Turkish Journal of Computer and Mathematics Education Vol.12 No.12 (2021), 1673-1678

Research Article

Stock Price Estimation Based On Historical Information And Textual Sentiment Analysis

Abhay Gupta[a], Aman Gupta[b] ,Anshul Chaudhary[c] , Rajesh Kumar Yadav[d]

[a,b,c] Department of Computer Science & Engineering, Delhi Technological University, New Delhi, India
[d]Assistant Professor Department of Computer Science & Engineering, Delhi Technological University, New Delhi, India
email: [a]nextresp@gmail.com, [b]amangupta.dtu@gmai l.com, [c]canshul100@gmail.co m

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 23 May 2021

Abstract: Stock Market Price estimating has been a subject of interest among experts and scientists for quite a while. Stock costs are difficult to foresee due to their unstable nature, which relies upon an assortment of political and financial variables, change of administration, speculator feelings, and numerous different things. Foreseeing costs dependent on verifiable or literary information alone has end up being lacking. A fruitful assessment of future share costs may capitulate a critical benefit. A stock trade hypothesis is the obligation to select future assessmeARDA_Publication_11897 Papernts of an entity's share or other currency-related items traded on a market. But there are methods and technologies that are supposed to allow us to get future price details. However, not a single good forecasting model has succeeded in beating the market trend further. As per the timetable information custom, theory is regularly founded on previous verifiable information and market patterns, authentic connection information and hypothesis can be determined

Keywords: Stock Price Estimating, Financial Blogs, Sentiment Analysis, Prophet

1. Introduction

In this precedented times, when people losing jobs, exhausting savings and struggling to cope with the economic challenges there are facing. Speculation is a protected choice for getting their future yet one doesn't have a clue where to contribute and the amount to contribute on the grounds that one couldn't say whether there will be a benefit or a misfortune on one's venture. This raises a fascinating issue on the grounds that a great many people for the most part wind up putting resources into any of the financial exchange areas. The answer for this issue permits us, find out about financial exchange choices and assisting with settling on it more precise choices. Stock market is most volatile and dynamic marketing system and to address this problem of chaotic and dynamic stock market we not only minimize our prediction based on technical factors[18] rather considering, analysing and pre-processing the fundamental factors such as customer ratings, brand value, financial news sentiment, etc. of companies.

Stock market forecasts[7] are an endeavour to anticipate what's to come, the estimation of an organization's stock or other monetary instrument available to be purchased to trade. stock exchange forecasts are moreover alluring tests. As indicated by a compelling business sector theory, stock costs ought to follow an arbitrary travel example and thusly ought not be the case can be anticipated with in excess of 50% precision.

Appropriate stock anticipating[8] can prompt extraordinary advantages for both the dealer and the purchaser. Frequently, it is expressed that estimates are more turbulent than irregular, which implies that they can be anticipated via cautiously

investigating securities exchange history. We in this way need a framework that can foresee the market esteem near the apparent worth, consequently expanding precision. Introduction of essential highlights like client suppositions, monetary sites, news, and so on nearby stock anticipating has pulled in numerous analysts as a result of its proficient and exact estimations.

In our project, we attempt to improve the exactness of stock worth gauges by get- together a ton of time game plan data. Apart from technical factors, we will also take into account fundamental factors like brand value, customer sentiments, ratings, performance graph, etc. to reflect upon the dynamic changes in the stock market and eventually provide stable and accurateresults.

An essential goal of this task is to share the scholarly comprehension of more exact and exact financial exchange expectation[15].

To develop an accurate automated system for precise stock value prediction to have higher return on investment and thusreducing risk and making greater profit.

To address the problem of chaotic and dynamic stock market we not only minimize our prediction based on technical factors rather considering, analysing and pre-processing the fundamental factors such as customer ratings, brand value, financial news sentiment, etc. of companies.

2.Related Works

Different procedures in the stock assumption space[16] can be orchestrated into two social affairs. The essential social event joins tallies that attempt to improve the introduction of figure by redesigning the

supposition models, an assortment of contraptions has been utilized, including Support Vector Machine[4], LSTM, etc., while the sub optimal of assessments bases on improving the highlights subject to which the supposition that is in the primaryget-together of the calculations thatemphasis on the guess models.

