

# A Survey on Stock Market Analysis using Supervised Machine Learning

Raju Bilwal, Rajeev Gupta

Department of CSE, SISTEC, RGPV, Bhopal, India

**Abstract**— The Machine Learning algorithm handles previously logged data as training samples and makes supports for forecasting the stock price for future trends. Supervised learning is a powerful machine learning task. The basic idea of supervised learning is to classify and process data using Machine Learning. The stock market is widely used in investment schemes promising high returns but has some risk. Stock returns are very fluctuating in nature. They depend upon factors like previous stock prices, current market trends, financial news, social media etc. Many practices, like technical analysis, fundamental analysis, time series analysis, statistical analysis, etc., are used to predict the stock value. Still, none of these procedures is proven as an allowable prediction tool. The technical and fundamental or time series analysis is used by most stockbrokers while making stock predictions. Python is the programming language used to predict the stock market using machine learning. In this paper, we propose a machine learning (ML) approach that will be trained from the available stock data and gain intelligence and then uses the acquired knowledge for an accurate prediction.

**KEYWORDS:** Machine learning, Supervised machine learning, finance, stock price prediction, sentiment analysis. Prediction, Classification, Optimisation, Stock Market, and Data Analysis.

## I. Introduction

Stock investment has been gradually growing popular lately. The increase in stock investors is propelled by the numerous stock prediction techniques formulated today. Most of these techniques use time series data. Utilising these techniques mainly has one objective: maximising profit and minimising loss. A stock market or equity market is a public entity (a loose network of economic transactions, not a physical facility or discrete entity) for trading company stock (shares) and derivatives at an agreed price. These are securities listed on a stock exchange and those only traded privately. The stocks are listed and traded on stock exchanges which are entities of a corporation or mutual organisations specialised in bringing buyers and sellers of the organisations together. Most of the trading in the Indian stock market takes place on its two stock exchanges: the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE). The BSE has been in existence since 1875. The NSE, on the other hand, was founded in 1992 and started trading in 1994. However, both exchanges follow the same trading mechanism, trading hours, settlement process, etc.

Financial market forecasting is based on certain principles, theories and models to study the financial markets and predict their future trends or course. Changes in stock prices largely depend on human opinions and expectations about a stock's future performance or share [2]. A Systematic approach for forecasting the stock market comprised the following operational steps: selection of input variables, data preprocessing, feature selection and extraction, training using prediction/classification model and evaluating proposed model performance. The first step in the stock market forecasting process consists of selecting input features modelled by CI methods and output to be predicted. In stock market forecasting, various fundamental and technical input variables are available. The selection of input variables is the main issue in stock market forecasting, and deciding which input variables to use is difficult. The next step in the process is preprocessing of data selected in the first step. Data is preprocessed to increase the models' prediction capability [3,4].

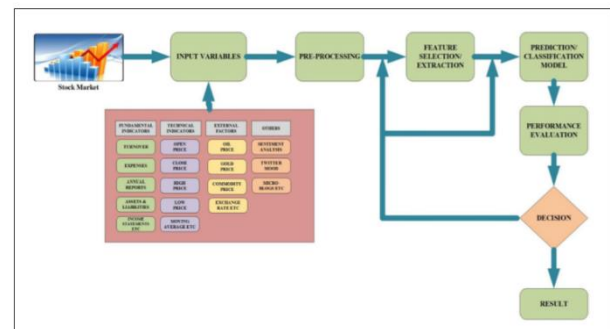


Fig1. A generalised framework for stock market forecasting

The preprocessing mechanism can be used for noise removal, outlier detection, missing value and normalising data. In the third phase, various feature selection or extraction techniques are applied to get the best representative input data variables to reduce the input data's dimensions and the model's computational complexity. The next phase involves determining the CI methods for forecasting and training the model using training data. The last phase of performance evaluation comprises selecting appropriate performance metrics, measuring the model's accuracy, and taking the right decision. Each selected paper has been discussed in this article according to a generalised framework for stock market forecasting [5].

**II. Related Work**

Many prediction models have been proposed to date to forecast stock prices and stock market trends. Some of the machine-learning techniques have been discussed in this paper. Table 1 displays a summary of all the techniques proposed by various researchers. All the techniques have been classified into subcategories: classification, regression, ensemble algorithms, evolutionary, deep learning, hybrid models, and other additional techniques. In 2016, Shobana et al. intricacy while analysing market trends depends on several external factors, some of which are not under one's control. This work analyses stock market trends using machine learning and nature-inspired techniques. These were first studied and then implemented (a few used in this paper are Decision Tree, PSO, Black-Hole, and Naïve Bayes).

