



Stock Market Prediction using Machine Learning: A Systematic Literature Review

Siddhartha Vadlamudi

Quixey Inc., Vintech Solutions, Comcast, Philadelphia, USA

*E-mail for correspondence: vadlamudisiddhartha@gmail.com

Received: July 05, 2017

Accepted: Nov 27, 2017

Published: Dec 26, 2017

Source of Support: Nil

No Conflict of Interest: Declared

ABSTRACT

Different machine learning algorithms are discussed in this literature review. These algorithms can be used for predicting the stock market. The prediction of the stock market is one of the challenging tasks that must have to be handled. In this paper, it is discussed how the machine learning algorithms can be used for predicting the stock value. Different attributes are identified that can be used for training the algorithm for this purpose. Some of the other factors are also discussed that can have an effect on the stock value.

Keywords: Stock Market, Machine Learning, Predictive Algorithms

INTRODUCTION

The stock market is the place where the buyers and sellers come for the purpose of buying the stocks and selling the stocks of a company. With the passage of time, more and more people are getting interested in the stock market. The increasing interest of the people makes this topic more important for the research. Machine learning provides the different predictive algorithms that can be very helpful in this domain. Different sectors are using the machine learning for their improvement. The use of machine learning enables the computers to learn without the involvement of any external program (Kumar, Krishna, Neelakanteswara, & Basha, 2020). These algorithms of the machine learning are first trained through a data set. This data set is called the training data set. The algorithm then makes the predictions according to the instructions that are given in the training data set.

People now want to invest their money in the stock market on the basis of the prediction (Sharma, Bhuriya, & Singh, 2017). From the past few years, the stock market prediction is one of the challenging tasks for the researchers (Gurav & Sidnal, 2018). The stock market prediction is the process of predicting the value of the stock of a company i.e. whether the value will be increased in future or it will be decreased (Hegazy, Soliman, & Salam, 2014). People buy the stocks, when there is the chance of an increase in the future value. This

creates the need of using the machine learning for predicting the future value of the stocks of a company.

The work is presented by the various researchers to predict the stock value. For the prediction purpose, the researchers created the different models using the predictive algorithms provided by the machine learning. Because machine learning has the ability to successfully forecast the stock price (Berry & Linoff, 2004; Vadlamudi, 2016). Researchers in (Shen, Jiang, & Zhang, 2012) used the Support vector machine (SVM) to forecast the stock value of the next coming day. The economic growth of a country depends on the stock market. According to the researchers, the economic growth of a county goes down when the stock market is falling (Fama, 1965). Researchers (Sheta, 2006) developed the fuzzy models for the prediction of the stock market.

This literature review is written to give a detailed summary of the different machine learning algorithms that are used by the various researchers. The work that is presented by various researchers will be described in this literature review. For the purpose of better understanding, the research question will be created and answered. The literature review is based on the papers from the various large databases. The further parts of this paper are formulated as: section 2 is about the methodology, section 3 show the research questions of this paper, in section 4 the search process is described, section 5 is about the results and discussion, 6th section shows the conclusion.



METHODOLOGY

The methodology that is used for writing this literature review is the Systematic literature review (SLR). This is the methodology that helps in writing the literature review while answering the particular research questions. The questions of this literature review will be answered based on the word presented by the researchers in the past.

Inclusion and Exclusion

The google scholar contains a number of research papers, conference papers, etc. The selection of the most relevant papers is necessary as well as a challenging task. For the inclusion of the papers in this literature review, there was an inclusion criteria. This criteria ensures that the papers are written in the English language, completely available to study and are relevant to stock market prediction. All the papers that do not fulfill this inclusion criteria were excluded.

Quality Assessment

To write a good quality literature review it is necessary that the included papers must have the good quality. The quality of the papers was assessed before including them in this literature review. To assess the quality, the contribution of each research paper author was analyzed.

