Stock market prediction based on social media

MD. Abdula rain
Abhay Bhargav
Abhishek Bishowakarma
Department of computer Science application
Sharda university Gr.noida Uttar Pradesh , India

Mr. Amit Kumar
Department of computer science engineering
Sharda university
Gr. Noida,uttar Pradesh India

Abstract:
Stock price forecasting is a well-known and crucial issue. With a successful stock prediction model, we may get insight into market behavior over time, detecting trends that else might not be found. Artificial intelligence will be a successful strategy to deal with this difficulty as a result of the computer's growing computational power. of prediction with use of Algorithms' premise is that if we have all of the knowledge regarding today's stock market, the price will be predictable. With the rise of the Internet, social media, and online social interactions, obtaining daily user projections has become a viable task. We focus mostly on the Social Media influence on stock market like Twitter and Instagram , we also Take Notice of How celebrities actions and words effects Stock Marketing

Keywords: Machinelearning, Algorithms, Social networking,social media
DOINumber:10.48047/nq.2022.20.22.NQ10328 **NeuroQuantology2022;20(22):3302-3306**

INTRODUCTION

The aim of predicting the stock market is to estimate how much a company's financial stocks will be worth in the future. Using the values of the most recent stock market indexes, artificial intelligence is a method used to create predictions. Stock market prediction is the process of attempting to predict the eventual value of a shares or other financial asset listed on a financial exchange. This study describes how stock performance predictions using algorithmic learning may be made. Among the most important activities in the banking industry is stock trading. Stock market prediction is the process of attempting to anticipate the future value of a stock or another economic instrument traded on a stock exchange.. Since online conversation as well as financial news data may affect investor behaviour, machine learning algorithms may be employed to forecast stock markets. In the present research, we use algorithms to investigate how social networking sites and finance news data affect stock market forecast accuracy over a predetermined time period, boosting a company's performance and satisfying its investors. We use algorithms to investigate how social networking sites and business news data affect the precision of stock market forecasts over a certain time frame.

I. LITERATURE REVIEW

New stock market techniques and models are being developed in an increasing body of literature. Previous research had a strong focus on a sizable population utilising social information to predict consumers' attitudes towards a brand[4]. A variety of research demonstrated that a mix strategy can enhance a categorization method [5]. The most typical use of emotional analysis is the examination of tweets from Twitter to highlight the most popular market trends [6]. Examining tweets and Facebook postings allows for the application of sentiment analysis in sales forecasting for a product[7]. Sentimental analysis was done on tweets with stock ties that were gathered over a 6-month period[8]..A selection of tweets with tags from the top 100 firms was taken into consideration in order to eliminate noise. A dataset of 2,500 tweets was used to train a Naive Bayes classification algorithm on each individual tweet. Results showed that stock volume and trading volume are

connected to sentiment indicators and anomalous returns, respectively[9].

To create new investment predictors, Sentimental Analysis was used to news headlines and tweets that were taken directly from Twitter[10]. They selected a random sample of the data and classified every message as either upward or downward depending on whether it contained those phrases. They demonstrated that two indicators of Twitter mood, the frequency of financial process[13].Building a social behaviour graph on people's online activity using sentimental analysis is a common way to uncover the Trading activity and stock price volume are correlated [14].Additionally, sentimental analysis was carried out.nData from SentiWordNet was retrieved using a mixed selection technique to demonstrate how consumer demand is influenced by market developments.[15].

METHODOLOGY

This project's goal is to create an website that generates precise suggestions in a quantitative way. Three modules are used for this, and they are as follows:

Machine learning Module

This module's goal is to produce the Stock Prediction value. The strength of the gap between the starting and closing prices determines the stock prediction value. To do this, we must forecast the stock's closing price. By using machine learning on the stock's historical data, this is accomplished.

Sentiment Analysis Module

This module's objective is to extract the sentiment value from the most recent news items about every stock then output the mean as mood value to the fuzzy module.

Data Processing

This Adj Close price of the top 5 technological companies—Apple, Google, Amazon, Alibaba, and Microsoft—were our initial five fundamental predictions. It is important to note that even if our predictions for Apple's ("AAPL") daily returns were accurate, we still wanted to include some information about the company in our model.After that, we calculate each of their daily returns because the measure is typically used to describe a return on investment [16]. It is calculated. as follow:

$$Ret = AdjR - AdjR - 1$$

$$AdjR - 1$$

The Fixed Share close Priceon the ith day less the Fixed stock Closing Price for the (i-1)th day multiplied by the phrases on the platform, and statistical significance as predictors of typical market returns. A microblogging service that is solely focused on the stock market was also subjected to sentimental analysis[11]. 62,100 blog entries were gathered from stocktwits.com. for three months' worth of time. J48 classifier, a machine learning algorithm, was used to classify the posts' sentiment in order to create a learning model. Services and an accuracy of .99% was found in the

