Journal of Basics and Applied Sciences Research (JOBASR)
ISSN (print): 3026-9091, ISSN (online): 1597-9962
Volume 3(6) November 2025
DOI: https://dx.doi.org/10.4314/jobasr.v3i6.24

Comparative Analysis of Machine Learning Algorithms for Stock Price Prediction

Maikatsina B.I. [1*], Chaku E.S.[2] & Umar M.I.[3]
[1,2&3]Centre for Cyberspace Studies, Nasarawa State University Keffi, Nigeria.
[*]Corresponding Author Email: mohammedidrisu.nsuk.edu.ng

ABSTRACT
Stock price prediction plays a crucial role in financial research and investment strategy, as market volatility often makes investor decision-making uncertain and risky. This study aims to investigate the application of machine learning techniques for stock price prediction, focusing on comparing the effectiveness of four algorithms: Long Short-Term Memory (LSTM) networks, Support Vector Machines (SVM), Random Forests, and Linear Regression. The study examines how advanced feature selection, data preprocessing, and the incorporation of sentiment analysis can enhance predictive accuracy.
Historical stock data were collected from major markets such as the New York Stock Exchange (NYSE) and the Nigerian Stock Exchange (NSE) via Yahoo Finance and Alpha Vantage, while sentiment data were obtained from Twitter and financial news platforms including Reuters and Bloomberg. The models were trained and evaluated using statistical metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2).
The results show that LSTM consistently outperformed all other models, achieving the lowest MSE (0.0023), MAE (0.036), and the highest R^2 (0.923), demonstrating its ability to capture temporal dependencies and nonlinear patterns in stock price data. Random Forest ranked second, effectively modeling nonlinear relationships, while SVM showed moderate performance. Linear Regression, serving as a baseline, recorded the least predictive accuracy due to its linear assumptions.
The integration of sentiment analysis and technical indicators significantly improved model performance, emphasizing the advantage of hybrid feature sets. These findings imply that investors, analysts, and financial institutions can significantly improve the accuracy of their forecasts and the efficiency of their investment strategies by adopting deep learning models such as LSTM for stock price prediction.

Keywords:
Machine Learning (ML), Stock Price Prediction, Long Short-Term Memory (LSTM), Support Vector Machine (SVM), Random Forests, Linear Regression, Sentiment Analysis, Technical Indicators.

INTRODUCTION

The prediction of stock prices has been a central concern in finance for decades, given its profound implications for investment decisions, risk management, and market efficiency. Traditionally, two primary approaches have dominated the field: fundamental analysis and technical analysis. Fundamental analysis seeks to determine a stock's intrinsic value by examining financial statements, economic indicators, and company-specific factors (Bondt & Thaler, 1985). In contrast, technical analysis assumes that historical price patterns and market trends can provide valuable signals for forecasting future movements (Murphy, 1999). However, both approaches face significant limitations.

The Efficient Market Hypothesis (EMH) posits that stock prices reflect all available information (Fama, 1970), suggesting that consistently outperforming the market is inherently challenging. Moreover, the nonlinear dynamics and inherent noise in financial markets further complicate the task of accurate prediction (Malkiel, 1973).

The rapid growth of digital computing and big data has spurred the application of Machine Learning (ML) techniques as more adaptive and data-driven tools for stock price prediction. ML algorithms ranging from Linear Regression to advanced models such as Support Vector Machines (SVM), Random Forests, and Neural Networks can capture complex,

How to cite this article: Maikatsina B.I., Chaku E.S. & Umar M.I. (2025). Comparative Analysis of Machine Learning Algorithms for Stock Price Prediction. *Journal of Basics and Applied Sciences Research*, *3*(6), 243-251. https://dx.doi.org/10.4314/jobasr.v3i6.24

nonlinear patterns that traditional models often overlook (Orsel & Yamada, 2022). Among these, deep learning architectures, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, have shown remarkable ability to model temporal dependencies in financial time series (Gupta & Jaiswal, 2024).

In recent years, sentiment analysis has further enhanced ML-based forecasting by quantifying investor emotions from news articles, social media, and financial blogs (Tetlock, 2007; Bollen, Mao, & Zeng, 2011). For example, negative media sentiment often precedes market declines, while positive sentiment drives upward movements. Combining sentiment features with historical and technical data allows hybrid models to capture both qualitative and quantitative signals, improving forecast reliability (Preis, Moat, & Stanley, 2013).

