# Stock Prediction using Machine-Learning Algorithms

**Senthil Jayave, Arpit Rathore, Jayakumar Sadhasivam**

*Abstract - The stock market is now days becomes very dynamic and liable to the external as well as internal factors, which can step-up or step-down the market. Nowadays it becomes important to understand the correlation between all the factors which can affect the market and so that we can achieve our primary objective. So, market trends prediction with achieving the high precision is now very necessitating by applying the machine learning algorithms to the historical data and analysing this with others factors like government policies, trending headlines, prices of the important commodities etc., which also play a very crucial role in directing the flow of the stock market and needed to keep beside while evaluating prices of the stock. Machine learning algorithm will help us to develop a model, which is going to analyze the stock prices patterns providing us a model, which is going to help us in the predicting of the stock prices. In this paper, I am comparing the two-machine learning algorithm i.e. Random forest and linear regression to create the training data model and going to test this model on the testing data set to predict the accuracy of the following algorithm's models.*

*Keywords – Dataset, Random Forest, MLP, Decision Tree, Training Dataset, Testing Dataset, NSE, BSE, SVR, SVM, BPNN.*

## I. INTRODUCTION

The stock market plays a colossal role in the economic aspect of any country development [1] as well as in the contribution in the Gross Develop Product (GDP) and provides a boost to economic development as well as the stability among the people who are connected directly or indirectly to the markets.

National Stock Exchange (NSE) is the one of the top stock exchange of India and world 10th largest stock exchange in the terms of market capital around $1.41 trillion in 2017 [2].

Housing Development Financial Corporation (HDFC) Bank Limited is an Indian banking and finance company. It is an India's one of the largest private sector bank [3] which is listed in the NSE, BSE [4] as well as in other stock exchanges also. Machine Learning is the subset of the data science which helps us in the prediction and classification of the of many real-life things like weather forecasting, cancer prediction etc.

Machine learning is paving a pathway providing close results while predicting stock price using various factors, which influence the market like the current situation like economic to various business policies, political affairs in the world, prices of the different equitable entities like crude oil, forex exchange etc., and trending news in the news feeds also.

There are various machine-learning algorithms, which can provide us precise results while developing the predictive model like Logistic Regression, Random Forest and Multilayer Perceptron (MLP) etc.

The statistician David Cox proposed Logistic Regression in 1959 [5]. It is probabilistic supervised learning algorithm. In this algorithm, a measurement of the relationship will take place between a categorical dependent variable and one or many independent variables using a function, which is called as the logistic function or sigmoid function.

It uses the same concept as Linear Regression by developing an equation for the development of the predictive or classification model.

Random Forest is also a type of supervised learning algorithm, which is an advanced type of decision trees in which many decision trees are generated and will merge together and produced the one decision tree, which can provide most promising and stable results.

Random Forest trees use bagging and binning techniques for the development of the trees and for the root node selection, it uses not only Information Gain and Gini index as well as randomness for this objective. The good part about random forest algorithm reduces the overfitting issues related to the predictive or classifying models and can handle missing values eminently.

Artificial Neural Networks (ANN) [6] in which there is a basic unit which is called as the artificial neurons on which whole neural network works is made up of the input number and a weight which is a floating point number which can be adjusted while training of the network for the best results.
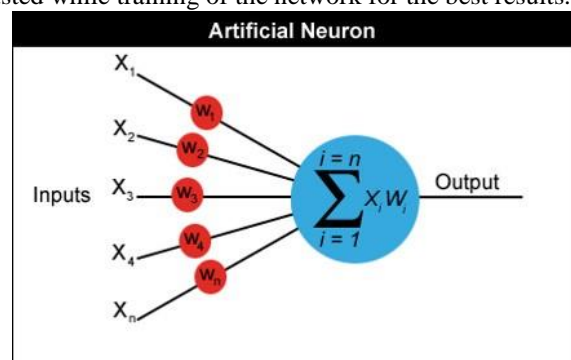


**Fig. 1.1**

The neural network can have any numbers of the artificial neurons which can collectively is a submission of the product of the set of the inputs and weights of each neuron.

