



# JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

## REVIEW OF MACHINE LEARNING ALGORITHM FOR STOCK MARKET PREDICTION

<sup>1</sup>Ashutosh Dubey, <sup>2</sup>Anmol Agarwal, <sup>3</sup>Preeti Naval

<sup>1,2,3</sup>CSE Department,

<sup>1,2,3</sup>SRMCEM, Lucknow, India

<sup>1</sup> myselfdubey99@gmail.com, <sup>2</sup> krishnagarg760@gmail.com, <sup>3</sup> er.preetinaval09@gmail.com

**Abstract :** Due to the inconsistency and highly instability nature of the Stock market, it has become a very tough task for investor to make accurate predictions. Polishing the accuracy of predictions has become the main task of many researchers. From the study of working with historical report to using the latest machine learning algorithms and deep learning techniques, developers are busy finding out the best possible model of accurate prediction. Many new models are designed that can make good prediction of stock prices. Investors are interested in calculating the immediate next-day prices and as well as future share prices in the long term. This review paper examine the algorithms and techniques that are useful for making accurate predictions.

**Keyword:** Closing Price, Market, Model, Predict, Price, Stock, Algorithm.

### I. INTRODUCTION

Stock market is trading platform where different investors buy and sell share to earn huge profit . Stock market movement effects the profit of shareholder. Buyers try to buy stocks with low prices and sell stocks at high prices and earn huge profit.

Predicting the stock market is one of the tough things to do. There are so many factors involved in the price movement – technical factors vs. fundamental factors , results, news etc. All these aspects combine to make share prices unstable and very difficult to predict with a good accuracy.

A stock (also known as shares) is a ownership of a business by any individual or a group of people. The attempt to calculate the future value of the share price is known as a stock market prediction.

### II. DESCRIPTION OF DATA SET

We will implement Linear Regression, Auto ARIMA, Prophet, LSTM algorithms on NSE TATA GLOBAL dataset, and we will also analyses the trends of data manipulation of . Dataset is obtained from Yahoo Finance.

#### 2.1. OBTAINING DATASET

The obtained data contained five features:

- Date: Trading Date of Stock Scrip.
- Opening price: When trading in scrip begins in a day this is opening price of stock.
- Closing price: When trading in scrip ends in a day this is closing price of stock.
- High: The highest price in a day at which the stock was traded (day).
- Low: The Lowest price in a day at which the stock was traded during a period(day).

|   | Date       | Open   | High   | Low    | Last   | Close  |
|---|------------|--------|--------|--------|--------|--------|
| 0 | 2018-10-08 | 208.00 | 222.25 | 206.85 | 216.00 | 215.15 |
| 1 | 2018-10-05 | 217.00 | 218.60 | 205.90 | 210.25 | 209.20 |
| 2 | 2018-10-04 | 223.50 | 227.80 | 216.15 | 217.25 | 218.20 |
| 3 | 2018-10-03 | 230.00 | 237.50 | 225.75 | 226.45 | 227.60 |
| 4 | 2018-10-01 | 234.55 | 234.60 | 221.05 | 230.30 | 230.90 |

Fig 1 : Description Of Dataset

### III. OBJECTIVE:

The main objective of this study is to:-

- Study about different methods(algorithms) to obtain more accurate stock prediction price and to evaluate them with some performance measures.
- To reduce the error percentage in predicting the future stock prices.
- It increases the chances for the investors to predict the prices more accurately by reducing error percentage and thus gain benefits in share markets.

After getting the idea about different methods of stock market forecasting techniques we can understand that which algorithm is best by using which methods we will get more accurate results. Then we will be able to reduce the amount of error by which investors can predict the price accurately.

The recent trend in stock market prediction methodology is the use of machine learning algorithm which makes predictions based on the values of current stock market indices by training on their previous closing values. Machine learning itself deploys different models to make prediction easier and authentic. The paper focuses on the use of Linear Regression, KNN, Auto ARIMA, Prophet, and LSTM to predict stock values. Factors considered are open price, close price, low, high and volume

### IV. LITERATURE REVIEW:

This consider attempted to make models for guess of the market trade and to select whether to buy/hold the stock using data mining and AI technologies. The ML algorithms like Naive Bayes, k-Nearest neighbors, Support Vector Machine, Artificial Neural Network and Random Forest has been used for prediction of stock market. Kim and Han in design a model as a union of artificial neural networks (ANN) and genetic algorithms (GA) with variational of features for predicting stock price index. The data used in their research includes the technical indicators as well as the direction of change in the daily Korea stock price index. Previous methods of stock forecasting include the use of Artificial Neural Networks and Convolution Neural Networks with an error rate of 20%. • Machine learning is an efficient way to represent process like prediction. It predicts a market value close to the noticeable value, thereby decreasing the error. Commencement of machine learning to the area of stock prediction has attract many researches because of its efficient and highly accurate measurements.

