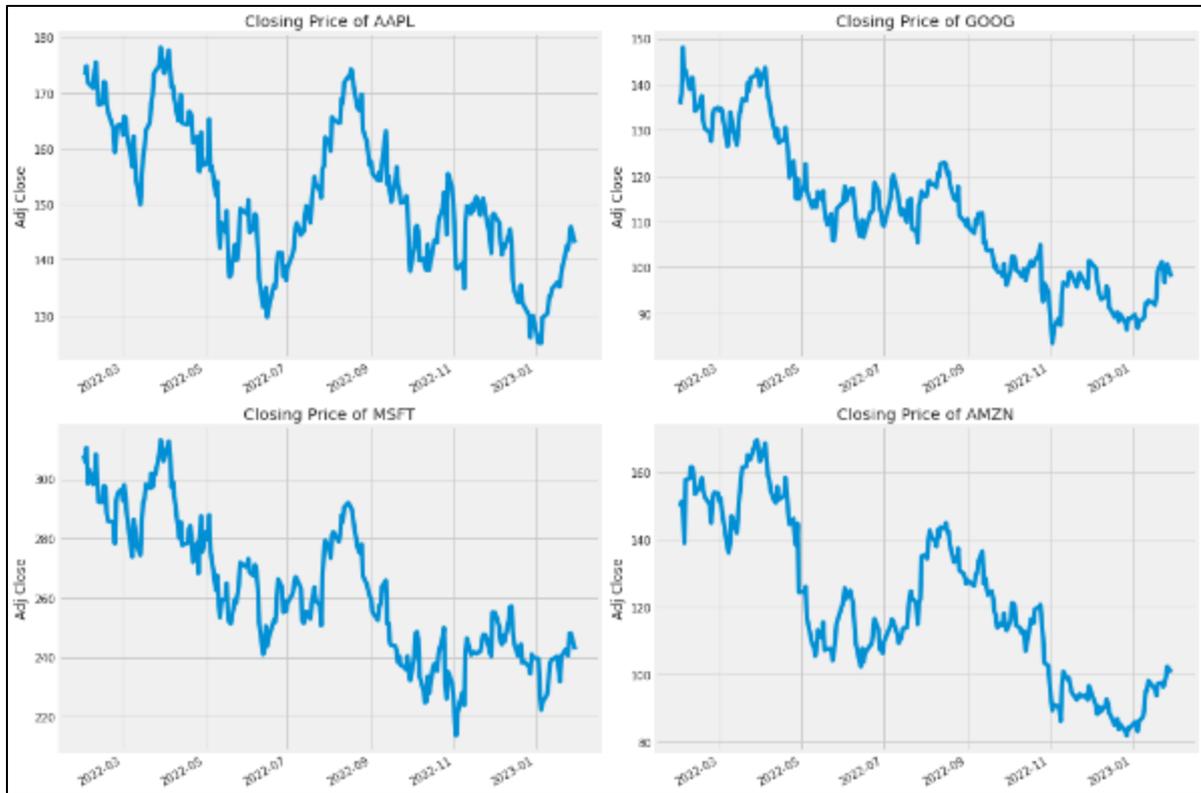


Figure 1 Closing Price

LSTM is the advanced version of Recurrent-NeuralNetworks (RNN) where the information belonging to previous state persists. These are different from RNNs as they involve long term dependencies and RNNs works on finding the relationship between the recent and the current information. This indicates that the interval of information is relatively smaller than that to LSTM. The main purpose behind using this model in stock market prediction is that the predictions depends on large amounts of data and are generally dependent on the long term history of the market. So LSTM regulates error by giving an aid to the RNNs through retaining information for older stages making the prediction more accurate [7]. Thus proving itself as much more reliable compared to other methods.

The LSTM consists of a remembering cell, input gate, output gate and a forget gate. The cell remembers the value for long term propagation and the gates regulate them.

In this work, a sequential model is constructed by stacking two LSTM layers. The first LSTM layer contains 128 units and is configured with `return_sequences=True`, allowing it to pass the full sequence output to the next LSTM layer. The input to this layer is shaped as `(time steps, 1)`, indicating a univariate time-series input. The second LSTM layer consists of 64 units and outputs only the final hidden state, as `return_sequences` is set to `False`.

