Stock Price Forecasting

Project report submitted in partial fulfilment of the requirements for the degree of B.Tech

by

Viknesh Rajaramon
(Roll No: COE18B060)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF INFORMATION TECHNOLOGY,
DESIGN AND MANUFACTURING, KANCHEEPURAM

December 2021

Certificate

I, **Viknesh Rajaramon**, with Roll No: **COE18B060** hereby declare that the material presented in the Project Report titled **Stock Price Forecasting** represents original work carried out by me in the **Department of Computer Science and Engineering** at the **Indian Institute of Information Technology, Design and Manufacturing, Kancheepuram** during the year **2021**. With my signature, I certify that:

- I have not manipulated any of the data or results.
- I have not committed any plagiarism of intellectual property. I have clearly indicated and referenced the contributions of others.
- I have explicitly acknowledged all collaborative research and discussions.
- I have understood that any false claim will result in severe disciplinary action.
- I have understood that the work may be screened for any form of academic misconduct.

Date: **08.12.2021** Student's Signature

In my capacity as supervisor of the above-mentioned work, I certify that the work presented in this Report is carried out under my supervision, and is worthy of consideration for the requirements of project work during the period October 2021 to December 2021.

Advisor's Name: Advisor's Signature

Abstract

Stock value forecasting is a difficult endeavour that needs a solid computational foundation in order to calculate share prices in the longer term. Because stock prices are connected by nature of the market, forecasting expenses will be challenging. The suggested approach uses market data and machine learning technique called Long Short Term Memory to forecast share price, with weights modified for each data point in the process. When compared to existing stock price prediction systems, this method will deliver accurate results.

Acknowledgements

I would extend my sincerest gratitude to **Dr. Pandiri Venkatesh** for being my guide in this project and helping me complete this project.

I would also like to thank the authors of all the papers that I have referenced.

Finally, I would want to express my gratitude to my institute faculty, friends, and family for their invaluable assistance and wisdom in assisting me in successfully completing my internship.

Contents

List of Figures

List of Tables

Abbreviations

kNN	**k** **N**earest **N**eighbour
LSTM	**L**ong **S**hort **T**erm **M**emory
MSE	**M**ean **S**quared **E**rror
RMSE	**R**oot **M**ean **S**quared **E**rror
RNN	**R**ecurrent **N**eural **N**etwork

Chapter 1

Introduction

1.1 Introduction

Stock investing provide some of the best returns on investment. Despite their volatility, share prices and other statistical elements may be visualised, allowing observant investors to make informed decisions about which firm to invest in. Machine learning-based stock price prediction assists you in determining the future worth of a company's shares and other financial assets traded on a stock exchange. The whole point of stock price forecasting is to make a lot of money. It's difficult to predict how the stock market will fare. Other aspects, such as physical and psychological characteristics, reasonable and illogical conduct, and so on, have a role in the forecast. All of these elements combine to create a vibrant and turbulent stock market. This makes it extremely difficult to accurately estimate stock prices.

1.2 Problem Statement

As discussed, stock prices of any company are subject to volatility because of various factors like quarterly results, dividend announcements, bonus announcements, stock splits, news about the company. Apart from the above factors, stock prices are dependent on various technical indicators, support and resistance and candlestick patterns. This makes forecasting stock price accurately one of the most difficult problems in machine learning domain. In this project, the aim is to forecast the stock prices accurately only using historical data. As a naive solution to this problem, the moving average method which uses the latest set of values to for each forecast. In other words, for each subsequent step, the oldest observed values are removed from the set while the forecasted values are added to the set. But this method does not give promising results as the predicted values are of the same range as the observed values in the train dataset and is not able to detect the uptrend earlier (delay between actual an forecasted data).

