Materials Today: Proceedings 49 (2022) 3187–3191

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

A systematic review of stock market prediction using machine learning and statistical techniques

Deepak Kumar [a], Pradeepta Kumar Sarangi [a], Rajit Verma [b]

[a] Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
[b] Chitkara Business School, Chitkara University, Punjab, India

ARTICLE INFO

Article history:
Received 21 September 2020
Accepted 3 November 2020
Available online 6 January 2021

Keywords:
Stock market prediction
Machine learning (ML)
Classification
Deep learning
Support vector machine (SVM)
Neural networks (NN)

ABSTRACT

The stock market prediction patterns are seen as an important activity and it is more effective. Hence, stock prices will lead to lucrative profits from sound taking decisions. Because of the stagnant and noisy data, stock market-related forecasts are a major challenge for investors. Therefore, forecasting the stock market is a major challenge for investors to use their money to make more profit. Stock market predictions use mathematical strategies and learning tools. This paper provides a complete overview of 30 research papers recommending methods that include calculation methods, ML algorithms, performance parameters, and outstanding journals. The studies are selected based on research questions. Hence, these selected studies are helping to find the ML techniques along with their dataset for stock market prediction. Most widely ANN and NN techniques are used to achieve precise predictions of the stock market. While much amount of work is done, the latest stock market-related prediction methodology has many limitations. It can be assumed in this study that stock market forecasting is an integrated process and distinctive parameters for forecast the stock market should be considered more accurate.

Selection and peer-review under responsibility of the scientific committee of the National Conference on Functional Materials: Emerging Technologies and Applications in Materials Science.

1. Introduction

One of the most fascinating inventions in the current time is the finance markets. These finance markets have a great significant [1] impact on many areas such as business, employment technology. Investors have used two main strategies to make decisions on the stock market to invest the money and earn more returns with low risks. The development in [2] stock market prediction has gained high significance among expert analysts and investors. Analyzing stock market movements [3] and price actions are extremely difficult due to the noisy environment in the market. The complication of stock prices changes many factors that include announcements of quarterly earnings and market news. The stock market [4] indices are calculated based on their market capitalization. Accurate forecasting of the stock market [5] is therefore a very difficult task by changing the market world. The researchers and market analysts [6,7] have been keen on developing and testing of stock market behavior. Therefore, different types of statistical techniques that include autoregressive integrated moving average and clustering are applied for stock market prediction. Since this model gives historical evidence and theories for normality postulates. Extensive research [8,9] was conducted in the area of stock market forecasting applications in the using SVM, NN, and Genetic adversarial network (GAN) ML techniques. By guessing the stock market price index [10] the data analyzer has used a neural network (ANN) and Support vector regression (SVR) ML algorithms. Every ML algorithm has a way to learn patterns. Some of the authors [11] have proposed a hybrid method and it is a combination of a basic set and a neural network (ANN) used to calculate relative strength index (RSI). RSI acts as a measuring oscillator and compares past price strength with the current price. The results show that the proposed hybrid technique could cope with high variations of stock market values with great accuracy as compared to other ML approaches. Generally, some of the researchers [12] apply a combination of long short term memory (LSTM) with a genetic algorithm (GA) on the Korea stock price index for stock market forecasting using available financing data. Many networks

E-mail addresses: Deepak.k@chitkara.edu.in (D. Kumar), Pradeepta.sarangi@-chitkara.edu.in (P.K. Sarangi), Rajit.verma@chitkara.edu.in (R. Verma)

https://doi.org/10.1016/j.matpr.2020.11.399
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the National Conference on Functional Materials: Emerging Technologies and Applications in Materials Science.

[13,14] use the feed-forward network for forecasted stock patterns. The main effect of this paper is to provide an overview of the stock market forecast strategies that are very helpful in predicting the stock market for the future. Therefore, this review observes several statistical and ML techniques for stock market prediction along with their datasets. The analysis should be carried out by taking several methodologies, performance matrices, datasets used, and techniques for stock market prediction.

