

RESEARCH ARTICLE | MARCH 21 2022

Stock market predictions using linear regression

Neha Titarmare ; Mrinmayi Wazarkar; Parth Palkar; Ravi Alwaria; Abhishek Shetye; Ritesh Khodiyar



AIP Conf. Proc. 2424, 050004 (2022)

<https://doi.org/10.1063/5.0076964>



CrossMark



APL Bioengineering
Special Topic:
Drug/Gene Delivery and Theranostics
Read Now!

 AIP
Publishing

Stock Market Predictions Using Linear Regression

Neha Titarmare^{a)}, Mrinmayi Wazarkar^{b)}, Parth Palkar^{c)}, Ravi Alwaria^{d)}, Abhishek Shetye^{e)} and Ritesh Khodiyar^{f)}

Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering and Research, Nagpur, India

a) Corresponding Author: nehatitarmare@gmail.com

b) mrinmayii.5419@gmail.com

c) palkarparth24@gmail.com

d) raviflux5@gmail.com

e) abhishekshetye2012@gmail.com

f) rkhodiyar14@gmail.com

Abstract. In the world of finance, activities related to stock exchange are perhaps considered important. The demonstration of trying to gauge the prospective assessment of a stock or other money related tool traded on a financial exchange is called as the stock market prediction or forecast. Share Market is a messy spot for anticipating since there are not any critical guidelines to assess or foresee the estimation of offer inside the share market. Numerous techniques like specialized investigation, principal examination, and factual examination, and so forth arrange to anticipate the value inside the share market nevertheless none of these approaches are incontestable as a faithfully worthy expectation instrument. This paper discusses about how AI methods can be utilized to anticipate the yield of a stock. Most stockbrokers utilize methodological and fundamental analysis, along with time series analysis, while making stock forecasts. The programming language is utilized to deliver stock market estimates. In this research work, we propose an AI (ML) methodology that will be trained from the available stock information and gain understanding and subsequently uses the obtained data for a definite estimate. In this context, this investigation uses an AI methodology called Linear Regression to anticipate stock prices for the gigantic and slight capitalizations, for example, using costs with both day by day and on the time frequencies.

Key Words: Stock Market, Machine Learning, Predictions, Linear Regression

INTRODUCTION

Stock Market forecasting and analysis is the demonstration of trying to come to a decision about the future or current estimation of organization stock or alternative financial instruments exchanged on a trade. The securities exchange may be a crucial piece of the economy of the country and assumes an imperative element in the expansion of the corporate and trade of a nation that eventually influences the economy of the nation. The investors and business are engaged with stock market and need to recognize whether some stock will rise or fall throughout a specific timeframe. The stock market is the most widely recognized route for a company to fund-raise for expansion. It depends on the idea of demand and supply. Assuming the interest for a corporation's stock is higher, the organization share cost increments and on the off chance that the interest for the company's stock is low, the organization share value decline.

A stock market forecast is an endeavor to anticipate the expected value of the stock exchange. The forecast is required to be hearty, precise, and effective. The framework should be intended to work in true conditions and ought to be appropriate to it. There are various methodologies available for carrying out the prediction system such as Fundamental Analysis, Technical Analysis, and Machine Learning, Market Mimicry, and Time-Series viewpoint. The

initial two procedures at estimating stock costs were principal and logical evaluations. Artificial neural networks (ANNs) are the most normally utilized method nowadays. Most of the time, ANNs experience the lacunas of an over-fitting concern due to the large number of boundaries to fix, and the slight earlier user statistics about the significance of the contributions to the studied issue.

In this research, we outline the design of the Neural Network model with its salient highlights and flexible boundaries. We have chosen a particular assembly of boundaries with reasonably an enormous effect on the share price of a business. Through the aid of statistical investigation, linkage between the chosen factors and share price is planned that may help in expecting precise results. In this paper, we talk about how various efforts have been taken in the use of ML to Stock forecasting and proposes new intense thoughts that can be worked upon. We additionally talk about a strategy for the estimation of stock-market opening and closing price. We endeavor to decide if the BSE market news in blend with the recorded statements can productively help in the computation of the BSE closing index for a given trading day.

