Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 11
TPCEE 2025

Feasibility of Machine Learning and Algorithms in Real World Stock Market Applications

Wenye Liang

Newport high school, Bellevue, Washington, United States of America

pineappletwo1@gmail.com

Abstract. The stock market is one of the most important parts of the global economy, and predicting its movements has long been a challenge. Traditional methods relied on mathematical models and historical data, but advances in computing have introduced machine learning and deep learning approaches. This paper reviews four common methods used in stock prediction: Random Forest, XGBoost, Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks. Each method has strengths, such as Random Forest's stability, XGBoost's strong pattern recognition, CNN's ability to process visual patterns, and LSTM's capability to capture temporal dependencies. However, challenges remain, including limited interpretability, poor generalization to new markets, and reliance on single-modality data. To address these issues, future research should focus on combining machine learning with expert systems for better interpretability, applying domain adaptation for cross-market generalization, and using multimodal learning to integrate numerical, textual, and sentiment data. These strategies may help improve the feasibility of machine learning in real-world stock market applications.

Keywords: Stock market prediction; machine learning; deep learning.

1. Introduction

The backbone of modern society is the economy, and within it, the stock market plays a central role. The stock market reflects not only the health of individual companies but also the overall economic stability of nations. Only forty years ago, stock market prediction relied heavily on mathematical models, financial indicators, and the intuition of experienced traders [1, 2]. Predictions were often based on technical analysis, macroeconomic data, and historical price patterns. However, as the computing power of machines advanced and data became more accessible, researchers began to explore algorithms and machine learning models for predicting market behavior. These approaches offered new possibilities, especially since they could process massive datasets faster and more efficiently than humans.

With the rapid development of computer science, many algorithms have been created that claim to forecast future stock movements. Classical statistical methods, such as autoregressive integrated moving average (ARIMA) [3], were widely used before the rise of deep learning. Later, more advanced models, including support vector machines (SVMs) [4], recurrent neural networks (RNNs) [5], and long short-term memory (LSTM) networks [6], became popular for financial prediction. These models demonstrated good performance in controlled experiments and on historical datasets. However, the transition from laboratory success to real-world applications is far more complex. In practice, markets are influenced by unpredictable human behaviors, sudden political events, and global crises, which make prediction extremely challenging.

One important limitation of many machine learning models is their inability to adapt to "black swan" events or edge cases. For example, during the COVID-19 pandemic, the global stock market faced sudden and severe disruptions. Many prediction models failed to capture the extent of the crash, as they were trained on past data that did not reflect such unprecedented conditions. The pandemic showed how interconnected the global economy has become, where supply chain shocks, health crises, and consumer confidence all interact and amplify disruptions. This revealed a critical weakness in many machine learning approaches: their dependency on historical data and assumptions that future trends will resemble the past.

Another example of market disruption is the influence of hype-driven factors, such as artificial intelligence itself. The so-called “Magnificent Seven” stocks—Apple, Microsoft, Amazon, Alphabet, Meta, Nvidia, and Tesla—experienced massive price increases due to the AI boom. As investors grew excited by the potential of AI, they poured capital into these companies, creating a feedback loop that further inflated stock prices. Traditional machine learning models often struggle with such situations because they rely on structured inputs like earnings reports or technical indicators. They cannot easily capture sentiment-driven factors like investor psychology, media influence, or sudden technological hype. This shows that real markets are not only shaped by numbers but also by human behavior, emotion, and collective expectations.

Given these challenges, it becomes critical to evaluate algorithms and machine learning models based on their robustness to extreme scenarios. Models that perform well under normal market conditions may fail under stress, which reduces their value in real-world applications. Therefore, when developing prediction systems, researchers must design stress tests and simulate unusual events to evaluate resilience. For instance, hybrid models that combine traditional financial indicators with alternative data—such as news articles, social media sentiment, or macroeconomic shocks—may perform better in capturing real-world complexity.

This paper reviews four common methods used for stock prediction, including traditional time-series approaches, machine learning models, deep learning networks, and hybrid systems. Each method has its strengths and weaknesses. Time-series models are interpretable but limited in capturing nonlinear dynamics. Machine learning models like SVMs offer flexibility but may overfit without careful tuning. Deep learning models, especially RNNs and LSTMs, can capture temporal patterns effectively, but they are often “black boxes” with limited interpretability. Finally, hybrid models that integrate financial theory, alternative data, and machine learning show promise in addressing the unpredictability of markets.

