OPTIMIZATION OF PARAMETER SELECTION FOR TIME SERIES DATA ANALYSIS THROUGH RECURRENT NEURAL NETWORK

[1]Vivek Uprit, [2]Dr.Pritej Yadav, [3]Dr. Sanjeev Gupta

[1]PhD Scholar, [2]Associate Professor, [3]Dean
Computer Science and Engineering (CSE),
Rabindranath Tagore University, Bhopal, MP, India

Abstract: Time Series Data is the most common but crucial type of data which is generated on day to day basis by humans. Every aspect for humanity is impacted due to this type of data. As the time series data can be univariate or multivariate thus the effective analysis and prediction of data depends on selection of most impacting parameters. So in this study we have tried to optimize the number of parameter that aids in the effective and efficient Data analysis by using the concepts of Deep learning. By the utilization of virtue of LSTM method instead of one multiple parameter were selected that enhanced the prediction by greater extent.

Index Terms - **Time Series Data, Optimization, Deep learning, RNN**

I. INTRODUCTION

In the Current era of Technology, each and every field of Humanity like Business, Medicine, Entertainment, etc are affected by data. Mostly this data is time series in nature. Organizations may better comprehend systemic patterns across time by using time series analysis. Business users may examine seasonal trends and learn more about their causes using data visualizations. These visualizations can do much more than just display line graphs using today's analytics solutions. Organizations may use time series forecasting to estimate the likelihood of future occurrences when they study data at regular intervals. Predictive analytics includes the prediction of time series data. It can indicate expected data changes, such as seasonality or cyclical behavior, which improves forecasting and gives a better understanding of data factors.

For the above-specified issue, computer science can play a vital role in solving the case, since by using the machine learning branch of Computer Science this problem can be tackled. For better understanding, we took the problem of Stock Price Prediction where the use of computer science for time series data analysis can be done easily. Since this problem relies on times series data where the current value of the stock depends on the previous value of the Stock. Therefore it can be solved using Deep Learning, more specifically RNN (Recurrent Neural Network) will be required to solve the problem. Here we are going to use the LSTM Algorithm i.e. Long Short Term Memory Algorithm which is a special case of Recurrent Neural Network. In the next section the available literature is reviewed to determine the step-by-step solution for the Stock Market prediction problem.

II. REVIEW OF LITERATURE

This section provides a detailed review of earlier work in the area of Time Series Analysis. The experiment's starting point is then established by a list of observed points. Here, the application of machine learning to time series data is established, along with the many methods that have already been studied.

The usage of a Deep learning framework for financial data forecasting was outlined by the author in [3]. The neural networks in this case were trained using a deep learning framework. The outcomes obtained using neural networks were then contrasted with those obtained using more conventional benchmark techniques for financial data forecasting. The findings of the neural network are discovered to be more reliable than those of the conventional approach.

The author has concentrated on the usage of machine learning algorithms in the sector of finance [5]. With the development of machine learning algorithms in finance, numerous important activities, including portfolio design, risk modeling, forecasting, and decision making, may be completed with considerably greater efficiency.

The deep learning-based approaches and the process of using them to forecast stocks are both covered in this paper by the authors [6]. Data from the Chinese stock market is collected, preprocessed, and then deep learning models are applied to the data for value prediction.

The application of ANN and Random Forest approach in the Stock Prediction issue is discussed in this study by [7] authors, where data from 5 distinct businesses are used. ANN and RF techniques are used to predict outcomes. then use RSME and MAPE as assessment indicators. ANN techniques are discovered to be more effective when used to forecast the value

