



Predicting the Unpredictable: An Application of Machine Learning Algorithms in Indian Stock Market

Ashwini Saini¹ · Anoop Sharma¹

Received: 6 June 2019 / Revised: 6 August 2019 / Accepted: 9 August 2019 / Published online: 22 August 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The stock market is a popular investment option for investors because of its expected high returns. Stock market prediction is a complex task to achieve with the help of artificial intelligence. Because stock prices depend on many factors, including trends and news in the market. However, in recent years, many creative techniques and models have been proposed and applied to efficiently and accurately forecast the behaviour of the stock market. This paper presents a comparative study of fundamental and technical analysis based on different parameters. We also discuss a comparative Analysis of various prediction techniques used to predict stock price. These strategies include technical analysis like time series analysis and machine learning algorithms such as the artificial neural network (ANN). Along with them, few researchers focused on the textual analysis of stock prices by continuous analysing the public sentiments from social media and other news sources. Various approaches are compared based on methodologies, datasets, and efficiency with the help of visualisation.

Keywords Stock market prediction · Machine learning algorithms · Artificial neural network · Sentiment analysis · Long short memory neural network

1 Introduction

The stock market is a source for organizations to raise additional funds for their business by selling shares of ownership of the company. Through issuing shares and bonds in the stock market, public companies can raise idle funds to solve the problem of temporary capital scarcity. For private investors, the stock is one of the easily available financial products. Through holding suitable stocks and doing appropriate portfolio

✉ Ashwini Saini
Prof.ashwinisaini@gmail.com

Anoop Sharma
Sharmaanoop001@gmail.com

¹ Singhania University, Pachari Bari, Jhunjhunu, Rajasthan, India

management, personal assets will be appreciated to some degree. Stock trading is the leading business for securities companies and investment banks. The price movements of the shares in a market are captured in price indices called stock market indices. e.g., FTSE and the S&P. The market index is calculated by taking the average of the prices of a set of companies.

Equally weighted price Index = $\text{Sum (Prices of } N \text{ companies)}/\text{divisor}$.

The stock price prediction has been an exciting and busy research domain for a long time. New technologies are playing an important part and making prediction more exciting and appropriate. Even though many studies have been done to predict the prices of stocks, but no reasonably perfect model has been developed to predict the stock market. Nowadays, every individual or group is looking towards the stock market to gain the maximum return with minimising the risk and financial loss.

There are many things which affect the stock market, but the most important factors are classified into three different categories: fundamental factors, technical factors, and the market sentiment. Social media posts, tweets, comments, and news on the Internet regarding any organisation can impact on the stock prices. Opinion mining or sentiment analysis is used to extract the information from the sentiments, which helps the researcher to forecast the stock market to gain higher returns [1].

In the last few decades, various tools and technologies are used to get the mood of the market before any movement in the stock prices. Investment banks evade funds, and insurance companies spend vital time and effort in predicting changes in the stocks. Facebook and Twitter are used as social media platforms around the globe, which are used for sentiment analysis. Sentiment Analysis is a process of finding the mood of a sentence. The mood can be positive, negative, or neutral. It is used to derive the opinion or sentiment of the speaker. Generally, it is used to determine how a person feels about a particular subject or topic.

The efficient market hypothesis (EMH) advocates that social media sentiments help to understand the stock market better. It exists in three forms [2]. Efficient Market Hypothesis (EMH) is a speculation theory that states it is difficult to “beat the market” in since stock market proficiency causes the current stock costs to consistently fuse and reflect all applicable data present in the market.

- *Weak EMH* It has covered only the past data
- *Semi-Strong EMH* In this, All public information is utilized.
- *Strong EMH* Publicly and privately available information is used in strong EMH.

Random walk theory—It Assumes that it is not possible to predict stock prices as stock prices don’t depend on past stock. It also considers that stock prices have high fluctuation, so it is not possible to predict the prices in the stock market.

