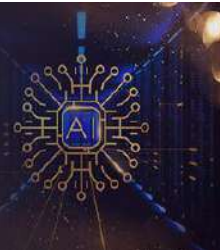


International Journal of Computing and Artificial Intelligence



E-ISSN: 2707-658X
P-ISSN: 2707-6571
IJCAI 2022; 3(2): 13-18
Received: 07-05-2022
Accepted: 10-06-2022

Manju Dhull
Department of computer
science and Engineering",
UIET, MDU Rohtak,
Haryana, India

Machine learning techniques and methodology analysis for stock market price prediction

Manju Dhull

DOI: <https://doi.org/10.33545/27076571.2022.v3.i2a.50>

Abstract

Prediction of market movements is a very emerging topic in research nowadays. The one handsome efficient hypothesis says that accurate stock prediction is impossible but, in another hand, if appropriate techniques and algorithms can be used then it can be predicted with high accuracy. Moreover, predicting profitable and accurate value is a challenging task for investors because price prediction is volatile in nature. It could be affected by various factors like the global economy, politics, disasters, etc. A lot of literature survey on technical analysis is available which identify the movement patterns of the market. Various hybrid machine learning approaches in combination have been used to forecast the beginning values of the market trends and also for the future long- term values.

This paper gives a comparative assessment of 40 research papers that recommend strategies such as calculating methodologies, stock prediction algorithms, Datasets, results, and more which leads to improved accuracy and lowers the error percentage of stock market prediction.

Keywords: ARIMA, CNN, stock value prediction, machine learning, support vectors machine (SVM)

Introduction

The stock market plays an important role in the financial growth of any country; it affects the country's growth economically and has a significant impact on a variety of professions, employment, industry, and so on. According to a survey, only 12% of people in each country dare to spend in the stock market ^[1]. Because of its complexity, investing in the stock market is extremely tough. Forecasting stock movements is complex due to volatile, noisy, and non-parametric data in nature it is also affected by political issues Financial and economic crises, as well as a slew of other causes ^[2], have all had an impact. Stock prices fluctuate due to supply and demand. When there are more shares available, the price of a stock decrease, and if more buyers are available to purchase then the price of the stock will rise. Patel, R. *et al.* ^[3]. It is an area where all types of people want to gain profit by investing in the market either trader, market participants, etc. Profit can be earned by the exact knowledge of market movements for the future and it is a very complicated task for the investors ^[4, 5]. So to make it easy for all kinds of people many data analysts and engineers are already working on various machine learning algorithms to develop software for prediction.

Stock prediction can be done in two ways: Technical and fundamental analysis ^[6], the technical part of analysis states the study of previous stock values to predict future movements. Basically, technical analysis helps in short-term prediction. Fundamental analysis frequently deals with unstructured data like financial data news and economic statistics or long-term prediction.

Time series analysis is the major tool for prediction. A few models are types of time series analysis ^[7, 8]. Such as ARIMA and ARCH are common linear models ^[9] but these cannot predict accurate data values and sometimes fails to predict large data values.

In the financial sector, machine learning (ML) implements its various approaches ^[10] to give a new mechanism that can help investors in making better investments and right decisions to improve the performance of the market ^[11]. Equity securities are one of the most popular fund-generating terms in the growth of the market. Equity security is a kind of liquid asset which helps in the wealth creation of any country. But it is a very challenging task which is why only 10% of people may get success. Wenjie Lu. *et al.* ^[12].

The p goal of this paper is to analyze various methodologies and tools already used in predicting the stock market trends. Section II of this paper contains a literature review in a similar area. The Section III. Contains a detailed introduction to various machine learning algorithms.

Corresponding Author:
Manju Dhull
Department of computer
science and Engineering",
UIET, MDU Rohtak,
Haryana, India

After that next IV section explains how to extract features and what should be the complete approach to stock prediction, the last section V discusses the challenges faced during the research work by researchers and concludes the paper, along with the future advancement required [13].

