Stock Market Prediction Using Machine Learning

Md. Ariful Islam
Department of Computer Science and Engineering
East West University
Dhaka,Bangladesh
2016-1-60-019

Department of Computer Science and Engineering
East West University
Dhaka,Bangladesh
2016-1-60-060

Sidratul Moontaha
Department of Computer Science and Engineering
East West University
Dhaka,Bangladesh
2015-1-60-033

***Abstract*— In Stock Market Prediction, the aim is to predict the future value of the financial stocks of a company. Now, machine learning algorithms can make better predictions than the stock market experience brokers. We have used Linear Regression and LSTM based machine learning algorithm to predict stock price. As a factor, we have considered closing price only.**

Keywords— Closing Price, Linear Regression, and LSTM

I. Introduction

Machine learning is an efficient way to predict a stock market condition. The vital part of machine learning is the dataset used. The dataset should be as concrete as possible because a little change in the data can perpetuate massive changes in the outcome. In this project, supervised machine learning has employed on a dataset obtained from Samsung. This dataset comprises of following seven variables: date, open, close, low, high, adj. close and volume. Open, close, low and high are different bid prices for the stock at separate times with nearly direct names. The volume is the number of shares that passed from one owner to another during the time period. The model is then tested on the test data. We have used Linear Regression and LSTM models separately. Finally, we have shown the resulted graph for the fluctuation of prices with the dates comparing actual and predicted price (for both Linear Regression and LSTM). At last we have compared the result from Linear Regression, LSTM AND actual price.

A correct prediction of stocks can lead to huge profits for the seller and the broker. Frequently, it is brought out that prediction is chaotic rather than random, which means it can be predicted by carefully analyzing the history of respective stock market. Machine learning is an efficient way to represent such processes. It predicts a market value close to the tangible value, thereby increasing the accuracy. Introduction of machine learning to the area of stock prediction has appealed to many researches because of its efficient and accurate measurements [1].

The vital part of machine learning is the dataset used. The dataset should be as concrete as possible because a little change in the data can perpetuate massive changes in the outcome [2]. In this project, supervised machine learning is employed on a dataset obtained from Yahoo Finance. This dataset comprises of following five variables: open, close, low, high and volume. Open, close, low and high are different bid prices for the stock at separate times with nearly direct names. The volume is the number of shares that passed from one owner to another during the time period. The model is then tested on the test data.

Regression and LSTM models are engaged for this conjecture separately. Regression involves minimizing error and LSTM [3][4] contributes to remembering the data and results for the long run. Finally, the graphs for the fluctuation of prices with the dates (in case of Regression based model) and between actual and predicted price are plotted.

The rest of the paper consists of following: Part 2 puts forward the two models used and the methods used in them in detail. Part 3 discusses the results produced with different plots for both the models in detail. Part 4 consists of conclusion and the last section involves the references.

II. RELEATED WORKS

From the literature survey, it was observed that the application of machine learning techniques to stock market prediction is being undertaken thoroughly throughout the world. Machine Learning techniques are proving to be much more accurate and faster as compared to contemporary prediction techniques.

Ishita parmar , Navanshu Agarwal, Sheirsh Saxena, Ridam Arora, Shikhin Gupta Himanshu Dhiman from National Institute of Technology Hamipur, INDIA did a great work on this filed using Regression and LSTM.

III. DATA SET

For stock market prediction, we have collected our data set form. In the data set, there have seven attribute and 2222 rows where number of row 2222 is empty.

Date:
Open:
High:
Low:
Close:
Adj Close:
Volume:

IV. METHODOLOGY

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.

A. REGRESSION BASED MODEL

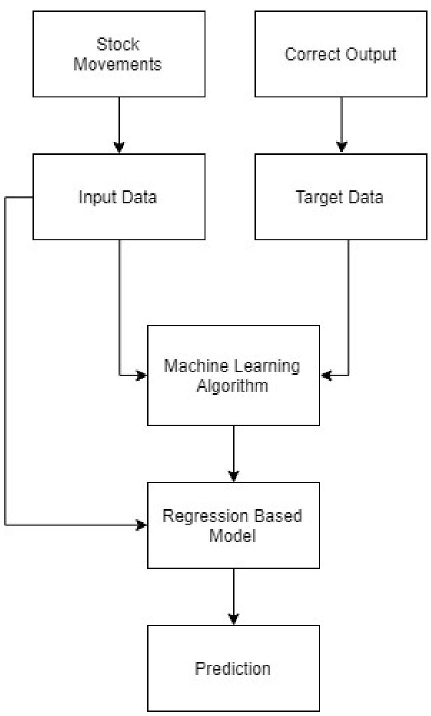

Figure 1. Flow Chart for Regression Based Model

Regression is used for predicting continuous values through some given independent values [5]. 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, *V*is continuous 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 [6].