LITERATURE REVIEW

Sadia et al. (2019) has discussed that, by estimating the exactness of different ML algorithms, the arbitrary forest algorithm is the utmost suitable for anticipating the market price of a stock dependent on numerous data points from recorded data. It will be a significant resource for stockbrokers and financiers expecting them to place resources into the securities exchange since it has been prepared on an immense variety of verifiable data and was picked in the wake of being tried on an example set of information. They utilized eleven attributes to depict the ascent and fall of stock costs. A portion of these credits are HIGH, LOW, OPENP, and CLOSEP. Different features incorporate LTP, YCP, VOLUME, Exchange and VALUE, yet the four referenced above assume a critical part in their discoveries [1].

Dhenuvakonda et al. (2020) in their study used a deep learning model and ANN to predict the stock price of INFRATEL organization. They used LSTM models to train four different types of networks. They anticipated that these models are sensibly effective at perceiving designs in the securities exchange space. This exhibits that there is a basic dynamic that is shared by all stock markets. They found that ARIMA and Auto Regression which are linear series models are univariate time-series based prediction models that are unequipped for perceiving the hidden elements utilizing multivariate time series. Based on the results; authors believe that deep learning models outperform AR and ARIMA models. As a result of the results obtained, ANN proved to be the superior performer when compared to the other three networks. Because only one window is used to predict the upcoming instance, ANN is quite capable of detecting unanticipated and unheralded changes in the system [2].

Muskaan et al. (2020) as per their findings described that, even though statistical methods are widely used for financial forecasting, there has been a significant shift toward the use of artificial intelligence and machine learning techniques. Among the different ML strategies, ANN and SVM are two of the most mainstream and generally utilized by researchers. At the same time, researchers have proposed combining these methods with GA and PSO to improve accuracy.

In the field of financial forecasting, statistical methods have their own importance and value; however, machine learning techniques are gaining popularity and, in many cases, providing better accuracy than other widely used methods [3].

Pahwa et al. (2017) conducted a review of work on stock prediction using Machine Learning. This paper talks about the best ML algorithms for such an application, just as the current tools and procedures that are most appropriate for its execution. The study recommends using linear regression and logistic regression for stock prediction and analysis, and SVM for precise results. One limitation of this hypothesis is that it must be deemed friendly by the dataset used in prediction. The paper provides a summary of the tools available for implementing algorithms in machine learning. All tools support regression and classification algorithms, so users can select any method based on their familiarity and convenience. The paper proposes a system for extracting knowledge from data and making predictions in order to advise consumers on investment decisions [4].

Sharma et al. (2017) has used Regression as the research technique in this study. Since monetary stock trades produce huge measures of information at some random time, an enormous volume of information should be analyzed before a forecast can be made. Every technique recorded under regression has benefits and demerits over its counterparts. The authors have mentioned linear regression as a valuable technique. Linear regression models are normally fitted using the least-squares methodology; in any case, they can moreover be fitted in another way, for

instance, by reducing the "lack of fit" in some other structure, or by decreasing a weakened variation of the least-squares loss function. The least-squares technique, of course, can be used to fit nonlinear models [5].

Zhang et al. (2018) proposed that though stock forecasting is a difficult task, there are several web techniques available to make it simple. Due to the interlinked data format, it will be easier to identify the different reasons between various variables and approximately outline the investment details. This pattern of investments in various workplaces has some similarities, and it leads to forecasting the stock market with accuracy using datasets. Stock forecasting can be done using analytical historical data and investor sentiment to analyze their mental conflict about investing in specific stocks [6].

The authors of this research work have used a ML algorithm to forecast the prices of a given stock. The information is derived from financial websites such as yahoo finance, Google finance, and others. Quandl is the name of the official website that provides data to researchers working on machine learning or neural networks. They have CSV files with datasets. For example, after entering data into the system, they receive predicted output.

