



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

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 6, June 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 using Random Forest Regression

Deepak Mandal<sup>1</sup>, Sushmit Vawal<sup>2</sup>, Muskan Shivnani<sup>3</sup>, Rutuja Dhamdhare<sup>4</sup>

U.G. Student, Department of Computer, Sinhgad Academy of Engineering, Pune, Maharashtra, India<sup>1</sup>

U.G. Student, Department of Computer, Sinhgad Academy of Engineering, Pune, Maharashtra, India<sup>2</sup>

U.G. Student, Department of Computer, Sinhgad Academy of Engineering, Pune, Maharashtra, India<sup>3</sup>

U.G. Student, Department of Computer, Sinhgad Academy of Engineering, Pune, Maharashtra, India<sup>4</sup>

**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 rises and falls in stock prices with the public sentiments in tweets.

**KEYWORDS:** Volatile environment, Supervised, Sentiment Analysis, linguistic technologies.

## 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 a 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 user's 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.

### A. Motivation

As stock market is very volatile, it is very unlikely to what the future stock prices could be. This is where machine learning comes into the picture. Real time stock price prediction using machine learning techniques could help the investors determine where to invest. This could gain them a significant profit.

**B. Problem statement**

Due to the complexity, variability and uncertainty of the stock market, the stock price formation mechanism presents the characteristics of complexity and unpredictability. Therefore, how to accurately analyze, judge and predict stock prices for investors to make decisions is very critical. Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. The successful prediction of a stock's future price could yield significant profit.

**II. BACKGROUND**

Prediction systems are designed to predict things to the user based on an analysis of their data. It has the capability of producing outputs by predicting the set of items based on the company's history. These systems generally work by dealing with huge amounts of information based on the data provided by the user along with their preferences. The design of such systems depends on the domain and the particular characteristics of the data available. There are some popular algorithms used in developing these systems. Support vector machine can be used in high dimensional space as it is relatively memory efficient. Most of the prediction system uses artificial neural networks, which uses the historical time-series data. A backpropagation neural networks was introduced to predict the stock market data, and then applied in real-time data to check about the further improvements in accuracy.

**III. 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"> <li>• More effective in high dimensional spaces.</li> <li>• Relatively memory efficient.</li> </ul>	<ul style="list-style-type: none"> <li>• Not suitable for large data sets.</li> </ul>
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"> <li>• Ability to train machine.</li> <li>• Parallel processing ability.</li> </ul>	<ul style="list-style-type: none"> <li>• Hardware dependence.</li> <li>• The duration of the network is unknown.</li> </ul>
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"> <li>• Does not require normalization of data.</li> <li>• Does not require scaling of data.</li> </ul>	<ul style="list-style-type: none"> <li>• Involves higher time to train the model.</li> <li>• Training is relatively expensive due to the complexity.</li> </ul>
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"> <li>• Simple to implement.</li> <li>• Works well irrespective of the dataset size.</li> </ul>	<ul style="list-style-type: none"> <li>• Prone to outliers.</li> <li>• Prone to multi-collinearity.</li> </ul>

#### IV. SYSTEM WORKFLOW

In this section, we present an overview of our proposed system as well as its workflow which includes a number of steps. The basic use case diagram is shown in Fig.