Literature survey

In recent years, a lot of researchers and practitioners proposed numerous machine learning and deep learning approaches and strategies to achieve accurate results and lower the risk in stock future value forecasting.

Idrees, S. M., Alam, M. A *et al.* [14] proposed an ARIMA approach to improve stock market price for next-day prediction. In this paper real-life data sets were used such as Nifty and Sensex. Results of the given model were compared with the existing models and found a 5% mean deviation overall.

Cheng-Lung Huang *et al.* [15] proposed a hybrid approach of to improve the accuracy and reduce the market risk. This hybrid model takes real dataset of Taiwan exchange and predicts the next day's price more accurately.

Kara, Y., Acar Boyacioglu, M *et al.* [16] attempted to predict the price movements using SVM support vector machine and ANN algorithms on datasets of the Istanbul stock exchange where 10 indicators were selected for input. This model shows that the performance of the ANN algorithm is more accurate as compared to SVM.

Ritika Singh1 *et al.* [18] Proposed a hybrid approach of two models Deep neural networks and PCA. It is a direct 2-dimensional principal approach that focuses on radial basis function Sheng Chen1 *et al.* [19] proposed a CNN algorithm to forecast the stock price movement of the Chinese stock market. And found that the CNN model is reliable and can use to make an errorless prediction. But it is tested only on the values of close, open, high, and low prices. Other than these features more data values cannot be considered.

Argyrios P. Ketsetsis1 *et al.* [20] conducted a survey on primary studies of the European union stock market which deal with deep learning techniques for prediction and gives accurate result. It stated that this paper involves 12 papers analysis and still not found intensive results so it is open research for the future.

Deepak Kumar *et al.* [20] proposed an understanding view of 30 research papers that recommend various methods like statics calculation, algorithms for performance measurement, and outstanding journals these studies are chosen in response to answering the open questions in research. As a result, these selected studies are assisting in the discovery of ML techniques and datasets for stock market prediction. It also found that the hybrid approach is suitable for stock prediction.

Jimmy Ming-Tai Wu1 *et al.* [21] proposed a new architecture to achieve more reliable stock market prediction, with a combination.

Aditya Bhardwaj *et al.* [52] proposed new algorithms for predicting the Indian stock market situation for Sensex and Nifty. This allows people, and investors to do a comparative study of stock market situations and can analyze the data properly by verifying the actual datasets. This paper also examines stock market sentiment by collecting data from the Sensex and Nifty live servers at different time intervals that can be used to predict accurate stock values.

Rakhi Batra *et al.* [23] proposed a new approach of SVM models which was used for stock marketing because it can handle for text classification. In this paper, the authors employ two datasets: tweets about Yahoo Finance. The suggested work yielded output with 75.22 percent training accuracy and 76.68 percent test accuracy. The authors believe that increasing the size of the dataset will enhance the performance of the text categorization test and its accuracy.

Idrees, S. M., Alam, M. A *et al.* [24] proposed an ARIMA approach to improve stock market price for next-day prediction. In this paper, real-life data sets were used such as Nifty and Sensex. Results of the given model were compared with the existing models and found a 5% mean deviation overall.

Rouf, N., Malik *et al.* [25] proposed an ANN model based on RMSE matric evaluation and predict values 93% accurately. In this paper NASDAQ datasets were used but the author believes that if input data is vague in this case results cannot be reliable. Machine learning algorithms that can be used in Stock value prediction

