

Stock Market Prediction

Subjects: **Computer Science, Artificial Intelligence**

Contributor: Nusrat Rouf , Arjun Somanathan

Stock Market Prediction (SMP) is an example of time-series forecasting that promptly examines previous data and estimates future data values. Financial market prediction has been a matter of worry for analysts in different disciplines, including economics, mathematics, material science, and computer science. Driving profits from the trading of stocks is an important factor for the prediction of the stock market.

generic review

machine learning

stock market prediction

support vector machine

1. Introduction

According to ^[1], there exist two main traditional approaches to the analysis of the stock markets: (1) fundamental analysis and (2) technical analysis.

Technical analysis is the study of stock prices to make a profit, or to make better investment decisions ^[2]. Technical analysis predicts the direction of the future price movements of stocks based on their historical data, and helps to analyze financial time series data using technical indicators to forecast stock prices. Meanwhile, it is assumed that the price moves in a trend and has momentum ^[3]. Technical analysis uses price charts and certain formulae, and studies patterns to predict future stock prices; it is mainly used by short-term investors. The price would be considered high, low or open, or the closing price of the stock, where the time points would be daily, weekly, monthly, or yearly. Dow theory puts forward the main principles for technical analysis, which are that the market price discounts everything, prices move in trends, and historic trends usually repeat the same patterns ^[4]. There are several technical indicators, such as the Moving Average (MA), Moving Average Convergence/Divergence (MACD), the Aroon indicator, and the money flow index, etc. The evident flaws of technical analysis as per ^[5] are that expert's opinions define rules in technical analysis, which are fixed and are reluctant to change. Various parameters that affect stock prices are ignored.

The prerequisite is to overcome the deficiencies of fundamental and technical analysis, and the evident advancement in the modelling techniques has motivated various researchers to study new methods for stock price prediction. A new form of collective intelligence has emerged, and new innovative methods are being employed for stock value forecasting. The methodologies incorporate the work of machine learning algorithms for stock market analysis and prediction.

One of the phenomena of current times that is changing the world is the global availability of the internet. The most-used platforms on the internet are social media. It is estimated that social media users all over the world will

number around 3.07 billion [6]. There is a high association between stock prices and events related to stocks on the web. The event information is extracted from the internet to predict stock prices; such an approach is known as event-driven stock prediction [7]. Through social networks, people generate tremendous amounts of data that is filled with emotions. Much of this data is related to user perceptions and concerns [8]. Sentiment analysis is a field of study that deals with the people’s concerns, beliefs, emotions, perceptions, and sentiments towards some entity [9][10]. It is the process of analyzing text corpora, e.g., news feeds or stock market-specific tweets, for stock trend prediction. The Stock Twits, Twitter, Yahoo Finance, and so on are well-known platforms used for the extraction of sentiments. There is a significant importance of using sentimental data for enhancing the prediction of volatility in the stock market. The ‘Wisdom of Crowds’ and sentiment analysis generate more insights that can be used to increase the performance in various fields, such as box office sales, election outcomes, SMP, and so on [11]. This suggests that a good decision can be made by taking the opinions and insights of large groups of people with varied types of information [12]. The information generated through social media allows us to explore vast and diverse opinions. Exploring sentiments from social media in addition to numeric time-series stock data would enhance the accuracy of the prediction. Using time-series data as well as social media data would intensify the prediction accuracy. Different approaches and techniques have been proposed over time to anticipate stock prices through numerous methodologies, thanks to the dynamic and challenging panorama of stock markets [13].

2. Generic Scheme for SMP

Figure 1 describes the generic process involved in SMP. The process starts with the collection of the data, and then pre-processing that data so that it can be fed to a machine learning model. The prediction models generally use two types of data: market and textual data. The literature of both types is discussed in the following section. The next section classifies the previous studies based on the type of data used. Furthermore, the next section surveys the previous studies based on the various data-preprocessing approaches applied. Moreover, the literature is further surveyed based on the machine learning algorithms used by different systems.

