© July 2020| IJIRT | Volume 7 Issue 2 | ISSN: 2349-6002

# Forecasting Stock Prices using LSTM and Web Sentiment Analysis

Karan Sharma
*Department of Computer Science and Engineering, Galgotias University*

***Abstract*— The craft of estimating the stock prices has been a troublesome task for huge numbers of researchers and analysts. Indeed, investors are profoundly interested in the exploration area of stock price prediction. For a good and fruitful investment, numerous investors are sharp in knowing the future circumstance of the stock market. An effective system for the stock market helps traders, investors, and analysts by giving strong information like the future direction of the stock market. In this work, I am presenting a Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) approach to predict the stock market indices.**

***Index Terms*— Long short-term memory (LSTM), Recurrent neural network (RNN), nifty 50, root mean square error (RMSE), prediction, stock prices, web sentiment analysis.**

## I. INTRODUCTION

There are a huge number of financial indicators and furthermore the variance of the stock market are exceptionally vicious. However, as the technology is getting advanced, the chance to increase a consistent fortune from the stock market is increased and it additionally encourages specialists to find the most informative pointers to make a better prediction. The forecast of the market value is of great importance in amplifying the profit of stock option buy while keeping the risk low.

Recurrent Neural Network (RNN) has demonstrated to be one of the most impressive models for handling sequential data.

Long Short-Term memory is one of the best RNNs architecture. LSTM presents the memory cell, a unit of computation that replaces conventional artificial neurons in the hidden layer of the network. With these memory cells, network are able to effectively associate memories and input remote in time, thus suit to get a handle on the structure of data dynamically after some time with high prediction capacity.

The paper that we have introduced demonstrated and predicted the stock returns of NIFTY 50 using LSTM. I have gathered 5 year of recorded data of NIFTY 50 and utilized it for the training and validation purposes for the model. The next segment of the paper will be methodology where we will clarify about each procedure in detail. From that point onward, we will have pictorial representation of the analysis that we have used and we will likewise reason about the outcomes accomplished.

## II. METHODOLOGY

Different kinds of neural networks can be developed by the mix of various factors like network topology, training method etc. For this experiment, we have considered Recurrent Neural Network and Long Short-Term Memory.

This section we will discuss the methodology of our system. Our system consists of several stages which are as follows:-

Stage 1: Raw Data

In this stage, the historical stock data is gathered from https://www.quandl.com/information/NSE and this data is utilized for the prediction of future stock prices.

Stage 2: Data Preprocessing

The pre-handling stage includes:-

a. Data discretization: Part of data reduction but with specific significance, particularly for numerical data.
b. Data Transformation: Normalization.
c. Data Cleaning: Fill in missing values.
d. Data Integration: Integration of data files.

After the dataset is changed into a clean dataset, the dataset is divided into training and testing sets in order to evaluate. Here, the training values are taken as the more recent values. Testing data is kept as 5-10 percent of the complete dataset.

Stage 3: Feature Extractions
In this layer, just the features which are to be fed to the neural networks are chosen. We will choose the feature from open, high, low, close, volume, date.
Stage 4: Training Neural Network
In this stage, the data is fed care of to the neural system and prepared for forecast appointing arbitrary predispositions and loads. Our LSTM model is made out of a successive info layer followed by 2 LSTM layers and dense layer with ReLU activation and afterward at last a thick output layer with linear activation function.

The code of the Neural Network implemented in Keras is as follows

```
model = Sequential()

model.add(LSTM(128, input_shape=(layers[1],
layers[0]), return_sequences=True))

model.add(LSTM(64, input_shape=(layers[1],
layers[0]), return_sequences=False))

model.add(Dense(16,init='uniform',activation='relu'))

model.add(Dense(1,init='uniform',activation='linear'))
```

Stage 5: Output Generation
In this layer, the output value generated by the output layer of the RNN is compared with the target value. The error or the contrast between the objective and the obtained output value is minimized by utilizing back propagation algorithm which adjusts the weights and the biases of the network.

