

Stock Market Trend Prediction Using Regression Model, RNNs, and Sentiment Analysis



Rajkumar Sujatha, Varad Abhyankar, Abhishek Gehlot, Priyank Gupta, and Sivakumar Subramaniam

Abstract Stock price prediction is the act of predicting the value of the stock of a particular company in the future to maximize an investor's profit. In this paper, we propose a machine learning model for stock price prediction. The machine learning model uses LSTM (Long short term memory networks) and Multiple regression algorithms. Along with the machine learning model we look at a couple of important ratios and sentiment analysis which are indicative of whether a stock is overvalued or undervalued. Our model is designed to be particularly helpful for short-term investors for deciding entry and exit points during stock trading.

Keywords Machine learning · Sentiment analysis · LSTM · Multiple regression · Stock value

1 Introduction

Stock value prediction is a very interesting topic that has been the center of attraction of several researchers since the accurate prediction of stock value can yield significant profits. However, stock value prediction has been proven to be a challenging task for analysts around the globe because of close to the random behavior of the stock value series. Stock value analysis or prediction can be broadly classified into two types- fundamental analysis and technical analysis. Fundamental analysts use financial reports of the company, various ratios, management effectiveness of the company to predict the future stock value. This method can be very effective for long term investors. Technical analysts use past stock values, patterns and trends to predict future stock values. We will be using technical analysis for future stock value

R. Sujatha (✉) · V. Abhyankar · A. Gehlot · P. Gupta
School of Electronics Engineering, Vellore Institute of Technology, Vellore, India
e-mail: sujatha.r@vit.ac.in

S. Subramaniam
Department of Electronics and Computer Engineering, Universiti Teknikal Malaysia, Melaka, Malaysia

prediction. This method can be effective for short term investors for determining entry and exit points in the stock market.

In this paper, we propose a machine learning model for stock value prediction using LSTM and Multiple regression along with Sentiment analysis. This model is expected to be particularly helpful for short-term investors for deciding entry and exit points during stock trading. It is also expected to be helpful for amateur investors and protect them from incurring significant losses before they gain a better understanding of the stock market.

2 Literature Review

A lot of research has been done on stock value prediction using machine learning. We went through multiple papers on the topic for gaining a better understanding of this topic. We also studied the previous work done on sentiment analysis. Here we give a brief description of our literature survey.

Hegazy et al. [1] proposed a method to predict the stock values of several S&P 500 stocks using Particle Swarm Optimization and Least Square Support Vector Machine. They achieved a significantly smaller mean squared error (MSEs) on several US stocks.

Khan et al. [2] use several algorithms for stock prediction such as Support vector machines, ANN, linear regression, K-NN, and Naïve Bayesian classifier.

Perwej et al. [3] use two types of models and compare them using historical data from the Bombay Stock Exchange. The first model was based on Deep learning parameters updated through particle swarm optimization. The second method was based on Deep learning parameters updated through the least mean square.

Jia et al. [4] discuss the effectiveness of long short term memory(LSTM) networks trained by backpropagation for stock price prediction. A range of different architecture LSTM networks are constructed trained and tested.

Chong et al. [5] applied a deep learning-based stock market prediction model and tested its ability to extract features from a large set of raw data without relying on prior knowledge of predictors. They tested it on data from the Korean stock market.

Rajput et al. [6] studied different methods for predicting stock prices using sentiment analysis from social media and data mining.

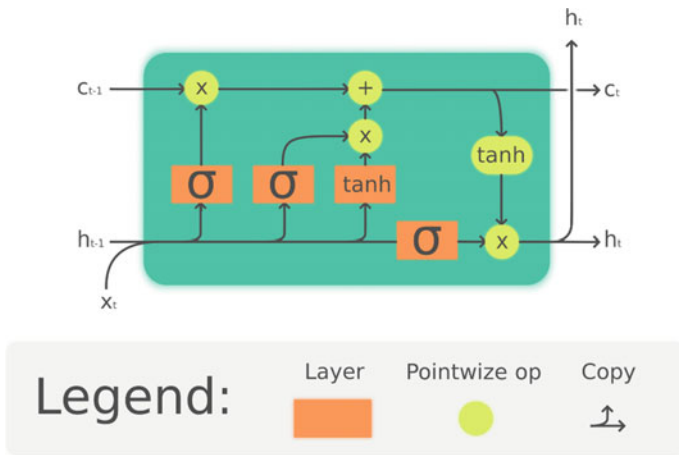
Ahlgren et al. [7] has used statistics to analyze the evolution of sentiment analysis over the last few years. The popularity and effectiveness of the algorithms have been also discussed.

