Stock Price Prediction using Machine Learning with Python

Pijus Pyke[1], Prajit Paul[2], Sayan Bhattacharya[3], Rima Dawn[4], Paramita Das[5], Koushik Chakraborty[6], Sourav Kumar Pandey[7]
[1,2,3,4,5,6,7] *Electronics and Communication Engineering, Asansol Engineering College, West Bengal, India*

***Abstract*— In Stock Market Prediction, the aim is to predict the future value of the financial stocks of a company. The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values. Machine learning itself employs different models to make prediction easier and authentic. The paper focuses on the use of Regression and LSTM based Machine learning to predict stock values. Factors considered are open, close, low, high and volume.**

***Index Terms:* future extraction, NIFTY value prediction, NSE share market, stock market, stock price prediction.**

1.INTRODUCTION

Online stock trading on the Web is increasing dramatically. Active trading con-tributes significantly to economic growth and leads to promotion of economic efficiency. However, it is said that more than 90% of newcomers as personal in-vectors withdraw within one year because of the difficulty of the trading. There-fore it is important for the personal investors to support the trading. Natural language processing is one useful approach for the task. Many researchers have proposed methods using text information for analysing the market. They generated models to predict the stock market, such as "UP" or "DOWN" of stock prices, by using texts and machine learning techniques.

A stock market is a public market for the trading of company stocks. Stock market allows us to buy and sell units of stocks (ownership)of a company. If the company's profits go up, then we own some of the profits and if they go down, then we lost profits with them.

2. OBJECTIVE

Machine learning is effectively implemented in forecasting stock prices. The objective is to predict the stock prices in order to make more informed and accurate investment decisions. This is important in our case because the previous price of a stock is crucial in predicting its future price.

2.1 Literature Survey

There were two important indicators in the literature for stock price forecasting. They are fundamental and technical analysis. Both were used to analyze the stock market

2.1.1 Prediction Techniques

Presented the recent methods for the prediction of stock market and give a comparative analysis of all these Techniques. Major prediction techniques such as data mining, machine learning and deep learning techniques used to estimate the future stock prices based on these techniques.

Holt-Winters

Holt-Winters is the appropriate or correct mode when the time series has trend and seasonal factors. The series was divided into three components or parts that are trend, basis and seasonality. Holt-Winters find three trend, level, and seasonal smoothening parameters. It has two variants: Additive Holt Winters Smoothening model and Multiplicative Holt-Winters model. The former is used for prediction and the latter is preferred if there are no constant seasonal variations in the series. It is mainly popular for its accuracy and in the field of prediction it has outperformed many other models. In short—term forecasts of economic development trends, Holt-Winters exponential smoothing method with the trend and seasonal fluctuations is usually used. After removing the seasonal trends from the data, the following function is taken as an input and in return,

the Holt-Winters makes the pre-calculations necessary for the purpose of forecasting. All parameters required for the forecasting purpose are automatically initialized based on the function data.

$$HWStock1\_ng = HoltWinters(ds, gamma = FALSE)\ predHW = predict(HWStock1\_ng\ n, ahead = 9)$$

Artificial Neural

Network An artificial neural network (ANN) is a technique inspired from biological nervous system, such as the human brain. It has a great ability to predict from large databases. On the basis of the back—propagation algorithm, ANN is generally used to forecast the stock market. In the back—propagation algorithm, a neural network of multilayer perceptron (MLP) is used. It consists of an input layer with a set of sensor nodes as input nodes, one or more hidden layers of computation nodes and computation nodes of the output layer. These networks often use raw data and data derived from the previously discussed technical and fundamental analysis. A Multilayer Feed Forward Neural Network is a neural network with an input layer, one or more hidden layers and an output layer. Inputs correspond to each training sample measured attributes. Inputs are passed to input layer simultaneously. The weighted outputs of these units are fed to the next layer of units that make up the hidden layer simultaneously. The weighted outputs of the hidden layers act as an input to another hidden layer, etc. The hidden layers number is an arbitrary design problem. The weighted output of the last the hidden layer acts as inputs to the output layer, which predicts the networks for certain samples. Important parameters of NN are learning rate, momentum and epoch. Back propagation is a neural network learning algorithm. The back propagation network learns by processing the sample set repeatedly and comparing the network prediction with the actual output. If the residual value exceeds the threshold value, the weight of the connections is modified to reduce the MSE between the forecast value and the original value. The weights are changed from the output layer to the first hidden layer in the opposite direction. Since the changes in the weights of the connections are made in the reverse direction, the name given to the algorithm is Back propagation. Use the back propagation algorithm to perform the calculations and compare the predicted output and target output. The predicted value is not closer to the actual value and the weights are modified.