The author has shown all of the material on LSTM here in [9]. (Long Short Term Memory Networks). It is a unique kind of RNN that can learn long-term dependencies, according to the author. Avoiding the issue of long-term reliance is the main driver for the creation of LSTM. As seen in Figure 1, it often has several levels. These layers are made up of several cells, each of which performs a variety of roles as shown in figure 2.
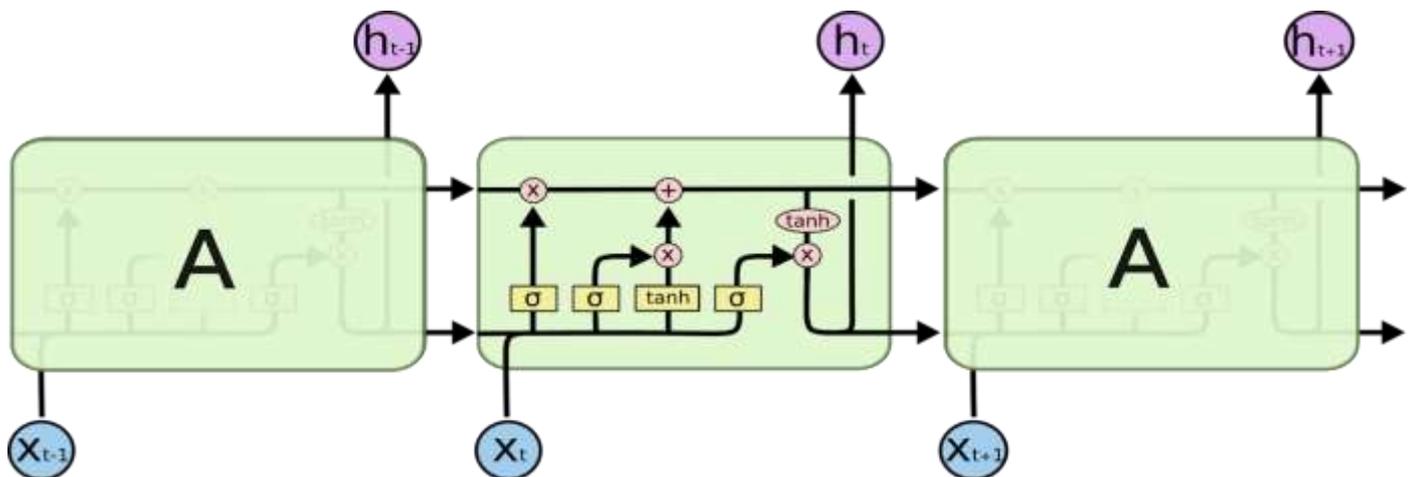

Figure 1 : Cell of the system LSTM

$$O_t : Output\ gate\ activation\ vector$$
$$f = \sigma_g(w_f x_t + U_f C_{t-1} + B_f)$$
$$i_t = \sigma_g(w_i x_t + U_i C_{t-1} + B_i)$$
$$O_t = \sigma_g(w_o x_t + U_o C_{t-1} + B_o)$$
$$C_t = f_t o\ C_{t-1} + i_t\ o\ \sigma_c(w_c x_t + B_c)$$
$$h_t = \sigma_h(O_t o\ C_t)$$
$$x_t : input\ vector\ to\ the\ LSTM$$

Figure 2 : LSTM Structure

III. METHODOLOGY

The procedure, dataset, and experimental setup that will be utilized to forecast a stock's value are described in this section. We have used the dataset from Kaggle for this experiment. The Nifty 50 businesses on the Indian Stock Market are all represented in this data by their daily records. The data used in this experiment spans a 5-year period from 2017 to 2021. where there are two portions, one lasting four and a half years and the other just a half year, to the five year term. The first component is used as the machine learning algorithm's training set, while the second part is used as the algorithm's testing set.

Table 1: Dataset field Description

Dataset Field Name	Specification
Date	Date specifies the Date of record for the Stock
High	Highest Value of the Stock
Low	Lowest Value of the Stock
Open	Opening Value of the Stock
Close	Closing Value of the Stock on the Date of Record
Volume	Volume of the Stock
Adj Close	Adjusted Closing Price of the Stock

. The experimental configuration employed here utilizes an Intel® core TM i5-9500U CPU running at 2.70GHz across four CPUs, together with 8GB of memory. For the creation of LSTM and other automation, Python is employed as the programming language. Important libraries used in this project are Pandas, Matplotlib, Numpy, and Keras.

The data from the stock must first be obtained in order to start the experiment, and only then may additional steps be taken. Here, the Nifty 50 Kaggle dataset is used. The Records are the subject of the experiment, as was previously stated in this section. Since the LSTM (Long Short Term Memory) Algorithm is particularly sensitive to data value,

processing of the common separated value file of the data begins as soon as it is uploaded in Python for the building of Data frame. The value is to be normalized in a range of 0 to 1.

