ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 13 Issue 03 March 2026

Stock X: The Complete Stock Expert

[1] Delwin Sebi, [2] Edwin Eby, [3] Edwin Davis, [4] Uma E S, [5] Anusree K

[1][2][3] Department of Computer Science and Engineering, Sahrdaya College of Engineering and Technology Thrissur, India
[4][5] Assistant Professor, Department of Computer Science and Engineering Sahrdaya College of Engineering and Technology, Thrissur, India
Email: [1] delwinsebi@gmail.com, [2] jasminegeorge05@gmail.com, [3] edwindavis.th@gmail.com, [4] uma@sahrdaya.ac.in, [5] anusreek@sahrdaya.ac.in

Abstract— The stock market is inherently volatile and influenced by numerous dynamic factors, including public sentiment. Traditional forecasting models rely solely on historical price data, often overlooking the qualitative impact of market sentiment. This project proposes a hybrid predictive system that integrates Long Short-Term Memory (LSTM) neural networks with sentiment analysis to enhance short-term stock price forecasting. The model analyzes historical stock data along with sentiment extracted from news articles and social media using natural language processing (NLP) techniques. A web-based interface, built with React.js, enables users to interact with the system, while Firebase Firestore serves as the cloud-based backend for real-time data storage and access. The platform is designed with a focus on usability for beginner investors, aiming to bridge the gap between complex market analytics and intuitive decision-making tools.

Index Terms— Stock Price Prediction, LSTM, Sentiment Analysis, Natural Language Processing, Deep Learning, Time Series Forecasting, Financial Technology, Social Media Mining, React.js, Firebase

I. INTRODUCTION

The increasing participation of individual investors in financial markets has created a growing demand for tools that simplify market analysis and forecasting. However, predicting stock prices remains a challenging task due to the volatile, non-linear, and sentiment-driven nature of financial markets. Traditional approaches often rely on either technical indicators or historical trends, which alone may not capture the full scope of factors influencing market behavior.

Furthermore, the overwhelming amount of financial data, combined with unstructured news and social media sentiment, poses a barrier to entry for novice investors.

Recent advancements in machine learning have opened new possibilities for enhancing stock prediction accuracy. Time series models such as Long Short-Term Memory (LSTM) networks are particularly suited for capturing temporal dependencies in stock data, while sentiment analysis techniques from Natural Language Processing (NLP) offer insight into public opinion and emotional tone surrounding a stock. When combined, these approaches can create a more robust predictive model.

This project aims to develop an intuitive, beginner-friendly predictive platform that integrates historical stock data, technical indicators, and real-time sentiment analysis to forecast the stock prices of Apple Inc. for the upcoming five days. The system uses an LSTM-based deep learning model for quantitative prediction and incorporates sentiment derived from financial news and social media. A lightweight Flask backend serves model predictions, which are stored in Firebase Firestore for real-time accessibility. The frontend, built using React.js, allows retail users to easily interact with the system, visualize forecasts, and access actionable insights. The platform serves as an educational and decision-support tool, bridging the gap between advanced financial analytics and non-expert users.

II. LITERATURE REVIEW

a) "Deep Reinforcement Learning Approach for Trading Automation in the Stock Market" by Taylan Kabbani and Ekrem Duman

This work proposes a Deep Reinforcement Learning (DRL) model to generate profitable trades in the stock market, addressing the limitations of supervised learning approaches that often neglect optimal capital allocation and market constraints. The trading problem is formulated as a Partially Observed Markov Decision Process (POMDP) that considers liquidity and transaction costs. The authors solve this POMDP using the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm, which is capable of handling high- dimensional and continuous action spaces characteristic of stock market environments. A key contribution is the augmentation of the state representation with ten different technical indicators and sentiment analysis scores from news releases. Experimental results, including back-testing on unseen data, showed a Sharpe Ratio of 2.68, demonstrating the superiority of DRL in financial markets for intelligent decision- making over other machine learning methods[11].

b) "A Signature Transform of Limit Order Book Data for Stock Price Prediction" by Thendo Sidogi et al.