As a trader, you will receive information about a specific date's open, close, low, and high. However, if one wishes to work with stock prediction, we can add some additional facts such as market mimicry to improve accuracy. The proposed model will compare all data to real-world data. The comparison will show the difference in accuracy and error between the real and predicted values. It may differ depending on the dataset because it may or may not be accurate. Typically, the accuracy of a system using linear regression must be greater than 82 percent. In this case, the system achieves greater than 97 percent accuracy in multiple tests [7].

Patil et al. (2016) proposed combining information from various worldwide business sectors with AI calculations to figure stock index movements. Their examination uncovers that the SVM algorithm works on enormous dataset of values collected from different financial markets of the world. Besides, SVM does not have the issue of over fitting. Correlation analysis uncovers a solid connection between the Marketplace stock index and worldwide business sectors that close just before or at the start of trading time. For anticipating the day-by-day pattern of Market stocks, different ML based models are proposed. The numerical results indicate a high efficiency. Their well-trained predictor serves as the foundation for a practical trading model. When compared to selected benchmarks, the model generates a higher profit [8].

Girija et al. (2015) have created facility techniques for exploitation, especially of online media information (social media), financial news, and investigation outcomes. A prediction or forecast model has been developed that utilizes big data analytical capacities, social media, analytics, and AI to forecast financial exchange drifts constantly. The model displays that sentiment investigation of social information supplements established technical examination approaches like regression analysis. It demonstrates that political and financial news, just as the impact of social media, sway the future performance and instability of the system's markets. With social media data in addition to numerical data improves estimates and the nature of the information. Big data technology usage enables real-time anticipations. Nevertheless, the sentiment analysis algorithm utilizes a cumulative assessment of the sentiments in a specific tweet or news bulletin story; this could be improved for enhanced end calculations, which would proliferate prediction precision [9].

Time series estimation issue was researched in different financial organizations' work centers. For stock market forecast, a prediction model reliant on independent and SVM analysis, named SVM-ICA, is suggested. ML is used to control an arrangement of time series analysis models. The SVM was made to deal with regression problems in time series analysis and nonlinear classification. An approximate function dependent on the risk lessening standard is utilized to limit the speculation error. As a result, the ICA technique extracts a variety of important features from the dataset. SVM is used to predict time series. The SVM model's output was compared to the ICA technique's output without any preprocessing [10].

Rajput et al. (2016) analyze the various procedures for stock market estimation using data mining and sentiment analysis in their examination. Opinion mining has become a functioning examination zone because of its value and individuals' requirements. As the volume of opinionated information develops, examining and summing up opinionated information turns out to be progressively significant. A wide range of kinds of opinionated summarization strategies are proposed to address these issues. Representative works incorporate probabilistic methodologies dependent on term statistics and heuristic methodologies dependent on predefined rules. Notwithstanding broad exploration, current stock forecast research has various constraints and opportunity to get better. At last, they infer that stock anticipating is an unpredictable assignment, and that different components ought to be considered to figure the market more precisely and effectively [11].

PROPOSED SYSTEM

System Design and Architecture

System Flow Diagram

The figure 1 depicts the flow diagram of our system which is explained as follows:

Data Source: In the figure 1, first the system is going to use Yahoo Finance data set as per the requirement. Here system use data of stocks up to the current date, which is useful for time series forecasting and after that it helps the system to predict the next price of the stocks and gives output to user.

Stream Lit: We are using stream lit for setting up webpage or web-application platform. Stream Lit it is an open-source application framework for machine learning and data-science teams. Stream Lit allows system for easy markdown style heading, text as well as embedding images and graphs too. Stream Lit makes it very easy and quick for system to build a machine learning application. To run stream lit files, we need to save as filename.py. In simple words, it explains, Stream lit helps machine learning project to run or host on a local web server.