B. LONG SHORT TERM MEMORT (LSTM)

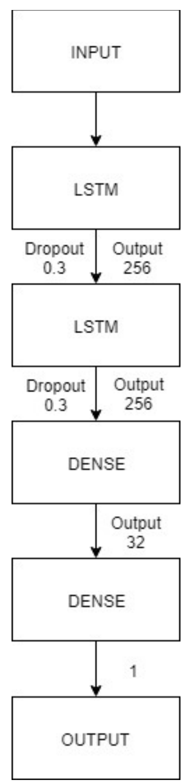

Figure 2. LSTM Layers

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 depends on large amounts of data and are generally dependent on the long term history of the market [6]. So LSTM regulates error by giving an aid to the RNNs through retaining information for older stages making the prediction more accurate [7].

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 [8].

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 [0] and layer[1]. A dropout value of 0.3 has been fixed which means that0.3 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 [9].

V. EXPERIMENTAL RESULT

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:

A. REGRESSION BASED MODEL RESULT

The plot in figure3 is the result of application of linear regression algorithm on the dataset to predict varying prices with respect to the time.

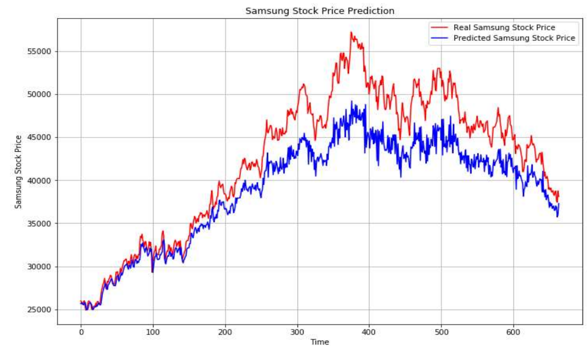

Figure 3. Result for Regression Based Model

B. LSTM BASED MODEL RESULT

The above graph 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 two lines tells, how efficient the LSTM based model is. The prediction approximates real trend when a considerable amount of time has passed. The more thesystem is trained the greater the accuracy which will be attained

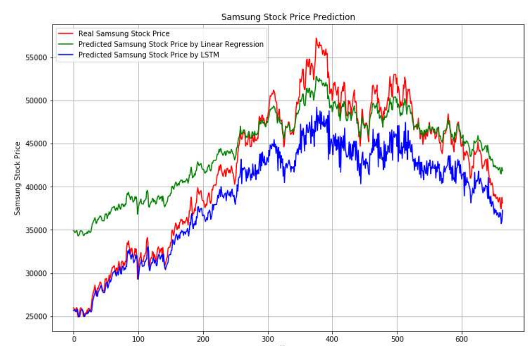

Figure 4. Result for LSTM model

VI. CONCLUSION

The primary contribution of the researchers being the application of the novel LSTM Model as a means of determining the stock prices. Both the techniques have shown an improvement in predictions, thereby yielding positive results with the LSTM model proving to be more efficient. The results are quite promising and has led to the conclusion that it is possible to better predict stock market using machine learning techniques. Sentiment analysis though Machine Learning on how news affects the stock prices of a company is also a very promising area. Other deep learning based models can also be used for prediction purposes

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.

VII. FUTURE ENHANCEMENT

Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employee. The use of traditional algorithms and data mining techniques can also help predict the corporation's performance structure as a whole.

REFERENCES:

[1] I. Parmar, N. Agarwal, S. Saxena, R. Arora, S. Gupta, H. Dhiman and L. Chouchan, "Stock Market Prediction Using Machine Learning", in First International Conference on Secure Cyber Computing and Communication(ICSCCC), 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), Karachi, 2017, pp. 1-1.

[3] H. Gunduz, Z. Cataltepe and Y. Yaslan, "Stock market direction prediction using deep neural networks," 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, 2017, pp. 1-4.

[4] M. Billah, S. Waheed and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, 2016, pp. 1-4.

[5] H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend," 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, 2012, pp. 1-5.

[6] K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," INTERACT-2010, Chennai, 2010, pp. 59-63.

[7] S. Liu, G. Liao and Y. Ding, "Stock transaction prediction modeling and analysis based on LSTM," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 2787-2790.

[8] T. Gao, Y. Chai and Y. Liu, "Applying long short term momory neural networks for predicting stock closing price," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 575-578. [9] K. A. Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and longterm stock market prediction," 2018 9th International Conference on Information and Communication Systems (ICICS), Irbid, 2018, pp. 151-156

[9] Prediction", IJPAM 2017.

[10] Pei-Yuan Zhou , Keith C.C. Chan, Member, IEEE, and Carol XiaojuanOu, "Corporate Communication Network and Stock Price Movements: Insights From Data Mining", IEEE 2018.