Various Techniques for Stock Price Prediction based on Twitter Data and Public Sentiments

Shivani Parekh[1], Siddhivinayak Kulkarni[2]

[1,2]Department of Computer Engineering, Dr.Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India

Abstract - **Prediction of stock market prices do not solely depend on only one or two factors, there are many other factors accountable for the same. Technical indicators such as moving averages, stochastic oscillators, RSI are few factors on which stock market prices depend. Sentiments of people are also playing a huge part in the stock prices going high or low. Public opinions about a stock or a product of a company can be taken from various platforms such as twitter or facebook etc. This study aims to predict the stock prices of few nifty 50 companies using twitter sentiment analysis and ensemble modeling. Firstly, how twitter sentiments are driving the stock prices up and down shall be analyzed and then prediction of closing price of stocks of companies listed in nifty 50 will be done.**

Index Terms - **Ensemble learning, Indian Stock market prediction, linear regression, machine learning, random forest, SVM, Twitter sentiment analysis.**

I.INTRODUCTION

A stock or stake is a financial tool that reflects a company or corporation's ownership and a roughly equal claim on its properties and revenue. Stock market is a place where individual and institutional buyers come together in a shared space to purchase and sell shares. The truth is that investing in the stock market entails risk, but it is one of the most successful ways of building up one's net worth when approached in a balanced manner.

According to a survey in 2017, 32% of the people suffering from depression, anxiety have faced the stock market loss. Whenever stock market crashes, a significant increase in rate of suicides is noticed. Assumption of stock growth rate is still one of the tasks which is not correctly implemented by any current systems. People can profit a lot if prediction of stock can be made correctly. This will give traders better idea of stock market and can think more

efficiently about their investment plan. Better prediction will enable companies to plan their product launch, business strategies and can help in avoiding losses or at least recovering from the losses. Public sentiment on social media about a particular company has seen to have an impact on the closing price of the company's shares on the following days. Analysis of emotion is a method used to determine a sentence's degree of positive or negative connotation. Using this methodology, it is possible to examine thousands of posts mined from microblogging sites such as twitter related to the stock of a certain business and the resulting observations can be used to train the model, which should work better in turn. Technical analysis can be carried out by the derivation of technical metrics from raw price data. Indicators play a crucial role in market forecasting, regardless of whether a trader is a beginner or expert. The indicators offer valuable market trend knowledge and help you increase your returns. The indicators indicate where the price will go next, as the name indicates. There are 4 indicators which indicate the market trend such as momentum, volatility, volume, and trend. Moving Averages provides statistics on the momentum of the market, market trends, pattern reversals, stop loss and stop loss points. The moving average helps traders to recognize trading possibilities in the context of the latest trend in the market. The volatility in the market is demonstrated by Bollinger bands. The Relative Strength index (RSI) is a measure of momentum. It is a continuous line that extends from 0 to 100, showing whether the stock in the market is overbought or oversold. The Commodity Channel Index describes the market's emerging trends. It contains 0, +100 and -100 values. The stochastic oscillator is also a measure of momentum. The oscillator measures the closing price of a stock over a period with a range of values. The key objective of this work is to establish

experimentally whether the sentiments drawn from tweets impact the market closing prices of Nifty 50 companies and to forecast the stock prices using this sentiment score and technical indicators as feedback to the model of the ensemble.

The remainder of this paper is arranged as follows: The relevant work is addressed in Section II. Our suggested approach is explained in Section III and IV.

II. LITERATURE REVIEW

Because of the various controlling factors such as investor sentiment, firm results, eco- nomic factors, and social media feelings, forecasting the stock market remains a difficult challenge. The literature argues that, considering the increase in the accuracy of stock prediction, prediction accuracy can be further enhanced beyond its current measure by searching for new sources of knowledge, especially on the Internet. In the forecasting area, numbers of renowned computer science researchers have often compared various related works with their results. Conducted their experiment using stock dataset of NSE from Quandl's API and tweets from twitter. 6 months stock data was taken into con- sideration from August 1st, 2018 to January 31st, 2019.and SVM and ExtRa trees were used as ML techniques, their outputs fitted to stacked regressor. Their studies found that the ensemble model performs best on a dataset spanning 5 months for the given case study. In most cases, the SVM meta-regressor is stronger than the ExtRa tree regressor. Analysis of sentiment has an overall negligible effect on the forecasts of the ensemble [1]. The research indicates that prices influence the stock market and that markets can be forecasted using those prices. Apple, Microsoft, Google, and Amazon companies' data are considered from yahoo finance API and are used to test effect of stock prices on market. They have used linear regression model for the prediction of stock prices and SVM algorithm for classifying each sentence as positive, negative, or neutral. The accuracy achieved for the Apple Company's is 81.3488 %, 98.2878 % for Google Company, 98.2186 % for Microsoft Company and 98.2186 % for Amazon Company [2]. Two separate textual representations, Word2vec and N gram, are used for the study of public feelings in tweets in the paper. In this article, a total of 2, 50,000 tweets on Microsoft were extracted from twitter API over the

