

# Stock Market Prediction Using Twitter Data Sentiment Analysis

Ankit Madhavi

Master of Computer Application

Bharati Vidyapeeth's Institute of Management and Information Technology

CBD Belapur, Navi Mumbai 400614, Maharashtra, India

madhaviankit0@gmail.com

Dr.Kirti Muley

Master of Computer Application

Bharati Vidyapeeth's Institute of Management and Information Technology

CBD Belapur, Navi Mumbai 400614, Maharashtra, India

Ikirti13@gmail.com

**Abstract** – Stock market forecasting is a well-known problem of interest. In today's virtual environment, social media platforms such as Twitter are ideal for expressing public opinion. Using the information accessible on these sites, highly accurate and relevant stock insights can be provided. This study examines various ways for dynamically learning about the market and its tendencies. For this article, three alternative models were employed and ran sentiment analysis on tweets about the company or the stock; the model with the least error is the best and most recommended way for prediction. The results have provided us with a clear polarity picture of the tweets as well as the stocks based on Twitter data, both statistically and graphically.

**Keywords** – Consumer key, Consumer Secret, Access token, Access token Secret, Long short-term memory (LSTM), Support vector regression (SVR), Random forest regression, Root Mean Squared Error, R-Squared, Twitter, Yahoo Finance,

## I. INTRODUCTION

The Dutch East India Company, the world's first stock exchange, was established in 1602. The world's largest stock markets are in the United States, United Kingdom, Japan, India, China, Canada, Germany (Frankfurt Stock Exchange), France, South Korea, and the Netherlands, with the world's largest stock markets in the United States, United Kingdom, Japan, India, China, Canada, Germany (Frankfurt Stock Exchange), France, South Korea, and the Netherlands Since 1602 and till now, the sole purpose of a stock trader has been to make the most money possible. People began testing and analyzing numerous tactics in order to build various trading strategies that would yield the best results. It is frequently stated that prediction is chaotic, implying that it may be predicted by thoroughly examining the history of the individual stock market. Machine learning is a good technique to express these types of operations. Social media has evolved into a mirror that reflects people's ideas and opinions on any given event or piece of news. Any public emotion toward a firm, whether positive or bad, can have a ripple effect on its stock prices. Stock market values can be forecasted for a specific firm by running sentiment analysis on all of the tweets accessible in real-time and extracting them in raw form from Twitter's real-time timeline.

To begin, get the Twitter developer account credentials, which include the Consumer key, Consumer Secret, Access token, and Access token Secret. After entering the developer's option through Twitter, all that's left is to extract the user's real-time tweet in its raw format.

Some yahoo financial data was also utilised. Using this

training data and the provided tweets, our model will be trained and develop the potential to make stock predictions in the future. Long short-term memory (LSTM), Support vector regression (SVR), and Random forest regression will be used.

## II. Literature survey

Following a review of the literature, one can infer that machine learning algorithms for stock market prediction are widely used around the world. Machine learning delivers considerably more accurate prediction since it is real-time and uses a large pool of data to do so, and it outperforms current prediction techniques.

Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, and Babita researched stock market prediction and their accuracy in findings after training the data received from Twitter data.[1]

Stock Market Prediction Using Machine Learning , a study by Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, and Lokesh Chouhan, used yahoo finance data available internationally to use and predict all stock prices after training the data models.

Tejas Mankar, Tushar Hotchandani, Manish Madhwani, Akshay Chidrawar, and Lifna C.S Stock Market Prediction based on Social Sentiments using Machine Learning discussed how to analyze data from various platforms and identify future stocks with the best returns.[2]

Stock Market Prediction Using Machine Learning , by Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, and Lokesh Chouhan, discussed the various innovative ways to apply machine learning algorithms to predict the stock market, as well as unique ways to get all relevant information.[3]

Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty, by [7] Aditya Bhardwaj, Yogendra Narayan, Vanraj, Pawana, and Maitreyee Dutta, explained how the stock market might be predicted using only Sensex and Nifty data.

### III. Proposed System

#### A. System overview

This paper describes in detail how a company's sentiment is recorded in raw form via tweets, how it is processed and cleaned, how data preprocessing is performed, how stock predictions are made based on the polarity rate of sentiment, and how a graphical representation is created on a website. Figure 1 depicts the flow chart for the fundamental process.

