International Journal of Linguistics and Computational Applications (IJLCA)
Volume 8, Issue 4, October – December 2021
ISSN 2394-6385 (Print)
ISSN 2394-6393 (Online)

Stock Prediction using Machine Learning

Dr. J. Dhilipan*1, D. B. Shanmugam#2, Imran Quraishi#3
[1]Professor& Head, Department of Computer Applications, SRM IST, Ramapuram Campus, Chennai
[2]Assistant Professor, Department of Computer Applications, SRM IST, Ramapuram Campus, Chennai
[3]MCA Final Year, Department of Computer Applications, SRM IST, Ramapuram Campus, Chennai
Corresponding Author mail id: hod.mca.rmp@srmist.edu.in

Abstract— Stock trading is one of the foremost activity in finance world. Stock market prediction is used to find the long run values of the stock and other financial factors influenced on a financial exchange. The technical and fundamental or the statistical analysis is employed by most of the stockbrokers while making the stock predictions. Python programming language in machine learning is used for the stock market prediction. In this paper we have proposed a Machine Learning (ML) approach which trains from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In stock market prediction, the aim is to predict the longer term value of the financial stocks of a corporation [1]. The recent trend in market prediction technologies is that the use of machine learning approach which makes predictions supported the values of current stock market indices by training on their previous values. Machine learning itself employs different models to form prediction easier and authentic. This paper focus on Regression and Long Short Term Memory (LSTM) based Machine learning to predict stock values. The factors that are being considered include re-open, close, low, high and volume [2,3].

Keywords* — *Long Short Term Memory; LSTM; Tensorflow; Neural Network Module.

1. Introduction

The introduction of an intelligent computer application into stock prices prediction will effectively reduce skepticism and conservative options among many investments' professionals and market participants. It will guide the investors in overcoming the uncertainty and imprecision in predicting the stock prices at the most times [4]. The characteristic in each stock markets have common uncertainty, imprecision which is said with their short and long-term future state. Prediction market has been a hot research area for several years. If any system which may consistently predict the trends of the dynamic stock exchange be developed, it might make the owner and therefore the users of the system wealthy. These data may be suffering from inflation or fluctuation of exchange rates especially in countries like Nigeria [5,6]. Back propagation neural network is usually used for price prediction [7]. This paper demonstrates the event of a neural network application for analyzing and predicting of stock exchange prices considering the technical and fundamental factors as our input against one factor employed by many papers reviewed. It is expected that the accuracy of the appliance are going to be high compared to similar applications

2. Existing Methods

The existing system proposes that a company's performance, in terms of its stock worth movement, is foreseen by internal communication patterns to urge early warning signals. We tend to believe that it is vital for patterns in company communication networks to be detected earlier for the prediction of great stock worth movement to avoid attainable adversities that a corporation could face within the stock exchange so as that stakeholders' interests is protected the utmost amount as attainable. Support Vector Machine may be a machine learning technique utilized in recent studies to forecast stock prices [8,9]. This model tries to predict whether the stock price is going to increase or decrease within a period in future.

2.1 Drawbacks of Existing Methods

- Accuracy would decrease when setting more levels of stock exchange movement.
- The typical of prediction accuracies using Decision Tree because the classifier are 43.44%, 31.92%, and 12.06% for "two levels," "three levels," and "five levels," respectively.
- These results indicate that the stock price is unpredictable while the traditional classifier is implemented.

3. Methodology

In this paper, we've a dataset containing stock prices of Google from January 2012 to December 2016 [10,11,12]. We are advised to use these stock prices for training the neural network and predict the stock prices for the month of January-2017. This is a Regression problem [13,14,15]. To achieve this goal, we trained a Recurrent Neural Network LSTM. We will use one of the deep learning libraries, Keras, to build the neural network.

DOI: 10.30726/ijlca/v8.i4.2021.84008

3.1 Feature Scaling

The performance of the neural network are going to be better if the whole training input is within the same range [16,17]. The stock prices aren't within the same range. So, we need to scale the training data such that they are in the same range. This process is called feature scaling. The two general methods for feature scaling are:

- *Standardization*

$$x'=x-\bar{x}\sigma$$

where, x is that the original feature vector, $\bar{x}$ is that the mean of that feature vector, and σ is its variance [18].

- *Normalization (Min-Max normalization)*

$$x'=x-\min(x)/\max(x)-\min(x)$$

where, x is an ingenious value, x′ is that the normalized value.

