Hybrid Intelligence for Financial Forecasting: Uniting LSTM with XG Boost

Krishna Mittal[1], M Akhil[2], Mohammed Bande Nawaz[3], Mohammed Mujeeb[4]

Abstract- *Predicting price of any stock is difficult to achieve because of volatility of market conditions. In fraction of seconds markets go up and down, fluctuations and mood of the investor these makes traditional methods less accurate to predict. Normal statistical methods do not detect small changes in stock market even though that are important for understanding the market and many of the times single ml model become unstable with market changes. In this study, we tried a prediction method that combines the 'Long Term Short Memory' network with the gradient boosting procedure (XG Boost). Long Term Short Memory (LSTM) helps to understand how the price changes with the time and it finds the underlying timely based pattern in past data(historical) related to stocks. These observations, combined with a group of produced market measures are given or passes to the XG Boost, which helps show patterns and irregular movements that the LSTM model alone skips or failed to notice. The objective of this mixed approach is to form predictions that remain accurate in various or changing conditions of the market and give more accurate changing pattern(nature) of stock data. Tests conducted on NIFTY 50 index and various NSE- listed stocks reveal that the integrated approach outperforms isolated LSTM, XG Boost, and baseline models, as measured by lower RMSE and MAE values, underscoring its reliability and real-world potential.*

Keywords: Long term sequence model, Gradient boosting model, Combined model, Sequence data study, Square Root-Error Metric, Absolute Error Metric.

I. INTRODUCTION

Financial markets behave like dynamic and highly reactive systems, shaped by a mix of economic signals, world events, investor psychology, and overall public sentiment.

Because these factors constantly interact and influence one another, price movements can shift unpredictably, making it difficult to forecast them accurately. Traditional approaches such as ARIMA or basic regression rely on fixed linear assumptions, so their performance often drops when markets become volatile or undergo major changes. Deep learning models like LSTMs are better at identifying time-based patterns in historical prices, but they may fail to capture deeper connections between multiple financial variables. In contrast, models like XG Boost are strong at detecting complex non-linear relationships, yet they cannot naturally account for the sequential nature of time-series data. So finally, to come out from this problem we came with a mixed approach that combines LSTM and XG Boost in network. The LSTM feature helps find the time-based visualization(patterns) in past price, then the XG Boost take this result with created features to discover extra connections. This combined method allows both models recompense for the ineffectiveness of one another, making assumptions or predictions remain more balanced and responsible in varying market situations.

II. LITERATURE SURVEY

1. Growth of Deep Learning in Financial Forecasting (2020)

Sezer, Mehmet Ugar Gudelek, and Ahmet Murat Ozbayoglu (2020) analyzed advanced techniques like deep learning to forecast the price of the stock over the time from the past or historical data. They came to know that LSTM model is mostly used because they are best to grasp the patterns of market movements.

2. Hybrid Deep Models for Improving Accuracy (2020)

Dash and Dash, 2020 found that combined models give achieve better results compared to single model prediction by improving steadiness and reducing mistakes.

3. XG Boost in High-Dimensional Market Data (2021)

Mahmud et al. (2021) noted that XG Boost tends to do very well in datasets rich in technical indicators due to its handling of nonlinearity and feature interactions.

4. LSTM Strength in Volatile Markets (2021)

Gao and Chai 2021 found that LSTM networks are effective in capturing abrupt changes during volatile market conditions and outperform traditional RNNs.

5. Two-Stage Prediction Frameworks (2021)

According to Gunduz and Cataltepe (2021), the multi models-stage-that first learn temporal patterns and then make refinements with feature-based learners-produce superior financial forecasts.

6. Technical Indicators Enhance ML Models (2022)

Indeed, Khan et al. (2022) confirmed that the inclusion of indicators, such as MACD, RSI, and Bollinger Bands, will raise the level of accuracy in machine-learning models' predictions.

7. Ensemble Learning for Financial Markets (2022)

According to Sohangir and Wang (2022), the ensemble methods which combined deep learning with gradient boosting showed more stable and consistent stock predictions compared to using single-model systems.

8. LSTM + Tree-Based Models Hybridization (2023)

Roy and Kumar (2023) proposed the LSTM-plus-tree model, demonstrating that tree-based learners improve residual correction, hence reducing RMSE in stock forecasting tasks.

9. The Importance of Feature Engineering in ML Finance (2024)

Alqahtani et al. (2024) emphasized the fact that features engineered from technical indicators distinctly improve predictive performance when coupled with boosting models.

10. Modern Deep Models vs. Traditional Methods (2025)

Li and Chen 2025 showed that modern deep-learning and hybrid models outperform ARIMA, SVM, and regression methods, especially in long-term forecasting for emerging markets.

