International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

# Stock Market Prediction using Machine Learning Techniques

**Vaibhavi[1], K. Viswavardhan Reddy[2]**

[1]*Student, Dept. Tele Communication, RV College of Engineering, Bengaluru, India*
[2]*Assistant Professor, Dept. Tele Communication, RV College of Engineering, Bengaluru, India*

---

**Abstract** - *Stock market prediction is act of trying to determine future value of stock. As study of stock market subject is very vast and huge; hence, it is difficult to predict accurate value of stock. So, in this work we will be using machine learning algorithms to predict the accurate value of stock. There are numerous ways of executing the forecast system like Fundamental Analysis, Technical Analysis, Market Mimicry and Time series aspect structuring. Machine learning is one of the trending and promising technologies in the field of stock market. In this paper we used Linear Regression (LR), Random Forest and Time Series Prediction algorithms. From the results, it is seen that Random Forest has shown promising results when the system is trained with 80-20%.*

***Key Words***: -Stock market, Machine Learning, Linear Regression, Random Forest, Time series Prediction.

## 1. INTRODUCTION

An individual owning a percentage of the share in a corporation has the ownership in the corporation proportionate to his share that is an inventory in preferred indicates rights on business enterprise with the aid of a selected individual or a set of people. The stock market is mainly a combination of various buyers of stock and sellers of stock, which signify ownership rights on businesses; these may comprise securities listed on a public stock exchange, as well as stock that are traded private only. The effort to define the future rate of the stock market is known as a stock market prediction. The system must predict as per the real-life conditions and should suit to real world applications. The model is also has to consider all the variables that may affect the stock's price and performance such as Stock bonus /split /dividend, Management change (weak/strong) Political and regulatory environment/statement by a political leader, new releases on earnings and profits etc.

Stock market forecasting has expanded massive popularity in financial markets across the world. The proficiency to predict the direction of a stock price or an index is very significant for various objectives. Few of old methods reduced the risk of investment for investor and supported in identifying prospects for investor.

From the study, it is observed that machine learning is widely used technique in the field of stock market prediction. Also it is found that most of the authors implemented Random Forest and Long Short Term Memory (LSTM) techniques to forecast the future value of stock [2], [3], [8].

In [1] authors implemented linear regression and LSTM (Long Short Term Memory) models. The built model is then evaluated by using test dataset of stock market which is taken from Yahoo Finance. And authors deduce that Regression encompasses minimizing error and Long Short Term Memory says to remember the data and results for the long run.

Authors have suggested a stock market forecast based on social media opinions [2]. Where authors seek to forecast the stock market prices of numerous organizations by carrying out sentiment analysis of the social media data such as tweets associated with the respective organization. The paper describes how Twitter will support you to become a better investor by making suitable investment choices with its information of the market emotions.

[7] Proposed a method of feature extraction for selection of useful stock indexes and proposes deep learning model to perform sentiment analysis of financial news as another influencing factor for influencing stock trend. Then it proposes model which forecasts share market price by using LSTM. Proposed paper method with feature compression and sentiment analysis for stock market forecasting and it improved accuracy than SVR (support vector regression) by about 20%. Market sentiments were trapped by sentiment analysis and it is a very significant factor for influencing stock trends and improves prediction accuracy by 5%.

## 2. PROPOSED METHODOLOGY

Figure 1 shows the process involving in stock market prediction.

