Metaheuristic Optimization Review

ISSN: 2692-4048 Vol: 02, No. 01, 2024 PP. 29–33
Article DOI: https://doi.org/10.54216/MOR.020105

REVIEW ARTICLE

Machine Learning in Stock Price Prediction: A Review of Techniques and Challenges

Doaa Sami Khafaga[1,*] **Sunil Kumar**[2]

[1] Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

[2] School of Computer Science, University of Petroleum and Energy Studies, Dehradun, 248001, India

Emails: dskhafga@pnu.edu.sa · skumar@ddn.upes.ac.in

Received: May 15, 2024 Revised: August 14, 2024 Accepted: November 24, 2024 ⋆ Corresponding author

ABSTRACT

Future stock price prediction is one of the most important and complex tasks in finance, mainly due to the characteristics of financial markets. Machine learning techniques have greatly improved this area: problems with frequent data and nonlinear processes, which cannot be solved using conventional models, have been addressed. This paper reviews how data preprocessing methodology and two modeling directions, namely high-frequency data models and sentiment analysis models, have helped improve the efficiency of stock price forecasts. Among the proposed techniques, Temporal Convolutional Networks (TCN), attention mechanisms, and Transformer-based architectures are highlighted because of their capability to distill complex market dynamics. However, issues such as data quality and market fluctuations remain persistent even as innovation accelerates, which reinforces the importance of model robustness and interpretability. Drawing on recent advances and mapping directions for future studies, this paper shows how machine learning may transform stock market prediction and investment decision-making in a continuously changing financial environment.

Keywords: High-frequency data ▪ Hybrid models ▪ Machine learning ▪ Stock price prediction ▪ Sentiment analysis ▪ Financial forecasting

1. INTRODUCTION

Stock price prediction has been widely explored because financial markets are volatile and provide a challenging environment for both practice and research. Due to innovations in machine learning and data analysis, this field has changed substantially. Advanced artificial intelligence is instrumental in addressing the inherent problems of predicting stock prices and understanding market behavior. This section introduces these developments and their consequences for financial forecasting.

1.1 Some Difficulties with Traditional Models

Traditional approaches to forecasting stock price movements mostly use low-frequency data and consequently fail to capture current global stock market dynamics. These models do not incorporate the features of high-frequency market data, where changes are abrupt and volatility occurs at higher rates. As a result, their predictions are often unstable and their usefulness is restricted for near real-time decision-making. These shortcomings have created a gap that needs to be filled by methodologies capable of handling higher frequencies and providing more detailed market trends [1].

1.2 Advancements in High-Frequency Data

Among the pioneering changes in this area is the use of high-frequency data for stock price forecasting. Higher frequency observations enable more efficient examination of changing market conditions and asset-price behavior. In preprocessing, histogram data has become popular for improving the quality of financial predictions. Histogram-based preprocessing converts raw high-frequency data into organized forms, supporting improved estimation of asset returns and better evaluation of market behavior [2].

1.3 Hybrid Models for Enhanced Prediction

The complexity and volatility of financial markets have also led to the creation of mixed forecasting models. These models combine predictive techniques to exploit the relative advantages of different approaches. For example, hybrid systems may combine machine learning techniques with statistical analysis, or integrate multiple machine learning procedures to build ensemble strategies. Such combinations have shown better forecasting results than typical approaches when analyzing complex markets and dealing with data variability [3].

1.4 Why the Stock Market is Important

The stock market has dual importance as an economic indicator and as a source of investment opportunities. Its impact extends from individual investors to global financial institutions, frameworks, and economic paradigms. Therefore, stock price prediction remains essential for decision-making, risk management, strategic business planning, and investment. Understanding how machine learning supports stock market forecasting and enables proactive action is central to broader adoption of the technology [4, 5].

From challenges to innovation and hybrid methodologies, this study highlights the evolving role of machine learning in financial markets. By bridging traditional weaknesses and taking advantage of high-frequency data and various modeling approaches, researchers and practitioners are laying the foundation for better forecasts of stock price movements.