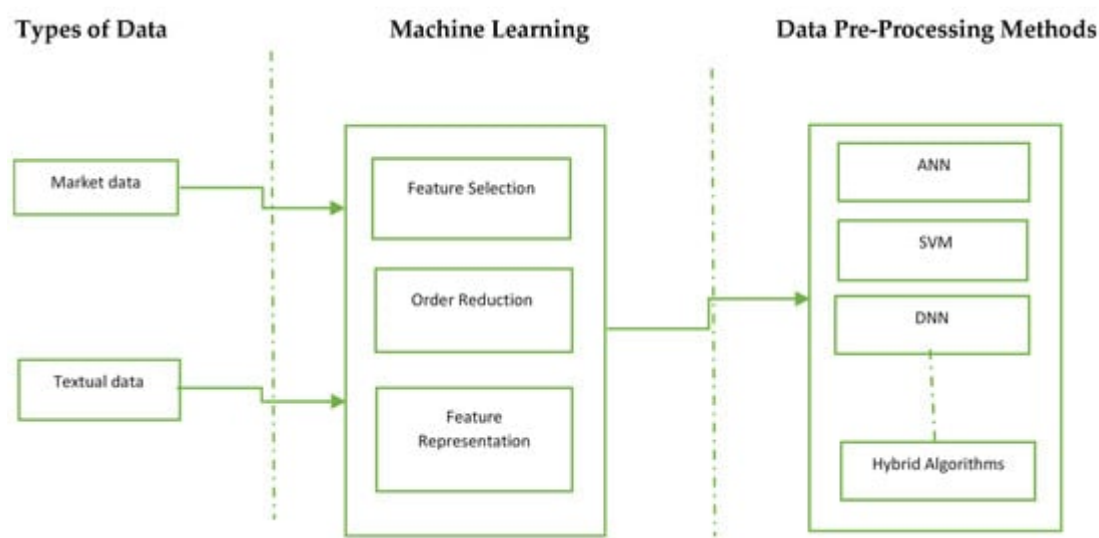


Figure 1. Generic Scheme for SMP (Stock Market Prediction).

3. Overfitting

One of the most well-known and challenging issues in machine learning models is overfitting. In this phenomenon, the model tries too hard to learn from training data. This means that the model picks up on noise or random fluctuations in the training data and learns them as ideas. These ideas don't apply to the new data that is to be predicted, thereby resulting in poor model generalization. Because stock market data is highly stochastic, it is imperative to explain the methods used to resolve this issue. The most common approach to mitigate the issue of overfitting is cross validation. A few studies have applied this approach, like in [14][15][16][17][18]. In a typical k-fold cross-validation, the data is partitioned into k subsets, or folds. The model is trained iteratively on k-1 folds, and the remaining fold—also known as the hold-out fold—is treated as a test set. Numerous studies have used the early stopping method to overcome overfitting [19]. Another method is to remove irrelevant features and noise from the data, which greatly increases the model's generalizability. A few studies have implemented these procedures to avoid overfitting, such as [15][20][21][22]. The most important preventive measure against overfitting is regularization. This technique removes the extra weights from the selected features and redistributes them uniformly. It discourages the learning of models that are complex or more flexible, hence avoiding the risk of overfitting. The majority of the reviewed studies applied regularization approaches to prevent overfitting [23][24][25]. A few recent studies applied the procedure of data augmentation to prevent overfitting [26][27].

4. Comparative Analysis

The distribution of the number of papers published in recent years is presented in **Figure 2**. The number of publications increased from 2009, and was at its peak in 2019, but over the previous two years, the publication number was low. The distribution of machine learning algorithms used for SMP is shown in **Figure 3**, where the SVM was the most popular technique used. However, the ANN and DNN have attracted the research community's attention for the last few years. Traditional neural network approaches may not make accurate SMPs as initially; the weight of the randomly selected problems may suffer from the local optimal, and results in incorrect predictions [28]. The deep learning approaches are used to analyze complicated patterns in the stock data, and provide much faster results [29]. Furthermore, there is no such single technique that can promise to give the optimum results. The comparative analysis between the type of data used and the performance of the models is represented in **Figure 4**. Data alone from social media do not perform better than using market data and technical indicators [30]. However, if data from textual sources is combined with them, then the model performance increases. These ensemble approaches in predictive model building has much advantages in terms of optimizing accuracy, rigor and availability of data within emerging economies [31][32]. Econometric models like Vector autoregression (VAR) is also promising technique if model can be provided with high frequency data and retraining procedures [33].