RESEARCH QUESTION

As it is mentioned above, that the systematic literature review must have to answer one or more research questions. In this part, the research questions of this literature review are created and presented. Below are the research questions of this systematic literature review.

RQ 1: "Why the stock market prediction is necessary and how it helps the investors in taking the right decision?"

RQ 2: "What are the different attributes and factors that can be used for predicting the stock market?"

RQ 3: "What are the different machine learning predictive algorithms that are used by the researchers for predicting the stock market?"

SEARCH PROCESS

Different important and famous digital libraries were used for collecting the right papers for this literature review. These digital libraries include the Elsevier, ACM, IEEE, etc. The papers that were collected were the 550. But not all of these papers were appropriate to be added to this literature review. For finding the most appropriate papers, a search process was followed. According to this search process, the papers were first included or excluded on the basis of their title. In the next phase of search process, the abstract of the remaining papers was studied. A lot of papers were excluded in this phase. In the last phase of the search process, complete papers were studied to be included or excluded. As a result, the remaining papers were 45.

RESULT AND DISCUSSION

Nowadays, predictions are made in different sectors to improve the progress. Making the prediction is a challenging task (D. Das, Uddin, & applications, 2013). The stock market prediction is one of the important and difficult activities because of the dynamic change in the values of the stock. This is one of the big issues in the finance field (Yoo, Kim, & Jan, 2005). This prediction can be done on the basis of the previous records of the stock market. Machine learning algorithms are very helpful in this regard. Researchers (Reddy & Technology, 2018) used the SVM algorithm for predicting the stock value. The programming language that is used by the researchers was python. They trained the algorithm with the previous data set of the stock market to predict the future value of the stock. Researchers (Huang, Nakamori, Wang, & research, 2005) also used the SVM for the purpose of stock market prediction. They compared its results with the other classification methods and they found this algorithm is better than the different classification methods. Researchers (Qian & Rasheed, 2007) trained the different machine learning algorithms for predicting the stock market. The achieved accuracy was the 65 percent that can be improved by using the more effective attributes and training the algorithm with a large data set. The answers to the different created research questions are given below.

RQ 1: "Why the stock market prediction is necessary and how it helps the investors in taking the right decision?"

People are getting more and more interested in the stock market. They want to invest their money to get the benefit. But there is also the chance of losing the invested money or a loss instead of the benefit. The person will face a loss if he will invest in a stock whose future value will be going to be decreased. There is a need for the prediction about the future value of the stock. This prediction will help in earning the more money (Deepak, Uday, Malathi, & Mathematics, 2017). The interest of the investors is increasing towards the stock market prediction (Khan et al., 2020). Some of the stock markets are difficult to predict. According to the researchers (Khan et al., 2020), random forest is the best classifier for this kind of prediction because the accuracy provided by this classifier is 83%.

Different kinds of events can affect the stock price. This makes the stock market prediction more challenging and difficult (Choudhry, Garg, & Technology, 2008). The investors want to gain more and more benefits but the human mind cannot give the correct prediction about the stock price. Machine learning must have to be used for this prediction purpose. According to the researchers (A.-S. Chen, Leung, Daouk, & Research, 2003; Vadlamudi, 2015), the more accurate prediction about the future stock value will result in more profit. This shows the importance and need of the stock market prediction for the investors. Researchers (Parmar et al., 2018), also said that the seller and broker can gain more benefits if the stock prices is predicted accurately. This prediction helps the investors to decide

when they should buy at the lowest price and when to sell at the highest prices (Boone, Giorno, Richardson, & DEPARTMENT, 1998).

RQ 2: “What are the different attributes and factors that can be used for predicting the stock market?”

