



Stock Market Price Prediction Using Machine Learning

Sachin Dhananjay Jadhav,

Guide: Prof. P. J. Kulkarni

Shivaji University Kolhapur, K. E. Society's Rajarambapu Institute of Technology

Islampur, Sangli, Maharashtra

Email - sjsachinjadhav12@gmail.com

Abstract

Investment corporations, hedge funds, and even individual investors need to study financial models to understand market behavior to build profitable investments and trades. Investors usually produce educated guesses by analyzing information of old stock prices, company's performance behavior, etc. The initial phase of revealing theories in the guesswork indicates that stock unit prices are fully random and unpredictable. In the betterment of the guesswork, quantitative analysts get deployed to make prophetic models. The paper focuses on the use of machine learning techniques in developing better models for enabling appropriate recommendations for financial investments.

Keywords—: Stock Price; Random Forest Regression; Machine Learning; Training; Testing

Introduction

The world's stock markets comprehend huge wealth. As long as markets have existed, investors have hunted for ways in which to amass data regarding the companies listed in the market.

In the past, investors relied upon their personal expertise to spot market patterns, but this is not possible nowadays. Easily applied math analysis of economic information provides some insights. However, in recent years, investment firms have used numerous varieties of artificial intelligence (AI) systems to seem for patterns in huge amounts of real-time equity and economic information. These systems support investment decision making and they have currently been used for a

Sufficiently long amount that their features and performance will be reviewed and analysed to

Identify those systems and improve prophetic performance compared with alternative techniques.

A correct prediction of stocks will result in large profits for the vendor and broker. Frequently, it's brought out that prediction is chaotic instead of random, which implies it is expected by analyzing the history of several securities markets. Machine learning is economical; thanks to representing such processes. It predicts a worth market price value near to the tangible value, thereby increasing the accuracy. The introduction of machine learning to the world of stock prediction has appealed to several types of research attributable to its economical and correct measurements.

The very important part of machine learning is the dataset that is used in the learning. The information set ought to be as concrete as potential as a result of which amendment within the data will uphold huge changes within the outcome. During this project, supervised machine learning is used on a dataset obtained from Yahoo



Finance. This dataset contains the following 5 variables: open, close, low, high, and volume. Open, close, low, and high area units indicate completely different bid costs for the stock at separate times with nearly direct names. The cost is the variety of shares that passed from one owner to a different throughout the fundamental measure. A model is then derived to look on test information. Regression and LSTM models area engaged for this conjecture, one by one. Regression involves minimizing error and LSTM contributes to the basic cognitive process of the information and the results for the end of the day. Finally, the graphs for the fluctuation of cost with the dates (in case of Regression-based model) and between actual and expected price (for the LSTM based model) are planned.

In Stock Market Prediction, the aim is to predict the longer-term price of the money stocks of a corporation. The recent trends available for market prediction technologies are based on the use of machine learning that makes predictions which are supported by the values of current exchange indices and coaching on their previous values. Machine learning employs different models to create prediction that are easier and more authentic. Our main focus is on the utilization of Regression and LSTM techniques of machine learning and deep learning to predict stock values. Factors considered are: area of unit open shut, low, high, and volume of stock values.

Here we used several techniques for calculating the prices like R factor, Quantitative Analysis.

R Factor - The chance/praise ratio, every so often called the "R/R ratio," compares the capability income of a change to its capability loss. It is a calculation that makes use of the distinction among the access factor of a change and the stop-loss order to decide chance, and the distinction among the income goal and the access factor to locate praise.

Quantitative Analysis - Quantitative evaluation (QA) in finance is a technique that emphasizes mathematical and statistical evaluation to assist decides the price of a monetary asset, along with an inventory or option. Quantitative buying and selling analysts (additionally recognized as "quanta") use a number of data—inclusive of historic funding and inventory marketplace data—to increase buying and selling algorithms and pc models.