Rajkumar et al. [8] has made use of Azure services along with Deep learning for predictive analysis of crop cultivation. The paper emphasizes the use of IoT technology for smart agriculture.

3 Algorithms Used

We used LSTM networks and Multiple regression to predict weekly stock prices of various companies listed on the National Stock Exchange of India(NSE). A brief overview of those algorithms is given below.

3.1 Long Short Term Memory (LSTM) Network



(https://www.google.com/url?sa=i&source=images&cd=&ved=2ahUKEwjL7rX99YnmAhWBxjgGHdyPBFwQjRx6BAgBEAQ&url=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FLong_short-term_memory&psig=AOvVaw3piovfrOBz3eHK6gPNUwZy&ust=1574927923718269)

LSTM networks are an improvement over the conventional Recurrent neural networks (RNN). They have been observed to provide an effective solution for most of the sequence prediction problems. Unlike RNN, an LSTM network can selectively remember patterns for a long duration of time. An LSTM unit consists of a cell, an input gate, an output gate and a forget gate. The cell selectively remembers values over arbitrary time intervals and the gates regulate the flow of information in and out of the cell.

We used an LSTM algorithm with 7 epochs, a time-step of 60 and a batch size of 1 on the stock price sequence of several companies. The time-step value indicates the number of data values the algorithm will look back into before predicting the next value. This means that in our model the algorithm looks at the past two months of stock values before predicting the future stock value. Batch size indicates the number of previous data values the network looks into before updating the weights. The predicted value by the LSTM network was again used to find the next prediction until the stock values of the entire week were predicted.

3.2 Multiple Regression

Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a dependent variable based on the value of two or more other independent variables. We used the multiple regression algorithm on the weekly closing price of stocks of various companies by considering multiple parameters as the dependent variables to predict the future price of the stock. The parameters included the 'RSI' index of the stock, closing price of the stock, closing price of the Nifty index and some ratios related to the financial health of the company. This method was used to predict weekly stock value.

4 Sentiment Analysis

We used twitter sentiment analysis to gain insight into the public opinion or perception about a company. The dataset consisted of trending tweets about the company. It was divided into Train and Test datasets for testing purposes. First of all the tweets containing emojis were classified directly into positive or negative by differentiating between happy and sad or angry emojis. Regular expressions were used for this purpose. For tweets without emojis distinct words were identified. Next stem and Lemma techniques were used to convert each word into the root word [9]. The words obtained were further cleaned and special characters were removed. More than two continuously the same characters in any word were truncated. On this clean data, a Multinomial Naive-Bayes classifier was used [10]. The model was tested on the Test dataset and was found to have around 85% accuracy.

5 Some Important Ratios

Before looking at the proposed model for stock value prediction let's glance at some very important ratios that are indicative of a stock is undervalued or overvalued [11].

5.1 Relative Strength Index(RSI)

RSI index is a ratio that ranges from 0 to 100. The stock is considered overbought when the value of the RSI is greater than 70. The stock is considered oversold when the value of the RSI is less than 30. RSI is very popular among stock market traders as its movement gives a good indication of stock value movement.

The formula for the RSI index is given as-

$RSI = 100 - [100 / (1 + (\text{Average of Upward Price Change for an arbitrary time interval} / \text{Average of Downward Price Change for an arbitrary time interval}))]$ [12].

5.2 Price to Equity Ratio

The price to equity ratio indicates if the company is undervalued or overvalued. It is the stock value of the company multiplied by the number of outstanding shares of the company divided by the book value of the company. A lower value of the Price to equity ratio indicates that the company is undervalued while a higher value of the Price to equity ratio indicates that the company is overvalued. Price to equity ratio is a relative term [13].

