IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105344 |

# Stock Price Prediction using LSTM on Quandl Dataset

**Varun Puram, Aruna Varanasi**

Dept. of Computer Science and Engineering, Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad, (TS), India.

**ABSTRACT:** Small businesses, brokerage firms, and the banking industry all rely on the stock market, also known as the share market, as a sophisticated and complex means of conducting business in order to make money and manage risks. However, it can be difficult for investors due to the market's unpredictable nature. In order to increase company predictability, this article suggests using machine learning algorithms to forecast future stock values using open-source libraries and already-existing algorithms. Although it should be emphasized that the outcome is entirely dependent on numbers and relies on many assumptions, the paper seeks to show how this strategy can produce acceptable results. It should be mentioned that these assumptions may or may not hold true in the real world at the time of prediction.
Keywords: investor, machine learning

## 1. INTRODUCTION

The stock market has been a traditional platform for regular individuals to trade stocks, make investments[8], and potentially earn profits by owning a part of companies that are listed on it. This system can be a lucrative investment opportunity if executed wisely. However, due to the unpredictable nature of stock prices and market liquidity, technology has become increasingly important in assisting investors. Machine learning is one such tool that can aid in predicting stock market trends.

The stock market is a big trade venue with significant effects on both people and countries [2]. In order to raise money, businesses essentially list their shares as commodities called stocks on the market. When these stocks are first made available to the public at a price known as the initial public offering (IPO), the company sells them to raise money. Following the initial public offering (IPO), the stocks are the owner's property and may be sold to a buyer at any time on a stock exchange like the Bombay Stock Exchange (BSE) for any amount. Following each successful transaction, the price of the share rises as a result of the ongoing buying and selling of these shares at various prices. However, if a business issues additional shares.

A graph can be easily extrapolated by the human brain with just a little observation, based on studies. Crowd computing can also be used to get a fair approximation of a real-life graph if a number of people strive to extend the graph by a set amount of time[1]. Crowd computing is a sluggish process, while being more successful in practice than in theory[5]. Consequently, a more mathematical and scientific approach is preferred, and computer simulations can be used for this. Crowdsourcing can be helpful when making stock market predictions, but because of its speed and accuracy, a computer-based method is preferred.

A common statistical method for visualizing the correlation between dependent and independent variables is linear regression [3][4]. Similar to this, in machine learning, this method is used to train classifiers to accurately predict the value of a label based on the given attributes. The quantity and caliber of the supplied data, together with the features chosen, all affect how accurate a classifier is. For developing a successful program, it is essential to have a solid understanding of data analysis, machine learning, and stock market principles. These core ideas will help us as we start to design our program.

## II. PREDICTION MODEL

In this stage, we will look over the raw data that is available and pinpoint the right attributes for forecasting the chosen label. Our data was gathered from www.quandl.com, a reputable datasets site. The dataset for GOOGL from WIKI that we chose for our program can be accessed by using the token "WIKI/GOOGL." A total of 14 years' worth of data were utilized to extract a number of variables, including Open, High, Low, Close, Volume, Split ratio, and Adjusted Open, High, Low, Close, and Volume. To extract features that will improve our predictions, we employed "Adj. Open, Adj. High, Adj. Close, Adj. Low, and Adj. Volume" with the "Close" attribute as our label. Instead of using raw data, we used modified values as processed and are free of typical data collection problems.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105344 |

The usage of OHLCV graphs for stock analysis and stock status information is well known. As a result, we chose the features for our classifier using the same graphing settings. Adj. Close and a derived feature are part of the collection of features we specified for use. HL_PCT. The percentage change formula is used to calculate HL_PCT, which helps to slash the number of features while keeping the crucial data. High-Low is an important characteristic that aids in formulating the OHLCV graph's shape. The Adjusted Close is important because it establishes the market opening price and expected volume for the next day.

What is the OHLCV graph?

A form of financial graph described as an OHLCV graph shows the Open, High, Low, Close, and Volume data for a certain stock or asset over a given time period. The "O" indicates the stock's opening price, the "H" the highest price it ever reached during that time, the "L" the lowest price it ever fell to, and the "C" the stock's closing price. The amount of shares traded during that time is shown by the "V" symbol. Traders and investors frequently use this kind of graph to assess market movements and decide when to purchase and sell stocks.