After analysing the trends with the help of standard techniques, we proposed an entirely new approach to analyse stock market indices over which accuracy is calculated and compared over different techniques and algorithms. We outline the design of the proposed model with its salient features and customisable parameters. We finally tested our model on the one-year Nifty stock index dataset in real time, where we analysed the values based on data from the past days for three months. In 2012, Preethi et al. that the price of the stocks is an important indicator for a company, and many factors can affect their values. Different events may affect public sentiments and emotions, affecting the stock market price trend. Because of dependency on various factors, the stock prices are not static but are instead dynamic, highly noisy and nonlinear time series data. Due to its great learning capability for solving nonlinear time series prediction problems, machine learning has been applied to this research area. Learning-based methods for stock price prediction are very popular, and many enhanced strategies have been used to improve the performance of the learning-based predictors. However, performing successful stock market predictions is still a challenge. News articles and social media data are also useful and important in financial prediction. Still, no good method exists to consider these social media to provide better financial market analysis. In 2016, Dang et al. gradually utilised the web to become the carrier of information in modern society.

Along with the rise and popularity of the web, the text initialised by common users is becoming increasingly popular. By analysing and data-mining this information, dynamic analysis can identify the emotional tendency- for or against, sometimes even the evolutionary rules of emotion. However, current information search technology takes common search engines as an example based on keywords and cannot support our needs. Emotions and opinions cannot be represented or sorted by keywords. General emotion analysis includes the speaker's attitude, while narrow sense emotion analysis identifies the speaker's opinions and classifies the opinion as for or against [5]. Text-based emotion analysis includes natural language dealing, information searching,

data mining and artificial intelligence. As is shown in figure 1, typical emotion analysis includes data collecting, data pretreatment, feature extraction and classification.

In 2019, Sable et al. made an accurate stock price prediction, and we must consider various factors affecting the stock price. A reasonable and effective judgment can be made only by fully understanding these factors' change trends and effects. At this stage, nonlinear methods are usually used to predict the stock price to deal with the randomness and frequent fluctuation of the price, which is influenced by the superposition of multiple factors. And the most common consideration is the use of nonlinear regression analysis. Regression analysis is a kind of analysis method which can confirm the interdependent statistical quantitative relationship between two or more two kinds of the variable. It can analyse the inherent law of the data, and it is a very effective method. However, considering the existing situation of the stock market, regression analysis is not suitable. As we all know, the impact factors of the stock price are very complex, involving the internal factors of the stock market, internal factors of the company, and macroeconomic factors. In 2019, Kunal Pahwa et al. described the stock or share market as one of the most complicated and sophisticated business methods. Small ownerships, brokerage corporations, and the banking sector depend on this very body to make revenue and divide risks, a very complicated model. However, this paper proposes using a machine learning algorithm to predict the future stock price for exchange using open-source libraries and preexisting algorithms to help make this unpredictable business format slightly more predictable. We shall see how this simple implementation will bring acceptable results. The outcome is completely based on numbers and assumes a lot of hypotheses that may or may not follow in the real world, so at the time of prediction.

In 2020, Parray et al. tried to compare the Machine Learning algorithms such as SVM, Perceptron, and Logistic regression for good prediction. Here, the author expressed that by using historical data of the stocks, Machine Learning Algorithms are trained to predict future data. In 2013, Kumar et al. pointed out that various statistical tools handled forecasting financial analysis in earlier days. But nowadays, we have many Neural Network models to find the exact solution to financial data. ANN is combined with a statistical model is given a better result. In 2016, Sreemalli et al. proposed the work of an artificial neural network is a very approved technique to predict the stock market price and support vector machines. They are using these models to list the pros and cons of all these models and compare the execution of the stock market.

Artificial Neural Network (ANN) has great potential for machine learning problems such as classification and prediction. Using the linear model to execute the nonlinear class partitions has occurred through some nonlinear mapping in which the input vector is fed into the high-dimensional feature space. ARIMA model deals

with time series data. This paper uses machine learning techniques like Support Vector Machine, Artificial Neural network and Auto-Regressive Integrated moving average to predict Nifty bank data. The dataset used here is the 2015 Nifty bank dataset. Implementing a neural network requires more time to perform computations than other techniques, whereas support vector machine has more error rate. Combining artificial neural networks and Genetic algorithms might result in high accuracy. In 2017, Indu Kumar et al. proposed machine learning techniques that have been applied to stock price prediction to overcome such difficulties. In this paper, five models have been developed, and their execution is compared in forecasting the stock market trends. The five supervised learning techniques are Support Vector Machine (SVM), Random Forest, K-Nearest Neighbor (KNN), Naive Bayes, and SoftMax. The probing results show that the Random Forest algorithm acts best for larger datasets.

