Stock Price Prediction using Machine Learning

Anshuman Behera
Cognizant Technology Solutions
anshuman.behera.371@gmail.com

Ayes Chinmay
Centre for Data Science, ITER, SoA (Deemed to be University), Bhubaneswar ayeschinmay@soa.ac.in

2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS) | 978-1-6654-5493-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/MLCSS57186.2022.00009

***Abstract*—The stock market has long attracted the attention of both researchers and ordinary people. Essentially, a typical person believes that investing in the stock market is akin to catching a falling knife because they believe that there is no way to foresee the price of any company over a certain length of time. So, it's all down to chance when it comes to making money in the stock market. However, according to academics, the stock market is based on profit, loss, opening price, closing price, neighboring closing price, and number of shares traded at the moment. The stock market always operates in cycles, and if someone can anticipate when the bull and bear runs will begin and end utilizing today's technologies, the stock market may be incredibly profitable, or it can be the polar opposite. That is why we want to apply machine learning to anticipate the future price of the stock so that we may make more money in a shorter period of time while minimizing risk.**

Keywords—stock, price prediction, machine learning.

I. Introduction (*Heading 1*)

SVM (Support vector regression), K-nearest neighbour, polynomial regression, and other approaches have been utilized by most academics to forecast the share of the price. However, the issue we discovered is that they used data from a single stock over a relatively short period of time. They have also remained worried about a certain stock in a single sector, as a consequence of which there is no variance in the results received, and it has also resulted in more accuracy because just a specific sector is considered. According to market attitudes, if a sector provides a good return, then nearly all of the companies in that area provide at least a decent return. As a consequence, we decided to collect data from a variety of stocks in various industries. As a result, there will be a fluctuation in the outcome. We've gathered all of the information on the stock price from Yahoo Finance. This data set contains all of the information for that day's trades, including the date, opening price, closing price, high, low, adjacent close, and number of shares moved. We gathered data on five NSE (National Stock Exchange) shares over the course of two years, from January 1, 2017 to December 31, 2019.

II. 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 amount 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.

As a result, in this article, we have attempted to propose an efficient approach to estimate the share of the price by evaluating more than two years of data from different firms from various sectors in the Indian market, such as starting price, close price, neighboring close price, and volume.

III. Model Used

Using several machine learning algorithms, we attempted to obtain various results. To begin, we used back tracking, decision tree regression, Monte Carlo simulation, and moving

978-1-6654-5493-3/22/$31.00 ©2022 IEEE
DOI 10.1109/MLCSS57186.2022.00009

average to estimate the price of a stock at a specific date using data from the previous two years.

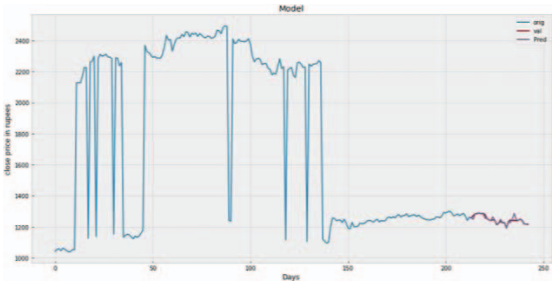
In backtracking, we attempted to solve the problem several times and to piece together a solution step by step. Then we attempted to eliminate any solutions that failed to meet the issues' cruxes at any point in time.

In a decision tree regression, the model is trained in the design of a tree to predict data in the future, resulting in meaningful and accurate continuous output.

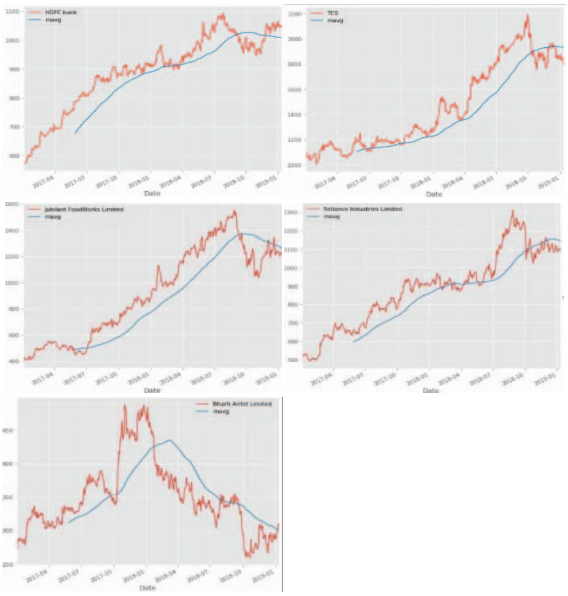
The Monte Carlo Simulation is a series of techniques for predicting the impact of uncertainty and risk in given data. To obtain numerical results, this method relies on continuous random sampling, and the simulator anticipates outcomes, giving users a greater possibility of minimizing risks and generating higher accuracy.

