ISSN (online): 2349-784X

# Trading Simulation and Stock Market Prediction

**Kavitha G**
*Asst. Professor*
*Department of CS Engineering*
*Sapthagiri College of Engg., Visvesvaraya Technological University, Bengaluru, India*

**Jayanth G**
*UG Student*
*Department of CS Engineering*
*Sapthagiri College of Engg., Visvesvaraya Technological University, Bengaluru, India*

**Jayanth K N**
*UG Student*
*Department of CS Engineering*
*Sapthagiri College of Engg., Visvesvaraya Technological University, Bengaluru, India*

**G SivaSankar Reddy**
*UG Student*
*Department of CS Engineering*
*Sapthagiri College of Engg., Visvesvaraya Technological University, Bengaluru, India*

**Darshan P**
*UG Student*
*Department of CS Engineering*
*Sapthagiri College of Engg., Visvesvaraya Technological University, Bengaluru, India*

## Abstract

Stock market prediction is one of the most attractive research topics since the successful prediction on the market's future movement could lead to significant profit. Traditional short term stock market predictions are based on the analysis of historical market data, such as stock prices, moving averages or daily returns. Stock market prediction on the basis of public sentiments expressed on social media has been an intriguing field of research. In an elaborate way, positive news and tweets in social media about a company will encourage people to invest in the stocks of that company and as a result the stock price of that company will increase. A company whose stock prices increase but has a negative image in the mindset of its investors will discourage current investors and not attract any new potential investors. The stock market is a platform where an enormous amount of data exists and constantly needs to be scrutinized for potential business opportunities. Since the stock market involves so much data, data sets get so large and complex that it becomes difficult to analyze using traditional data processing applications. In order to overcome these challenges, we can extract the useful information from the stock market trading data to an understandable structure using Data Mining, and also use algorithms that learn from this data and automatically predict further trends. Using Deep Learning algorithms and Sentiment Analysis as the approaches, we implement and simulate a brokerage system and analyze the stock market while at the same time learning the fundamentals of investment, how sentiment can influence a company's shares and most importantly, gain practical insight to the application of Deep Learning in the field of large-scale trading and financial enterprises.

**Keywords: Stock market prediction, social media, Sentiment Analysis, Deep Learning algorithms**

## I. INTRODUCTION

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 can yield significant profit. The efficient market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable. Later studies have debunked this theory. Prediction methodologies fall into three broad categories which may (and often do) overlap. They are fundamental analysis, technical analysis (charting) and technological methods. Traditional short term stock market predictions are usually based on the analysis of historical market data, such as stock prices, moving averages or daily returns. Stock market prices are largely fluctuating. The efficient market hypothesis (EMH) states that financial market movements depend on news, current events and product releases and all these factors will have a significant impact on a company's stock value. The EMH also states that any attempts to predict the future price, any future predictions, or any events scheduled in the future, such as upcoming products, developments cannot be considered. However, because of the lying unpredictability in news and current events, stock market prices follow a random pattern and cannot be predicted with more than 50% accuracy. Stock market prediction on the basis of public sentiments expressed on social media is an intriguing field of research. 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. More than a million users post over 140 million tweets every day. This situation makes Twitter like a corpus with valuable data for researchers .Each tweet is 140 characters long and speaks public opinion on a topic concisely. In an elaborate way, positive news and tweets in social media about a company will encourage people to invest

in the stocks of that company and as a result the stock price of that company would increase. With the advent of the digital computer, stock market prediction has since moved into the technological realm. The most prominent technique involves the use of artificial neural networks (ANNs), Genetic Algorithms (GA) and Sentimental Analysis.

## II. METHODOLOGY

### *A. Algorithms required for predicting future of stock using historical dataset:*

Stock Market Prediction involves requirement of previous years of stock data (historical data) of a particular company which can be retrieved from a financial website.

The algorithm we use for predicting the future price is the LSTM – RNN.

