iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IX **Month of publication:** September 2021

DOI: https://doi.org/10.22214/ijraset.2021.38231

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

Stock Prediction Using Sentiment Analysis and LSTM

Ms. Anjima K. S
Department of Computer Science, Thejus Engineering College

Abstract: ***The stock market is a difficult area to anticipate since it is influenced by a variety of variables at the same time. The stock exchange is where equities are exchanged, transferred, and circulated. This research proposes a hybrid algorithm that predicts a stock's next day closing prices using sentiment analysis and Long Short Term Memory. The LSTM model seems to be quite popular in time-series forecasting, which is why it was selected for this project. Our proposed methodology makes use of the temporal association between public opinion and stock prices. Part-of-speech tagging is used to do sentiment analysis, and Long Short Term Memory is utilized to predict the stock's next day closing price. When these two factors are combined, we get a good picture of the stock's future. In this project, two main datasets have been used: HCLTECH company stock data and the news related to each stock of the HCL company for each day. The project is implemented by using the python programming language. The python programming language has been used to execute the project. This also incorporates machine learning along with public feedback. Sentiment analysis enables us to evaluate a diversity of political and economic factors, which have a significant impact on the stock market.***

Keywords: LSTM, sentiment analysis, RNN, Back propagation neural network.

I. INTRODUCTION

Scholars and researchers from many academic domains have been drawn to the issue of financial market modeling and forecasting. Forecasting the trend or price of stocks using machine learning techniques and artificial neural networks is the most fascinating subject to research in the current financial trading environment, particularly in the stock market. Neural networks are well-studied and have been effectively utilized for forecasting and modeling financial markets among all machine learning approaches. LSTM is meant to anticipate, predict, and categorize time series data, even when there are significant time gaps between important occurrences. LSTMs have been used to manage a variety of issues, the most well-known of which being handwriting recognition and speech recognition. When compared to traditional back-propagation neural networks and conventional recurrent neural networks, LSTM brings various advantages.

The ability of LSTM to overcome large time delays in the event of issues similar to those mentioned above is due to the constant error back propagation inside memory blocks; LSTM can manage noise, dispersed representations, and continuous values; There is no need to perfect LSTM parameters because it works well with a wide range of them, including learning rate, input gate bias, and output gate bias. This was accomplished by employing a hybrid algorithm that combines sentiment analysis and LSTM to forecast the following day's stock prices and public sentiment, allowing us to better link market circumstances and public emotion. The sentiment analysis performed on the obtained news data is then applied to the HCLTECH dataset.

II. METHODOLOGY

With the advancement of neural network and natural language processing techniques, the prediction of future price of stocks in stock market has become more easier. This proposes an algorithm which automatically predict the future price of individual stocks and also analyze public response for this stocks from a particular company. It is achieved by using combining the machine learning and the sentiment analysis, i.e here construct an hybrid algorithm which contain LSTM to predict the future price of individual stocks and also use sentiment analysis to analyze the public response . The development of the model has been divided into different sections based on context. In this section, the overall system architecture of the proposed technique is described.

Machine learning and public opinion are combined in this research. This was accomplished through the deployment of a hybrid algorithm that combines sentiment analysis and LSTM to forecast the next day's stock prices and public sentiment, allowing us to link market circumstances and public mood. Sentiment research is performed using publicly available Twitter data, while stock values are obtained from yahoo finance. The python programming language is used to carry out the project.

This also incorporates machine learning as well as public feedback. Sentiment analysis allows us to analyze a variety of political and economic factors, which have a significant impact on the stock market. LSTM, on the other hand, has shown to be the greatest algorithm for predicting stock prices since it considers prior values.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

A. LSTM

Long Short Term Memory Networks (LSTMs) are a kind of RNN that can learn long-term dependencies. LSTMs are particularly aimed at preventing the problem of long-term dependency. All recurrent neural networks are made up of a series of repeated neural network modules. This repeating module in conventional RNNs will have a relatively basic structure, such as a single tanh layer. LSTMs have a chain-like structure as well, but the repeating module is different. Instead of a single neural network layer, there are four, each of which interacts in a unique way.