2. LITERATURE REVIEW

The intense advances in data analytics and machine learning have significantly affected financial markets, especially asset-price forecasting. Traditional models that operate on low-frequency data can hardly capture severe fluctuations and fine details of high-frequency market changes. Recent research has therefore focused on histogram data as a preprocessing tool, adding a different perspective to asset-return estimation and market-behavior analysis. The growing interest in high-frequency data has also driven development of hybrid models, in which two or more forecasting techniques are combined to enhance forecasting quality.

Stock trend forecasting often requires extensive data analysis and robust predictive techniques. Talazadeh and Perakovic [6] show that integrating sentiment analysis using the FinGPT generative AI model with Random Forest enhances stock price prediction. Their Sentiment-Augmented Random Forest (SARF) approach incorporates sentiment features into the Random Forest framework, achieving an average accuracy improvement of 9.23% over conventional Random Forest and LSTM models while reducing prediction errors.

The stock market reflects economic conditions, offers investment opportunities, and shapes global financial trends. Gil et al. [7] address the challenges of predicting nonlinear and stochastic market trends through advanced deep-learning models applied to daily and hourly closing prices from the S&P 500 index and the Brazilian ETF EWZ. The study evaluates Temporal Convolutional Networks, N-BEATS, Temporal Fusion Transformers, N-HiTS, TiDE, and xLSTM-TS. Wavelet denoising is used to smooth fluctuations, and xLSTM-TS achieves strong performance, including 72.82% test accuracy and 73.16% F1-score on the EWZ daily dataset.

Ge [8] investigates predictive models designed to improve forecasting accuracy and address market volatility. The MEME-AO-LSTM model employs time-series decomposition and hyperparameter optimization, achieving an RMSE of 27.12, MAE of 19.43, and correlation value of 0.992. The model is evaluated across major markets such as NASDAQ 100, Nikkei 225, and FTSE, and demonstrates robustness under market shocks including the COVID-19 period and geopolitical tensions.

Attention mechanisms are effective for capturing patterns in sequential data and outperforming several traditional models. Kelvin et al. [9] incorporate attention into an enhanced stock prediction model that standardizes and decomposes stock data into high- and low-frequency components. The attention mechanism is applied to high-frequency signals, reducing RMSE from 0.5824 to 0.5709 and improving predictive accuracy.

Stock price prediction also benefits from social-media sentiment. Mcdonald [10] predicts future closing prices of the top five NASDAQ100 stocks by integrating Twitter sentiment data with machine learning models. ARIMA, RNN, LSTM, GRU, and Transformer models are evaluated, and the Transformer model, leveraging attention mechanisms, outperforms recurrent models in predictive accuracy and computational efficiency.

Sustainable investing requires integrating multiple sources of information. Yu et al. [11] develop a deep-learning model that combines stock data and textual information using LSTM and GloVe embeddings. Their results show that semantic information can improve forecasting, particularly where nonlinear relationships are present in financial data.

Social-media discussions can also affect bank stocks. Fellah and Koualed [12] examine Twitter and YouTube conversations and show that negative sentiment correlates with market declines during critical financial events. This line of work emphasizes the need to manage volatile sentiment information when deploying machine learning in finance.

Generative and time-series pretraining methods are increasingly relevant. Cao et al. [13] introduce TEMPO, a prompt-based generative pre-trained Transformer for time-series forecasting, demonstrating the importance of time-domain and frequency-domain consistency. Lin [14] uses wrapper feature selection and Gradient Boosting Machines for key financial indicator analysis, supporting buy, sell, and hold decisions through a confidence-voting strategy.

Beyond equity prices, related work extends machine learning to crude-oil and index prediction. Xu et al. [17] combine multivariate selection, machine learning, and nonlinear strategies for crude-oil price forecasting, while Garikapati and Yeddula [18] compare SVM and LSTM for NIFTY 50 prediction, finding that LSTM better captures trends and patterns. Qiu [1] provides a broader analysis of market trends and practical applications of machine learning in stock prediction.

High-frequency financial econometrics remains central to this field. Degiannakis and Floros [23] discuss modeling and forecasting of high-frequency financial data, while Ozupek et al. [24] present a hybrid EMD-TI-LSTM model for financial forecasting. Buchetti et al. [25] and Arksey and O'Malley [26] further emphasize emerging research trends and systematic mapping methods relevant to reviewing financial machine learning literature.