2.1..Stock Price Prediction Based On Ann

In 2011, a connection between the introduction of ANN plus, SVM was finished. Every count has its technique for learning models and subsequently expecting. Artificial Neural Network (ANN) is a notable and later procedure which similarly combine particular examination for making estimates in financial business areas. ANN[5] incorporates a bunch of edge capacities. These capacities prepared on recorded information subsequent to associating each other with versatile loads and they are utilized to make future expectations. ANN fuses a lot of edge limits. These limits arranged on recorded data ensuing to partner each other with flexible burdens and they are used to make future assumptions. ANN can consider as a computation or a mathematical model which is animated by the utilitarian or essential ascribes of natural neural associations. These neural associations are made so it can remove plans from riotous data. ANN[4] first train a system using a huge illustration of data known as getting ready stage then itfamiliarizes the association with the data which was avoided from the arrangement

stage, this stage known as endorsement or assumption stage.

2.2..Stock Price Prediction Based On Lstm

LSTM is that where the data having a spot with the past state drives forward. The main purpose behind using this model in protections trade figure is that the assessments depend on a ton of data and are all around dependent on the drawn-out history of the market. So, LSTM[3] guides jumble up by giving a manual for the RNNs through holding information for additional organized stages making the measure more careful. In this way approving itself as by and large more solid wandered from different strategies.

3.Proposed Method

Our approach is to examine the historical data of stocks of different companies and using that dataset to train our proposed model. After obtaining the processable data we select relevant features from this large dataset that may impact the prediction. It's an optimisation step and has a lot of importance as a good feature engineering leads to lesser time and space complexity ofthe project and more manageable system[14].

Also, our model expects feature vectors derived from the processed data as input to get trained and make predictions. To yield excellent results, a set of technical as well as fundamental features could be used so that our prediction could get an overall exposure to all expectations that could affect the price of the stock.

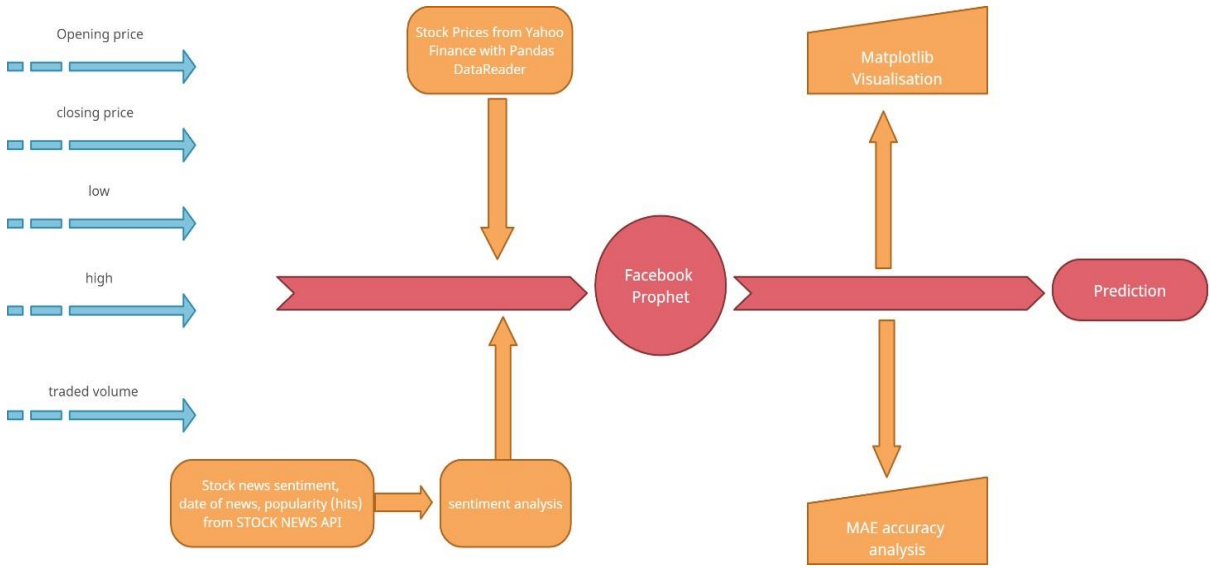

Figure 1 : Model Flow Chart

3.1.Pre-Necessities

- Timeline testing can be done on a PC / workplace or on servers.
- To carry out the mentioned work, we will utilize the Jupyter Notebook. In the event that the product can now be imported, you need to launch and configure the Jupyter Notebook Python 3 pad.
- Operating on larger databases has increased memory and will need a system with any 2GB of working primary storage capacity to execute computations.