## III. ANALYSIS

For analyzing the effectiveness of the system we have used the Root Mean Square Error(RMSE). The error or the difference between the target and the obtained value by utilizing by the RMSE value. RMSE is the square root of the mean/averagel of the square of all the error. The utilization of RMSE is highly common and it makes a magnificent general purpose error metric for numerical predictions. Contrasted with the comparative Mean Absolute Error, RMSE intensifies and seriously punishes the large error.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum\left(\hat{Y}_i - Y_i\right)^2}$$

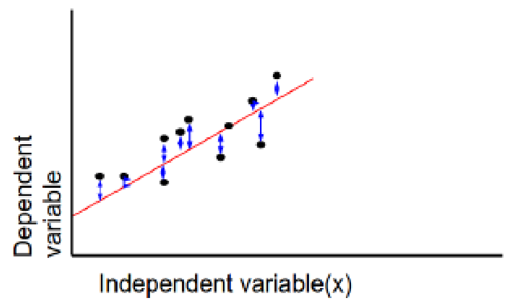


**Figure 1:** RMSE Value Calculation.

## IV. EXPERIMENTAL WORK

- Dataset description: We gained the data from https://www.quandl.com. We have gathered the historical stock information of NIFTY 50 from the National Stock Exchange. We have gathered every day's dataset and kept a window size of 22 days. Data ranges from 01.01.2011 to 31.12.2016.
- Sequence Data: We got 1312 groupings from 01.01.2011 to 31.12.2016. From these dataset we utilized 1180 examples for training reason and 132 examples for testing reason.
- Training Detail: For preparing the model we utilized RMSprop as the analyzer and normalised every vector of the arrangement. We utilized Google cloud engine as atraining platform [Machine type: n1-standard (2 vCPUs, 7.5 GB memory), CPU stage: Intel Ivy Bridge] and utilized Ubuntu 18.04, Keras (Frontend) and Tensorflow (Backend) as the learning environment. For our explore, we have utilized a various set of parameters with an alternate number of epochs to measure the RMSE of Training and Testing dataset.

## V. EXPERIMENTAL RESULTS

Table 1: Comparing the Results by Using Different Parameters and Epochs.

| Parameters | No. of Epoch | Training RMSE | Testing RMSE |
|---|---|---|---|
| Open/Close | 250 | 0.01491 | 0.01358 |
| Open/Close | 500 | 0.01027 | 0.00918 |
| High/Low/Close | 250 | 0.01511 | 0.014 |
| High/Low/Close | 500 | 0.01133 | 0.01059 |
| High/Low/Open/Close | 250 | 0.0133 | 0.01236 |
| High/Low/Open/Close | 500 | 0.00983 | 0.00859 |