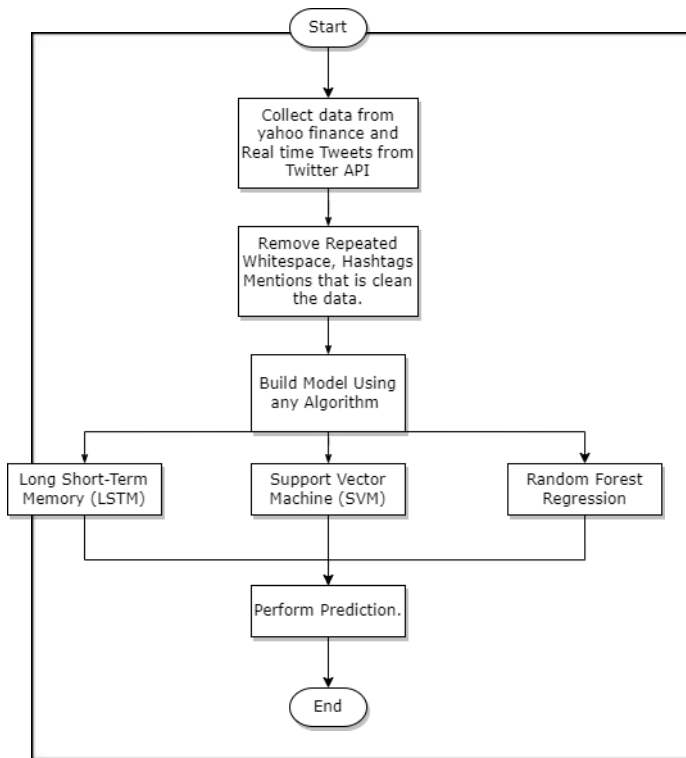


Fig. 1. System overview (Overview of process)

#### B. The Proposed Method

1) *Collection of data:* Information can be gathered in a variety of ways. It must be ensured that the data collected should be accurate and of high quality. It's also critical to have a vast pool of data; it's been proven that having a significant amount of legitimate data produces extremely accurate outcomes. In this post, it is examined the data using Yahoo Finance, which allows us to define the time range for which data is needed. Real-time raw data from users' are acquired and trends' Twitter timelines, excluding rebroadcast content and only including tweets in English or a single language.

2) *Data Preprocessing:* Once all of our data gathered, the next and most crucial stage is data preparation, which involves cleaning up the raw data so that it can be used. This is a critical step since clean data has a direct impact on the model's prediction accuracy. All of the tweets we've gathered via the Twitter API need to be cleaned up; we'll delete any whitespace that won't help us achieve our goal. The existence of hashtags and mentions in the text lengthens the phrase, making the algorithm less capable of classifying the sentiments. Even emoticons in tweets aren't taken into account in order to achieve more accurate results.

Text pre- processing was done using Regular Expressions after the removal of such tweets. The model extracts all tweets that contain either the firm name or the stock name for our trials.

3) *Testing the accuracy of different models :* We utilized three alternative techniques to test our models: Support Vector Machine, Long-short term memory (LSTM), and random forest regression. We found the value of Root mean squared error and R-Squared of varied time durations using all three algorithms. All of the time periods produced extremely diverse results. We used recent time durations of 6 months and 1 year to get more reliable results. To compare recent market movements to general patterns about a stock, the stock's total performance was taken into account. For Root Mean Squared Error, the rate for SVR was 0.19, whereas the rate for Random Forest was 0.2007 and the highest for LSTM was 0.201, and for R-Squared, the rate for SVR was 0.27, the highest, LSTM was 0.2365, and Random Forest was 0.2393. The statistics for a year show that the lowest Root Mean Squared Error for SVR was 0.09, the highest 0.15 for Random forest and the lowest 0.09 for LSTM, and the lowest 0.09 for R-Squared SVR, LSTM 0.53, and Random forest 0.78. SVR is 0.05, LSTM is 0.10, and Random Forest is 0.09 for all-time prediction, and SVR, LSTM, and Random Forest are 0.94, 0.81, and 0.822 for R-Squared.

#### C. Result and Evaluation

After reading over the data, we can see the model outcomes for the various algorithms that we employed in figure 2 down below. In terms of Root Mean Squared Error, LSTM provided the best long-term results, whereas SVR provided the best short-term results. In terms of short-term performance, we can state that LSTM and SVR are superior.