The set of features that we will utilize in our prediction model are as follows:

- Adj. Close: This attribute holds significant information as it determines the opening price of the market for the next day and the expected volume of trades.

- HL_PCT: This feature is calculated as a percentage change between the high and low values, and helps us to understand the shape of the OHLCV graph.

  It is given as:

$$HL\_PCT = \frac{\text{Adj. High} - \text{Adj. Low}}{Adj.\,Close} \times 100$$

- PCT_change: Another derived feature, which is also expressed as a percentage change. Its purpose is to reduce the number of features while retaining the relevant information.

  It is defined by:

$$PCT\_Change = \frac{Adj.Close - Adj.Open}{Adj.Open} \times 100$$

  To minimize redundancy and improve our prediction model, we apply the same treatment to Open and Close as we did with High and Low. This results in fewer features. Additionally, we include

- Adj. Volume is a decision parameter as it has a direct impact on future stock prices, making it a crucial component of our model. We do not modify this attribute and use it in its original form.

We were able to successfully retrieve the data required for classification. Any blunder or error in the procedure could result in the prediction model failing. It is crucial to remember that the features collected are subject-specific and may change depending on the subject. Generalization might be feasible if the data for a different subject are gathered in a coherent manner similar to those of the first subject. To ensure the accuracy and effectiveness of the classifier, this step must be handled very carefully.

**3. Training and Testing State:**

The obtained features and labels will now be applied to our machine-learning model. We will be utilizing Python tools like SciPy, Scikit-learn, and Matplotlib to achieve this. Our model will first be programmed by utilizing the newly acquired characteristics and labels to train it, and it will then be tested using the same data.

We will initially pre-process the data to generate the following data:

- Changed label attribute values by the desired prediction percentage.
- Numpy array format is converted from data frame format.
- All NaN data values are eliminated before the classifier receives them.
- The data is divided into test data and train data depending on its kind, i.e., label and feature, and is scaled to ensure accuracy for every value.

Our prepared data is now ready to be fed into a classifier. For this, we have used the Python Scikit-learn package's Linear Regression classifier, which is a component. This classifier was chosen since it was straightforward and suited our needs. A common technique for data analysis and forecasting is linear regression. By utilizing important

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105344 |

traits and their dependencies on other aspects, it determines the relationship between variables. Supervised machine learning is a term used to describe this kind of prediction.

Supervised learning is a machine learning method in which the input data is labeled, with each feature paired with its corr ding label. The purpose of training the classifier is to teach it to recognize the patterns and relationships betwee erent combinations of features and their corresponding labels.

In supe d machine learning, each feature is matched with a label, and the classifier is given labeled data. The classifi bsequently discovers the pattern of features that result in a certain label. The label in our situation is the stock p few days later. We input a set of features to test our classifier, then see if the result matches the label. This ai evaluating the classifier's accuracy, which is crucial for our model. For our objectives, an accuracy of less tha % is deemed impractical. For a machine learning model to be effective, it is essential to comprehend the idea of correctness and how to enhance it.

Te
m
pe
rat

## IV.RESULTS

After developing the machine learning model, we can use it to generate the desired outputin the desired format. In our case, we will be generating a graph of the predicted stock prices.

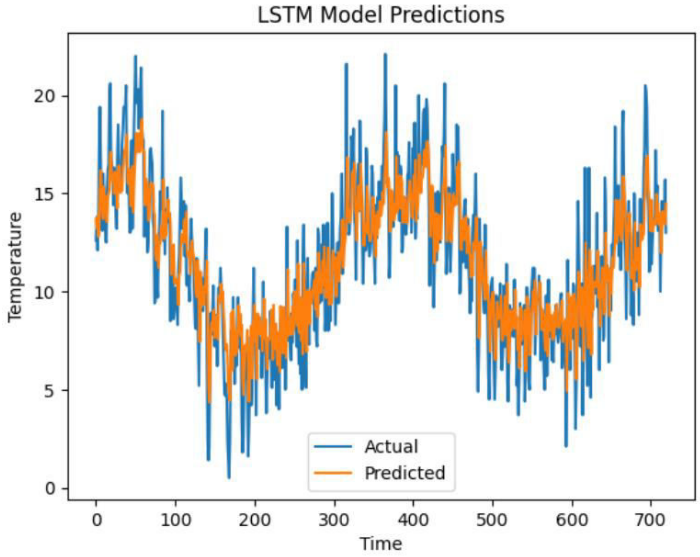


Fig.1. Graph showing the performance of Google's shares over the period of time. The blue line reflects the training value of the stock, and the red line represents the test or prediction data that is currently available.