A fully connected dense layer with 25 neurons is added to further transform the learned temporal features. Finally, an output dense layer with a single neuron is used to generate the final prediction value. The model is optimized using the Adam optimizer and trained by minimizing the mean squared error (MSE) loss function, which is suitable for regression-based prediction tasks.

The closing price is the last price at which the stock is traded during the regular trading day. A stock's closing price is the standard benchmark used by investors to track its performance over time.

3.1. Volume of Sales

Volume is the amount of an asset or security that changes hands over some period of time, often over the course of a day. For instance, the stock trading volume would refer to the number of shares of security traded between its daily open and close. Trading volume, and changes to volume over the course of time, are important inputs for technical traders.

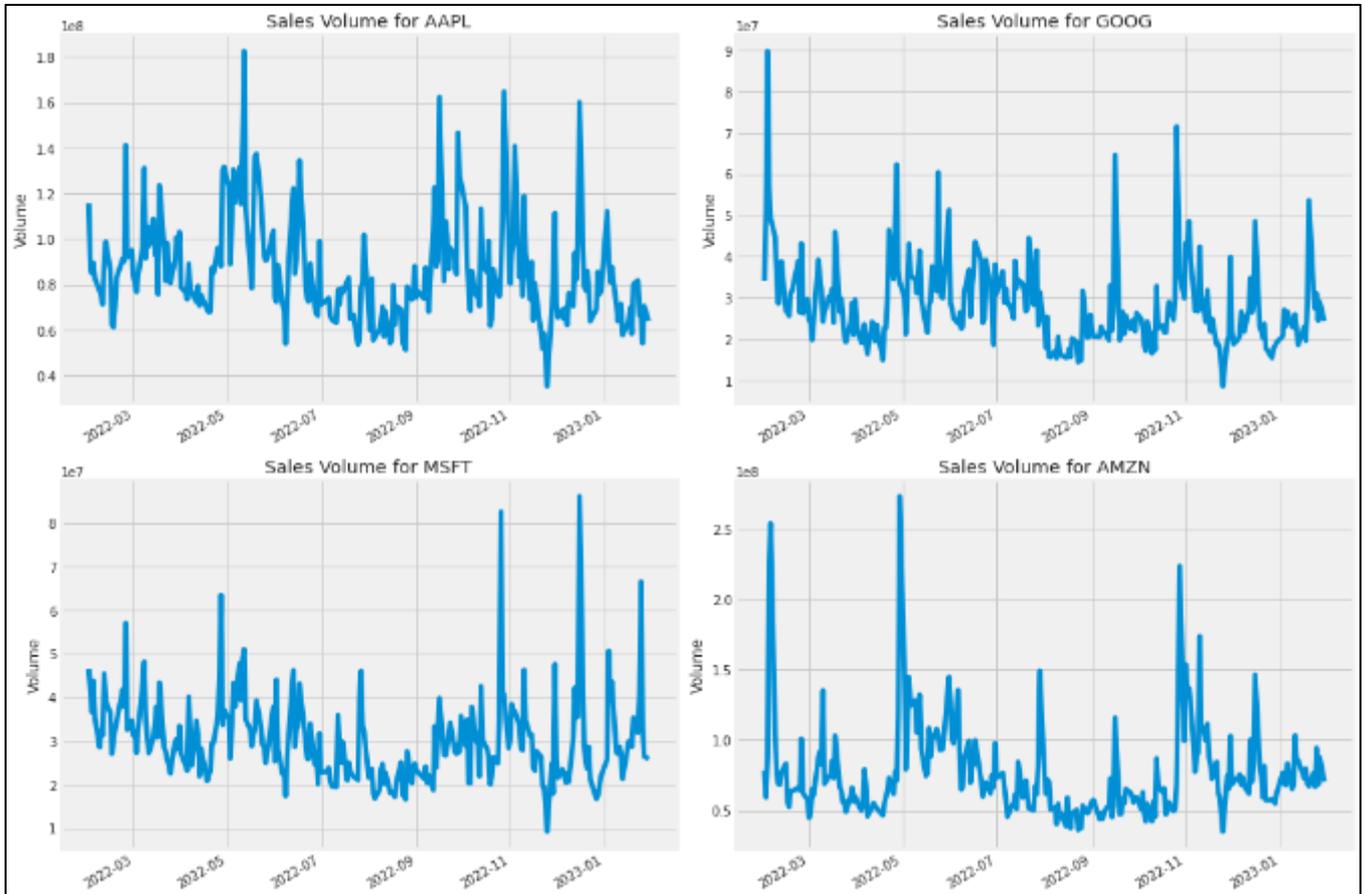


Figure 2 Volume of Sales

3.2. Moving average of the various stocks

The moving average (MA) is a simple technical analysis tool that smooths out price data by creating a constantly updated average price. The average is taken over a specific period of time, like 10 days, 20 minutes, 30 weeks, or any time period the trader chooses. Figure 3 shows the moving average of the various stocks.

