Stock trend prediction and analysis using LSTM neural network and dual moving average crossover algorithm

Rahul Pradhan[1], Kanishk Bharadwaj[2], Aman Saxena[3] and Vikram Rajpoot[4]

GLA University, 17km Stone, NH-2, Mathura-Delhi Road Mathura, Chaumuhan, Uttar Pradesh 281406, India

[1] rahul.pradhan@glau.ac.in, [2] kanishk.bharadwaj_cs18@glau.ac.in, [3] aman.saxena_cs18@glau.ac.in, [4] vikram.rajpoot@glau.ac.in

Abstract. In today’s world where data is the new currency, computer machines are capable enough to find and predict the future by using the previous data and the algorithms which are becoming more precise day by day. The proposed solution by us includes the pre-processing of the 10 years of Bombay Stock Exchange (BSE) stock market data of HDFC Bank. Our solution implemented the engineering techniques of deep learning and machine learning which are combined with the LSTM neural network algorithm and Dual Moving Average Crossover algorithm.

1. Introduction

Investing in the stock market has always been a dream of common men who want to invest money but as we all are familiar with the fact that banks usually provide an interest rate of 3% to 4% on savings accounts which is very low but still people use this as their primary choice.

Now here the question is, why they choose banks over other choices of investments. This has a very common and traditional answer, they have the security of money in banks. This means they will get their total invested money whenever they want and with some percent of simple interest added to the total amount as well.

During investing in the share market people are usually afraid of losing money as they have almost zero knowledge about the share market and the one who is having knowledge about it is also somewhere afraid that either he would gain or totally lose money.

After analyzing the stock market, we have found that there somewhere exist some trends in the stock price of the companies in the share market. We found that there are 2 major stock markets in India where people are interested in investing which are Bombay Stock Exchange (BSE) and National Stock Exchange (NSE), NSE has 1328 listed companies out of which the top 50 companies stock price decide the performing index (NIFTY) whereas BSE has 5000 listed companies and here only the top 30 companies decide the percentage (SENSEX analysis). Hence it is clear that SENSEX is a better performance index than NIFTY.

Technical analysts and brokers generally believe that the prices proceed in trends and it tends to repeat themselves after a period when it comes to the market’s overall mindset. Using the previous year’s data of companies, we aim to use the past performance to predict the future trends in the stock.
market [1]. We have used HDFC Bank for this analysis which is a part of BSE and SENSEX. We have taken around 10 years of previous data (2011-2020) of HDFC Bank and performed the analysis on them. The algorithms we have used are DMAC [2] and LSTM[3] to predict the behavior of the market.

We chose the LSTM[3] neural network over RNN[4] because RNN has a drawback of looking at the recent data to perform the current task. It can only predict based on recent information and it also has a problem called long term dependency problem. That’s why we used the LSTM neural network, it is designed to avoid long-term dependency problems. It can remember the information for a long period of time, basically, it’s the default behavior of LSTM[5,6]. So, it can predict easily based on the previous 10 years of the dataset.

2. Related Work

| Author | Approach | Dataset Used | Pros | Cons |
|--------|----------|--------------|------|------|
| Kim and Han[2] | ANN and Genetic Algorithm (GA) | They used the Korean Stock Price Index (KSPI) | They use GA to optimize the model of ANN | Issue with ANN models is, it needs lots of data because of the large number of connections in ANN. |
| | | The dataset contains 2928 days of trading data from Jan 1989 to Dec 1998 | They use feature discretization to optimize or reduce the dimension of input data | | |
| Jingyi Shen and M. Omair Shafq[1] | PCA algorithm to reduce the dimensions of input data and LSTM model to predict the stock price trend | 3558 stocks from Chinese Stock Market | Research Paper uses LR algorithms which achieve the same result as SVM and MLP but take less training time as compared to other algorithms. | Research Paper only focus on short term trend prediction |
| | | | Their LSTM model achieves 93.25% accuracy. | | |
| Anton H.Hofman and Gerrit ten Brinke[3] | They used GA optimized and extracted features with a fuzzy | Stock Market dataset of 200 largest companies listed in the | They proposed an optimized model for selection and scoring of stocks | They did not compare the model with the pre-existing |
membership function with a model, for optimizing stock scoring.

apart and distinguished with the traditional prediction model. They majorly focused on rankings, selection, and performance evaluation which are the factors one should keep in mind.

Manoj Thakur and Deepak Kumar[7]

They used Random Forest (RM) for pruning of features in the approach. The algorithms were with a WMGESVM to generate the “Buy/Hold/Sell” signals for the analysis.

They used Random Forest, NIFTY 50, S&P 500, DOW JONES, and NIFTY BANK.