5.3 Price to Sales Ratio

Price to sales ratio is the yearly sales of the company divided by the multiplication of the number of outstanding shares of the company and the stock value of the company. A lower value of the Price to equity ratio indicates that the company is undervalued while a higher value of the Price to equity ratio indicates that the company is overvalued. Price to sales ratio is a relative term.

5.4 Price to Earnings Growth Ratio

The price to earnings growth ratio is Price to Earnings Ratio (P/E Ratio) divided by the Earnings per Share Growth. A lower value of the Price to earnings growth ratio indicates that the stock has a positive financial future. This also is a relative term.

6 Proposed Model

Our proposed model for future stock price prediction takes into account the prediction by the LSTM network, the Multiple regression algorithm, RSI index and Sentiment analysis of the company before making a prediction. The model makes weekly predictions of the closing price of a stock. We tested this model on the stock values of various companies, and the results of these tests have been given in the results and discussion section (Fig. 1).

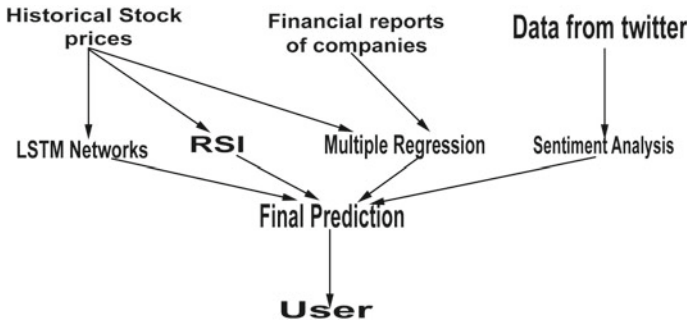


Fig. 1 Flow diagram of the model

7 Results and Discussions

All the historical stock data for the tests was taken from Yahoo Finance while the financial data about various companies was obtained from the annual and quarterly reports available on each company’s official website.

For testing the LSTM network the historical stock data was classified into train and test datasets with a train: test ratio of 4:1. Graphs of the Closing price of the stock and predicted stock price for a couple of companies are shown (Figs. 2 and 3).

We combined the prediction by the LSTM network with Multiple Regression and Sentiment analysis to provide a final prediction model for the stock price. The model predicts weekly Stock values. This model was tested on the data of multiple companies listed on the National Stock Exchange of India(NSE). The results of some of the tests are tabulated in Table 1.

We see from the results of the tests conducted that the model gives a more accurate prediction for the stock price when compared to individual LSTM or multiple

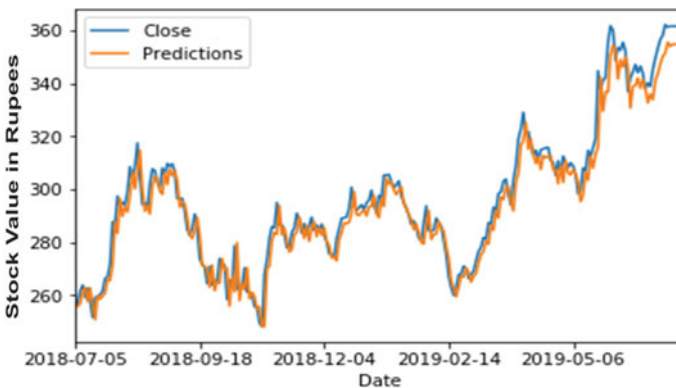


Fig. 2 Actual closing price versus predicted stock price by LSTM network for the stock of State Bank of India (SBI)