### V. ALGORITHMS USED:

We have tried to designed a model based on different ML Algorithms to predict the next day closing price and try to analyze which model best fit the data and generate best prediction. The different algorithms used are:

#### 5.1. LINEAR REGRESSION

It is a ML algorithm based on supervised learning. It carry out regression task. Regression models a target prediction value dependent on independent variables. It is mostly used for determining the relationship between variables .

$$Y = \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_n X_n$$

Y= Predicted Variable

$\theta_1, \theta_2, \dots, \theta_n$ = Coefficients

$X_1, X_2, \dots, X_n$ = independent variable

#### 5.2. Auto ARIMA

ARIMA, stands for 'Auto Regressive Integrated Moving Average' is actually a class of models that based on time series that predict future value. It is specified by three important parameters: p ,d, q.

- \* p= previous value used to predict future value
- \* q= previous errors used to predict next value
- \* d= order of differencing

### 5.3. Random Forest

It is a Supervised Machine Learning Algorithm that is used commonly in Classification and Regression problems. It builds decision trees on samples and takes their majority vote for classification and average case for regression. One of the most important features of this Algorithm is that it can handle the data set containing continuous variables in the case of regression.

Random Forest is a classifier that contains decision trees on various subsets of the given data and takes the average to increase the predictive accuracy of that dataset. Instead of depending on single decision tree, the random forest takes the prediction from each tree and depending on the majority votes of predictions, and it predicts the final data.

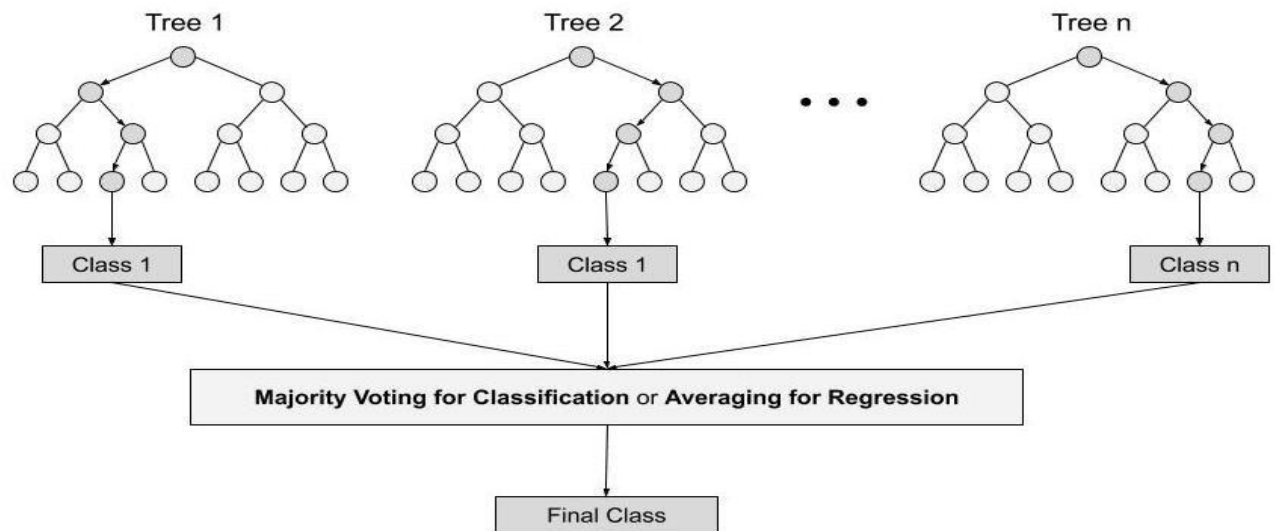


Fig 2. Random Forest

### 5.4. Prophet

The number of time series techniques that can be execute on the stock price prediction , but majority of these techniques require extreme data preprocessing before fitting the model. Prophet, designed a Facebook, is a time series forecasting algorithm that requires no data preprocessing and is very simple to implement.

$$y(t) = g(t) + s(t) + h(t) + e(t)$$

$g(t)$  denotes trend (long term changes)

$s(t)$  denotes seasonality (short term changes)

$h(t)$  denotes effects of holidays to the prediction

$e(t)$  denotes the unconditional changes (error term).

$y(t)$  is the prediction.

### 5.5. LONG SHORT TERM MEMORY(LSTM):

It is a special type of RNN, which can learn long-term dependence. It carries information in memory, alike computer memory. It is capable to read, write and delete data in its memory. This memory can be consider as a closed cell, with a closed description, the cell decides to reserve or remove data. LSTM uses three gates: input, forget and exit gate. These gates examine whether new input (input gate) should be allowed or not.