The issue with their work was that they didn’t mention any kind of time or space complexity of their work which showed that irrespective of the accuracy of their algorithm the time and space factor is somehow lost.

McNally S, Roche J, Caton S[8]

They used RNN and LSTM models and Boruta Algorithm which works the same as a random forest classifier.

They used the Bitcoin trading dataset. The dataset contains data from 19 Aug 2013 to 19 July 2016.

They used various optimization to improve deep learning algorithm.

They majorly treated questions as a time sequence problem.

Their dataset contains various noise which directly affects the prediction

algorithms we use which made it more challenging to identify if GA would perform better than the other pre-existing algorithms.
Cheng-Lung Huang and Cheng-Yi Tsai [6] They used filter-based feature selection which is assembled with a hybrid Self-Organizing Feature Map (SOFM) [5] Taiwan index futures (FITX) for prediction They also applied feature selection before the processing for improving the prediction accuracy and to reduce the computational complexity of processing the stock market data. The algorithm could only predict for next day index price which is not feasible for long-time market investors. Moreover, the evaluation was not strong enough as they set a single SVR model as the baseline, and they didn’t compare previous works with their model’s performance.

Fischer T, Krauss C [7] They use Artificial Recurrent Neural Network architecture LSTM They used the S & P 500 Index dataset. The dataset contains data from Dec 1989 to Sept 2015 They used the latest artificial recurrent neural network architecture. Their LSTM model performs far better than Standard Deep Neural Network (DNN) and regression algorithms. They too much relied on LSTM. They didn’t mention how they trained LSTM model with long time prediction They didn’t show when to buy and sell the stock. They just predicted the future price of the stock market.

Thakkar and Chaudhari [9] By using cross-reference they predicted the price of one market with the trained data of another share market They used BSE and NSE data. They used the cross-reference technique which is not often used and it is a good choice They did not use the data of companies other than IT companies. All types of companies show different trends.

M. Ananthi & K. Vijayakumar [10] They used candlestick regression and market trend prediction (CKRM) They used tech companies datasets They did a good comparison among many algorithms. They did not focus on trends analysis and they also did not take the dataset of companies other than tech companies.
3. Proposed Approach
In this work we took the last 10 years of the data of HDFC Bank from www.yahoo.com [11] and then we cleaned the dataset. We then extracted the weighted features which showed correlation. We then reshaped the data according to our needs and according to the need of the LSTM neural network.

We fed that data to the LSTM model and tested our model on that dataset followed by training our model with another splitted part of the dataset.

We then used DMA algorithm to generate buy and sell signals and visualized those signals.

This diagram shows our proposed approach for this research work. Here we have made a flowchart to illustrate the approach.

4. Experimental Setup and Result Analysis
This section explains the ML models and algorithms used in this prediction.

We used datasets of HDFC Bank. The data used was from 01-01-2012 to 17-12-2019 dates. The dataset was imported by using .data_reader() which is a method of pandas library in Python. Furthermore we analyzed the correlation between the attributes of the dataset and reached the conclusion that taking the `close` attribute is overall a good choice to analyze and predict the stock market trends (figure 2).

We reshaped the dataset three-dimensionally and then we transformed those values into 0 and 1 scale followed by splitting the data into 2 parts, train data, and test data. Furthermore, we split the train and test data into 2 parts, x and y train data & x and y test data.

We built the LSTM model consisting of 50 neural combined with 25 dense layers. Afterward, we trained the data by using 80% of the total data in the dataset, for all this we used the timestamp of 60 days to predict the further values.

Explanation of Dual Moving Average Crossover Algorithm[12] - Dual moving crossover is one of the most used trading strategies. It is exactly like the sliding windows technique which is widely used in computer programming. In this, we fix the size and slide it over the whole list, and we keep taking averages (mean value) of all the values coming inside the window.
Suppose we have a list $[1,5,7,8,2,5,8,5,1,5,7,9,2,5]$ and size is 6, then the average will be like. 
\[
\frac{(1+5+7+8+2+5+8+5+1+5+7+9+2+5)}{6} \quad \ldots \quad \text{up to the size of the list.}
\]

DMAC algorithm is used by 2 windows one of size 30 and another of size 100.

Simple Moving Average (SMA) is a technique of DMAC. Whenever the SMA30 becomes greater than SMA100, then it is a signal to buy the stocks and if it becomes less than SMA100, then it signals to sell the stocks.

![Closing Price History](image)

**Figure 2:** Date vs. Closing Price in INR

5. Result
The trend was analyzed by using the predicted 30 days’ values and previous values of the stock market.
Figure 3: Days of Months vs. Stock Price

This picture figure 3 shows the days (x-axis) vs stock price (y-axis) of predicted 30 days price.