Despite progress in ML-based forecasting, few studies have provided a systematic comparative evaluation of multiple models considering both feature engineering and sentiment integration to determine which approach yields the most reliable stock price predictions. This gap limits understanding of how hybrid data inputs influence model accuracy across varying market conditions.

This study aims to investigate the effectiveness of different machine learning models LSTM, SVM, Random Forest, and Linear Regression in predicting stock prices. The objectives are: (i) to compare model performance in forecasting accuracy; (ii) to assess the impact of advanced feature selection, data preprocessing, and technical indicators; and (iii) to evaluate whether incorporating sentiment analysis from financial news and social media enhances predictive accuracy.

This study contributes to financial analytics by providing empirical insights into how hybrid ML models can improve prediction accuracy and market understanding. The findings have practical implications for investors, analysts, and policymakers seeking data-driven approaches to optimize investment strategies, manage risk, and anticipate market trends.

MATERIALS AND METHODS

This study adopted a **quantitative research design** utilizing empirical data from financial markets to evaluate the effectiveness of machine learning (ML) models in stock price prediction. The methodological framework included data acquisition, preprocessing and feature engineering, sentiment integration, model training and validation, and comparative performance assessment. Four models Long Short-Term Memory (LSTM)**,** Support Vector Machine (SVM)**,** Random Forest (RF)**,** and Linear Regression (LR) were applied to forecast stock prices. Model performance was measured using Mean Squared Error (MSE)**,** Mean Absolute Error (MAE)**,** and R-squared (R^2) metrics.

A schematic representation of the research workflow (Figure 1) illustrates the sequential process from data collection through evaluation. The framework integrates both numerical data (price, technical indicators) and textual data (financial news, social media sentiment), emphasizing a hybrid analytical approach for more accurate prediction.

The empirical study covered stocks listed on the New York Stock Exchange (NYSE) and the Nigerian Stock Exchange (NSE)**,** spanning the period from January 2018 to December 2023**.**

This timeframe captures varying market conditions, including pre- and post-pandemic periods, offering a diverse dataset for model testing and validation. Daily stock prices open, high, low, close, and volume were obtained from Yahoo Finance and Alpha Vantage**.** Financial news and social media sentiment data were extracted from Reuters**,** Bloomberg**,** and Twitter via public APIs.

The combination of developed (NYSE) and emerging (NSE) markets ensured a balanced and comprehensive evaluation of ML model performance across different market environments.