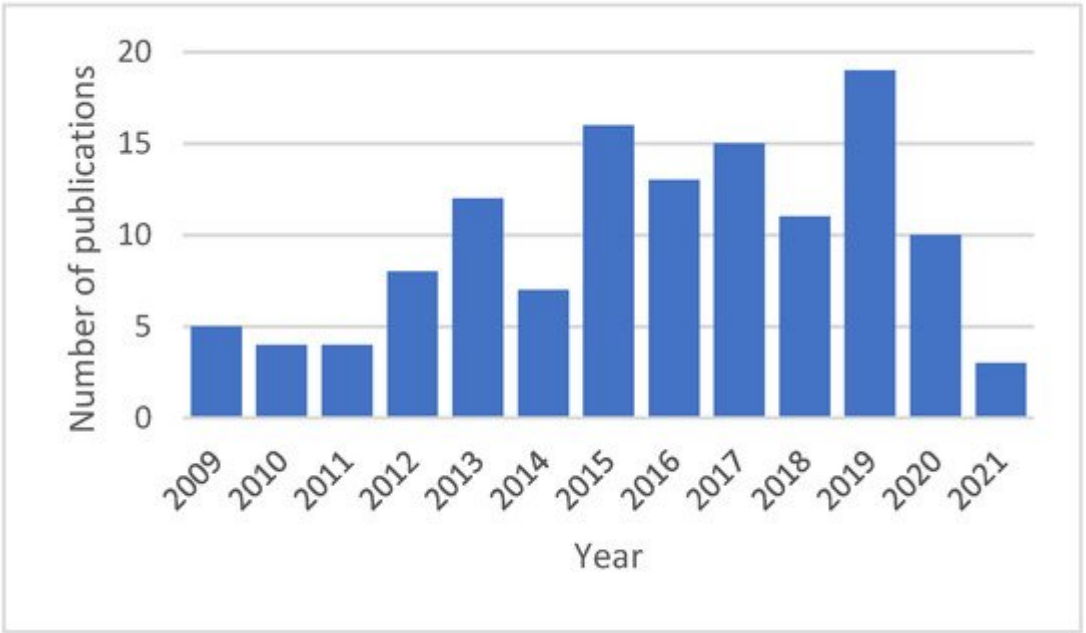


Figure 2. Number of publications per year.

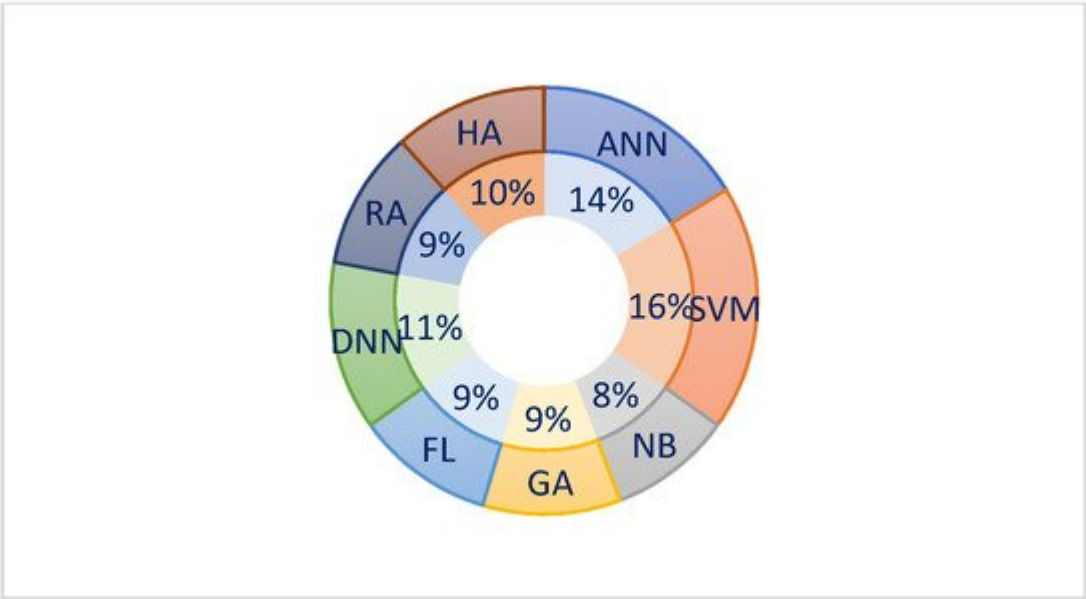


Figure 3. Distribution of the SMP techniques.