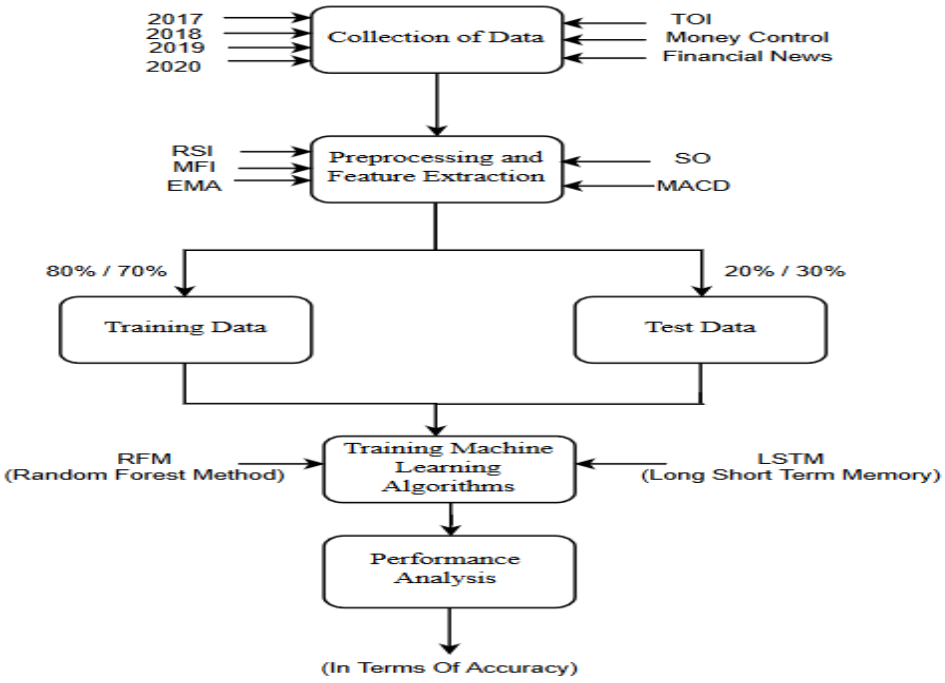


**Fig -1**: System Architecture

For this work input data is collected from yahoo finance. We have collected data of 5 years from 2016 to 2020. The collected dataset has four features that is Open, Close, High and Low. Further the dataset is divided into two parts, that is training and testing dataset. We examined algorithms for three cases, 1. 80% training and 20% testing, 2. 70% training and 30% testing and 3. 60% training and 40% testing to check the accuracy of algorithms.

### 2.1 Linear Regression Model

Basically regression based models are used in evaluation of continuous data. It is predictive modelling technique which is used to determine the relationship between independent and dependent variables as discussed in equation (1). It also determines the correlation between two variables. Equation for linear regression is as follows,

$$y = mX + c \quad (1)$$

Where, X is independent variable and y is dependent variable, m is slope of the line as discussed in equation 2 and c is intercept. Least square method is used to determine the regression line (the one with least error). Slope of the line is calculated by following equation,

$$m = \frac{\sum(x-x_{bar})(y-y_{bar})}{(x-x_{bar})^2} \quad (2)$$

In this work X is date, y is open, close, high and low price of stock. Where, open price is stock at which opens at the start of market, close price is price at the closing hour, high price is highest bid of the day, and low price is lowest bid of the day.

Accuracy of the linear regression algorithm is calculated by using R-Square method as discussed in equation 3. R2 is used to determine best fit of the line which is given as follows,

$$R^2 = \frac{\sum(y_p - \bar{y})^2}{\sum(y-\bar{y})^2} \quad (3)$$

Where, yp is predicted value and y is actual value.

### 2.2 Random Forest Model

Random forest is one of the widely used model in prediction of stock [9]. It works on the principle of decision tree. It is a type of ensemble learning. The figure 2 shows the basic diagram of random forest where it uses many decision trees to predict the outcomes therefore it is also called as ensemble learning technique. Steps involved in creating tree are as follows:

i. Consider a parent node and split the parent node into two daughter nodes.

ii. Splitting of parent node into daughter nodes is depends on splitting principle as in equation (4).

iii. Base (bottom) nodes are known as terminal nodes.

iv. For regression the predicted value at a node is calculated by taking the average response of variable for all observations in the node.

v. Splitting principle: For Regression splitting of nodes is calculated by using residual sum square

$$RSS = \sum_{left}(y_i - y_L^*)^2 + \sum_{right}(y_i - y_R^*)^2 \quad (4)$$

Where, $y_L^*$ = mean y-value for left node

$y_R^*$ = mean y-value for right node

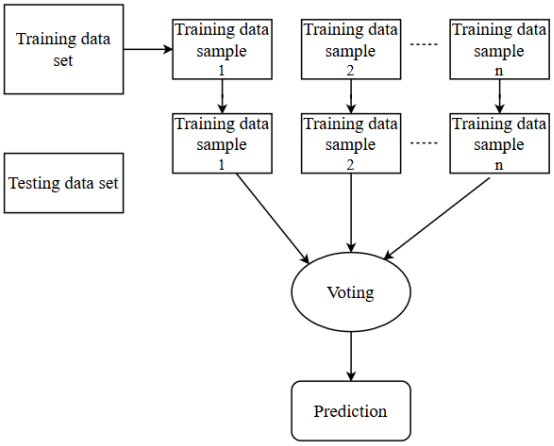


**Fig -2**: Random Forest's Decision Tree

## 2.3 Time Series Prediction Model

In time series prediction there are many methods to analyze input stock data in order to extract useful information and characteristics of data. In this, for prediction of future stock price previous year data is used. An autoregressive integrated moving average (ARIMA) model is an overview of an autoregressive moving average (ARMA) model. Both of these models are used to predict the future values in the series as in equation 5 and to understand the data thoroughly. Essentially the ARIMA model is a "cascade" of two models. The first is non-stationary and the second is wild stationary,