Affected stock Closed Value on the (i-1)th day equals the amount returned on the it day.A stock's adjusted stock closing price is the close price of the stock after dividends have been taken into account. This measure is frequently used in Returns calculations. 'AAPL' is utilized as an example to highlight our work in this article and as the primary focus of the proposed research solution.However, any individual This programme can be used by stock market companies.focusing on in indexes like the Dow or S&P 500 would prove ineffective in this situation since there is a lack of pertinent text data produced from Twitter. The historical daily pricing of these five firms are relatively simple to obtain.To manage the download, Python offers practical libraries, like as pandas. The following format may be used to draw down and format the data into a data frame, as shown in Table 1: Following the download of five data frames, the daily return was calculated We combine the daily return data from each of them to create our own market pricing data frame, which looks like this: As can be seen, there is a lot of information, and some tweets even missing important details like created_at. The

Table 2. Processed data Frame

Date	Return_AAPL	Return_GOOG	Return_BABA	Return_MSFT	Return_AMZN
2017-08-14	0.015050	0.009055	0.019183	0.0150340	0.015816
2017-08-15	0.010948	-0.000488	0.020309	0.000273	-0.000570
…	…	…	…	…	…

challenges arise while trying to individual Twitter statistics. The sole language utilised in our research is English, and the only subject we focus on is the date that tweets are posted. We have eliminated

noise, such as useless symbols and characters, and only kept the pertinent data, mostly the date and content, to make data analysis easier.

Table 1. Historical Prices of AAPL

Date	Open	High	Low	Close	Adj Close	Volume
Sep 06, 2017	162.71	162.99	160.52	161.91	161.91	21,651, 762
Sep 05, 2017	163.75	164.25	160.56	162.08	162.08	29,468, 500
…	…	…	..	…	…	…

Fuzzy logic Module

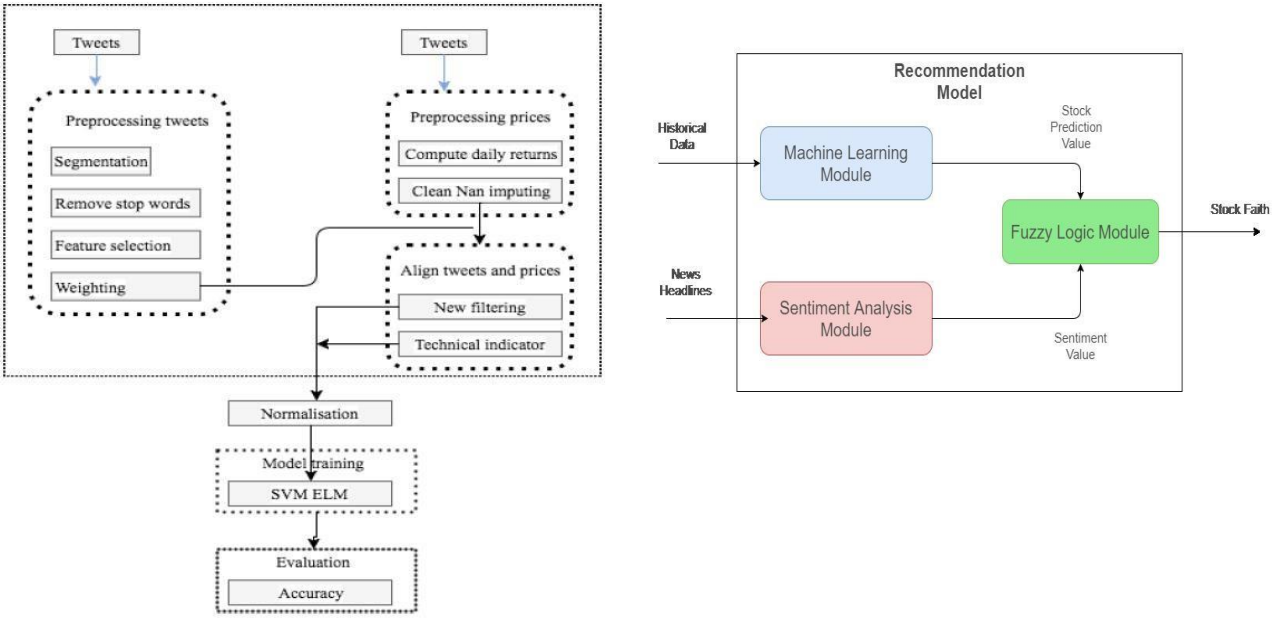

Figure 1. Stock volatility prediction framework

EXPERIMENTAL RESULTS

This module's goal is to provide Stock Faith, a measure of the strength of a recommendation. This module's activation guidelines are as follows:The stock will have high confidence IF the reported sentiment was positive or the stock forecast value was positive.The stock faith will be average IF the stock's prediction value was average. If stock predictions were weak and news sentiment was negative, stock confidence will be low.

