An analysis of forecasting methods based on sentiment analysis and deep learning for stock market price movements

Satish Kumar Singh[1], Dr. Sheeba Praveen[2]
[1,2] Department of Computer Science and Engineering, Integral University, Lucknow, India.

Abstract: *Predicting the stock market or share market price movement is an important and effective activity. This prediction is a serious challenge as the availability of the data and its types are very volatile. Earlier the movement of the price was subjected to analysis of historical data patterns but recently daily news and sentiments are also important criteria for predicting market trends. The sources of the data are much categorized and need the latest tools and techniques to process it for any analysis. Investors face difficulty in generating profit from the market, leading to the development of prediction theories. This study highlights the review of such mathematical models that include algorithms of Deep Learning and Machine Learning, calculation methods, and performance parameters using historical data and sentiment analysis that are helping in forecasting the price movements of the stock market till date. Accuracy has been the most significant criterion for the prediction algorithms, and neural networks and LSTM have been widely utilised in conjunction with sentiment analysis.*

Keywords: *Stocks, price prediction, market analysis, sentiment analysis, machine learning, deep learning, price movements*

1. Introduction

A share market is a financial marketplace where shares of publicly listed companies are traded. It provides a platform where buyers and sellers meet to exchange equity shares of public corporations. The evolution of stock market forecasting has attracted considerable attention from professional analysts and investors. Due to the chaotic environment in the market, it is exceedingly challenging to analyse price actions and stock market movements. The complexity of stock prices affects a variety of variables related to stocks, including quarterly earnings reports and market headlines.

Stock market price prediction is a complex task that involves a variety of factors such as economic indicators, company performance, market trends, news, and more. While there are several approaches to predicting stock prices, one of the most common methods for predicting stock prices is through technical analysis and fundamental analysis based on historical data. But with the advent of machine learning and deep learning the prediction process has been automated up to great extent. These algorithms implement various statistical models that analyse vast amounts of historical data and public sentiments received from different sources like news, social media, and financial reports etc. to identify potential future movements.

Sentiment analysis can be used as a tool to complement traditional stock market analysis in predicting stock prices. Sentiment analysis is the process of analyzing and classifying text data,

such as news articles, social media posts, and financial reports, to ascertain the general sentiment toward a specific company or the stock market as a whole. It involves the use of natural language processing (NLP), machine learning, and deep learning techniques. While negative mood can cause stock prices to fall, positive sentiment can cause them to rise. Investors can obtain insight into the expectations and views of the market by analyzing the sentiment of news and social media, and they can then modify their investment decisions as necessary.

This paper is arranged into parts or sections as follows: The full introduction to stock market forecasting is defined in part 1. The driving forces behind stock market forecasting are described in part 2. The Objective of the study and process is described in Section 3. The detailed literature review is given in Section 4, the performance parameters are given in Section 5, section 6 includes the discussion and results of the study, and Section 7 covers the read publication and challenges encountered in the process.

2. Driving forces behind stock market forecasting

Figure 1: The driving forces behind stock market forecasting

3. Objective of the study

Our survey's main objective is to collect empirical data on stock market forecasts utilizing sentiment analysis and machine learning models. The aim is to study the following as of now:

1. The study of available algorithms based on deep learning and machine learning
2. To study the algorithms of sentiment analysis for stock market prediction
3. To highlight and study hybrid approach if available