It is always advised to use Normalization in the case of Recurrent Neural Network (RNN) [19]. Therefore, we use Min-Max normalization here.

The fit method only calculates the min and max values [20,21]. It doesn't apply the formula on the training set. The fit transform method applies the min-max formula on the training set [22]. Once applying the min-max formula, the transformed values will be in the range 0 and 1 i.e., the training data (features) will be in the range 0 and 1.

3.2 Reshaping

The input shape of the LSTM is 3Dimensional tensor with shape (batch_size, timesteps, input_dim). Batch_size represents the number of iterations required to traverse through the entire training data. Timesteps represent the number of inputs required for each prediction. In our scenario it is 60.

If no activation function is specified in the dense layer then the linear activation is performed by default wherein, the dense layer takes a weighted sum of its inputs which corresponds to the predicted output.

3.3 Data Collection

A data set (or dataset) may be a collection of knowledge. In tabular data, a knowledge set corresponds to at least one or more database, where every column represents a specific variable and every row corresponds to a record of the info set in question. The dataset lists values for all variables, like height and weight of an object, for every member of the info set. Each value is known as a datum. Data sets also can contains a set of documents or files.

Table 1. Knowledge dataset

Date	Open	High	Low	Close
01-03-2017	237.5	251.7	237.5	251.7
02-03-2017	258.4	271.85	251.3	271.85
03-03-2017	256.65	287.9	256.65	282.5
04-03-2017	289	300.7	289	294.35
05-03-2017	295	317.9	293	314.55
06-03-2017	317.4	318.7	305.3	308.5
07-03-2017	307.95	310.95	283.85	288.5
08-03-2017	289	305	282.15	301.7
09-03-2017	306	316.4	304.1	311.85
10-03-2017	309.5	321.65	309.5	316.3
11-03-2017	318.25	322.9	307.05	308.75
12-03-2017	309	318.95	305.55	314.2
13-03-2017	322.85	324.9	316.2	319.6
14-03-2017	320	324.9	310.55	312.85
15-03-2017	312.7	328.4	312	326.35
16-03-2017	325.5	331.85	322	326.15
17-03-2017	325.05	325.3	318.25	319.95
18-03-2017	329	336.8	322.7	334.95
19-03-2017	335.9	341.9	331.6	338.35
20-03-2017	334.45	336	326.55	328.15
21-03-2017	330.8	337.45	328	334.05
22-03-2017	342.9	346.9	338.55	340.15

3.4 Proposed System

Accuracy plays a pivotal role in stock market prediction. Although many algorithms are available for this purpose, selecting the foremost accurate one continues to be the elemental task in getting the simplest results.

In order to achieve this, in this project we have compared and analyzed the performance of various available algorithms such as Recurrent Neural Network (LSTM). This involves training the algorithms, executing them, getting the results, comparing various performance parameters of those algorithms and eventually obtaining the foremost accurate one.

3.5 Benefits of the Proposed System

- The successful prediction will maximize the benefit of the customer. In this project we have discussed various algorithms to predict the same.
- In this project we used stock data of Alphabet Inc Class A (GOOGLE) dataset consisting of data ranging from 2016 to 2019 to train different machine learning algorithms.
- Hence, we compared the accuracy of different machine learning algorithm.
- By using LSTM algorithm, it shows the highest prediction accuracy.
- We used algorithms on social media to find the impact of this data on stock market to predict accuracy for ten subsequent days. In order to improve both the

DOI: **10.30726/ijlca/v8.i4.2021.84008**

ISSN 2394-6385 (Print)
ISSN 2394-6393 (Online)

performance and quality of predictions, feature selection and spam tweets reduction are performed on the info sets.

- Moreover, we perform experiments to seek out such stock markets that are difficult to predict and people that are more influenced by social media and financial news. We compared results of various algorithms to seek out a uniform classifier. Finally, for achieving maximum prediction accuracy, deep learning is employed.