Figure 4: Years vs. Closing Price in INR (with train, Val, and Predicted Data)

The above picture figure 4 shows the date (x-axis) vs closing value (y-axis) where the blue color line shows the train values, orange color lines show the actual values and yellow color shows the predicted values. Here we found the accuracy of our LSTM model to be 94.8%.
The above graph figure 5 is a DMAC graph which shows the values of SMA30 and SMA100 with the values of HDFC Bank’s original values of stocks and HDFC Bank’s predicted values of the stock price.

The above graph (figure 5) shows the legends to buy and sell the stocks by the use of the Dual Moving Average Crossover (DMAC) algorithm.

In this graph (figure 5), the red color down arrow shows when to sell the stocks and the green color up arrow shows when to buy the stocks. These both generated signals are at the intersection points of SMA30 and SMA100. The detailed conclusion from this graph is in the conclusion section.

5.1 Result Table

| Topic                   | Detailed Explanation                                                                                                                                                                                                 | Remarks and Results                                                                 |
|-------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| LSTM neural Network     | We used this to perform the prediction for stock market prices. We first predicted the values for the years’ data we already had and then we found the accuracy of the model to be 94.8%. We then used that model to predict the prices of the months of which we were not having data. | We found that accuracy of the model is 94.8% which shows that the model is very efficient. |
| DMAC Algorithm          | We used this algorithm to generate signals. For further detailed explanation of the DMAC algorithm refer to [9].                                                                                                      | We found that this algorithm is very efficient and there was not any point in the graph where we can see that generated signals are wrong. |

6. Conclusion

Our work was based on predicting the stock market price with the use of the DMAC algorithm and LSTM neural network. We achieved an accuracy of 94.8% from which we can conclude that analyzing trends in the Stock Market is also a good choice and a lot of research work can be done in this aspect of the stock market.
6.1 Justification of Results

By visualizing the figure 5 we found that Dual Moving Average Crossover (DMAC) is also an accurate and efficient way to analyze trends in the Stock Market. We also found that investing in the stock market using trends analysis (or we can say long trends) is a very low risk task or an almost zero risk task as the signals are very accurate and we couldn’t find any case where the generated signals are showing wrong indications of buying and selling, which could result into the loss of money.

![Graph showing Years vs. Closing Price in INR](image)

**Figure 6:** Years vs. Closing Price in INR (with train, Val, and Predicted Data)

It is very clear from this graph (figure 6) that the predicted values and original values are very near to each other which shows that trends also play a major role in predicting the prices of the stock market just exactly like sentiments do.

7. Future Scope

The future work which we will do in this research work is, we will compare other machine learning algorithms like Random forest classifier and multiple regression etc. and other trend analysis algorithms like EMA, MACD and WMA etc. to find the differences between them and also we will find which is the most efficient algorithm for different markets.

We will also use datasets of other stock markets and datasets of different types of companies to find if there exists no pattern or some unique patterns in those companies and markets.

References

[1] Shen J and Shafiq M O 2020 Short-term stock market price trend prediction using a comprehensive deep learning system J. Big Data 7

[2] Kim K-J and Han I 2000 Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index Expert Syst. Appl. 19 125–32

[3] Olah, C., 2021. *Understanding LSTM Networks -- colah's blog*. [online] Colah.github.io. Available at: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> [Accessed 10
February 2021].

[4] Hofman A H, ten Brinke G and Loos K 2016 Hierarchical structure formation in supramolecular comb-shaped block copolymers Polymer (Guildf.) 107 343–56

[5] Mahrishi, M., Hiran, K. K., Meena, G., & Sharma, P. (2020). Machine Learning and Deep Learning in Real-Time Applications. IGI Global.

[6] Huang C-L and Tsai C-Y 2009 A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting Expert Syst. Appl. 36 1529–39

[7] Fischer T and Krauss C 2018 Deep learning with long short-term memory networks for financial market predictions Eur. J. Oper. Res. 270 654–69

[8] Thakur M and Kumar D 2018 A hybrid financial trading support system using multi-category classifiers and random forest Appl. Soft Comput. 67 337–49

[9] Thakkar A and Chaudhari K 2020 Predicting stock trend using an integrated term frequency–inverse document frequency-based feature weight matrix with neural networks Appl. Soft Comput. 96 106684.

[10] Ananthi M and Vijayakumar K 2020 Stock market analysis using candlestick regression and trend prediction (CKRM) J. Ambient Intell. Humaniz. Comput.

[11] McNally S, Roche J and Caton S 2018 Predicting the price of bitcoin using machine learning 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP) (IEEE) pp 339–43

[12] Kumar G, Jain S and Singh U P 2020 Stock market forecasting using computational intelligence: A survey Arch. Comput. Methods Eng.