4. Performance parameters of different algorithms used in price prediction
5. Different literature sources used

4. Literature Review:

4.1 The Methodology: The Overview of the stock market price prediction process may be represented as:

Figure 2: The generic stock market price prediction process

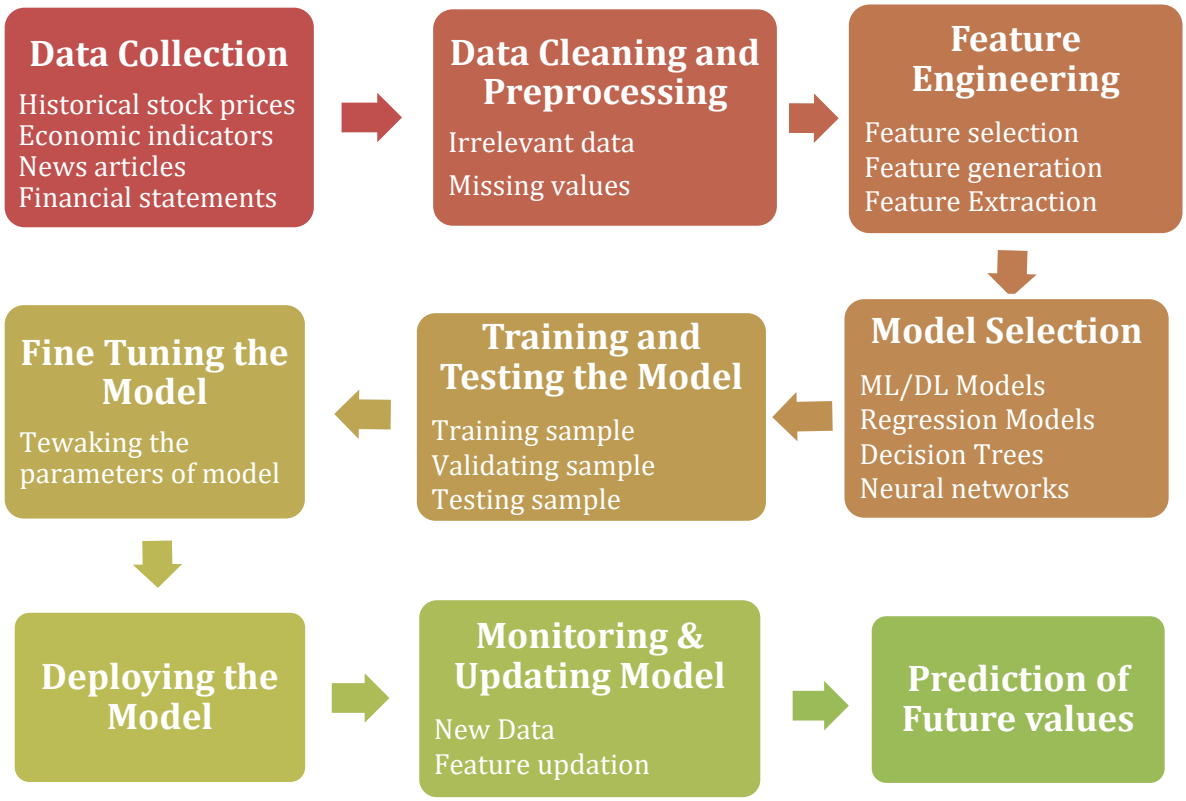

Figure 3: The detailed stock market price prediction process

4.2 The ML and DL algorithms for stick market price prediction and sentiment analysis

The state-of-the-art methods used in stock market forecasting employing historical data and sentiment analysis are elaborated on in the section that follows.

A. Linear Regression Algorithms
B. Support Vector Machine (SVM)
C. Neural Networks (NN)
D. Artificial neural network (ANN)
E. Convolutional Neural Network (CNN)
F. Recurrent Neural Network (RNN)
G. Support vector regression (SVR)
H. Generative adversarial network (GAN)
I. Naïve Bayes (NB)
J. Long Short-Term Memory (LSTM)