Thus this paper is following as: Section 1 defines the detailed introduction of stock market forecasting. Section 2 outlines the need regarding the stock market prediction. Section 3 describes the research methodology for selecting subjects. Section 4 provides an overview of the findings and discussions and section 5 provides research findings.

2. The need for stock market forecasting:

In the stock market, the investor shows interest in profit by investing some money in the stock market. The stock market has shows investor interest due to advanced applications where prediction may lead to prosperous market forecasting. Predicting movements [15] of the stock market precisely depends on advance information. The tools which are used for stock market forecasting [16,17] can track and control the market which can be used to make the right decisions. The stock market needs to handle several information [18] on industrial stocks which covers the entire financial market. These are adjusted according to the business status investors [19–21] who consider sales and purchase. Several factors affect the market position are the future estimation income, a news release on earnings and changes in management, etc. Therefore, accurate prediction [22] of the stock market helps investors in making better decisions. Through ML techniques the investor can earn more money with high risk. Fig. 1 describes the process of the stock market.

In Fig. 1, firstly real-time data is collected from various sources either websites or previous datasets such as NASDAQ [23] based on their price index. The price index is a subset of the stock market and it grants investors to compare the current price flatten with past market prices for performance calculation. After collecting data, the collected data is preprocessed [24] for removing the noise and other parameters. Then preprocessed data can be useful for stock market forecasting. The feature selection methods select some features [25] from a huge amount of data. Some of the data analyzer function or user-friendly application divides the dataset into two subcategories namely as current and prediction details. These details are very useful for making better stock market decisions. After a strong decision, a notification [26,27] is sent to investors about the prices index. This notification is very useful for investors because this notification [28] tells the status of profit or loss for price index. If the status generated by the application [29] is profit, then the investor can use the shares for high sales and if the prices index is low [30] then development is more attention to make better decisions.

3. Research approach

Te main aim of our survey is to summarize the empirical evidence of stock market prediction based on ML models. In this approach, five research questions (RQ1, RQ2, RQ3, RQ4, RQ6) are under the narrative synthesis method and one research question (RQ4) is under the vote-counting method. The research approach includes the research questions which are helping for extracting information. From, the selected studies, we construct some research questions which are listed as:

RQ1. What are the various statistical tools are used in a stock market analysis?
RQ2. What are the different types of ML algorithms used to predict the stock market?
RQ3: What are the distinct types of datasets used for stock market forecasting?
RQ4. Is any hybrid approach of ML model has been used for stock market forecasting or not?
RQ5: What is the different performance parameters used in the stock market forecasting?
RQ6: What are the most dominant journals for stock market forecasting?

4. Results and discussions

We are selecting some research articles based on research questions. In this section, we are reviewing the research question which is described in the above section. The research questions described as:

I. RQ1. What are the various statistical tools are used in a stock market analysis?

After selected studies, we are analyzing and extract the information. For more, we are studied some statistical tools which are used in stock market analysis. The various statistical methods were used in the analysis containing basic descriptive for interpretation of the stock market. Some of the selected studies use ARIMA, Regression, and clustering techniques for stock market forecasting. Each technique is described as:

a. ARIMA: The Arima [11] is a statistical tool and used in time series to predict future trends to understand the dataset better.
b. Clustering: The clustering method [13,14] is used to group sets of objects that share similar properties. The stocks that are having high correlations fall into one basket those are less correlated into another. This process is continued until each stock is placed in every group.

