



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 1, January 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.512

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Real Time Stock Price Prediction

Deepak Mandal¹, Sushmit Vawal², Muskan Shivnani³, Rutuja Dhamdhare⁴

U.G. Student, Department of Computer, Sinhgad Academy of Engineering, Pune, Maharashtra, India¹

U.G. Student, Department of Computer, Sinhgad Academy of Engineering, Pune, Maharashtra, India²

U.G. Student, Department of Computer, Sinhgad Academy of Engineering, Pune, Maharashtra, India³

U.G. Student, Department of Computer, Sinhgad Academy of Engineering, Pune, Maharashtra, India⁴

ABSTRACT: The stock market is a very volatile environment; it can be influenced by positive or negative stock price releases. Different events may affect public sentiments and emotions differently, which may have an effect on the trend of stock market prices. Development of linguistic technologies and penetration of social media provide powerful possibilities to investigate users' moods and psychological states of people. In this paper we have discussed the possibility to improve accuracy of stock market indicator predictions by using data about sentiments of Twitter users. Understanding author's opinion from a piece of text is the objective of sentiment analysis. The thesis of this work is to observe how well the changes in stock prices of a company, the rises and falls, are correlated with the public opinions being expressed in tweets about that company. Here, we have applied sentiment analysis and supervised machine learning principles to the tweets extracted from twitter and analyze the correlation between stock market movements of a company and sentiments in tweets. In an elaborate way, positive news and tweets in social media about a company would definitely encourage people to invest in the stocks of that company and as a result the stock price of that company would increase. At the end of the paper, it is shown that a strong correlation exists between the rise and falls in stock prices with the public sentiments in tweets.

KEYWORDS: Stock Market Prediction, Sentiment Analysis, Twitter API.

I. INTRODUCTION

Stock market is an important division of the economy of a country. Complex behavior and unstable nature are difficult and essential tasks for the stock market decision. As tremendous amount of the data is produced by stocks each day, it is difficult task for an individual to analyze every past and present stock price information for predicting the future trend of stocks. Stock trends are affected by many factors, and one of which is daily stock price articles. Daily stock articles can play an vital role for prophets while judging the stock prices. Hence, it is become necessary to efficiently analyze the information to hold up and assist the investors to make smart trading resolution before making the real investments.

With the advent of social media, the information about public feelings has become abundant. Social media is transforming like a perfect platform to share public emotions about any topic and has a significant impact on overall public opinion. Twitter, a social media platform, has received a lot of attention from researchers in the recent times. Twitter is a micro-blogging application that allows users to follow and comment other users thoughts or share their opinions in real time. This situation makes Twitter like a corpus with valuable data for researchers. Each tweet is 140 characters long and therefore speaks public opinion on a topic concisely. This information exploited from tweets is very useful for making predictions.

In this paper, we contribute to the field of sentiment analysis of twitter data. The tweets are classified into positive, negative and neutral based on the sentiment of the stock price using Machine learning algorithm like support vector classifier and the technique used is Linear Regression model. This sentiment analysis on classification is used to determine the future trend of a stock market. If the positive sentiment is found, there is more probability that price of the stock will increase further and if negative sentiment is found, then price of the stock may decrease.



II. RELATED WORK

Sr. No.	Method	Title	Description	Advantages	Disadvantages
1	Support Vector Machine	Sentiment Based Stock market Prediction	Sentiment Analysis is the NLP technique that performs on the text to determine whether the author's intentions towards a particular topic, product, etc. are positive, negative, or neutral.	<ul style="list-style-type: none"> • More effective in high dimensional spaces. • Relatively memory efficient. 	<ul style="list-style-type: none"> • Not suitable for large data sets.
2	Neural Networks	Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis	Convolution Neural Networks (CNN) are used for Natural Language Processing in Neural Networks	<ul style="list-style-type: none"> • Ability to train machine. • Parallel processing ability. 	<ul style="list-style-type: none"> • Hardware dependence. • The duration of the network is unknown.
3	Decision Trees	Stock market prediction based on social sentiments using machine learning.	A decision tree is a flowchart-like tree structure where an internal node represents feature (or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node.	<ul style="list-style-type: none"> • Does not require normalization of data. • Does not require scaling of data. 	<ul style="list-style-type: none"> • Involves higher time to train the model. • Training is relatively expensive due to the complexity.
4	Linear Regression	Stock market prediction using a linear regression.	Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable.	<ul style="list-style-type: none"> • Simple to implement. • Works well irrespective of the dataset size. 	<ul style="list-style-type: none"> • Prone to outliers. • Prone to multicollinearity.