One of the highly researched topics is the search for a machine learning model that can predict the stock price with more accuracy (Henrique, Sobreiro, & Kimura, 2019). To train the machine learning algorithm, there must be a previous data set. This data set can contain the data in the form of various variables or attributes. A data set is required for the training of a machine learning algorithm so that it can be able to predict the future value of the stock. This data set can be based on a number of attributes. Researchers have trained their models with different attributes. However, some of the attributes are commonly used. A piece of news can also affect the stock price. News is considered as positive if it is the reason behind the stock value (Shah & Spring, 2007). The news can also have a negative impact or no impact. Researchers (Raza, 2017) created the machine learning model that is based on the different factors. Some of these factors are the prices of fuel, foreign exchange, commodity, foreign exchange and the news. Researchers studied the each factor independently to check its importance i.e. why this factor is helpful in predicting the stock price. They came to know that the change in the prices of fuel has a huge impact on the market performance. Researchers (Lee, Cho, Kwon, & Sohn, 2019) applied the network indicators as an input for predicting the stock value. These indicators proved more important during the crisis period of the market. According to the researchers (Usmani, Adil, Raza, & Ali, 2016), there are the six different factors that affect the stock market. The value of the stock market change according to these factors. These are the “Market History, The News, General Public Mood, Commodity Price, Interest Rate, Foreign Exchange”. Researchers (Kohara, Ishikawa, Fukuhara, Nakamura, & Management, 1997), combined the event knowledge with the machine learning algorithm to predict the stock price. The experiment has shown that the combination of event knowledge with other attributes increased the accuracy by five percent. A positive event can be the reason behind the increase in stock price. The event information can be extracted from the news. As shown in the below figure:



Figure 1: “Event Knowledge Extraction”

Various researchers said that combining the event information with neural network algorithm can improve the prediction accuracy (Gabriel Pui Cheong Fung, Yu, & Lam, 2002) (Hong & Han, 2002) (Hong & Han, 2004) (Kohara, 2003) (Mittermayer, 2004) (G Pui Cheong Fung, Yu, & Lam, 2003). The below table shows the different commonly used attributes.

Table 1: “Commonly used attributes”

Oil Rates	Gold Rates
FEX	SMA
Silver Rates	ARIMA
News	Twitter

Different macroeconomic factors can also influence the trend of the stock. Researchers (Hu, Zhu, & Tse, 2013) highlighted some of these macroeconomic factors. These include the “consumer spending, consumer investment, unemployment rate, inflation rate, federal funds rate”.

RQ 3: “What are the different machine learning predictive algorithms that are used by the researchers for predicting the stock market?”

Different algorithms of machine learning are used by the various researchers for the stock market prediction. These algorithms are trained and applied to the different data sets. Different attributes are used by the researchers for improving the prediction accuracy. The prediction accuracy is very important for the investors. The accuracy of the algorithm can save the investors from the loss and helps them to earn more profit. Below is the table that shows the various algorithms that are used by the researchers.

Table 2: “Algorithms for predicting stock market”

Linear Regression	Neural Network
Genetic Algorithm	Support Vector Machine (SVM)
K-Nearest Neighbor	Random Forest

Linear Regression

This is a machine learning algorithm that belongs to the supervised category. This algorithm predicts the values that are in a range instead of predicting the category. For example, it can predict the sales, prices etc. Researchers (Bhuriya, Kaushal, Sharma, & Singh, 2017) used this algorithm for predicting the stock market. The researchers said that the prediction of daily stock value is a serious challenge that must be handled. With the prediction of this algorithm, the investors can invest the money with more confidence.

Neural Network

This is the unsupervised machine learning algorithm that consists of the number of nodes and each node performs some computation. Various researchers have used this algorithm for the stock price prediction. Researchers (Naeini, Tarehian, & Hashemi, 2010) used this algorithm for predicting the stock market. Researchers (Adebiyi, Ayo, Adebiyi, Otokiti, & Sciences, 2012) trained this algorithm with different variables to predict the stock value so that the investors can take the correct decisions. This algorithm can also be trained for this purpose, through the news from different social media platforms (W. Chen, Zhang, Yeo, Lau, & Lee, 2017).