A summary of selected and reviewed papers on machine learning in stock price prediction is presented in Table 1. The table highlights hybrid approaches, sentiment analysis, and sophisticated frameworks such as Transformers and attention mechanisms. It also outlines common challenges such as data quality, market fluctuation, model complexity, and sentiment variability.

3. DISCUSSION

Machine learning has transformed stock price prediction by enabling models to process large, complex, and nonlinear datasets. The reviewed studies show that models based on high-frequency data, hybrid architectures, attention mechanisms, and sentiment analysis can improve forecasting accuracy compared with conventional approaches. However, these improvements must be assessed together with challenges related to data quality, interpretability, robustness, and computational cost.

3.1 Strengths of Machine Learning in Stock Price Prediction

Machine learning models are well suited to detecting hidden relationships in financial data. They can combine historical prices, technical indicators, social-media sentiment, news, and macroeconomic variables. Hybrid and ensemble approaches further increase predictive power because they combine complementary model strengths. Deep-learning architectures, especially LSTM, GRU, attention-based networks, and Transformers, capture temporal dependence and nonlinear market behavior more effectively than many traditional statistical methods.

High-frequency data offers another advantage by improving the granularity of market analysis. When raw high-frequency observations are transformed into structured representations such as histograms, models can better capture intraday volatility, correlations, and market responses to events. Sentiment analysis adds a behavioral dimension by incorporating investor opinion and public reaction into forecasting models.

3.2 Challenges and Limitations

Despite strong results, several limitations remain. Financial markets are noisy, nonstationary, and influenced by unpredictable events, including economic policy changes, geopolitical tensions, investor psychology, and global crises. Models trained on historical data may fail when market regimes change. Data quality is also critical: missing values, class imbalance, biased sentiment sources, and noisy social-media text can reduce forecasting reliability.

Interpretability is another persistent challenge. Many deep-learning models operate as black boxes, making it difficult for investors, regulators, and analysts to understand why a prediction was made. Computational complexity can also limit real-time deployment, particularly when models use high-frequency data or complex ensemble structures.

3.3 Real-Life Deployment and Consequences

Real-life deployment requires more than high predictive accuracy. Financial institutions must consider risk management, latency, transaction costs, regulatory compliance, and model governance. A model that performs well in back-testing may not achieve the same results after deployment if data pipelines, market liquidity, or decision rules are not carefully designed. Therefore, practical systems should include continuous monitoring, model updating, and risk controls.

3.4 Future Research Directions

Future studies should develop models that are more robust to market shocks and regime changes. Promising directions include adaptive learning, explainable artificial intelligence, hybrid models that integrate behavioral finance and macroeconomics, and transparent evaluation protocols. Combining financial variables with textual and sentiment sources may provide a richer qualitative assessment of market trends. Models capable of continuous learning may also improve forecasting stability over time.

3.5 Other Effects that Affect the Financial Industry

Using machine learning to forecast stock prices is one example of the broader digitalization and automation of financial markets. As these technologies continue to evolve, they are expected to reshape investment management, policies, and market operations. Although challenges remain, the potential advantages—from higher accuracy to improved risk assessment—point to the significant role machine learning can play in modern finance [25, 26].

This discussion highlights the dual nature of machine learning in finance: it is powerful and effective, but only when the associated problems and controversies are carefully managed. Continued research and methodological development can help machine learning reveal new insights about stock price prediction that were previously difficult to observe.

4. CONCLUSION

Combining high-frequency data and sophisticated machine learning approaches has improved stock price prediction and provided researchers and practitioners with robust tools for studying intricate and dynamic market phenomena. Data frequency has become crucial for refining the analysis of market fluctuations because granular high-frequency data can reveal short-term movements that are hidden in low-frequency observations. Hybrid and ensemble techniques have also handled the nonlinear and stochastic nature of financial markets more effectively, producing forecasts that are more practical and accurate.