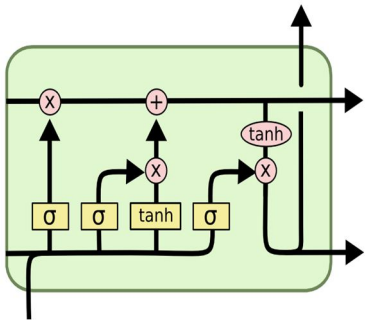


Figure 1. LSTM Architecture

B. Sentiment Analysis

Sentiment analysis is a fascinating branch of Natural Language Processing (NLP) that is used to assess the language used in a body of text. We can apply sentiment analysis to determine whether hundreds of tweets about a company are generally positive or negative. The stock market is one of the most volatile, and public sentiment can influence the market's overall trajectory. Trader mood is one of several elements that influence the stock market's movement. Stock market investing is now an unavoidable part of the financial industry, since a high stock market value is regarded as a sign of a strong economy. Sentiment analysis is the process of evaluating people's feelings and views on various platforms such as social media and similar websites, where individuals may freely express their feelings and ideas on whatever they choose. Because of the broad range of layers and their uses in Neural Networks, some of the layers are combined to form the model's overall design. It can be divided into four layers as shown in the figure below

1) *Input Layer:* This is the first layer in the model's architecture to be built. This layer receives the data that was obtained from the dataset. This data is often an input sequence that has been examined from a dataset and has gone through a number of preprocessing steps. This preprocessed input sequence is given to the architecture's next layer as an input.

Data collection: There are two main datasets are used.

a) HCLTECH dataset which is from 2012 to 2021.this dataset include

b) Second data corresponds to the news about the stock o the company HCL.

Data Preprocessing*:* Data Analysis, Data Cleaning, Data Splitting are performed in this module. The data collected from the dataset is evaluated in this phase. The dataset is analysed and all of the dimensions are identified. The data in the dataset may be impure because it contains a lot of unwanted or unnecessary words and characteristics. As a result, these meaningless words are referred to as noise. Finally, the dataset must be divided for model training and testing.

2) *Sentiment Analysis Layer:* In this project two main datasets are used. One is HCLTECH dataset and another data is collected from website. That is corresponds to the news about HCLTECH company for particular days. This news data is only given for the sentiment analysis, and calculate the sentiment of each news on each day. This calculated sentiment score then attatched to the HCLTECH dataset along with its parameters. Now the dataset contain five fields. One corresponds the calculated sentiment value. This layer consists of 4 steps.

a) Fetch News for the Current Stock and calculate its sentiment value.

b) Convert the text to sentiment score.

c) Convert news to date with score.

d) Apply sentiment score to data frame.

3) *Hidden Layer:* The LSTM unit constructs the hidden layer, which is influenced by the current and previous moment's input data. We need to import a couple of Keras modules in order to create the LSTM. The LSTM unit's forgotten gate decides which cell state data is deleted from the model. The input gate controls how much of the current time network input x_t is reserved for the cell state Ct, preventing irrelevant data from reaching the memory cells.

4) Output Layer: After the model has been trained, the stock time series is input for prediction, i.e., stock data for N days is entered to forecast the stock trend on the N+1 day. The trained model predicts the closing price of the fifth trading day based on the trading data from the first four trading days.