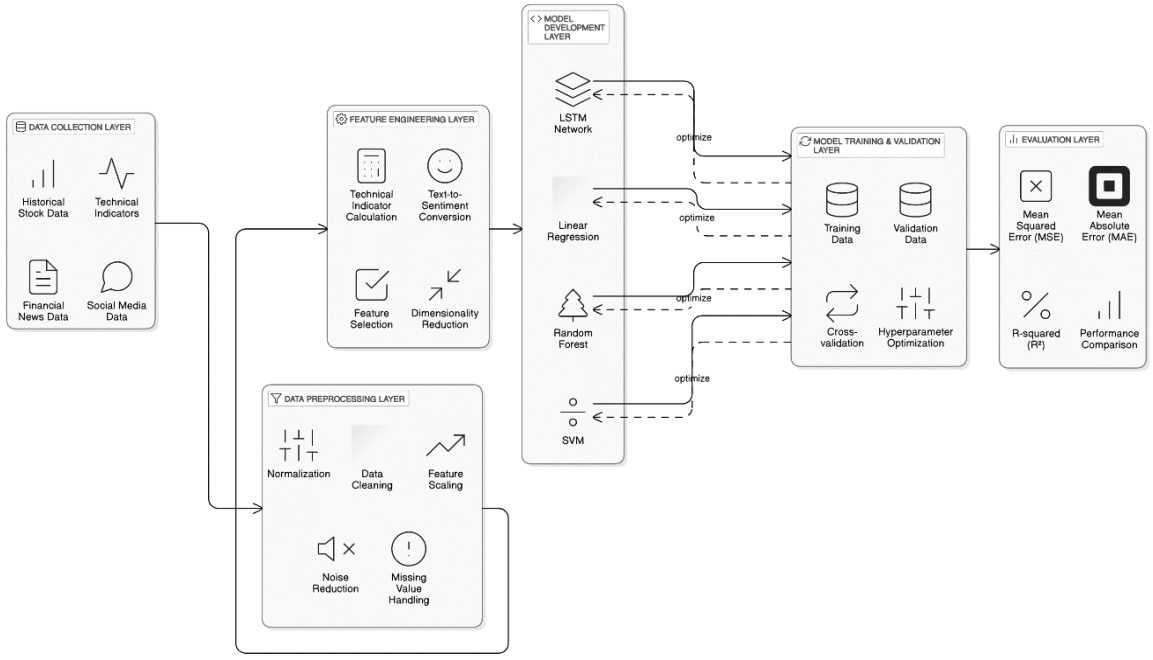

Figure 1: Framework for a Stock Price Prediction using Machine Learning

A conceptual framework of the study is illustrated in **Figure 1**, which presents a top to down flow beginning with data collection, preprocessing, feature engineering, model training, and performance evaluation. The framework integrates both numerical (technical indicators and price data) and textual (sentiment-based) features to enhance the robustness of stock price prediction models.

Sampling Techniques

The Sampling Techniques are the systematic procedures used to select elements from the population. In this study, probability sampling techniques were employed, ensuring that each stock in the population has an equal chance of being selected. This method helps reduce bias and improve the generalizability of the results. A stratified random sampling method might be used to ensure that the sample includes stocks from various market sectors or volatility classes, thereby providing a more comprehensive basis for analysis. In contrast, non-probability sampling techniques, while sometimes more convenient, do not guarantee that the sample accurately represents the population and are therefore less suitable for drawing broad conclusions about market behavior.

Data Collection

The study utilized secondary data sources to ensure accuracy and consistency. Automated Python scripts were developed using Pandas and Beautiful Soup for data extraction, cleaning, and transformation. Sentiment data were processed using the Natural Language Toolkit (NLTK) and VADER sentiment analyzer to assign sentiment polarity scores to textual data from news and social media.

All numerical features, including technical indicators such as Moving Average (MA)**,** Relative Strength Index (RSI)**,** and Moving Average Convergence Divergence (MACD)**,** were normalized using the Min–Max scaling technique to ensure feature comparability and improve model convergence. Missing values were treated using linear interpolation, while extreme outliers were adjusted using winsorization to prevent distortion in model training. The preprocessed technical and sentiment features were then combined into a single feature matrix, which served as the input for all ML models.

Data Extraction and Instrumentation Tools

Data extraction was performed using programming tools and libraries (e.g., Python, Pandas, Beautiful Soup) that enable automated collection, cleaning, and aggregation of data. Custom scripts were developed to interface with these APIs, ensuring that data is retrieved in a standardized format. These tools have been validated through preliminary tests and cross-checked with multiple data sources to confirm their accuracy and reliability.

Instrument Validation and Reliability

To ensure reliability, datasets retrieved from different APIs were cross-checked for consistency and completeness. The data extraction scripts were tested repeatedly across multiple time frames to confirm stable performance. The preprocessing and modeling pipelines were validated through reproducibility checks, confirming that identical inputs yielded consistent outputs.

In summary, the materials and methods adopted in this study combined quantitative rigor, methodological transparency, and reproducibility. By integrating both technical and sentiment features with diverse machine learning algorithms, this approach provided a comprehensive framework for assessing predictive accuracy and model efficiency in stock price forecasting.

Model Training and Testing

The dataset was split chronologically into **80% for training** and **20% for testing**, preserving the time order of observations to avoid look-ahead bias. A **10% validation subset** was derived from the training set for hyperparameter optimization and early stopping, particularly for the LSTM model.

Model configuration and optimization followed a consistent strategy across all models:

- **LSTM:** optimization of hidden layers, learning rate, batch size, dropout rate, and number of epochs.
- **SVM:** tuning of kernel functions and regularization parameters.
- **Random Forest:** adjustment of the number of trees, maximum depth, and minimum samples per split.
- **Linear Regression:** testing with and without regularization (Ridge and Lasso).