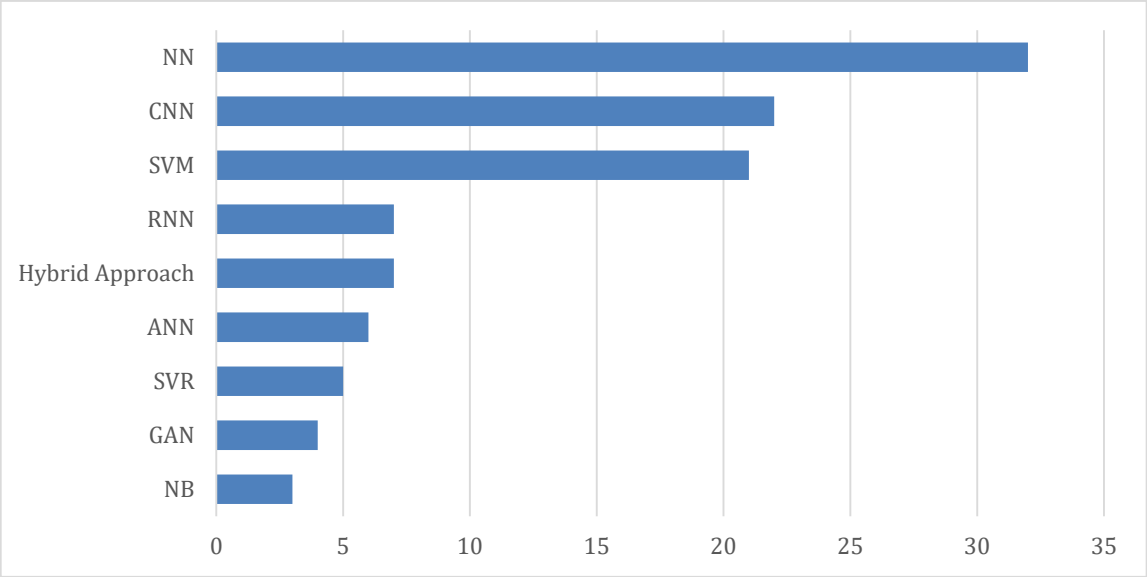

Figure 4: Stock market prediction methods

Furthermore, some of the chosen studies use ML or deep learning methods to forecast the stock market. These algorithms were put through their paces utilising performance measurements and the real-time dataset and its attributes.

The table 1 represents the various ML and DL algorithms used in stock market price prediction year by year. It shows that SVM and LSTM are the two major algorithms that contributed most in price prediction.

Contributor	Algorithms and Model used	Year	Ref.
Bing et al.	BPNN on Composite Index based on Shanghai Stock Exchange	2012	1
Wensheng et al.	Asia stock market Analysis (Non-linear independent component analysis (NLICA) & BPNN)	2012	2
Patel et al.	ANN, SVM, RF, and Naive Bayes	2013	3
Ticknor	Modified Bayesian ANN	2013	4
Olivera et al.	Modified ANN	2013	5

Hegazy. et al.	Comparison of particle swarm optimization (PSO) and least-squares SVM (LS-SVM).	2013	6
Lin et al.	Support Vector Machine (Correlation-based) with quasi-linear SVC	2013	7
Kazem et al.	support vector regression (SVR)	2013	8
Alkhatib et al.	KNN algorithm-based model	2013	9
Ding et al.	Large-scale public news analysis through SVM (sentiment analysis)	2014	10
Bailings et al	AdaBoost, Random forest (RF), kernel factory, k- nearest neighbours (KNN) and NN, SVM	2015	11
Vui et al	ANN and its variant models	2015	12
Chen et al.	LSTM model	2015	13
Ding et al.	Event-embedding CNN through Neural tensor network (NTN)	2015	14
Qiu et al	A Model based on Genetic algorithms and ANN	2016	15
Akita et al.	Application of paragraph vector using LSTM method	2016	16
Li et al.	Comparative study of SVM, BPN and Extreme Learning Machine (ELM)	2016	17
Selvin et al.	LSTM, RNN, CNN-Sliding window	2017	18
Nelson et al	Comparative analysis of multilayer perceptron (MLP) and RF with LSTM	2017	19
Chang et al.	Model using connected neural networks (partially) (EPCNN)	2017	20
Vargas et al	Application of Recurrent CNN (RCNN)	2017	21
Roondiwala et al.	LSTM and RNN	2017	22
Sharma et al.	Regression based models	2017	23
G. et al.	MLP, RNN, LSTM and CNN	2018	24
Ren et al.	Study on sentiment analysis accuracy through SVM	2018	25
Kang et al.	Implementation of LSTM with Generative adversarial networks (GAN), and MLP	2019	26
Moghar et al.	RNN-based LSTM	2020	27

.

Table 1: ML and DL used for price prediction

The algorithms are evaluated at different parameters by other researchers to check their performance. The performance evaluation at different parameters is discussed in table 2 given below:

4.3 Performance Parameters of Algorithms:

In order to determine if Machine Learning and Deep Learning can anticipate stock markets prices and forecasts with greater accuracy, a variety of performance indicators are used. These metrics rate algorithms according on their methodology and dataset. The numerous performance metrics utilized by the selected studies to assess their success include the following:

A. Accuracy
B. Root-mean-square error (RMSE)
C. Mean Absolute Error (MAE)
D. Mean Squared Error (MSE)
E. Mean Absolute Percentage Error (MAPE)

Following table respresents the comperative performace anaylisys of different atock market price prediction algorithms based on evaluation metrices and pridiction targets.