Genetic Algorithm

Researchers (Bonde & Khaled, 2012) used this algorithm for predicting whether the value of the stocks is going to

be increased or decreased in the next coming day. They used the six attributes to train the algorithm for this purpose. Researchers (Naik, Ramesh, Manjula, Govardhan, & Systems, 2012) also used this algorithm for predicting the price movement of the stocks.

SVM

Researchers (Lin, Guo, & Hu, 2013) used an approach based on this algorithm to predict the future trend of the stock. Researchers (Y. Chen & Hao, 2017) also used this algorithm for the prediction of stock market indices. Researchers (S. P. Das & Padhy, 2012) compared the two different algorithms and they found SVM is better in predicting the future value of the stocks.

K-Nearest Neighbor

This algorithm can also help the investors in investing the money at the right time by predicting the stock value at an initial stage (Alkhatib, Najadat, Hmeidi, Shatnawi, & Technology, 2013). Researchers (Imandoust, Bolandraftar, & Applications, 2013) said that the KNN can be used in the finance sector for predicting the stock values and to decide the best time to buy and sell the stocks.

Random Forest

This algorithm can be helpful to minimize the risk while investing in the stock market (Khaidem, Saha, & Dey, 2016). The investors can invest the more money if there is a minimal risk of investment. Researchers (GHOLAMIAN & DAVOODI, 2018) achieved 64% accuracy in predicting the future stock value by using this algorithm.

Below is the diagram that shows how a machine learning algorithm is trained and used for predicting the future value of the stock.

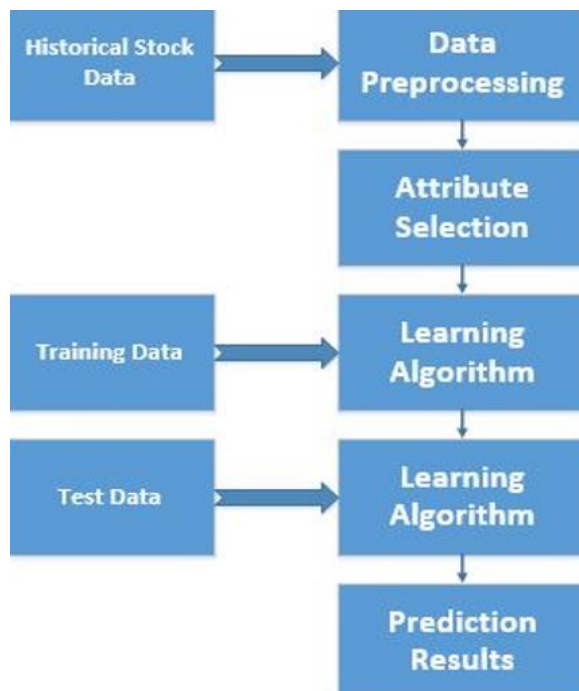


Figure 2: Algorithm training for stock market prediction

CONCLUSION

Prediction of the stock value is very important for investors. With the help of an accurate prediction, the investors can gain more and more profit. Taking the decision about when to buy and when to sell is very hard for the investors. The algorithm through the machine learning algorithms can be very helpful in making this decision. Investors can take the risk of investment easily when they have an idea about the future value of the stocks. Different researchers have worked for predicting the future values of the stock. This work is discussed in this paper so that the reader should not have to read all the papers.