Y. Finance: At Yahoo Finance, the framework can get free stock cites, most recent news, portfolio administration assets and worldwide showcase information. Y. Finance is an open-source library which provides financial and stock data for free, so our system uses Y. Finance dataset for setting up with the research.

Visual Studio: Visual Studio and Plotly directly links together through a straight connector as per representation. Here utilizing Microsoft visual studio, we create computer programs as well as websites, web applications, and web administrations. It is re-imagined and optimized for building and investigating mode web applications. It could be a streamlined code editor with development operations. Visual Studio Code could be a freeware source code editor.

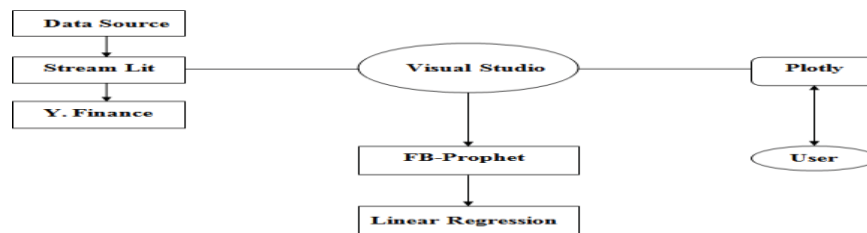


FIGURE 1. Flow Diagram of System

5) **FB-Prophet:** Face-book prophet and Machine learning data structure are working together in system, which helps for future forecasting. Face-book open-sourced its time-series forecasting tool called Prophet, which produces accurate forecasts as produced by skilled analysts with a minimum number of human efforts. The Face-book prophetis available in the form of API in Python.

6) **Linear Regression:** Linear Regression belongs to the category of supervised machine learning algorithm where

the expected yield is tenacious and highlights a consistent incline. It is utilized to calculate values inside a continuous range (for example: sales, price) rather than endeavoring to arrange them into classes. Here in our framework, we have applied linear regression. FB-prophet which is as an open-source system and can be customized is used in such way, so that it can apply basic machine learning algorithm uses linear regression.

7) **Plotly:** Directly links to visual studio it gives us graphical representation to user. However, system needs to import graph from Plotly. Imports give them access to imports packages.

8) **User:** Plotly and user relate to a bidirectional connector i.e., they are one to one as shown in figure 1. User or Viewer can access the web application as per their requirement.

MODEL CREATION

```

Welcome  main.py x
main.py
1 # pip install streamlit fbprophet yfinance plotly
2 import streamlit as st
3 from datetime import date
4
5 import yfinance as yf
6 from fbprophet import Prophet
7 from fbprophet.plot import plot_plotly
8 from plotly import graph_objs as go
9

```

FIGURE 2. Code

Figure 2 refers to a visual studio code terminal directory named as "stock predict" and file named "main.py" by using python extensions are created. Here, "#pip" "command" is used to import the development kits.

Raw data

	Date	Open	High	Low	Close	A
7...	2021-03-01T00:00:...	2,056.5...	2,086.5...	2,046.1...	2,081.5...	2
7...	2021-03-02T00:00:...	2,076.1...	2,104.3...	2,071.2...	2,075.8...	2
7...	2021-03-03T00:00:...	2,067.2...	2,088.5...	2010	2,026.7...	2
7...	2021-03-04T00:00:...	2,023.3...	2,089.2...	2,020.2...	2,049.0...	2
7...	2021-03-05T00:00:...	2,073.1...	2,118.1...	2,046.4...	2,108.5...	2

FIGURE 3. Raw Data

In the given figure 3, "Raw data " extracts current data of the present Stock scenario is shown with well labeled tags like (date , open price , close price, high price and low price) in table of rows differences and columns with tag contents.