2 Related Work

Chen and Lazer [3] have researched the connection between the movement of the stock market and twitter feeds. They used both classification and regression strategy

to predict the movement of the stock market. After investigation, they identified several features affecting the sentiment values. Zhang et al. [4] have discovered the relationship between Twitter sentiment and stock prices. They analysed the aftereffect of three machine learning systems and the viability of the methods on the correlation. Adebisi Ayodele et al. [5] has given a hybridized approach for prediction by combining fundamental and technical analysis.

Bollen et al. [6] have also analysed the behaviour of the stock market on the basis of the mood of the people on twitter. Twitter text analysed for a specific period for finding the changes in mood. They found that the accuracy of the Dow Jones Industrial Average (DJIA) predictions improved using specific public mood.

Rechenthin et al. [7] show the ability to predict daily price direction by combining the sentiment of a favourite stock message board with historical price movements. Various authors used a variety of supervised learning algorithms to collect sentiment information and used artificial neural networks (ANN) to determine that the markets were predictable.

3 Stock Market Prediction Approaches

The stock market prediction has two conventional approaches. These are Fundamental and Technical analysis.

3.1 Fundamental Analysis

The performance and profitability of an organization can be analysed to measure its intrinsic value. The intrinsic value can be determined by understanding and styling physically by way of its sales, Staffing quality, infrastructure, and return on investment. Fundamental analysis uses various factors like sales revenues, returns on equity, profit margins, and future growth. It is the right approach for long term investment and growth. This approach is advantageous because of its systematic approach and its ability to predict changes early before showing on the charts.

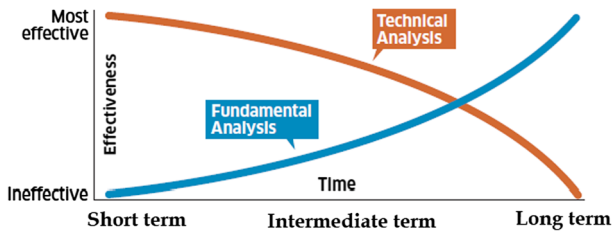
3.2 Technical Analysis

Technical analysis is one another method used to evaluate stocks by analysing statistics generated by market activity, past prices, and volume. Technical Analysis covers significant activities related to stock price movement like trends, patterns, peak price, lowest price, etc. [8]. Forecasting of any stock price often depends on past behaviour of the stock and its correlated variables. It finds the patterns and indicators on stock charts to determine the future performance of the stock. However, sometimes, this approach is criticized because of its highly subjective nature (Table 1).

Fundamental and Technical analysis are compared in the above table, results show that the technical analysis approaches are most effective for short to medium term trades, whereas Fundamental analysis approach is better to make long term investments (Fig. 1).

Table 1 Analysis between technical and fundamental approaches based on different parameters [9, 10]

Parameters	Technical analysis	Fundamental analysis
Description	Prices are predicted by observing the price movements of a security	Prices are predicted by analysing the economic and financial factors affecting business
Tools	In technical analysis, analysis starts with stock charts, as all relevant information is included in stock prices	In fundamental analysis, analysis starts with the company's financial statements like income, balance sheet, and cash flow statement
Time effectiveness	Short term approach	Long term approach
Trading versus Investing	Identify short to intermediate term trades	Make long term investments

**Fig. 1** Stock market prediction using machine learning algorithms

4 Machine Learning Algorithms Used

4.1 Regression Analysis

Regression Analysis is used for stock market prediction by using non-linear methods. Regression is used in some situations where the relationship between the outcome and the predictor variables is not linear. It is based on analyzing the market variables, and afterward, these equations are utilized as the predictive model to the adjustments in the number of variables and to predict the dependent variable relationships during the forecast period.

4.2 Hidden Markov Model

This model is also one of the models used for predicting stock market future. It analyses the hidden state variables to predict future output and state variables.