3.2.News Information:

Presently, official information sources to get authentic monetary news are restricted. Most sources like News API[2] just permit admittance to the most recent couple of long stretches of information free of charge and it would cost roughly ₹70000 to assemble all the information important for this venture. I had the option to discover a news Programming interface, *The Stock News API*, giving news to stocks at similarly lesser cost and is a basic HTTP

REST Programming interface that empowers us to get the most recent stock news from different uplifting news sources from web. We used the API to get news, date of publish of the news, sentiment of the news[2] and its popularity of companies in the stock market.

3.3.Numerical Measures:

Opening Value - The underlying expense is the expense at which the security is first sold opening of the exchange upon the appearance of trading. The beginning exchanging cost of any stock is its own day by day opening cost. Cost open key imprint for the business movement of the day, particularly for those intrigued estimating transient outcomes like informal investors.

Closing Value - The end cost of the stock is the standard seat used to follow its exhibition over the long haul. The end cost is considered as the last an incentive at which the stock sold during the ordinary exchanging period day.

Low Price - The lowest price is the least selling cost in stock at the most recent day

exchanging. Today's minimum is the lowest internal security trade Price.

High Price - Higher stock price means higher stock trading price. Today's top price is where the stock is sold at highest during trading day course. Top prices are often greater than closing or starting prices of stocks.

Volume - Volume alludes to the quantity of offers exchanged a given time span. A stock's volume alludes to the quantity of offers that are sold, or exchanged,throughout a specific timeframe (normally every day).

Stock Split - A stock split[10] is the point at which an organization separates the current portions of its stock into numerous newoffers to help the stock's liquidity.

A stock split or stock separation expands the quantity of offers in an organization. A stock split causes a decline of market cost of individual offers, not causing a difference in absolute market capitalization of the organization.

Dividend - It is a movement of advantages by an undertaking to its financial backers. Exactly when an association secures an advantage or overabundance, it can pay a degree of the advantage as a benefit to financial backers.

Here, including fundamental features is the tricky part as to get the sentiment of the stakeholders is not an easy task to execute. Albeit certain other basic highlights are effectively accessible in the library yfinance itself, for example, net revenues, profit, revenue on valuation, additional information for deciding an organization's capability for subsequent development,incomes and basic worth.

Our implementation includes sentimentanalysis of the stakeholders and the news

related to the stock to get hold of the market and make predictions to higher accuracies even when external factors affect the stock price.

It is seen that generally ARIMA[1] is best utilized for expectation yet it isn't most appropriate for non-straight information designs. It gives best outcomes for time strategy which has a couple season(s) of authentic information and solid occasional impacts. Incredibly, the center fundamental Information Science group of Facebook organization distributed a fascinating new strategy as of late called by name Prophet, connects with information

fashioners and analysts to test or perform surveying in Python at scale. Prophet[1] is really a strategy to perform determining of time arrangement information mostly dependent on added substance model dissimilar to non-straight patterns that in everyday works with day by day, week after week, and yearly too for irregularity, in addition to occasions.

Thus, we are using fbprophet[1] library which gives a Prophet which is a method for determining measurement information maintained an added substance model where non-direct examples are fit with step by step, after quite a while after week, and yearly abnormality, notwithstanding occasion impacts. Prophet gives precise outcomes even if there should be an occurrence of missing data and movements in the pattern, and usually handles inconsistencies well. It works best with measurements that have strong occasional effects and different various times ofrecorded data.

Prophet is really an added substance framework that has the accompanying:

$$\mathbf{m(t) = n(t) + o(t) + p(t)+ \epsilon} \quad (1)$$

- **n(t)** represents the trend(s), which errands long stretch decay or extension in data. Prophet has two example related model(s), a piecewise straight model and a drenching improvement model, which depends upon the sort of guess.

- **o(t)** represents Fourier approach with inconsistency, that chooses how the information will be affected considering season related factor(s) like the season

- **p(t)** represents the special times of year effect[11] or enormous occasions which profoundly impacts business time arrangement information (e.g., The day after Thanksgiving, New Item Dispatch, Superbowl and so forth)

- **ϵ** addresses a mistake term which is final.