4. Results

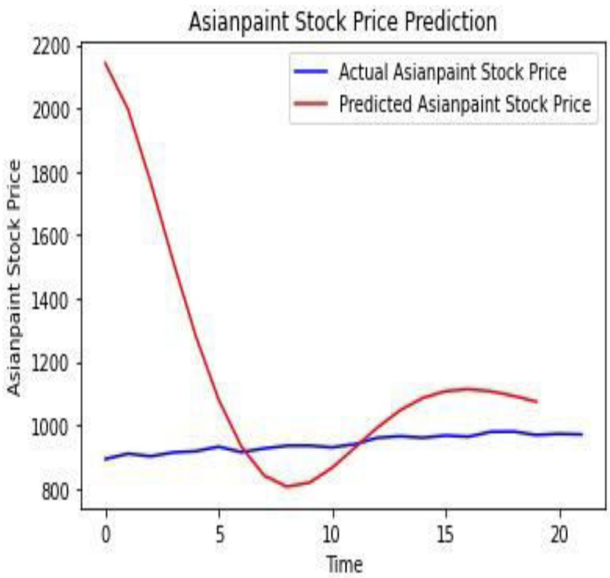

Fig.1: Stock Price Prediction (Asian Paint)

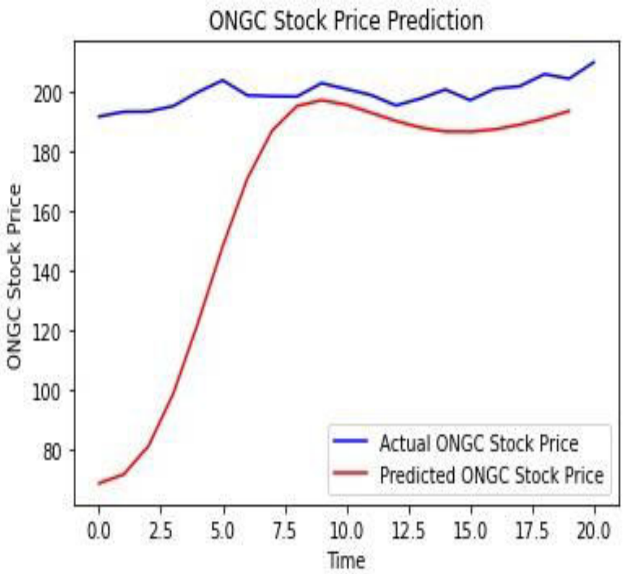

Fig.2: Stock Price Prediction (ONGC)

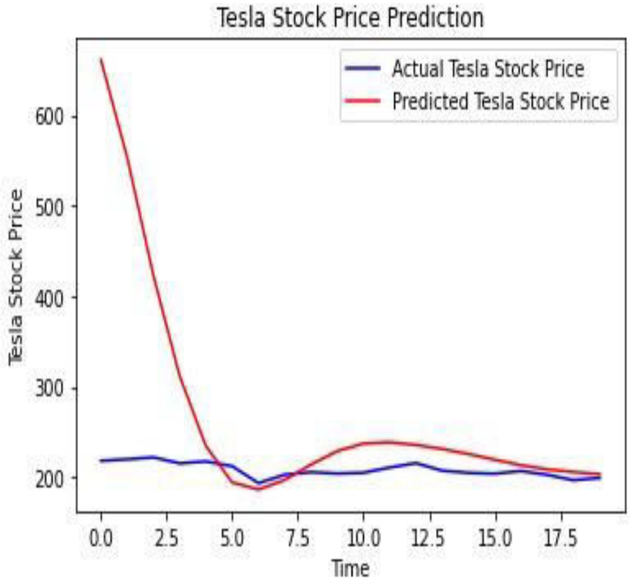

Fig. 3: Stock Price Prediction (Tesla)

We have seen that our model closely predicts the trend of the actual stock prices. The model can be further improved and experimented by considering the following (but not limited to):

- Training the model with more data. Eg: Here we have used 5 years of stock prices but you can train the model with 10 years of data.
- Increasing the number of time steps.
- Adding more LSTM layers.
- Increasing the units in the LSTM layer.
- Adding some other indicators. Eg: If you have the financial instinct that the stock price of some other companies might be correlated to the one of Google, you could add this other stock price as a replacement indicator within the training data.

5. Conclusion

Over the years, various machine learning techniques are utilized in stock exchange prediction, but with the increased amount of knowledge and expectation of more accurate prediction, the deep learning models which are being used nowadays have proven their advantage over traditional machine learning methods in terms of accuracy and speed of prediction. In this article, we'll discuss the Long-Short-Term Memory (LSTM) Recurrent Neural Network, one among the favored deep learning models, utilized in stock exchange prediction. In this task, we'll fetch the historical data of stock automatically using python libraries and fit the LSTM model on this data to predict the longer term prices of the stock. The prediction are often more accurate if the model will train with a greater number of knowledge set.