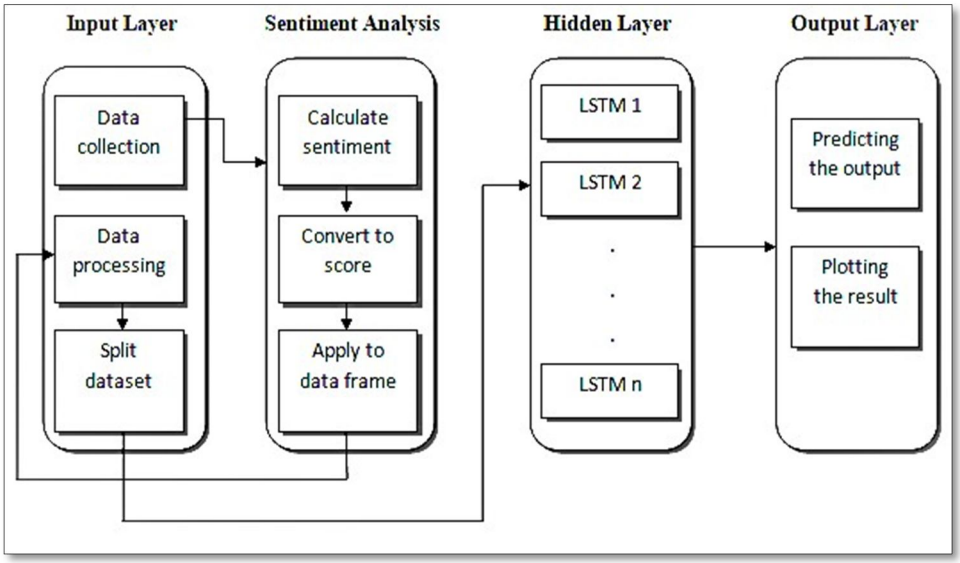


Figure 2. Overall architecture

III. RESULTS AND ANALYSIS

After the model was built and integrated with LSTM and sentiment analysis, the test dataset was used to classify them into their respective classes. The implementation results and the respective performance analysis is displayed using screenshots in the following sections.

Figure 3 is the result after performing the sentiment analysis. This news data is only given for the sentiment analysis, and calculate the sentiment of each news on each day.

```
{'2021-01-18': 0.5,
 '2021-01-19': 0.0,
 '2021-01-25': 0.0,
 '2021-01-28': 0.0,
 '2021-02-03': 0.0,
 '2021-03-10': 0.4,
 '2021-03-19': 0.0,
 '2021-03-29': 0.5714285714285714,
 '2021-04-05': 0.0,
 '2021-04-07': 0.0,
 '2021-04-19': 0.0,
 '2021-04-22': 0.0,
 '2021-04-23': 0.24444444444444446,
 '2021-04-24': 1.0,
 '2021-04-26': 0.35,
 '2021-04-27': 0.0}
```

Figure 3: sentiment analysis

This calculated sentiment score then attached to the HCLTECH dataset along with its parameters. Now the dataset contain five fields. One correspond the calculated sentiment value. The values represent the sentiment score of the particular news. Sentiment analysis (or opinion mining) is a natural language processing technique used to determine whether data is positive, negative or neutral. The value greater than .5 means that it is a positive sentiment. The value less than .5 means that it is a negative news. Value near to .5 means it is neutral.

In figure 4, the predicted price value has shown. The red line indicate the real HCL stock price. The blue colored line represents the predicted HCL stock price. Here we can see that when the real stock price moves higher the predicted stock price also moves higher. When the real stock price moves lower the predicted stock price also moves lower.