Algorithm		SVR	LSTM	Random Forest
Root Mean Squared Error:	6 month	0.1955339257	0.20111691	0.2007478071
	1 year	0.09267928968	0.1574356427	0.1058215327
	Full year	0.05521824889	0.1016819029	0.09927538587
R-Squared	6 month	0.278342369	0.2365437678	0.2393434857
	1 year	0.8376090441	0.5313996941	0.7882884556
	Full year	0.9451735811	0.814086062	0.8227820177

Fig. 2. Result of Various algorithm in various time frame

We can plainly see how SVR predicts prices in the graph below, where blue represents real prices and orange represents a prediction made using the SVR algorithm for one year with a high prediction accuracy. Take a look at Figure 3. We

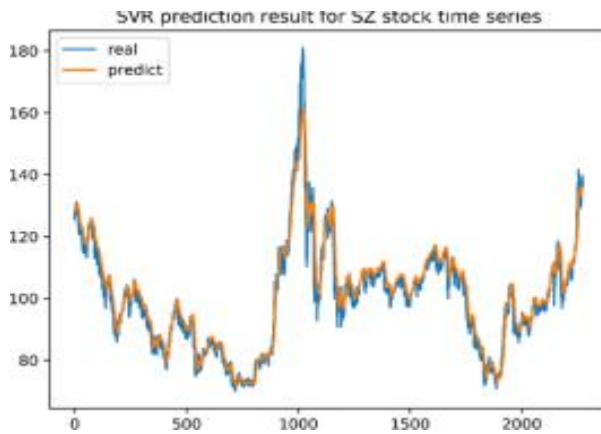


Fig. 3. Result of SVR algorithm

Figure 4 shows that the outcomes of LSTM data for Actual vs. Prediction are not as excellent as SVM, but they are better than Random forest. The data in Figure 4 is for a year, with red representing real prices and blue representing predictions, and as seen that the findings for actual and prediction are not correct.



Fig. 4. Result of LSTM algorithm

When real and expected data are compared, it is clear that Random Forest does not produce the desired results (see Figure 5). This graph also takes a year's worth of data into account, with red representing real prices and blue representing predictions. As seen that there is a significant discrepancy between the actual and predicted values, indicating that this model is not as accurate as other models.

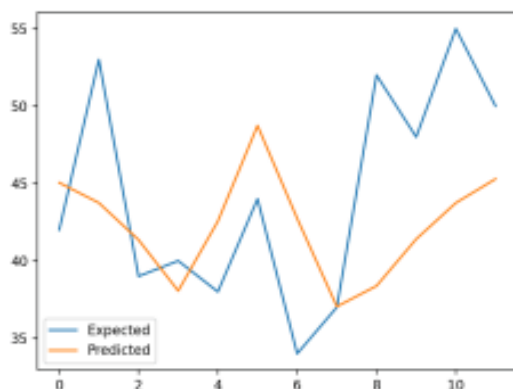


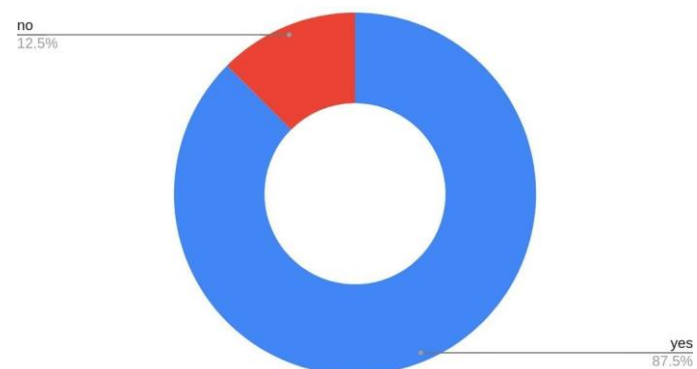
Fig. 5. Result of Random Forest algorithm in

#### IV. Conclusion

The accuracy of three stock price prediction algorithms, SVR, LSTM, and Random Forest, was investigated and compared in this research. To gain access, Tweepy is used, a Python package. By gaining access to token keys, the Twitter API can first do sentiment analysis on tweets. The polarity rate is calculated after the data from the Twitter timeline has been analysed. Yahoo Finance data is used in this paper to get a better and more accurate prediction. The sentiment analysis of tweets in combination with the projected vs actual prices of a specific stock.