Naïve Bayesian Classifier is the best for smaller datasets. The proposed architecture for the implemented work mainly consists of four steps: feature extraction for the given dataset, supervised classification of the training dataset, supervised classification of the test dataset, and result evaluation. In 2013, Venkata Sasank et al. proposed the work to observe how well the changes in stock prices of a company, the rises and falls, are correlated with the public opinions being revealed on Twitter about that company. Appreciating the author's judgement from a text is sentiment analysis's purpose. In this paper, they have applied sentiment analysis and supervised machine learning principles to the tweets pulled out from Twitter and examined the relationship between the stock market movements of a company and sentiments on Twitter. Straightforwardly, positive news and tweets on social media about a company would uplift people to invest in the stocks of that company. As a result, the stock price of that company would expand. At the end of the survey, it is shown that a strong association exists between the rise and fall in stock prices and the public sentiments in tweets.

### III. CONCLUSION

Machine learning, as they have seen till now, is a very powerful tool and, as evitable, it has some great applications. They have seen till now that machine learning is very much dependent upon data. Thus, it is important to understand that data is invaluable, and as simple as it may sound, data analysis is not easy. This paper shows the elementary awareness and ideologies of supervised machine learning. It also shows the basic algorithm of supervised machine learning, including RF, Knn, SVM, MLR, etc. From this survey, they conclude that various supervised machine learning algorithm produces more optimised results. Finally, they describe a comparative study of current supervised machine learning methods and future patterns. Stock investments have been of interest to many investors around the world. However, making decisions is difficult and complex as

numerous factors are involved. For successful investment, investors are keen to forecast the future situation of the stock market. Even small improvements in predictive efficiency can be very profitable. A good prediction system will help investors make investments more accurate and profitable by providing supportive information such as the future direction of stock prices. The python tool proposes method analysis, making stock market analysis more accurate and profitable.

### REFERENCES

- [1]. Soni, D., Agarwal, S., Agarwal, T., Arora, P., & Gupta, K., Optimised Prediction Model for Stock Market Trend Analysis. 2018 Eleventh International Conference on Contemporary Computing (IC3),2018.
- [2]. Tang, X., Yang, C., & Zhou, J., Stock Price Forecasting by Combining News Mining and Time Series Analysis. 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology,2009.
- [3]. Lauren, S., & Harlili, S. D, Stock trend prediction using simple moving average supported by news classification. 2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA),2014.
- [4]. Soni, D., Agarwal, S., Agarwal, T., Arora, P., & Gupta, K., Optimised Prediction Model for Stock Market Trend Analysis. 2018 Eleventh International Conference on Contemporary Computing (IC3),2018.
- [5]. Wang, Z., Ho, S.-B., & Lin, Z., Stock Market Prediction Analysis by Incorporating Social and News Opinion and Sentiment. 2018 IEEE International Conference on Data Mining Workshops (ICDMW). ,2018.
- [6]. Shobana, T. & Umamakeswari, Arumugam, A Review on Prediction of Stock Market using Various Methods in the Field of Data Mining. Indian Journal of Science and Technology, 2016.
- [7]. Preethi, G. & Santhi, B., "Stock market forecasting techniques: A survey", Journal of Theoretical and Applied Information Technology. 46. 24-30,2012.
- [8]. Dang, L. Minh & Duong, Duc. (2016). Improvement methods for stock market prediction using financial news articles. 125-129. 2016.
- [9]. Sable, Rachna & Goel, Shivani & Chatterjee, Pradeep, Empirical Study on Stock Market Prediction Using Machine Learning. 1-5.2019.
- [10]. Kunal Pahwa, Neha Agarwal, "Stock Market Analysis using Supervised Machine Learning", 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 14th -16th Feb 2019.

- [11]. I. R. Parray, S. S. Khurana, M. Kumar, and A. A. Altalbe, "Time series data analysis of stock price movement using machine learning techniques," *Soft Comput*, Apr. 2020,
- [12]. D. A. Kumar and S. Murugan, "Performance analysis of Indian stock market index using neural network time series model," in *2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*, Salem, pp. 72–78, Feb. 2013.
- [13]. Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements", 2016.
- [14]. Mariam Moukalled, Wassim El-Hajj, Mohamad Jaber "Automated Stock Price Prediction Using Machine Learning", 2017.
- [15]. Osman Hegazy, Omar S, Soliman and Mustafa Abdul Salam" A Machine Learning Model for Stock Market Prediction, 2013.