REFERENCES

- Adebiyi, A. A., Ayo, C. K., Adebiyi, M. O., Otokiti, S. O. J. J. o. E. T. i. C., & Sciences, I. (2012). Stock price prediction using neural network with hybridized market indicators. 3(1), 1-9.
- Alkhatib, K., Najadat, H., Hmeidi, I., Shatnawi, M. K. A. J. I. J. o. B., Humanities, & Technology. (2013). Stock price prediction using k-nearest neighbor (kNN) algorithm. 3(3), 32-44.
- Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*: John Wiley & Sons.
- Bhuriya, D., Kaushal, G., Sharma, A., & Singh, U. (2017). *Stock market prediction using a linear regression*. Paper presented at the 2017 international conference of electronics, communication and aerospace technology (ICECA).
- Bonde, G., & Khaled, R. (2012). *Stock price prediction using genetic algorithms and evolution strategies*. Paper presented at the Proceedings of the International Conference on Genetic and Evolutionary Methods (GEM).
- Boone, L., Giorno, C., Richardson, P. J. W. P.-O. F. E. C., & DEPARTMENT, D. E. (1998). *Stock Market Fluctuations and Consumption Behaviour: Some Recent Evidence* Economics Department Working Papers No. 208.
- Chen, A.-S., Leung, M. T., Daouk, H. J. C., & Research, O. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. 30(6), 901-923.
- Chen, W., Zhang, Y., Yeo, C. K., Lau, C. T., & Lee, B. S. (2017). *Stock market prediction using neural network through news on online social networks*. Paper presented at the 2017 international smart cities conference (ISC2).
- Chen, Y., & Hao, Y. J. E. S. w. A. (2017). A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. 80, 340-355.
- Choudhry, R., Garg, K. J. W. A. o. S., Engineering, & Technology. (2008). A hybrid machine learning system for stock market forecasting. 39(3), 315-318.
- Das, D., Uddin, M. S. J. I. j. o. a. i., & applications. (2013). Data mining and neural network techniques in stock market prediction: A methodological review. 4(1), 117.
- Das, S. P., & Padhy, S. J. I. J. o. C. A. (2012). Support vector machines for prediction of futures prices in Indian stock market. 41(3).
- Deepak, R. S., Uday, S. I., Malathi, D. J. I. J. o. P., & Mathematics, A. (2017). Machine learning approach in stock market prediction. 115(8), 71-77.

