

7th International Conference on Computer Science and Computational Intelligence 2022

Machine learning approaches in stock market prediction: A systematic literature review

Latrisha N. Mintarya^a, Jeta N. M. Halim^a, Callista Angie^a, Said Achmad^{a,*}, Aditya Kurniawan^{a,*}

^aComputer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia

Abstract

Predicting the stock market has been done for a long time using traditional methods by analyzing fundamental and technical aspects. With machine learning, stock market predictions are made more accessible and more accurate. Various machine learning approaches have been applied in stock market prediction. This study aims to review relevant works about machine learning approaches in stock market prediction. To achieve this aim, we did a systematic literature review. This study review 30 studies regarding machine learning approaches/models in stock market prediction. Approaches that were used included neural networks and support vector machines. The result of this study is that neural networks are the most used model for stock market prediction. However, this does not mean that other models cannot be used for predicting the stock market.

© 2023 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the 7th International Conference on Computer Science and Computational Intelligence 2022

Keywords: Machine Learning; Machine Learning Approach; Machine Learning Model; Stock Market; Stock Market Prediction.

1. Introduction

Predicting the stock market is a complex task, and many factors influence the market. Therefore, before investing in a stock, investors perform two types of analysis. First is the fundamental analysis. For example, investors look at

* Corresponding author.

E-mail address: said.achmad@binus.edu; adkurniawan@binus.edu

the intrinsic value of stocks, the industry's performance, the economy, etc. Second, in technical analysis, investors look at stock valuations and statistics generated by market activity, such as past prices and volumes [1].

Nowadays, with technological advances, predicting the stock market can be done using many techniques that are not done manually. Many algorithms have been used for predicting the stock market, initially classical regression. However, the artificial neural network (ANN) and support vector machine (SVM) are two algorithms that are widely used. Each algorithm has its way of learning patterns and predicting stock market trends [1].

Vui et al. have discussed various ANN models that can be used to predict the stock market. They discuss these models, including feedforward neural networks (NN), backpropagation NN (BPNN), and hybrid NN. Although ANN has given acceptable results, much research is still being conducted to find new models to increase the accuracy of this approach [2].

The main object of this paper is to review relevant works about machine learning (ML) approaches in stock market prediction. This research conduct a systematic literature review (SLR) to achieve the aim. The research questions of this SLR are:

- Overview of ML in stock market prediction
- Understanding the trend of ML application in stock market prediction
- Frequently used models/approaches of ML in stock market prediction

We hope this paper will contribute to the stock market prediction area by providing a set of information and domain understanding which can support another researcher to expand the study and to develop a new machine learning approach for stock market prediction.

The paper is organized as follows: Section I is an introduction, Section II is a literature review, Section III explains the methodology used for this paper, Section IV consists of answers to the research questions and discussion, and Section V is the conclusion for this systematic literature review.

2. Literature Review

Many machine-learning approaches can be used to predict the stock market. Nayak et al. compared some supervised machine-learning approaches. Bailings et al. used random forest (RF), AdaBoost, kernel factory, NN, SVM, and k-nearest neighbors (KNN) to predict the stock market's direction for a year. Patel et al. discussed several machine-learning models, which are ANN, SVM, RF, and Naïve Bayes [1], as well as made stock market index predictions using ANN, SVM, and RF [3]. ANN is one of the most widely used models. ANN itself has a variety of approaches, as discussed by Vui et al. [2]. In addition, several researchers discussed approaches using those ANN models, like Bing et al. used BPNN to make predictions on the Shanghai Stock Exchange Composite Index [4]; Wensheng et al. compared nonlinear independent component analysis (NLICA) and BPNN for the Asia stock market [5]; Selvin et al. discussed approaches using long short-term memory (LSTM), recurrent neural network (RNN), and convolutional neural network-sliding window (CNN-sliding window) [6]; Chen et al. discussed the use of the LSTM model [7]; Nelson et al. compared LSTM with RF and multilayer perceptron (MLP) [8]; and G. et al. discussed approaches using MLP, RNN, LSTM and CNN [9].

A considerable amount of researches modified the ANN models, for instance, Vargas et al. used word-embedding and sentence-embedding with recurrent convolutional neural network (RCNN) [10]; Ding et al. used neural tensor network (NTN) and event-embedding CNN [11]; Chang et al. discussed evolving partially connected neural networks (EPCNN); Persio and Honchar discussed the accuracy of the ANN models and then proposed a wavelet CNN (W-CNN) model [8]; Qiu et al. made a genetic algorithm ANN (GA-ANN) model [12]; Ticknor used a modified bayesian ANN [13]; Moghar et al. used a RNN-based LSTM [14]; Roondiwala et al. made a model using LSTM and RNN [15]; Kang et al. used a generative adversarial networks (GAN) model combined with MLP and LSTM [16]; Akita et al. used the LSTM approach with paragraph vector [17]; Olivera et al. used a modified ANN to predict market behaviour and stock market trend [18]; and Li et al. compared Extreme Learning Machine (ELM) with SVM and BPN, the result was that kernelized ELM and SVM had higher precision than BPNN, and normal ELM [19].