Time Series data with Rangeslider

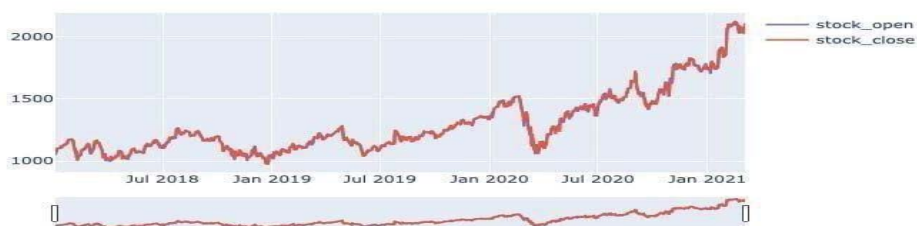


FIGURE 4. Analysis of the Opening and Closing prices from data frame

The above figure 4 depicts the variation of the opening and closing price with time interval. This stock chart is basically a price chart. It illustrates stock price which are plotted over a time frame, along with other important information. It displays price data (high, low, open and close) over a specific period. Looking at a daily stock chart this reflects the day's price change compared to previous closing price.

We are showing differences of 6 months. In Jan2021 stock is high and close price in 2000 and open between 1500 up to 2000 as compared to Jul 2020 which is closed at price below 1500 i.e. 1300 and open between 1000 and 1500. Similarly, we can see differences in Jul2020 and Jan2020.


```

9
10 START = "2018-01-01"
11 TODAY = date.today().strftime("%Y-%m-%d")
12
13 st.title('Stock Forecast App')
14
15 stocks = ('GOOG', 'AAPL', 'MSFT', 'GME', 'TCS', 'RIL')
16 selected_stock = st.selectbox('Select dataset for prediction', stocks)
17
18 n_years = st.slider('Years of prediction:', 1, 4)
19 period = n_years * 365
20
21 @st.cache
22 def load_data(ticker):
23     data = yf.download(ticker, START, TODAY)
24     data.reset_index(inplace=True)
25     return data
26
27
28 data_load_state = st.text('Loading data...')
29 data = load_data(selected_stock)
30 data_load_state.text('Loading data... done!')
31

```

FIGURE 5. Sorting the dataset on date time and filter “Date” and “Time”

Figure 5 represents the backend coding that describe the Datasets and stock scripts selected to provide the desired output with the checked time gap.

```

45
46 # Predict forecast with Prophet.
47 df_train = data[['Date', 'Close']]
48 df_train = df_train.rename(columns={"Date": "ds", "Close": "y"})
49
50 m = Prophet()
51 m.fit(df_train)
52 future = m.make_future_dataframe(periods=period)
53 forecast = m.predict(future)
54

```

FIGURE 6. Building and training with Prophet

Figure 6 represents the data frame train that is "df_train" which commands the program to execute the forecasting process with the help of prophet toolkit. Prophet then starts the algorithmic calculation and analyzing the data of the stocks provided on the internet which then helps to forecast the stock price [12].

```

54
55 # Show and plot forecast
56 st.subheader('Forecast data')
57 st.write(forecast.tail())
58
59 st.write(f'Forecast plot for {n_years} years')
60 fig1 = plot_plotly(m, forecast)
61 st.plotly_chart(fig1)
62
63 st.write("Forecast components")
64 fig2 = m.plot_components(forecast)
65 st.write(fig2)

```

FIGURE 7. Visualizing the forecast data

As per the Figure 7, the Plotly starts it works on the given code. It represents "st" as structure which drives the Plotly tool kit analysis. Plotly predicts data and forecast components and creates a chart or graph of objects in a systematic visualized interface.

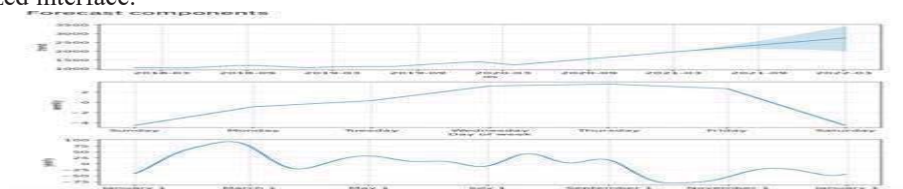


FIGURE 8. Analysis of the forecast component

In figure 8, forecast component shows all three charts. It shows the trend line of the stock price, whereas the shaded region shows the variation as the program is not 100% accurate. All three charts are dilation of the time interval with respect to (presently, weekly, and yearly).