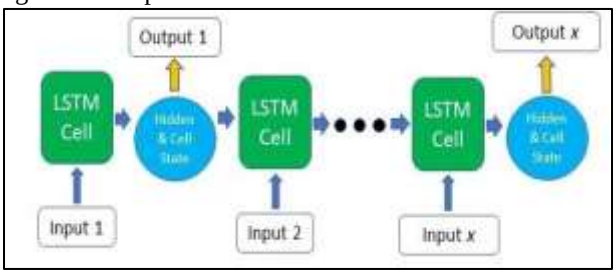


Fig. 1: Overall architecture of the LSTM neural network

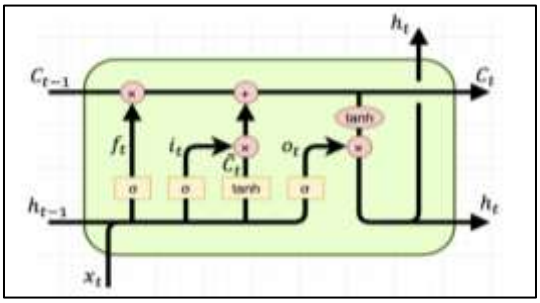


Fig. 2: Architecture of an individual LSTM cell

*) Long short-term memory (LSTM)*

is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. A recurrent neural network takes the output of the previous layer and the current input as the overall input to the current layer. A recurrent neural network is used because it overcomes the vanishing gradient problems while training artificial neural networks. Typically, in backpropagation algorithm or in standard feed–forward neural networks, the weights have to be updated once the first iteration of training is performed to reduce the error between the actual value and the obtained value. However, sometimes, the updates to the weights may be so disproportionately small that the weights may not be updated at all. This will lead to the training of the neural network being stopped.

A common LSTM unit is composed of a **cell**, an **input gate**, an **output gate** and a **forget gate**. The cell remembers values over arbitrary time intervals and the three *gates* regulate the flow of information into and out of the cell.

### *B. Sentimental analysis for predicting the future of stock using twitter sentiments:*

Natural language processing mainly deals with the context of the statement, unlike human beings who need to read the entire statement to process the sentiment of the statement. However, the statements need to be preprocessed before applying natural language processing to remove unwanted data. The input must consist of only textual data and separators used grammatically.

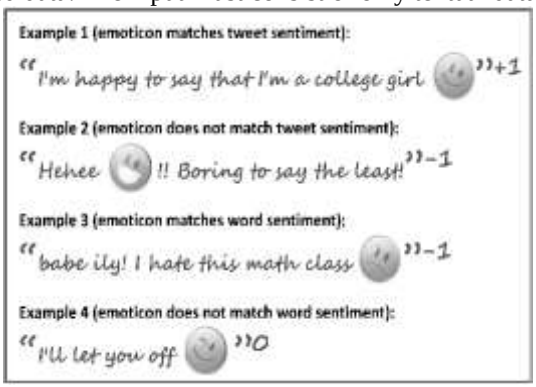

Fig. 3: Some samples of textual data, and the resultant polarities

Textual representations are done using n-grams.

*2) N-gram Representation:*

The N in N – grams is used to specify the number of words to be present in the split sequences of the statement. All punctuations and emoticons are discarded before dividing the statement into the respective sized N – grams. If one word is chosen, the N – gram is known as unigram, if two words are chosen, the N – gram is known as bigram, if three words are chosen, the N – gram is known as trigram, and so on. Ideally, bigrams are always preferred, because it is more efficient in obtaining the context and predicts the sentiment much more accurately. Longer sequences can be considered, however it becomes much more difficult to determine the context accurately.

## III. SYSTEM ARCHITECTURE

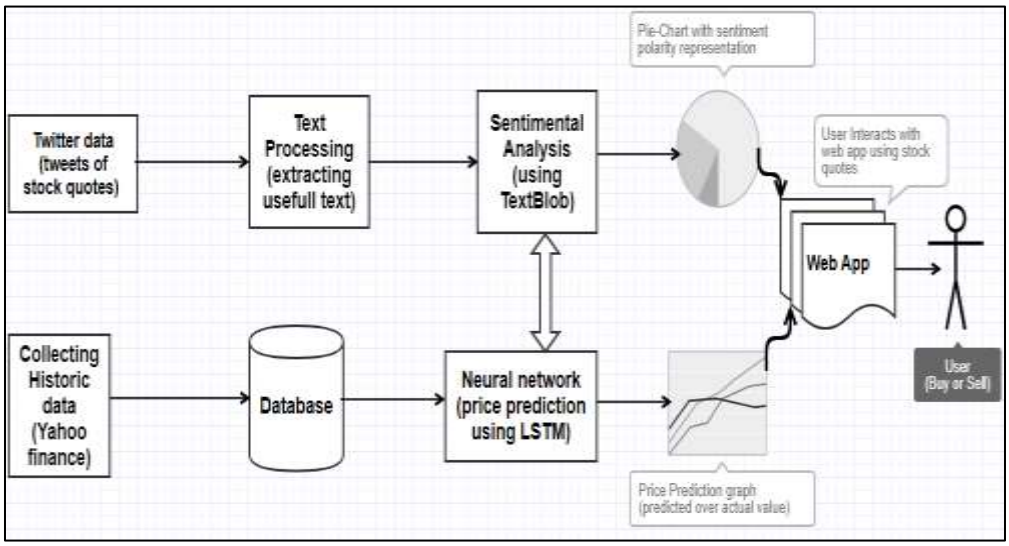


Fig. 4: System Architecture

### *A. Components of Architecture:*

*1) Twitter data –*

The stock quote entered by the user is taken as input and related tweets are fetched.

