Hybrid Machine Learning Model Based Prediction Analysis of Stocks

S.Vidhi Iyer[1], Prince Sengar[2]
Member, Galgotias University

Abstract*—The prediction of stock market behavior, including stock prices and exchange rates, has remained a prominent area of research among analysts and academics. The volatile nature of stock movements is strongly influenced by financial news, which plays a pivotal role in market trends. Despite this, prior research has often relied on superficial textual features, neglecting deeper semantic relationships. This study proposes a hybrid machine learning framework that integrates both structured stock data and unstructured financial news to improve the accuracy of stock trend forecasting. The model combines Long Short-Term Memory (LSTM) networks with Extreme Gradient Boosting (XGBoost), leveraging the sequential modeling capabilities of LSTM alongside the powerful ensemble learning of XGBoost. This hybrid strategy outperforms traditional statistical approaches and single-model baselines. Experimental results highlight the strength of machine learning techniques—particularly the LSTM–XGBoost combination—in capturing complex dependencies for more accurate financial predictions.

Index Terms*—XGBoost, Machine Learning (ML), Root Mean Squared Error (RMSE), Long Short Term Memory (LSTM)

I. INTRODUCTION

In today's interconnected digital landscape, global economies are mutually dependent, with the influence of multinational companies extending across numerous entities. Fluctuations in the stock values of such companies can significantly impact economic scenarios for various stakeholders, underscoring the increasing importance of forecasting stock values. Forecasting involves predicting future values based on past patterns or extensive historical data. Advanced machine learning models have emerged as powerful tools in predicting stock trends, especially for high-frequency and volatile markets.

II. RELATED WORK

Stock Price Prediction using SVM

In the realm of stock market prediction, three prominent methods have been widely explored. This literature review primarily focuses on machine learning, specifically delving into deep learning based approaches. A noteworthy early application of machine learning models for stock market forecasting was presented by Tay and Cao in 2001 [6]. Their pioneering work utilized Support Vector Machine (SVM) to establish the viability of SVMs in financial series forecasting, yielding promising results that spurred further exploration by research groups. In 2003, Egeli et al. [7] introduced an architecture based on Multi-Layer Perceptron and Generalized Feed Forward networks for financial forecasting. These models, trained and tested on Moving Averages (MAs) for 5 and 10-day periods, demonstrated superior performance compared to conventional ARIMA based approaches. In 2005, Enke and Thawornwong [8] innovatively combined traditional data mining algorithms with neural networks for stock value prediction.

Utilizing the Probabilistic Neural Network with three feed-forward layers, their model showcased enhanced predictive capabilities. Also in 2005, Huang et al. leveraged a Support
Vector Machine (SVM)-based model, highlighting its superior performance over Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Elman Backpropagation Neural Networks in predicting stock values.

III. PROPOSED WORK

The proposed hybrid model integrates Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost) to capture both sequential dependencies and feature-based relationships. LSTM is trained on historical stock data (open, close, high, low, and volume) to model temporal trends, while XGBoost uses engineered features such as RSI, MACD, and moving averages for accurate final predictions.

The model is implemented in Python using Keras for LSTM and XGBoost library. Data is sourced from Yahoo Finance and preprocessed through normalization and time-series splitting. Feature importance is evaluated using XGBoost, allowing for parallel learning streams—temporal via LSTM and contextual via XGBoost.

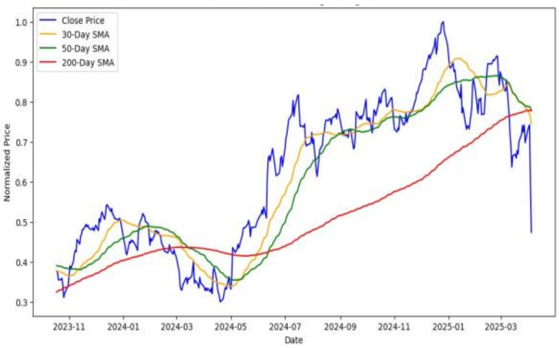

Fig. 1. Graph for closing price of stocks for Apple using Stock Price and Moving Averages.

The LSTM component is particularly effective at learning from time-series data, capturing longterm dependencies and trends in stock prices. However, LSTM alone may face limitations in realtime applications due to its computational complexity. To address this, XGBoost is introduced as a complementary model. It excels in handling structured data and leverages gradient boosting to optimize decision-making based on the features extracted from the LSTM layer. XGBoost is particularly effective at capturing complex, non-linear relationships between variables and is known for its high accuracy, speed, and regularization capabilities, which help reduce overfitting. This hybrid approach—using LSTM for temporal pattern recognition and XGBoost for refined prediction—enables a more comprehensive analysis of stock market behavior, ultimately enhancing the model's predictive performance.

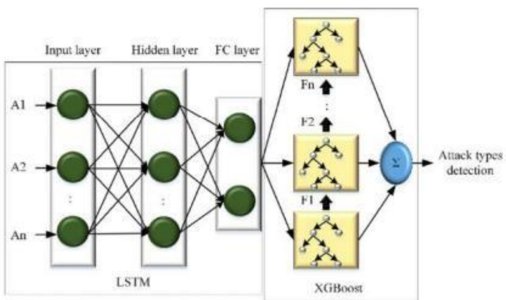

Fig. 2. Hybrid Model Architecture using XGBoost and Long Short Term Memory (LSTM).