$$Y_t = (1 - L)^d X_t \quad (5)$$

## 3. EXPERIMENT RESULTS

In this work, data is collected from yahoo finance and the collected data is divided into training and testing set.

Chart 1 shows the HUL stock performance of last five years. From graphs it is observed that, stock market rate of HUL varies randomly. In above figures open, close, high and low prices of stocks are indicated by red, yellow, orange and green colors respectively. Open price is day's opening price, close price is day's closing price, high price is day's highest bid, and low price is day's lowest bid.


**Chart -1**: 5 Years stock performance of HUL

From yahoo finance, we have collected previous five years data for eight stocks that is Asian Paints, Axis Bank, Bajaj Finance, HUL, Titan, TCS, Titan, Tech Mahindra.

Chart 2 shows the graph between actual price and predicted price of Hindustan Unilever Limited of Linear Regression for 3 cases 80% training 20% testing shown in 2(a), 70% training 30% testing shown in 2(b) and 60% training 40% testing shown in 2(c). In the below charts red line is indicating the regression line which maps predicted values with actual data points. Moreover, it is seen that from regression line, the actual data points are far away and this indicates the error. Therefore, accuracy obtained for 80-20% is 94.35, for 70-30% is 94.16 and for 60-40% is 93.46. And it is seen that for Hindustan Unilever Limited, Linear Regression gives better accuracy for 80% training and 20% testing set.

**Chart 2(a)**: Hindustan Unilever Limited[80-20]

**Chart 2(b)**: Hindustan Unilever Limited[70-30]

**Chart 2(c)**: Hindustan Unilever Limited[60-40]

Table 1 shows the accuracy calculated by Linear Regression model for 3 cases. From the table, it is observed that Linear Regression model of Machine Learning technique gives better accuracy when we consider 80% training and 20% testing set.

**Table -1:** Accuracy Calculated by LR

| Stocks | Linear Regression Model | | |
|---|---|---|---|
| | 80%-20% | 70%-30% | 60%-40% |
| Reliance Ir | 95.29 | 94.60% | 94.12% |
| TCS | 85.63 | 83.38 | 85.59 |
| Axis Bank | 73.01 | 72.37 | 72 |
| Tech Mahi | 72.79 | 72.05 | 71.49 |
| Asian Pair | 84.97 | 84.46 | 84.36 |
| Bajaj Fina | 93.34 | 92.59 | 92.04 |
| HUL | 94.35 | 94.16 | 93.46 |
| Titan | 91.34 | 93.07 | 92.5 |

Chart 3 shows the graph between actual price and predicted price of Hindustan Unilever Limited of Random Forest for 3 cases 80% training 20% testing shown in 3(a), 70% training 30% testing shown in 3(b) and 60% training 40% testing shown in 3(c). In the below charts red curve indicates predicted points and blue points are actual data points. From charts it's observed that predicted data points exactly matches to actual data points. Accuracy obtained for 80-20 is 99.9, for 70-30 is 99.88 and for 60-40 is 99.77. And it is seen that for Hindustan Unilever Limited, Random Forest model gives better accuracy for 80% training and 20% testing set.