*2) Text processing-*

The extracted tweets are passed through pattern parser and n-grams to find stop word such as the,is,a,an etc. and only related words which are adjacent to sentiment analysis is retained.

*3) Sentimental analysis –*

By using built in library package TextBlob we gather the required subjectivity and polarity of the tweets.

*4) Collecting Historic data –*

It is nothing but previous year performance of a stock of a company i.e. opening and closing prices.

*5) Database –*

The collected data from the previous steps is stored in memory for computational purposes.

*6) Neural network –*

LSTM an RNN is used to predict the price by taking historic data price as input.

*7) Graph (outputs) –*

The final computed results are visualized using graphs for better understanding by the user.

*8) Web app and User –*

By considering the predicted price and current sentiment of stock in market the user decides to buy or sell the stock.

### *B. Implementation:*

*1) Sentimental analysis:*

We get the access key and tokens from developers/twitter.com and use it with tweepy package. After authenticating to our developers account we enter the keyword/ticker of a stock company to get related tweets and then we preprocess the tweets according to our needs. We store tweets in file and pass it to TextBlob() for sentimental analysis. We have used regular expressions to remove the links,'@','#' symbols from tweets. TextBlob then compares the tweets with the pre-defined library of words which has pre-defined sentiments using n-grams (dividing words into n words). Then we get the polarity of the tweets as

shown in the figure by sentiment.polarity(). Finally we visualize the sentiments in the pie chart with positive, negative and neutral.

*2) LSTM implementation:*

First we import the necessary libraries and LSTM from keras sequential model. We import the datasets using pandas and check for missing values and replace them with mean. We normalize the datasets using minmax scaler to a range (0, 1). We divide the scaled datasets into training and testing in 70:30 ratios. We create a dataset that takes x as no. of days and y as price and reshape it before feeding it to the LSTM model. The network has a visible layer with 1 input, a hidden layer with 4 LSTM blocks or neurons, and an output layer that makes a single value prediction. The sigmoid function is used as the activation function for the LSTM blocks. The sigmoid function accepts values in the range of 0 to 1. The network is trained for 500 epochs and a batch size of 1 is used. We then import RMSE from sklearn.metrics to get the error rate for the train and test set. At the end we plot the graph for predicted model considering trained and tested data and analyze the graph.

*3) Prophet:*

Although LSTM can be used to predict the data with sufficient accuracy, it is very time consuming, because the neural network has to be trained for a long duration over a large dataset for the model to achieve satisfiable accuracy. The training also puts an enormous strain on the computer's hardware resources. Prophet uses a simpler dataset, is faster, more accurate and fully automated when compared to LSTM. Since LSTM only plots the sequential data of actual and predicted values, we need a future data of stock so we use prophet to get the future datagram closing price. We first get the closing price of the stock and convert it to pandas dataframe. We will model by using Prophet() and pass the dataframe to fit the model using model.fit(df) , we specify the number of days to which we require the future datagram for. We get the future price of the stock using the model.make_future_dataframe(periods=num_days). After getting the future datagram we predict the price using model.predict(). Finally we visualize the predicted price for the number of days specified using a line chart in flask frame work.

## IV. EXPERIMENTAL RESULTS

We have taken TCS as the sample company for experimentation purpose.

