

# Short Term Stock Market Prediction by using Hybrid Approach

Chetan Gondaliya  
Ganpat University  
Ganpat Vidyanagar  
Kherva

Ajay Patel, PhD  
Ganpat University  
Ganpat Vidyanagar  
Kherva

Satyen Parikh, PhD  
Ganpat University  
Ganpat Vidyanagar  
Kherva

## ABSTRACT

Nowadays, finance market has become the most prevalent sector in the world. In finance market, the stock market is a main pillar which represent the major economy of the Country. The stock market nature is random which is dependent on the so many factors like fundamental, technical, overseas news, domestic news, Government policies, global demand and supply etc. Therefore, it is necessary consider each factor which are lies under timeline of the forecast. Most of the researcher have just used technical parameters for stock market prediction. It may happen that stock has good technical although it is not giving the good results. On the other side, stock has poor technical but good sentiment, given good result. Nowadays, the most of the peoples are expressing their views on the social media platforms. This news can be taken to process and discover the features which can be used for the stock market predictions. The main aim of this research paper is to develop hybrid model which can be used technical as well as sentiment parameters for stock market prediction in the short-term duration.

## Keywords

ML algorithms, Stock market prediction, Sentiment analysis, Technical analysis, Indian stock market, short term prediction

## 1. INTRODUCTION

Stock market prediction is act of trying to forecast the future value of the stock. Nowadays, it is very critical to predict the future stock movements because of continuously change in the nature of the stock market. Newcomer investors are facing problem when, how much and in which stock to invest for getting the maximum return on investment (ROI). In addition to that, stock market produces larger amounts of the data on the daily basis. It is very difficult for the newcomer investors to process these large amounts of data within fraction of time and derived the conclusion which can further used for the stock market prediction.

Fundamental and technical analysis are two main approaches which can used for stock market prediction. The fundamental approach main based on intrinsic value of the company. The intrinsic value is derived by analysing the various financial records of the company. It is very lengthy process and sometimes it may happen that technical analyst has bias on any particular sector or stock [6-7]. On the other side, the technical analysis is depending on the historical data of the stock. The historical data means past stock transactions detail such as open, high, low and closing price. These base technical parameters are used to find the patterns and trends in the stock market. The technical analysis works-based history repeat itself theory in which technical analyst believe that patterns repeat in the future [6-7]. Nowadays, it may happen

that stock has good technical but sentiment of the stock is not good therefore stock is not giving good result. On the other side, the stock has poor technical but stock is giving good result because of sentiment of the stock is good. Hence, it is necessary to consider technical as well as sentiment parameters for developing the stock market prediction model for the short-term duration [8-9].

This research focus on to develop a model which is based on the technical and sentiment factors. The technical factor is used to identify the trends and patterns in the stock market whereas sentiment factors is used to calculate the impact of news articles on the stock prices. In addition to that, the researcher is endeavors to used hybrid approach to build the stock prediction model which can be used for the short-term duration.

The paper is divided into five sections. The first section is deliberated about the current scope of the stock market prediction. In second section, the comprehensive study of various literature which are based on the stock market prediction. The third section describes about the data collection and methodology of the proposed model. The fourth section discusses about result and their discussion. At last, in fifth section discuss about the conclusion of the proposed study.

## 2. LITERATURE REVIEW

The literature review is an important phase which can be carried out before doing any type of research to find the gap and further scope into any domain. As per random walk theory (RWH) and efficient market hypothesis (EMH) theory, stock price reflects all the available information in the market hence it is almost impossible to predict movement of stock price. The author has also suggested that past movement of trends of a market cannot be taken into consideration and same trends will not repeat in the future movement of the stock [1].