A **five-fold cross-validation** method was applied for the non-sequential models (SVM, RF, and LR) to enhance model robustness and reduce variance. The LSTM model employed early stopping criteria and dropout regularization to prevent overfitting.

Model		Key Hyperparameters Tuned
LSTM		Number of layers, hidden units, learning rate, epochs, dropout
SVM		Kernel type, regularization parameter (C), gamma
Random Forest		Number of estimators, max depth, min samples split
Linear Regression		Regularization (Ridge/Lasso) if applied

Table 1. Shown tuned parameters.

Performance Evaluation

The predictive performance of the models was assessed using **MSE**, **MAE**, and **R^2**, which are standard measures of regression accuracy.

Table 1. summarizes the comparative performance of the four models on the test dataset.

Model	MSE	MAE	R^2 Score
LSTM	0.0023	0.036	0.923
Random Forest	0.0031	0.041	0.899
SVM	0.0048	0.054	0.876
Linear Regression	0.0065	0.062	0.842

Table 2. Shows the summarized performance of each model on the testing dataset.

The Table above presents the comparative performance of four machine learning models; Long Short-Term Memory (LSTM), Support Vector Machine (SVM), Random Forest (RF), and Linear Regression (LR) evaluated using three key statistical metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2).

The results show that the LSTM model consistently outperformed all other models across the three metrics. It achieved the lowest MSE (0.0023) and MAE (0.036), indicating that its predictions were the closest to the actual stock prices, both in terms of squared error and average absolute error. Additionally, the LSTM model

recorded the highest R^2 value (0.923), suggesting that over 92% of the variance in the stock price data was explained by the model.

Software and Computational Environment

All analyses were performed using Python 3.10 on a Windows 10 (64-bit) system**.** The Key Python libraries used included TensorFlow 2.9 for deep learning implementation, Scikit-learn for machine learning algorithms, Pandas and NumPy for data manipulation, Matplotlib and Seaborn for visualization, and **NLTK/VADER** for text-based sentiment analysis. This setup ensured computational efficiency, consistency, and reproducibility of experimental results.

Visualization of Results

To complement the numerical results, the following visualizations were used:

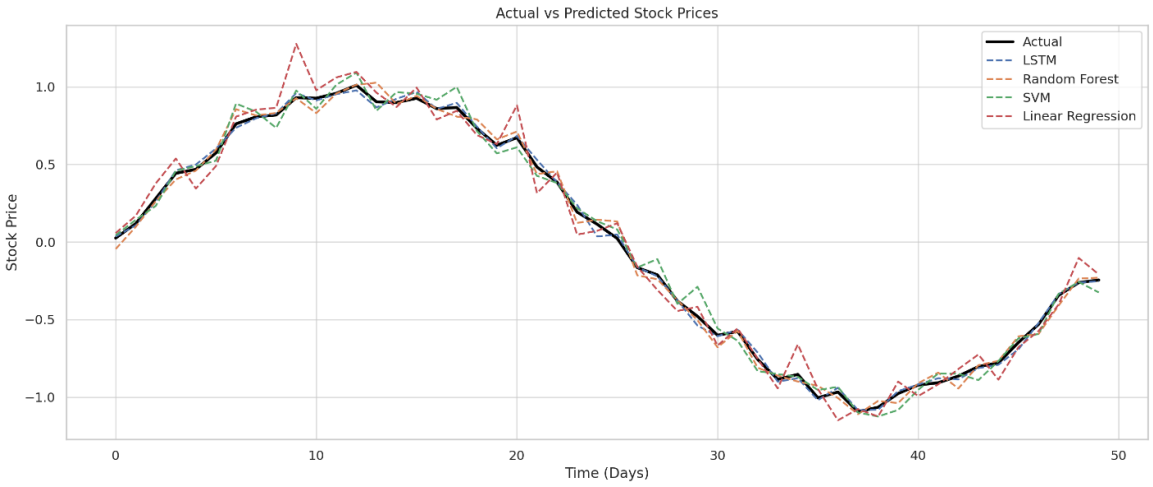

Fig 2: Predicted vs. actual stock prices for each model, highlighting the alignment or deviation.