From the analysis of the Table 1, some of the selected topics use statistical methods used to predict the stock market. Only 18% of

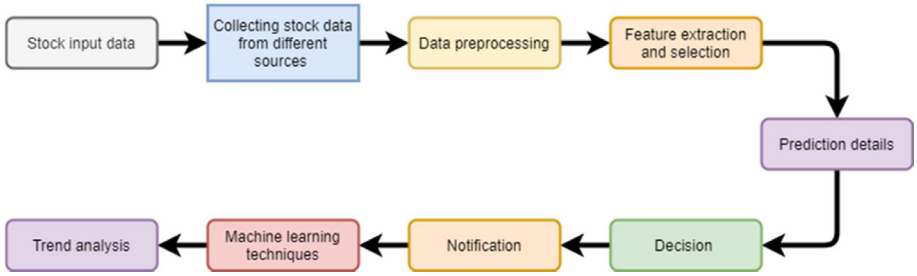

Fig. 1. Stock market prediction process.

Table 1
Statistical techniques used by selected studies.

Tools	Selected studies	Percentage (%)
An autoregressive integrated moving average (ARIMA)	S8, S3, S8, S13, S14, S16	18
Clustering	S17	3

the subjects are popular especially in stock market predictions but only one study is used in the merger process.

II. RQ2. What are the different types of ML algorithms used to predict the stock market?

Most selected subjects use ML or in-depth learning methods to predict the stock market. A few selected studies use the hybrid approach for better prediction accuracy in stock market forecasting. Mostly the stock market prediction approaches are elaborated in this section. The most common prediction techniques are described as:

a. SVM: SVM is one of the better effective methods for forecast time series. SVM can be used for regression and classification purposes. SVM consists of plotting [15,16] data as a point in n-dimensional space. Such stock market measurements are characterized and plotted on different coordinate planes. SVM is the most powerful and predictive tool in the financial market.
b. NN: It is a sequence of algorithms [17] that identifies the changing dataset over a mechanism and describes a way that how the human brain works. Chung and Shin [18] developed an outer break NN approach to stronger stock market forecasting. The data is taken in real-time from the livestock market. The deep LSTM dependent NN was established here for the usage of the embedded layer. The neural LSTM network encoder to predict trends in stock.
c. Artificial neural network (ANN): ANN captures the structural relation [19] of stock more specifically output and its determinants than lots of other statistical techniques. Many of the researchers applied the ANN model before preprocessing the data. There are so many performance parameters used in stock market forecasting which is discussed in RQ5.
d. Convolutional NN (CNN): The CNN [20] is feed up the neural network. In CNN, there is more number of hidden layers as correlated to traditional neural network techniques. CNN is the name of a comprehensive learning algorithm and used in stock market predictions.
e. Recurrent NN (RNN): RNN [20] in the form of ANN in which nodes are interconnected in graph shape which is directed along with their temporal chain. Therefore, it allows showing complex dynamic behavior.
f. Support vector regression (SVR): The SVR [5] adopts the concepts of SVM but there is only a small difference between an SVM and SVR. The SVR is used in stock market prices forecasting but SVM is used in stock market forecasting according to their time-series.
g. Generative adversarial network (GAN): The GAN [23] is a new framework and it is training two versions like zero fun game. In the antagonist cycle, the generator can be called a fraud to produce data close to real data while the racist takes on the role of judges to separate real data from data processing.
h. Naïve Bayes (NB): The NB is a classification algorithm [29] that generates the Bayesian networks for a specific Bayes based dataset theorem. It supposes that the specified dataset contains a unique function that is unrelated to any other function in a class. It is a simple algorithm and outer perform for highly grading strategies for large datasets.

Moreover, some of the selected studies use either ML or deep learning techniques for stock market forecasting. These algorithms have taken the real-time dataset along with their features and evaluated through performance parameters. The description of each selected study along with their implementation ML algorithm is listed in Table 2.

From the analysis of Table 2, most of the selected studies use the NN approaches frequently. Fig. 2 shows the percentage of technique.

III. RQ3: What are the distinct types of datasets used for stock market forecasting?

A selected study uses various types of a dataset in stock market forecasting. According to selected studies some datasets are publicly available. Most selected subjects have used public data sets to predict the stock market. These data sets are used for classification or predictive purposes. Table 3 shows various types of datasets used by various selected studies are described as:

The above table shows most of the selected studies used the NASDAQ dataset in stock market prediction/forecasting.