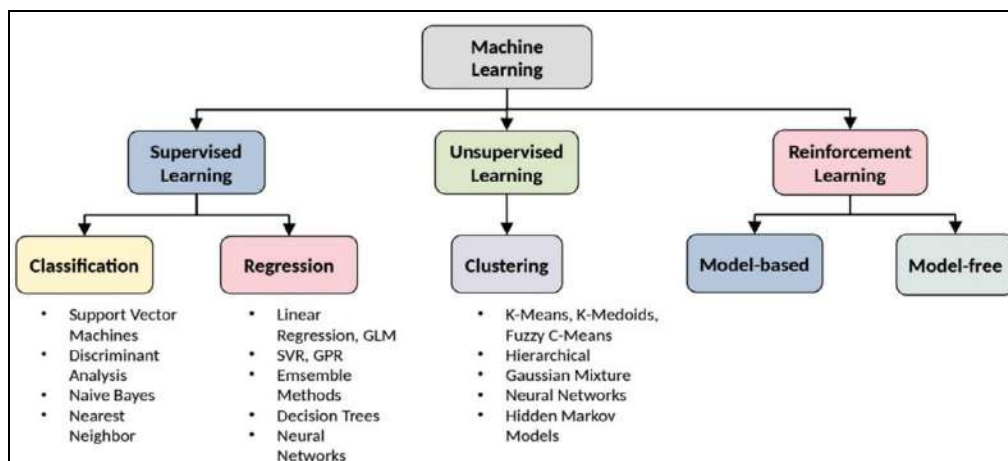


Fig 1: Categorization of machine learning algorithms [25].

SVM: The support vector machine algorithm is one of the more efficient machine learning algorithms which can be used for forecasting time series [26, 27]. It enables to plot of raw data on an n-dimensional graph. Basically, SVM is used

for classification and regression tasks. Here, stock market metrics are used to plot multidimensional coordinate planes [28]. This algorithm is the most potent and efficient financial instrument that can be used in stock market forecasting.

NN: A series of algorithms used to understand the functioning of the human brain and identify changing datasets over a time period. To improve stock market forecasting [29], created the outer break NN technique for which information is gathered immediately from the current stock market. Using this methodology, the deep long short-term -dependent NN was developed for the embedded layer's use [30].

Artificial neural network: ANN is A learning that can give accurate results and captures the stock movements more effectively in comparison to other statistical techniques. Most of the researchers applied this model before pre-processing the data for the characterization of data. For stock market prediction numerous performance measurement parameters can be used. Patel, R [31].

Convolutional NN (CNN): It is a machine learning algorithm mainly used in classification and pattern recognition, time series, etc. CNNs do not require manual feature extraction because they learn the features on their own [32]. The results of categorization using CNNs are quite precise. You may expand on pre-existing networks by training CNNs for new classification techniques. It has no hidden layers as compared to traditional algorithms [33].

Recurrent NN (RNN): Recurrent Neural Network (RNN) are a kind of machine learning algorithm where the result from the previous step is supplied as input to the next step. It works under complex behavior but gives better results as compared to traditional models [34, 35].

The SVR [36] is a subtype of SVM with a small difference between them. The SVR is used in market price prediction but SVM works on time series-based datasets.

GAN (Generative adversarial network): it is a technique used for generative modeling that makes use of deep learning techniques like convolutional neural networks [37]. The GAN is a smart technique of training models by changing the problem as a supervised learning problem with two subtypes: the generator model, which is used to retrain

new models, and the discriminator model that tries to identify the problem as either real (from the domain) or fake generated [38].

Naïve Bayes (NB): The NB is a machine learning algorithm used to classify meshy datasets that creates the Bayesian networks based on probability for a specific Bayes-based dataset theorem [39, 40]. It works on independent features and does not work with dependent variables. This algorithm creates independent results.

Logistics Regression: This algorithm estimates discrete values from the datasets and also works on the probability of completion of any task [41]. Mostly it is used to eliminate extra features from the dataset in preprocessing of data.

Stock market prediction approach

Pre-Processing and Feature Selection: As the following flow diagram figure 2 shows the complete structure of price prediction. Initially, past stock data is collected from real datasets such as the BSE dataset [46], and then Data pre-processing is used to clean out after they have been gathered.

The pre-processing step is performed in machine learning to remove repetitious, obsolete, and poorly structured input. After removing all unnecessary data, and noise from the dataset useful information was extracted. As a consequence, data pre-processing is performed to clean the dataset [45].