On August 17, Ali baba released its latest financial report, noting that the rise of its online sales had allowed it to exceed revenue projections. This Motivating news may become popular subjects among regular people who will tweet about it and offer their own predictions regarding the organization. As a result, BABA_P/N is expected to considerably grow in the days to come. We are yet unable to integrate these results from the aforementioned analysis to make a final forecast. Before using the Support Vector Machine (SVM) classifier to generate a forecast because we have two distinct types of data (once a day returns and positive/negative), we must normalize the data. For pa'sklearn.processing' package.

Prediction Table

Date	Returns	Predicted returns
15/08/2017	1	1
16/08/2017	-1	-1
17/08/2017	-1	-1
18/08/2017	-1	1
21/08/2017	1	-1
22/08/2017	1	-1
….	…	…

CONCLUSION

We have suggested a novel learning model in this research to analyse stock trend employing tweets and past information, and we have also compared it with existing traditional approaches, which frequently focus on just one element (such as historical information or mood analysis). We conduct the tests using tweets from the previous two months and historical pricing from the previous three years. Results show that when accuracy in prediction is considered, the recommended model regularly outperforms the competition.. This model also investigates the possibility of using additional characteristics to boost prediction accuracy. Since fluctuations in time are not taken into account in this research, certain monetary parameters, such as 2 . -day goes back or 5-day refunds, can be gave as new features.This methodology may also be used to analyse and apply certain other relevant information not related to the market for stocks, such as corporate news and professional analysis.

REFERENCES

[1] "About Twitter, Inc", https://twitter.com/?lang=en, accessed on 25th, September, 2017.

[2] "Twitter Passed 500M Users in June 2012, 140M of Them in US; Jakarta ''Biggest Tweeting' City". TechCrunch. July 30, 2012.

[3] "200 million Tweets per day." Twitter Engineering, Twitter Official Blog. June 30, 2011.

[4] "Behavioural Economics", LifeTrends Issue 18, September 2016.

[5] Fama, Eugene, F, e. a. International Economic Review 10, 1-21. 1969

[6] Fama, E. F. (1991) Journal of Finance 46, 1575-617.

[7] H. Cootner, P. The random character of stock market prices.
MIT, 1964

[8] D. Pyle, Data Preparation for Data Mining. Morgan Kaufmann Publishers, Inc, 1999.

[9] B. Baumohl, The Secrets of Economic Indicators: Hidden Clues to Future Eco- nomic Trends and Investment Opportunities. FT Press, 2012.

[10] L. S. Goldberg and C. Grisse, Time variation in asset price responses to macro announcements. Technical report, 2013. [11] Fama, E. F. , The Journal of Business 38, 34- 105, 1965 [12] Johan Bollen and Huina Mao, e. a., Twitter mood predicts the stock market. arXiv, June 2015.

[13] Deepti Chopra, Master Natural language Processing with python, Packt Publishing, 2016, ISBN 978-1-78398-904-1 [14] http://scikit-learn.org/stable/tutorial/machine_ learning_map/, accessed on 25th, September, 2017. [15] "Twitter Developer Documentation", https://dev.twitter.com/rest/public/search, accessed on 25th, September, 2017.

[16] "return: definition of return in Oxford dictionary (British & World English)"

[17] C. Hsu and C. Lin, A comparison of methods for multiclass support vector machines, IEEE Transactions on Neural Networks, vol. 13, no. 2, pp. 415-425, Mar 2002. [18] L. Gonzalez-Abril, F. Velasco, C. Angulo, J.A. Ortega, A study on output normalization in multiclass SVMs, Pattern Recognition Letters, Vol. 34, no. 3, pp. 344-348, 2013.

[19] V. NarayananIshan, A, Bhatia, Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model, IDEAL 2013, pp 194-201.

[20] Y. Ma and G. Guo (eds), Support Vector Machines Applications, Chapter 2, Springer International Publishing Switzerland 2014.

[21] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008. ISBN: 0521865719.

[22] Porshnev, Machine learning in Prediction of Stock Market
Indicators Based on Historical Data and Data from Twitter Sentiment Analysis, 2013 IEEE 13th International Conference on Data Mining Workshops, Dec. 2013, pp.440-444

[23] Pagolu and Venkata, e. a., Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. Arxiv ID: 1610.09225. Cornell University.

[24] Anthony R.C and Ariel, E. A., Prediction Model of the Stock
Market Index Using Twitter Sentiment Analysis, International Journal of Information Technology and Computer Science, 01 October 2016, Vol.8(10), pp.11-20.

[25] F. Wang and Y. Zhang, Q. Rao, K. Li and H. Zhang, Exploring mutual information-based sentiment analysis with kernel-based extreme learning machine for stock prediction