European Economic Letters
ISSN 2323-5233
Vol 15, Issue 4 (2025)
http://eelet.org.uk

Hybrid Machine Learning and Econometric Models for Predicting BSE-IT Index Prices

Dr. K. Siva Nageswara Rao[1]

[1]Department of Management Studies, VFSTR (Deemed to be university) Vadlamudi, Guntur, Andhra Pradesh, India, ksnr.vignan@gmail.com

Dr. G. Padmavathi[2]

[2]Department of BS&H, Chalapathi Institute of Engineering and Technology, Lam, Guntur, Andhra Pradesh, India, Padma.goriparthi68gmail.com

Abstract

The forecasting ability of ARIMA, GARCH, XGBoost, and LSTM model is compared in this research in the scenario of forecasting the BSE-IT index stock price. Historical monthly data were trained and tested on these models with significant performance metrics such as MAE, MSE, RMSE, MAPE, and R^2 score. Findings suggest that XGBoost performs best among other models with the smallest error rates and highest predictive capability. Machine learning algorithms (XGBoost, LSTM) have better flexibility towards stock market patterns than statistical models. The research indicates that hybrid methods for financial prediction are necessary and propose the integration of sentiment analysis and macroeconomic data for increased accuracy.

Keywords: Stock Price Prediction, BSE-IT Index, ARIMA, GARCH, XGBoost, LSTM, Machine Learning, Financial Forecasting.

1. Introduction

Stock price forecasting is a difficult task because financial markets are volatile, nonlinear, and stochastic in nature. Conventional econometric models, including Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH), are extensively applied for time series forecasting and volatility modeling. These models are not, however, efficient in detecting complex nonlinear relationships between stock price movements. On the other hand, Machine Learning (ML) algorithms like XGBoost, Long Short-Term Memory (LSTM) networks, and Transformers are extremely good at detecting latent patterns in financial data but are not interpretable to a great extent (Li et al., 2020; Wang et al., 2024).

Hybridization of econometric models with ML algorithms can combine the interpretability of econometric models with the predictive capability of ML. The current research considers hybrid models to forecast BSE-IT index prices with emphasis on accuracy and robustness in financial forecasting (Challa et al., 2018; Malepati et al., 2019).

This research explores a hybrid econometric-ML approach to forecasting BSE-IT index prices with monthly data ranging from April 2014 to March 2024. It combines ARIMA, GARCH, XGBoost, LSTM, and Transformers to provide higher accuracy forecasts. The research assists investors in assessing risk, especially in India's IT sector trend and volatility (Latha et al., 2022).

Forecasting stock price is crucial for both investors and policymakers. Classical econometric models detect trends, while ML models recognize patterns. The current research combines both in pursuit of more predictive precision, easing risk management, investment decisions, and market stability. It fills methodological gaps, enhancing forecasting for the volatile BSE-IT index.

The BSE-IT index is highly volatile because of technological advancements, world market movements, and macroeconomic conditions. Classical econometric models such as ARIMA and GARCH can recognize historical trends well but cannot manage nonlinear trends. However, Machine Learning (ML) models such as XGBoost, LSTM, and Transformers can recognize underlying relationships but are not interpretable and do not have economic significance (Latha et al., 2024; Malepati et al., 2021; Venkataramanaiah et al., 2018).

The issue is to design a robust predictive model integrating econometric and ML models to enhance the precision of stock price prediction. Both techniques together can bridge the gap between statistical complexity and AI-driven intelligence, improving financial decision-making, risk analysis, and investment planning for the BSE-IT sector.

Objectives:

- Develop a hybrid econometric-ML model to predict stock prices.
- Compare ARIMA, GARCH, and ML model performances in trend detection and volatility identification.
- Compare ML models (XGBoost, LSTM, Transformers) in model optimization of stock price predictions.
- Compare the enhancement in performance between hybrid models and standalone models.

2. Literature Review

The application of econometric models together with machine learning (ML) techniques has been comprehensively explored to enhance the accuracy of financial time series forecasting.