According to Mehar et. al. (2019), the prediction of future stock price is very challenging task due to random nature, volatility and non-linear nature of the stock market in the finance market. With the invention of latest technologies, nonlinear types of data can be predicted by using the machine learning (ML) algorithms. The author has used artificial neural network (ANN) and random forest (RF) algorithm to predict next day stock price by using the various technical parameters. As per result derived from the experiment of the study, ANN gives better prediction of the future stock price compared to RF. The author has also suggested that financial news articles can be consider with technical parameters for better prediction result of the model [2]. As per Bruno et. al. (2018), Stock market is continuously changing his nature and

patten because of the various external factors like overseas news, domestic news, Government policies and global demand. Nowadays, social information is a valuable resource which should be taken into consideration while developing a predictive model of the stock market. The author has used various technical parameters such as simple moving average (SMA), weighted moving average (WMA), relative strength index (RSI) and average true range (ATR) for feeding into the support vector regression and support vector machine algorithm for developing the model. Experimental study shows that support vector machine gives good result compared to support vector regression algorithms. Author has suggested to used fundamental indicators along with technical parameters for developing the better prediction model [3].

As per Aatish et. al. (2020), stock market is very common buzz word among the end users and trader in which buys and sells of the stock by investing money into the market. Traders are continuously watch the stock movements in which they invest their money. It is very difficult for traders to find hidden fluctuations and find the relevant news from huge amounts of social sentiment to data and analysis it within fraction of time which can be used for taking further decision in the stock market. Therefore, the researcher has focused on to solve out these problems by using the machine learning algorithm to discovered the non-linear patterns of the stock market. In study, support vector machine (SVM) algorithm has used in the experiment of the study. The data is collected from yahoo finance website by using data reader library of the python. Author has suggested to used sentiment analysis to finds out core features from the social sentiment data to achieve the better result of the model [4]. Some of works like Rupesh has used the various technical indicators such as MACD, RSI, KDJ and BB in decision tree (DT) and random forest (RF) algorithms to capture the short-term trends in the stock market. The study has collected the three-year technical transactions data from the BSE website. The various data pre-processing and data sampling techniques has used to reform the data for better development of the proposed model in the study. The main delimitation of this study is that, it relies only on technical data which is not enough for accurate prediction of the stock movement. He has also suggested to integrate the news-based system in this model for better prediction of the model [5][10-11].

### 3. DATASET AND PROPOSED METHODOLOGY

#### 3.1 Data Collection

The data is a core element of any research work so without proper data research should not be possible. In this work, there are mainly two types of data is required for the development of the model. The first type of data is technical data which usually referred as open, high, low and closing price of the stock. The yahoo finance portal is used to fetch OHLC data for the proposed research work. The researcher has built one script in python language to fetch the three-year of the technical data. The second type of data is required is called sentiment data. The sentiment data source is forum discussion and news articles. These data is collected by using web crawling. In addition, the author has developed a various python script to collected the sentiment data from the different sources of the sentiment. The data has been collected for 1-jan-2018 to 31-dec-2020. Figure-1 describes the data sources which is given below;

**Table 1. Data collection sources**

RSS feed	Forum discussion	News portals
Money Control	Money control	Marketmojo
Economics times	Yahoo conversations	Outlook India
United news of India	Traderji	Money Control
Business standard	Valuepickr	Economic times
Business Line		United news of India
		Millenniumpost
		Htsyndication
		The Himalayan
		Business standard
		Free press
		Equity bulls
		Business line
		The statesman
		Investing
		Stock adda
		Bse2nse
		Reuters

Sentiment data has been divided into three categories like positive, negative and neutral. The figure-2 reveals the sample words of each category.

**Table 2. Sample phrases of each category**

Sr. No	Positive phrases	Negative phrases	Neutral phrases
1	Buy	Sell	Hold
2	Up move	Down move	Neutral
3	Strong	Bad	Equal
4	Only buyer	Only seller	Consolidation

#### 3.2 Proposed Framework

The proposed model aims to helps to investors to taking investment decisions to avoid risks and financial crisis. In proposed model, the stock market news and past historical stock transaction details are used to built the hybrid model in order to get better classification accuracy of the stock market.