period from 31 August 2015 to 25 August 2016, tweets extracted from twitter were subjected to sentiment analysis and supervised machine learning concepts over the period from 31 August 2015 to 25 August 2016 and analyzed the correlation between the stock price fluctuations of a company and tweet feelings. It is seen at the end of the paper that there is a clear link between the rise and decline of stock markets and the feelings of the public of tweets [3]. In this paper, to forecast the Indian stock market movements, a regular prediction model is suggested using historical data and news stories. To categorize the news text with negative or positive emotions, Classifier Naïve Bayes is used. For predictive purposes and an accuracy ranging from 65.30 to 91.2 percent obtained with different machine learning strategies, the count of the positive and negative emotions of news stories for each day and variation of neighboring days close price along with historical data is used [4]. The time series data prediction in this paper is carried out by the classification techniques SVM, AdaSVM, Naïve Bayes, AdaNaive. The data obtained from www.datamarket.com were used for the study of the Historical Time Series. The training data consisted of 3777 records and the dataset for testing consisted of 2500 records. The data set is split into several classes, one for preparation and another for evaluating the algorithms for classification. The findings reveal that both the SVM and Naïve Bayes algorithms have higher output values when matched with AdaBoost than the SVM and Naïve Bayes algorithms [5].

Using sentiment analysis on the tweets gathered using the Twitter API and even the closing values of different stocks obtained from yahoo API, they have attempted to construct a method that predicts the movement of stock prices for different companies. Such a prediction would significantly assist a future stockholder in making smart decisions that would lead directly to his gain. In predicting the stock price trend, in lieu of the emotions of the tweets, SVM proved to be the most effective and feasible model [6]. Sentimental analysis is performed in this paper on the data for top five companies i.e. Apple, Microsoft, Oracle, Google, and Facebook are collected in form of comments and tweets from Stock Twits. To measure the mood of the user's tweet, the data is analyzed. Such opinions are split into four categories that are happy, up, down and rejected. To forecast the outcomes, the polar- ity index along with market data is supplied to