I. LITERATURE REVIEW

Osman Hegazyetal, Omar S. Soliman, Mustafa Abdul Salam 2013 [1]

Stock price prediction can be predicted using AI and machine learning models in machine learning fields. Using the SVM model for stock price prediction. SVM is one of the machine learning algorithms which works on classification algorithms. It is used to get a new text as an output. Applying Multiple Linear Regression with Interactions to predict the trend in stock.

Kai-Yin Woo, Chulin Mai, Michael McAleer, Wing-Keung Wong 2020

Review on Efficiency and Anomalies in Stock Markets [2]

Scholars could use data from stock markets all over the world to check whether the markets are efficient, as well as find whether there is any market anomaly. When there is any anomaly being discovered, scholars first confirm the existence of the market anomaly and thereafter look for any existing model to explain the anomaly. If scholars cannot estimate, evaluate, and forecast any model to explain the anomaly, scholars will then explain



the anomaly by using quantitative analysis, modeling, or even building up a new theory to explain the anomaly that built up the theory of Behavioral Finance. However, if there is any unexplained anomaly, one may grasp the methods to profiteer by using the anomaly. On the one hand, this is a good way to offer investors valuable investment advice

Boriss. Siliverstovs, Manh Ha Duong JUNE 9 2006

On the role of stock market for real economic activity [3]

The relationship between the stock market and the real economic activity, represented by the real Gross Domestic Product, for the five European countries: Germany, France, Italy, the Netherlands, and the UK. In addition to the variables that are commonly used in such an analysis like the stock market returns, the measures of real economic activity, and the interest rate. The authors have included the composite leading indicator in our empirical VAR models.

Fahad Almudhaf Yaser A. AlKulaib OCTOMBER 2012

“ARE CIVETS STOCK MARKETS PREDICTABLE” [4]

The weak-form potency and stochastic process behavior of the CIVETS stock markets throughout the amount 2002–2012. We have a tendency to applied unit root tests, serial autocorrelation, and variance quantitative relation tests. Our unit root results imply that CIVETS follow a stochastic process.

Hiransha M, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P 2018 [5].

NSE Stock Market Prediction Using Deep Learning Models

In this work the authors tend to used four deciliter architectures for the stock value prediction of NSE and securities market, which are 2 totally different leading stock markets within the world. Here we tend to trained four networks MLP, RNN, LSTM and CNN with the stock value of TATA MOTORS from NSE. The models obtained were used for predicting the stock value of MARUTI, HCL and AXIS BANK from NSE stock exchange and conjointly for predicting the stock value of BANK OF AMERICA (BAC) and CHESAPEAK ENERGY (CHK) from securities market. From the result obtained, it's clear that the models are capable of characteristic the patterns existing in each the stock markets.

Pranav Bhat 2020 [6]

1 JAN.-MAR 2020

A Machine Learning Model for Stock Market Prediction

Anticipating the securities exchange cost is exceptionally well known among financial specialists as speculators need to know the arrival that they will get for their ventures. Generally the specialized experts and intermediaries used to foresee the stock costs dependent on chronicled costs, volumes, value designs and the essential patterns. Today the stock value expectation has turned out to be mind boggling than before as stock costs are influenced because of organization's money related status as well as due to socio practical state of the nation, political environment and cataclysmic events and so on.

II. A PROPOSED SYSTEM

We have developed LSTM (Long Short Term Memory) model in stream-lit which will predict the values based on old dataset. The Prediction values are High, Low, Open, Close. It is reliable application for student and beginners who wants to trade and they can easily identify the trends in the market whether the market is going upward or downward or else it will remain in sideways.

The previous paper which we have used for study purpose [1] they have used the Support Vector.

Machine and Particle Swarm Optimization algorithm and we are going to use LSTM (Long Short Term Memory) algorithm with linear regression and Random Forest.

We have used confusion matrix for classification report.

We have used two methods that is regression and classification for stock market prediction. In regression method closing price of company stock is predicted. The classification method it will predict company stocks closing price that will increase or decrease in upcoming days.