Contributors	Prediction target	Evaluation metrics	Performance	Year	Ref
Haq et al.	Daily stock prices prediction	Accuracy	59.44	2021	28
Labiad et al.	Forecasting of 10 min ahead prices		90%	2016	29
Kumar et al.	Price prediction of one-day forward		44.22–62.72	2016	30
Alsubaie et al	Direction of the stock returns		0.05–80.79	2019	31
Nabi et al	Forecasting of monthly price		100%	2019	32
Yuan et al.	Prediction of excess returns		52%	2020	33
Shen et al	Forecasting of stock price		0.93	2020	34
Singh et al	Direction of 10 days ahead		83.62%	2021	35
Qolipour et al.	Prediction of stock return		0.947–1.0	2021	36
Rana et al	Daily stock price	RMSE	0.0151	2019	37
Chen et al.	Direction of stock indices price		0.0143–0.0239	2017	38
Siddique et al	Predicting the next day close price		4.3	2019	39
Das et al	Prediction of 1,3,5,7,15,30 days ahead price		143.1104	2019	40
Li et al.	Predicting one-day ahead close price value	MAE	0.054	2022	41
Botunac et al	Direction of close price		0.01606	2020	42
Siddique et al	Next day close price		2.76	2019	43
Das et al	Prediction of 1,3,5,7,15,30 days ahead price		121.8011	2019	44
Farahani et al.	Close price		13.499	2021	45
Dami et al	Returns prediction		0.022–0.039	2021	46
Li et al.	Predicting closing priced on day advance	MSE	0.006	2022	47
Botunac et al	Direction of close price		0.00046	2020	48
Li et al.	Forecasting one day ahead closing prices	MAPE	1.092	2022	49
Chen et al.	Prediction of direction of stock indices		0.646–1.06	2017	50
Siddique et al	Next day close price		0.63	2019	51
Das et al	Prediction of 1,3,5,7,15,30 days ahead price		0.308	2019	52

Table 2: Comparative performance analysis of algorithms

5. Results and Discussions

The use of NN accounts for 41% of all the papers reviewed, SVM for 16%, LSTM for 18%, RNN for 14% and other approaches account for 11%. The use of NN is 228% higher than that of LSTM, 256% higher than that of SVM, 292% higher than that of RNN and 372% higher than that of other models as given in the figure1 above.

LSTM usage surged in 2015, whereas SVM usage was primarily prior to 2016. NN, on the other hand, has become more prevalent in the last ten years. Other approaches, such as Random Forest and KNN, are employed in share market forecasting less frequently.

All these algorithms use some datasets collected from different sources that are available on the web. In statistical analysis, this data is very helpful in predicting price movements based on past data e.g. Linear regression. For preparing any mathematical model these data are very useful. In sentiment analysis, different sources like social media posts, blogs, news, and financial reports have been considered if it was needed.

The relationships and data patterns between input and output can be understood, generalized, and modified via neural networks. As a result, NNs are advantageous for stock market forecasting and are more precise than other approaches. NNs are the most popular model used in stock market prediction due to these factors. In recent years, price prediction has used the bi-directional LSTM more and more.