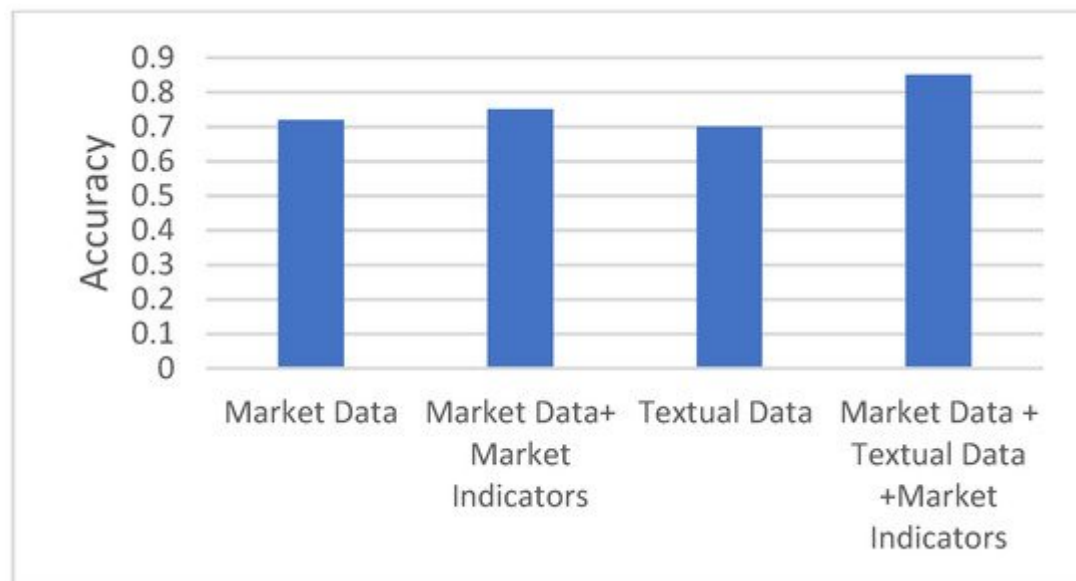


Figure 4. Comparison of the accuracies with different types of data.

References

1. Park, C.-H.; Irwin, S.H. What Do We Know about the Profitability of Technical Analysis? *J. Econ. Surv.* 2007, 21, 786–826.
2. Zhu, Y.; Zhou, G. Technical analysis: An asset allocation perspective on the use of moving averages. *J. Financ. Econ.* 2009, 92, 519–544.
3. Peachavanish, R. Stock selection and trading based on cluster analysis of trend and momentum indicators. *Lect. Notes Eng. Comput. Sci.* 2016, 1, 317–321.
4. Hulbert, M. Viewpoint: More Proof for the Dow Theory. Available online: <https://www.nytimes.com/1998/09/06/business/viewpoint-more-proof-for-the-dow-theory.html> (accessed on 17 October 2021).
5. Deboeck, G. *Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*; Wiley: New York, NY, USA, 1994.
6. Number of Social Network Users Worldwide from 2017 to 2025. Available online: <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/> (accessed on 30 May 2021).
7. 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*, Doha, Qatar, 25–29 October 2014; pp. 1415–1425.