DOI: **10.30726/ijlca/v8.i4.2021.84008**

Moreover, within the case of prediction of varied shares, there could also be some scope of specific business analysis. We can study various pattern of the share price of various sectors and may analyse a graph with more different time span to fine tune the accuracy. This framework broadly helps in marketing research and prediction of growth of various companies in several time spans. Incorporating other parameters (e.g. investor sentiment, election outcome, geopolitical stability) that aren't directly correlated with the price may improve the prediction accuracy.

References

[1] M. Hagenau, M. Liebmann, and D. Neumann, “Automated news reading: Stock price prediction based on financial news using contextcapturing features,” Decision Support Systems, vol. 55, no. 3, pp. 685– 697, 2013.

[2] X. Zhao, J. Yang, L. Zhao, and Q. Li, “The impact of news on stock market: Quantifying the content of internet-based financial news,” in Proceedings of the 11th International DSI and 16th APDSI Joint meeting, 2011, pp. 12–16.

[3] S. S. Groth and J. Muntermann, “Supporting Investment Management Processes with Machine Learning Technique,” in Business Services: Konzepte, Technologien, Anwendungen - 9, Internationale Tagung Wirtschaftsinformatik, 2009.

[4] G. Fung, J. Yu, and H. Lu, “The predicting power of textual information on financial markets,” IEEE Intelligent Informatics Bulletin, vol. 5, no. 1, 2005.

[5] G. Fung, J. Yu, and W. Lam, “News sensitive stock trend prediction,” in Proceedings of the 6th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, 2002, vol. 2336, pp. 481–493.

[6] R. P. Schumaker and H. Chen, “Textual analysis of stock market prediction using breaking financial news,” ACM Transactions on Information Systems, vol. 27, no. 2, pp. 1–19, 2009.

[7] M. Mittermayer and G. Knolmayer, “Text mining systems for market response to news: A survey,” vol. 41, no. 184. University of Bern, 2006.

[8] S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai, “Combining Technical Analysis with Sentiment Analysis for Stock Price Prediction,” in Proceedings of the IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing, 2011, pp. 800–807.

[9] S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai, “Multiple Kernel Learning on Time Series Data and Social Networks for Stock Price Prediction,” in Proceedings of the 10th International Conference on Machine Learning and Applications and Workshops, 2011, pp. 228–234.

[10] X. Li, C. Wang, J. Dong, and F. Wang, “Improving stock market prediction by integrating both market news and stock prices,” Database and Expert Systems Applications, Lecture Notes in Computer Science, vol. 6861, pp. 279–293, 2011.

[11] Chi, L.C., Tang, T.C., and Chin, M. (2012), Corporate transparency as a defense against a stock price plunge: Evidence from a market-crash context. Chiao Da Management Review, 32, 137-162.

[12] Fromlet, H. (2001), Behavioral finance-Theory and practical application, Business Economics, 36(3), 63-69.

[13] Kutan, A.M. and Yuan, S. (2002), Does public information arrival matter in emerging markets?: Evidence from stock exchanges in China, Working paper, Department of Economics and Finance, Southern Illinois University.

[14] Lee, C.W. (1986), Information content of financial column, Journal of Economics and Business, 38, 27-39.

[15] Mitchell, M.L. and Mulheri, J.H. (1994), The impact of public information on the stock market, Journal of Finance, 49(3), 923-950.

[16] Nofsinger, J.R. (2001), The impact of public information in investors, Journal of Banking & Finance, 25(7), 1339-1366.

[17] Tetlock, P.C. (2007), Giving content to investor sentiment: the role of media in the stock market, Journal of Finance, 62(3), 1139-1168.

[18] Tang, T.C. and Chi, L.C. (2014), See what you are searching for?. The proceeding of IAOS 2014 Conference on Official Statistics Meeting the demands of a changing world.

[19] Tetlock, P.C., Maytal, S., and Macskassy, S, (2008), More than words: Quantifying language to measure firms’ fundamentals, Journal of Finance, 63, 1437-1467.

[20] Schumaker, R.P. and Chen, H. (2009), Textual analysis of stock market prediction using breaking financial news: The AZFin text system, ACM Transactions on Information Systems, 27(2), 1-19.

[21] Womack, K.L. (1996), Do brokerage analysts’ recommendations have investment value? Journal of Finance, 51, 137-167.

[22] Womack, K.L. (2012), Do Brokerage Analysts' Recommendations Have Investment Value? The Journal of Finance, 51(1), 137-167.

DOI: **10.30726/ijlca/v8.i4.2021.84008**