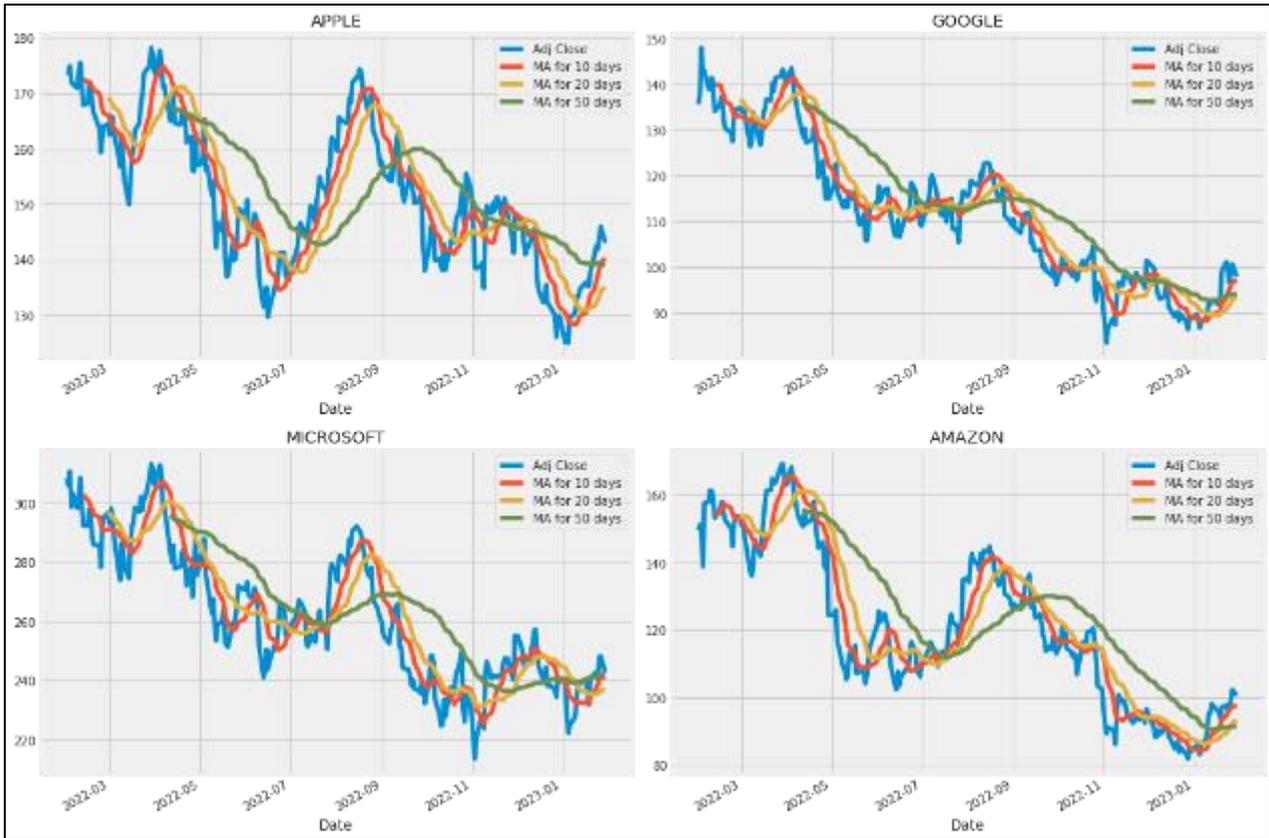


Figure 3 Moving average of the various stocks

3.3. Daily return of the stock on average

Figure 4 shows the daily stock returns for four companies: Apple, Google, Microsoft, and Amazon. Each plot displays how the stock return changes from one trading day to the next over a given time period. The horizontal axis represents the date, while the vertical axis represents the daily return value.