In the next phase, some techniques are used for feature selection after cleaning data is again a challenging task to find an appropriate technique for feature selection [47]. This is the most important phase of stock prediction. Feature selection is basically used to lower the cost of modeling, and also used to reduce of the variables from the dataset. The number of input variables can be reduced to lower the computational expense of modeling in certain situations.[48] After completion of the classification of data finally trends analysis was done using various indicators and techniques such as ARIMA etc [49].

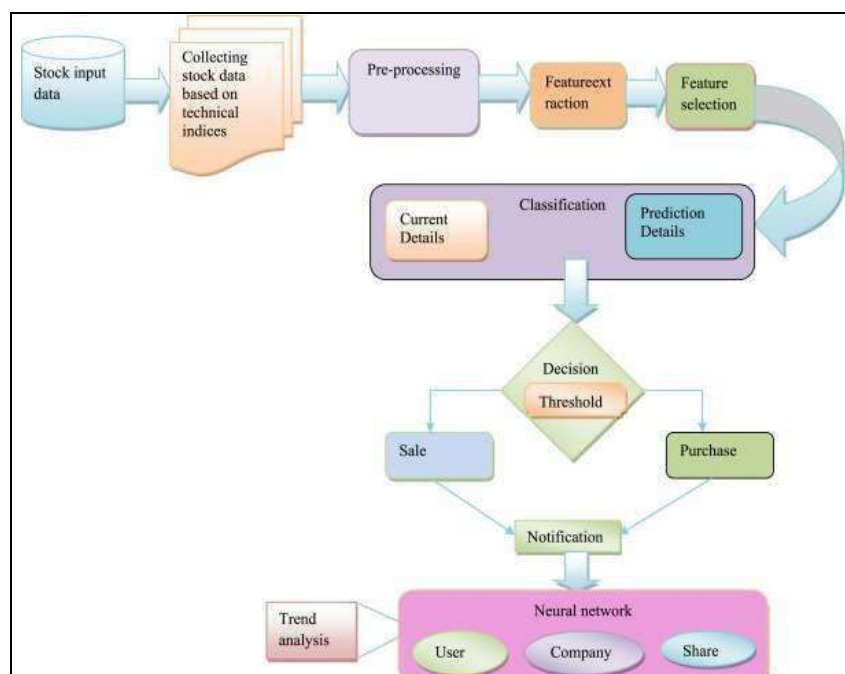


Fig 2: Flow diagram of the stock market process system. Kumar, D *et al.* [52]

Challenges

Stock Prediction is a continuous process

Due to the daily changes in the stock market's movements predicted values vary accordingly. To overcome this obstacle, we must retrain our model online regularly so that we can learn the predictions for each change ^[13].

When to update the learning model

Uncertainty exists around the appropriate time frame for retraining our model to get the best possible forecast.

Choose the appropriate learning model for the concern prediction

A. to the availability of the datasets selection of appropriate Training algorithms is a very complex task. Such as if no of the available data values are lesser pre-processing and feature extraction is also very complex due to their ambiguity in nature. Analysis table of the stock market

Conclusion

The aim of this research study is to give benefit to those investors who invest capital in the stock market. Although it is a very complicated task for investors because of fluctuations and other issues which were discussed above in the paper. So, the purpose behind this paper survey is to classify new technologies and methodologies, datasets and various matrices. This review paper examines a number of machine learning techniques, including Natural Language Processing (NLP), Linear Regression, KNN, SVM, LSTM, Artificial Neural Networks, and many more this survey found that to improve forecasting accuracy and to lower the error rate many studies use the hybrid approach. On the other side big challenge in stock market prediction is that if the dataset contains false knowledge and incorrect news, the stock price estimate will be inaccurate.

In the future to increase stock prediction accuracy and lower the error risk we intend to use new hybrid models and techniques to make prediction more reliable and trustworthy.