Besides ANN, the SVM model is widely used to take a research approach. Ding et al. used SVM to predict stock market price from large-scale public news [20], and Hegazy. et al. also tried to compare the least-squares SVM (LS-

SVM) algorithm with particle swarm optimization (PSO) for the financial sector [21]. Some researchers modified the SVM model. For example, Lin et al. evaluated the performance of correlation-based SVM with quasi-linear SVM [22], and Ren et al. studied the accuracy of SVM with sentiment analysis [23].

Aside from the two models above (ANN and SVM), several other models, such as regression, can be used to predict the stock market. Regression has several models, as discussed by Sharma et al. [24]. In addition to regression, the use of support vector regression (SVR) that is optimized with a chaos-based firefly algorithm is discussed by Kazem et al.

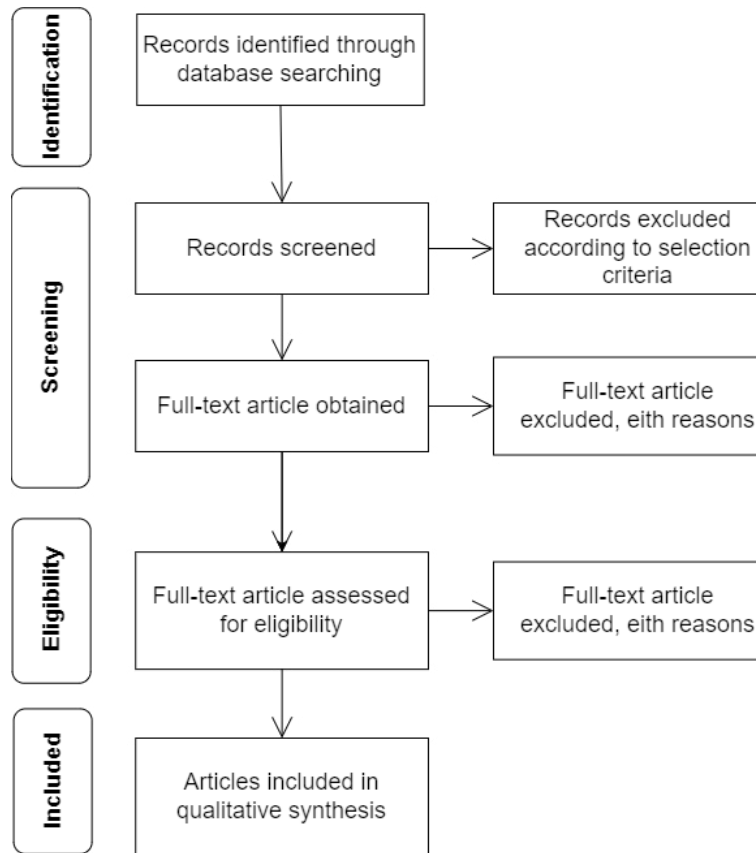


Fig. 1. PRISMA Flowchart

[25]. There is also a discussion on the use of the KNN model for stock market prediction by Alkhatib et al. They used the KNN algorithm and non-linear regression approach to perform stock price prediction for six major companies registered on the Jordanian stock exchange [26].

3. Methodology

The methodology used for this research is SLR. Research questions are formed first, and related data are collected from journal articles and literature reviews. To evaluate the papers, this research uses a modified PRISMA checklist. The PRISMA flowchart can be seen in Fig 1.

The filter criteria for this research are papers written in English, duplicates, full-texts, citations, and when the papers are released. The following terms were used to search for papers related to this study on Google Scholar with publishing year starting from 2012 :

- “machine learning” AND “for” AND “stock market” AND “prediction*” AND “approach*”,
- “machine learning” AND “for” AND “stock market” AND “prediction*” AND “model*”, and

- machine learning approach for stock market prediction.

4. Result and Discussion

The result presented in Table 1, Table 2, Fig 2 is based on the papers that have been chosen to be included in this study.

RQ1: Overview of ML in Stock Market Prediction

Technology development has been overgrown in the past few decades, and its use has spread into every aspect, field, and sector in this world, including the stock market. For example, the application of ML in the stock market includes predicting stock prices, predicting stock returns, and predicting the stock market's price direction movement. ML has changed the stock market in ways few could imagine a long time ago. ML has helped those in the stock market business by helping investors to predict stock market price indexes with the accuracy they need. Not only predict stock prices, but ML also helps to predict market behavior and even provides new models to help with data processing and predict the market indexes. It has irrevocably changed the way stock market predictions are made.