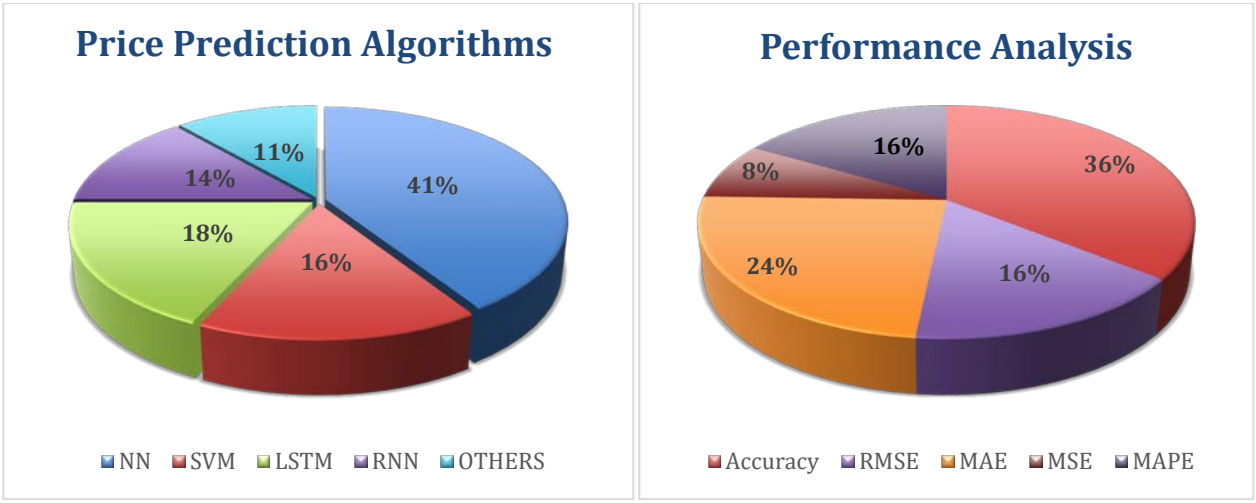

Figure5: Algorithms used and their performance analysis parameters

Figure 5 represents the percentage of evaluation metrics covered in the study given in the previous table. Figure 6 represents the paper's distribution year by year for the algorithms used for prediction and Figure 7 highlights the performance analysis metrics distribution year wise.