RESULT

Our model can select number of years to be predicted through a bar for the script of stock database. It is a slider to a command, which is number of years to be predicted. The component of the stock is related to current scenarios. The data frame Plotly chart help us to determine the current changes in opening and closing prices of stock and gives us detail analysis which is shown in figure 9.



FIGURE 9. Changes in opening and closing prices

Forecast component drives us to analyze the time synchronization differentiation trend on weekly and yearly basis as shown in figure 9.

CONCLUSION

Deciding the Stock exchange estimates have constantly been testing work for business investigators. We device the Artificial Neural Network to the task of stock trade estimation and ANN model and striking component. Our underlying examination shows a huge relationship between various information boundaries. The expectation is genuinely exact except if there is an enormous and abrupt variety inside the real information. This, research work applies the ML technology of neural networks to stock price forecasts and gets a best outcome, which will give the examination of the stock market development a novel idea and we endeavored to utilize huge text-based information to foresee the securities exchange indices. By assessing the exactness of the numerous computations, we suggest that the best possible fitting estimation for foretelling the cost of a stock dependent on number of information focuses on the irrefutable information is the random forest algorithm. As it is trained on a wide choice of recorded statistics and has been picked after being checked on a model data, the algorithm can be a crucial tool for investors and financial bankers, brokers with regards to placing assets into the securities exchange [13,14]. The research shows the ML model to foresee the stock worth with more precision when diverged with the recently executed machine.

REFERENCES

1. Stock Market Prediction Using Machine Learning Algorithms International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-4, April2019.
2. Padmaja Dhenuvakonda1, R. Anandan2, N. Kumar3, “Stock Price Prediction Using Artificial Neural Networks” Journal of Critical Reviews, Vol 7, Issue 11, 2020, ISSN- 2394-5125.
3. Muskaan, Pradeepta Kumar Sarangi, “A Literature Review on Machine Learning Applications in Financial Forecasting”, [Journal of Technology Management for Growing Economies](#), Vol. 11, No. 1, April 2020, pp. 23–27
4. Pahwa, N., & Vora, D. “Stock Prediction using Machine Learning a Review Paper” [International Journal of Computer Applications](#), 163(5), 36-43 (2017).
5. Sharma, D. Bhuriya, U. Singh, “Survey of Stock Market Prediction Using Machine Learning Approach”, International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017.
6. Xi Zhang, SiyuQu, Jieyun Huang, Binxing Fang, Philip Yu, “Stock Market Prediction via Multi-Source Multiple Instance Learning.” IEEE 2018.
7. N. Verma, B. Mohapatra, “Stock Market Predication Using Machine Learning”, International Conference on Communication and Information Processing, ICCIP-2020.

9. S.Patil, K. Patidar, M. Jain, "A Survey on Stock Market Prediction Using SVM", International Journal of Current Trends in Engineering & Technology, Volume: 02, Issue: 01 (JAN-FAB 2016).
10. Girija V Attigeri, Manohara Pai M, Radhika M Pai, Aparna Nayak. "Stock market prediction: A big data approach", TENCON 2015 - 2015 IEEE Region 10 Conference, 2015.
11. Hakob Grigoryan "Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA)", DSJ 2016.
12. Vivek Rajput, Sarika Bobde, "Stock Market Forecasting Techniques: Literature Survey", IJCSMC, Vol. 5, Issue. 6, June 2016, pg.500 – 506
13. Sukhman Singh, Tarun Kumar Madan, Jitendra Kumar, Ashutosh Kumar Singh. "Stock Market Forecasting using Machine Learning: Today and Tomorrow", 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT), 2019.
14. Nisal Waduge, Upeksha Ganegoda "Forecasting Stock Price of a Company Considering Macroeconomic Effect from News Events", 2018 3rd International Conference on Information Technology Research (ICITR), 2018.