Dataset: - Yahoo Finance. (<https://www.kaggle.com/achintyatripathi/eda-autoviz-class-one-line-code-yahoo-stock-price?scriptVersionId=42446951>)

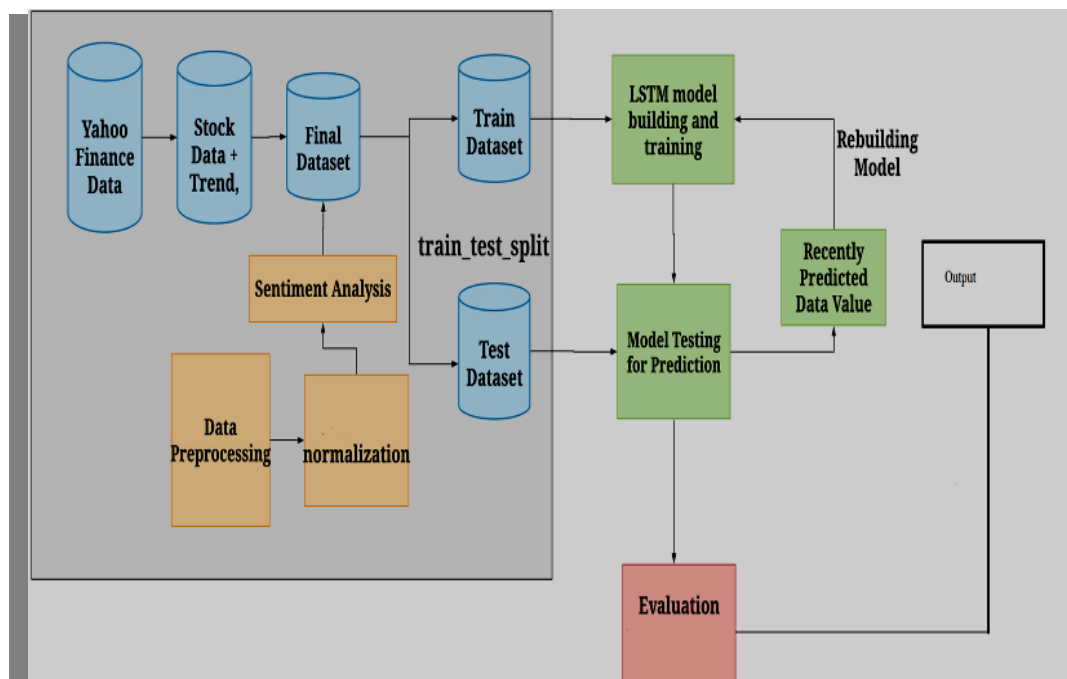


Fig 1.Proposed System Design

The figure1 shows proposed system Design. Here we have taken datasets from yahoo finance data. The first step is to train the data and in the second step, testing of data is done and with the LSTM model we forecast the values to get predation value.



The Outcome of LSTM Algorithm which used by this research.

Table 1 indicates the results generated with LSTM algorithm

1. Before Training – Old Dataset

The Table1 shows old datasets used from Yahoo finance which is having time series, open value, low value, high value, close value. This are the few columns showed in the below dataset.

	Open	High	Low	Close
2016-02-17	58.6000	58.6000	57.1000	58.1700
2016-02-18	58.1700	58.9600	57.9800	58.6000
2016-02-19	58.5000	59.3100	57.6300	59.3100
2016-02-22	59.6500	60.4750	59.2600	60.1700
2016-02-23	60.2600	60.7400	59.8900	60.2100
2016-02-24	59.7400	60.6950	59.2000	60.4000
2016-02-25	60.6800	62.3800	60.6299	62.3800
2016-02-26	62.6400	63.4900	62.4657	62.6000
2016-02-29	62.4600	62.6800	61.5800	61.5900
2016-03-01	61.9700	63.1000	61.7500	62.9200

Table 1. LSTM Result

2. Test Data – Linear regression

Table 2 shows the outcome of tested data using linear regression algorithm for NKE (Nike) company. Here we can see the actual value and the predicted value on the same date.