2. Methods

2.1. Random Forest

Random Forest has been widely applied in stock market prediction due to its ensemble nature and ability to reduce overfitting compared to single decision trees [7]. These findings suggest that Random Forest can effectively capture stable price movements in markets where fluctuations are largely influenced by local rather than global policies. The relatively controlled nature of such markets may have contributed to the strong performance of the model.

However, research has also highlighted important limitations of Random Forest in financial prediction. While the algorithm performs well under stable conditions, it struggles to adapt to sudden and unpredictable market changes. Sharp fluctuations, often triggered by external shocks or policy changes, reduce its effectiveness, as Random Forest does not inherently capture temporal dependencies or long-term dynamics. This restricts its use in highly volatile environments such as global stock exchanges, where unexpected events frequently disrupt price patterns.

2.2. XGBoost

XGBoost has gained attention for its gradient boosting framework and strong performance in classification and regression tasks, including financial prediction [8]. It can identify recurring short-term patterns in historical stock data and has shown effectiveness in trend prediction. Its flexibility and regularization mechanisms make it powerful in reducing bias and variance compared to traditional tree-based methods.

However, XGBoost also presents significant challenges. It relies heavily on historical data, which restricts its ability to predict abrupt market changes caused by external events. Additionally, the algorithm requires extensive hyperparameter tuning, which is computationally expensive and time consuming. Overfitting remains a concern if tuning is not carefully performed. Compared with other approaches such as Random Forest, XGBoost does not consistently outperform and may underperform in highly volatile conditions.

2.3. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have been explored for financial forecasting by treating stock data as visual patterns, such as line graphs or candlestick representations [9]. Studies reported high accuracy, with CNN models achieving up to 100% on training data and strong validation performance for certain structured input formats. Additional research on the Chinese stock market has also shown CNN models achieving accuracies around 73–75%, indicating potential for stable application.

However, CNNs face challenges of overfitting, especially when trained on limited or simplified graphical inputs. Their performance often declines in complex, real-world scenarios where financial data exhibits noise, non-linear dependencies, and sudden shocks. Moreover, CNNs primarily excel at spatial pattern recognition and may not fully capture the sequential nature of financial time series. This suggests that while CNNs are promising for stock market analysis, further refinement is required for them to compete with sequence-based models.

2.4. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are among the most widely used deep learning models for time series forecasting, including financial markets [10]. Their ability to capture long-term dependencies enables them to model both short-term fluctuations and longer market trends effectively. LSTM has demonstrated strong predictive capability in many studies, making it a powerful candidate for stock market applications.

However, the model comes with considerable drawbacks. LSTMs are computationally intensive, requiring significant training resources and time. They are also prone to overfitting, especially with small datasets, and often operate as "black-box" models with limited interpretability. These challenges make it difficult to adopt LSTMs in real-time or resource-limited trading environments. Despite these drawbacks, LSTM remains one of the most effective methods for capturing sequential dependencies and modeling dynamic market behavior.

3. Discussion

3.1. Challenges

Although machine learning and deep learning methods have shown promise in stock market prediction, several important challenges remain. These challenges limit the practical use of such models in real-world financial systems.

3.1.1. Lack of interpretability.

One of the most serious problems with machine learning models is the lack of interpretability. Models such as deep neural networks, CNNs, and LSTMs can achieve high accuracy, but they often act as "black boxes." Traders, policymakers, and investors may not trust predictions when the decision-making process is unclear. For example, when an LSTM predicts that a stock price will drop, the model does not provide clear reasoning. In finance, interpretability is especially important because investment decisions involve high risks. Without clear explanations, even accurate predictions may not be accepted in practice.

3.1.2. Limited generalization.

Another major issue is that models often do not generalize well across different markets or changing conditions. Many models are trained on historical datasets from specific countries or stock exchanges. While they may perform well in those settings, they often fail when applied to new environments. For example, a Random Forest trained on the Indonesian market may not work well for the U.S. or European markets, where stock movements are more affected by global news and policy changes. Similarly, models trained on data before major events, such as the COVID-19 pandemic, fail to predict extreme disruptions. This shows that current models are not flexible enough to handle the diversity and volatility of real-world financial markets.

3.1.3. Reliance on single-modality data.

Most stock prediction studies rely heavily on structured numerical data such as historical prices, volume, or technical indicators. While this data is useful, it does not capture the full picture of market dynamics. Real-world markets are influenced by multiple factors, including political news, economic policies, social media sentiment, and investor psychology. Relying only on price and volume limits the ability of models to capture sudden changes caused by these external factors. This single-modality approach reduces robustness and makes models less effective in practice.