Data that has undergone preprocessing has been separated into training and testing sets. Then, using an approach where the closure price is dependent upon the previous closing price, an LSTM Model is built with several layers. In this case, 100 layers are thought to exist. Once this model has been created, training is given to it so that it may educate itself using the Training Set and Testing Set.

Following the prediction of this future value on a predetermined training set, the accuracy of the results is determined by comparing them to the equivalent value in the testing test. The Model is now prepared to forecast more future values. For pattern analysis and decision support, the observed value and the expected value might be shown on a graph.

1. Collection of data in CSV format
2. Preprocessing of data
3. Selection the parameter to be used
4. Dividing the data to training and testing set
5. Create LSTM Model
6. Pass the Training Set to LSTM to train it
7. Predict the future value
8. Now compare the observed value with the corresponding value in testing set
9. Plot the observed value and predicted value in a graph

IV. RESULTS AND DISCUSSION

This part records and analyzes the outcomes of the algorithm that was used in the methodology section. The graph for the stock for the opening and closing prices is shown in the figure. The graph that was drawn for the forecast made and the observed Actual Value is then specified in the figure.
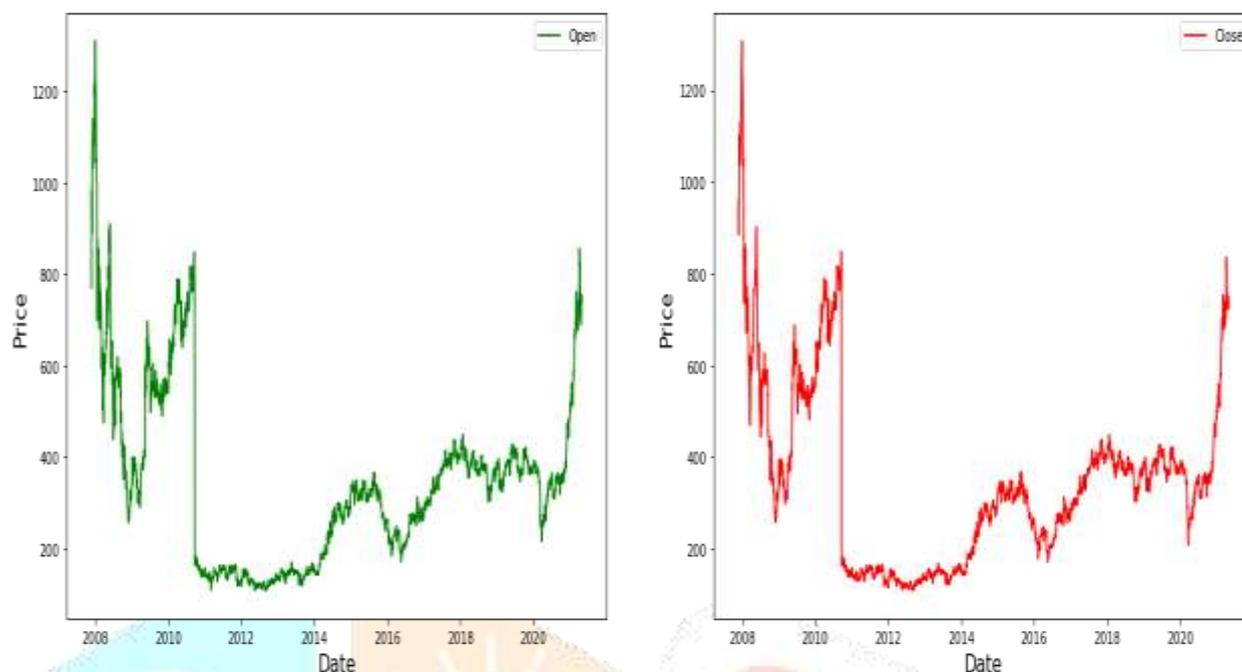

Figure 3 : Plotted Data for two different Parameters

Date	Open	Close	Open_predicted	Close_predicted
2018-11-06	331.00	325.60	332.002716	328.242889
2018-11-07	329.00	326.20	329.272766	325.423706
2018-11-09	328.00	337.60	327.819519	323.849854
2018-11-12	339.00	328.65	334.373657	330.394562
2018-11-13	327.65	334.10	333.620575	329.705597
2018-11-14	335.30	332.00	333.974396	330.042725
2018-11-15	334.00	346.20	334.216064	330.280457
2018-11-16	346.00	347.45	342.582947	338.686218
2018-11-19	348.20	350.55	348.445923	344.629547
2018-11-20	349.85	355.25	351.628693	347.770935