References

1. Pai PF, Lin CS. A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*. 2005;33(6):497-505. <https://doi.org/10.1016/j.omega.2004.07.024>
2. Huang W, Nakamori Y, Wang SY. Forecasting stock market movement direction with support vector machine. *Computers and Operations Research*. 2005;32(10):2513-2522. <https://doi.org/10.1016/j.cor.2004.03.016>
3. Ince H, Trafalis TB. Short term forecasting with support vector machines and application to stock price prediction. *International Journal of General Systems*. 2008;37(6):677-687. <https://doi.org/10.1080/03081070601068595>
4. Chong TTL, Ng WK. Technical analysis and the London stock exchange: Testing the MACD and RSI rules using the FT30. *Applied Economics Letters*. 2008;15(14):1111-1114. <https://doi.org/10.1080/13504850600993598>
5. Huang CL, Tsai CY. A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting. *Expert Systems with Applications*, 36(2 PART 1), 2009, 1529-1539. <https://doi.org/10.1016/j.eswa.2007.11.062>
6. Kara Y, Acar Boyacioglu M, Baykan ÖK. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*. 2011;38(5):5311-5319. <https://doi.org/10.1016/j.eswa.2010.10.027>
7. Bhardwaj A, Narayan Y, Vanraj Pawan, Dutta M. Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty. *Procedia Computer Science*. 2015;70:85-91. <https://doi.org/10.1016/j.procs.2015.10.043>
8. Singh R, Srivastava S. Stock prediction using deep learning. *Multimedia Tools and Applications*. 2017;76(18):18569-18584. <https://doi.org/10.1007/s11042-016-4159-7>
9. Bharathi S, Geetha A. Sentiment analysis for effective stock market prediction. *International Journal of Intelligent Engineering and Systems*. 2017;10(3):146-154. <https://doi.org/10.22266/ijies2017.0630.16>
10. Chen S, He H. Stock Prediction Using Convolutional Neural Network. *IOP Conference Series: Materials Science and Engineering*. 2018;435(1). <https://doi.org/10.1088/1757-899X/435/1/012026>
11. Jeon S, Hong B, Chang V. Pattern graph tracking-based stock price prediction using big data. *Future Generation Computer Systems*. 2018;80:171-187. <https://doi.org/10.1016/j.future.2017.02.010>
12. Gandhmal DP, Kumar K. Systematic analysis and review of stock market prediction techniques. In *Computer Science Review (Vol. 34)*. Elsevier Ireland Ltd, 2019. <https://doi.org/10.1016/j.cosrev.2019.08.001>
13. Idrees SM, Alam MA, Agarwal P. A Prediction Approach for Stock Market Volatility Based on Time Series Data. *IEEE Access*. 2019;7:17287-17298. <https://doi.org/10.1109/ACCESS.2019.2895252>
14. Idrees SM, Alam MA, Agarwal P. A Prediction Approach for Stock Market Volatility Based on Time Series Data. *IEEE Access*. 2019;7:17287-17298. <https://doi.org/10.1109/ACCESS.2019.2895252>
15. Ketsetsis AP, Kourounis C, Spanos G, Giannoutakis KM, Pavlidis P, Vazakidis D, *et al*. Deep Learning Techniques for Stock Market Prediction in the European Union: A Systematic Review. *Proceedings - 2020 International Conference on Computational Science and Computational Intelligence, CSCI 2020*, 605-610. <https://doi.org/10.1109/CSCI51800.2020.00107>
16. Gu Y, Shibukawa T, Kondo Y, Nagao S, Kamijo S. Prediction of stock performance using deep neural networks. *Applied Sciences (Switzerland)*, 2020;10(22):1-20. <https://doi.org/10.3390/app10228142>
17. Gu Y, Shibukawa T, Kondo Y, Nagao S, Kamijo S. Prediction of stock performance using deep neural networks. *Applied Sciences (Switzerland)*. 2020;10(22):1-20. <https://doi.org/10.3390/app10228142>
18. Rao PS, Srinivas K, Mohan AK. A Survey on Stock Market Prediction Using Machine Learning Techniques. *Lecture Notes in Electrical Engineering*. 2020;601:923-931. https://doi.org/10.1007/978-981-15-1420-3_101
19. Lu W, Li J, Li Y, Sun A, Wang J. A CNN-LSTM-based model to forecast stock prices. *Complexity*, 2020. <https://doi.org/10.1155/2020/6622927>