	Date	Actual	Predicted
0	2019-11-22T00:00:00	93.3400	92.7280
1	2019-11-25T00:00:00	92.9000	92.8143
2	2019-11-26T00:00:00	93.3300	93.0370
3	2019-11-27T00:00:00	94.1400	94.0403
4	2019-11-29T00:00:00	93.4900	93.4918
5	2019-12-02T00:00:00	93.5600	93.3453
6	2019-12-03T00:00:00	92.4600	92.0188
7	2019-12-04T00:00:00	93.7200	93.6354
8	2019-12-05T00:00:00	95.7900	94.9154
9	2019-12-06T00:00:00	97.0000	96.5986

Plotting Actual vs Predicted for - Linear Regression

Table 2. Linear Regression Result

3. Test Data – Random Forest

Table3 shows the outcome of tested data using Random Forest algorithm for NKE (Nike) company. Here we can see the actual value and the predicted value on the same date.

Stock Prediction on Test Data for - NKE			
	Date	Actual	Predicted
72	2020-03-10T00:00:00	88.3000	83.5968
73	2020-03-11T00:00:00	84.0000	83.2930
74	2020-03-12T00:00:00	74.2000	74.8072
75	2020-03-13T00:00:00	75.5800	73.4053
76	2020-03-16T00:00:00	66.7900	65.8859
77	2020-03-17T00:00:00	69.8400	65.7417
78	2020-03-18T00:00:00	68.0400	63.8903
79	2020-03-19T00:00:00	70.3400	66.1174
80	2020-03-20T00:00:00	67.4500	68.0143
81	2020-03-23T00:00:00	62.8000	64.8301

Plotting Actual vs Predicted for - Random Forest

Table 3. Result of Random Forest Algorithm

4. Total Count of Training Testing

Table 4 shows the total Count of Training and testing data of NKE (Nike) Company with using two algorithms that is linear regression and random forest.

		Train_MAE	Train_MAPE	Train_RMSE	Test_MAE	Test_MAPE
NKE	Linear Regression	0.2100	0.3100	0.2700	0.5800	0.6600
NKE	Random Forest	0.1300	0.2000	0.1800	4.2600	4.3300

Table 4. Training & Testing Data

5. New Data After Training Testing

The Table5 shows the new data after Training & Testing. Here we can see the date, open value, high value, close value, low value, and volume of the stock and also dividends.

Stock Data with Date and High Low Stock						
	Open	High	Low	Close	Volume	Dividends
2022-01-26T00:00:00	147.7800	148.9700	142.7300	143.9900	7444600	0
2022-01-27T00:00:00	144.9900	147.3500	143.5800	144.6400	6396400	0
2022-01-28T00:00:00	144.1200	146.0300	140.6100	145.9100	6130200	0
2022-01-31T00:00:00	146.2700	148.1800	145.2200	148.0700	6932200	0
2022-02-01T00:00:00	149.5000	149.6800	146.8000	148.2200	6695000	0
2022-02-02T00:00:00	148.3500	149.4600	147.5300	148.7100	5491700	0
2022-02-03T00:00:00	147.4000	148.4600	144.9700	145.3100	5240200	0
2022-02-04T00:00:00	145.3700	146.4800	143.6400	145.3900	4312500	0
2022-02-07T00:00:00	145.3000	146.1200	144.2200	145.1400	3877500	0
2022-02-08T00:00:00	144.7400	145.5000	142.3000	143.5300	5829400	0

Table 5. Stock After training & Testing

6. New Actual Predicted Result

The Below Plotted Graph shows the actual price & Predicted Price of stock.