Many researchers have proposed various approaches and conducted experiments to increase the accuracy of predicting stock market prices compared to traditional methods.

Before, stock market prediction used traditional methods. There are two significant theories: the efficient market hypothesis (EMH) and the random walk theory. Then, there are Two conventional approaches. The approaches are technical analysis and fundamental analysis. Using these approaches can predict the stock market. However, it could be more effective now because computers can process large amounts of data faster and more precisely. [2]

RQ2: Understanding the trend of ML application in Stock Market Prediction

The distribution of the 30 papers by publication year used for this SLR can be seen in Table I and visualized in Fig.2 As displayed, research on the use of ML in stock market prediction has increased by 133.3% in 2013 compared to 2012, while a decrease of 85.7% happened in 2014. From 2015 to 2017, a stable amount of research was conducted. There has been another decrease in 2018. However, the amount after has been relatively similar.

There has been more interest in using ML in predicting the stock market during the early- to mid-2010s while it has dwindled during the late-2010s. If the distribution of papers is divided into three periods, each consisting of three years, then the time in which the most papers were published was 2015 to 2017, with 15 papers, followed by 2012 to 2014, with 11 papers, and lastly, 2018 to 2020 with only four papers.

Table 1. Distribution of Papers by Year

Year	Amount of Papers	Reference
2012	3	[27, 4, 5]
2013	7	[2, 22, 13, 18, 25, 21, 26]
2014	1	[20]
2015	5	[11, 7, 1, 3, 28]
2016	5	[17, 29, 30, 12, 19]
2017	5	[24, 6, 8, 15, 10]
2018	1	[9]
2019	2	[23, 16]
2020	1	[14]

RQ3: Frequently used Models/Approach of ML in Stock Market Prediction

Table 2 shows that the most used method for stock market prediction is neural networks, followed by its derivative, LSTM. The use of NN constitutes 46.3% of the total number of various approaches, SVM constitutes 17.1%, LSTM constitutes 19.5%, and other approaches constitute 17.1%. NN usage is 137.5% more than LSTM, 171.43% more than SVM, and 171.43% more than other models.

As seen in Fig. 2b, the use of LSTM became more frequent in 2015, while the use of SVM was mostly before 2016. On the other hand, NN is relatively frequent throughout the ten years. Other methods, such as RF and KNN, are less frequently used in stock market prediction.

NNs can comprehend non-linear data trends, generalize it, and adjust to the relationships and data patterns between input and output. This makes NNs convenient for predicting the stock market and able to predict more accurately compared to other approaches [2]. For these reasons, NNs are the most frequently used model in stock market prediction.

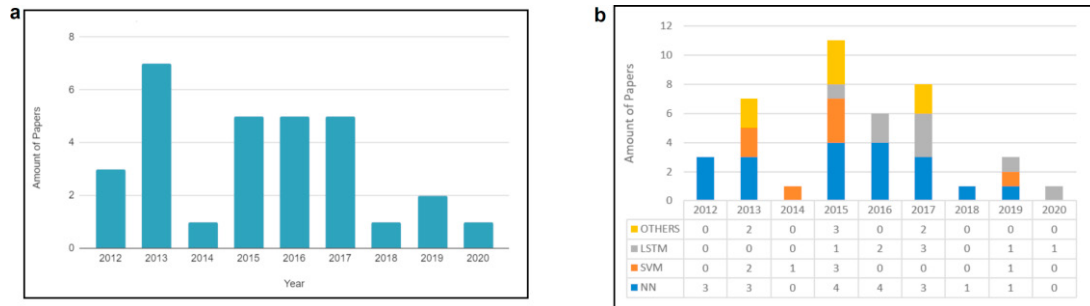


Fig. 2. (a) Distribution of Papers by Year; (b) Distribution of ML Models/Approaches by Year.

