A Robust Framework for Stock Market Prediction Using Hybrid Feature Selection and Advanced Predictive Models

Smruti Rekha Das[1], Divya Midhun[2], Rajesh Dey[3]

[1]GITAM Deemed to be University, Andhra Pradesh

[2]Lincoln University College, Malaysia

[3]Gopal Narayan Singh University, India

[1]pdf.smruti@lincoln.edu.my, [2]divya@lincoln.edu.my, [3]rajesh.dey@gnsu.ac.in

Abstract

Stock market prediction is a complex task due to its non-linear dynamics and the influence of multifactorial indicators. Traditional models often fall short in capturing the intricate dependencies that affect market behaviour. This study proposes a robust prediction framework that leverages an integrated dataset comprising financial variables, macroeconomic indicators, ESG scores, sentiment, and technical indicators. Feature selection is executed through a hybrid approach using both statistical methods (RFE, embedded, filter) and optimization algorithms (PSO, GA, DE, Grey Wolf). Advanced predictive models including Long Short-Term Memory (LSTM), Gradient Boosting Decision Trees (GBDT), and Natural Language Processing (NLP) techniques are employed to forecast stock prices across various time horizons (1 to 45 days). Results indicate that hybrid feature selection enhances prediction accuracy, and that model performance varies with the forecast window. Notably, ESG features significantly improve mid- to long-term forecasts. This framework is applicable in algorithmic trading, portfolio management, and risk mitigation strategies.

Keywords: Stock Market Prediction; Feature Selection; ESG; LSTM; Optimization Algorithms

1. Introduction

Stock price prediction remains a central challenge in financial forecasting due to the volatile, non-linear, and multifactorial nature of market movements. Financial variables alone fail to provide a holistic picture of market dynamics. As sustainable investing and sentiment-based trading rise in popularity, additional variables such as ESG scores, macroeconomic indicators, and news sentiment have become critical in modern forecasting frameworks. Motivated by this complexity, this work aims to build a comprehensive and robust prediction framework by integrating diverse datasets and deploying hybrid feature selection and machine learning models.

2. Related Work

Various studies have attempted to forecast stock prices using machine learning (ML) and deep learning (DL) models. For instance, Sonkavde et al. [1] presented a comprehensive review of ML and DL models for financial forecasting, noting the superior performance of LSTM in capturing time-series dependencies. Lim [2] explored how ESG factors, combined with AI, can influence financial predictions. Rezaei et al. [3] emphasized the importance of macroeconomic and ESG indicators in achieving sustainable forecasting accuracy. Giudici and Wu [4] further demonstrated the integration of ESG features into financial AI frameworks.

Table 1 compares previous work with the current study based on three major aspects: use of ESG, optimization-based feature selection, and prediction over varying horizons.

Table 1. Comparison of Related Works

Reference	ESG Included	Optimization-Based Feature Selection	Multi-Horizon Prediction
[1]	No	Yes	No
[2]	Yes	No	No
[3]	Yes	Yes	No
This Work	Yes	Yes	Yes

3. Key Contribution

- Integration of financial, macroeconomic, sentiment, and ESG data into a unified prediction dataset.
- Application of both statistical (RFE, embedded, filter) and optimization-based (PSO, GA, DE, Grey Wolf) feature selection methods.
- Comparative analysis of multiple models (LSTM, GBDT, NLP-based) across forecasting horizons ranging from 1 to 45 days.
- Demonstration of improved accuracy using ESG scores and hybrid feature selection.
- Practical application in portfolio optimization, financial planning, and algorithmic trading strategies.

4. Method

The methodology involves the creation of a dataset comprising:

- Financial metrics (OHLC, Volume)

- Macroeconomic indicators (Inflation, Interest Rate)
- ESG scores (Environmental, Social, Governance factors)
- Technical indicators using a 12-day window (SMA, EMA, RSI, MACD, Stochastic Oscillator)

Feature Selection

Feature selection is carried out using:

- **Statistical Methods**: Recursive Feature Elimination (RFE), filter methods (e.g., mutual information), and embedded methods (e.g., tree-based importance).
- **Optimization Techniques**: Particle Swarm Optimization (PSO), Differential Evolution (DE), Genetic Algorithm (GA), and Grey Wolf Optimizer (GWO).

Predictive Modelling

Three classes of models are used:

- **LSTM**: Captures temporal dependencies in sequential data.
- **GBDT**: Effective for tabular data and non-linear interactions.
- **NLP-Based Models**: Incorporate sentiment data for contextual prediction.

5. Discussions

The study demonstrates that hybrid feature selection significantly enhances model performance by filtering out noise and irrelevant variables. ESG scores, when used alongside macroeconomic and technical indicators, improve the model's ability to generalize over medium and long horizons. LSTM models provide superior short-term predictions due to their recurrent architecture, while GBDT and NLP models excel with more contextual and non-linear relationships.

Moreover, optimization-based feature selection (especially GA and Grey Wolf) yielded better feature subsets than purely statistical methods. These findings align with existing literature but extend the scope by integrating more data dimensions and broader forecasting horizons.

6. Conclusions

This study presents a unified framework for stock market prediction that integrates comprehensive data sources, dual-layered feature selection, and advanced predictive modelling. The inclusion of ESG scores and macroeconomic indicators alongside traditional technical and financial data provides a holistic view of market behaviour. The experimental results demonstrate that combining statistical and optimization-based feature selection with ensemble and deep learning models leads to enhanced forecasting accuracy. This framework holds promise for both academic researchers and financial practitioners aiming to improve market forecasting and investment strategies in an increasingly complex and data-rich financial ecosystem.

Limitations and Future Work:

Limitations: Model performance may vary across different sectors or countries; results are dataset-dependent.

Future Work: Integration of real-time streaming data, sector-specific modeling, and transformer-based NLP models.

References

1. G. Sonkavde, D. S. Dharrao, A. M. Bongale, S. T. Deokate, D. Doreswamy, and S. K. Bhat, "Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications," *Int. J. Financial Stud.*, vol. 11, no. 3, p. 94, 2023. https://doi.org/10.3390/ijfs11030094

2. T. Lim, "Environmental, social, and governance (ESG) and artificial intelligence in finance: State-of-the-art and research takeaways," *Artif. Intell. Rev.*, vol. 57, no. 4, p. 76, 2024. https://doi.org/10.1007/s10462-023-10576-7

3. A. Rezaei, I. Abdellatif, and A. Umar, "Towards Economic Sustainability: A Comprehensive Review of Artificial Intelligence and Machine Learning Techniques in Improving the Accuracy of Stock Market Movements," *Int. J. Financial Stud.*, vol. 13, no. 1, p. 28, 2025. https://doi.org/10.3390/ijfs13010028

4. P. Giudici and L. Wu, "Sustainable artificial intelligence in finance: impact of ESG factors," *Front. Artif. Intell.*, vol. 8, p. 1566197, 2025. https://doi.org/10.3389/frai.2023.1566197