*Retrieval Number B10831282S18/18©BEIESP*
*Journal Website: www.ijeat.org*

402

*Published By:*
*Blue Eyes Intelligence Engineering*
*and Sciences Publication (BEIESP)*
*© Copyright: All rights reserved.*

The Multilayer Perceptron (MLP) is a combination of the mainly three important layers i.e.
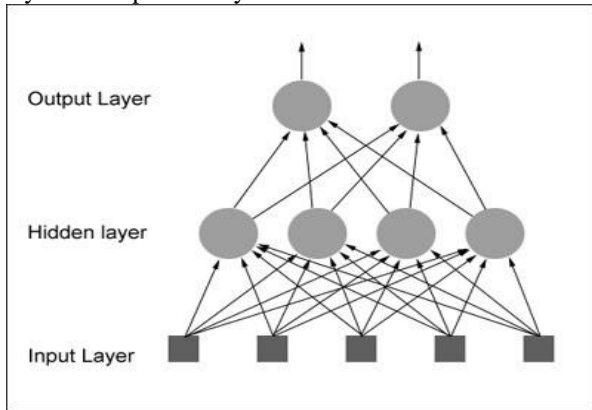


**Fig. 1.2**

Input layer, which used to provide the inputs to the neural networks and output layer, which produces the output using the N- numbers of hidden layers, which changes the inputs according to the user's requirements, or training of the perceptions or artificial neurons.

## II. RELATED WORK

In the previous work, Ryo Akita and his team [7] used the deep learning technique in which they are converting newspaper articles using paragraph vector and using deep learning model like Long Short-Term Memory (LSTM) on the opening prices of the companies of the Tokyo Stock Exchange.

Whereas, Mingze Xu and his team of authors [8] used an unsupervised technique like Non-negative Matrix Factorization (NMF) and using this with regression technique like logistic regression provided them powerful prediction model. There are some other works on this objective is also proposed like Feasal Mithani [9] proposed a Back Propagation Neural network (BPN) techniques of the neural networks which he used on the 78 attributes containing dataset which contained almost 1002000 records.

Rohit and his team [10] have used a hybrid system containing the genetic algorithm alongside with the Support Vector Machine (SVM) and they implemented on the National Stock Exchange (NSE) stocks to produce a predictive model.

Mehak and her team [11] used multiple machine learning algorithms on the Karachi Stock Exchange (KSE) along with the prices of the different commodities like oil prices, Foreign Exchange prices etc. to provide an insight how stock prices will behave under so many different factors which usually controls the market trading.

Ratnayaka and his team [12] of authors used time series analysis and forecasting techniques to prepare a new model for the Colombo Stock Exchange (CSE) and utilized the techniques like ARIMA and neural networks for more precise results.

Some of them used more complex and multiple algorithms system for the preparation of the predictive model. Jonathan and Farhana [13] have used the Auto Regressive Integrated Moving Average (ARIMA) and Back Propagation Neural Networks (BPNN) but still did not have the up to the mark results.

Jintak Park and his fellow [14] used a combination of the historical data and sentimental analysis of the tweets, which were posted on that day for particular keywords and as the results were quite improved.

Some other hybrid models also came into the light like Bashar Al-Hnaity and Mayasam Abbod [15] came up with the hybrid model featuring the SVR, SVM and BPNN along with the Genetic Algorithm for more accuracy in the development of the model.

Whereas, Xia Liang and her team of authors [16] also used a hybrid model, which consists the data mining techniques for smoothing of the time series, data and they have used the wavelet transformation for removing the noise in the time series data and Support Vector Machine (SVM) for the preparation of the prediction model development.