(Kuo & Chiu, 2024) proposes a hybrid approach that separates stock price series into periodic and non-periodic components, employing ML models like Gradient Boosting Decision Trees and Random Forest to improve the precision of prediction. (Dioubi et al., 2024) present a comparative analysis of the performance of ARIMA, LSTM, and Transformer models using historical stock price data, debating the relative strength and weakness of each model in identifying different kinds of patterns in the data. (Alzaman, 2024) presents various deep learning structures, their application, strengths, and weaknesses for financial time series forecasting. (Rather, 2021) recommends a hybrid process combining ML algorithms, e.g., LSTM networks and econometric models, in order to enhance stock market behavior prediction.

(Dave et al., 2021) explores the application of ARIMA-GARCH, LSTM, and Wavelet-LSTM models, contrasting predictive performance to highlight the importance of accurate forecasting towards market prediction. The concern of (Kakade et al., 2022) is using ML models like LSTM and Wavelet-LSTM for the forecasting of tomorrow's stock price and for decision making.

(Chen et al., 2022) suggests a new financial time series predictive model with an ensemble of LSTM, Transformer, and MLP models to enhance robustness of prediction. (Yang et al., 2024) construct a hybrid model merging econometric and ML methods to describe the manner in which financial assets affect realized volatility with the objective of providing best daily forecasts.

(Gülmez, 2025) hypothesizes a combination of Genetic Algorithms and Time Series Analysis for stock market prediction to achieve the highest accuracy. (Ashrafzadeh et al., 2023) present a new financial time series forecasting model based on a deep learning ensemble with a combination of CNN, LSTM, and ARMA models to acquire spatiotemporal and autocorrelation features. Financial time series forecasting has been widely explored using both econometric and ML models. ARIMA and GARCH Models: ARIMA performs well in identifying linear trends, whereas GARCH models volatility clustering. They do not perform well with nonlinearity in stock price movements (Hu et al., 2020; Mutinda & Langat, 2024).

XGBoost is a good fit for forecasting in structured data, LSTM networks are good to find sequential patterns, and Transformers are increasingly a good financial forecasting tool (Yun et al., 2021). It has been proven that a combination of econometric models and ML techniques increases predictability. For instance, ARIMA + XGBoost improves time series forecasting, and GARCH + LSTM identifies volatility and nonlinear relationships (Jing et al., 2021). Overall, the literature shows the potency of hybrid models combining econometric and ML methods in enhancing the accuracy and stability of financial time series forecasting.

3. Methodology:

Research Design

The research utilizes an exploratory and analytical research design to forecast BSE-IT index prices based on a hybrid econometric and machine learning methodology. It utilizes a quantitative research approach, utilizing past stock price data from April 2014 to March 2024. The research combines econometric models (ARIMA, GARCH) with ML algorithms

(XGBoost, LSTM, Transformers) for improved forecasting accuracy. Data pre-processing, feature selection, and model validation by RMSE and MAPE are used for robustness. Comparative analysis is made between hybrid and standalone models. The study is conducted following a systematic, empirical framework, and Python and EViews are used for statistical modeling and performance measurement.

Data Collection

The research uses monthly BSE-IT index prices from April 2014 to March 2024, obtained from BSE India, Yahoo Finance, and RBI. Other macroeconomic variables, including interest rates, inflation, and exchange rates, are added to improve model performance. Pre-processing of data includes missing value handling, outlier removal, and normalization of variables. Technical indicators like Moving Averages, RSI, and Bollinger Bands are computed to improve predictive performance. The data is split into training (80%) and test (20%) sets to test the model. Libraries such as Pandas, NumPy, and Scikit-Learn are used for data cleaning and feature engineering.

Machine Learning Techniques

The research uses several machine learning (ML) algorithms to optimize the precision of BSE-IT index price forecasts. Some of the ML techniques used are:

- XGBoost (Extreme Gradient Boosting): A high-performance ensemble learning algorithm that works well with big data and avoids overfitting.
- LSTM (Long Short-Term Memory): A recurrent neural network (RNN) that can extract long-term relationships in time-series data.
- Transformer Models: These models apply self-attention mechanisms to identify intricate patterns and enhance sequential forecasting.
- Hybrid ARIMA-GARCH-LSTM Model: Combines econometric models with deep learning to capture linear trends and nonlinear dependencies.