III. METHODOLOGY

To predict the stock price our proposed model consists of 5 stages as listed below

- Dataset Creation
- Dataset pre-processing
- Model creation
- Stock Prediction
- Plotting

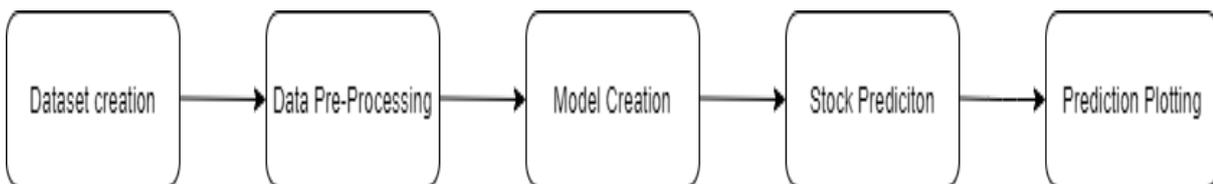


Fig. 1 Workflow of Proposed Methodology

1. Dataset Creation

Our Dataset consists of twitter data and the stock prices for a particular company for a week. Twitter data consists of tweets and stock prices have variables open, low, high, close data for a company over a week.

2. Data Pre-Processing:

This step consists of cleaning our data which handles the stop words. Handling the missing data. After cleaning we aggregate our tweets based on the date and add them in a single column along with the closing rate for that day. We assign the polarity for each statement. That is how much positive, negative, neutral statement is. And also assign the compound value that is overall sentiment of the statement.

3. Model Creation:

In this stage we create several models and find the model which has the best accuracy. The models that will be used are:

a. Random Forest Regressor :

Random forest is a supervised learning algorithm. The "forest" it builds, is an ensemble of decision trees, usually trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result.