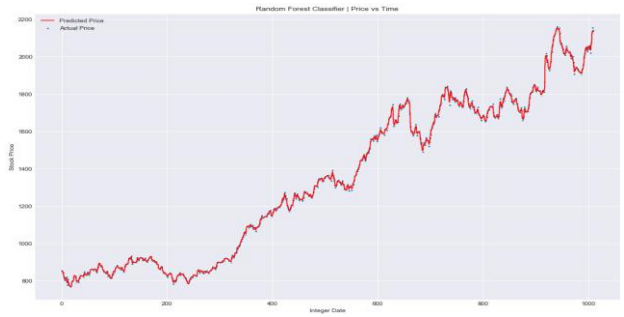

**Chart 3(a):** Hindustan Unilever Limited[80-20]

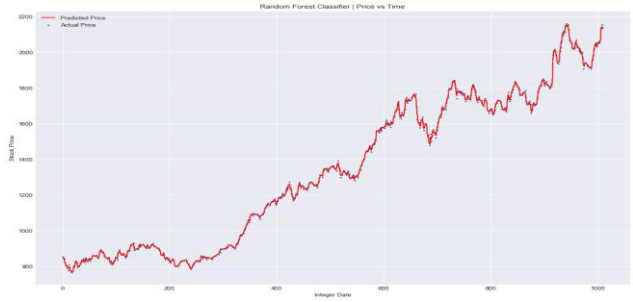

**Chart 3(b):** Hindustan Unilever Limited[70-30]

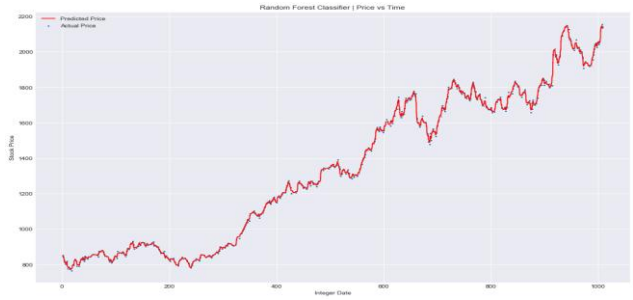

**Chart 3(c):** Hindustan Unilever Limited[60-40]

Table 2 shows the accuracy calculated by Random Forest model for 3 cases. From the table, it is observed that Random Forest model of Machine Learning technique gives better accuracy when we consider 80% training and 20% testing set.

**Table -2:** Accuracy Calculated by RF

| Stocks | Random Forest Model | | |
|---|---|---|---|
| | 80%-20% | 70%-30% | 60%-40% |
| Reliance Ir | 99.86 | 99.81 | 99.79 |
| TCS | 99.81 | 99.8 | 99.79 |
| Axis Bank | 99.39 | 99.4 | 99.34 |
| Tech Mahi | 99.57 | 99.55 | 99.3 |
| Asian Pair | 99.69 | 99.66 | 99.59 |
| Bajaj Fina | 99.82 | 99.81 | 99.8 |
| HUL | 99.9 | 99.88 | 99.77 |
| Titan | 99.86 | 99.77 | 99.7 |

Chart 4 shows the graph between actual price and predicted price of Hindustan Unilever Limited of Time Series Prediction for 3 cases 80% training 20% testing shown in 4(a), 70% training 30% testing shown in 4(b) and 60% training 40% testing shown in 4(c). In the above figures red curve indicates predicted points and green points are actual data points. From figures it's observed that predicted data points exactly matches to actual data points. Accuracy obtained for 80-20 is 97.82, for 70-30 is 97.89 and for 60-40 is 98.09. And it is seen that for Hindustan Unilever Limited, Time Series Prediction gives better accuracy for 60% training and 40% testing set.