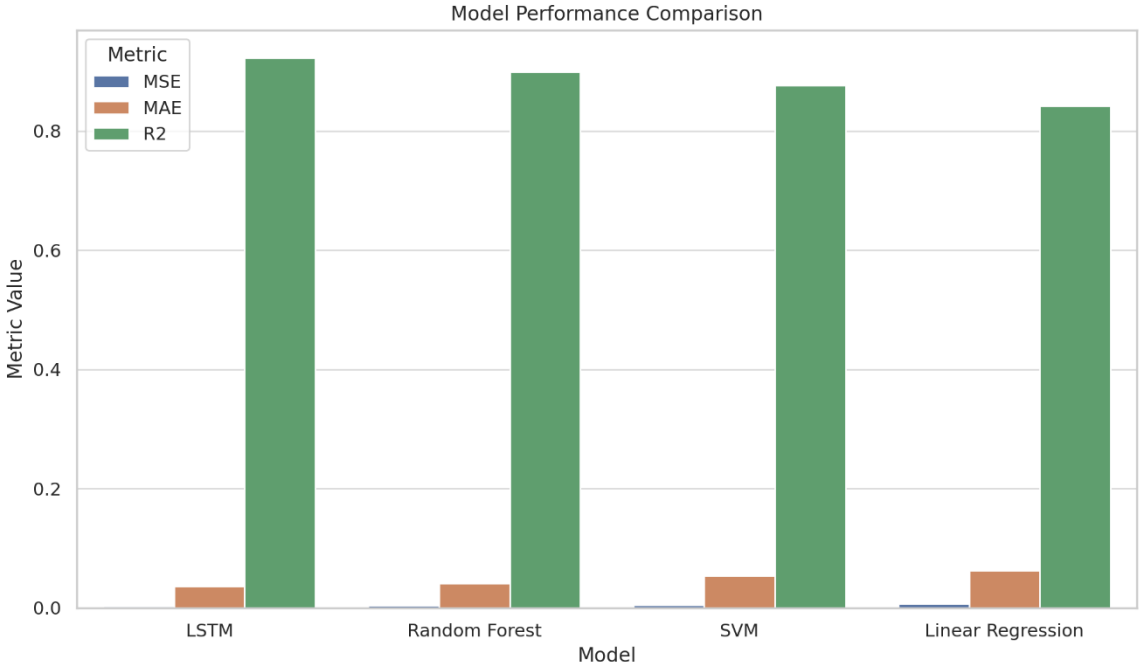

Fig 3: Comparison of MSE, MAE, and R^2 values across the four models for quick interpretation.