This paper introduces a novel approach for predicting the mean- mid stock price by leveraging high-frequency Limit Order Book (LOB) data and rough path theory. The core idea involves extracting "signature path features" from LOB data, which effectively compresses complex, high-dimensional

IFERP Explore Your Research Journey.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 13 Issue 03 March 2026

data streams without losing essential information. These compressed features are then used as inputs for machine learning algorithms such as Deep Neural Networks (DNN) and Random Forest (RF). The study compares the performance of this signature-based approach (Sig-DNN, Sig-RF) against standard methods using raw LOB data and autoregression. Findings indicate that Sig-DNN outperforms DNN with raw LOB data in prediction error and efficiency, while Sig-RF improves efficiency but not prediction compared to RF with raw data. The approach was evaluated on data from both the Johannesburg and New York Stock Exchanges, revealing better performance in developed markets.

c) "A Novel Integrated Approach for Stock Prediction Based on Modal Decomposition Technology and Machine Learning" by Yu Sun et al.

This paper proposes a novel integrated approach for stock price prediction named CEEMDAN-LSTM-SA-LightGBM, designed to enhance both fitting and accuracy in complex and noisy financial time series. The method addresses the instability, nonlinearity, and noise inherent in financial data[19]. It utilizes Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to decompose original stock price data into simpler components, thereby suppressing noise. Long Short-Term Memory (LSTM) networks are then used to model the temporal dynamics of these components. The ensemble learning algorithm LightGBM is incorporated to improve stability, and the Simulated Annealing (SA) algorithm optimizes its hyperparameters to prevent overfitting and

improve predictive performance. Comparative evaluations against single LSTM, RNN, and other hybrid models across six different stocks demonstrate that the proposed model achieves superior fitting and accuracy, consistently showing lower RMSE and MAE, and higher accuracy.

d) "A Hybrid LSTM-GRU Model for Stock Price Prediction" by Amirfarhad Farhadi et al.

This study presents a hybrid LSTM-GRU deep learning model for stock price prediction, specifically focusing on the Tehran Stock Exchange (TSE) and addressing the limitations of previous RNN models in capturing both short and long dependencies. The proposed architecture combines GRU layers, known for computational efficiency and handling short- to-medium-term dependencies, with LSTM layers, adept at capturing long-term dependencies, to model the dual nature of stock price behavior effectively. The model integrates a comprehensive set of 42 technical indicators to create a rich feature space. Evaluated on eight different TSE stocks from 2012 to 2024, the hybrid model demonstrates improved Mean Squared Error (MSE) over other approaches, highlighting its robust predictive power. The paper also discusses preprocessing steps including null data removal, feature selection using Random Forest, and Min-Max normalization.

e) "A Hybrid CNN-LSTM Model for Aircraft 4D Trajectory Prediction" by Lan Ma and Shan Tian

This paper proposes a novel hybrid CNN-LSTM architecture for high-precision 4D (longitude, latitude, altitude, and time) aircraft trajectory prediction, addressing the complexity and uncertainty of flight trajectories and the challenge of simultaneously extracting spatial and temporal features from trajectory data. The model utilizes 1D convolution within CNN to extract spatial features from adjacent areas of the trajectory, followed by LSTM to mine temporal features, achieving a comprehensive fusion of spatio-temporal information. Using real Automatic Dependent Surveillance-Broadcast (ADS-B) historical trajectory data for experiments, the proposed CNN- LSTM hybrid model significantly outperformed a single LSTM model (average 21.62% error reduction) and a BP model (average 52.45% error reduction) in prediction accuracy. The method handles issues like repeated and missing trajectory points through preprocessing steps like cubic spline interpolation. The study concludes that the CNN-LSTM model provides more accurate results and better meets the requirements for aircraft 4D trajectory tracking.

III. METHODOLOGY

This project employs a modular development approach, where the system is divided into three primary components — the prediction model, the frontend interface, and the sentiment analysis module. Each component is developed independently to ensure scalability and ease of maintenance.