4.3 Artificial Neural Network

ANN is a data-driven model. It is a network of computational entities (neurons) that uses parallel distributed computing to predict the prices. This ANN Model can be trained using correct data to be applicable in fields such as regression and classification

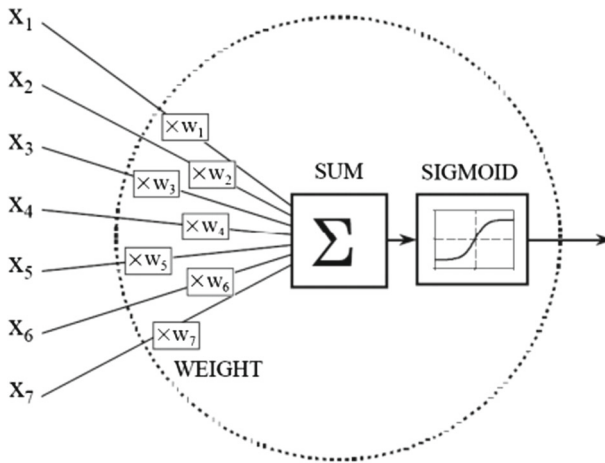


Fig. 2 Seven-input one-output ANN representation (Smith 1997) [12]

[11]. Neural networks can tackle an issue without earlier learning of the connection amongst input and output, so these are also called as self-adjusting methods. A generic ANN with input, processing, and output units is showing in the below figure (Fig. 2).

4.4 Naïve Bayesian Classifier

Naïve Bayes is a classification technique which generates Bayesian Networks for a given dataset based on Bayes theorem. Naïve Bayesian classifier is based on a supervised learning algorithm. A supervised learning algorithm is a kind of machine learning in which supervision in learning comes from labelled examples in the training dataset. [13] It is also known as learning with the help of a teacher just because here, class labels are already defined. The naïve Bayesian classifier approach is one of the best techniques for data classification. Classification is a two-step process, the very first step is learning, where the training set is analyzed for model construction, whereas another one is classification step, in which class labels are predicted for the dataset according to the model used.

4.5 Decision Tree Classifier

Decision trees are the easiest way to perform the classification. In this classifier, the class label is denoted by terminal nodes, internal nodes are denoted by the test on an attribute, and the desired outcome of the test is represented by branches of a tree. These classifiers are very popular because they can handle multidimensional data and doesn't require any domain knowledge.

Table 2 Comparative analysis between technical analysis, fundamental analysis, time series analysis, and machine learning techniques on different criteria

Criteria	Technical analysis	Fundamental analysis	Traditional time series analysis	Machine learning techniques
Data used	Price, volume, highest, lowest prices	Growth, dividend payment, sales level, interest rates, tax rates, etc	Historical data	Set of sample data
Learning methods	Extraction of trading rules from charts	Simple trading rules extraction	Regression analysis on attributes is used	Inductive learning is used
Type of tools	Charts are used	Trading rules	Simple regression and multivariate analysis used for time series	Nearest neighbor and neural networks are used
Implementation	Daily basis prediction	Long-term basis prediction	Long-term basis prediction	Daily basis prediction

4.6 Random Forest

Random forest algorithm is a supervised machine learning algorithm. It is very flexible and easy to use because it can be used for both regression and classification task. The forest it builds, is an ensemble of Decision Trees, most of the time trained with the bagging method. In this algorithm, there is no requirement for the pruning of trees that are not sensitive to outliers in training data.

5 Stock Market Analysis

After considering all the algorithms and methods involved in the prediction of stock prices, the efficiency and accuracy of prediction are analysed. We found that the Short-Term Memory (LSTM) Neural network always produce better results as compared to other methods like K-Nearest Neighbor (KNN), Naive Bayesian Classifier, Regression and Support Vector Machines (SVM).

5.1 Comparative Analysis of Various Prediction Techniques

In the above table, we compared different analysis methods used to predict the stock market. These methods are compared to different criteria like dataset, training methods, tools, and algorithm used. Various authors proposed models based on different algorithms are also analysed with dataset and accuracy level (Tables 2, 3 and Figs. 3, 4).