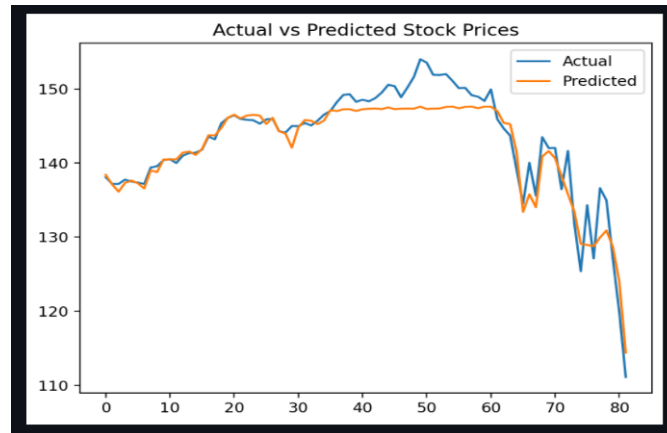


Fig 2. Actual Vs Predicted Stock Prices

In above Fig 2 we see actual value which is denoted by blue color and predicted value is denoted by orange color.

Methodologies Used for Project

1. R – Factor

- Equity market risks can be broadly classified as systematic and unsystematic risks.
- The supply of systematic hazard is the marketplace or international elements such as rising oil prices, currency movements, changing government policies.
- Unsystematic risks, however, are owed to factors unique to a company or an industry.
- Management and labor relations, increased competition, entry of new players, and customers’ preference for a company’s products are some of the factors that generate unsystematic risk.

$$R_p = \alpha + \beta R_M + \epsilon$$

α – Excess Return
 β – Excepted Move Value
 R_M = Market Return
 R_p = Portfolio Return
 ϵ = Error Term

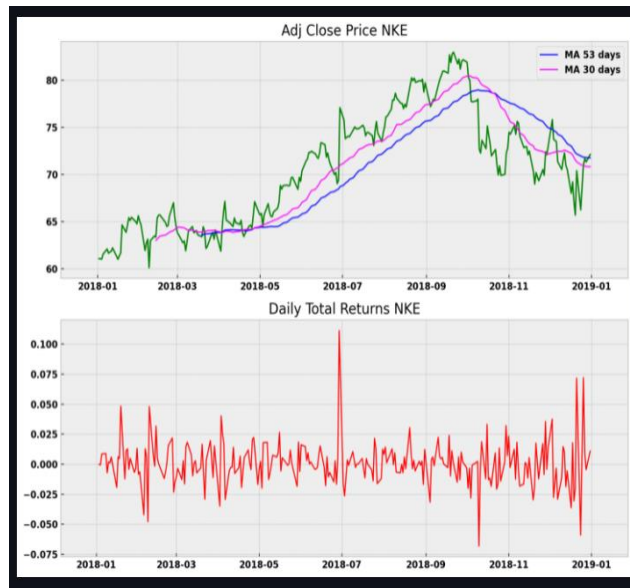


Fig 3. R Factor result

In above Fig 3 we see close Price of NKE (Nike) company with actual value denoted in green color and compared with MA (Moving Average) which is having value of MA53 days denoted with blue color and MA30 days is denoted with red color.

Here we also see the daily total returns of NKE

(Nike) company. On y-axis value is denoted and on x-axis year is denoted.

2. Stock Analysis Candle Stick Chart

- A candlestick chart is a style of financial chart used to describe price movements of a security, derivative, or currency. It is similar to a chart in that each candle represents all four important pieces of information for that day: open and close in the thick body; high and low in the “candle wick”.
- Candlesticks allow traders to visualize buying and selling pressure in two ways.



Fig 4. Candle Stick Chart

In above Fig 4 we see the candle stick chart which is compared with Boll (Bollinger Band) denoted with red color, RSI (Relative Strength Index) denoted by blue color and volume by gray color.

On y-axis price of stock is denoted and on x-axis month and year is denoted.

3. Quantitative Analysis

- A quantitative analyst is the one who designs a complex framework for financial institutions that aids them to price and trade securities in the financial market. Quants can be of two types:
- Front office quants - These are the ones who directly provide the trader with the price of the financial securities or the trading tools.
- Backoffice quants - These quants are there to validate the framework and create new strategies after conducting through research.

Following are the contents of Quantitative Analysis of NKE(Nike) company.