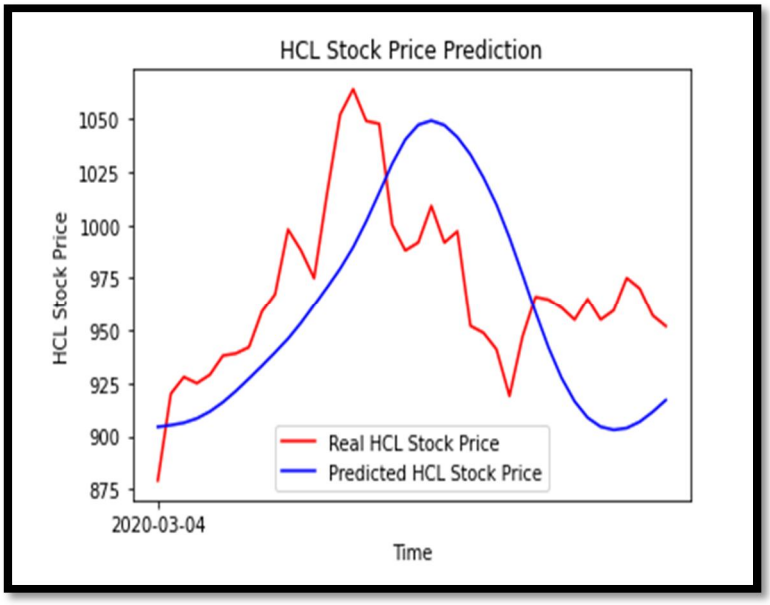


Figure 4: Predicted stock value

The model that was built for predicting the future price of financial stocks was evaluated with variousperformance measures like Accuracy, loss, error rate.

1) *Accuracy:* The most commonly used metric for measuring the performance of a classifier is accuracy. It subtracts the number of samples that were correctly predicted from the total number of samples. It displays the percentage of samples that were successfully predicted out of the total number of sample inputs.
2) *Loss:* Poor prediction results in a loss. To put it another way, loss is a statistic that indicates how inaccurate the model's forecast was for a particular case. The loss is 0 if the model's forecast is flawless; otherwise, the loss is larger. The aim of training a model is to identify a set of weights and biases that have a low loss across all cases on average.
3) *Mean square error (MSE):* Over the whole dataset, it is the average squared loss per case. MSE is calculated by adding up the squared losses for individual cases and then dividing by the number of examples. While MSE is widely used in machine learning, it is not the only practicable nor the optimal loss function for all situations.

IV. CONCLUSION

This paper presents a hybrid approach that employs a neural network LSTM to forecast stock prices and sentiment analysis to confirm our predictions. Sentiment research allows us to analyse a variety of political and economic issues, which have a significant impact on the stock market. Our findings demonstrate that there is a connection between public opinion and stock prices. On the other hand, LSTM has shown to be the best algorithm for predicting stock prices since it considers prior values but also utilises forget gates to delete extremely old data that are unlikely to affect the future outcome, making it very efficient. There are number of further directions that can be investigated beginning from this project. Long Short Term Memory algorithm was used to optimise the model in the form of small step iteration taken by small batch gradient descent algorithm in Attention-LSTM algorithm training, so that the model could be approximated at a quicker speed with a lower error.

V. ACKNOWLEDGEMENT

I am thankful to my project coordinator, Mr Valanto Alappatt, Assistant Professor of Computer Science and Engineering Department for providing all the necessary facilities. I sincerely express my gratitude to Dr. K Vijayakumar, Principal of Thejus Engineering College, Vellarakkad and Mrs.Vineetha A.V, Assistant Professor & Head of Department of Computer Science and Engineering, for encouraging and supporting me throughout the project.

REFERENCES

[1] Di Persio, L., Honchar, O. (2016). Artificial neural networks architectures for stock priceprediction: Comparisons and applications. International journal of circuits, systems and signal processing, 10(2016), 403-413.

[2] Mittal, A., Goel, A. (2012). Stock prediction using Twitter sentiment analysis. Standford University, CS229 (2011http://cs229.stanford.edu/proj2011/GoelMittal- StockMarketPredictionUsingtwitterSentimentAnalysis.pdf), 15.

[3] Shah, V.H. (2007). Machine learning techniques for stock prediction. Foundations of Machine Learning— Spring, 1(1), 6-12.

