JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF STOCK PRICE MARKET USING LONG SHORT-TERM MEMORY

Author: Lindelwa, S. Khumalo

Abstract

Stock price markets are very uncertain due to the many factors that could affect it. By investing your money in the stock market, you take a risk that you might win or lose. During the Covid-19 pandemic, one never know when the stock market might crash. Henceforth it is necessary for investors to be reassured that investments will bear forth financial gains. The study uses stock price data from several companies to predict the future stock price. It used a machine language algorithm called Long Short-Term Memory (LSTM) to analyse historical stock prices to make a prediction. The findings indicates that the Long Short-Term Memory model successfully predicted the stock prices of Apple and Microsoft with data from Yahoo Finance. The model was able to make use of historic stock prices and learn the trends and patterns from them to predict highly accurate results for the next year.

Key words: Stock price, Price Data, Long Short-Term Memory. Apple and Microsoft, Yahoo Finance

Introduction

Stock market is a market that facilitates the sale of stock, bonds, and other securities. Every day millions of economic transactions occur with investors losing or gaining money by buying or selling the stock at a rate called stock value determined by a company's financial success (Anmol, 2019). The stock traded enables the owners to own shares in a business and partake in major company decisions like voting (Afrianto *et al*., 2013). The possibility of gaining or losing shares is unpredictable.

The stock market is an important part of an economy. It is erratic and affected by a lot of factors including global news, company news, demand and supply of that day and company mergers. These factors can cause the stock prices to rise and fall according to how they negatively or positively affect the company (Chandradas, 2019). Understanding and knowing the stock price is a requirement for any person interested in investing. Investors need to understand the stock market and the factors that influence it to make wise decisions about where to invest. Investors need to accurately predict the stock market movements to be successful in their investments (Wagg, 2020)

With advances in technology, researchers have been trying to come up with tools to predict stock. Data mining has proven to be a focus for researchers when it comes to classifying and predicting data. Data mining refers

to extracting or mining knowledge from a large data set (Jian, 2020). Data mining comes in handy in stock market analysis because it provides many techniques to use and extract data for finding trends or patterns.

Many research papers have tried to incorporate data mining in their research method to try and achieve an accurate prediction of stock prices, but these kinds of methods did not work (Roy, 2020). A method that promises to give the highest percentage of accuracy is Artificial Neural Network (ANN). An artificial neural network which is mostly referred to as a neural network is a set of interconnected links that have weights associated with them Nasdaq, 2021). Artificial Neural Network is one of the best because it involves regression and classification processes of data mining.

Investors are always looking for companies to invest in and always hope their investments will be fruitful. Predicting the stock market is a demanded task worldwide (Dwivedi, 2020). A prediction system that is 100% correct is impossible in this day and age. With multiple factors affecting stock prices, people have been wishing to know the right time to risk their money and invest in the stock market (Baeldung, 2020). This study, therefore, aims to develop an application that will analyse the historical data and then predict the stock price to help investors know when to invest in the stock market or buy or sell stock without risk. The application will be designed using Python language and the algorithm used to develop this application is the Long Short-Term Memory. This study is significantly important to the business world because stock price plays a vital role in the economy of a country

Methodology

This study used the Quantitative research technique. The technique involves numeric data being manipulated or analysed using a specific technique to get some sort of result that satisfies the objectives of the study and the goal of this method is to test, predict or control (Apuke, 2017).

The target population was the stock market which is a collection of markets where the public can buy or sell shares of companies (Chen, 2020). This means that the target market of this research was all the companies that are a part of the stock market which are referred to as population elements.

Non-Probability sampling consists of techniques such as Convenience sampling, Quota sampling, Purposive sampling, and Expert selection sampling. Convenience sampling involves following the judgment of the researcher and picking a sample that the researcher deems convenient (Smigel, 2020). Quota sampling is picking a sampling from a socio-demographic or predetermined group. In Purposive sampling, the researcher seeks to choose a sample that is diverse or a rare target population while in Expert selection sampling the sample is picked by experts in that subject or field (Nivethitha, 2019).

The non-probability sampling method was used and the convenience sampling technique was used in this research as each company in the stock market is a population element. Two elements were used in the research as samples, namely elements Apple Incorporated, and Microsoft. The Convenience sampling technique was used to generally pick the sample and not have it randomly decided.

A secondary data method was used in this study because the data is coming from the internet. Raw data was retrieved from Yahoo Finance which is an application that provides financial news and data. The information from Yahoo is freely available to anyone in the public for download.

Data were analysed using the coding language called Python. Python is an interpreted, object-oriented programming language. It contains built-in data structures and supports modules and packages which makes it appropriate for Rapid Application Development (Python Software Foundation, 2021). Python also contains multiple libraries which are methods or functions that can be used imported to analyse or edit your code.

RESULTS AND DISCUSION

The section shows the results produced by the Long Short-Term Memory system to predict the stock price of Apple and Microsoft. The results were discussed and displayed in the form of graphs and figures. The factors that affected the accuracy of the model will also be discussed.

In figure 1, the blue line represents the actual stock price from January 2015 to December 2019, the orange line represents validation which is the data used to evaluate the model's ability to learn and predict. The green line represents the prediction made by the model after learning from the training and validation dataset. The prediction is the output of the experiment.

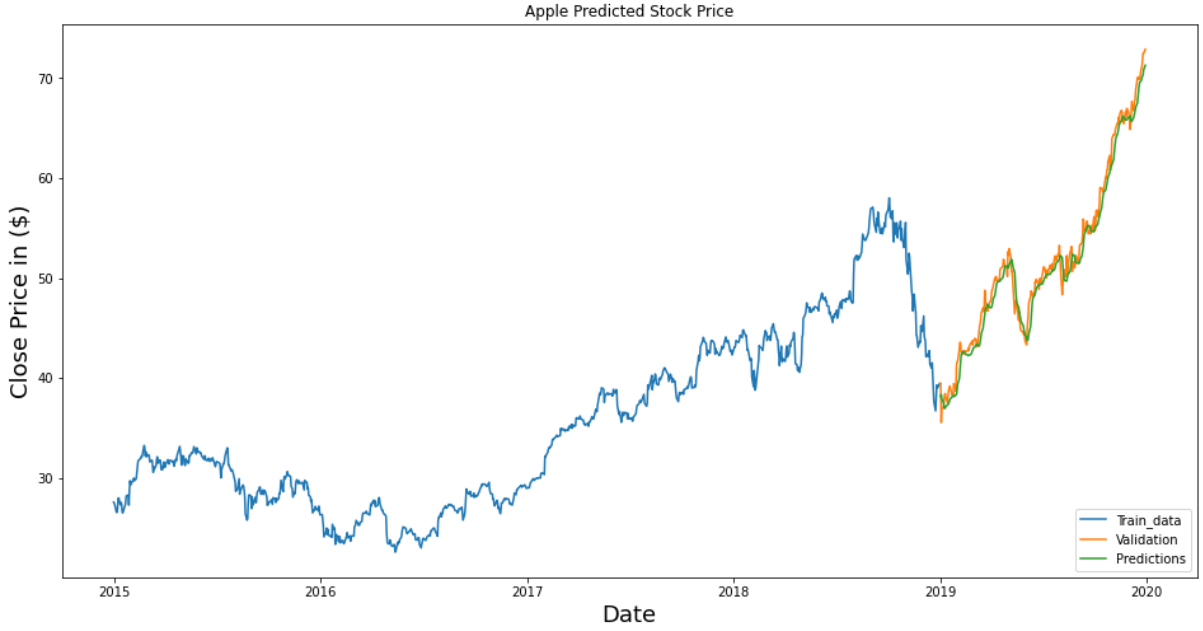


Figure 1 : Apple Prediction graph

Apple Prediction Table comparing the closing price to the prediction

Date	Close	Predictions
2019-01-02	39.480000	38.131023
2019-01-03	35.547501	38.348099
2019-01-04	37.064999	37.939911
2019-01-07	36.982498	37.523186
2019-01-08	37.687500	37.142914
...	...	...
2019-12-23	71.000000	69.762390
2019-12-24	71.067497	70.047386
2019-12-26	72.477501	70.323700
2019-12-27	72.449997	70.799667
2019-12-30	72.879997	71.271301

Figure 2: Apple Prediction

Microsoft Corporation

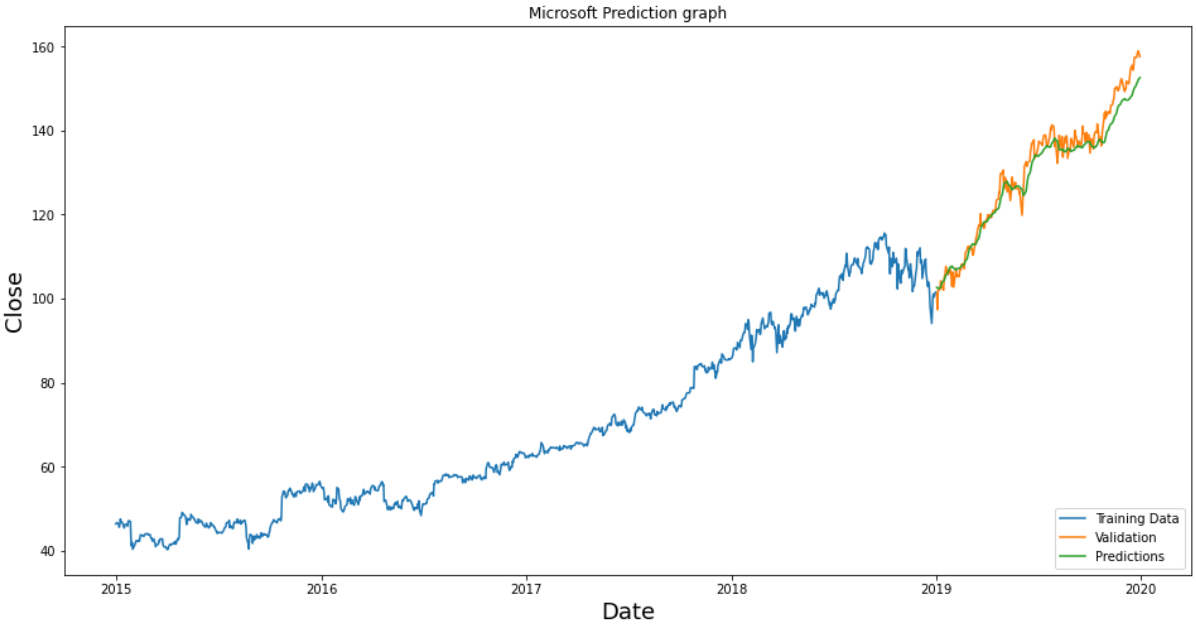


Figure 3: Microsoft prediction graph

Microsoft Prediction Table comparing the closing price to the prediction

Date	Close	Predictions
2019-01-02	101.120003	102.688904
2019-01-03	97.400002	102.699226
2019-01-04	101.930000	102.426010
2019-01-07	102.059998	102.440300
2019-01-08	102.800003	102.610703
...	...	...
2019-12-23	157.410004	150.666275
2019-12-24	157.380005	151.171875
2019-12-26	158.669998	151.625473
2019-12-27	158.960007	152.118515
2019-12-30	157.589996	152.601257

Figure 4: Microsoft Prediction

The model proved to be very effective because the prediction results it generated were closely like the actual results. The Root mean squared error method (RMSE) was used to test the accuracy of the model and the results for the RSME are 1.792 for Apple Inc and 6.026 for Microsoft.

While designing and implementing the model, several factors influenced the accuracy of the model. The first factor that had an influence on the model was the source of the data. When the data was downloaded from Yahoo Finance and saved on the computer, the model had difficulties accessing and analysing the data and would fail to give any results due to the format of the data. On the other hand, after changing the source of the data and retrieving it straight from Yahoo Finance, the model worked well and gave good results (Kumar, 2018). Another factor that affected the accuracy of the model is the size of the data. A model is more likely to give better and more accurate results if there is more data (Gupta, 2019). In the case of this research, the data used was 5 years' worth of data for each company which gave the model enough data to be able to learn the patterns and trends for each company and to be able to give a good prediction.