- Fama, E. F. J. T. j. o. B. (1965). The behavior of stock-market prices. 38(1), 34-105.
- Fung, G. P. C., Yu, J. X., & Lam, W. (2002). *News sensitive stock trend prediction*. Paper presented at the Pacific-Asia Conference on Knowledge Discovery and Data Mining.
- Fung, G. P. C., Yu, J. X., & Lam, W. (2003). *Stock prediction: Integrating text mining approach using real-time news*. Paper presented at the 2003 IEEE International Conference on Computational Intelligence for Financial Engineering, 2003. Proceedings.
- GHOLAMIAN, E., & DAVOODI, S. M. R. (2018). Predicting the direction of stock market prices using random forest.
- Gunav, U., & Sidnal, N. (2018). Predict Stock Market Behavior: Role of Machine Learning Algorithms. In *Intelligent Computing and Information and Communication* (pp. 383-394): Springer.
- Hegazy, O., Soliman, O. S., & Salam, M. A. J. a. p. a. (2014). A machine learning model for stock market prediction.
- Henrique, B. M., Sobreiro, V. A., & Kimura, H. J. E. S. w. A. (2019). Literature review: Machine learning techniques applied to financial market prediction. 124, 226-251.
- Hong, T., & Han, I. J. E. S. (2004). Integrated approach of cognitive maps and neural networks using qualitative information on the World Wide Web: the KBNMiner. 21(5), 243-252.
- Hong, T., & Han, I. J. E. s. w. a. (2002). Knowledge-based data mining of news information on the Internet using cognitive maps and neural networks. 23(1), 1-8.
- Hu, Z., Zhu, J., & Tse, K. (2013). *Stocks market prediction using support vector machine*. Paper presented at the 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering.
- Huang, W., Nakamori, Y., Wang, S.-Y. J. C., & research, o. (2005). Forecasting stock market movement direction with support vector machine. 32(10), 2513-2522.
- Imandoust, S. B., Bolandraftar, M. J. I. J. o. E. R., & Applications. (2013). Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. 3(5), 605-610.
- Khaidem, L., Saha, S., & Dey, S. R. J. a. p. a. (2016). Predicting the direction of stock market prices using random forest.
- Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., Alfakeeh, A. S. J. J. o. A. I., & Computing, H. (2020). Stock market prediction using machine learning classifiers and social media, news. 1-24.
- Kohara, K. (2003). *Selective-learning-rate approach for stock market prediction by simple recurrent neural networks*. Paper presented at the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems.
- Kohara, K., Ishikawa, T., Fukuhara, Y., Nakamura, Y. J. I. S. i. A., Finance, & Management. (1997). Stock price prediction using prior knowledge and neural networks. 6(1), 11-22.
- Kumar, V. U., Krishna, A., Neelakanteswara, P., & Basha, C. Z. (2020). *Advanced Prediction of Performance of a Student in an University using Machine Learning Techniques*. Paper presented at the 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC).
- Lee, T. K., Cho, J. H., Kwon, D. S., & Sohn, S. Y. J. E. S. w. A. (2019). Global stock market investment strategies based on financial network indicators using machine learning techniques. 117, 228-242.
- Lin, Y., Guo, H., & Hu, J. (2013). *An SVM-based approach for stock market trend prediction*. Paper presented at the The 2013 international joint conference on neural networks (IJCNN).
- Mittermayer, M.-A. (2004). *Forecasting intraday stock price trends with text mining techniques*. Paper presented at the 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the.
- Naeini, M. P., Tareman, H., & Hashemi, H. B. (2010). *Stock market value prediction using neural networks*. Paper presented at the 2010 international conference on computer information systems and industrial management applications (CISIM).
- Naik, R. L., Ramesh, D., Manjula, B., Govardhan, A. J. C. E., & Systems, I. (2012). Prediction of stock market index using genetic algorithm. 3(7), 162-171.
- Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. (2018). *Stock market prediction using Machine Learning*. Paper presented at the 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC).
- Qian, B., & Rasheed, K. J. A. I. (2007). Stock market prediction with multiple classifiers. 26(1), 25-33.
- Raza, K. (2017). *Prediction of Stock Market performance by using machine learning techniques*. Paper presented at the 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT).
- Reddy, V. K. S. J. I. R. J. o. E., & Technology. (2018). Stock market prediction using machine learning. 5(10), 1033-1035.
- Shah, V. H. J. F. o. M. L., & Spring. (2007). Machine learning techniques for stock prediction. 1(1), 6-12.
- Sharma, A., Bhuriya, D., & Singh, U. (2017). *Survey of stock market prediction using machine learning approach*. Paper presented at the 2017 International conference of electronics, communication and aerospace technology (ICECA).
- Shen, S., Jiang, H., & Zhang, T. J. D. o. E. E., Stanford University, Stanford, CA. (2012). Stock market forecasting using machine learning algorithms. 1-5.
- Sheta, A. (2006). *Software effort estimation and stock market prediction using takagi-sugeno fuzzy models*. Paper presented at the 2006 IEEE International Conference on Fuzzy Systems.
- Usmani, M., Adil, S. H., Raza, K., & Ali, S. S. A. (2016). *Stock market prediction using machine learning techniques*. Paper presented at the 2016 3rd international conference on computer and information sciences (ICCOINS).
- Yoo, P. D., Kim, M. H., & Jan, T. (2005). *Machine learning techniques and use of event information for stock market prediction: A survey and evaluation*. Paper presented at the International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06).
- Vadlamudi, S. (2015). Enabling Trustworthiness in Artificial Intelligence - A Detailed Discussion. *Engineering International*, 3(2), 105-114. <https://doi.org/10.18034/ei.v3i2.519>

Vadlamudi, S. (2016). What Impact does Internet of Things have on Project Management in Project based Firms?. *Asian Business Review*, 6(3), 179-186.
<https://doi.org/10.18034/abr.v6i3.520>

--0--

SOCIAL SCIENCE RESEARCH NETWORK

2171 Monroe Avenue, Suite 203, Rochester, NY 14618, USA

<http://www.ssrn.com/en/>

AJTP Link: <http://www.ssrn.com/link/American-Journal-Trade-Policy.html>