an artificial neural network [7]. Their experimental setup with stock data (January 2010 to September 2019) of three (3) companies listed on the Ghana stock exchange showed that using public sentiments the stock market is predictable. Based on individual and combined datasets, the improvement in accuracy achieved by the proposed model in forecasting future market price for 1 day, 7 days, 30 days, 60 days and 90 days ahead indicates that the accuracy of stock prediction models can be significantly improved with market data amalgamation [8]. They showed sentiment analysis for the stock market in this paper by fetching Sensex and Nifty live server data values at various time intervals that can be used to predict stock market status. Python script with sleep count time interval of 1 second for the fetching data, and values were calculated for different intervals. Their results showed that for a particular time interval the fetched values of Sensex and Nifty remained con- stant [9]. In this paper, an exhaustive analysis was undertaken between the flow of news opinion and stock indices. Lexical based sentiment analysis was selected for the performance of Sentiment Analysis [10]. AFINN lexicons were used. AFINN list is   a web-focused collection that comprises web-jargons, slang and obscene expressions. The list includes manually gathered and graded valence-rated English terms from minus five (negative) to plus five (positive). It analyzed the interaction of opinion with other macro-economic influences. They used 30 + twitter handles to gather news. The study showed that Sensex and Nifty are well correlated and that exchange rate and oil and gas rates are moderately correlated [11]. This paper continues with Big Data Analytics, the notion of big data. This paper then included the diverse aspects of Data Analytics and classical and technical stock price prediction techniques. Then there is a study of numerous approaches that can be implemented along with Big Data Analytics to forecast stock price prediction. Survey concludes a sufficiently large dataset containing a mixture of historical and opinion data to produce test data then extracted and normalized with the right methodology and to collect the data according to the required predictive analysis process, machine learning system or deep learning method and finally com- pared it with a good number of test data and compared the outcome with the relevant method [12]. To forecast the market performance of the National Stock Exchange, this study uses the feedforward network (NSE). This study further strengthens the claims made by previous studies that the Nifty-100 index can be predicted by using appropriate market parameters and training the ANN using the Error Backpropagation Training Algorithm. For the testing dataset, the backpropagation algorithm achieved a maximum accuracy of 77.80 % [13]. In this paper, using eight regression and eight classification methods, they have presented several approaches to stock price and movement pre- diction on a weekly forecast horizon. Machine learning and deep learning approaches are the basis of these models. These models were built, fine-tuned, and tested using NIFTY 50 daily historical data from January 2, 2015 to June 28, 2019. The efficiency of the predictive framework focused on the fuzzy neural network SOFNN algorithm is found to be the highest among all models in its ability to reliably forecast the stock price movement of NIFTY 50. The research has conclusively demonstrated that social media public opinions act as a very powerful feedback for stock price movement in predictive model construction [14]. In this article, they focused on finding ties between a single company/Twitter-based index's sentiment analysis and its short-term market performance using large-scale tweet data collection. Their studies indicate a high as- sociation between stock values and Twitter emotions (up to 0.88 for returns). To show that their estimated returns offer a high value of Rsquare (0.952) with low Max Absolute Percentage Error (MaxAPE) of 1.76 percent for Dow Jones Industrial Average (DJIA), they have introduced the Expert Model Mining Method (EMMS) [15]. In this study, a taxonomy of computational approaches to stock market analysis and forecasting was suggested, a comprehensive literature review of state-of-the-art algorithms and methods widely applicable to stock market prediction was presented, and some of the ongoing problems in this field were addressed that demand further effort and offer opportunities for future growth and research [16]. Further literature survey is shown in table below:

Table I. Summary of various techniques

Paper Title	Techniques	Predicted Parameter	Performance Metrics
Efficacy of News Sentiment for Stock Market	KNN, SVM, Naïve Bayes,	future price movement	65.30 to 91.2 % Accuracy

Prediction [4].	Neural network		
Prediction Models for Indian Stock Market [17]	Boosted Decision Tree, Logistic Regression, Sentiment Analysis, SVM.	Trend for the next day	70% of accuracy.
Stock Market Prediction Using Machine Learning Techniques [18]	Recurrent Neural Network (RNN), Long-short Term Memory Neural Networks	Closing Price of Stocks	For 30 days of prediction-Twitter sentiment: 0.6292, web news sentiment: 0.6367search engine query models: 0.6702Ensemble model: 0.6702.
The Prediction of Stock Index Movements Based on Machine Learning [19]	Support Vector Machine, Random Forest	S&P Stock index prediction	F1 score SVM-0.91RF-0.92
Stock Market Prediction with Historical Time Series Data and Sentimental Analysis of Social Media Data [20]	Sentiment Analysis, LSTM.	Closing Price of Stocks	96.95 accuracy and 3.05 percentage error.
Prediction of Stock Market using Machine Learning Algorithms [21]	Supervised learning algorithms: Logistic regression, Random forest, KNN, ARIMA.	Future stock market prices	Accuracy in forecasting the stock market directions. Logistic Regression: 74%, Random Forest: 70%, KNN: 64%, ARIMA model: 47%
Prediction of Stock Market using Ensemble Model [5]	SVM, Naive Bayes, boosting algorithm-Adaboost.	Closing Price of Stocks	SVM- 93.86% AdaSVM-94.33% NB-88.32% AdaNB-97.19

III.PROPOSED ARCHITECHTURE

A. Data Sources

Data for stocks will be extracted using Yahoo finance API [22]. Tweets from twitter will be taken using Twitter API using tweepy python library [24]. Tweet preprocessor along with NLP techniques will be used for sentiment analysis [23]. Open, High, Low, Close, Volume, Dividends and Splits will be the stock data extracted from yahoo API. Required features will be kept and remaining will be preprocessed. Other additional features fill be technical indicators which will be calculated based on its formula.