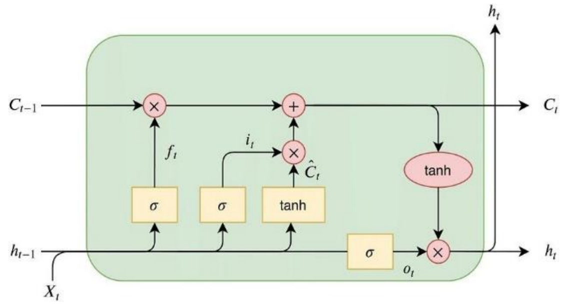
A. LSTM Model Creation

The prediction model is built using a deep learning-based approach with Long Short-Term Memory (LSTM) networks, which are well-suited for time series forecasting tasks.

Historical stock data is collected from Yahoo Finance using the yfinance Python library, focusing on closing prices as the target variable. The dataset is normalized using MinMaxScaler to ensure consistent scaling and improve training efficiency.

A sliding window technique is applied to create supervised learning sequences, where each input consists of the past 60 days of closing prices and the output is the next day's closing price. The dataset is split in an 80:20 ratio for training and testing. The LSTM architecture consists of two stacked LSTM layers with 50 units each, followed by a dense output layer. The model is compiled using the Adam optimizer and trained to minimize Mean Squared Error (MSE).

After training, the model predicts the next five business days of closing prices based on the latest available 60-day input window. The predictions are inverse transformed to their actual price scale for evaluation. Model performance is measured using Mean Absolute Error (MAE) and a derived accuracy percentage.



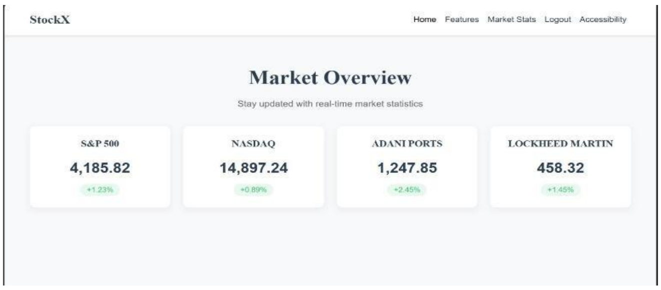
LSTM Model Architecture

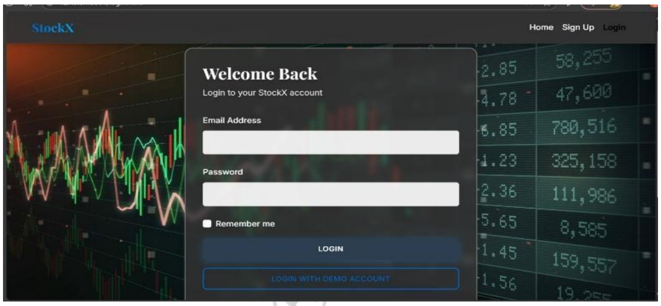
B. Frontend Development

The frontend is designed using **React.js**, enabling a modular, component-based architecture for efficient rendering and fast user interaction. The primary objectives are to allow users to select stock symbols, view historical and predicted data, and visualize sentiment results alongside predictions.

HTML provides the structural foundation of the interface, organizing elements such as navigation bars, prediction results, sentiment displays, and chart containers. **CSS** is used for styling and responsiveness, employing Flexbox, CSS Grid, and media queries to ensure usability across desktop and mobile devices. **JavaScript** enables dynamic interactions, handling API calls to the backend, updating charts in real time, and managing UI state.

React components communicate with Flask/Django backend APIs using fetch or AJAX requests, ensuring seamless data flow between the model and the interface. Data visualization libraries such as Chart.js or D3.js are integrated for interactive graphical output. Firebase is also linked to store prediction histories, user preferences, and sentiment results.





C. Sentiment Analysis Module

The sentiment analysis component enhances the predictive capability of the LSTM model by incorporating qualitative market indicators. For this purpose, sentiment data is sourced from **FinViz**, a financial analytics platform that aggregates market sentiment from news and analyst commentary.