Sornpon and Surin [17] improvised the traditional Auto Regression Integrated Moving Average (ARIMA) into a Cross Correlation Auto Regression Integrated Moving Objects (CARIMA) and applied it on the Stock Exchange of the Thailand (SET).

Zhao and his team of authors [18] were proposed new data mining technique for the trend prediction using novel outlier mining algorithm and they have implemented this using Chinese Stock Market and really appreciated their results.

## III. PROPOSED WORK

In this research paper, I am going to use three machine learning algorithm i.e. Logistic Regression, Random Forest and Multilayer Perceptron (MLP) for building a predictive model for stock price prediction and for gaining more precise results including sentimental analysis on the news headlines on the prices of the stock value.

Along with the one year of historical data of the HDFC Bank listed in the National Stock Exchange (NSE) is obtained from Yahoo Finance Website [19] along with performing the sentimental analysis of the news headlines, which is collected from the New York Times. For the lexicons, I am going to use NLTK's Vader Lexicon and sentimental analyser for getting the positive, negative, neutral and compound values. Python platform is used for implementing the proposed work.

*Dataset details:-*

| | |
|---|---|
| 1.Date | Date stamp of trading in the market. |
| 2.Open | Price of the stock while trading starts. |
| 3.High | Highest price of the stock gained while trading. |
| 4.Low | Lowest price of the stock dropped while trading. |
| 5.Close | Price of the stock when market halt trading for a day. |
| 6.Adjacent Close | Last price at which a stock trades during a regular trading session. |
| 7.Volume | Volume is the quantity of shares or contracts traded in a security or an entire market during a given period. |

403

## IV.    RESULTS

The machine learning algorithm's implementation provides the amazing insights into the stock behavior according to the sentimental influences on the stocks on the daily basis which producing the predictive model.

The Logistic Regression algorithm provides a quite scattered result. According to the Fig. 4.1, the plot of the average literal price and average of forecasted values showing the differential behaviour as the same happened between the literal price and the forecasted values.
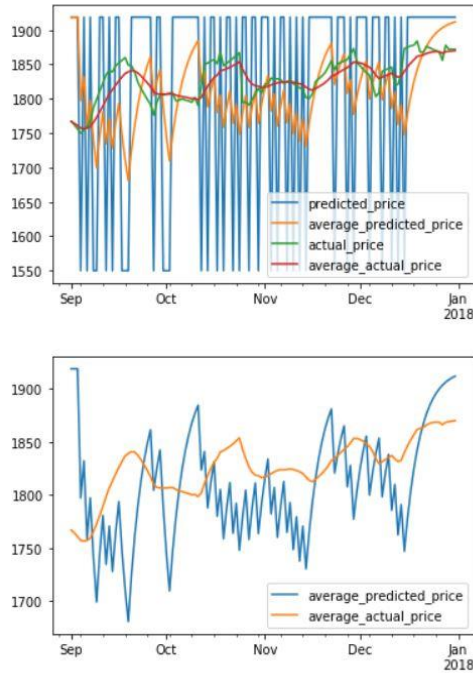


**Fig. 4.1**

The Random Forest Regressor algorithm provides us a comparably a promising results as compare of the Logistic Regression algorithm. As the Fig. 4.2, is illustrating the plot of the literal price against the forecasted price, as well as the average of the literal price against the forecasted price, is now showing less movement as compared to the previous model based on the Logistic Regression.
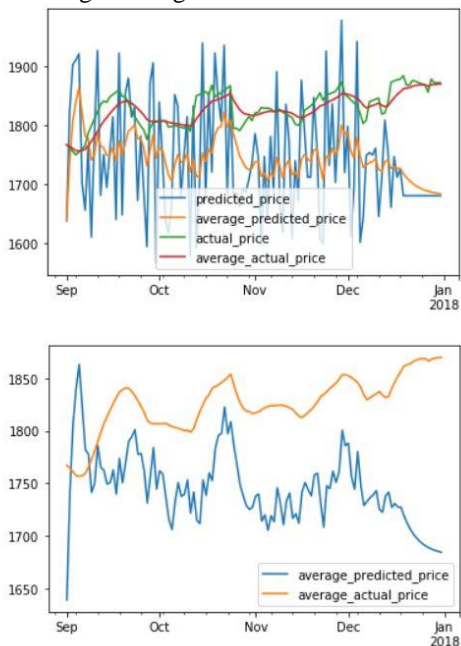