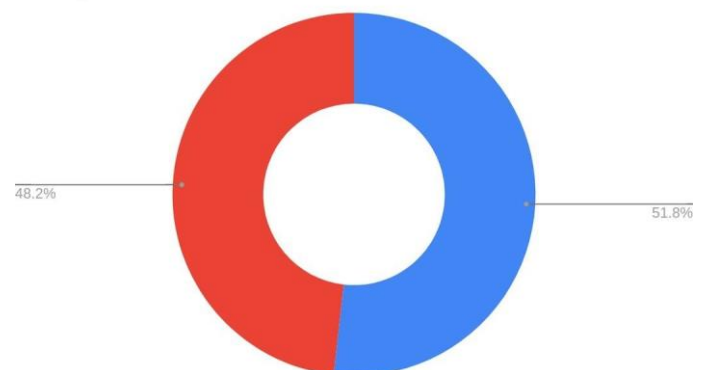
#### V. Public Responses

##### Do You Invest in Stock market?



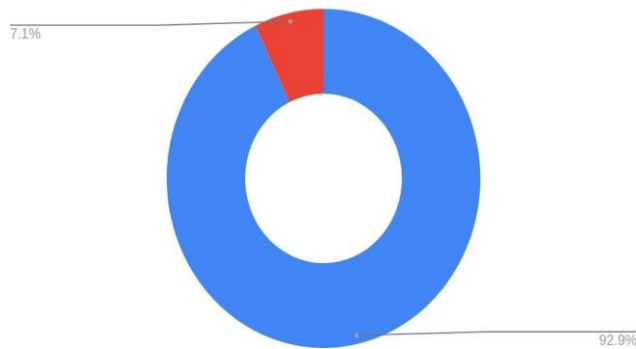
When asked to different people about Do You Invest in stock market only 12.6% people answered as NO and rest 87.5% out of 56 people answered YES.

##### Do you use Twitter to make stock market investments?



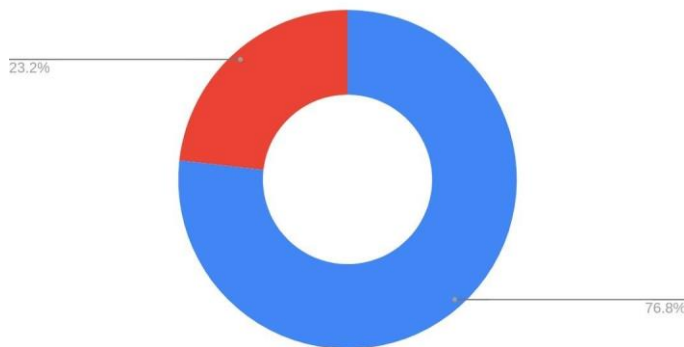
When asked to different people about Do You use Twitter to make stock market investments only 48.2% people answered as NO and rest 51.6% out of 56 people answered YES.

**Have you been influenced for investments by recent tweets regarding electric vehicles?**



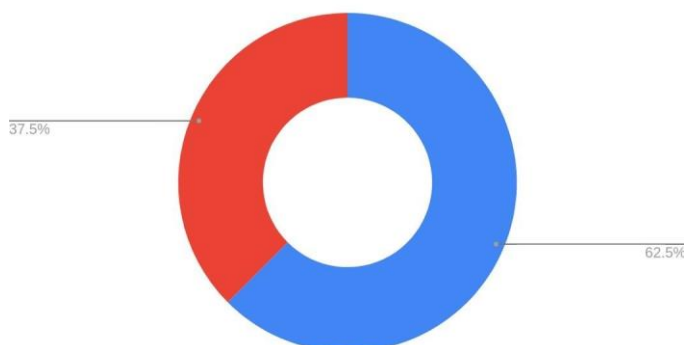
When asked to different people about Have you been influenced for investments by recent tweets regarding electric vehicles only 7.1% people answered NO and rest 92.9% out of 56 people answered YES.

**Do you like to study the market or follow the stock market influencers?**



When asked to different people about Do you like to study the market or follow the stock market influencers only 23.2% people answered They Study about the markert and rest 76.8% out of 56 people answered they follow the investors.

**Do you believe Twitter has an impact on the stock market or on investors?**



When asked to different people about Do you believe Twitter has an impact on the stock market or on investors only 37.5% people answered NO and The rest 62.5% out of 56 people answered YES.

**VI.Refrances**

[1] V. S. Pagolu, K. N. Reddy Challa, Ganapati Panda, Babita Majhi, Sentiment Analysis of Twitter Data for Predicting Stock Market Movements, International conference on Signal Processing, Communication, Power and Embedded System (SCOPEs)-2016

[2] T. Mankar, Tushar Hotchandani, Manish Madhwani, Akshay Chidrawar, Lifna C.S Stock Market Prediction based on Social Sentiments using Machine Learning(2020)

[3] Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta, Himanshu Dhiman, Lokesh Chouhan, Stock Market Prediction Using Machine Learning, 2018 First International Conference on Secure Cyber Computing and Communication(ICSCCC)

[4] Aditya Bhardwaj, Yogendra Narayan, Vanrajc, Pawana, Maitreyee Dutta, Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty, 4thInternational Conference on Eco-friendly Computing and Communication Systems. Procedia Computer Science 70 ( 2015 ) 85 – 91.