III. PROPOSED METHODOLOGY

A. Data Acquisition and Preparation Daily stock metrics (opening, peak, trough, closing prices, and trading volume) were gathered for NIFTY 50 constituents and chosen NSE equities spanning 2016 to 2024Additional attributes were created, including momentum oscillators like RSI, trend signals such as MACD, volatility measures via Bollinger Bands, and smoothed averages Scaling techniques were employed to standardize the dataset, promoting consistent LSTM performance

B. LSTM Architecture Inputs consist of scaled sequences of prior price data.

The setup features stacked LSTM layers with regularization via dropout layers to curb overfitting risks.
Outputs yield forecasted prices that reflect learned temporal structures.

C. XG Boost Integration Inputs encompass crafted features plus LSTM- derived forecasts as supplementary variables the goal focuses on reducing mean squared residuals to polish initial LSTM estimates and uncover lingering trends.
Parameter optimization occurred through systematic grid search for peak efficiency.

D. Integrated Workflow Gather and clean raw stock information.
Compute technical metrics and form sliding sequence windows.
Fit the LSTM model using past data timelines.
Feed LSTM results combined with feature sets into XG Boost.
Assess final outputs with RMSE and MAE criteria.
This sequential process merges time-based pattern recognition with attribute-driven refinement effectively.

IV. EXPERIMENTAL RESULTS

A. Data Overview The study utilized NIFTY 50 index components and specific NSE stocks Time frame 2016 through 2024Data origin Yahoo Finance platform.

The last 5 years of analysis and prediction of a particular stock are shown in fig 1 and fig 2 respectively.

Fig 1 : Graph of past 5 years stock Analysis.

B. Performance Measures Root Mean Squared Error (RMSE), Mean Absolute Error (MAE)

C. Model Comparison The combined system shows marked superiority, adeptly integrating sequential insights with feature dynamics for enhanced accuracy.

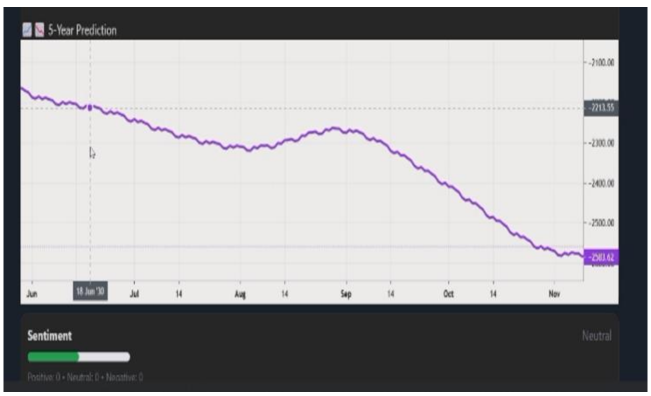

Fig 2 : Graph of next 5 years prediction.

V. CONCLUSION

This research outlines an innovative fusion of LSTM and XG Boost tailored for stock price forecasting.
LSTM manages the nuances of price evolution over time, with XG Boost elevating accuracy via tailored features and prior model outputs.
Validation through experiments affirms the systems edge over conventional stats-based tools and single ML variants.
Looking ahead, expansions could incorporate sentiment from news sources, broader economic factors, and adaptive learning methods to boost forecasting capabilities further.

REFERENCES

[1] Sezer, O. B., Mehmet Ugar Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial Time Series Forecasting with Deep Learning: A Systematic Literature Review. Applied Soft Computing.

[2] Dash, R., & Dash, P. (2020). A Hybrid Deep-Learning Model for Stock Price Prediction. Journal of Intelligent Systems.

[3] Mahmud, M., Kaiser, M. S., Rahman, M. A., et al. (2021) Performance of XG Boost in High-Dimensional Financial Data. IEEE Access.

[4] Gao, Y., & Chai, J. (2021). Stock Market Volatility Prediction Using LSTM Networks. Computational Economics.

[5] Gunduz, H., & Cataltepe, Z. (2021). A Multi-Stage Deep Learning Framework for Forecasting Financial Time Series. Expert Systems with Applications.

[6] Khan, S., Raza, F., & Ahmed, N. (2022). Effect of Technical Indicators on Machine Learning-Based Stock Forecasting Models. International Journal of Finance & Economics.

[7] Sohangir, S., & Wang, D. (2022). Ensemble Approaches for Deep Financial Prediction. Neural Computing and Applications.

[8] Roy, A., & Kumar, R. (2023). A Hybrid LSTM–Tree-Based Model for Stock Price Forecasting. Journal of Forecasting.

[9] Alqahtani, F., Alshamrani, A., & Hussain, S. (2024). Feature Engineering Approaches for Machine-Learning-Based Financial Forecasting. Computational Finance Review.

[10] Li, H., & Chen, Y. (2025). Evaluating Modern Deep-Learning Models for Emerging-Market Stock Prediction. International Journal of Data Science & Analytics.