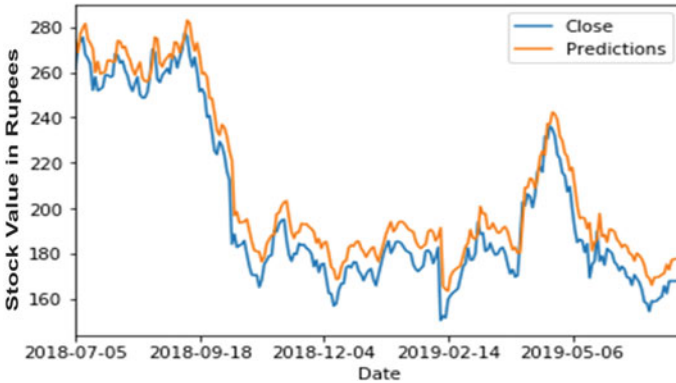


Fig. 3 Actual closing price versus predicted stock price by LSTM network for the stock of Tata-motors

Table 1 Results of some tests

Name of the company	Stock value on 1 July 2019 (in rupees)	Sentiment analysis (% of positive opinion) (%)	RSI	The predicted value of the stock on 8 July 2019 (in rupees)	Actual stock value on 8 July 2019 (in rupees)
TCS	2239.5	72	46	2175.4	2224.9
Tata motors	164	43	69.9	173.5	154.5
Nestle India	11868.8	67	77.6	11847.2	11801.5
SBI	361.5	38	91.8	330.2	355.3
HDFC	2245.9	74	73.6	2232.8	2179.6

regression-based models. More importantly the model most of the time correctly predicts whether the stock price will increase or decrease in the coming week.

8 Conclusion and Future Work

This paper proposes a unique model for stock price prediction consisting of the LSTM network, Multiple regression and Sentiment analysis. This model achieves significant improvement in accuracy over models predicting stock prices using LSTM or Multiple regression individually. This model predicts weekly Stock values. We hope that this model will be useful for investors and especially amateur investors for determining entry and exit points during stock trading.

Future work can include improving the accuracy of the LSTM network or increasing the accuracy of the Multiple regression algorithm by increasing the number of parameters. This model also can be implemented in a mobile application with a

good user interface so that more and more investors can get the benefit of this model for stock price prediction.

References

1. Hegazy O, Soliman OS, Salam OS (2014) A machine learning model for stock market prediction. arXiv preprint arXiv, pp 1402.7351
2. Khan W, Ghazanfar M, Asam M, Iqbal A, Ahmad S, Khan JA (2016) Predicting trend in stock market exchange using machine learning classifiers. *Sci Int* 28(2):1363–1367
3. Perwej A, Yadav KP, Sood V, Perwej Y (2018) An evolutionary approach to bombay stock exchange prediction with deep learning technique. *IOSR J Bus Manag (IOSR-JBM)* 20(12):63–79
4. Jia H (2016) Investigation into the effectiveness of long short term memory networks for stock price prediction. arXiv preprint arXiv: 1603.07893
5. Chong E, Han C, Park FC (2017) Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Syst Appl* 83:187–205
6. Vivek R, Sarika B (2016) Stock market forecasting techniques: literature survey. *IJCSMC* 5(6):500–506
7. Ahlgren O (2016) Research on sentiment analysis: the first decade. In: 2016 IEEE 16th international conference on data mining workshops (ICDMW), Barcelona, pp 890–899
8. Rajkumar S, Arun M, Hirwani J, Sanjeev SS (2019) Predictive analysis of crops cultivation for a smart green environment using azure services. *Int J Recent Technol Eng (IJRTE)* 7(5S2). ISSN: 2277–3878
9. Liu B (2012) *Synthesis lectures on human language technologies: sentiment analysis and opinion mining*. Morgan & Claypool Publishers, California
10. Dey L (2016) Sentiment analysis of review datasets using naïve bayes' and K-NN Classifier. *Int J Inf Eng Electron Bus* 8(4):54–62
11. Abbad J, Obeidat S (2010) Determinants of the intrinsic value of common stocks and the application of industrial companies in the ASE. *Eur J Econo Finan Adm Sci*
12. <https://admiralmarkets.com/education/articles/forex-indicators/relative-strength-index-how-to-trade-with-an-rsi-indicator>
13. Sarikhani M, Ebrahimi F (2012) An empirical evaluation of using the residual income model for prediction of a stock price. *Afr J Bus Manag* 6(5):2043–2047