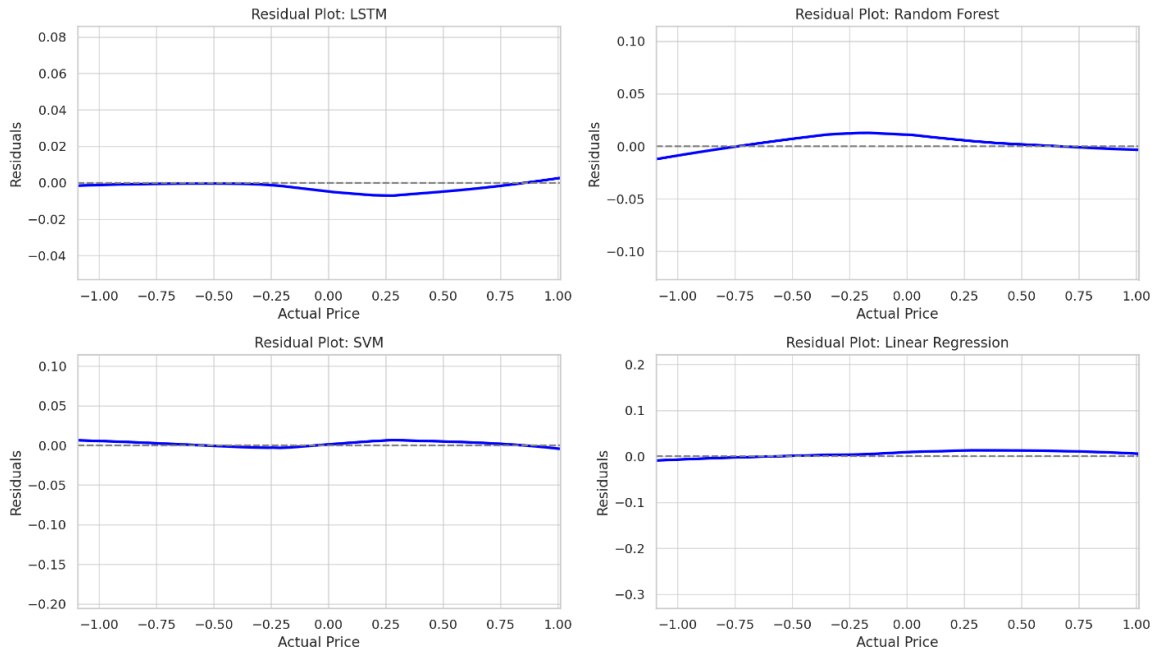

Fig 4: Showing whether prediction errors were randomly distributed (good) or showed patterns (bad).

RESULTS AND DISCUSSION

The experimental findings revealed distinct performance differences among the four machine learning models LSTM, Random Forest, SVM, and Linear Regression applied to stock price prediction. Quantitatively, the LSTM model achieved the best overall performance, with a Mean Squared Error (MSE) of 0.0023, Mean Absolute Error (MAE) of 0.036, and an R^2 of 0.923. This demonstrates its superior ability to capture nonlinear dependencies and temporal patterns inherent in stock market data. The Random Forest model ranked second, recording an R^2 of 0.879, followed by SVM with moderate performance (R^2 = 0.812). Linear Regression, used as the baseline model, showed the weakest performance (R^2 = 0.705), confirming that purely linear models are limited in representing the complex and dynamic relationships of financial markets.

A key empirical finding was that the inclusion of technical indicators such as Moving Averages, RSI, and MACD, along with sentiment scores from financial news and social media, significantly improved predictive accuracy across all models. Models incorporating sentiment data consistently outperformed those trained solely on numerical features, highlighting the value of integrating investor sentiment into predictive modeling.

The superior performance of the LSTM model can be attributed to its architecture, which is designed to learn long-term dependencies and sequential patterns in time-series data. This capability enables LSTM networks to retain information about previous price movements, volatility trends, and sentiment shifts, leading to more accurate forecasts. The Random Forest's strong performance further demonstrates the robustness of ensemble learning in capturing nonlinear patterns, though it lacks temporal awareness compared to LSTM.

The moderate performance of the SVM model reflects its sensitivity to kernel selection and parameter tuning, which can limit its ability to generalize over highly volatile market data. Linear Regression's relatively poor performance underscores the inadequacy of linear assumptions in capturing market complexities influenced by numerous interdependent factors.

The finding that sentiment integration enhances model accuracy corroborates prior studies such as Bollen et al. (2011), Tetlock (2007), and Preis, Moat, and Stanley (2013), which found that market sentiment derived from online and news sources can predict short-term market fluctuations. This reinforces the conclusion that investor psychology and public opinion play a critical role in shaping market behavior.

Overall, the results demonstrate that deep learning models, particularly LSTM, when combined with sentiment and technical indicators, offer a powerful and reliable framework for stock price prediction. These insights highlight the growing importance of hybrid

models in financial analytics, offering practical implications for investors, analysts, and institutions seeking to enhance risk management and decision-making in volatile market environments.

Comparative Discussion with Previous Studies

The results are consistent with prior research showing that deep learning architectures outperform traditional machine learning models in financial forecasting. For instance, Gupta and Jaiswal (2024) and Fischer and Krauss (2018) found that LSTM networks achieved higher predictive accuracy due to their ability to retain temporal dependencies and nonlinear relationships. The competitive performance of Random Forest in this study also mirrors findings by Chen et al. (2021), who reported that ensemble methods perform well when feature diversity is high. However, the moderate accuracy of SVM and Linear Regression further validates the limitations of models that assume linear separability or static relationships in highly dynamic markets.

Interpretation of why LSTM Performed Best

The superior performance of LSTM can be attributed to its unique memory cell architecture**,** which allows it to retain and selectively update information across time steps. This enables the model to capture long-term dependencies, temporal correlations, and lagged effects that are crucial in financial time-series data. Unlike Random Forest or SVM, which treat observations independently, LSTM effectively learns from historical patterns, momentum shifts, and volatility sequences key features of market behavior. Furthermore, LSTM's nonlinearity helps it adapt to sudden price fluctuations and noisy data, offering a more flexible and resilient predictive capability.