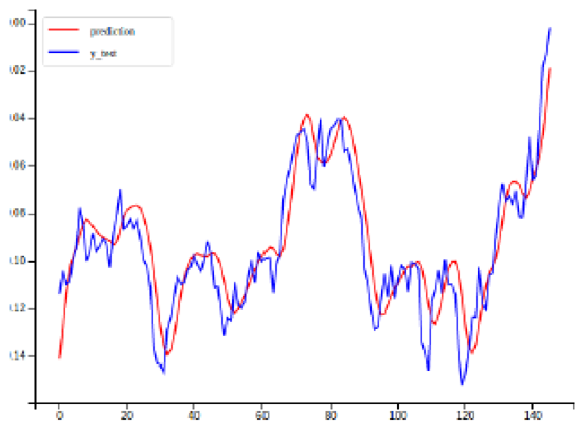

Figure 2: Open/Close with 250 epochs

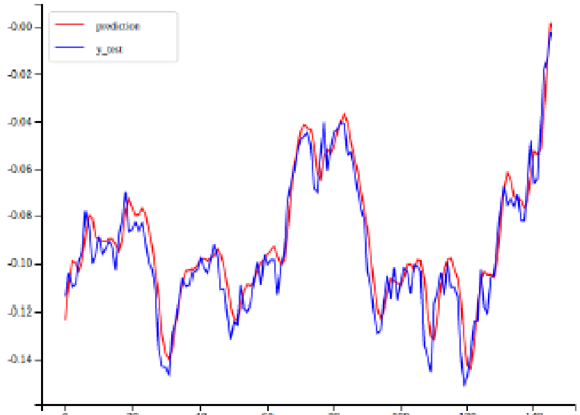

Figure 3: Open/Close with 500 epochs

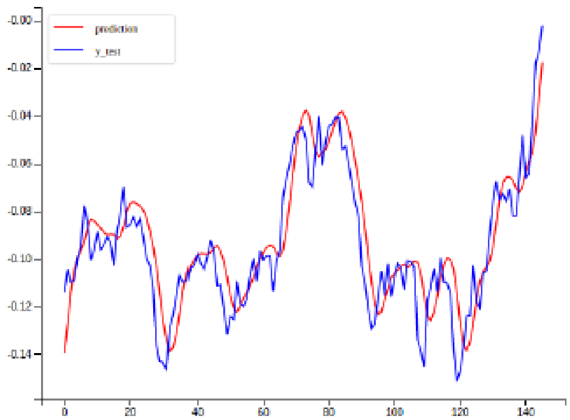

Figure 4: High/Low/Close with 250 epochs

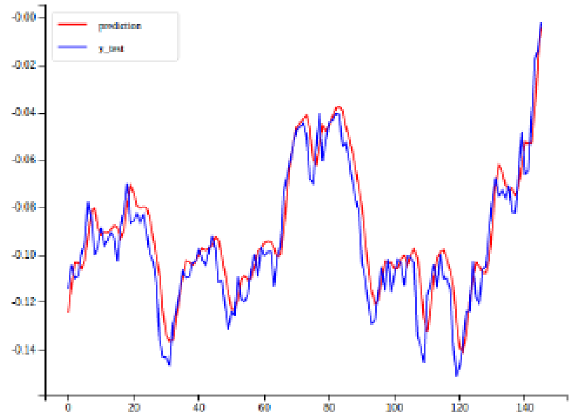

Figure 5: High/Low/Close with 500 epochs

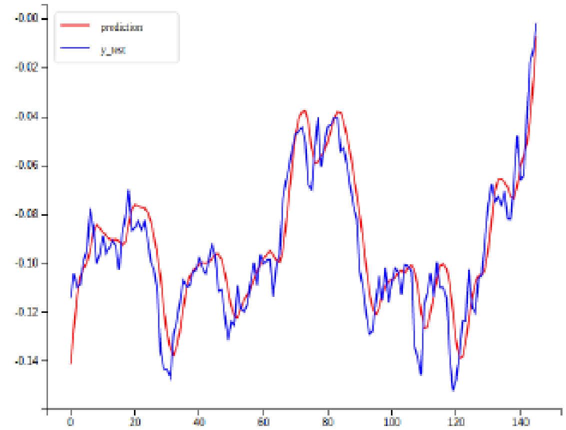

Figure 6: High/Low/Open/Close with 250 epochs

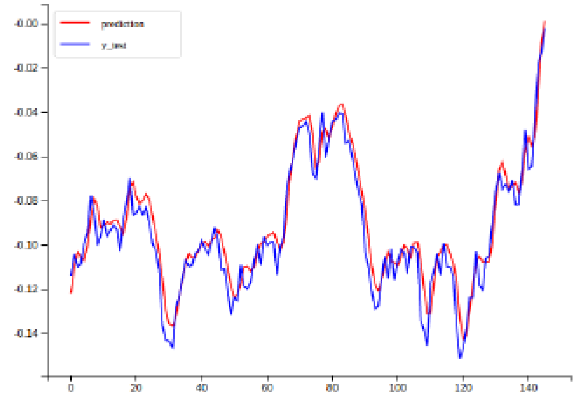

Figure 7: High/Low/Open/Close with 500 epochs

Subsequent to performing different reproductions with an alternate number of boundaries and ages, we have seen that
by taking 4 highlights set (High/Low/Open/Close) with 500 ages we accomplish the best outcomes with preparing RMSE of 0.00983 and testing RMSE of 0.00859.

## VII. CONCLUSION

The fame of stock market trading is developing rapidly, which is urging researchers to discover new techniques for the prediction using new methods. The forecasting method is not just helping the researchers moreover it helps investors and any individual dealing with the stock market. So as to help anticipate the stock indices, a forecasting model with great precision is required.

In this work, we have utilized one of the most precise forecasting technology utilizing Recurrent Neural Network and, Long Short-Term Memory unit which helps investors, analyst or any personal keen in investing into the stock market by giving them a decent information of future circumstance of the stock market.

## REFERENCES


[1] Zhi S U, Man L U, Dexuan L I. Deep Learning in Financial Empirical Applications: Dynamics, Contributions and Prospects [J]. Journal of Financial Research, 2017..

[2] Kai Chen, Yi Zhou and FangyanDai ―A LSTM-based method for stock returns prediction: A case study of China stock market, ‖IEEE International Conference on Big Data, 2015.

[3] Xu T, Zhang J, Ma Z, et al. Deep LSTM for Large Vocabulary Continuous Speech Recognition [J]. 2017.

[4] Prashant S. Chavan, Prof. Dr. Shrishail. T. Patil ―Parameters for Stock Market Prediction,‖ Prashant S Chavan et al, Int.J.Computer Technology & Applications, Vol 4 (2),337-340.

[5] Shih C H, Yan B C, Liu S H, et al. Investigating Siamese LSTM networks for text categorization[C]// Asia-pacific Signal & Information Processing Association Summit & Conference. 2017.

[6] SharvilKatariya, Saurabh Jain―Stock Price Trend Forecasting using Supervised Learning Methods.

[7] Hochreiter, Sepp, and Jürgen Schmidhuber―Long short-term memory,‖ Neural computation 9.8 (1997): 1735-1780