IV. RQ4. Is any hybrid approach of ML model has been used for stock market forecasting or Not?

The Fig. 2 shows that only three of the selected studies use the hybrid approach for stock market forecasting. Selected study S3 proposed a hybrid method that is an ANN compound with a rough approach and S8 proposed a hybrid method that combines ANN with GA to improve GA performance in stock market forecasts. Another study S13 combined the statistical technique discrete wavelet transform with ML algorithm ANN (DWT-ANN) for stock market prediction.

V. RQ5: What is the different performance parameters used in the stock market forecasting?

Different performance parameters are used to test ML's ability to better predict stock markets/exchanges/forecasts. These performance parameters are evaluating the particular algorithm based on their technique and dataset. Various performance parameters used by the selected studies to measure their performance is described as:

a. Accuracy: Accuracy [21] is one metric used to assess the model classification. Informal accuracy is part of the prediction that our model is correct.

Table 2
Percentage of each technique used by selected studies.

# Studies	Techniques	Percentage (%)
S2, S3, S4, S9, S12, S15, S22	SVM	21
S1, S2, S10, S12, S17, S18, S21, S24, S25, S27, S29	NN	33
S5, S6, S11, S14, S16, S20, S28, S30	ANN	24
S24, S26	CNN	6
S24, S26	RNN	6
S5	SVR	3
S23	GAN	3
S29	NB	3
S7, S8, S13	Hybrid approaches	9

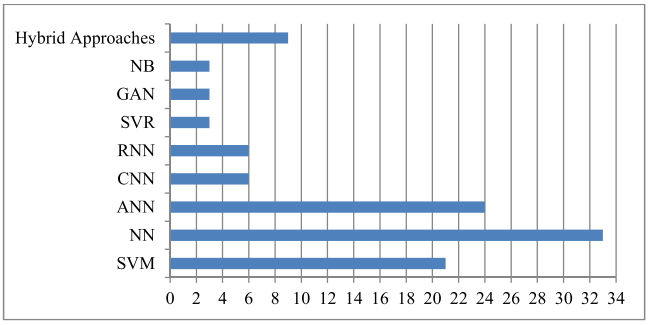

Fig. 2. Most frequently ML techniques.

Table 3
Dataset used by selected studies.

#Studies	Year	Target	Dataset values (days)/source
S1	2007	Dow Jones industrial average index	1024
S2	2016	Stock market	500
S3	2019	Stock market	NASDAQ
S4	2019	Stock market	1659/ www.moneycontrol.com
S5	2014	Stock market	www.nseindia.com
S6	2019	Stock market	NASDAQ
S7	2012	DHAKA stock market	www.dse.com.hd
S8	2014	Stock market	OGDCL Pakistan
S9	2002	Stock market	104/FASM
S10	2016	Stock market	TFIDM
S11	2003	Stock market	100/INSTABUL STOCK EXCHANGE
S12	2005	Stock market forecasting	MCcardy cd mahen
S13	2016	Stock market	1414/tatasteel, Cisco
S14	2013	Stock market forecasting	734/ Goldman sacks Inc.
S15	2017	Stock market	108
S16	2015	Stock market	1024
S17	2011	Stock market	360
S18	2018	Stock market	4203/Korea stock index price
S19	2014	Stock market	www.finet.hk
S20	2017	Stock market	38/kospi market
S21	2018	Stock market	2691
S22	2016	Stock market	Crawler
S23	2019	Stock market	5000/ new york stock exchange
S24	2017	Stock market	1721
S25	2017	Stock market	NASDAQ
S26	2017	Stock market	600
S27	2018	Stock market	500
	2016	Stock market	Dhaka stock exchange
S29	2017	Stock market	CS1300
S30	2011	Stock market	NASDAQ

b. Root mean square error (RMSE): The RMSE is used at the level used to calculate [22] the difference between the expected model values and the retained values. RMSE is very close to the training and assessment database.
c. Mean absolute error (MAE): MAE [24,25] is used in regression values. In this case, error prediction is the sum of the differences between the expected and actual variables, divided by the number of data points above all data points. MAE refers to calculating the difference between two continuous variables.
d. Mean squared error (MSE): MSE is square average error used [26,27] as a loss function for calculating the minimum square regression. Also, it is the sum of the differences between the expected and actual variables, divided by the number of data points above all data points.