IV. Analysis of Proposed Model

The output of prediction by the model using back tracking with a split of 80% training set and 20% test set is as follows:

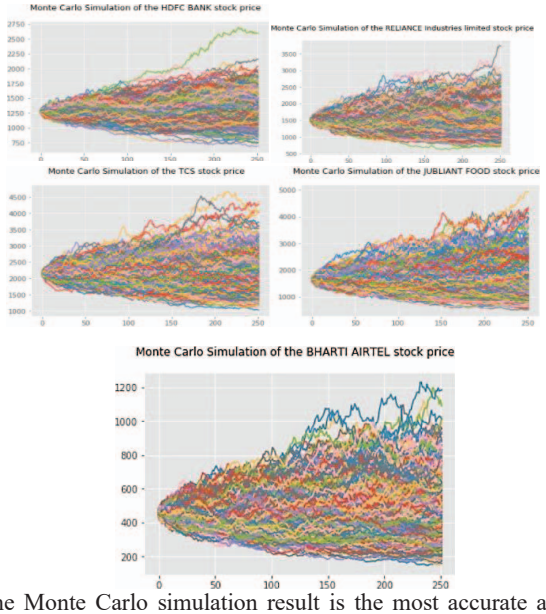


The output prediction by using moving average method for five different sectors in Indian stock market:



The result obtained by using Monte Carlo simulation over the same five stocks are the following:

When it comes to stock price prediction, there is no single strategy that provides the same level of accuracy for all stocks. We explored numerous methods for different equities and found that the accuracy varied. Out of all of these outcomes,



the Monte Carlo simulation result is the most accurate and may be utilized for analysis since it offers a share price forecast by considering a variety of scenarios.

V. Future Scope

As the stock market attracts more and more researchers' attention, and more data becomes available to everybody, there are issues such as the risk of data loss due to hacker access. As more people spend their money in the stock market, their personal information, such as brokerage account information, is at risk of being stolen by hackers. As a result of these factors, machine learning cyber security is becoming increasingly critical. In much of the country, the government, in addition to education, healthcare, and military, invests tax revenue in index funds, mutual funds, and equities to increase wealth. These are some of the factors that are drawing major tech companies' attention to the stock market. Investing in fixed bank deposits and gold now only provides a fraction of the return that the stock market provides over the same time span.

VI. Challenges

Even while there is a wealth of information available on the internet about various companies, it is insufficient to precisely anticipate the price of a share. This is owing to the market's current state of uncertainty. Most individuals participate in the stock market for a short period of time and desire to turn a few pennies into hundreds of dollars in a single day. People invest based on guesses rather than researching the company's appropriate fundamentals because of this greed. These are one of the most dangerous threat tour approaches because if the market moves as a consequence of people's greed and fear, then all models may be unable to grasp the reason for price rises and falls. And it's possible that you'll wind up supplying incorrect information.

VII. Conclusion

Using machine learning to anticipate stock prices may be expensive, and most small investors cannot afford to pay

that price. This is the primary motivation for large brokerage firms to promote algorithmic trading. Algorithm trading is a more advanced type of machine trading in which a machine uses the machine learning technique to automate the buy and sale of shares in bulk. This lowers the operating costs of algorithm trading, making it more accessible to the general public. Here, algorithm trading offers advantages such as lower costs, less reliance on human emotions, and more accuracy. The stock market is a gamble for most individuals since they don't follow the basic investment idea of studying the company's fundamentals and financial profit and loss statements. Because India is one of the world's fastest-growing economies, the government encourages citizens to invest in the stock market. These investments enable the corporation to expand its operations and contribute to the country's economic growth.

REFERENCES

[1] C. N. Ochotorena, C. A. Yap, E. P. Dadios, E. P. Dadios and E. Sybingco, "Robust stock trading using fuzzy decision trees", Computational Intelligence for Financial Engineering & Economics (CIFEr), March 2012.

[2] K. Pahwa and N. Agarwal, "Stock Market Analysis using Supervised Machine Learning", International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 10 October 2019.

[3] S. Yadav, "Stock Market Volatility - A Study of Indian Stock Market", Global Journal for Research Analysis, vol. 6, pp. 629-632, 2017.

[4] Ms. Anju bala, "Indian stock market - review of literature", Asian Journal of Marketing & Management Research, vol.2, issue 7, July 2013.

[5] S. Papadamou and S. Tsopoglou, "Investigating the profitability of technical analysis systems on foreign exchange markets," Managerial Finance, vol. 27, no. 8, pp. 63-78, 2001.

[6] E. Hurwitz and T. Marwala, "Common Mistakes when Applying Computational Intelligence and Machine Learning to Stock Market modelling", 2012.

[7] H. P. Pan, "A joint review of Technical and Quantitative Analysis of Financial Markets Towards a Unified Science of Intelligent Finance", Hawaii International Conference on statistics and Related, 2003.

[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.