Table 2. Distribution of ML Approaches by Year

		ML Models Approaches			
		NN	SVM	LSTM	OTHERS
Year	2012	[27, 4, 5]	-	-	-
	2013	[2, 13, 18]	[22, 21]	-	[25, 26]
	2014	-	[20]	-	-
	2015	[11, 1, 3]	[1, 3, 28]	[7]	[3, 28]
	2016	[29, 30, 12, 19]	-	[17, 29]	-
	2017	[6, 8, 10]	-	[6, 8, 15]	[24, 8]
	2018	[9]	-	-	-
	2019	[16]	[23]	[16]	-
	2020	-	-	[14]	-
Total		19	7	8	7

4.1. Discussion

In this paper, we compared and reviewed the most used approach among numerous ML approaches. This research shows that the method used most often is NN, with a reasonably significant difference compared to other methods. Surprisingly, the year with the highest number of research papers discussing NNs is 2015 and 2016, each with four papers. In contrast, 2013, which has the highest number of published papers in this study, only has 3 discussing NNs. NN's effectiveness in predicting the stock market is proven by one of the researchers predicting the daily highest and lowest price, as well as the closing price of the Shanghai Stock Exchange in the shortest time [4]. The NN model may be beneficial in further research. LSTM is the second-highest approach to be used. Despite this fact, LSTM's use only started in 2015 as the use of SVM is diminishing. In 2017, 3 papers studied LSTM, making it the year with the highest number of papers researching LSTM. In the same year, the number of papers discussing LSTM equals the number of papers discussing NN.

5. Conclusion

In conclusion, ML helps investors to predict stock market price indexes more efficiently than traditional means because computers can now process large amounts of data faster and more precisely. The ML trend for stock mar-

ket prediction peaked in 2013, even though the 2015-2017 period yielded the most, with 15 papers. Many machine learning approaches can provide stock market prediction at high accuracy. Some propose novel ML models to achieve higher accuracy. Most ML models are NNs, although SVM and LSTM are also used. We hope these findings will contribute to developing researchers' understanding of what approach could be used to predict the stock market.

References

- [1] 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.
- [2] 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.
- [3] Patel J, Shah S, Thakkar P, Kotecha K. Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*. 2015;42(4):2162-72.
- [4] 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.
- [5] 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.
- [6] Selvin S, Vinayakumar R, Gopalakrishnan E, Menon VK, Soman K. Stock price prediction using LSTM, RNN and CNN-sliding window model. In: 2017 international conference on advances in computing, communications and informatics (icacci). IEEE; 2017. p. 1643-7.
- [7] 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.
- [8] 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.
- [9] Hiransha M, Gopalakrishnan EA, Menon VK, Soman K. NSE stock market prediction using deep-learning models. *Procedia computer science*. 2018;132:1351-62.
- [10] 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.
- [11] 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. .
- [12] 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.
- [13] Ticknor JL. A Bayesian regularized artificial neural network for stock market forecasting. *Expert systems with applications*. 2013;40(14):5501-6.
- [14] Moghar A, Hamiche M. Stock market prediction using LSTM recurrent neural network. *Procedia Computer Science*. 2020;170:1168-73.
- [15] Roondiwala M, Patel H, Varma S. Predicting stock prices using LSTM. *International Journal of Science and Research (IJSR)*. 2017;6(4):1754-6.
- [16] 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.
- [17] 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.
- [18] De Oliveira FA, Nobre CN, Zárate 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.
- [19] Li X, Xie H, Wang R, Cai Y, Cao J, Wang F, et al. Empirical analysis: stock market prediction via extreme learning machine. *Neural Computing and Applications*. 2016;27(1):67-78.
- [20] Ding X, Zhang Y, Liu T, Duan J. Using structured events to predict stock price movement: An empirical investigation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*; 2014. p. 1415-25.
- [21] Hegazy O, Soliman O, Salam M. A Machine Learning Model for Stock Market. *arXiv Prepr arXiv1402.7351*:17-23.
- [22] 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.
- [23] Ren R, Wu DD, Liu T. Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Systems Journal*. 2018;13(1):760-70.
- [24] Sharma A, Bhuriya D, Singh U. Survey of stock market prediction using machine learning approach. In: 2017 International conference of electronics, communication and aerospace technology (ICECA). vol. 2. IEEE; 2017. p. 506-9.
- [25] Kazem A, Sharifi E, Hussain FK, Saberi M, Hussain OK. Support vector regression with chaos-based firefly algorithm for stock market price forecasting. *Applied soft computing*. 2013;13(2):947-58.
- [26] Alkhatib K, Najadat H, Hmeidi I, Shatnawi MKA. Stock price prediction using k-nearest neighbor (kNN) algorithm. *International Journal*

ofBusiness, Humanities and Technology. 2013;3(3):32-44.

- [27] Chang PC, et al. A novel model by evolving partially connected neural network for stock price trend forecasting. *Expert Systems withApplications*. 2012;39(1):611-20.
- [28] 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.
- [29] Nayak A, Pai MM, Pai RM. Prediction models for Indian stock market. *Procedia Computer Science*. 2016;89:441-9.
- [30] Di Persio L, Honchar O. Artificial neural networks architectures for stock price prediction: Comparisons and applications. *International journalof circuits, systems and signal processing*. 2016;10(2016):403-13.