Figure 4 : Predicted Data
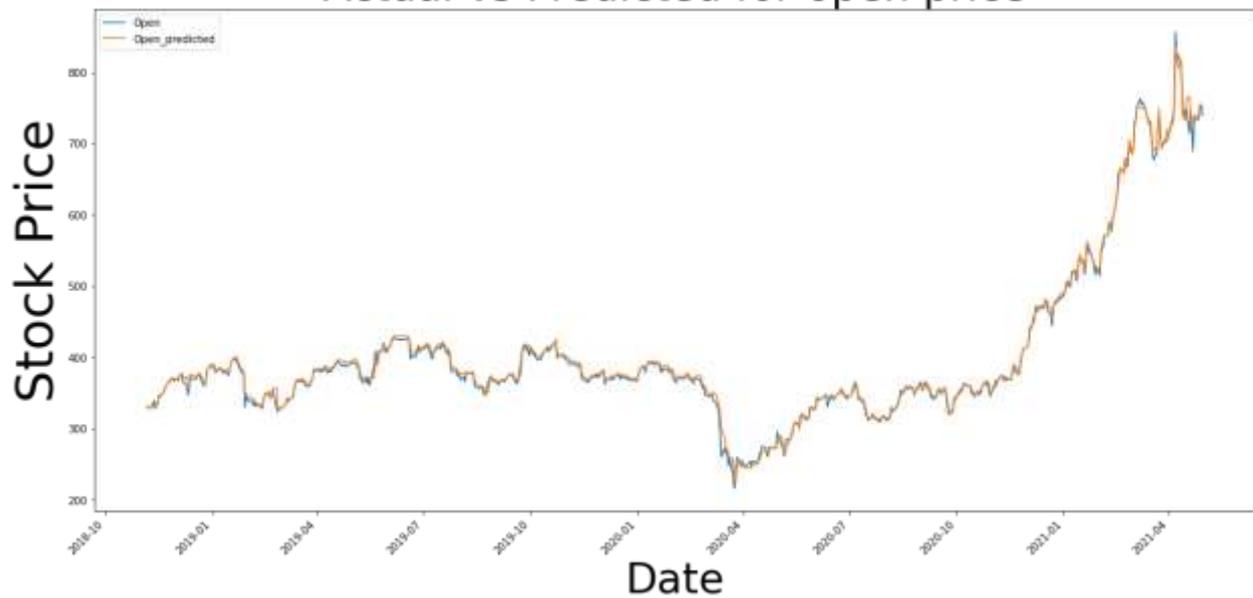
Figure 5 : Actual Vs Predicted Opening Price