In all four plots, most daily return values are close to zero. This means that on many days, stock prices change by a small amount. Both positive and negative values appear, which shows that stock prices move up and down from day to day. This behavior reflects the uncertain nature of the stock market.

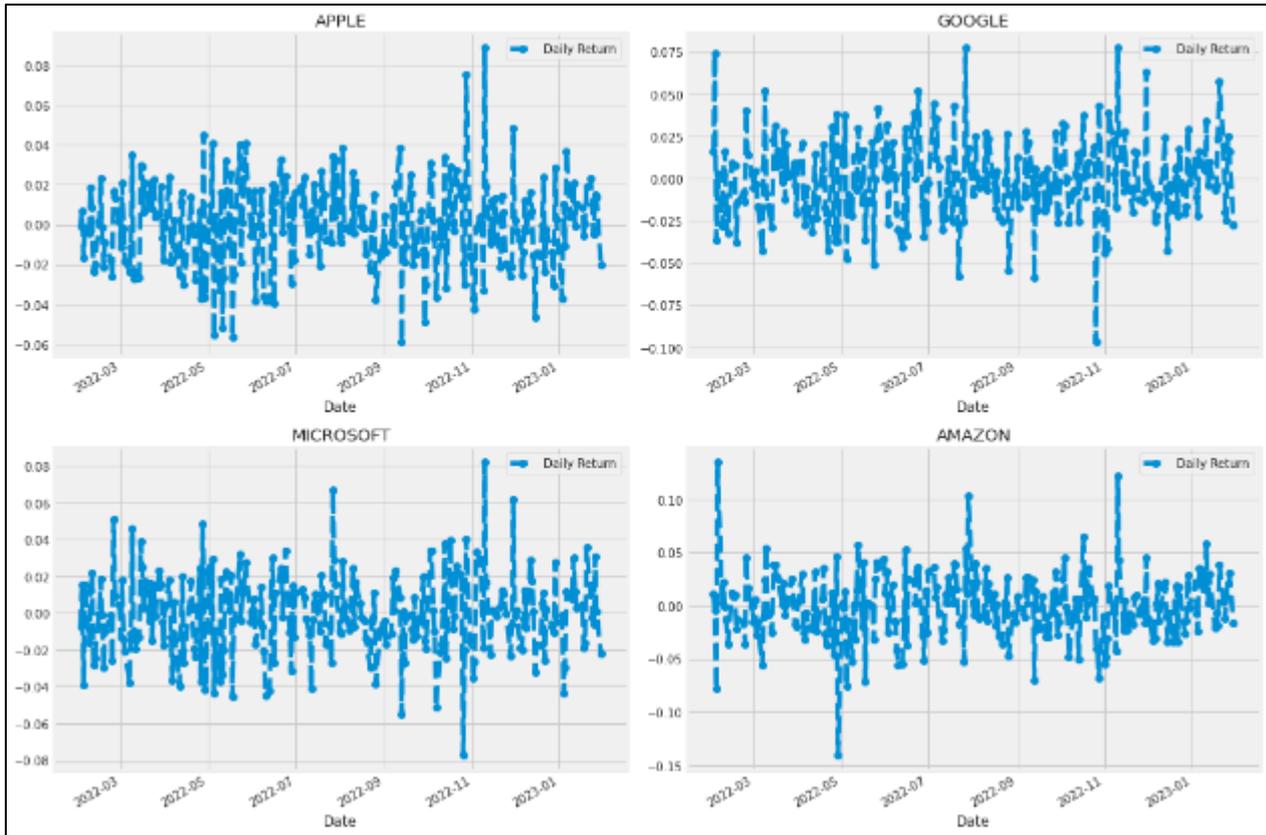


Figure 4 Daily return of stock on average

4. Predicting the closing price stock price of APPLE

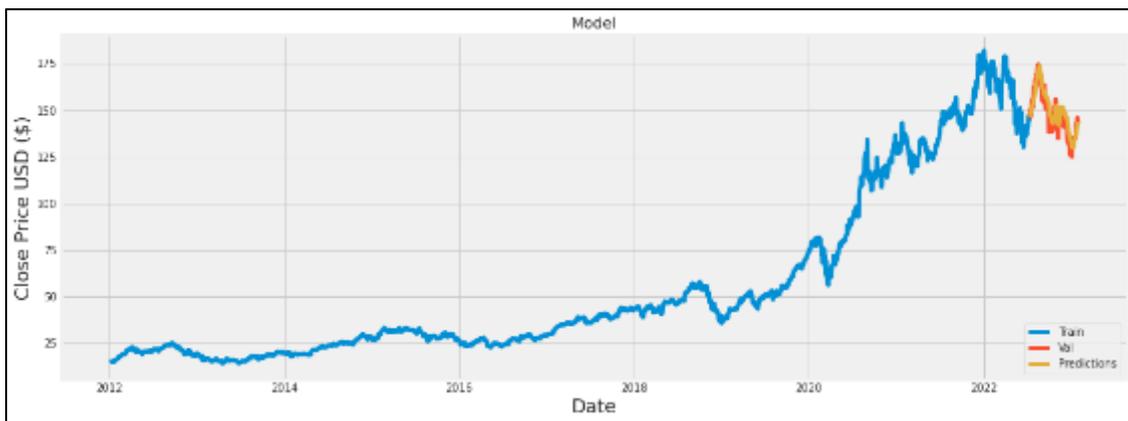


Figure 5 Stock Price prediction using LSTM

Figure 5 shows Predicting the closing price stock price of APPLE inc suing LSTM. Table 1 shows the Valid and predicted prices. RMSE is 4.98294.

Table 1 Valid and predicted prices

Date	Close	Predictions
2022-07-13 00:00:00-04:00	145.490005	146.457565
2022-07-14 00:00:00-04:00	148.470001	146.872879
2022-07-15 00:00:00-04:00	150.169998	147.586197
2022-07-18 00:00:00-04:00	147.070007	148.572937
2022-07-19 00:00:00-04:00	151.000000	148.995255
2023-01-24 00:00:00-05:00	142.529999	138.565536
2023-01-25 00:00:00-05:00	141.860001	140.022110
2023-01-26 00:00:00-05:00	143.960007	141.225128
2023-01-27 00:00:00-05:00	145.929993	142.469315
2023-01-30 00:00:00-05:00	143.000000	143.833130

5. Conclusion

This paper is an attempt to determine the future prices of the stocks of a company with greater accuracy and reliability using machine learning techniques. The primary contribution of the researchers being the application of the novel LSTM Model as a means of determining the stock prices. In the future, the accuracy of the stock market prediction system can be further improved by utilizing a much bigger dataset than the one being utilized currently. Furthermore, other emerging models of Machine