Put simply: Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

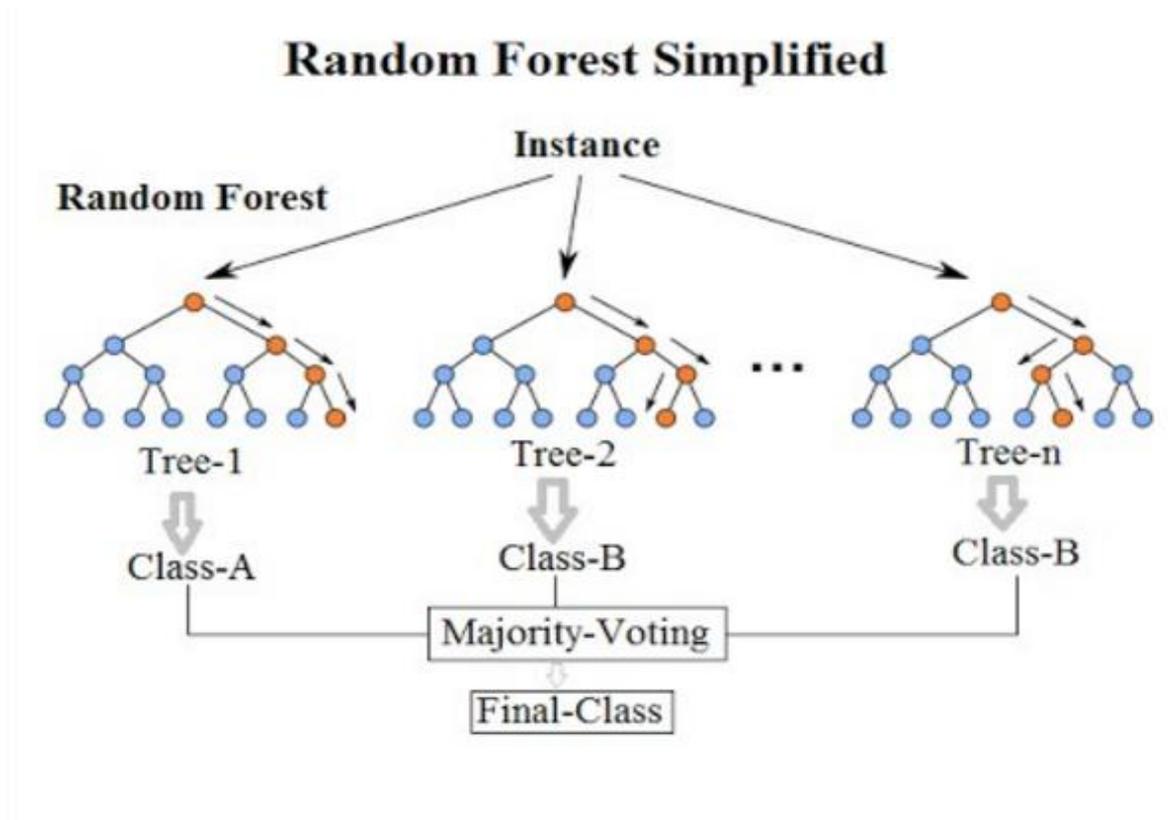


Fig. 2 Random Forest Model

b. LSTM:

Long Short Term Memory networks – usually just called “LSTMs” – are a special kind of Recurrent Neural Networks (RNN), capable of learning long-term dependencies. A Recurrent Neural Network is a type of neural network that contains loops, allowing information to be stored within the network. In short, Recurrent Neural Networks use their reasoning from previous experiences to inform the upcoming events.

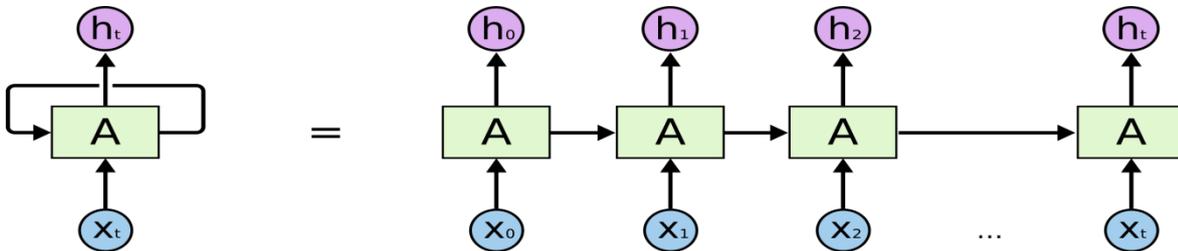


Fig. 3 Unrolled Recurrent Neural Network

In the case of RNN, the learned model always has the same input size, because it is specified in terms of transition from one state to another. Also, the architecture uses the same transition function with the same parameters at every time step. LSTM is a special kind of RNN, introduced in 1997 by Hochreiter and Schmidhuber. In the case of LSTM architecture, the usual hidden layers are replaced with LSTM cells. The cells are composed of various gates that can control the input flow. An LSTM cell consists of input gate, cell state, forget gate, and output gate. It also consists of sigmoid layer, tanh layer and point wise multiplication operation. The various gates and their functions are as follows

- Input gate: Input gate consists of the input.
- Cell State: Runs through the entire network and has the ability to add or remove information with the help of gates.
- Forget gate layer: Decides the fraction of the information to be allowed.
- Output gate: It consists of the output generated by the LSTM.
- Sigmoid layer generates numbers between zero and one, describing how much of each component should be let through.
- Tanh layer generates a new vector, which will be added to the state. The cell state is updated based on the outputs from the gates. Mathematically we can represent it using the following equations.

$$\begin{aligned}
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
 c_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\
 o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
 h_t &= o_t * \tanh(c_t)
 \end{aligned}$$

Where x_t : input vector, h_t : output vector, c_t : cell state vector, f_t : forget gate vector, i_t : input gate vector, o_t : output gate vector and W, b are the parameter matrix and vector.