The first phase is focuses on the sentiment analysis in which first step is to collection all social sentiment data from the various sources such as forum discussion, news articles and RSS feed. There are no any standard rules to write views about any topic or stock so these types of data usually referred as unstructured data. The next step is to apply pre-process technique on textual data to remove noise and duplication from the data. The one of the main obstacles to process textual

data is that we cannot directly feed these types of data into machine learning algorithms. Hence, these types of data should be converted into proper structured format so we can directly use to feed to machine learning algorithms. There are no of techniques used to do so in the proposed research work. The first step is tokenisation in which user views are splinted into different tokens by using space delimiters. These tokens may contain some unnecessary tokens which is not representing any special meaning toward to user's view. Therefore, it should be removed from the user's view to reduce overhead in the learning phase of the algorithms. The next step is stemming in which all similar types of words are converted into their root format to reduce the token corpus of sentiment. The document representation is last step in the sentiment analysis which used to represent user's view by using the vector space model.

The main aim of technical phase is to identify the patterns and trends by using the past historical data of the stock and integrate the sentiment features with technical feature to build the hybrid model which can be further used to make better accurate prediction in the stock market. The first step is to find missing values in the technical data to reduce the inconsistency of the data. The next step is to build the technical features from the OHLC features which can further used for discover the trends and patterns in the stock market. The next step is integration in which all the features of sentiment and technical are combined by using date parameter. At last, the various ML algorithms are used to train in the training process and predict the future forecast of the stock movement in the testing phase. The next section discussed about the results of various ML algorithms and their outcomes.

#### 4. RESULT AND DISCUSSION

The proposed model is endeavors with two experiments: 1) Hybrid model by using KNN, Decision tree and NN as base learner and logistic regression is used as hybrid learner. 2) Hybrid model by using logistic regression, random forest and SVM as base learner and logistic regression is used as hybrid learner.

This section provides detail insights of experiment performed in the proposed research work. As per the experiment result of both experiments while comparing the experiment-1 and experiment-2 for the given data, it was observed that the experiment-2 has showing higher classification accuracy in 70 percent of the cases. Hence, experiment-2 taken in to consideration for further result discussion here onwards. Figure-3 represents the classification accuracy of pharma sector in experiment-2.

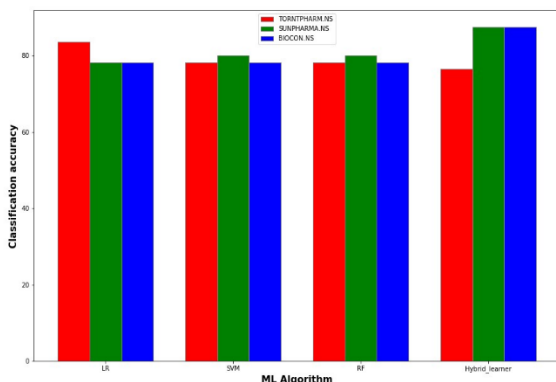


Figure 2 Classification accuracy of Pharma sector

In Figure-3, proposed model showing 87.45 percent of accuracy compare to LR, SVM and RF gives accuracy of 78.18, 80 and 80 respectively for the sun pharma stock while in Torrent stock, proposed model showing 76.55 of accuracy compared to LR, SVM and RF give accuracy of 83.64, 78.18 and 78.18 respectively. This confirms on an average hybrid learner is best predictor of price moment in pharma sector. Figure-4 represents the classification accuracy of real estate sector in experiment-2.

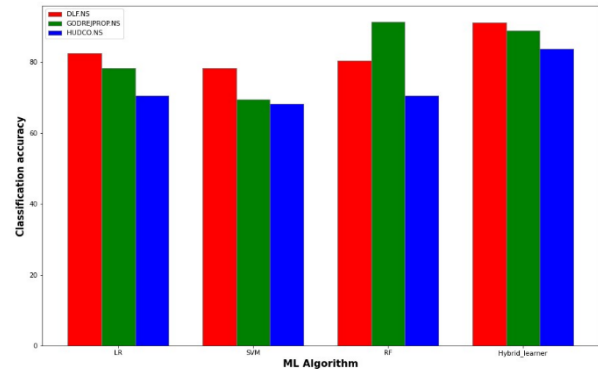


Figure 3 Classification accuracy of Real estate sector

In Figure-4, proposed model showing 91.13 percent of accuracy compared to LR, SVM and RF gives accuracy of 82.61, 78.26 and 80.44 respectively for DLF stock while in hudco stock, proposed model showing 83.82 of accuracy compared to LR, SVM and RF give accuracy of 70.46, 68.18 and 70.46 respectively. This confirms on an average hybrid learner is best predictor of price moment in real estate sector. Figure-5 represents the average classification accuracy in all sector in experiment-2.