3.2. Future prospects

3.2.1. Expert systems and domain knowledge for interpretability.

One way to improve interpretability is to combine machine learning models with expert systems and domain knowledge. Financial experts can design rule-based layers or add financial indicators that explain predictions in human-understandable terms. For example, a model could link its prediction of a stock drop to rising interest rates or declining earnings reports. By connecting model outputs to financial reasoning, trust and usability can be improved. Hybrid systems that integrate explainable AI (XAI) methods, such as SHAP or LIME, can also highlight which features (e.g., trading volume, volatility) most influenced the decision. This approach makes predictions more transparent and useful for investors.

3.2.2. Domain adaptation for generalization.

To address the problem of limited generalization, future research can focus on domain adaptation techniques. Domain adaptation allows models trained in one market to adapt to another by aligning distributions or transferring learned patterns. For example, a model trained on U.S. market data could be adapted to the Chinese stock market by reweighting features or fine-tuning with local data. This would make models more flexible and useful across global markets. In addition, models can be stress-tested with simulated "black swan" events, ensuring that they maintain robustness under extreme conditions. Domain adaptation could make financial prediction systems more reliable in practice, where sudden changes are common.

3.2.3. Multimodal learning for richer data representation.

Another promising direction is the use of multimodal learning, which combines different sources of data. Instead of relying only on numerical stock indicators, models can incorporate textual data from financial news, social media sentiment, or macroeconomic reports. For example, integrating Twitter sentiment with historical prices may help predict sudden stock movements triggered by public opinion. Similarly, news headlines about trade policies or earnings announcements could provide early signals of market shifts. Multimodal learning can give models a more complete understanding of market dynamics, making predictions more accurate and robust. Recent progress in natural language processing (NLP) and transformer-based models makes this integration increasingly feasible.

4. Conclusion

This review shows that while machine learning and deep learning have achieved strong results in stock prediction, their use in real markets is still limited by interpretability, adaptability, and data diversity. Random Forest and XGBoost offer stable performance but struggle with sudden market shocks. CNNs and LSTMs provide deeper insights but are often resource-intensive and hard to explain. Future research must move beyond accuracy and focus on building models that are transparent, adaptable, and capable of using multiple data sources. By combining financial knowledge with AI methods, and by testing models under extreme conditions, prediction systems can become more practical and reliable for real-world financial decision-making.

References

[1] Gandhmal DP, Kumar K. Systematic analysis and review of stock market prediction techniques. Computer Science Review. 2019 Nov 1; 34: 100190.

[2] Kumar D, Sarangi PK, Verma R. A systematic review of stock market prediction using machine learning and statistical techniques. Materials Today: Proceedings. 2022 Jan 1; 49: 3187-91.

[3] Ariyo AA, Adewumi AO, Ayo CK. Stock price prediction using the ARIMA model. In2014 UKSim-AMSS 16th international conference on computer modelling and simulation 2014 Mar 26 (pp. 106-112). IEEE.

[4] Huang W, Nakamori Y, Wang SY. Forecasting stock market movement direction with support vector machine. Computers & operations research. 2005 Oct 1; 32 (10): 2513-22.

[5] Samarawickrama AJ, Fernando TG. A recurrent neural network approach in predicting daily stock prices an application to the Sri Lankan stock market. In2017 IEEE International Conference on Industrial and Information Systems (ICIIS) 2017 Dec 15 (pp. 1-6). IEEE.

[6] Moghar A, Hamiche M. Stock market prediction using LSTM recurrent neural network. Procedia computer science. 2020 Jan 1; 170: 1168-73.

[7] Khaidem L, Saha S, Dey SR. Predicting the direction of stock market prices using random forest. arXiv preprint arXiv: 1605.00003. 2016 Apr 29.

[8] Gumelar AB, Setyorini H, Adi DP, Nilowardono S, Widodo A, Wibowo AT, Sulistyono MT, Christine E. Boosting the accuracy of stock market prediction using xgboost and long short-term memory. In2020 International Seminar on Application for Technology of Information and Communication (iSemantic) 2020 Sep 19 (pp. 609-613). IEEE.

[9] Chen S, He H. Stock prediction using convolutional neural network. InIOP Conference series: materials science and engineering 2018 Nov 5 (Vol. 435, p. 012026). IOP Publishing.

[10] Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions. European journal of operational research. 2018 Oct 16; 270 (2): 654-69.