ARIMA model

This ARIMA model was introduced by Box and Jenkins in 1970. The Box—Jenkins's methodology is also referred to as a set of activities to identify, estimate and diagnose ARIMA models with time series data. The model is the most important financial forecasting method. Models from ARIMA have been shown to be effective in generating short-term forecasts. The future value of a variable in the ARIMA model is a linear combination of past values and past errors.

3. MATERIALS AND METHODS

Stock market prediction seems a complex problem because there are many factors that have yet to be addressed and it doesn't seem statistical at first. But by proper use of machine learning techniques, one can relate previous data to the current data and train the machine to learn from it and make appropriate assumptions. Machine learning as such has many models but this paper focuses on two most important of them and made the predictions using them.

Regression is used for predicting continuous values through some given independent values. The project is based upon the use of linear regression algorithm for predicting correct values by minimizing the error function as given in Figure1. This operation is called gradient descent. Regression uses a given linear function for predicting continuous values: Where, Vis acontinuous value; K represents known independent values; and, a, b are coefficients. Work was carried out on csv format of data through panda library and calculated the parameter which is to be predicted, the price of the stocks with respect to time. The data is divided into different train sets for cross validation to avoid over fitting. The test set is generally kept 20% of the whole dataset. Linear regression as given by the above equation is performed on the data and then predictions are made, which are plotted to show the results of the stock market prices vs time.

$$V = a + bK + error$$

Where, Vis acontinuous value; K represents known independent values; and, a, b are coefficients. Work was carried out on csv format of data through panda library and calculated the parameter which is to be predicted, the price of the stocks with respect to time. The data is divided into different train sets for cross validation to avoid over fitting. The test set is generally kept 20% of the whole dataset. Linear regression as given by the above equation is performed on the data and then predictions are made, which are plotted to show the results of the stock market prices vs time.