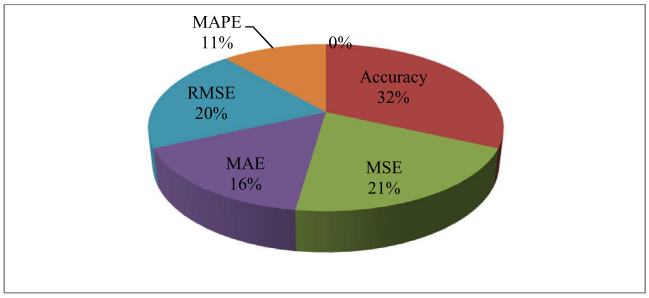

Fig. 3. Selected studies used performance parameters.

Table 4
Most dominant journals/conferences.

# Studies	Conference/Journal	%	Publisher
S1	Applied Intelligence	3	Springer
S2	International conference on computer and information sciences	3	IEEE
S3	International journal of financial studies	3	MDPI
S4	Computational intelligence in data mining	3	Springer
S5, S14, S20, S30	Expert systems with applications	12	ScienceDirect
S6	International conference on information technology and systems	3	Springer
S7	International conference on computer and information technology	3	IEEE
S8	International journal of computer science and information security	3	Scopus
S9	International conference on intelligent data engineering and automated learning	3	Springer
S10	International conference on ICT in business industry and Government	3	IEEE
S11	Neural computing and applications	3	Springer
S12	Pakistan multitopic conference	3	IEEE
S13	Indian journal of science and technology	3	Scopus
S15	International conference of electronics, communication and aerospace technology	3	IEEE
S16	International research journal of engineering and technology	3	Scopus
S17, S22, S23	Procedia computer science	9	ScienceDirect
S18	Sustainability	3	MDPI
S19	Neural computing and applications	3	Springer
S21	Journal of supercomputing	3	Springer
S24	International conference of advances in computing, communications and informatics	3	IEEE
S25	Multimedia tools and applications	3	Springer
S26	International conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications	3	IEEE
S27	International conference on information and communication systems	3	IEEE
S28	International conference on Electrical, Computer & Telecommunication Engineering	3	IEEE
S29	International conference on service systems and service management	3	IEEE

e. Mean absolute percentage error (MAPE): MAPE is most widely used in KPI [28] for calculating the stock market forecasting. It is the sum [30] of absolute individual errors separated by the demand. It is a percentage error average.

Besides, selected subjects have used these performance parameters and their database to predict the stock market. Stock market exchange rates increase/decrease monthly or annually.

The Fig. 3 shows most of the selected studies use the accuracy performance parameter used to evaluate their model along with their dataset. But only 11% of the selected studies used the MAPE parameter for prediction.

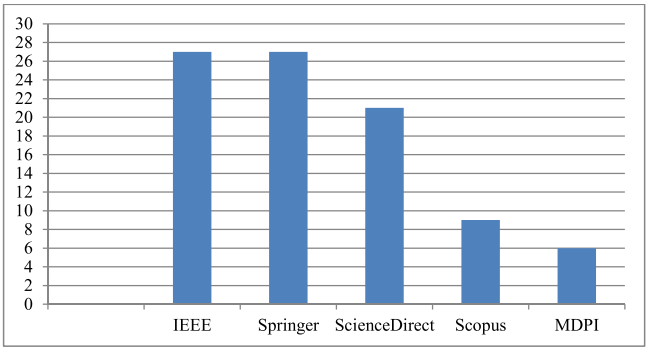

Fig. 4. Most frequently journals.