Investing in the stock market is becoming popular giving researchers the incentive to find a way to new techniques of predicting future stock prices. To successfully predict stock price, researchers need to come up with an accuracy technique (Mohamed, 2015). The Long Short-Term Memory model proposed in this study has successfully provided a prediction of the stock prices of Apple and Microsoft with data from Yahoo Finance. The model was able to make use of historic stock prices and learn the trends and patterns from them to provide highly accurate prediction results for the next year.

The findings are in line with (Nivethitha, 2019,) who reported concluded that the LSTM model was highly accurate. The results showed a close resemblance between the actual stock price and the predicted stock price (Levy, 2021). Panthula (2016) reached a similar conclusion that the model was able to establish clear results during prediction while working with a limited size of data.

References

Afrianto, R., Tjandrasa, H., & Arieshanti, I. (2013). Stock Exchange Prediction using Back Propagation Neural Network. *Simantec*.

Anmol. (2019). Probability Sampling Techniques: A Literature Review. *International Journal of Science and Research, 8*(7).

Apuke, O. (2017). Quantitative Research Methods : A Synopsis Approach. *Arabian Journal of Business and Management Review (6).*

Baeldung. (2020). *Feature Scaling*. Retrieved from Baeldung CS: https://www.baeldung.com/cs/feature-scaling

Chandradas, A. (2019). *All you need to know about yfinance : Yahoo! Finance Library*. Retrieved from Medium: https://medium.com/nerd-for-tech/all-you-need-to-know-about-yfinance-yahoo-finance-library-fa4c6e48f08e

Chen, J. (2020). *Stock Market*. Retrieved from Investopedia: https://www.investopedia.com/terms/s/stockmarket.asp

Crossman, A. (2020). *An Overview of Qualitative Research Methods*. Retrieved from ThoughtsCo: https://www.thoughtco.com/qualitative-research-methods-3026555

Dwivedi, R. (2020). *How Does Support Vector Machine (SVM) Algorithm Works In Machine Learning?* Retrieved from Analytics Steps: https://www.analyticssteps.com/blogs/how-does-support-vector-machine-algorithm-works-machine-learning.

Gupta, A. (2019). *Python Libraries – Python Standard Library & List of Important Libraries*. Retrieved from DataFlair: https://data-flair.training/blogs/python-libraries/#:~:text=A%20Python%20library%20is%20a%20reusable%20chunk%20of,then%2C%20a%20library%20is%20a%20collection%20of%20modules.

Jain, D. (2019). September 9). *Data Preprocessing in Data Mining*. Retrieved from GeeksforGeeks: https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/#:~:text=%20Steps%20Involved%20in%20Data%20Preprocessing%3A%20%201,suitable%20for%20mining...%203%20Data%20Reduction%3A%20More%20

Kumar, H. (2018)). *Scaling vs Normalization*. Retrieved from Technical Friday: https://kharshit.github.io/blog/2018/03/23/scaling-vs-normalization#:~:text=Feature%20scaling%28also%20known%20as%20data%20normalization%29%20is%20the,in%20data%20preprocessing%20while%20using%20machine%20learning%20algorithms.

Mohamed, A. (2015). Research Population. 1-9.

Nasdaq, K.(2021). Retrieved from Nasdaq website: https://www.nasdaq.com/about

Nivethitha, D. (2019). Future Stock Price Prediction using LSTM Machine Learning Algorithm. *International Research Journal of Engineering and Technology, 6*(3), 1182-1186.

Panthula, G. A. (2016). Stock Price Trends Prediction Using LSTM Neural Networks.

Roy, B. (2020). *All about feature scaling*. Retrieved from Towards data science: https://towardsdatascience.com/all-about-feature-scaling-bcc0ad75cb35

Smigel, L. (2020). *What is Open High Low Close in Stocks?* Retrieved from Analyzing Alpha: https://analyzingalpha.com/open-high-low-close-stocks#:~:text=The%20high%20is%20the%20highest%20price%20at%20which,a%20stock%E2%80%99s%2052-week%20high%20and%2052-week%20low%20listed.

Wagg, L. (2020). *The Relationship Between Interest Rates & Stock Prices* . Retrieved from DailyFX: https://www.dailyfx.com/education/understanding-the-stock-market/interest-rates-and-stock-