3.1. Architecture

Stock Data of few companies of NIFTY50 are extracted. This Data collected from the yahoo finance API will be pre-processed i.e., missing values will be filled up using pandas fillna method/imputer library of python sklearn package.
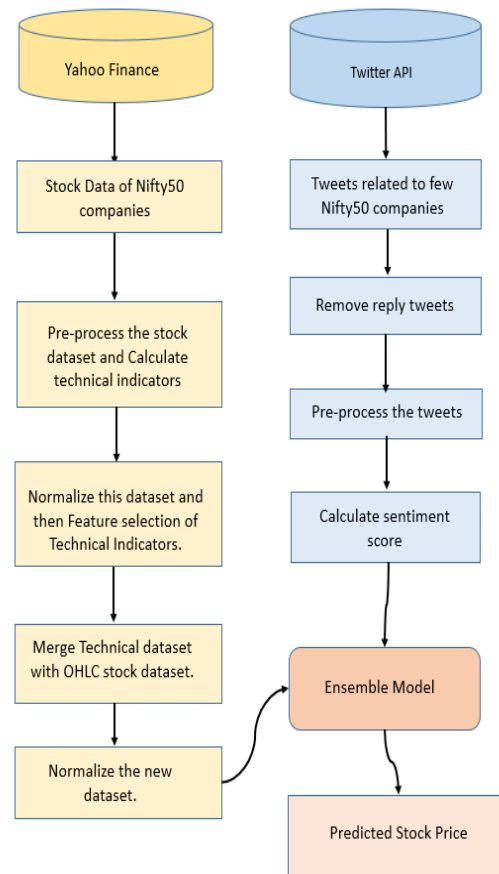

Fig. 1. Proposed architecture

After pre-processing stage, technical indicators such as ROC, Bollinger Bands, Stochastic oscillators, moving averages, Channel Commodity Index, Relative Strength Index, and Price Volume Trend will be calculated. After calculating technical indicators, only few technical indicators have impact on closing price of stocks. So, feature selection technique will be used to select only those features which are more contributing to closing price of stock. Then merge the technical indica- tors dataset with original fetched dataset. Now normalize the features which are not in normalized form. These features are fed to Ensemble model.

Tweets on everyday basis will be extracted using twitter API. Tweets extracted in the raw form are quite unclean. It contains URLs, emoticons, regex, mentions etc. So, we need to preprocess it. URLs, regex, mentions, emoticons etc. will be removed. Also, tweets containing replies shall be removed too.to remove redundancy of tweets. Next, sentiment score of tweets will then be calculated using NLP techniques. Sentiment score of tweets from every day will be calculated and then the sentiment score and stock dataset will be merged and then analysis of sentiment score and closing prices will be done. Merged dataset will then be given as input to ensemble model. Supervised Machine Learning Techniques such as SVM, Random Forest, KNN, and Naïve Bayes will be used as prediction algorithms for ensemble model.

IV. CONCLUSION

Predicting the stock prices is a challenging task when there are a lot of factors affecting the price of the stock. Many researchers have concluded in their study that sentiments of the people play an important role in driving the closing price of company shares. Our study is also focused in that area, analysis of public sentiments will be done to see how the stock prices are affected by public opinion. Stock price prediction of company is our second goal. Using sentiment analysis and various technical indicators, prediction of closing price of stock shall be done. Prediction of stock price is a regression problem which requires greater accuracy for model to be successful as this will solve the problem of many investors. So, a model which can correctly predict the closing prices will be helpful. Many researchers have applied different methodology for predicting the stock price so by using ensemble model, which uses more than one algorithm, will give combined output of two or more ML algorithms thus accuracy will be better.

REFERENCES

[1] U. Pasupulety, A. Abdullah Anees, S. Anmol and B. R. Mohan, "Predicting Stock Prices using Ensemble Learning and Sentiment Analysis," 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), Sardinia, Italy, 2019, pp. 215-222,

[2] B. L. Pooja, S. Kanakaraddi and M. M. Raikar, "Sentiment Based Stock Market Prediction," 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Belgaum, India, 2018, pp. 12-17,

[3] V. S. Pagolu, K. N. Reddy, G. Panda and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements," 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, India, 2016, pp. 1345-1350,

[4] S. Kalra and J. S. Prasad, "Efficacy of News Sentiment for Stock Market Prediction," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 491-496, doi: 10.1109/COMITCon.2019.8862265.,

[5] Narayanan, B. & Govindarajan, M. (2015). Prediction of Stock Market using Ensemble Model. International Journal of Computer Applications.128.18-21.10.5120/ijca2015906412

[6] T. Mankar, T. Hotchandani, M. Madhwani, A. Chidrawar and C. S. Lifna, "Stock Market Prediction based on Social Sentiments using Machine Learning," 2018 International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, India, 2018, pp. 1-3, doi: 10.1109/ICSCET.2018.8537242.