Table 1. Summary of Literature Review

Study	Focus Area	Methods/Models Used	Key Findings	Challenges Identified
[6]	Sentiment-Augmented Random Forest	Random Forest integrated with sentiment analysis	Improved prediction accuracy by 9.23% over LSTM models and reduced prediction errors	Data integration; sentiment variability
[7]	Deep learning for stock trend prediction	TCN, N-BEATS, TFT, N-HiTS, xLSTM-TS	xLSTM-TS achieved 72.82% accuracy; wavelet denoising enhanced preprocessing	Nonlinear market trends
[8]	Hybrid model for market forecasting	MEME-AO-LSTM	RMSE of 27.12 and MAE of 19.43; effective during COVID-19 and geopolitical shocks	Data variability; model complexity
[9]	Attention mechanisms in prediction	Bidirectional LSTM, XCEEMDAN, spline	Reduced RMSE to 0.5709 from 0.5824 and better captured sequential relationships	Handling multidimensional data
[10]	NASDAQ stock prediction with sentiment	ARIMA, RNN, LSTM, GRU, Transformer	Transformer model outperformed recurrent models for several stocks	Sentiment causality verification
[11]	Sustainable stock prediction	LSTM with GloVe embeddings	Integrated text and stock data improved semantic forecasting	Nonlinear data patterns
[12]	Bank stock losses and social media	Sentiment analysis	Negative sentiments correlated with market declines during critical events	Volatile sentiment integration
[13]	Time-frequency consistency	Time-domain and frequency-domain analysis	Enhanced prediction with risk control and low-risk decision strategies	Traditional model limitations
[14]	Key financial indicators	Wrapper feature selection; Gradient Boosting Machines	Supported buy, sell, and hold decisions through confidence voting	Data incompleteness; chaotic data nature
[17]	Crude-oil price forecasting	Hybrid ML models with histogram data	Achieved average absolute percentage errors of 2.9962% and 2.4314%	Dependence on historical data and external factors
[18]	NIFTY 50 stock prediction	SVM and LSTM	LSTM outperformed SVM with lower error metrics for trend capture	Data preprocessing; external factor integration
[2]	High-frequency data modeling	Histogram-based preprocessing	Improved prediction accuracy for ASEAN-5 stock markets during significant events	High-frequency volatility and market interdependence

Nevertheless, market risk and dynamics remain critical issues for future research. Financial developments are characterized by uncertainty and are affected by economic and political conditions, investor opinion, and policy changes. Machine-learning models must therefore adapt to market changes and shocks, including COVID-19-sensitive market disturbances. Additional challenges concern data quality, interpretability, and computational efficiency, especially when models are applied to large datasets and real-time decision-making.

Future investigations should address these difficulties by combining machine learning with fields such as behavioral finance and macroeconomics. Integrating financial variables with qualitative information can provide a richer assessment of market trends, while continuous-learning models can improve forecasting stability. Transparency should also be increased so stakeholders can make appropriate decisions when using predicted data.

In conclusion, stock price prediction using machine learning is a rapidly developing area of study. If current limitations are overcome and interdisciplinary approaches are followed, these technologies can reveal hidden potential within financial markets. Integrating high-frequency data, efficient algorithms, and new techniques is poised to transform conventional stock-price forecasting and create opportunities for financiers, regulators, and policymakers to understand the contemporary financial environment.

REFERENCES

[1] H. Qiu, "Leveraging Machine Learning for Stock Market Prediction: A Comprehensive Analysis of Market Trends and Practical Applications," *Transactions on Computational and Scientific Methods*, vol. 11, no. 4, Nov. 2024.

[2] A. Koukorinis, "Kernel-based Machine Learning Methods for High-Frequency Financial Time Series Analysis," Doctoral thesis, UCL (University College London), Sep. 2024.

[3] R. Kalra et al., "An efficient hybrid approach for forecasting real-time stock market indices," *Journal of King Saud University—Computer and Information Sciences*, vol. 36, no. 8, p. 102180, Oct. 2024.

[4] "ResearchGate | Find and share research," accessed Dec. 11, 2024.

[5] "IEEE Xplore," accessed Dec. 11, 2024.