**Fig. 4.2**

Whereas, the predictive model in Fig 4.3, which is developed by using Multilayer Perceptron (MLP), showed most promising results as compared to the previous two models developed using Logistic Regression and Random Forest. The following results showing the promising insights about the models which is developed by using machine learning algorithm with the side of the sentimental analysis of the news feed to check the influence on the behaviour of the stock prices.
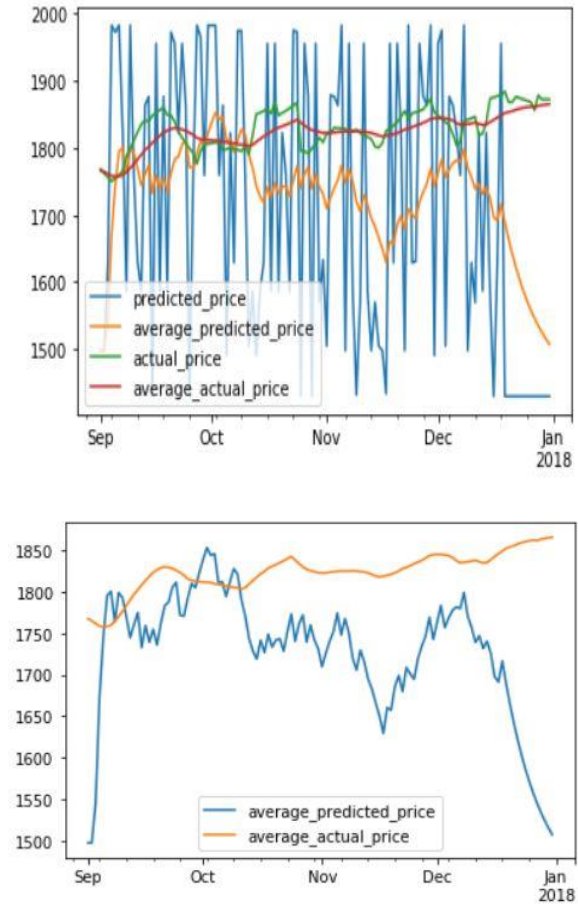


**Fig. 4.3**

## V.    CONCLUSION

Multilayer Perceptron (MLP) showed the most precise results as compared to the Logistic Regression and Random Forest by showing less deviation between the plots of the literal price against the forecasted price as well as the averages of literal price against the forecasted price. For more precise results can be achieved by using more factors in the consideration while the development of the prediction models and which can change the behaviour of the stock prices and the stock market accordingly.

### REFERENCES

1. Elliott, Larry. "The Stock Market Turmoil Was All about Good Economic News." The-Stock-Market-Turmoil-Was-All-about-Good-Economic-News, The Gaurdian, 11 Feb. 2018, www.theguardian.com/business/2018/feb/11/the-stock-market-turmoil-was-all-about-good-economic-news.

2. "National Stock Exchange of India." Wikipedia, Wikimedia Foundation, 14 Mar. 2018,en.wikipedia.org/wiki/National_Stock_Exchange_of_India.

3. "HDFC Bank." Wikipedia, Wikimedia Foundation, 16 Mar. 2018, en.wikipedia.org/wiki/HDFC_Bank.

4. "Stock Share Price Housing Development Finance Corpltd | Get Quote Hdfc | BSE." BSE Ltd, www.bseindia.com/stock-share-price/housing-development-finance-corpltd/hdfc/500010/.

5. "Logistic Regression." Wikipedia, Wikimedia Foundation, 17 Mar. 2018, en.wikipedia.org/wiki/Logistic_regression.