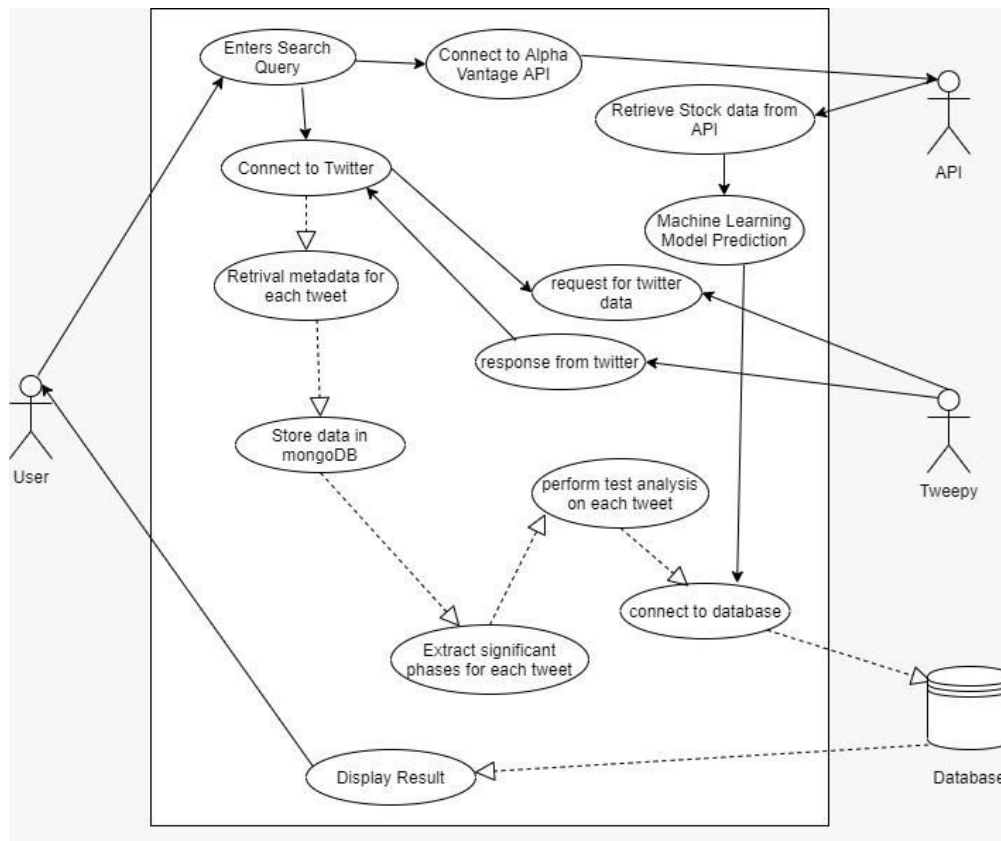


Fig 1: Use Case Diagram

##### A. Proposed System:

Real time Stock Price Prediction is a prediction model which allows the user to analyze future stock prices for a particular company. Here, we are using random forest regression as our machine learning model to predict the stock prices. We are using historical as well real time data for data collection.

##### B. Work Flow:

###### Step 1: Data Collection

The collection of data sets was one of the important tasks for our proposed system. Most of the data sets contained very few samples that did not fulfil our requirements. Therefore, we decided to use historical as well real time data.

Tweepy is open-sourced, hosted on GitHub and enables Python to communicate with Twitter platform and use its API. Using Tweepy we extract the stock price that is open, closed, high or low from the tweet of a particular company.

Yahoo Finance has been used to collect historical data of stock prices. The Yahoo Finance provides free JSONaccess to the stock market, plus a comprehensive set of technical indicators.

###### Step 2: Data Processing:

The data collected in Step 1 was inconsistent and noisy data. This could have led to poor quality of collected data and building a low-quality model. Gathering accurate and high-quality data is important for producing good results. Hence it was necessary for us to process it in order to overcome this issue. Data processing organizes the data in a proper form. We performed data cleaning to fill in the missing values and correct the inconsistencies in our data. The merging of data from multiple sources into a coherent database was done in the data integration task. The next technique



performed was data transformation where we performed smoothing and aggregation of data for storing and presenting it in a summary format. Data reduction was helpful in analyzing the representation of our data set without compromising its original integrity.

#### Step 3: Recommendation Generation

In our proposed system we have used random forest regressor algorithm for predicting prices. Here first, the data is split into different partitions. Then a certain number of random features is used to create and train a decision tree. This will repeat n number of times, where n is the number of trees to grow. Each tree will then output a prediction. Each prediction will then be calculated for number of votes, and the prediction with the highest number of votes will be the final prediction.

### V. SYSTEM IMPLEMENTATION

In this section, we present a brief explanation of the implementation of our proposed system.

#### Step 1: Data Collection

We collected the data using Tweepy and Yahoo Finance. Most companies put up tweets stating their updated stock prices. We implemented sentiment analysis on these tweets collected from Tweepy API in order to classify the statements into positive, negative or neutral. We have also used Yahoo Finance to extract historical data of stocks for a particular company.

#### Step 2: Data pre-processing

The data collected was noisy and inconsistent. Therefore, we had to organize the data in a proper form. We performed data cleaning to correct the inconsistencies and fill in the missing values in our data. We sorted the tweets according to the date when they were posted. Similarly the prices from Yahoo Finance are picked and then merged into a single data frame; which consisted the closing prices and the tweets put up on the same date.

We implemented sentiment analysis to segregate our data into positive, negative and neutral statements. The pie chart given below shows the breakdown of the same.

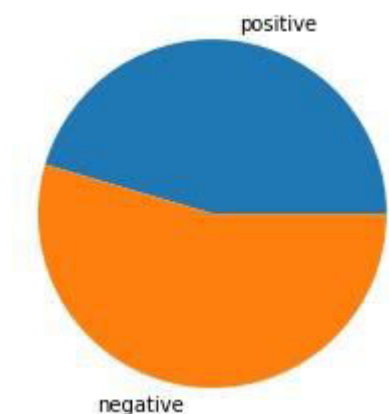


Fig 2: Pie Chart of Segregated Data

#### Step 3: Prediction

The processed data is then split into training and testing datasets. We applied various machine learning models to analyze the prediction accuracy. Following are the machine learning models which we used on our training dataset.