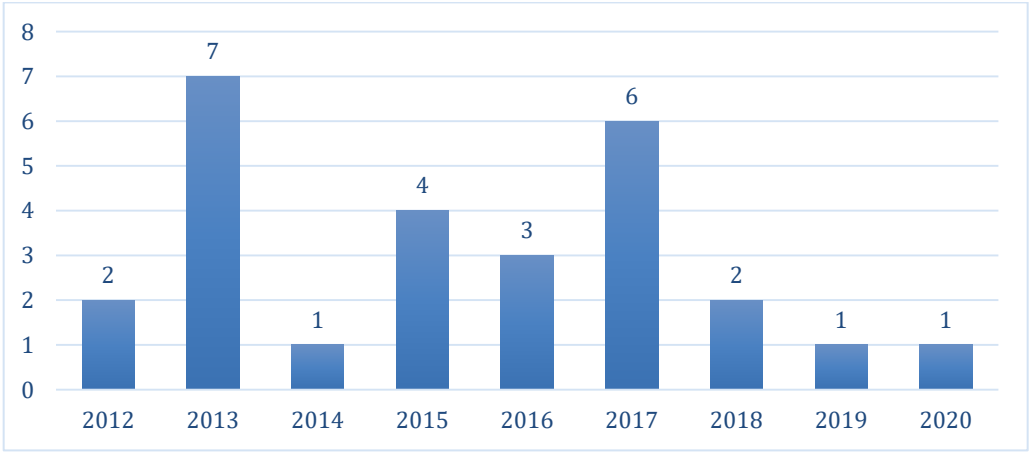

Figure 6: Prediction algorithms papers distribution year wise

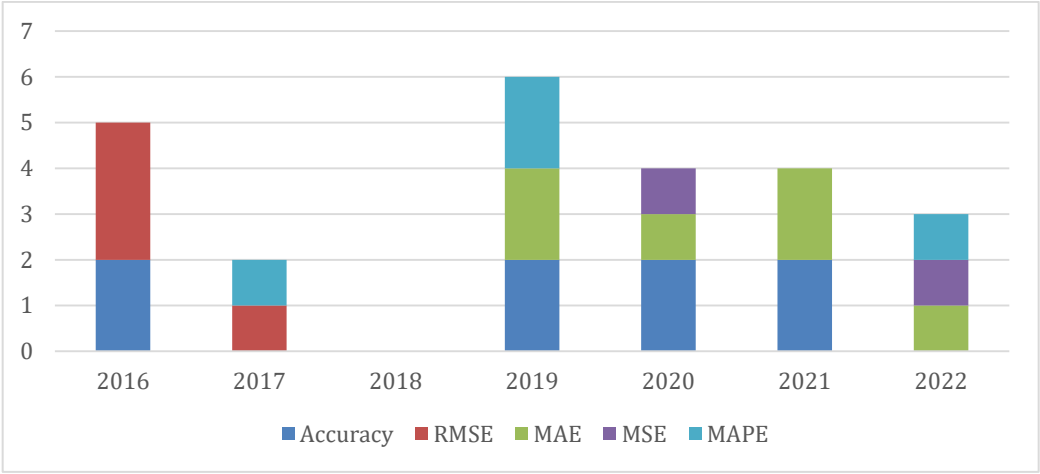

Figure 7: Algorithms distribution on evaluation metrics year wise

The above figure shows that accuracy and MAE was the most dominant metric for the performance evaluation for the given period. But since 2020 MSE have been a very popular metric and together with MAE it has been widely used for the performance analysis of the price prediction algorithms.

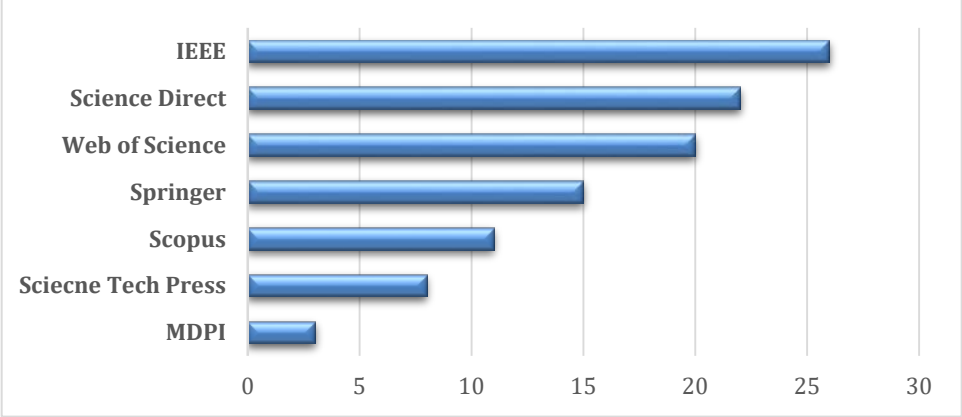

Figure 8: Literature review sources

6. Conclusion

In price prediction algorithms Neural Network based algorithms were used initially which included mainly ANN, CNN and SVM but with the advances in the technologies later they were replaced by LSTM and LSTM based other methods combined with ANN, CNN, RNN and RNN based LSTM. In performance evaluation, accuracy criteria has been used extensively to evaluate the performance of prediction algorithms for short durations for less than 15 days. Other criteria RMSE, MAPE and MAE were very helpful in testing the algorithms in longer duration prediction process for 15 to 30 days in advance.

Acknowledgement

We would like to extend our sincere appreciation to Integral University for their invaluable support in facilitating our research endeavours. The Manuscript Control Number (MCN) [IU/R&D/2023-MCN0002127], issued by Integral University, serves as a testament to their commitment to advancing academic research. This recognition underscores the significance of our work, and we are grateful for the university's dedication to fostering scholarly contributions in our field."

References

[1] Bing Y, Hao JK, Zhang SC. Stock market prediction using artificial neural networks. In: Advanced Engineering Forum. vol. 6. Trans Tech Publ; 2012. p. 1055-60.
[2] Dai W, Wu JY, Lu CJ. Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. Expert systems with applications. 2012;39(4):4444-52.
[3] Patel J, Shah S, Thakkar P, Kotecha K. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert systems with applications. 2015;42(1):259-68
[4] Ticknor JL. A Bayesian regularized artificial neural network for stock market forecasting. Expert systems with applications. 2013;40(14):5501-6.
[5] De Oliveira FA, Nobre CN, Za´rate LE. Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index–Case study of PETR4, Petrobras, Brazil. Expert systems with applications. 2013;40(18):7596-606
[6] Hegazy O, Soliman O, Salam M. A Machine Learning Model for Stock Market. arXiv Prepr arXiv1402;7351:17-23.
[7] Lin Y, Guo H, Hu J. An SVM-based approach for stock market trend prediction. In: The 2013 international joint conference on neural networks (IJCNN). IEEE; 2013. p. 1-7.
[8] Kazem A, Sharifi E, Hussain FK, Saberi M, Hussain OK. Support vector regression with chaos-based firefly algorithm for stock market priceforecasting. Applied soft computing. 2013;13(2):947-58.
[9] Alkhatib K, Najadat H, Hmeidi I, Shatnawi MKA. Stock price prediction using k-nearest neighbor (kNN) algorithm. International JournalofBusiness, Humanities and Technology. 2013;3(3):32-44.
[10] Ding X, Zhang Y, Liu T, Duan J. Deep learning for event-driven stock prediction. In: Twenty-fourth international joint conference on artificial intelligence; 2015.