20. Shen J, Shafiq MO. Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of Big Data*. 2020;7(1). <https://doi.org/10.1186/s40537-020-00333-6>
21. Vijh M, Chandola D, Tikkiwal VA, Kumar A. Stock Closing Price Prediction using Machine Learning Techniques. *Procedia Computer Science*. 2020;167:599-606. <https://doi.org/10.1016/j.procs.2020.03.326>
22. Sam DS, Pokhariyal GP, Rogo K, Ndhine EO. Otoi-NARIMA model for forecast seasonality of COVID-19 waves: case of Kenya. *Int. J. Stat. Appl. Math*. 2021;6:48-58.
23. Shen J, Shafiq MO. Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of Big Data*. 2020, 7(1). <https://doi.org/10.1186/s40537-020-00333-6>
24. Wu JMT, Li Z, Herencsar N, Vo B, Lin JCW. A graph-based CNN-LSTM stock price prediction algorithm with leading indicators. *Multimedia Systems*, 2021. <https://doi.org/10.1007/s00530-021-00758-w>
25. Patel R, Choudhary V, Saxena D, Singh AK. Review of Stock Prediction Using Machine Learning Techniques. *Proceedings of the 5th International Conference on Trends in Electronics and Informatics, ICOEI, 2021, 840–846*. <https://doi.org/10.1109/ICOEI51242.2021.9453099>
26. Hu Z, Zhao Y, Khushi M. A survey of forex and stock price prediction using deep learning. In *Applied System Innovation*. 2021;4(1):1-30. MDPI AG, <https://doi.org/10.3390/ASI4010009>
27. Jiang W. Applications of deep learning in stock market prediction: Recent progress. In *Expert Systems with Applications*. 2021, 184. Elsevier Ltd., <https://doi.org/10.1016/j.eswa.2021.115537>
28. Rouf N, Malik MB, Arif T, Sharma S, Singh S, Aich S, *et al.* Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. In *Electronics (Switzerland)* 10, 21. MDPI. <https://doi.org/10.3390/electronics10212717>
29. Liu Q, Tao Z, Tse Y, Wang C. Stock market prediction with deep learning: The case of China. *Finance Research Letters*, 2022, 46. <https://doi.org/10.1016/j.frl.2021.102209>
30. Kumar D, Sarangi PK, Verma R. A systematic review of stock market prediction using machine learning and statistical techniques. *Materials Today: Proceedings*. 2022;49:3187–3191. <https://doi.org/10.1016/j.matpr.2020.11.399>
31. Soni P, Tewari Y, Krishnan D. Machine Learning Approaches in Stock Price Prediction: A Systematic Review. *Journal of Physics: Conference Series*. 2022, 2161(1). <https://doi.org/10.1088/1742-6596/2161/1/012065>
32. Chandar SK. Convolutional neural network for stock trading using technical indicators. *Automated Software Engineering*. 2022, 29(1). <https://doi.org/10.1007/s10515-021-00303-z>
33. Kumar D, Sarangi PK, Verma R. A systematic review of stock market prediction using machine learning and statistical techniques. *Materials Today: Proceedings*. 2022;49:3187–3191. <https://doi.org/10.1016/j.matpr.2020.11.399>
34. Obthong M, Tantisantiwong N, Jeamwattanachai W, Wills G. (N.D.). A Survey on Machine Learning for Stock Price Prediction: Algorithms and Techniques. <https://orcid.org/0000-0002-4622-0493>
35. Sen J, Mehtab S. (N.D.). Stock Price Prediction Using Convolutional Neural Networks on a Multivariate Timeseries Ensuring Security and Privacy in Wireless Sensor Networks: Algorithms, Implementations and Applications View project Security and Privacy in Computing and Communications: A Cryptography-Based Approach View project Sidra Mehtab NSHM Knowledge Campus XXX-X-XXXX-XXXX-X/XX/\$XX.00 ©20XX IEEE Stock Price Prediction Using Convolutional Neural Networks on a Multivariate Time Series. <https://www.researchgate.net/publication/338477393> 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). (N.D.). IEEE.
36. Obthong M, Tantisantiwong N, Jeamwattanachai W, Wills G. (n.d.). A Survey on Machine Learning for Stock Price Prediction: Algorithms and Techniques. <https://orcid.org/0000-0002-4622-0493>
37. Obthong M, Tantisantiwong N, Jeamwattanachai W, Wills G. (n.d.). A Survey on Machine Learning for Stock Price Prediction: Algorithms and Techniques. <https://orcid.org/0000-0002-4622-0493>
38. Quah TS, Srinivasan B. (n.d.). Improving returns on stock investment through neural network selection. www.elsevier.com/locate/eswa
39. Chakrabarti S, Saha HN. University of Nevada, Institute of Electrical and Electronics Engineers. Region 1, Institute of Electrical and Electronics Engineers. Region 6, IEEE-USA, & Institute of Electrical and Electronics Engineers. (N.D.). 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC) : 7th-9th January, 2019, University of Nevada, Las Vegas, NV, USA.
40. Sen J, Mehtab S, Nath G. (n.d.). Stock Price Prediction Using Deep Learning Models, 2022. <https://www.researchgate.net/publication/345253844>
41. National Institute of Technology (Punjab, I., National Institute of Technology (Punjab, I. D. of C. S. & E., Institute of Electrical and Electronics Engineers. Delhi Section, & Institute of Electrical and Electronics Engineers. (n.d.). ICSCCC. International Conference on Secure Cyber Computing and Communication : December 15-17, 2018.
42. Kumar M. (n.d.). Forecasting Stock Index Movement: A Comparison of Support Vector Machines And Random Forest.
43. A Predictive Meta Model for Forecasting Stock Price using Time Series Data, 2021. *International Journal of Advanced Trends in Computer Science and Engineering*, 10(1), 345–356. <https://doi.org/10.30534/ijatcse/2021/501012021>
44. Pranav AM, Babu J, Chandran A, Assistant Professor AS. (n.d.). StockClue: Stock Prediction using Machine Learning. *International Journal of Engineering Research & Technology*. www.ijert.org
45. Gomathy CK. The Stock Market Prediction System The Graphical Secret Code In Internet Banking For Improved security transaction view project e-file system view project, 2021. www.irjet.net