Performance is measured using RMSE, MAPE, and R-squared measures to compare models.

4.Result

This research compares the performance of different forecasting models—ARIMA, GARCH, XGBoost, and LSTM—to forecast stock prices. The models were compared using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R^2 Score.

The findings point out that XGBoost is superior to the other models by showcasing the lowest error rates and the best R^2 score, which testifies to its better capability of modeling market trends. Although LSTM is also very good, the older models such as ARIMA and GARCH exhibit comparatively higher error levels. The findings stress the superiority of machine learning models in stock price forecasting compared to statistical methods.

Figure 1: BSE-IT Index Price Trend Over 2014 to 2024

Figure 1 is the BSE-IT Index, which reflects consistent growth between 2014 and 2018, with a sudden jump after 2020 due to digitalization, possibly. Higher volatility from 2022 indicates market realignment. The pattern reflects IT industry growth, pandemic-led demand, and economic factors dictating price action during the decade.

Figure 2 shows the relative performance of four forecasting models (ARIMA, GARCH, XGBoost, and LSTM) on the basis of five different performance metrics.

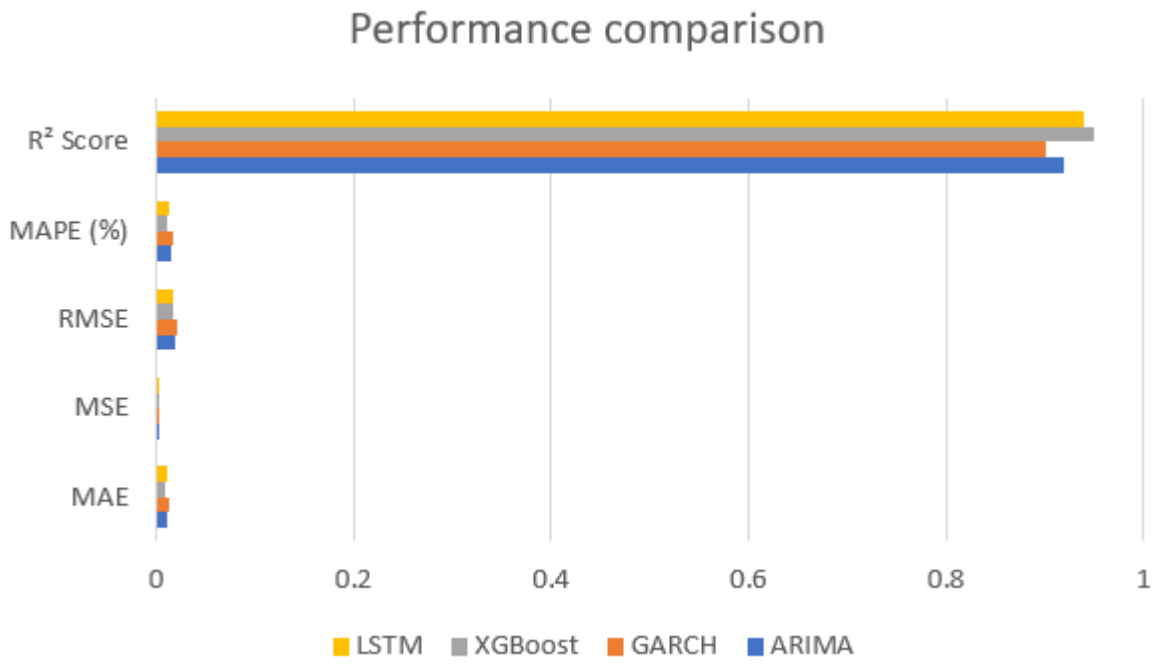

Figure 2: Performance comparison on LSTM, XGBoost, GARCH and ARIMA models

Mean Absolute Error (MAE): Measures average absolute difference between observed and forecasted values. XGBoost has the lowest MAE (0.01), which means higher precision. Mean Squared Error (MSE) & Root Mean Squared Error (RMSE): Smaller values represent better performance. The lowest MSE (0.0003) and RMSE (0.017) are from XGBoost. Mean Absolute Percentage Error (MAPE) is the average error expressed as a percentage. XGBoost's minimum MAPE is 1.20%, with virtually zero relative prediction errors. R^2 Score: To what extent the model explains data variance (the higher the value, the closer to 1 it is). XGBoost has the highest R^2 (0.95), with LSTM coming second (0.94). Among the models, XGBoost has the smallest error and highest R^2, making it the most reliable predictive model in this experiment.