[11] Ballings M, Van den Poel D, Hespeels N, Gryp R. Evaluating multiple classifiers for stock price direction prediction. Expert systems withApplications. 2015;42(20):7046-56.
[12] Vui CS, Soon GK, On CK, Alfred R, Anthony P. A review of stock market prediction with Artificial neural network (ANN). In: 2013 IEEE international conference on control system, computing and engineering. IEEE; 2013. p. 477-82.
[13] Chen K, Zhou Y, Dai F. A LSTM-based method for stock returns prediction: A case study of China stock market. In: 2015 IEEE international conference on big data (big data). IEEE; 2015. p. 2823-4.
[14] Ding X, Zhang Y, Liu T, Duan J. Deep learning for event-driven stock prediction. In: Twenty-fourth international joint conference on artificial intelligence; 2015.
[15] Qiu M, Song Y. Predicting the direction of stock market index movement using an optimized artificial neural network model. PloS one.2016;11(5):e0155133.
[16] Akita R, Yoshihara A, Matsubara T, Uehara K. Deep learning for stock prediction using numerical and textual information. In: 2016 IEEE/ACIS15th International Conference on Computer and Information Science (ICIS). IEEE; 2016. p. 1-6.
[17] Li X, Xie H, Wang R, Cai Y, Cao J, Wang F, et al. Empirical analysis: stock market prediction via extreme learning machine. Neural Computingand Applications. 2016;27(1):67-78.
[18] Selvin S, Vinayakumar R, Gopalakrishnan E, Menon VK, Soman K. Stock price prediction using LSTM, RNN and CNN-sliding windowmodel. In: 2017 international conference on advances in computing, communications and informatics (icacci). IEEE; 2017. p. 1643-7.
[19] Nelson DM, Pereira AC, De Oliveira RA. Stock market's price movement prediction with LSTM neural networks. In: 2017 International jointconference on neural networks (IJCNN). Ieee; 2017. p. 1419-26.
[20] Chang PC, et al. A novel model by evolving partially connected neural network for stock price trend forecasting. Expert SystemswithApplications. 2012;39(1):611-20.
[21] Vargas MR, De Lima BS, Evsukoff AG. Deep learning for stock market prediction from financial news articles. In: 2017 IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA). IEEE; 2017. p. 60-5.
[22] Roondiwala M, Patel H, Varma S. Predicting stock prices using LSTM. International Journal of Science and Research (IJSR). 2017;6(4):1754-6.
[23] Sharma A, Bhuriya D, Singh U. Survey of stock market prediction using machine learning approach. In: 2017 International conferenceofelectronics, communication and aerospace technology (ICECA). vol. 2. IEEE; 2017. p. 506-9
[24] Hiransha M, Gopalakrishnan EA, Menon VK, Soman K. NSE stock market prediction using deep-learning models. Procedia computer science.2018;132:1351-62.
[25] Ren R, Wu DD, Liu T. Forecasting stock market movement direction using sentiment analysis and support vector machine. IEEE SystemsJournal. 2018;13(1):760-70.
[26] Zhang K, Zhong G, Dong J, Wang S, Wang Y. Stock market prediction based on generative adversarial network. Procedia computer science.2019;147:400-6.
[27] Moghar A, Hamiche M. Stock market prediction using LSTM recurrent neural network. Procedia Computer Science. 2020;170:1168-73.
[28] Haq AU, Zeb A, Lei Z, Zhang D (2021) Forecasting daily stock trend using multi-filter feature selection and deep learning. Expert Syst Appl 168:114444.
[29] Labiad B, Berrado A, Benabbou L (2016) Machine learning techniques for short term stock movements classification for moroccan stock exchange, In 11th SITA, Mohammedia, Morocco.
[30] Kumar D, Meghwani SS, Thakur M (2016) Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets. J Comput Sci 17:1–13.