VI. RQ6: What are the most dominant journals for stock market forecasting?

To predict the stock market, the authors selected subjects based on research questions. These research questions are extracting the information from particular selected studies. But some of the selected studies have been published in conferences as well as in journals. So the authors have selected the selected studies to form dominant journals and conferences which are listed as:

Table 4 shows most of the journals/conferences by the selected studies. Fig. 4 shows the most dominant publisher frequently. Most of the expert systems & applications and Procedia computer science are the dominant journals.

5. Conclusion

This paper provides research on the various strategies used in stock market divisions divided by mathematical strategies and ML strategies. The purpose behind this survey is to classifying the current techniques related to adapted methodologies, used various datasets, performance matrices, and applying techniques, most dominant journals using 30 investigative articles. The techniques used in the stock market prediction are categorized in different ML algorithms. For improving the prediction accuracy, some of the selected studies use the hybrid approaches in the stock market. ANN and NN techniques are widely used approach for achieving the successful stock market prediction. These techniques can design for monitoring and surveillance of the whole stock market. The big challenge that stock market prediction face is that most current techniques cannot be identified with the aid of historical data on stocks. Hence stock markets are influenced by other factors such as policy decisions by government and consumer sentiments. In the future, we will strive to improve the system for making a reliable stock market system that is more reliable and accurate.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B.o. Qian, K. Rasheed, Stock market prediction with multiple classifiers, Appl. Intell. 26 (1) (2007) 25–33.
[2] N.A.A. Hussain, S.S.A. Ali, M.N.M. Saad, N. Nordin, 2016, December. Underactuated nonlinear adaptive control approach using U-model for multivariable underwater glider control parameters. In 2016 IEEE International Conference on Underwater System Technology: Theory and Applications (USYS) (pp. 19-25). IEEE.
[3] D. Shah, H. Isah, F. Zulkernine, Stock market analysis: A review and taxonomy of prediction techniques, Int. J. Financial Stud. 7 (2) (2019) 26.
[4] Pathak, Ashish, Nisha P. Shetty., 2019. Indian stock market prediction using ML and sentiment analysis. In Computational Intelligence in Data Mining, pp. 595-603. Springer, Singapore, pp. 595-603.
[5] J. Patel, S. Shah, P. Thakkar, K. Kotecha, Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques, Expert Syst. Appl. 42 (1) (2015) 259–268.
[6] Hernández-Álvarez, Myriam, Edgar A. Torres Hernández, Sang Guun Yoo., 2019. Stock Market Data Prediction Using ML Techniques.“ In International Conference on Information Technology & Systems, Springer, Cham, pp. 539-547.
[7] S. Banik, A.K. Khan, M. Anwer, 2012, December. Dhaka stock market timing decisions by hybrid machine learning technique. In 2012 15th International Conference on Computer and Information Technology (ICCIT) (pp. 384-389). IEEE.
[8] Ahmad Waqas, Analyzing different ML techniques for stock market prediction, Int. J. Comput. Sci. Inform. Sec. 12 (2014) 12–17.
[9] H. Yang, L. Chan, I. King. Support vector machine regression for volatile stock market prediction. In International Conference on Intelligent Data Engineering and Automated Learning 2002 Aug 12 (pp. 391-396). Springer, Berlin, Heidelberg.