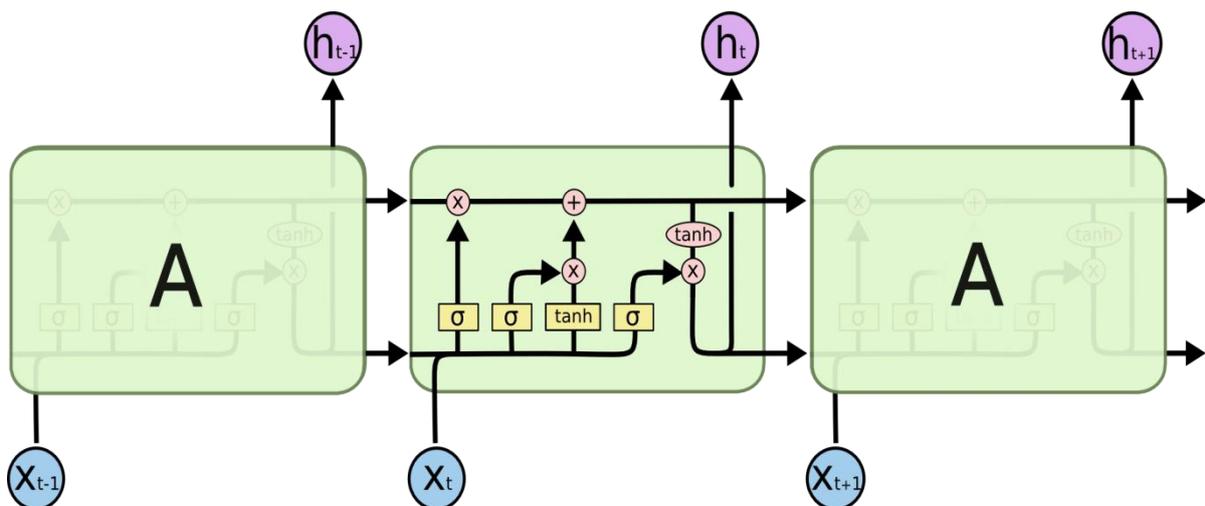


Fig. 4 LSTM model



IV. CONCLUSION

In this paper, we have shown that a strong correlation exists between rise/fall in stock prices of a company to the public opinions or emotions about that company expressed on twitter through tweets. The main contribution of our work is the development of a sentiment analyzer that can judge the type of sentiment present in the tweet. The tweets are classified into three categories: positive, negative and neutral. At the beginning, we claimed that positive emotions or sentiment of public in twitter about a company would reflect in its stock price. Our speculation is well supported by the results achieved and seems to have a promising future in research.

REFERENCES

1. CHUN YUAN LAI , RUNG-CHING CHEN, REZZY EKO CARAKA, " PREDICTION STOCK PRICE BASED ON DIFFERENT INDEX FACTORS USING LSTM": Proceedings of 2019 International Conference on Machine Learning and Cybernetics (ICMLC).
2. Sreelekshmy Selvin, Vinayakumar R, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P "STOCK PRICE PREDICTION USING LSTM,RNN AND CNN-SLIDING WINDOW MODEL": Proceedings of 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI).
3. Zhaoxia WANG, Seng-Beng HO, Zhiping LIN, "Stock Market Prediction Analysis by Incorporating Social and News Opinion and Sentiment": Proceedings of 2018 IEEE International Conference on Data Mining Workshops (ICDMW).
4. B. L. Pooja, Suvarna Kanakaraddi, Meenaxi .M. Raikar, "Sentiment Based Stock Market Prediction": Proceedings of International Multi-Conference on Information Processing, 2016.
5. Alexander Porshnev, Ilya Redkin, Alexey Shevchenko, "Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis": Proceedings of 2013 IEEE 13th International Conference on Data Mining Workshops.
6. Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, Babita Majhi, "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements": Proceedings of International conference on Signal Processing, Communication, Power and Embedded System (SCOPE)-2016.
7. Tejas Mankar, Tushar Hotchandani, Manish Madhwani, Akshay Chidrawar, Lifna C.S, "Stock Market Prediction based on Social Sentiments using Machine Learning": Proceedings of Neural Computing and Applications 27.1 (2016).
8. Vrishabh Sharma, Rajgauri Khemnar, Renu Kumari, Dr.Biju R Mohan, "Time Series with Sentiment Analysis for Stock Price Prediction": Proceedings of 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT) Manipal University Jaipur, Sep 28-29, 2019.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor

Impact Factor:
7.512

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details