[31] Alsubaie Y, Hindi KE, Alsalman H (2019) Cost-sensitive prediction of stock price direction: selection of technical indicators. IEEE Access 7:146876–146892.
[32] Nabi RM, Saeed SAM, Harron HB, Fujita H (2019) Ultimate prediction of stock market price movement. J Comput Sci 15(12):1795–1808.
[33] Yuan X, Yuan J, Jiang T, Ain QU (2020) Integrated long-term stock selection models based on feature selection and machine learning algorithms for china stock market. IEEE Access 8:22672–22685.
[34] Shen J, Shafiq MO (2020) Short-term stock market price trend prediction using a comprehensive deep learning system, J Big Data, 766.
[35] Singh J, Khushi M (2021) Feature learning for stock price prediction shows a significant role of analyst rating. Appl Syst Innov. https://doi.org/10.3390/asi4010017.
[36] Qolipour F, Ghasemzadeh M, Mohammad-Karimi N (2021) The predictability of tree-based machine learning algorithms in the big data context. Inter J Eng 34(01):82–89.
[37] Rana M, Uddin MM, Hoque MM (2019) Effects of activation functions and optimizers on stock price prediction using LSTM recurrent networks, CSAI, Beijing, China, pp 354–358.
[38] Chen Y, Hao Y (2017) A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. Expert Syst Appl 80:340–355.
[39] Siddique M, Panda D (2019) A hybrid forecasting model for prediction of stock index of tata motors using principal component analysis, support vector regression and particle swarm optimization. I J Eng Adv Tech 9:3032–3037.
[40] Das SR, Mishra D, Rout M (2019) Stock market prediction using firefly algorithm with evolutionary framework optimized feature reduction for OSELM method". Expert Syst Appl X 4:100016.
[41] Li G, Zhang A, Zhang Q, Wu D, Zhan C (2022)Pearson correlation coefficient-based performance enhancement of Broad Learning System for stock price prediction, IEEE Trans Circuits Syst II, (Early Access).
[42] Botunac I, Panjkota A, Matetic M (2020) The effect of feature selection on the performance of long short-term memory neural network in stock market predictions, In 31st DAAAM ISIMA, Vienna, Austria, pp 0592-0598.
[43] Siddique M, Panda D (2019) A hybrid forecasting model for prediction of stock index of tata motors using principal component analysis, support vector regression and particle swarm optimization. I J Eng Adv Tech 9:3032–3037.
[44] Das SR, Mishra D, Rout M (2019) Stock market prediction using firefly algorithm with evolutionary framework optimized feature reduction for OSELM method". Expert Syst Appl X 4:100016.
[45] Farahani MS, Hajiagha SHR (2021) Forecasting stock price using integrated artificial neural network and metaheuristic algorithms compared to time series models. Soft Comput 25:8483–8513.
[46] Dami S, Esterabi M (2021) Predicting stock returns of Tehran exchange using LSTM neural network and feature engineering technique. Multimed Tools Appl 80:19947–19970.
[47] Li G, Zhang A, Zhang Q, Wu D, Zhan C (2022)Pearson correlation coefficient-based performance enhancement of Broad Learning System for stock price prediction, IEEE Trans Circuits Syst II, (Early Access).
[48] Botunac I, Panjkota A, Matetic M (2020) The effect of feature selection on the performance of long short-term memory neural network in stock market predictions, In 31st DAAAM ISIMA, Vienna, Austria, pp 0592-0598.
[49] Li G, Zhang A, Zhang Q, Wu D, Zhan C (2022)Pearson correlation coefficient-based performance enhancement of Broad Learning System for stock price prediction, IEEE Trans Circuits Syst II, (Early Access).
[50] Chen Y, Hao Y (2017) A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. Expert Syst Appl 80:340–355.

[51] Siddique M, Panda D (2019) A hybrid forecasting model for prediction of stock index of tata motors using principal component analysis, support vector regression and particle swarm optimization. I J Eng Adv Tech 9:3032–3037.
[52] Das SR, Mishra D, Rout M (2019) Stock market prediction using firefly algorithm with evolutionary framework optimized feature reduction for OSELM method". Expert Syst Appl X 4:100016.