Learning could also be studied to check for the accuracy rate resulted by them. Sentiment analysis though Machine Learning on how news affects the stock prices of a company is also a very promising area. Other deep learning based models can also be used for prediction purposes.

References

- [1] M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.
- [2] K. Raza, "Prediction of Stock Market performance by using machine learning techniques," 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, 2017, pp. 1-1
- [3] Bipasha, Y. A. (2022). Market efficiency, anomalies and behavioral finance: A review of theories and empirical evidence. *World Journal of Advanced Research and Reviews*, 15(02), 827-839.
- [4] Bipasha, Y. A. (2023). Blockchain technology in supply chain management: transparency, security, and efficiency challenges. *International Journal of Science and Research Archive*, 10(01), 1186-1196.
- [5] Mahmud, T. (2023). Applications for the Internet of medical things. *International Journal of Science and Research Archive*, 10(02), 1247-1254.
- [6] Mahmud, T. (2022). ML-driven resource management in cloud computing. *World Journal of Advanced Research and Reviews*, 16(03), 1230-1238.
- [7] S. Liu, G. Liao and Y. Ding, "Stock transaction prediction modelling and analysis based on LSTM," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 2787-2790
- [8] M. F. M. Osborne, "Brownian Motion in the Stock Market", *Operations Research*, 1959.
- [9] M. D. Godfrey, C. W. J. Granger and O. Morgenstern, "The Random-Walk Hypothesis of Stock Market Behavior", *Kyklos*, February 1964.
- [10] P. K. Clark, "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", vol. 41, no. 1, pp. 135-155, January 1973.