6. Haykin, Simon. Neural Networks. 2nd ed., vol. 1 1, Prentice Hall, 1998.

7. Akita, R., Yoshihara, A., Matsubara, T., & Uehara, K. (2016). Deep learning for stock prediction using numerical and textual information. 2016 IEEE/ACIS 15th International Conference on Computer and Information Science, ICIS 2016 - Proceedings. https://doi.org/10.1109/ICIS.2016.7550882

8. Xu, M., Lan, Y., & Jiang, D. (2016). Unsupervised Learning Part-Based Representation for Stocks Market Prediction. Proceedings - 2015 8th International Symposium on Computational Intelligence and Design, ISCID 2015, 2, 63–66. https://doi.org/10.1109/ISCID.2015.300

9. Mithani, F., Machchhar, S., & Jasdanwala, F. (2017). A modified BPN approach for stock market prediction. 2016 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2016, 0–3. https://doi.org/10.1109/ICCIC.2016.7919718

10. R. Choudhry and K. Garg, "A Hybrid Machine Learning System for Stock Market Forecasting," World Acad. Sci. Eng. Technol., vol. 2, no. 15, pp. 315–318, 2008.

11. Usmani, M., Adil, S. H., Raza, K., & Ali, S. S. A. (2016). Stock market prediction using machine learning techniques. 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), 322–327. https://doi.org/10.1109/ICCOINS.2016.7783235

12. Rathnayaka, R. M. K. T., Seneviratna, D. M. K. N., Jianguo, W., Arumawadu, H. I., Rathnayaka, K. T., Seneviratne, D. M. K. ., … Jianguo, W. (2015). A hybrid statistical approach for stock market forecasting based on Artificial Neural Network and ARIMA time series models. In 2015 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC) (pp. 54–60).https://doi.org/10.1109/BESC.2015.7365958

13. Creighton, J., & Zulkernine, F. H. (2017). Towards building a hybrid model for predicting stock indexes. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 4128–4133). https://doi.org/10.1109/BigData.2017.8258433

14. Park, J., Leung, H., & Ma, K. (2017). Information fusion of stock prices and sentiment in social media using Granger causality. 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2017), (Mfi), 614–619.

15. Park, J., Leung, H., & Ma, K. (2017). Information fusion of stock prices and sentiment in social media using Granger causality. 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2017), (Mfi), 614–619.

16. Liang, X., Liang, X., Xu, W., & Wang, X. (2015). A hybrid model for stock price based on wavelet transform and support vector machines. In 2015 12th International Conference on Service Systems and Service Management (ICSSSM) (pp. 1–7). https://doi.org/10.1109/ICSSSM.2015.7170268

17. Wichaidit, S., & Kittitornkun, S. (2016). Predicting SET50 stock prices using CARIMA (Cross Correlation ARIMA). ICSEC 2015 - 19th International Computer Science and Engineering Conference: Hybrid Cloud Computing: A New Approach for Big Data Era.https://doi.org/10.1109/ICSEC.2015.7401453

18. Zhao, L., & Wang, L. (2015). Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm. Proceedings - 2015 IEEE 5th International Conference on Big Data and Cloud Computing, BDCloud 2015, 93–98. https://doi.org/10.1109/BDCloud.2015.19

19. [19] "Yahoo Finance - Business Finance, Stock Market, Quotes, News." Yahoo! Finance, Yahoo!, finance.yahoo.com/.

20. J. Sadhasivam and R. B. Kalivaradhan, "STUDY AND COMPARISON OF VARIOUS ALGORITHMS USED IN TWITTER FOR SENTIMENT ANALYSIS," Int. J. Pure Appl. Math., 2018.

21. J. Sadhasivam and R. B. Kalivaradhan, "Sentiment Analysis of Amazon Products Using Ensemble Machine Learning Algorithm," Int. J. Pure Appl. Math., 2018.