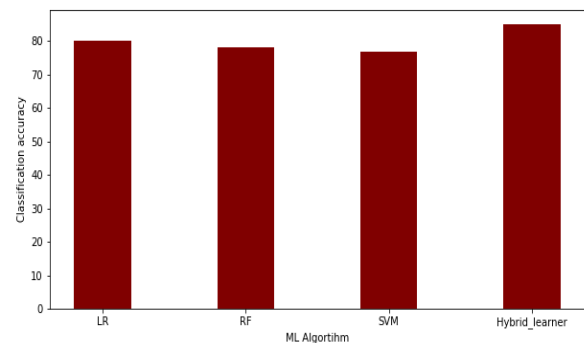


Figure 4 Average Classification accuracy of all sector stock's

In Figure-5, proposed model showing 84.71 percent of accuracy compare to LR, SVM and RF gives accuracy of 80.18, 78.34 and 76.96 respectively. Finally, this graph confirms on an average hybrid learner is best predictor of price moment overall.

#### 5. CONCLUSION AND FUTURE SCOPE

Proposed study explored the effect of different news on stock movement by using the sentiment analysis. Moreover, the study has used hybrid approach in which technical as well as sentiment factors are combined to train the algorithm and predict the future movement of the stock for the short-term duration.

In expetiemnt-2, proposed model showing 84.56 percent of accuracy compare to base learner 1, 2 and 3 gives accuracy of 80.18, 78.34 and 76.96 respectively. It concludes that the proposed model giving better accuracy compared to other conventional algorithms. The proposed study concludes that use of sentiment analysis in hybrid approach where 3 base learners (LR, RF, SVM) will resulted into best accuracy for predicting signal of equity trading. The researcher has stated that news has strong relationship with short-term stock movement in the stock market. In future work, one can used overseas news to analysis the global effect on the Indian market along with technical and sentiment of the stock which is just based on the domestic news.

## **6. REFERENCES**

- [1] Malkiel, B. G. and Fama, E. F.(1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- [2] Vijn, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia Computer Science*, 167, 599-606.
- [3] Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2018). Stock price prediction using support vector regression on daily and up to the minute prices. *The Journal of finance and data science*, 4(3), 183-201.
- [4] John, J., Kumar, A., Abhishek, A., Dhule<sup>4</sup>, T. A., Roy, A., & Jha, A. (2020). STOCK MARKET PREDICTION USING MACHINE LEARNING.
- [5] Kamble, R. A. (2017, June). Short and long term stock trend prediction using decision tree. In 2017 International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1371-1375). IEEE.
- [6] Reilly, F, and Brown, K. (2012). *Investment Analysis and Portfolio Management*. 12th Edition. Cengage Learning Publication.
- [7] Wellington Garikai, Bonga. (2015). *The Need for Efficient Investment: Fundamental Analysis and Technical Analysis*. Finance & development. 10.2139/ssrn.2593315.
- [8] Shantanu Pacharkar, Pavan Kulkarni, Yash Mishra, Amol Jagadambe, S.G.Shaikh, (2018, March). Predicting Stock Market Investment Using Sentiment Analysis. *International Journal of Advanced Research in Computer and Communication Engineering* (2278-1021)"
- [9] Khatri, S. K., & Srivastava, A. (2016, September). Using sentimental analysis in prediction of stock market investment. In 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 566-569). IEEE.
- [10] Minh, D. L., Sadeghi-Niaraki, A., Huy, H. D., Min, K., & Moon, H. (2018). Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network. *IEEE Access*, 6, 55392-55404.
- [11] Paredes-Valverde, M. A., Colomo-Palacios, R., Salas-Zárate, M. D., & Valencia-García, R. (2017). Sentiment Analysis in Spanish for Improvement of Products and Services: A Deep Learning Approach. *Scientific Programming*, 2017, 1-6. doi:10.1155/2017/1329281.