[6] S. Talazadeh and D. Perakovic, "SARF: Enhancing Stock Market Prediction with Sentiment-Augmented Random Forest," Sep. 2024.

[7] G. L. Gil, P. Duhamel-Sebline, and A. McCarren, "An Evaluation of Deep Learning Models for Stock Market Trend Prediction," Aug. 2024.

[8] Q. Ge, "Enhancing stock market forecasting: A hybrid model for accurate prediction of S&P 500 and CSI 300 future prices," *Expert Systems with Applications*, vol. 260, p. 125380, Jan. 2025.

[9] Kelvin, F. M. Sinaga, and S. Winardi, "Exploring New Frontiers: XCEEMDAN, Bidirectional LSTM, Attention Mechanism, and Spline in Stock Price Forecasting," 2024.

[10] B. Mcdonald, "Predicting Future Stock Price with Sentiment Analysis: Recurrent vs. Attention Based Learning for Regression Tasks," University of the Witwatersrand, Johannesburg, 2023.

[11] H. Yu, T. Zhang, and M. Kizildag, "Advancing sustainable investing: A deep learning model for multi-source stock prediction," *Journal of Global Business Insights*, vol. 9, no. 2, pp. 176–176, Sep. 2024.

[12] A. S. Fellah and A. M. S. Koualed, "Exposure of Bank Run to Social Media: Predicting bank stock market losses from Twitter and YouTube conversation," 2024.

[13] D. Cao et al., "TEMPO: Prompt-Based Generative Pretrained Transformer for Time Series Forecasting," *ICLR 2024*, 2024.

[14] C. Lin, "Key Financial Indicators Analysis and Stock Trend Forecasting Based on a Wrapper Feature Selection Method," pp. 755–759, Oct. 2024.

[15] S. Talazadeh and D. Perakovic, "SARF: Enhancing Stock Market Prediction with Sentiment-Augmented Random Forest," Sep. 2024.

[16] S. Talazadeh and D. Perakovic, "SARF: Enhancing Stock Market Prediction with Sentiment-Augmented Random Forest," Sep. 2024.

[17] Y. Xu, T. Liu, Q. Fang, P. Du, and J. Wang, "Crude oil price forecasting with multivariate selection, machine learning, and a nonlinear combination strategy," *Engineering Applications of Artificial Intelligence*, vol. 139, p. 109510, Jan. 2025.

[18] A. Garikapati and D. R. Yeddula, "A Comparative Analysis of SVM and LSTM: Stock market prediction Nifty 50 index," 2024.

[19] H. Qiu, "Leveraging Machine Learning for Stock Market Prediction: A Comprehensive Analysis of Market Trends and Practical Applications," *Transactions on Computational and Scientific Methods*, vol. 11, no. 4, Nov. 2024.

[20] H. Qiu, "Leveraging Machine Learning for Stock Market Prediction: A Comprehensive Analysis of Market Trends and Practical Applications," *Transactions on Computational and Scientific Methods*, vol. 11, no. 4, Nov. 2024.

[21] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *Proceedings of EMNLP 2014*, pp. 1724–1734, 2014.

[22] N. Ngoc Thach, V. Kreinovich, and N. D. Trung, Eds., *Data Science for Financial Econometrics*, vol. 898, 2021.

[23] S. Degiannakis and C. Floros, *Modelling and Forecasting High-Frequency Financial Data*, pp. 1–278, Jan. 2015.

[24] O. Ozupek, R. Yilmaz, B. Ghasemkhani, D. Birant, and R. A. Kut, "A Novel Hybrid Model (EMD-TI-LSTM) for Enhanced Financial Forecasting with Machine Learning," *Mathematics*, vol. 12, no. 17, p. 2794, Sep. 2024.

[25] B. Buchetti, F. R. Arduino, and S. Perdichizzi, "A literature review on corporate governance and ESG research: Emerging trends and future directions," *International Review of Financial Analysis*, vol. 97, Jan. 2025.

[26] H. Arksey and L. O'Malley, "Scoping studies: Towards a methodological framework," *International Journal of Social Research Methodology: Theory and Practice*, vol. 8, no. 1, pp. 19–32, Feb. 2005.