8. Howells, K.; Ertugan, A. Applying fuzzy logic for sentiment analysis of social media network data in marketing. *Procedia Comput. Sci.* 2017, 120, 664–670.
9. Liu, B. *Sentiment Analysis and Opinion Mining*; Morgan & Claypool Publishers: San Rafael, CA, USA, 2012; p. 167.
10. Li, W. Improvement of Stochastic Competitive Learning for Social Network. *Comput. Mater. Contin.* 2020, 63, 755–768.
11. Devi, K.N.; Bhaskaran, V.M. Semantic Enhanced Social Media Sentiments for Stock Market Prediction. *Int. J. Econ. Manag. Eng.* 2015, 9, 684–688.
12. Hill, S.; Ready-Campbell, N. Expert Stock Picker: The Wisdom of (Experts in) Crowds. *Int. J. Electron. Commer.* 2011, 15, 73–102.
13. Chen, T.-L.; Chen, F.-Y. An intelligent pattern recognition model for supporting investment decisions in stock market. *Inf. Sci.* 2016, 346–347, 261–274.
14. Guresen, E.; Kayakutlu, G.; Daim, T.U. Using artificial neural network models in stock market index prediction. *Expert Syst. Appl.* 2011, 38, 10389–10397.
15. Weng, B.; Ahmed, M.A.; Megahed, F. Stock market one-day ahead movement prediction using disparate data sources. *Expert Syst. Appl.* 2017, 79, 153–163.
16. Khan, W.; Ghazanfar, M.A.; Azam, M.A.; Karami, A.; Aloubi, K.H.; Alfakeeh, A.S. Stock market prediction using machine learning classifiers and social media, news. *J. Ambient. Intell. Humaniz. Comput.* 2020, 1–24.
17. Bergmeir, C.; Hyndman, R.; Koo, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Comput. Stat. Data Anal.* 2018, 120, 70–83.
18. Ghiassi, M.; Saidane, H.; Zimbra, D. A dynamic artificial neural network model for forecasting time series events. *Int. J. Forecast.* 2005, 21, 341–362.
19. Binkowski, M.; Marti, G.; Donnat, P. Autoregressive convolutional neural networks for asynchronous time series. In *Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; Volume 2*, pp. 933–945.
20. Hagenau, M.; Liebmann, M.; Hedwig, M.; Neumann, D. Automated News Reading: Stock Price Prediction Based on Financial News Using Context-Specific Features. In *Proceedings of the 2012 45th Hawaii International Conference on System Sciences, Maui, HI, USA, 4–9 January 2012*; pp. 1040–1049.
21. Zhong, X.; Enke, D. Forecasting daily stock market return using dimensionality reduction. *Expert Syst. Appl.* 2017, 67, 126–139.

22. Gao, P.; Zhang, R.; Yang, X. The Application of Stock Index Price Prediction with Neural Network. *Math. Comput. Appl.* 2020, 25, 53.
23. Ballings, M.; Poel, D.V.D.; Hespeels, N.; Gryp, R. Evaluating multiple classifiers for stock price direction prediction. *Expert Syst. Appl.* 2015, 42, 7046–7056.
24. Srivastava, D.K.; Bhambhu, L. Data classification using support vector machine. *J. Theor. Appl. Inf. Technol.* 2010, 12, 1–7.
25. Zhang, X.; Zhang, Y.; Wang, S.; Yao, Y.; Fang, B.; Yu, P.S. Improving stock market prediction via heterogeneous information fusion. *Knowl.-Based Syst.* 2018, 143, 236–247.
26. Baek, Y.; Kim, H.Y. ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. *Expert Syst. Appl.* 2018, 113, 457–480.
27. Zheng, H.; Zhou, Z.; Chen, J. RLSTM: A New Framework of Stock Prediction by Using Random Noise for Overfitting Prevention. *Comput. Intell. Neurosci.* 2021, 2021, 8865816.
28. Pang, X.; Zhou, Y.; Wang, P.; Lin, W.; Chang, V. An innovative neural network approach for stock market prediction. *J. Supercomput.* 2018, 76, 2098–2118.
29. Arjun R., Suprabha K.R., Majhi R. (2021) Deep Learning for Stock Index Tracking: Bank Sector Case. In: Bhateja V., Peng S.L., Satapathy S.C., Zhang Y.D. (eds) *Evolution in Computational Intelligence. Advances in Intelligent Systems and Computing*, vol 1176. Springer, Singapore. https://doi.org/10.1007/978-981-15-5788-0_29
30. Arjun R & Suprabha KR, 2018. "Predictive modeling of stock indices closing from web search trends," Papers 1804.01676, arXiv.org.
31. Arjun Remadevi Somanathan; Suprabha Kudigrama Rama; A Bibliometric Review of Stock Market Prediction: Perspective of Emerging Markets. *Applied Computer Systems* **2020**, 25, 77-86, 10.2478/acss-2020-0010.
32. Arjun R., Suprabha K.R. (2020) Modeling Hybrid Indicators for Stock Index Prediction. In: Abraham A., Cherukuri A., Melin P., Gandhi N. (eds) *Intelligent Systems Design and Applications. ISDA 2018 2018. Advances in Intelligent Systems and Computing*, vol 940. Springer, Cham. https://doi.org/10.1007/978-3-030-16657-1_18
33. R. Arjun; K.R. Suprabha; Forecasting banking sectors in Indian stock markets using machine intelligence. *International Journal of Hybrid Intelligent Systems* **2019**, 15, 129-142, 10.3233/HIS-190266.

Retrieved from <https://encyclopedia.pub/entry/history/show/40218>