- [11] I. G. Morgan, "Stock Prices and Heteroscedasticity", vol. 49, no. 4, pp. 496-508, October 1976.
- [12] J. M. Westerfield, "An examination of foreign exchange risk under fixed and floating rate regimes", *Journal of International Economics*, vol. 7, no. 2, pp. 181-200, May 1977.
- [13] P. C. Jain and G.-H. Joh, "The Dependence between Hourly Prices and Trading Volume", vol. 23, no. 3, pp. 269-283, September 1988.
- [14] R. Mahanty, S. Mahapatra, A. Nayak and A. Chinmay, "Comparative Study of Various Image Captioning Models", *International Conference on Recent Advances in Energy-efficient Computing and Communication (ICRAECC)*, 13 February 2020.
- [15] A. Mehra, P. Tripathy, A. Faridi and A. Chinmay, "Ensemble Learning Approach to Improve Existing Models", *International Journal of Innovative Science and Research Technology*, vol. 4, no. 12, December 2019.
- [16] S. Mishra, N. Sethi and A. Chinmay, "Various Data Skewness Methods in the Hadoop Environment", *International Conference on Recent Advances in Energy-efficient Computing and Communication (ICRAECC)*, 13 February 2020.
- [17] A. Srivastava, A. Khare, P. Satapathy and A. Chinmay, "Investigating Various Cryptographic Techniques Used in Cloud Computing", *Advances in Data Science and Management, Lecture Notes on Data Engineering and Communications Technologies*, vol. 37, pp. 263-272, 14 January 2020.
- [18] P. Rai, A. Prasad, S. M. Reddy and A. Chinmay, "Evolution of Optical Storage in Computer Memory", *Advances in Data Science and Management, Lecture Notes on Data Engineering and Communications Technologies*, vol. 37, pp. 489-495, 14 January 2020.
- [19] N. Mohapatra, K. Shreya and A. Chinmay, "Optimization of the random forest algorithm", *Advances in Data Science and Management, Lecture Notes on Data Engineering and Communications Technologies*, vol. 37, pp. 201-208, 14 January 2020.
- [20] Hossain, M. B., & Rahman, R. (2022). Federated learning: Challenges and future work. *World Journal of Advanced Research and Reviews*, 15(02), 850-862.
- [21] MD, B. H., & KHANDOKER, H. (2022). Machine Learning approaches in IDS. *INTERNATIONAL JOURNAL*, 7(2), 706-715.
- [22] Rahaman, Md Zahedur, et al. "Assessment of thrombolytic, antioxidant and analgesic properties of a medicinal plant of Asteraceae family growing in Bangladesh." *Discovery Phytomedicine* 7.1 (2020): 47-52.
- [23] Ahammed, M.S., Khanom, M.M., Kobir, M.E., Akhter, S., Hossain, M.S., Begum, S., Bayen, S., Islam, M.R., Mondal, K.K., Mottakin, M., Mahmud, M.K. 2020. A Study on Hevea Brasiliensis for evaluation of phytochemical and pharmacological properties in Swiss Albino Mice. *Discovery Phytomedicine* 7(2): 72-75. DOI: 10.15562/phytomedicine.2019.122.
- [24] Das, D., Billah, M. M., Nath, A. D., Sharif, N. B., & Mondal, K. K. (2022). Breast Cancer Classification using LGBM and SVM. *International Journal of Science and Research Archive*.
- [25] Hossain, M. B., Rahman, R., & Hoque, K. (2021). Feature-Driven Supervised Learning for Detecting DDoS Attack. *International Journal of Science and Research Archive*, 4(01), 393-402.
- [26] Md Masum Billah, Amit Deb Nath, Denesh Das, Tanvir Mahmud and Rashedur Rahman, "Skin Cancer Classification using NasNet", *World Journal of Advanced Research and Reviews*, 2023, 19(01), 1652-1658.
- [27] Md. Maniruzzaman, Md Shihab Uddin, Md Boktiar Hossain, Khandoker Hoque, "Understanding COVID-19 Through Tweets using Machine Learning: A Visualization of Trends and Conversations", *European Journal of Advances in Engineering and Technology*, 2023, 10(5):108-114.
- [28] Rahman, F., Das, D., Sami, A., Podder, P., & Michael, D. L. (2024). Liver cirrhosis prediction using logistic regression, naïve bayes and KNN. *International Journal of Science and Research Archive*, 12(1), 2411-2420.