[10] H.R. Patel, S.M. Parikh, D.N. Darji. Prediction model for the stock market using news based different Classification, Regression, and Statistical Techniques: (PMSMN). In2016 International Conference on ICT in Business Industry & Government (ICTBIG) 2016 Nov 18 (pp. 1-5). IEEE.
[11] K.J. Kim, W.B. Lee, Stock market prediction using artificial NN with optimal feature transformation, Neural Comput. Appl. 13 (3) (2004) 255–260.
[12] P.D. Yoo, M.H. Kim, T. Jan. Financial forecasting: advanced machine learning techniques in stock market analysis. In2005 Pakistan Section Multitopic Conference 2005 Dec 24 (pp. 1-7). IEEE.
[13] S.K. Chandar, M. Sumathi, S.N. Sivanandam, Prediction of the stock market price using a hybrid of wavelet transform and artificial neural network, Indian J. Sci. Technol. 9 (8) (2016) 1–5.
[14] Jonathan L. Ticknor, A bayesian regularized artificial neural network for stock market forecasting, Expert Syst. Appl. 40 (14) (2013) 5501–5506.
[15] A. Sharma, D. Bhuriya, U. Singh, 2017, April. Survey of stock market prediction using a machine learning approach. In 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA) (Vol. 2, pp. 506-509). IEEE.
[16] A. Rao, S. Hule, H. Shaikh, E. Nirwan, Daflapurkar PM. Survey: stock market prediction using statistical computational methodologies and artificial neural networks. Int. Res. J. Eng. Technol. (08). 2015.
[17] D. Enke, M. Grauer, N. Mehdiyev, Stock market prediction with multiple regression, fuzzy type-2 clustering, and neural networks, Procedia Comput. Sci. 1 (6) (2011) 201–206.
[18] H. Chung, K.S. Shin, Genetic algorithm-optimized long short-term memory network for stock market prediction, Sustainability 10 (10) (2018) 3765.
[19] X. Li, H. Xie, R. Wang, Y. Cai, J. Cao, F. Wang, X. Deng, Empirical analysis: stock market prediction via extreme learning machine, Neural Comput. Appl. 27 (1) (2016) 67–78.
[20] E. Chong, C. Han, F.C. Park, Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies, Expert Syst. Appl. 83 (2017) 187–205.
[21] X. Pang, Y. Zhou, P. Wang, W. Lin, V. Chang, An innovative neural network approach for stock market prediction, J. Supercomput. 76 (3) (2020) 2098–2118.
[22] M.M. Pai, A. Nayak, R.M. Pai, Prediction models for the Indian stock market, Procedia Comput. Sci. 89 (2016) 441–449.
[23] K. Zhang, G. Zhong, J. Dong, S. Wang, Y. Wang, Stock market prediction based on the generative adversarial network, Procedia Comput. Sci. 147 (2019) 400–406.
[24] S. Selvin, R. Vinayakumar, E.A. Gopalakrishnan, V.K. Menon, K.P. Soman. Stock price prediction using LSTM, RNN, and CNN-sliding window model. In2017 international conference on advances in computing, communications, and informatics (icacci) 2017 Sep 13 (pp. 1643-1647). IEEE.
[25] Ritika Singh, Shashi Srivastava, Stock prediction using deep learning, Multimedia Tools Appl. 76 (18) (2017) 18569–18584.
[26] M.R. Vargas, B.S. De Lima, A.G. Evsukoff. (2017, June). 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) (pp. 60-65). IEEE.
[27] K.A. Althelaya, E.S.M. El-Alfy, S. Mohammed. (2018, April). Evaluation of bidirectional lstm for short-and long-term stock market prediction. In 2018 9th international conference on information and communication systems (ICICS) (pp. 151-156). IEEE.
[28] R. Ray, P. Khandelwal, B. Baranidharan, 2018, December. A survey on stock market prediction using artificial intelligence techniques. In 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 594-598). IEEE.
[29] J. Li, H. Bu, J. Wu. (2017, June). Sentiment-aware stock market prediction: A deep learning method. In 2017 international conference on service systems and service management (pp. 1-6). IEEE.
[30] E. Guresen, G. Kayakutlu, T.U. Daim, Using artificial neural network models in stock market index prediction, Expert Syst. Appl. 38 (8) (2011) 10389–10397.