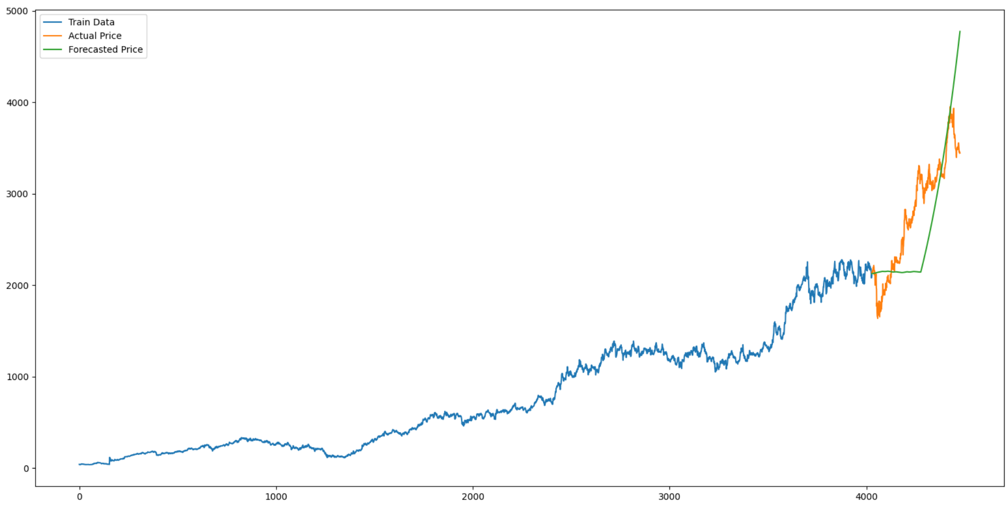

FIGURE 1.1: Forecast Stock Price vs Actual Stock Price - Moving Average Method

Chapter 2

Literature Survey

2.1 Studies on Stock Market Forecast

This section consists of the details of various research by various authors relating to our study of interest and discuss them.

2.1.1 Moving Average Method

Average is most commonly used for any type of data. The average of the set of previously observed values is the forecasted stock price for each day. We use moving average rather than simple average as it captures the recent set of values for each forecast.

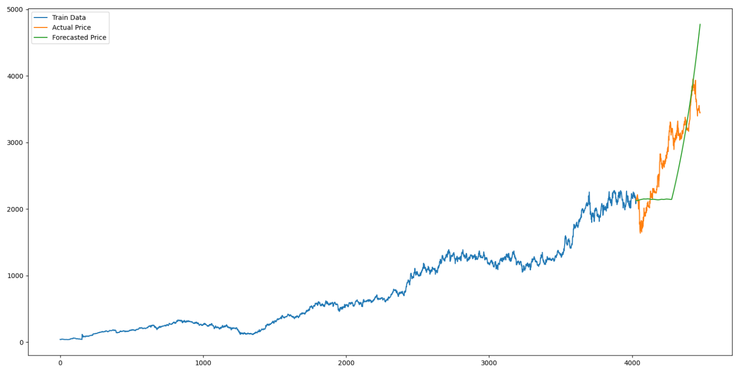

Figure 2.1: Forecast Stock Price vs Actual Stock Price - Moving Average Method

2.1.2 Linear Regression

Linear regression model establishes the relationship between the independent variables and the dependent variable. But for stock price forecast, all the variables are dependent on each other. Considering Date as a parameter, the model does not take into account the recent values from the date of prediction. The model rather considers the value from the same date a week ago, a month ago, or the same date/month a year ago. Linear regression performs better when the variables are independent of each other.