.1 Regression Based Model

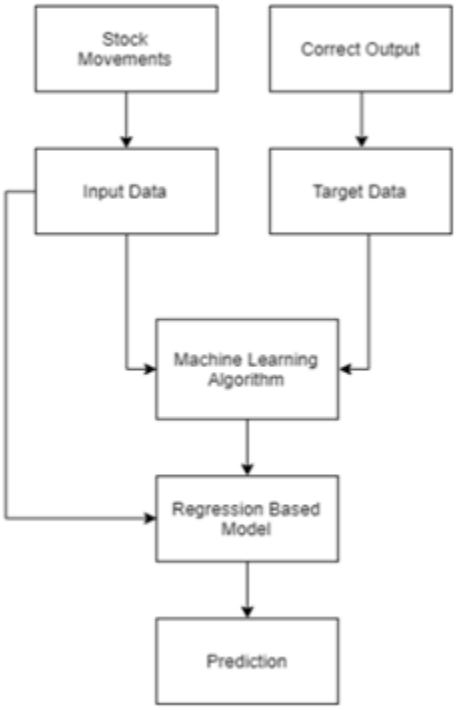

Figure 1 Flow Chart for Regression Based Model

LSTM is the advanced version of Recurrent-Neural Networks (RNN) where the information belonging to previous state persists. These are different from RNNs as they involve long term dependencies and RNNs works on finding the relationship between the recent and the current information. This indicates that the interval of information is relatively smaller than that to LSTM. The main purpose behind using this model in stock market prediction is that the predictions depend on large amounts of data and are generally dependent on the long-term history of the market. So, LSTM regulates error by giving an aid to the RNNs through retaining information for older stages making the prediction more accurate. Since stock market involves processing of huge data, the gradients with respect to the weight matrix may become very small and may degrade the learning rate of the system. This corresponds to the problem of Vanishing Gradient. LSTM prevents this from happening. The LSTM consists of a remembering cell, input gate, output gate and a forget gate. The cell remembers the value for long term propagation and the gates regulate them. In this paper, a sequential model has been made which involves stacking two LSTM layers on top of each other with the output value of 256. The input to the layer is in the form of two layer and layer. A dropout value of 0.3 has been fixed which means that 0.3 out of total nodes will be frozen during the training process to avoid over-fitting of data and increase the speed of the training process. At last, the core dense layer where each neuron is connected to every other in the next layer is added providing input of 32 parameters to the next core layer which gives output as 1. The model is compiled with a mean square cost function to maintain the error throughout the process and accuracy is chosen as a metric for the prediction.

Software used

Machine Learning with Python.

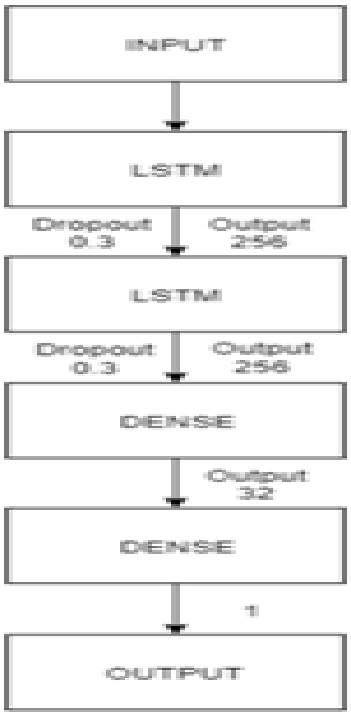

Figure 2 LSTM Layers

4. RESULTS AND DISCUSSION

The proposed system is trained and tested over the dataset taken from Yahoo Finance. It is split into training and testing sets respectively and yields the following results upon passing through the different models:

Regression Based Model Results The plot in figure3 is the result of application of linear regression algorithm on the dataset to predict varying prices with respect to thetime.

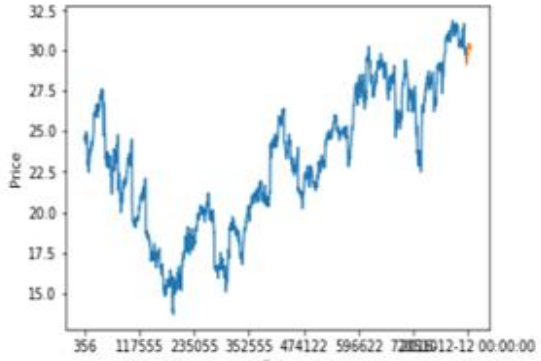

Figure 3 Plot between Price and Data Using Regression

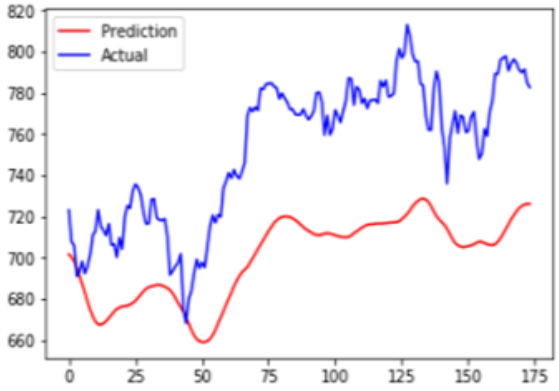

Figure 4 Plot between Actual and Predicted Trend of LSTM

The above figure 3 is plot over the data having batch size 512 and 90 epochs. The prediction is shown by red line and the actual trend is shown by blue. The proximity of these lines tells , how efficient the LSTM based model is. The prediction approximates real trend when a considerable amount of time has passed. The more the system is trained the greater the accuracy which will be attained. Attai