[7] S. K. Khatri and A. Srivastava, "Using sentimental analysis in prediction of stock market investment," 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2016, pp. 566-569,

[8] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "Predicting Stock Market Price Movement Using Sentiment Analysis: Evidence from Ghana," Applied Computer Systems, vol. 25, no. 1, pp. 33–42, 2020.

[9] A. Bhardwaj, Y. Narayan, Vanraj, Pawan, and M. Dutta, "Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty," Procedia Computer Science, vol. 70, pp. 85–91, 2015.

[10] S. Kumar, S. Maskara, N. Chandak, and S. Goswami, "Empirical Study of Relationship between Twitter Mood and Stock Market from an Indian Context," International Journal of Applied Information Systems, vol. 8, no. 7, pp. 33–37, 2015.

[11] F. Å. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," arXiv.org, 15-Mar-2011. [Online]. Available: https://arxiv.org/abs/1103.2903

[12] A. Awate and B. Nandwalkar, "A Survey on Various methods for Stock Prediction using Big Data Analytics", AJCT, Apr. 2019.

[13] A. Kale, O. Khanvilkar, H. Jivani, P. Kumkar, I. Madan and T. Sarode, "Forecasting Indian Stock Market Using Artificial Neural Networks," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697724.

[14] S. Mehtab and J. Sen, "A Robust Predictive Model for Stock Price Prediction Using Deep Learning and Natural Language Processing," SSRN Electronic Journal, 2019.

[15] T. Rao and S. Srivastava, "Analyzing Stock Market Movements Using Twitter Sentiment Analysis," Analyzing Stock Market Movements Using Twitter Sentiment Analysis | Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), 01-Aug-2012. [Online]. Available: https://dl.acm.org/doi/10.5555/2456719.2456923

[16] D. Shah, H. Isah, and F. Zulkernine, "Stock Market Analysis: A Review and Taxonomy of Prediction Techniques," International Journal of Financial Studies, vol. 7, no. 2, p. 26, 2019.

[17] A. Nayak, M. M. M. Pai, and R. M. Pai, "Prediction Models for Indian Stock Market," Procedia Computer Science, vol. 89, pp. 441–449, 2016.

[18] N. Sirimevan, I. G. U. H. Mamalgaha, C. Jayasekara, Y. S. Mayuran and C. Jayawardena, "Stock Market Prediction Using Machine Learning Techniques," 2019 International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka, 2019, pp. 192-197, doi: 10.1109/ICAC49085.2019.9103381.

[19] S. Wang, "The Prediction of Stock Index Movements Based on Machine Learning," Proceedings of the 2020 12th International Conference on Computer and Automation Engineering, 2020.

[20] M. Kesavan, J. Karthiraman, R. T. Ebenezer and S. Adhithyan, "Stock Market Prediction with Historical Time Series Data and Sentimental Analysis of Social Media Data," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 477-482, doi: 10.1109/ICICCS48265.2020.9121121.

[21] N. Bharadwaj and A. Ansari, "Prediction of Stock Market using Machine Learning Algorithms.," International Research Journal of Engineering and Technology, pp. 5994–6005, May 2019.

[22] Yahoo Finance API, https://in.finance.yahoo.com/

[23] tweet-preprocessor, https://pypi.org/project/tweet-preprocessor/

[24] "Tweepy":https://tweepy.readthedocs.io/en/v3.5.0/api.html

[25] S. Bird, E. Klein, and E. Loper, Natural language processing with Python. Beijing: OReilly, 2009.

[26] "pytrends," PyPI. [Online]. Available: https://pypi.org/project/pytrends/.

[27] T. M. Mitchell, Machine learning. New York: McGraw Hill, 2017.