We generate Visualisations[6] of the predictions in the form of graphs and data frames to get a better insight of the predictions and analysis. This step is not a necessary one but is important in practical usage.

3.4.Performance Analysis using MAE:

This is the final step in the sequence wherethe evaluation of the model[12] takes place which is very helpful in:

1. Determining the risk factor involved in using the prophet and

2. Analysing our model in comparison with other models based on technical analysis only or any other algorithm.

In the case of our model, we are using Mean Absolute Error as a performance metric to determine the accuracy of the model.

Mean absolute error is calculated as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |x_i - x| \quad (2)$$

where x represents the actual price and xi represents the anticipated cost and n represents the quantity of questions or expectations made.

4.Result and Analysis

We trained our model using technical factors such as opening and closing values, previous day prices etc. and sentiment analysis of fundamental properties to capture dynamic nature of stocks. Here, we were able to make predictions for the share values of Google for a span of 30 days starting from 16th October 2020 till 27th November 2020. The result is depicted in the below graph:

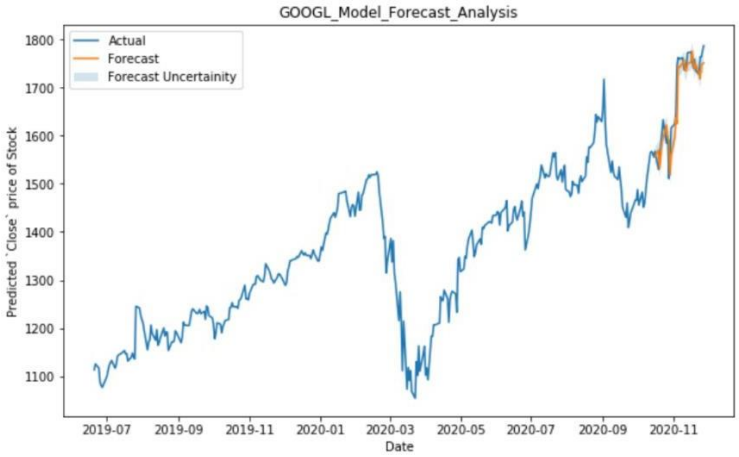

Figure 2 : Predicted closing price vs Dates mentioned

Below table depicts the forecasted values of all the factors we took into consideration for making speculations. Although time periodwas from 16th October 2020 to 27th

November 2020 as mentioned above but below figure displays result for last ten days:

```
In [112]: forecast.tail(10)
```

Out[112]:

	ds	trend	yhat_lower	yhat_upper	trend_lower	trend_upper	Close_lag_1	Close_lag_1_lower	Close_lag_1_upper	Close_lag_2	...	weekly	w
20	2020-11-13	679.423509	1731.560805	1767.647013	679.423509	679.427412	393.493299	393.493299	393.493299	62.947036	...	63.242683	
21	2020-11-16	679.495258	1733.556780	1772.087405	679.495258	679.499727	404.683666	404.683666	404.683666	62.680167	...	63.558914	
22	2020-11-17	679.519174	1758.103089	1794.449862	679.519174	679.525006	405.356462	405.356462	405.356462	64.461708	...	63.885502	
23	2020-11-18	679.543091	1739.741495	1776.091396	679.543091	679.553962	400.654544	400.654544	400.654544	64.568819	...	64.546251	
24	2020-11-19	679.567007	1719.298726	1754.389763	679.567007	679.578265	392.664693	392.664693	392.664693	63.820259	...	63.891305	
25	2020-11-20	679.590924	1730.305532	1765.629516	679.590924	679.603487	399.479981	399.479981	399.479981	62.548250	...	63.242683	
26	2020-11-23	679.662673	1714.747425	1749.802974	679.658277	679.683326	391.045434	391.045434	391.045434	63.633265	...	63.558914	
27	2020-11-24	679.686589	1699.744034	1736.771137	679.680467	679.710099	387.692913	387.692913	387.692913	62.290459	...	63.885502	
28	2020-11-25	679.710505	1730.260097	1763.216656	679.703039	679.738286	401.505979	401.505979	401.505979	61.756728	...	64.546251	
29	2020-11-27	679.758338	1733.074671	1768.332612	679.745873	679.790455	401.593396	401.593396	401.593396	63.955810	...	63.242683	

10 rows × 94 columns

Figure 3 : Predicted values of all the factors considered

5.Conclusion and Future Work

The mean absolute error for our model is 27% initially, that is, we were able to forecast the share values for Google for said period with an accuracy of 73%. But later by including fundamental features such as profit margins, return on equity, earnings, analysis of stock values in news, etc.; we further increased the accuracy of our modelto 78.2% , that is, reducing the absolute errorto 21.8%.