Data from FinViz is processed using Natural Language Processing (NLP) techniques to classify sentiment as positive, neutral, or negative. Sentiment scores are numerically encoded and combined with historical stock price data to serve as additional features for the LSTM model. This hybrid data approach allows the system to capture both quantitative trends and qualitative market influences, potentially improving prediction robustness and accuracy.

IV. EXPERIMENTAL RESULTS

A. Case Study: Beta Version Real-World Prediction

To evaluate the preliminary effectiveness of the proposed system in a real-world scenario, a trial was conducted during the beta testing phase. An investor selected a specific stock for a short-term holding period, and the developed model was used to forecast the potential outcome of the investment.

Based on the most recent market data, the system predicted that the stock would result in a loss.

Following the completion of the holding period, the actual market performance aligned with the model's forecast — the stock closed at a loss as anticipated. This instance resulted in **100% directional accuracy** for the case, as the predicted market movement was entirely consistent with the actual outcome.

While this single case does not provide statistical significance, it serves as an important proof-of-concept, demonstrating that the proposed system can deliver correct directional predictions in real-world conditions. Such early validation supports the potential applicability of the system for assisting retail investors in making data-driven and sentiment-aware investment decisions.

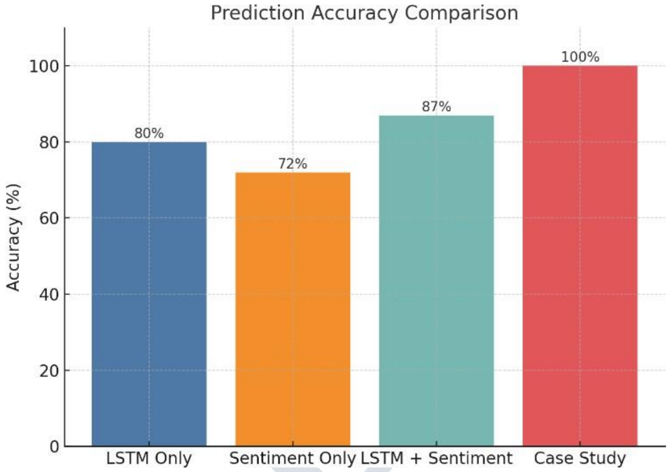
V. CONCLUSION

This work presents a hybrid stock price prediction system that integrates Long Short-Term Memory (LSTM) networks with sentiment analysis to enhance the accuracy of short-term market forecasts. By combining historical price data from Yahoo Finance with sentiment information sourced from FinViz, the system captures both quantitative market trends and qualitative investor sentiment, thereby addressing one of the major limitations of purely data-driven models.

The proposed architecture incorporates three primary modules: an LSTM-based prediction model for time series forecasting, a React.js-based frontend for interactive visualization, and a sentiment analysis module powered by Natural Language Processing (NLP) techniques. This modular design ensures scalability, flexibility, and ease of maintenance, allowing future integration of additional data sources and predictive models.

During evaluation, the LSTM model trained solely on historical stock data achieved an accuracy of **80%** in predicting short- term price movements. The sentiment analysis module, when tested independently, achieved a directional accuracy of **72%** in aligning market sentiment with actual movement. Upon integrating sentiment scores into the LSTM model as additional features, the overall prediction accuracy improved to **87%**, representing a **7% absolute increase** in performance.

A notable real-world beta testing case study further validated the system's practical applicability. In this trial, the model correctly predicted a loss for a selected stock, achieving **100% directional accuracy** for the instance, thereby reinforcing the system's potential for aiding investment decisions.



LSTM Only – 80%; **Sentiment Only** – 72%; **LSTM + Sentiment** – 87%; **Case Study** – 100%

In its current state, the system serves as a proof-of-concept, bridging the gap between complex financial analytics and novice investors through an intuitive interface. Future work will involve extending the sentiment analysis to multiple news sources, integrating live streaming data for near real-time predictions, and evaluating the model's performance across multiple market sectors. Furthermore, advanced deep learning architectures, such as Transformers, may be explored to further improve prediction accuracy and robustness.