- 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.



Fig 3: Linear Regression

b. Decision trees: 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. In Decision Trees, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with the record's attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.

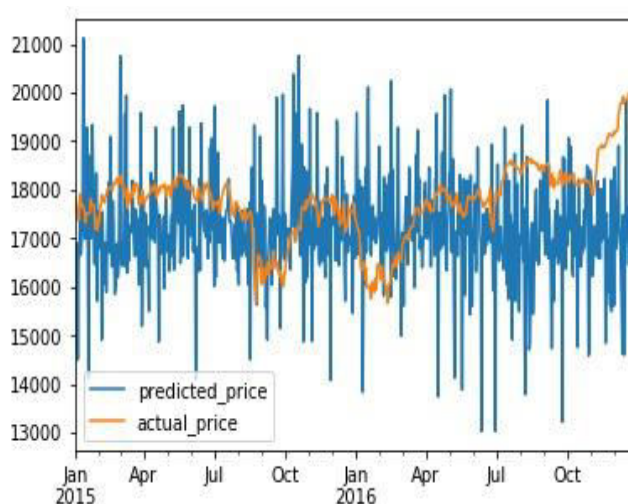


Fig 4: Decision Trees

c. Neural Network: Convolutional neural networks are composed of multiple layers of artificial neurons. Artificial neurons, a rough imitation of their biological counterparts, are mathematical functions that calculate the weighted sum of multiple inputs and outputs an activation value.

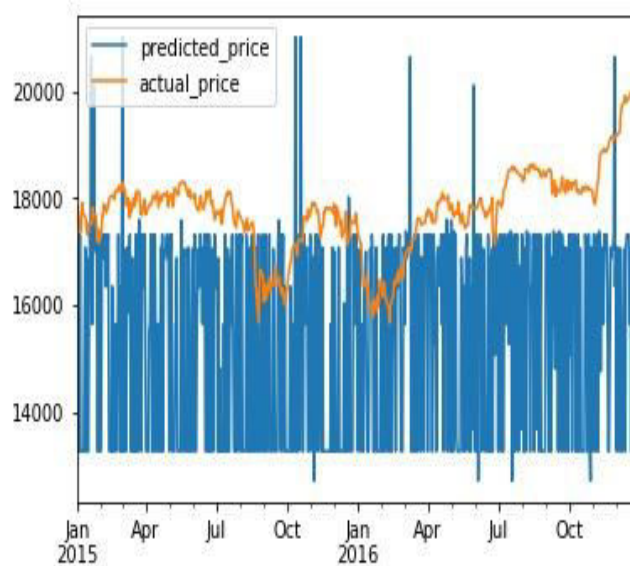


Fig 5: Neural Network

Following were the algorithms used to predict the output for our dataset. Out of all, we found that Random forest regression gave us the most accurate results and therefore we have used the same for our project.

Step 4: End.

## VI. FUTURE SCOPE

In our proposed system we have used the random forest algorithm. As future work, we can extend the functionalities of the project by applying better machine learning algorithms to generate more accurate results. We can also have an option to select from various companies. This will be helpful to enable sharing and improvement in the system. We can also use larger data sets which will help in getting better prediction results. We can also improve the validation of data for our system. We can add more detailed functions such as indicators to the graph. Buy and sell indicator which will help the user to buy and sell the stock at the right time. Support and resistance indicator will read the graph and provide exact supports and resistance for a particular graph. We can also provide moving averages to see the trend in the stock price. RSI can be used to see when a particular stock was overbought or oversold. We can provide an option to add multiple indicators at the same time for better experience in trading.

## VII. 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. Muhammad AimanAsrafRoslan, MohdHezriFazalulRahiman, "Stock Prediction Using Sentiment Analysis in Twitter for Day Trader": Proceedings of 2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC 2018), 3 - 4 August 2018, Shah Alam, Malaysia.



2. Saloni Mohan, SahityaMullapudi, SudheerSammata, ParagVijayvergia and David C. Anastasiu, "Stock Price Prediction Using News Sentiment Analysis": Proceedings of 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService).
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, SuvarnaKanakaraddi, Meenaxi .M. Raikar, "Sentiment Based Stock Market Prediction": Proceedings of International Multi-Conference on Information Processing, 2016.
5. Alexander Porshnev, IlyaRedkin, 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. VenkataSasankPagolu, Kamal Nayan Reddy Challa, Ganapati Panda, BabitaMajhi, "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements": Proceedings of International conference on Signal Processing, Communication, Power and Embedded System (SCOPEs)-2016.
7. Vrishabh Sharma, RajgauriKhemnar, RenuKumari, 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.





**Impact Factor:  
7.512**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

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



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details