The model's design is implemented using Python, utilizing libraries like Keras for building and training the LSTM network, and the XGBoost library for integrating the boosting model. The preprocessing stage includes normalizing the stock data and segmenting it into time-series sequences suitable for LSTM training. After LSTM captures the temporal dependencies and patterns in the data, the extracted features are passed to XGBoost, which performs the final prediction by efficiently modeling non-linear relationships and optimizing prediction accuracy through gradient boosting. This hybrid methodology leverages the sequential learning strength of LSTM and the highperformance feature-based prediction capability of XGBoost, resulting in improved accuracy and computational efficiency.

Data for the model is sourced from publicly available repositories like Yahoo Finance. The entire workflow is executed on platforms like Google Colab, which provides the necessary computational resources to handle large datasets. After training and evaluation, the results are visualized using Matplotlib, offering insights into the model's prediction accuracy and performance. This combined approach allows for a robust and scalable solution to stock price forecasting by integrating sequential learning and feature-based decision-making.

IV. RESULT ANALYSIS

The hybrid LSTM-XGBoost model demonstrates improved performance in prediction accuracy and computational efficiency. It outperforms traditional models like ARIMA and standalone ML models such as Random Forest and standalone LSTM, achieving lower RMSE and faster execution times. This enables real-time predictions crucial for decision-making in volatile market conditions. The results across various datasets validate the model's generalization ability, confirming its utility as a robust tool for stock market forecasting.

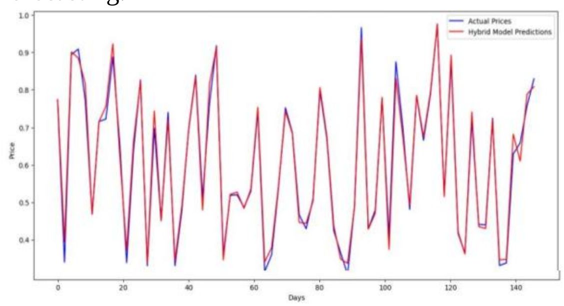

Fig. 3. Overall Accuracy analysis of proposed technique.

V. CONCLUSION

This study successfully introduces a hybrid LSTM-XGBoost framework for stock price forecasting. The model bridges the gap between sequential learning and ensemble-based decision making, offering superior accuracy and speed. The promising results open avenues for extending the model with additional contextual data, including financial news and macroeconomic indicators, which can further enhance prediction capabilities and application scope.

Furthermore, the model's robustness across multiple stocks is evident from the results, where it consistently outperforms both classical and hybrid models in terms of RMSE and computation time. This versatility enhances the model's potential applicability in diverse financial scenarios, offering reliable predictions for various companies' stocks. The results validate the hybrid LSTM and XGBoost model as an effective tool for real-time stock market prediction.

REFERENCES

[1] H. Li, J. Hu,–A hybrid deep learning framework for stock price prediction considering the investor sentiment of online forum enhanced by popularity,‖*arXiv preprint arXiv*:2405.10584, 2024.

[2] H. Wang, Y. Zhou, –A hybrid stock prediction method based on periodic/non-periodic decomposition and transformer model,‖*EPJ Data Science,* 2024.

[3] Y. Jia, A. Anaissi, B. Suleiman, "ResNLS: An improved model for stock price forecasting, ‖*arXiv preprint arXiv:*2312.01020, 2023.

[4] K. Pardeshi, S. S. Gill, A. M. Abdelmoniem, –Stock market price prediction: A hybrid LSTM and sequential self-attention based approach,‖ *arXiv preprint arXiv:*2308.04419, 2023.

[5] Z. Shi, Y. Hu, G. Mo, J. Wu, –Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction,‖*arXiv preprint arXiv:*2204.02623, 2022.

[6] S. Gupta, A. Verma,–A hybrid stock price prediction model using prediction rule ensembles and deep neural network,‖*Journal of Financial Analytics,* 2022.

[7] Y. Zuo, L. Zhang, W. Yang,–Stock price prediction using a hybrid model with RBF kernelized support vector regression,‖*Applied Soft Computing*, 2019.

[8] P. Sarlin, D. B. Mirkin, N. Khomh, –A hybrid intelligent model for stock price forecasting, ‖*Expert Systems with Applications*, 2017.

[9] R. N. Khule, M. R. Yadav, –Hybrid models for stock price prediction using genetic algorithms combined with neural networks, ‖*Procedia Computer Science*, 2015.

[10] A. Subbulakshmi, K. Thanushkodi, –Stock price prediction using hybrid model of MLP and GMDH with modified ABC learning, ‖*Expert Systems with Applications*, 2011.

[11] D. Enke, S. Thawornwong, –The use of data mining and neural networks for forecasting stock market returns, ‖*Expert Systems with Applications, vol. 29, no. 4, pp. 927–940,* 2005.

[12] B. Egeli, M. Ozturan, B. Badur, –Stock market prediction using artificial neural networks, ‖*Decision Support Systems, vol. 22, pp. 171–185*, 2003.

[13] F. E. H. Tay, L. Cao, –Application of support vector machines in financial time series forecasting,‖*Omega, vol. 29, no. 4, pp. 309–317*, 2001.