We intended to train our model on both technical factors as well as basic indicators like financial analysis in news, year of establishment, turnover, public sentiments etc.

Our work can be additionally broadened utilizing deep learning for predicting stock values for multi-national organization yet this may fuse huge space and time complexity than our proposed approach

References

1. M. B. Patel and S. R. Yalamalle, “Stock price prediction using artifi-cial neural network,”International Journal of Innovative Research inScience, Engineering and Technology, vol. 3, no. 6, pp. 13 755–13 762,2014.
2. E. J. De Fortuny, T. De Smedt, D. Martens, and W. Daelemans, “Eval-uating and understanding text-based stock price prediction models,”Information Processing & Management, vol. 50, no. 2, pp. 426–441,2014.

3. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of arimaand artificial neural networks models for stock price prediction,"Journalof Applied Mathematics, vol. 2014, 2014.
4. K. Kohara, T. Ishikawa, Y. Fukuhara, and Y. Nakamura, "Stock priceprediction using prior knowledge and neural networks,"IntelligentSystems in Accounting, Finance & Management, vol. 6, no. 1, pp. 11–22,1997.
5. Y. E. Cakra and B. D. Trisedya, "Stock price prediction using linearregression based on sentiment analysis," in2015 international confer-ence on advanced computer science and information systems (ICACSIS).IEEE, 2015, pp. 147–154.
6. S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai,"Combining technical analysis with sentiment analysis for stock priceprediction," in2011 IEEE ninth international conference on dependable,autonomic and secure computing. IEEE, 2011, pp. 800–807.
7. X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, "News impact on stockprice return via sentiment analysis,"Knowledge-Based Systems, vol. 69,pp. 14–23, 2014.
8. Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based onsentiment analysis and lstm,"Neural Computing and Applications, pp.1–17, 2019.
9. S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anas-tasiu, "Stock price prediction using news sentiment analysis," in2019IEEE Fifth International Conference on Big Data Computing Serviceand Applications (BigDataService). IEEE, 2019, pp. 205–208.
10. M. Lam, "Neural network techniques for financial performance predic-tion: integrating fundamental and technical analysis,"Decision supportsystems, vol. 37, no. 4, pp. 567–581, 2004.
11. Adebiyi, C. K. Ayo, M. O. Adebiyi, and S. O. Otokiti,"Stock priceprediction using neural network with hybridized market indicators,"Journal of Emerging Trends in Computing and Information Sciences,vol. 3, no. 1, pp. 1–9, 2012.
12. R. Madhuri, M. Chinta, and V. P. Kumar, "Stock market predictionfor time-series forecasting using prophet upon arima," in2020 7thInternational Conference on Smart Structures and Systems (ICSSS).IEEE, 2020, pp. 1–5.
13. S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D.C. Anas-tasiu, "Stock price prediction using news sentiment analysis," in2019IEEE Fifth International Conference on Big Data Computing Serviceand Applications (BigDataService). IEEE, 2019, pp. 205–208.
14. Parmar, N. Agarwal, S. Saxena, R. Arora, S. Gupta, H. Dhiman,and L. Chouhan, "Stock market prediction using machine learning," in2018 First International Conference on Secure Cyber Computing andCommunication (ICSCCC). IEEE, 2018, pp. 574–576.
15. B. Qian and K. Rasheed, "Stock market prediction with multipleclassifiers,"Applied Intelligence, vol. 26, no. 1, pp. 25–33, 2007.
16. L. Sayavong, Z. Wu, and S. Chalita, "Research on stock price predictionmethod based on convolutional neural network," in2019 InternationalConference on Virtual Reality and Intelligent Systems (ICVRIS). IEEE,2019, pp. 173–176.