1. Mean
2. Median
3. Mode
4. Maximum
5. Minimum
6. Variance



Fig 5. Quantative Analysis

Pie Chart



Fig 6. Quantative Analysis pie chart

In above Fig 6 we see Variance value denoted in blue color, Maximum value in red color, Mean value in green color, Median value in violet color, Mode value in orange color and Minimum value in light blue color.

4. Forecasting Value

Forecasting is used for the time series to predict the value of stock with the terms of day like: - Short term (5-10 days), medium term (20-100 days) and long term (200 days).

NKE stock forecasts on testing set, testing error 17.41%.

Here we have produced result using forecasting values.

The Below Graph Show the predicted value for the previous data.

The Graph Shows value up to 24 April 2022.

Forecast Values

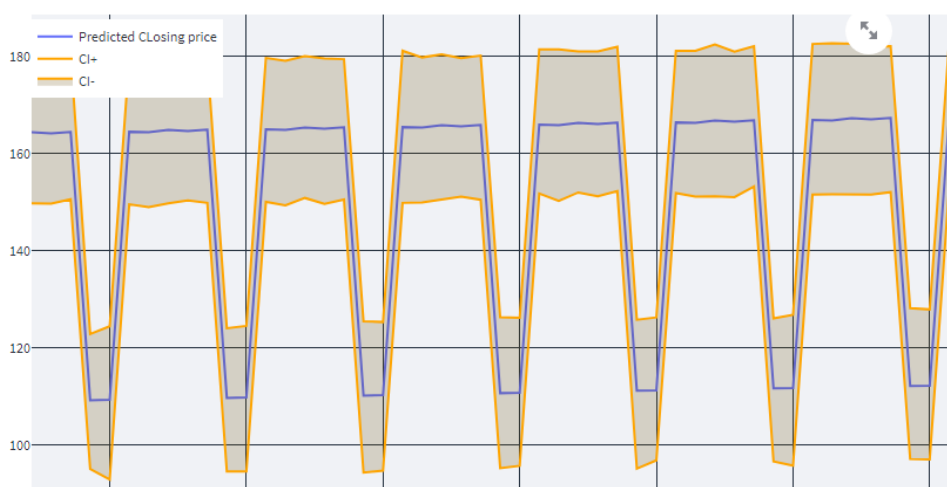


Fig 7. Forecast Value of next days



In above Fig 7 we see Predicted closing price denoted with blue color line, CL+ (closing positive) price which is above the predicted price and denoted with orange color and also, we see CL- (closing negative) price which is below the predicted price and denoted with orange color.

III. CONCLUSION

The prediction may be improved if we can train with a larger range of knowledge sets. Moreover, within the case of prediction of varied shares, there is also some scope of specific business analysis. We are able to study the various pattern of the share value of various sectors and might analyze a graph with a lot of completely different time span to fine-tune the accuracy. This framework loosely helps in marketing research and prediction of growth of various corporations in several time spans. Incorporating alternative parameters (e.g capitalist sentiment, election outcome, and government stability) may improve the prediction accuracy.

IV. REFERENCES

1. C Osman Hegazy, Omar S. Soliman, Mustafa Abdul Salam, "A Machine Learning Model for Stock Market Prediction" International Journal of Computer Science and Telecommunications [Volume 4, Issue 12, December 2013].
2. Kai-Yin Woo, Chulin Mai, Michael McAleer, Wing-Keung Wong "Review on Efficiency and Anomalies in Stock Markets" 22 December 2019; Accepted: 4 March 2020; Published: 12 March 2020.
3. Boriss Siliverstovs, Manh Ha Duong "On the role of stock market for real economic activity" JUNE 9 2006.
4. Fahad Almudhaf, Yaser A. AlKulaib OCTOMBER 2012
"ARE CIVETS STOCK MARKETS PREDICTABLE".
5. Hiransha M, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P "NSE Stock Market Prediction Using Deep Learning Models" 2018.
6. Pranav Bhat "A Machine Learning Model for Stock Market Prediction".