Table 3 Proposed models by authors with respective datasets and accuracy results obtained

Proposed models	Referenced by	Dataset used	Accuracy (%)
Convolutional neural network (CNN)	Gudelek et al. [14]	Google finance 14	72
Based on time series analysis and improved text mining techniques	Dang and Duong [15]	vietstock.vn, hsx.vn, hsn.vn	73
WB-NN	Ding et al. [16]	Standard and poor's 500 stock (S and P 500) index	60.25
WB-CNN			61.73
E-CNN			61.45
EB-NN			62.84
EB-CNN			65.08
SVM	Zhang et al. [17]	China A-share	55.3
PCA+SVM			57.5
CMT-Z-X			59.03
TeSIA			60.6
CMT-Z			60.25
Long short-term memory (LSTM) neural network	Li et al. [18]	CS1300 index from wind database	87.86
ANN, MLP, feed forward, back propagation	Yunus et al. [19]	NAS DAQ stock market index	81.0
Single layer perceptron, LP, RBF and SVM	Mehak et al. [20]	KSE-100 index	77.0

ACC Accuracy, SVM support vector machines, PCA principal component analysis, TeSIA tensor-based learning approach, CMT two auxiliary matrices and a tensor are factorized together, X, Z auxiliary matrices, WB-NN word embedding input and standard neural network, WB-CNN word embedding input and convolutional neural network, E-CNN structured events tuple input and convolutional neural network, EB-NN event embedding input and standard neural network, EB-CNN event embedding input and convolutional neural network

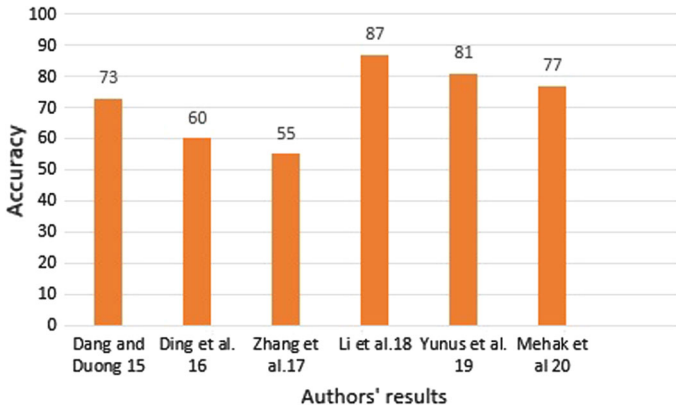


Fig. 3 Authors' results and accuracy graph

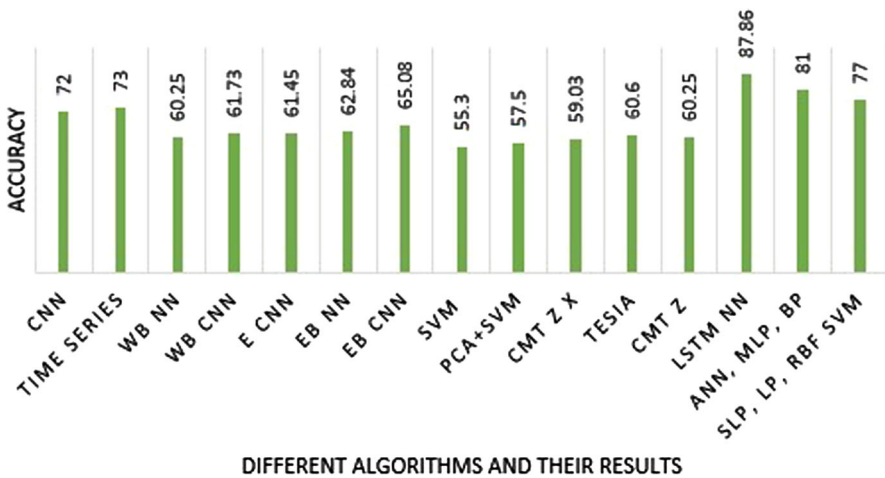


Fig. 4 Algorithms and accuracy results graph

6 Conclusion

The prediction of the stock market is a complex and challenging task. Many authors suggested different prediction approaches for the stock market like Non-linear regression, Naïve Bayes Classifier, Random Forest Method, Artificial Neural Networks and support vector Machine. We study fundamental and technical analysis and compare them to get the results. A comparative analysis of various different algorithms used for predicting future stock market prices and found that Long short Memory Neural network (LSTM NN) producing better results as compared to other techniques. In the future, this research would help to improve the efficiency and accuracy of prediction. Deep learning classifiers would also be analysed in the future to predict and gain the maximum benefit from the stock market.