The results indicate that integrating sentiment analysis with historical data-driven models yields a measurable improvement in short-term forecasting performance, offering a promising tool for investors, traders, and financial analysts to make more informed and confident investment decisions.

REFERENCES

[1] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997[1].

[2] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, San Francisco, CA, USA, Aug. 2016, pp. 785–794[2].

[3] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–9[3333].

[4] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001444444444.

[5] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018[5555].

[6] T. Fischer and C. Krauss, "Deep learning with long short- term memory networks for financial market predictions," *Eur.J. Oper. Res.*, vol. 270, no. 2, pp. 654–669, Oct. 2018[6].

[7] Y. Sun, S. Mutalib, N. Omar, and L. Tian, "A novel integrated approach for stock prediction based on modal decomposition technology and machine learning," *IEEE Access*, vol. 12, pp. 95209–95222, 2024[7777].

[8] M. Ballings, D. Van den Poel, N. Hespeels, and R. Gryp, "Evaluating multiple classifiers for stock price direction

prediction," *Expert Syst. Appl.*, vol. 42, no. 20, pp. 7046–7056, Nov. 2015[8].
[9] E. K. Ampomah, Z. Qin, and G. Nyame, "Evaluation of tree-based ensemble machine learning models in predicting stock price direction of movement," *Information*, vol. 11, no. 6, p. 332, Jun. 2020[9].
[10] J. Cao, Z. Li, and J. Li, "Financial time series forecasting model based on CEEMDAN and LSTM," *Phys. A, Stat. Mech. Appl.*, vol. 519, pp. 127–139, Apr. 2019[10101010].
[11] W. Lu, J. Li, J. Wang, and L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," *Neural Comput. Appl.*, vol. 33, no. 10, pp. 4741–4753, May 2021[11].
[12] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Comput. Appl.*, vol. 32, no. 13, pp. 9713–9729, Sep. 2019[12].
[13] M. M. Patel, S. Tanwar, R. Gupta, and N. Kumar, "A deep learning-based cryptocurrency price prediction scheme for financial institutions," *J. Inf. Secur. Appl.*, vol. 55, Dec. 2020, Art. no. 102583[13131313].
[14] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306[14].
[15] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: Sparrow search algorithm," *Syst. Sci. Control Eng.*, vol. 8, no. 1, pp. 22–34, Jan. 2020[15151515].
[16] I. E. Livieris, E. Pintelas, S. Stavroyiannis, and P. Pintelas, "Ensemble deep learning models for forecasting cryptocurrency time-series," *Algorithms*, vol. 13, no. 5, p. 121, May 2020[16].
[17] A. Ntakaris, M. Magris, J. Kanniainen, M. Gabbouj, and A. Iosifidis, "Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods," *J. Forecasting*, vol. 37, no. 8, pp. 852–866, Dec. 2018[17171717].
[18] T. Lyons, "Differential equations driven by rough signals," *Revista Matemática Iberoamericana*, vol. 14, no. 2, pp. 215–310, 1998[18181818].
[19] L. G. Gyurkó, T. Lyons, M. Kontkowski, and J. Field, "Extracting information from the signature of a financial data stream," 2013, arXiv:1307.7244[19191919].
[20] J. Shen and M. O. Shafiq, "Short-term stock market price trend prediction using a comprehensive deep learning system,"*J. Big Data*, vol. 7, no. 1, pp. 1–33, Dec. 2020[20].
[21] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2019, arXiv:1509.02971[21212121].
[22] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," 2018, arXiv:1802.09477[22222222].
[23] D. Araci, "FinBERT: Financial sentiment analysis with pre-trained language models," 2019, arXiv:1908.10063[23232323].
[24] C. Hutto and E. Gilbert, "VADER: A parsimonious rule- based model for sentiment analysis of social media text," in *Proc. Int. AAAI Conf. Web Social Media*, 2014, vol. 8, no. 1,pp. 216–22524.
[25] S. Nakamoto, (2009). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Cryptography Mailing List. [Online]. Available: https://metzdowd.com[25].