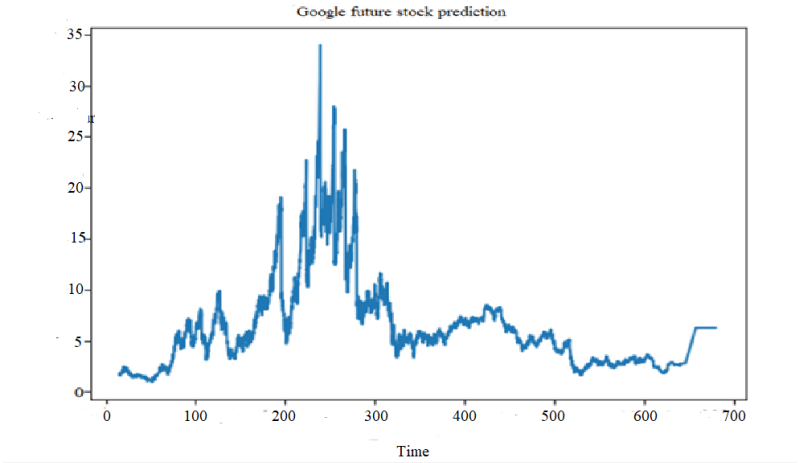


Figure 2: Represents the future stock value with the respective time period from 2019 to 2020.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105344 |

## V.CONCLUSION AND FUTURE SCOPE

Despite it has numerous uses, machine learning is a powerful technology that mainly depends on data. However, data analysis is a difficult undertaking that should not be taken lightly. Despite the development of deep learning and neural networks, the basic concept of machine learning has not changed. This study offers some insight into how supervised machine learning is put into practice. This study is restricted to outlining the fundamentals of the intricate process of supervised machine learning, despite the fact that there are several approaches and techniques to handle and resolve a variety of difficulties in various situations.

By including a larger dataset than the one it now uses, the stock market prediction system's accuracy may be improved in the future. The accuracy rates of additional upcoming machine learning models can also be investigated. The use of sentiment analysis via machine learning to ascertain how news affects a company's stock prices is another potential field. Models based on deep learning can also be used to make predictions.

## REFERENCES

[1] Andrew McCallum, Kamal Nigam, Jason Rennie, Kristie Seymore "AMachinelearningapproachtoBuildingdomain-specificSearchengine",IJCAI, 1999-Citeseer

[2] Yadav,Sameer.(2017).STOCKMARKETVOLATILITY-ASTUDYOFTHE INDIANSTOCKMARKET.GlobalJournalforResearch Analysis.6.629-632.

[3] Montgomery, D.C., Peck, E.A. and Vining, G.G., 2012. Introductiontolinearregression analysis (Vol. 821). John Wiley&Sons.

[4] Draper, N.R.; Smith, H. (1998). Applied Regression Analysis (3rded.). JohnWiley.ISBN0-471-17082-8

[5] RobertS. Pindyckand Daniel L. Rubinfeld(1998, 4h ed.). EconometricModelsandEconomicForecasts

[6] Nandakumar, R., Uttamraj, K. R., Vishal, R., &Lokeswari, Y. V. (2018). Stock price prediction using long short-term memory. International Research Journal of Engineering and Technology, 5(03).

[7] Roondiwala, Murtaza, Harshal Patel, and Shraddha Varma. "Predicting stock prices using LSTM." International Journal of Science and Research (IJSR) 6.4 (2017): 1754-1756

[8] "Linear Regression", 1997-1998, Yale Universityhttp://www.stat.yale.edu/Courses/1997-98/101/linreg.htm

[9] Agarwal(July14,2017)."IntroductiontotheStockMarket".Intelligent Economist.Retrieved December18,2017.

[10] Ghosh, Achyut, Soumik Bose, Giridhar Maji, Narayan Debnath, and Soumya Sen. "Stock price prediction using LSTM on the Indian share market." In Proceedings of 32nd international conference on, vol. 63, pp. 101-110. 2019.

[11] Wei, Dou. "Prediction of stock price based on LSTM neural network." 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM). IEEE, 2019.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details