5 Conclusion

The current research compared the performance of ARIMA, GARCH, XGBoost, and LSTM models in predicting stock prices. According to the results, machine learning models (XGBoost and LSTM) are superior to classical statistical models (ARIMA and GARCH) in accuracy and predictability. Amongst all the models, XGBoost performed best in terms of lowest error values (MAE, MSE, RMSE, MAPE) and highest R^2 value (0.95), and therefore it is the most precise stock price prediction model. LSTM also performed well, confirming the superiority of deep learning techniques. The results support the improved predictability given by advanced machine learning methods and how they could serve as accurate tools for predictions in finance. Work can persist in testing the combined models and other sources of market variables for further refining the prediction accuracy.

Further work can proceed with improving stock price prediction based on hybrid models that combine statistical techniques (ARIMA, GARCH) and machine learning (XGBoost, LSTM) for their benefits. Adding macroeconomic factors such as inflation and interest rates can be useful to frame a general market environment. Sentiment analysis of news and social media can reflect investor sentiment, enhancing predictive power. Sophisticated feature engineering incorporating technical indicators and non-traditional data sources might enhance model inputs. Explainable AI methods can boost decision-

making transparency. Real-time forecasting with adaptive learning models can, in turn, enhance effective implementation in practice, guaranteeing dynamic market adaptability for optimal financial forecasts.

References:

1. Alzaman, C. (2024). Deep learning in stock portfolio selection and predictions. *Expert Systems with Applications*, *237*, 121404. https://doi.org/https://doi.org/10.1016/j.eswa.2023.121404
2. Ashrafzadeh, M., Taheri, H. M., Gharehgozlou, M., & Hashemkhani Zolfani, S. (2023). Clustering-based return prediction model for stock pre-selection in portfolio optimization using PSO-CNN+MVF. *Journal of King Saud University - Computer and Information Sciences*, *35*(9), 101737. https://doi.org/https://doi-org.elibpondiuni.remotexs.in/10.1016/j.jksuci.2023.101737
3. Challa, M. L., Malepati, V., & Kolusu, S. N. R. (2018). Forecasting risk using auto regressive integrated moving average approach: an evidence from S&P BSE Sensex. *Financial Innovation*, *4*(1), 24. https://doi.org/10.1186/s40854-018-0107-z
4. Chen, Y., Wu, J., & Wu, Z. (2022). China's commercial bank stock price prediction using a novel K-means-LSTM hybrid approach. *Expert Systems with Applications*, *202*, 117370. https://doi.org/https://doi.org/10.1016/j.eswa.2022.117370
5. Dave, E., Leonardo, A., Jeanice, M., & Hanafiah, N. (2021). Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM. *Procedia Computer Science*, *179*, 480–487. https://doi.org/https://doi-org.elibpondiuni.remotexs.in/10.1016/j.procs.2021.01.031
6. Dioubi, F., Hundera, N. W., Xu, H., & Zhu, X. (2024). Enhancing stock market predictions via hybrid external trend and internal components analysis and long short term memory model. *Journal of King Saud University - Computer and Information Sciences*, *36*(10), 102252. https://doi.org/https://doi.org/10.1016/j.jksuci.2024.102252
7. Gülmez, B. (2025). GA-Attention-Fuzzy-Stock-Net: An optimized neuro-fuzzy system for stock market price prediction with genetic algorithm and attention mechanism. *Heliyon*, *11*(3), e42393. https://doi.org/https://doi.org/10.1016/j.heliyon.2025.e42393
8. Hu, Y., Ni, J., & Wen, L. (2020). A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction. *Physica A: Statistical Mechanics and Its Applications*, *557*, 124907. https://doi.org/https://doi-org.elibpondiuni.remotexs.in/10.1016/j.physa.2020.124907
9. Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, *178*, 115019. https://doi.org/https://doi.org/10.1016/j.eswa.2021.115019
10. Kakade, K., Jain, I., & Mishra, A. K. (2022). Value-at-Risk forecasting: A hybrid ensemble learning GARCH-LSTM based approach. *Resources Policy*, *78*, 102903. https://doi.org/https://doi-org.elibpondiuni.remotexs.in/10.1016/j.resourpol.2022.102903
11. Kuo, R. J., & Chiu, T.-H. (2024). Hybrid of jellyfish and particle swarm optimization algorithm-based support vector machine for stock market trend prediction. *Applied Soft Computing*, *154*, 111394. https://doi.org/https://doi.org/10.1016/j.asoc.2024.111394
12. Latha, C. M., Bhuvaneswari, S., & Soujanya, K. L. S. (2022). Stock Price Prediction using HFTSF Algorithm. *2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 1053–1059. https://doi.org/10.1109/I-SMAC55078.2022.9987378
13. Latha, C. M., Bhuvaneswari, S., & Soujanya, K. L. S. (2024). *Optimized FOREX Rate Prediction Using Hybrid Machine Learning Algorithm* (pp. 31–41). https://doi.org/10.1007/978-981-97-5412-0_3
14. Li, Y., Bu, H., Li, J., & Wu, J. (2020). The role of text-extracted investor sentiment in Chinese stock price prediction with the enhancement of deep learning. *International Journal of Forecasting*, *36*(4), 1541–1562. https://doi.org/https://doi.org/10.1016/j.ijforecast.2020.05.001
15. Malepati, V., Challa, M. L., & Kolusu, S. N. R. (2019). *Stock Market Volatility Modeling and Forecasting with a Special Reference to BSE Sensex* (pp. 105–118). https://doi.org/10.1108/S0196-382120190000035005
16. Malepati, V., Latha, C. M., & Nageswararao, K. S. (2021). Determinants of Access to Finance by MSMEs in Andhra Pradesh. *SEDME (Small Enterprises Development, Management & Extension Journal): A Worldwide Window on MSME Studies*, *48*(2), 203–219. https://doi.org/10.1177/09708464211066630

17. Mutinda, J. K., & Langat, A. K. (2024). Stock price prediction using combined GARCH-AI models. *Scientific African*, *26*, e02374. https://doi.org/https://doi.org/10.1016/j.sciaf.2024.e02374
18. Rather, A. M. (2021). LSTM-based Deep Learning Model for Stock Prediction and Predictive Optimization Model. *EURO Journal on Decision Processes*, *9*, 100001. https://doi.org/https://doi.org/10.1016/j.ejdp.2021.100001
19. Venkataramanaiah, M., Madhavi Latha, Ch., & Siva Nageswara Rao, K. (2018). Determinants of Dividend Policy in the Indian Corporate Sector : A Study of Companies Listed on Nifty 50, NSE. *Indian Journal of Finance*, *12*(1), 37. https://doi.org/10.17010/ijf/2018/v12i1/120740
20. Wang, J., Liu, J., & Jiang, W. (2024). An enhanced interval-valued decomposition integration model for stock price prediction based on comprehensive feature extraction and optimized deep learning. *Expert Systems with Applications*, *243*, 122891. https://doi.org/https://doi.org/10.1016/j.eswa.2023.122891
21. Yang, Q., Yu, Y., Dai, D., He, Q., & Lin, Y. (2024). Can hybrid model improve the forecasting performance of stock price index amid COVID-19? Contextual evidence from the MEEMD-LSTM-MLP approach. *The North American Journal of Economics and Finance*, *74*, 102252. https://doi.org/https://doi.org/10.1016/j.najef.2024.102252
22. Yun, K. K., Yoon, S. W., & Won, D. (2021). Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. *Expert Systems with Applications*, *186*, 115716. https://doi.org/https://doi.org/10.1016/j.eswa.2021.115716