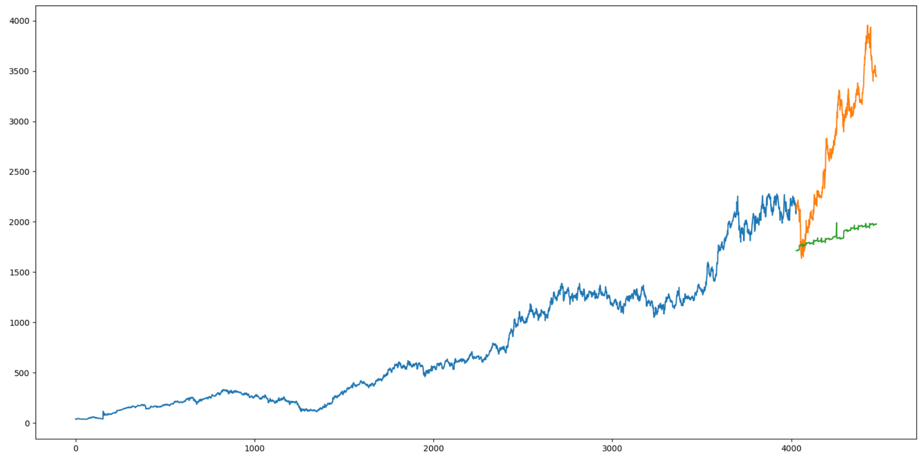

FIGURE 2.2: Forecast Stock Price vs Actual Stock Price - Linear Regression

2.1.3 k-Nearest Neighbour

kNN finds similarity between old and new data points based on independent variables. Similar to Linear Regression, kNN algorithm also performs better in case of independent variables. In general, regression algorithms perform poorly in the case of stock price data as it contains dependent variables.

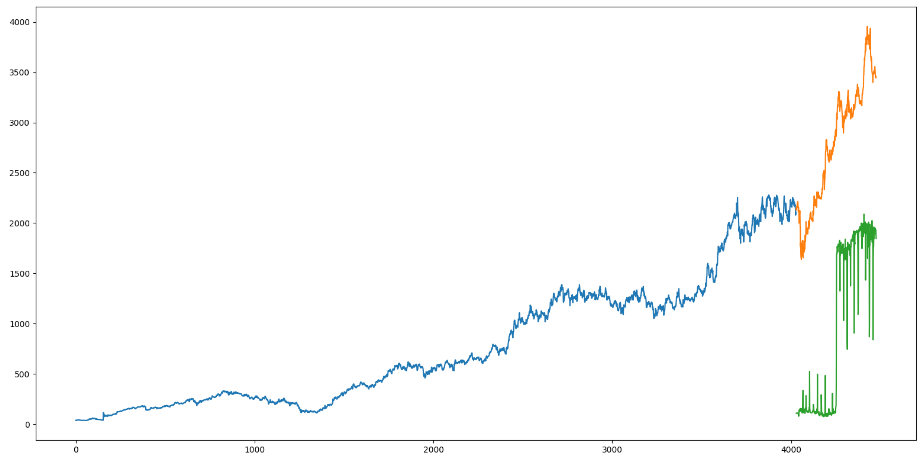

FIGURE 2.3: Forecast Stock Price vs Actual Stock Price - k Nearest Neighbour

Chapter 3

Work Done

3.1 LSTM Based Approach

As illustrated in the literature survey section, there are some research works done to forecast stock data based only on historical data. These methods, however are unable to store past information and unable to forget the information that is not required. The methodology used for this project is LSTM based, which is able to store and retrieve past data that is important as well as forget the data that is not required.

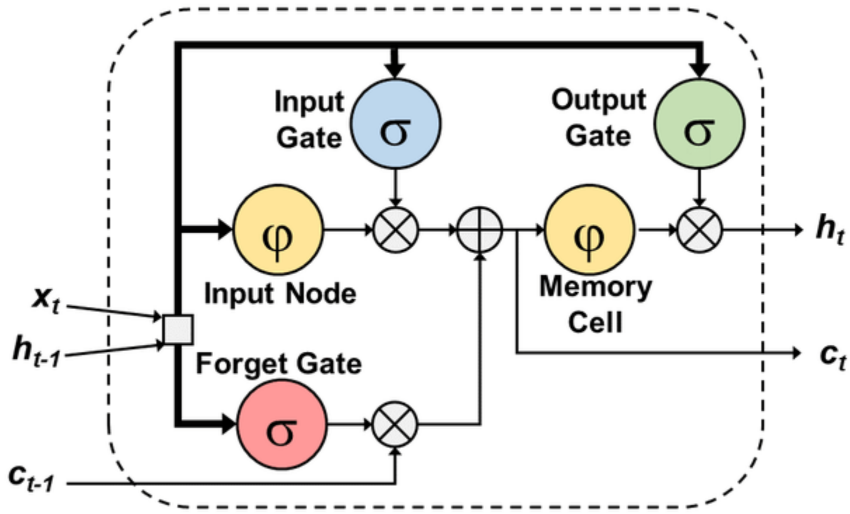

FIGURE 3.1: LSTM Cell Architecture

LSTM is a RNN architecture modified to remember past information over arbitrary intervals. RNN takes the last hidden state's output for its computations and doesn't remember any information before the hidden state. LSTM network has three gates:

- **Input Gate :** Adds data to the cell state
- **Forget Gate :** Removes the data that is no longer required by the LSTM model
- **Output Gate :** Selects the data to be shown as output

3.2 Dataset Gathering

The stock symbol (hereafter referred to as ticker symbol) and the exchange where the symbol is listed is taken as input for retrieving the data. With the ticker symbol and the stock exchange, the historical data is scraped from Yahoo Finance website using Selenium package in Python. The scraped data is then parsed using BeautifulSoup package in Python. The parsed data is then preprocessed to convert the data to their respective datatypes (Date column is converted from String to Date and Open, High Low, Close, etc. columns are converted from String to Float) and the processed data is then stored as a dataframe. The number of days to forecast is taken as input to forecast the stock price for that number of days.

3.3 Preprocessing Phase

The columns which are not being forecasted are dropped from the dataframe (To predict the closing price, Date and Close columns are retained in the dataframe. All other columns are dropped). The entries with NULL values are discarded to ensure a clean dataset for training. The data is then sorted based on date in ascending order as stock price dependent on time. With all the preprocessing done, the LSTM network is now ready for training.

3.4 Training Phase

With the acquired dataset, the dataset is first scaled down using the MinMaxScaler module with the feature range between [0, 1] before training a LSTM network with two layers for 200 epochs. Early Stopping was used prevent the model from being overfitted as the model began to diverge from the minima after a certain number of epochs. The dataset is split into 90% training data and 10% test data. The recent 10% of the dataset is used as test dataset. The hyperparameters, such as batch size, optimizer and loss function are adjusted to yield optimal results. Since this is not a classification problem, MSE is used as loss function to validate the model. The first sesssion of training gave a RMSE value of 826.32 which is further improvised based on various inferences. With all the improvements in training, the last session of training gave a RMSE value of 101.07.

Chapter 4

Results and Discussions

4.1 Observation

The training is carried with the dataset of 4476 samples. The batch size was set to 1 and Early Stopping was used to prevent the model from being overfitted. Early stopping is implemented by monitoring the loss and the patience level was set to 5. The loss function used is RMSE, as it is not a classification problem.

Name	**Last**		**Min**	
	Value	**Step**	**Value**	**Step**
Train Loss	0.000016935202438617125	47	0.0000163258591783233	42
Test Loss	0.000081051141023635586	47	0.000059723475715145 47	23