Figure 6 : Actual Vs Predicted Closing Price
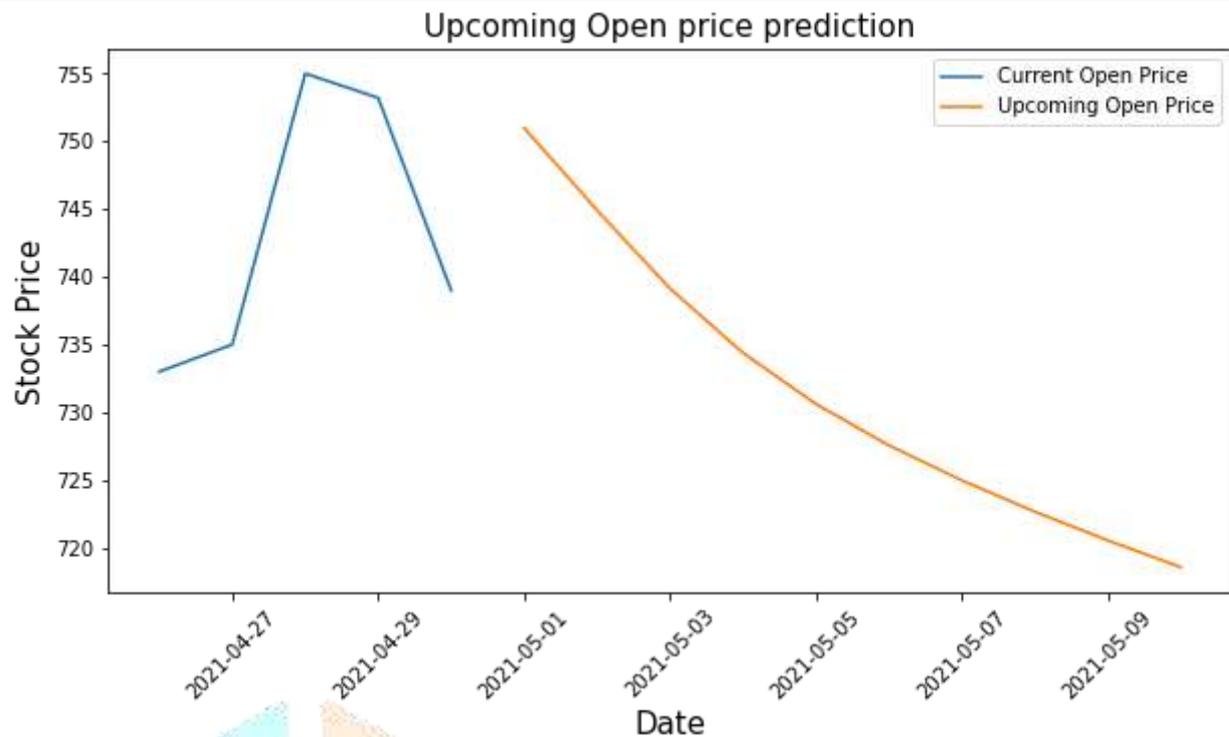
Figure 7 : Forthcoming Trend Predicted

In order to test the accuracy of our model, the predicted values are subjected to RSME (Root Square Mean Error) Method using the formula

Where Yi is the Actual Value of the Stock, Pi is the Predicted Value of the stock, n is the No. of Iterations for Date. After subjecting 150 Different rows of Data consisting of the above specified values The Calculations were made and the RSME value is found to be

V. CONCLUSION

Because of the numerous and sometimes uncontrollable elements that impact the stock market, making predictions about it is typically challenging. In this essay, we've used the closing price as a metric to estimate the stock's potential worth. As a consequence of the results' about 57% accuracy, they were deemed adequate. Additionally, the variance in prediction was rather little with an RSME value of 5.7652. However, there is also room for improvement in accuracy in the future by include other important aspects like volume, P/E Statement, the impact of social events, etc.

REFERENCES

[1] Securities and Exchange Board of India, SEBI Annual Report 2021,https://www.sebi.gov.in/reports-and-statistics/publications/aug-2021/annual-report-2020-21_51610.html

[2] R.Vopai, "NIFTY-50 Stock Market Data (2000 - 2021)" , https://www.kaggle.com/rohanrao/nifty50-stock-market-data

[3] S.Emerson, R. Kennedy, L.O'Shea, J. O'Brien,"Trends and Applications of Machine Learning in Quantitative Finance" (May 30). 8th International Conference on Economics and Finance Research (ICEFR 2019), , **2019** Available at SSRN: https://ssrn.com/abstract=3397005

[4] Nayak, M. M. Manohara Pai, R. M. Pai, "Prediction Models for Indian Stock Market",Procedia Computer Science, Volume **89**, Pages **441-449**, ISSN 1877-0509, **2016**

[5] M. Vijh, D.Chandola, V.Anand Tikkiwal, A.Kumar, "Stock Closing Price Prediction using Machine Learning Techniques", Procedia Computer Science, Volume **167**, Pages **599-606**, ISSN 1877-0509, **2020**

[6] J.Shen, M.O.Shafiq, "Short-term stock market price trend prediction using a comprehensive deep learning system". J Big Data 7, 66 ,**2020**.

[7] A.Moghar, M.Hamiche, "Stock Market Prediction Using LSTM Recurrent Neural Network", Procedia Computer Science, Volume **170**, , Pages **1168-1173**, **2020**

[8] S.Banyal, P.Goel, D.Grover, "Indian Stock-Market Prediction using Stacked LSTM AND Multi-Layered Perceptron" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-**9** Issue-**3**, **2020**

[9] C.Olah, (2015). Understanding lstm networks–colah's blog. Colah. github. Io