46. Saurabh N. LSTM-RNN Model to Predict Future Stock Prices using an Efficient Optimizer. International Research Journal of Engineering and Technology, 2020. www.irjet.net
47. Kumar D, Sarangi PK, Verma R. A systematic review of stock market prediction using machine learning and statistical techniques. Materials Today: Proceedings. 2022;49:3187–3191. <https://doi.org/10.1016/j.matpr.2020.11.399>
48. Parmar I, Agarwal N, Saxena S, Arora R, Gupta S, Dhiman H, *et al.* Stock Market Prediction Using Machine Learning. ICSCCC 2018 - 1st International Conference on Secure Cyber Computing and Communications, 2018, 574-576. <https://doi.org/10.1109/ICSCCC.2018.8703332>
49. Rouf N, Malik MB, Arif T, Sharma S, Singh S, Aich S, *et al.* Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. In Electronics (Switzerland) 2021, 10(21). MDPI. <https://doi.org/10.3390/electronics10212717>
50. Fisher KL, Statman M, Klimek G, Thank W, Christian C, Fernandez R, *et al.* Consumer confidence and stock returns, 2002.
51. Fisher KL, Statman M, Klimek G, Thank W, Christian C, Fernandez R, *et al.* Consumer confidence and stock returns, 2002.
52. Strader TJ, Rozycki JJ, Root TH, Huang YH. Machine Learning Stock Market Prediction Studies: Review and Research Directions. In Journal of International Technology and Information Management, 2020, 28.