Table 4.1: Loss Summary

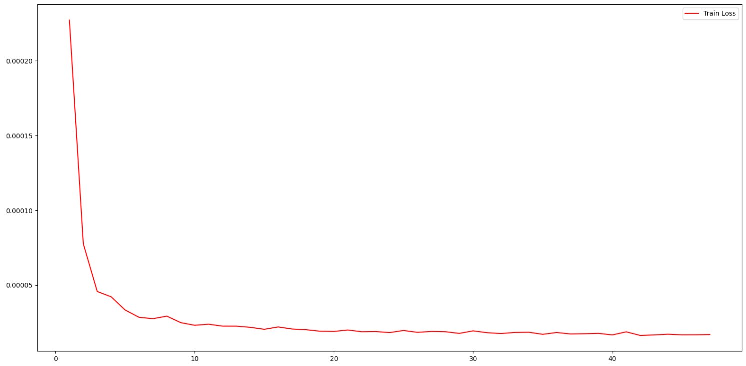

FIGURE 4.1: Training Loss

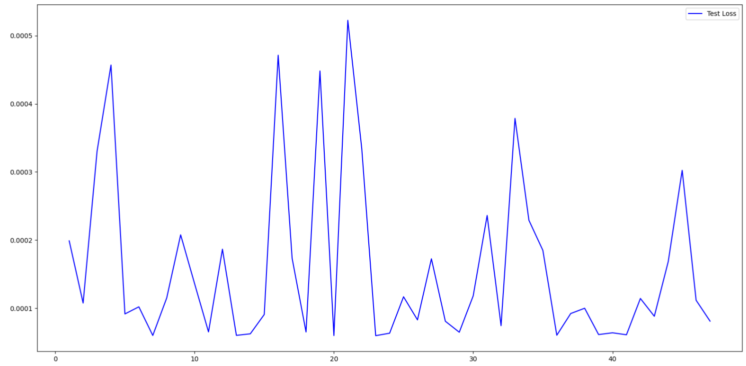

FIGURE 4.2: Testing Loss

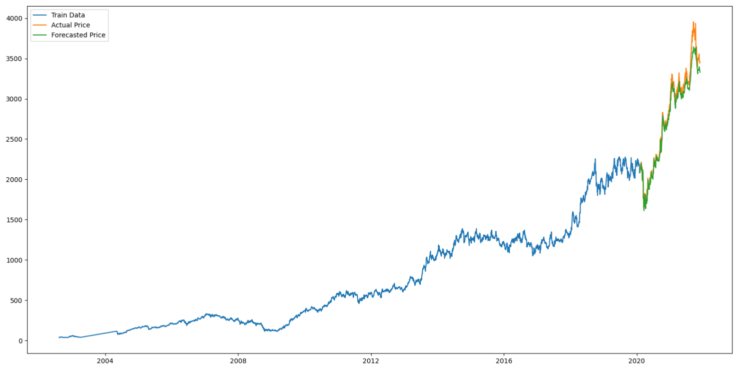

FIGURE 4.3: Actual Stock Price vs Forecast Stock Price

The forecasted stock price was found to almost coincide with the actual stock price. The RMSE value was 101.07.

4.2 Possible Improvements

- The model could be trained for more number of epochs by increasing the patience level in Early Stopping.
- The number of units in each layer can be increased to make the model more accurate.
- The number of hidden layers can be increased.
- Dropout can be added after every layer to avoid overfitting of the model.

Chapter 5

Conclusions and Future Works

5.1 Future Scope

- The model can be made more accurate by incorporating Sentiment Analysis as any news about the organization or any news that impacts the organization has a direct affect on the stock price.
- Accuracy to forecast the stock price on a daily interval can be greatly improved if the training dataset consists of tick-by-tick data instead of 1D interval data

5.2 Conclusion

The application designed to forecast stock prices is more reliable and promising than the conventional RNN and regression methods. Also, the hyperparameter tuning is done keeping in mind the time taken for training and testing without comprimising the accuracy of the forecast. The model can be further improvised to perform Sentiment Analysis as stock market is irrational and prone to human emotions. The obtained results prove to be of great assistance as the model is able to forecast the price trend with an accuracy of 80%.

Bibliography

[1] Batres-Estrada, B., *Deep learning for multivarate financial time series*, 2015.

[2] Raut Sushrut Deepak, Shinde Isha Uday, Dr. D. Malathi, *Machine Learning Approach in Stock Market Prediction*, IJPAM 2017.

[3] T. Kim and H. Y. Kim, *Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data*, PloS 2019.

[4] Loke.K.S, *Impact Of Financial Ratios And Technical Analysis On Stock Price Prediction Using Random Forests*, IEEE 2017.

[5] X. Shao, D. Ma, Y. Liu, Q. Yin., *Short-term forecast of stock price of multi-branch LSTM based on K-means*, ICSAI 2017.

[6] K. Chen, Y. Zhou, and F. Dai, *A LSTM-based method for stock returns prediction: a case study of China stock market*, IEEE 2015.

[7] S. Hochreiter and J. Schmidhuber, *Long short-term memory*, 1997.

[8] Yaohu Lin, Shancun Liu, Haijun Yang, Harris Wu, *A Deep Learning Framework for Stock Prediction Using LSTM*, DECON 2019.