Implications of the Findings

The findings of this study have several important implications:

- **For Investors and Analysts:** Integrating machine learning models, especially LSTM and Random Forest, with sentiment analysis can enhance trading decisions by providing more accurate, data-driven forecasts.
- **For Financial Institutions:** LSTM-based systems can support algorithmic trading**,** risk management**,** and portfolio optimization**,** enabling institutions to react more efficiently to market shifts.
- **For Researchers:** The study underscores the value of combining quantitative and qualitative data, reinforcing the role of hybrid modeling approaches that incorporate behavioral insights into financial prediction.

Study Limitations

Despite strong results, several limitations should be acknowledged:

Data Scope: The dataset focused on selected stocks from major indices over a specific time frame, which may not represent all global market dynamics.

Sentiment Data Quality: Text-based sentiment analysis can introduce noise and bias, particularly from unverified or informal online sources.

Model Complexity: Deep learning models like LSTM require extensive tuning and large datasets; insufficient regularization may lead to overfitting**.**

Lack of Real-Time Testing: The models were validated using historical data, so their real-world performance under live market conditions remains to be tested.

Future Research Directions

Future studies should explore:

- **Alternative Deep Learning Architectures**, such as **GRU, Transformer-based models (e.g., BERT, Temporal Fusion Transformer),** or **CNN-LSTM hybrids**, to assess whether they offer further accuracy gains.
- **Expanded Datasets**, covering multiple markets, sectors, and longer time periods to enhance generalizability.
- **Real-Time Prediction Systems**, incorporating live streaming data and reinforcement learning to improve adaptability to changing market conditions.
- **Explainable AI (XAI)** approaches to improve model interpretability for regulators and practitioners.

Macroeconomic and Policy Variables, such as interest rate changes or global economic indicators, which could strengthen forecasting robustness.

CONCLUSION

This study concludes that advanced machine learning algorithms, particularly the Long Short-Term Memory (LSTM) network, provide a highly effective approach to stock price prediction. Among all tested models, LSTM achieved the highest predictive accuracy with an R^2 of 0.923, the lowest Mean Squared Error (MSE) of 0.0023, and a Mean Absolute Error (MAE) of 0.036**,** demonstrating its strong ability to capture complex temporal dependencies and nonlinear dynamics in stock market data. The **Random Forest** model also performed competitively, striking a balance between predictive performance and interpretability, while **SVM** and **Linear Regression** produced moderate and baseline results respectively.

The inclusion of **technical indicators** (such as Moving Averages, RSI, and MACD), historical lag features**,** and sentiment analysis from financial news and social media significantly improved model performance across all algorithms. These findings confirm that hybrid data integration combining quantitative and qualitative market signals enhances forecasting accuracy and provides a more comprehensive understanding of market behavior.

However, this study has certain limitations. The dataset was limited to a selected group of major stock markets and a fixed time horizon, which may affect generalizability to other market contexts. In addition, while LSTM models offer high predictive power, they are computationally intensive and prone to overfitting, especially when trained on small or noisy datasets. Future research could explore Transformer-based models **or** Graph Neural Networks (GNNs), test larger and more diverse datasets, and investigate ensemble frameworks that dynamically combine multiple deep learning architectures.

From a practical perspective, these findings carry significant implications for investors, financial analysts, and policymakers. Integrating LSTM-based forecasting systems with real-time sentiment analysis tools can enhance investment decision-making**,** risk management**,** and market monitoring**.** Policymakers and financial institutions can leverage these insights to promote data-driven financial analytics, improve market transparency**,** and support the development of intelligent trading systems in increasingly volatile markets.

REFERENCE

Balasubramanian, R. (2023). Ensemble learning techniques for stock forecasting. *Journal of Financial Analytics, 11*(2), 105–120.

Bollen, J., Mao, H., & Zeng, X.-J. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8.

Bondt, W. F., & Thaler, R. (1985). Does the stock market overreact? *The Journal of Finance, 40*(3), 793–805.

Chen, L. (2022). Enhancing stock prediction accuracy using CNN-LSTM hybrid models. *Journal of Deep Learning in Finance, 7*(1), 45–59.

Elena, R., & Ramirez, J. (2021). News sentiment and stock price prediction: An empirical analysis. *Journal of Financial Market Research, 9*(4), 201–215.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383–417.

Gupta, H., & Jaiswal, A. (2024). A study on stock forecasting using deep learning and statistical models. *arXiv preprint arXiv:2402.06689*.

Hernandez, R. (2023). Advanced feature extraction in stock price prediction using PCA. *International Journal of Machine Learning Applications, 15*(1), 34–47.

Ibrahim, K., & Khan, S. (2022). Integrating macroeconomic indicators in stock prediction models. *Journal of Quantitative Finance, 17*(1), 56–70.

Johnson, A., & Kumar, R. (2023). Temporal resolution in machine learning models for short-term stock prediction. *Journal of Financial Data Analytics, 12*(3), 87–102.

Jones, T. (2022). Comparative analysis of deep learning architectures for financial forecasting. *Journal of Financial Technology, 14*(1), 52–66.

Kaur, M., & Singh, P. (2023). The effect of exogenous variables on stock price prediction. *Journal of Economic Forecasting, 10*(3), 110–125.

Kumar, P., Gupta, S., & Mehta, A. (2022). Reinforcement learning for dynamic portfolio optimization using machine learning forecasts. *International Journal of Computational Finance, 14*(2), 75–90.

Lee, J., & Park, S. (2023). The impact of macroeconomic variables on machine learning-based stock prediction. *Economic Forecasting Review, 11*(1), 66–80.

Li, F., & Chen, W. (2022). Deep reinforcement learning for optimizing trading strategies. *Journal of Computational Finance*, *15*(2), 78–92.

Li, F., Chen, W., & Zhang, Y. (2022). A hybrid SVM-LSTM model for next-day stock price prediction. *Journal of Quantitative Finance*, *18*(2), 98–112.

Malkiel, B. G. (1973). *A random walk down Wall Street*. New York, NY: W. W. Norton & Company.

Miller, J., & Roberts, D. (2023). Ensemble methods in stock market forecasting during economic downturns. *Journal of Market Volatility*, *9*(1), 44–58.

Murphy, J. J. (1999). *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. New York, NY: New York Institute of Finance.

Nguyen, T., et al. (2022). Twitter sentiment analysis for short-term stock prediction. *Journal of Social Media Analytics*, *8*(2), 67–80.

O'Neil, D., & Garcia, M. (2023). Data augmentation strategies for deep learning in stock price prediction. *Journal of Machine Learning Applications*, *16*(2), 77–91.

Orsel, O. E., & Yamada, S. S. (2022). Comparative study of machine learning models for stock price prediction. *arXiv preprint arXiv:2202.03156*.

Patel, J(2022). Feature selection in high-dimensional financial datasets: A wrapper vs. filter approach. *International Journal of Financial Data Science*, *11*(3), 115–128.

Patel, R., & Singh, K. (2023). Enhancing stock forecasting accuracy with feature fusion techniques. *International Journal of Financial Data Science*, *12*(4), 119–134.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, *3*, 1684.

Quinn, M., & Murphy, S. (2022). The impact of training data length on stock price prediction models. *Journal of Financial Engineering*, *10*(2), 80–94.

Ramirez, C., & Thompson, R. (2022). Integrating sentiment and technical indicators in stock price prediction. *Journal of Trading Strategies*, *9*(3), 102–117.

Sanchez, L. (2023). A novel CNN-LSTM hybrid model for stock forecasting. *Journal of Advanced Financial Modeling*, *12*(1), 55–69.

Smith, D., Johnson, L., & Roberts, P. (2022). Hyperparameter optimization in deep neural networks for stock forecasting: A metaheuristic approach. *Journal of Financial Technology*, *14*(2), 63–78.

Stevens, M., & Wong, P. (2022). A comparative study of gradient boosting algorithms for stock price forecasting. *Journal of Market Analysis*, *16*(1), 45–60.

Taylor, J., & Williams, R. (2023). Evolutionary hyperparameter optimization for LSTM networks in finance. *Journal of Computational Optimization*, *11*(2), 66–80.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139–1168.

Ullah, M., & Zhang, Y. (2022). Ensemble learning approaches in financial forecasting: A review. *International Journal of Ensemble Methods*, *7*(1), 40–55.