5.CONCLUSION

Two techniques have been utilized in this paper: LSTM and Regression, on the Yahoo finance dataset. Both the techniques have shown an improvement in the accuracy of predictions, thereby yielding positive results. Use of recently introduced machine learning techniques in the prediction of stocks have yielded promising results and thereby marked the use of them in profitable exchange schemes. It has led to the conclusion that it is possible to predict stock market with more accuracy and efficiency using machine learning techniques. In the future, the stock market prediction system can be further improved by utilizing a much bigger dataset than the one being utilized currently. This would help to increase the accuracy of our prediction models. Furthermore, other models of Machine Learning could also be studied to check for the accuracy rate resulted by them

Conflicts of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Funding Statement

This research received no external funding

ACKNOWLEDGEMENTS

It is our great privilege to express our profound and sincere gratitude to our Project Supervisor, Assistant Prof. Prajit Paul for providing us a very cooperative and precious guidance at every stage of the present project work being carried out under his supervision. His valuable advice and instructions in carrying out the present study has been a very rewarding and pleasurable experience that has greatly benefited me throughout the course of work.

We would like to convey my sincere gratitude towards

Prof. (Dr) Arunava De, Head of the Department of ECE, Asansol Engineering College for providing us the requisite support for time completion of our work.

We would also like pay my heartiest thanks and gratitude to all the teachers of the Department of ECE, Asansol Engineering College for various suggestions being provided in attaining success in our work.

REFERENCES

[1] Gareja Pradip, Chitrak Bari, J. Shiva Nandhini, “Stock market prediction using machine learning” International Journal of Advance Research and Development, Volume 3, Issue 10, 2018
[2] K. Raza, “Prediction of Stock Market performance by using machine learning techniques”, 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT),
[3] K. Hiba Sadia, Aditya Sharma, Adarrsh Paul, Sarmistha Padhi, Saurav Sanyal, "Stock Market Prediction Using Machine Learning Algorithms", International Journal of Engineering and Advanced Technology (IJEAT), Volume-8 Issue-4, 2019
[4] Raut Sushrut Deepak, Shinde Isha Uday, Dr. D. Malathi, “MACHINE LEARNING APPROACH IN STOCK MARKET PREDICTION”, International Journal of Pure and Applied Mathematics, Volume 115, No. 8, 2017.
[5] Mehtab, S., Sen, J., A robust predictive model for stock price prediction using deep learning and natural language processing. In: Proceedings of the 7th International Conference on Business Analytics and Intelligence
[6] M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, “Stock market prediction using machine learning techniques", 2016 3rd International Conference on Computer and Information Sciences (ICCOINS)
[7] Tang, J., Chen, X., Stock market prediction based on historic prices and news titles. In: Proceedings of the International Conference on Machine Learning Technologies (ICMLT)
[8] Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017
[9] Sachin Sampath Patil, Prof. Kailash Patidar, Asst. Prof. Megha Jain, “A Survey on Stock Market Prediction Using SVM”, IJCTET 2016.
[10] https://marutitech.com/ai-and-ml-in-finance/
[11] Guresen E, Kayakutlu G, Daim TU (2011) Using artificial neural network models in stock market index prediction.
[12] https://towardsdatascience.com/machine-learning-basics-random-forest-regression-be3e1e3bb91a
[13] https://www.javatpoint.com/machine-learning-polynomial-regression
[14] https://www.scribbr.com/statistics/multiple-linear-regression
[15] Yodele et al (2012) Stock Price prediction using neural network with hybridized market indicators. Journal of Emerging Trends in Computing and Information Sciences 3(1)
[16] https://colah.github.io/posts/2015-08-Understanding-LSTMs/
[17] https://www.sas.com/en_us/insights/analytics/machine-learning.html
[18] https://groww.in/us-stocks/googl
[19] https://www.investopedia.com/terms/d/deep-learning.asp
[20] https://www.nasdaq.com/market-activity/stocks/