- a. Forget gate: The forget gateway determines when certain parts of the cell will be inserted with information that is more recent.
- b. Input gate: Based on the input, this network category reads the conditions under which any data should be stored (or updated) in the cell.
- c. Output gate: Depending on the input mode and the cell, this gate determines which information is carried to the next location in the network

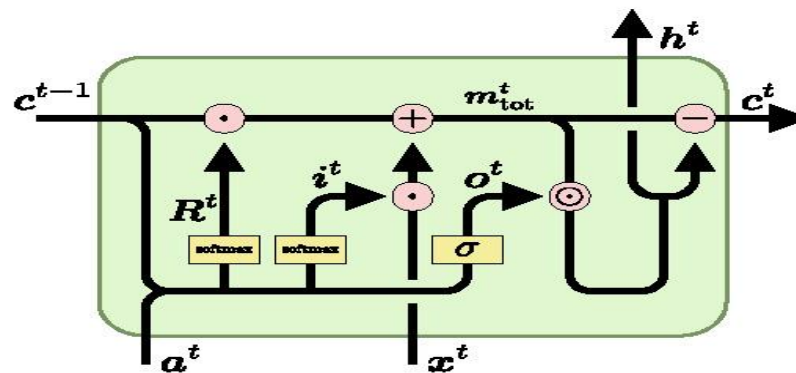


Fig 3 . LSTM Memory Cell

## VI. ERROR ANALYSIS

To evaluate the accuracy of the different algorithms, the same dataset is trained for different algorithm. Predicted closing price are subjected to Root Mean Square Error(RMSE) for finding the minimized error in predicted price and help to know the best algorithm for stock price prediction.

❖ **RMSE:**

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}}$$

where,

n = total window size.

$x_i$  = original closing price.

$y_i$  = predicted closing price .

## VII. RESULT

The above discussed different algorithms are used to predict the future price of the stock and the different results are obtained. The accuracy of different models are calculated with the RMSE(Root Mean Square Error) which helps us to conclude that with model predict the result with good precision.

Table 1. Error Analysis

| Algorithm (Model) | Error Analysis |
|-------------------|----------------|
| Linear Regression | 7.55           |
| Auto Arima        | 2.07           |
| Random Forest     | 3.58           |
| Prophet           | 2.84           |
| LSTM              | 0.62           |

According to research and study the domain range of RMSE values between 0.2 and 0.5 shows that the model can relatively predict the data accurately and is a best fit model.

## VIII. CONCLUSION

Predicting stock market price is a complex task due to rapidly changing stocks values which are dependent on multiple parameters which form a complex pattern. The Historical dataset available on website consist of only few features like high, low, open, close, of stock prices, volume of shares traded etc. which are not sufficient. Machine learning algorithms is used for predicting the next Day Closing price of the stock and for a comparative analysis.

Predicting stock market with 100% accuracy is not at all possible as multiple factor impact the price of stock. All models are designed comparatively yield good result. The comparative analysis is based on RMSE values clearly indicate that the model gives good prediction of stock price. But we can conclude that the time series forecasting based algorithms resulted in the good accuracy

such as Auto ARIMA, Prophet, LSTM. However, LSTM came out to be the best algorithm which predict the price with the highest and good accuracy. The techniques have shown an improvement in the accuracy of predictions, thereby predicting accurate results. In future multiple factors such as revenue, debt, results can be added to obtain higher accuracy.

## REFERENCES:

- [1] Tae Kyun Lee et al. "Global stock market investment strategies based on financial network indicators using machine learning techniques", *Expert Systems with Applications*, 117(2019):228-242
- [2] Olivier C., Blaise Pascal University: "Neural network modeling for stock movement prediction, state of art". 2007 .
- [3] M. Tsanga, Paul Kwoka, S.O. Choya, Reggie Kwanb, S.C. Nga, Jacky Maka, Jonathan Tsangc, Kai Koongd, Tak-Lam Wong (2007). Design and implementation of NN5 for Hong Kong stock price forecasting, *Engineering Applications of Artificial Intelligence* 453– 461
- [4] Kim K, Han I. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Syst Appl*. 2000;19:125–32 .
- [5] 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
- [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] Billah M, Waheed S, Hanifa A (2015) Predicting closing stock price using artificial neural network and adaptive neuro fuzzy inference system: the case of the dhaka exchange. *International Journal of Computer Application* 11(129):975–8887
- [9] Ashutosh Kale, Omkaar Khanvilkar (2018) 'Forecasting Indian Stock Market Using Artificial Neural Networks'. Published in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA).
- [10] Kai Chen, Yi Zhou and Fangyan Dai —A LSTM-based method for stock returns prediction: A case study of China stock market, *IEEE International Conference on Big Data*, 2015.