References

1. Liu B (2012) Sentiment analysis and opinion mining. Morgan and Claypool Publishers, San Rafael
2. Attigeri GV, Manohara PMM, Pai RM, Nayak A (2015) Stock market prediction: a big data approach. In: Proceedings of the IEEE region 10th conference on TENCON 2015, Macao, China, pp 1–5
3. Chen R, Lazer M (2013) Sentiment analysis of twitter feeds for the prediction of stock market movement, Stanford. edu. Retrieved January. 2013
4. Zhang L (2013) Sentiment analysis on twitter with stock price and significant keyword correlation. The University of Texas at Austin, Austin
5. Adebisi Ayodele A, Charles AK, Adebisi MO, Otokiti SO (2012) Stock price prediction using neural network with hybridized market indicators. *J Emerg Trends Comput Inf Sci* 3(1):1–9
6. Bollen J, Mao J, Zeng X (2010) Twitter mood predicts the stock market, *CoRR*. vol. abs/1010.3003
7. Rechenthin M, Street WN, Srinivasan P (2013) Stock chatter: using stock sentiment to predict price direction. *Algorithm Finance* 2(3–4):169–196
8. Zabir HK, Tasnim SA, Md. Hussain MA (2011) Price prediction of share market using artificial neural network (ANN). *Int J Comput Appl* 22(2):42–47
9. Setty DV, Rangaswamy TM, Subramanya KN (2010) A review on data mining applications to the performance of stock marketing. *Int J Comput Appl* 1:33–43
10. Rechenthin MD (2014) Machine-learning classification techniques for the analysis and prediction of high-frequency stock direction. Ph.D. thesis, University of Iowa, USA
11. Buche A, Chandak MB (2016) Stock market prediction using text opinion mining: a survey. *Int J Adv Res Comput Sci Softw Eng* 6:566–569
12. Twomey JM, Smith AE (1997) Validation and verification. In: Kartam N, Flood I, Garrett JH (eds) *Artificial neural networks for civil engineers: fundamentals and applications*. ASCE, New York, pp 44–64
13. Cheng CH, Chen YS (2007) Fundamental analysis of stock trading systems using classification techniques. In: International conference on machine learning and cybernetics, vol 3, Hong Kong, China, pp 1377–1382
14. Gudelek, MU, Boluk SA, Ozbayoglu AM (2017) A deep learning based stock trading model with 2-D CNN trend detection. In: *IEEE (SSCI)*, Honolulu, HI, USA, pp 1–8
15. Dang M, Duong D (2016) Improvement methods for stock market prediction using financial news articles. In: Proceedings of 3rd national foundation for science and technology development conference on information and computer science (NICS), Danang, Vietnam, pp 125–129
16. Ding X, Zhang Y, Liu T, Duan J (2015) Deep learning for event-driven stock prediction. In: International joint conference on artificial intelligence, AAAI Press, pp 2327–2333
17. Zhang X, Zhang Y, Wang S, Yao Y, Fang B, Philip SY (2018) Improving stock market prediction via heterogeneous information fusion. *Knowl-Based Syst* 143:236–247
18. Li JH, Bu, Wu J (2017) Sentiment-aware stock market prediction: a deep learning method. In: International conference on service systems and service management (ICSSSM' 17), IEEE, Dalian, China, pp 1–6
19. Yunus Y, Halid K, Jamshidi M (2014) Stock market prediction by using artificial neural network. In: World automation congress (WAC), pp 718–722
20. Mehak U, Syed HA, Kamran R, Syed SAA (2016) Stock market prediction using machine learning techniques. In: International conference on computer and information sciences (ICCOINS), pp 322–327

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.