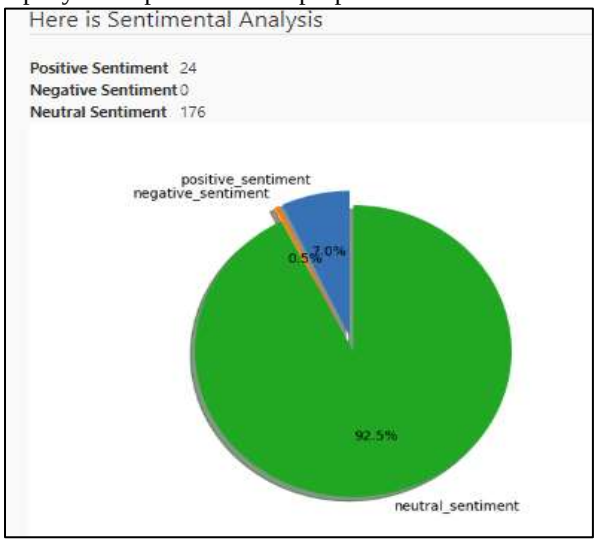


Fig. 5: Pie Chart showing Sentiment Analysis for TCS

After the sentiments are fetched, we analyze the positive, neutral and negative sentiments and represent the result using a pie-chart plot.

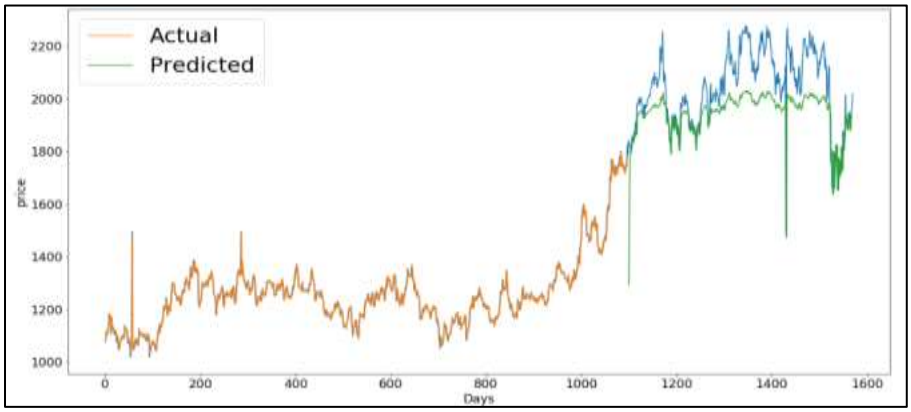


Fig. 6: Graph showing the actual versus predicted values using LSTM

The graph plot of actual versus predicted of TCS.NS considering data from (2014-2020). We have obtained 100% accuracy during training and around 70% accuracy during validation.


Fig. 7: Graph showing the actual versus predicted values using Prophet

The above graph shows prediction of real time data using prophet actual versus predicted of TCS.NS. It displays real-time future price for 10-days.

## V. CONCLUSIONS

Stock market is a volatile field, it is impossible to predict the exactly how the market fluctuates as it depends on various factors ex: public sentiments, historic price, business partnership etc. In this system we have considered the sentiments and historic price of stock to predict how market actually fluctuates. By considering pie-chart we get current sentiment of a person towards a company and graph plots of LSTM and prophet we get the future price of the stock. The predicted price came out to be 97.7722% accurate compared with the original price.

## VI. FUTURE WORKS

This system is used to show how prices can be predicted using neural networks and the influence of social media in the share market. It can be adopted and enhanced by using a larger dataset for training purpose, and increasing the number of cells in the neural network to obtain higher accuracy. Although the numbers of tweets used in our system for Sentiment Analysis are very few, investors can get authorized to obtain a larger corpus of textual data for processing, thereby obtaining better results from Sentiment Analysis and make a wise decision before investing in the stocks of any company.

## REFERENCES

[1] Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, Babita Majhi – Sentiment Analysis of Twitter Data for Predicting Stock Market Movements (28 Oct 2016), arXiv:1610.09225v1 [cs.IR]
[2] Sunil Kumar Khatri, Ayush Srivastava – Using Sentimental Analysis in Prediction of Stock Market Investment (2016)
[3] Rohan Srivastava, Soumil Agarwal, Dhruv Garg, Jagdish Chandra Patni – Capital Market Forecasting By Using Sentimental Analysis (2016)
[4] Nonita Sharma, Akanksha Juneja – Combining of Random Forest Estimates using LSboost For Stock Market Index Prediction (2017)
[5] Pranjal Chakraborty, Ummay Sani Pria, Md. Rashad Al Hasan Rony, Mahbub Alam Majumdar – Predicting Stock Movement using Sentiment Analysis of Twitter Feed (2017)
[6] Jeevan B, Naresh E, Vijaya Kumar, Prashanth Kambli - Share Price Prediction using Machine Learning Technique (2018)
[7] Roger Achkar, Fady Elias-Sleiman, Hasan Ezzidine, Nourhane Haidar - Comparison of BPA-MLP and LSTM for Stocks Prediction (2018)