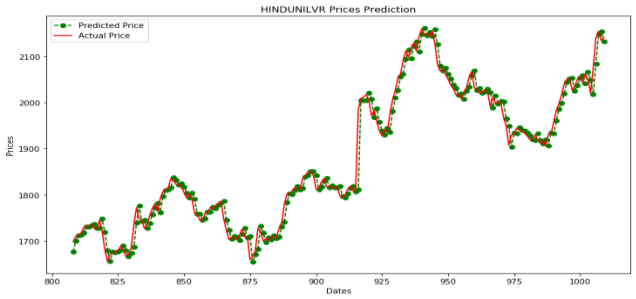


**Chart 4(a):** Hindustan Unilever Limited[80-20]

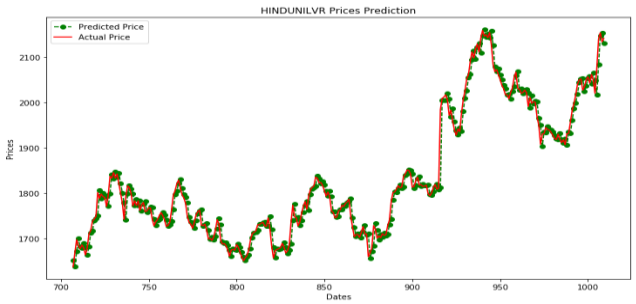


**Chart 4(b):** Hindustan Unilever Limited[70-30]

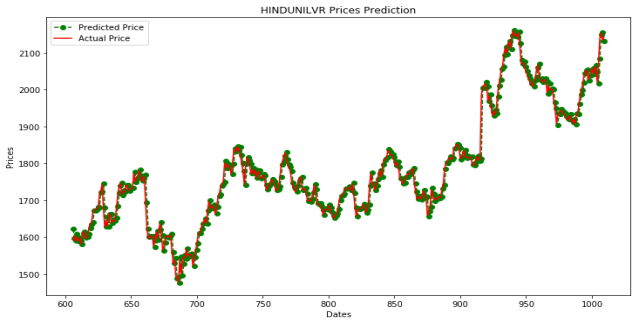


**Chart 4(c):** Hindustan Unilever Limited[60-40]

Table 3 shows the accuracy calculated by Time Series prediction for 3 cases. From the table, it is observed that Time Series Prediction model of Machine Learning technique gives better accuracy when we consider 60% training set and 40% testing set.

**Table -3:** Accuracy Calculated by TSP

| Stocks | Time Series Prediction | | |
|---|---|---|---|
| | 80%-20% | 70%-30% | 60%-40% |
| Reliance In | 97.55 | 98.18 | 98.31 |
| TCS | 86.1 | 94.57 | 95.23 |
| Axis Bank | 93.45 | 96.12 | 98.13 |
| Tech Mahi | 94.95 | 95.61 | 96.01 |
| Asian Pain | 98.38 | 98.61 | 98.75 |
| Bajaj Finar | 97.22 | 98.91 | 98.98 |
| HUL | 97.82 | 97.89 | 98.09 |
| Titan | 90.42 | 96.68 | 98.31 |

By comparing accuracy of different algorithms, it is noticed that Linear Regression model and Random Forest of Machine Learning technique gives better accuracy when we given with 80% training set and 20% testing set. Also it is seen that Time Series Prediction model of Machine Learning technique gives better accuracy when we given with 60% training and 40% testing set. From this we can conclude that Random Forest gives better accuracy compared to Linear Regression model and Time Series Prediction algorithm in prediction of stock prices.

## 4. CONCLUSIONS

Proposed models are used to predict the accurate value of stock index from the given inputs as well as used to calculate the accuracy of each algorithm. We study the use of Linear Regression, Random Forest and Time Series Prediction models to predict price of different stocks.

Of these algorithms, we saw that Random Forest model gave us accurate results with an accuracy of (98 ±1) %. So we conclude that Random Forest model gives better accuracy and it is most suitable algorithm for forecasting the price of stock. From this work we have got satisfactory results by predicting the price of stock from historical data. And as the proposed models can only predict the price of stock for next trading day.

In the future we will be trying to predict price of stock by using real time signals, and we will try to implement different algorithms which will gives better accuracy than Random Forest Algorithm.

## REFERENCES

[1] I. Parmar et al., "Stock Market Prediction Using Machine Learning," in the proceeding of First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 574-576.

[2] T. Mankar, T. Hotchandani, M. Madhwani, A. Chidrawar and C. S. Lifna, "Stock Market Prediction based on Social

Sentiments using Machine Learning," in the proceeding of International Conference on Smart City and Emerging Technology (ICSCET), Mumbai, 2018, pp. 1-3.

[3] R. Batra and S. M. Daudpota, "Integrating Stock Twits with sentiment analysis for better prediction of stock price movement," in the proceeding of International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, 2018, pp. 1-5.

[4] Hegazy, Osman et al., "A Machine Learning Model for Stock Market Prediction," in the proceeding of International Journal of Computer Science and Telecommunications, Egypt, 2016, volume 4,pp. 17-23.

[5] H. Gündüz, Y. Yaslan and Z. Çataltepe, "Stock market prediction with deep learning using financial news," in the proceeding of 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4.

[6] Mruga Gurjar et al., " Stock market prediction using ANN," in the proceedings of International Research Journal of Engineering and Technology (IRJET), Pune, India, 2018, volume 5.

[7] S. S. Maini and K. Govinda, "Stock market prediction using data mining techniques," in the proceeding of International Conference on Intelligent Sustainable Systems (ICISS), Palladam, 2017, pp. 654-661.

[8] B. Jeevan, E. Naresh, B. P. V. kumar and P. Kambli, "Share Price Prediction using Machine Learning Technique," in the proceeding of 3rd International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, 2018, pp. 1-4.

[9] Hakob GRIGORYAN, "A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA)", in the proceeding of DSJ 2016.