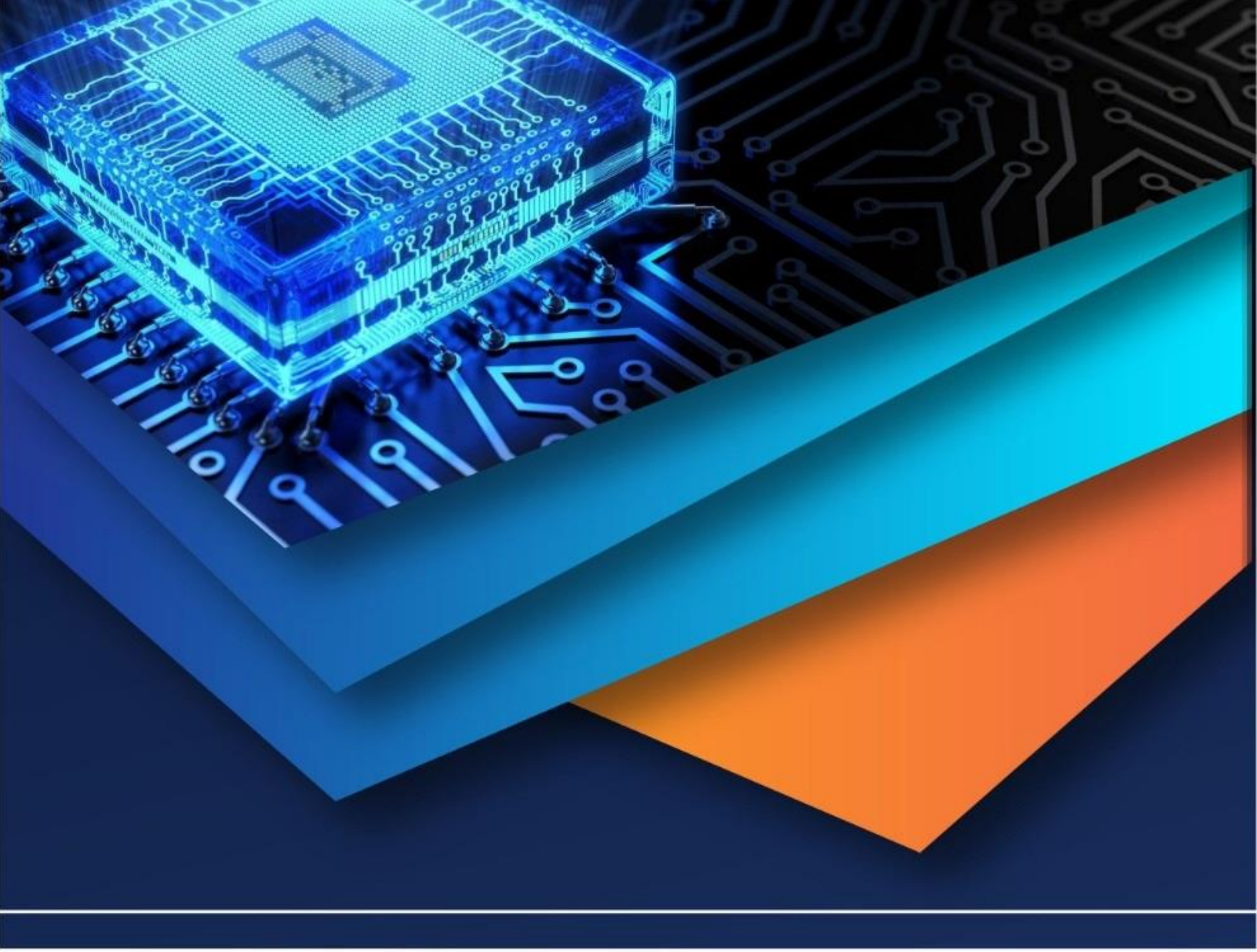
[4] Khan, Z.H., Alin, T.S., Hussain, M.A. (2011). Price prediction of share market using artificial neural network (ANN). International Journal of Computer Applications, 22(2), 42-47.

[5] Shen, S., Jiang, H., Zhang, T. (2012). Stock market forecasting using machine learning algorithms. Department of Electrical Engineering, Stanford University, Stanford, CA, 1- 5.

[6] https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=865327

[7] Shen, S., Jiang, H., Zhang, T. (2012). Stock market forecasting using machine learning algorithms. Department of Electrical Engineering, Stanford University, Stanford, CA, 1- 5.

[8] Hegazy, O., Soliman, O.S., Salam, M.A. (2014). A machine learning model for stock market prediction. *arXiv preprint arXiv:1402. 7351*



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY