Stock Market Prediction Using Machine Learning

1st Ishita Parmar
Department of Computer Science and Engineering
National Institute Of Technology
Hamirpur, INDIA
cs14mi515@nith.ac.in

2nd Navanshu Agarwal
Department of Computer Science and Engineering
National Institute Of Technology
Hamirpur, INDIA

3rd Sheirsh Saxena
Department of Computer Science and Engineering
National Institute Of Technology
Hamirpur, INDIA

4th Ridam Arora
Department of Computer Science and Engineering
National Institute Of Technology
Hamirpur, INDIA

5th Shikhin Gupta
Department of Computer Science and Engineering
National Institute Of Technology
Hamirpur, INDIA

6th Himanshu Dhiman
Department of Computer Science and Engineering
National Institute Of Technology
Hamirpur, INDIA

7th Lokesh Chouhan
Department of Computer Science and Engineering
National Institute Of Technology
Hamirpur, INDIA
lokesh@nith.ac.in

***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.**

Keywords-Close, high, low,LSTM model, open, regression, and volume.

I. INTRODUCTION

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] [2].

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 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 thefluctuation of prices with the dates (in case of Regression based model) and between actual and predicted price (forthe LSTM based model) are plotted. The rest of the paper consists of following: Section II discusses the related work. Section III puts forward the two models used and the methods used in them in detail. Section IV discusses the results produced with different plots for both the models in detail. While Section V consists of conclusion and the last section involves the references.

II. RELATED WORK

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.

Significant work has been done throughout the world in this field. A testament to which is the work of M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali [1] and that of K. Raza [2] who have surveyed the application of machine learning techniques and presented the current advancements in this field.

H. Gunduz, Z. Cataltepe and Y. Yaslan [3] predicted stock prices using deep neural network techniques. Similarly, M. Billah, S. Waheed and A. Hanifa [4] suggested further improvements to stock prediction using neural networks through the use of a training algorithm which they designed on their own. K. V. Sujatha and S. M. Sundaram [6] suggested insightful techniques on handling non-normal situations which may often arise during the working of the system and cause disruptions or lead to inaccurate predictions. Liu, G. Liao and Y. Ding [7]

conducted similar work and designed a model for applying LSTM to stock prediction with lots of scope for improvements to prediction accuracy. K. A. Althelaya, E. M. El-Alfy and S. Mohammed [9] further contributed to the field by staging experiments and simulations to assess the feasibility of applying deep learning techniques to prediction of stock prices.

III. 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.

The dataset being utilized for analysis was picked up from Yahoo Finance. The dataset consisted of approximately 9 lakh records of the required stock prices and other relevant values. The data reflected the stock prices at certain time intervals for each day of the year. It consisted of various sections namely date, symbol, open, close, low, high and volume. For the purpose of simulation and analysis, the data for only one company was considered. All the data was available in a file of *csv* format which was first read and transformed into a data-frame using the *Pandas*library in Python. From this, the data for one particular company was extracted by segregating data on the basis of the symbol field. Following this normalization of the data was performed through usage of the *sklearn* library in Python and the data was divided into training and testing sets. The test set was kept as 20% of the available dataset.

Although machine learning as such has many models but this paper focuses on two of the most important amongst them and made the predictions using these.

A. Regression Based Model

In general, the Regression based Model is used for predicting continuous values through some given independent values [5]. Regression uses a given linear function for predicting continuous values:

$$V = a + bK + error \quad (1)$$

Where, $V$is acontinuous value; K represents known independent values; and, a, b are coefficients.

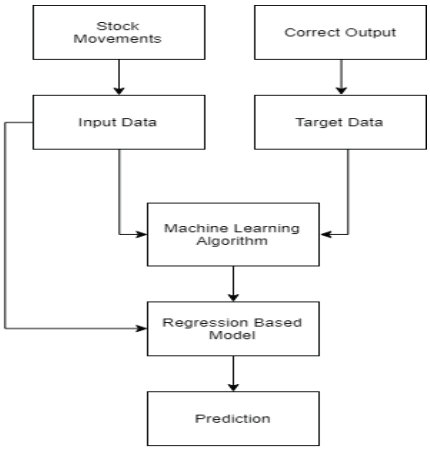

Fig. 1 Flow Chart for Regression Based Model

The paperutilizes the gradient descent linear regression algorithm for predicting correct values by minimizing the error function as given in Fig.1.

Linear Regression [6]as governed by the above equation is performed on the data and then the relevant predictions are made. The factors considered for the regression were low, open, high, close and volume. The R-square confidence test was used to determine the confidence score and the predictions were plotted to show the results of the stock market prices vs time.

B. Long Short Term Memory (LSTM) Network Based Model

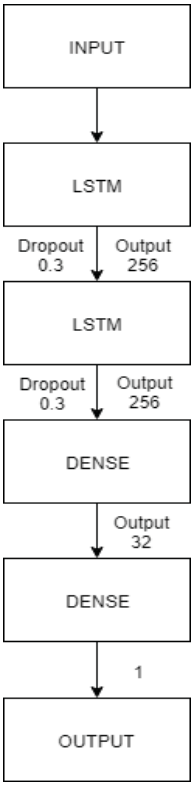

Fig. 2 LSTM Layers

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]. Thus proving itself as much more reliable compared to other methods.

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.[8].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 [0] and layer[1]. 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 [9].

IV. EXPERIMENTAL RESULTS

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 Results

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.

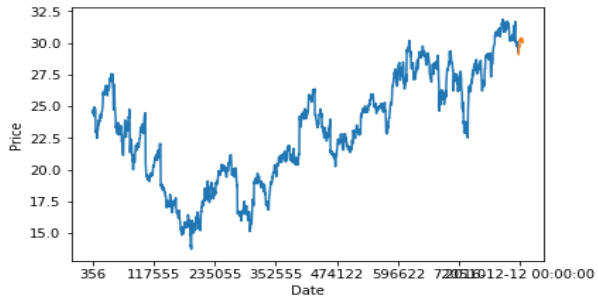

Fig.3 Plot between Price and Date Using Regression

The above graph Fig. 3 is plot over the data having batch size 512 and 90 epochs. The R-square confidence test resulted in a confidence score of 0.86625.

B. LSTM Based Model Results

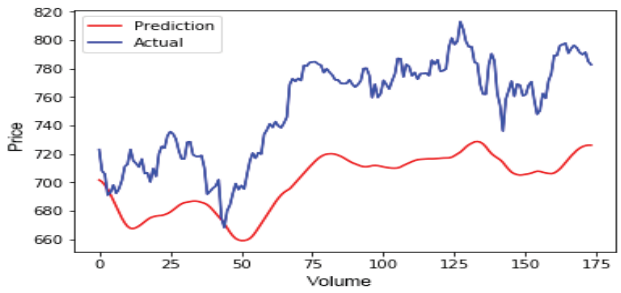

Fig. 4 Plot between Actual& Predicted Trend of LSTM

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 model resulted in a Train Score of 0.00106 MSE (0.03 RMSE) and a Test Score of 0.00875 MSE (0.09 RMSE). The more thesystem is trained and the greater the size of the dataset utilized the greater the accuracy which will be attained. The LSTM Model offered more accuracy than the Regression based Model.

V. CONCLUSION

This paper was an attempt to determine the future prices of the stocks of a company with greater accuracy and reliability using machine learning techniques. 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 the accuracy of 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 predict stock market with more accuracy and efficiency using machine learning techniques.

In the future, the accuracy of the stock market prediction system can be further improved by utilizing a much bigger dataset than the one being utilized currently. Furthermore, other emerging models of Machine Learning could also be studied to check for the accuracy rate resulted by them. 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.

REFERENCES

[1] 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), Kuala Lumpur, 2016, pp. 322-327.

[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 modelling 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 memory 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 long-term stock market prediction," 2018 9th International Conference on Information and Communication Systems (ICICS), Irbid, 2018, pp. 151-156