A Hybrid SARIMAX–LSTM Framework for Predicting Price Volatility in High-Tech Digital Markets: Evidence from NVIDIA

Flurey Martin[1,*]

[1]School of Computer Science and Electronic Engineering, University of Essex, United Kingdom

ABSTRACT

This study develops a Hybrid SARIMAX–LSTM model to improve the accuracy and robustness of stock price forecasting in digital and volatile financial markets. The model combines the linear and seasonal forecasting strengths of the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) with the nonlinear learning capability of the Long Short-Term Memory (LSTM) network. Using historical data from NVDIA Corporation, the hybrid framework was optimized through smart weighting to balance the contribution of both components. The results show that the model achieved a Root Mean Square Error (RMSE) of 8.59 and a coefficient of determination (R^2) of 0.9166, indicating that over 91 percent of price variance was accurately explained. Residual analysis confirmed unbiased predictions with normally distributed errors, demonstrating high stability and adaptability under volatile market conditions. Compared with individual models, the hybrid approach produced smoother and more consistent forecasts. Overall, the Hybrid SARIMAX–LSTM framework offers an interpretable and reliable tool for digital market forecasting and AI-based financial decision-making.

Keywords Hybrid SARIMAX–LSTM, Stock Forecasting, Digital Markets, Artificial Intelligence, Time Series

INTRODUCTION

In recent years, the rapid digitalization of financial systems and the expansion of algorithmic trading have reshaped the landscape of global capital markets. The increasing reliance on digital trading platforms, real-time data streams, and AI-driven analytics has elevated the importance of accurate and adaptive forecasting methods for financial decision-making. Stock price prediction, in particular, plays a critical role in digital finance by enabling investors, portfolio managers, and automated systems to anticipate market behavior and optimize asset allocation. High-tech firms such as NVIDIA Corporation exemplify the volatility and complexity of digital-era equities, where price fluctuations are driven not only by macroeconomic factors but also by micro-level events, sentiment dynamics, and technological innovation cycles. These challenges demand predictive models that are capable of learning complex, nonlinear relationships while maintaining transparency and reliability across diverse market conditions [1].

Traditional statistical models, such as the Autoregressive Integrated Moving Average (ARIMA) and its seasonal extensions like SARIMA and SARIMAX, have long been utilized for financial time series forecasting. Their appeal lies in their interpretability, mathematical rigor, and ability to

Submitted: 5 October 2025
Accepted: 15 December 2025
Published: 23 May 2026

Corresponding author
Flurey Martin,
fleum@essex.ac.uk

Additional Information and Declarations can be found on page 129

DOI: 10.47738/jdmdc.v3i2.63

How to cite this article: F. Martin, “A Hybrid SARIMAX–LSTM Framework for Predicting Price Volatility in High-Tech Digital Markets: Evidence from NVIDIA,” *J. Digit. Mark. Digit. Curr.*, vol. 3, no. 2, pp. 115-130, 2026.

model autoregressive and moving average patterns within stationary data [2]. Numerous studies have demonstrated their utility in predicting macroeconomic indicators, stock indices, and exchange rates. However, these models exhibit structural limitations when applied to volatile and nonlinear financial environments. They rely on linear assumptions and fixed-lag dependencies, which restrict their ability to capture the complex dynamics and sudden regime shifts common in modern stock markets. As digital markets evolve with increasing data complexity and trading frequency, the limitations of purely statistical approaches have become more apparent.

In parallel, advances in Artificial Intelligence (AI) and Deep Learning (DL) have revolutionized forecasting techniques by introducing models capable of learning temporal dependencies directly from data. Among these, the LSTM network has emerged as one of the most powerful architectures for sequential data modeling. Studies have shown that LSTM-based models outperform traditional time series methods in capturing nonlinearities and temporal correlations in stock price data [3]. While these models deliver superior predictive accuracy, they suffer from notable limitations such as a lack of interpretability, the risk of overfitting, and the need for large training datasets. Moreover, deep learning models are often data-hungry and computationally intensive, which restricts their scalability in real-world financial applications. These limitations underline a key theoretical and practical challenge in the field: finding a balance between the interpretability of statistical models and the adaptability of AI-based methods.

The state of the art in financial forecasting reflects a growing interest in hybrid modeling approaches that aim to integrate statistical and deep learning techniques [4]. Such models combine the structured interpretability of econometric frameworks with the nonlinear pattern recognition capabilities of neural networks. Prior research has highlighted the effectiveness of hybrid ARIMA–LSTM and similar frameworks in improving forecasting accuracy. However, existing studies have primarily focused on generic time series or macroeconomic data rather than technology-sector stocks that exhibit unique volatility characteristics. Furthermore, most existing hybrid approaches employ static weighting schemes between components, which limit their adaptability to changing market regimes. The lack of dynamic hybrid models that can automatically adjust their learning emphasis based on real-time performance remains an unresolved gap in the literature [5].

To address this gap, the present study introduces a Hybrid SARIMAX–LSTM model designed to combine the advantages of both statistical and deep learning approaches. The SARIMAX component captures long-term dependencies, seasonal structures, and exogenous influences such as market volume and volatility, while the LSTM component models short-term nonlinear dynamics and rapid market fluctuations. A key innovation in this framework is the incorporation of a smart weighting mechanism that dynamically adjusts the contribution of each component according to their predictive accuracy. This ensures that the model remains adaptive under varying market conditions. Using NVIDIA stock data as a representative case of high-volatility digital assets, the model is evaluated based on predictive accuracy, robustness, and residual behavior. The results demonstrate that the hybrid system significantly outperforms single-model benchmarks, achieving a high coefficient of determination ($R^2 = 0.9166$) and a low Root Mean Square Error (RMSE = 8.59).

In summary, this study contributes to both the theoretical and practical advancement of AI-assisted financial forecasting by providing an interpretable and adaptive framework for stock price prediction in digital markets. By bridging the gap between classical econometric modeling and modern deep learning techniques, the Hybrid SARIMAX–LSTM model offers a new methodological perspective for financial analytics. It provides a robust foundation for further exploration of hybrid forecasting architectures that can be applied to digital assets, cryptocurrencies, and other complex financial instruments. The research findings not only enhance academic understanding of hybrid time series modeling but also deliver practical implications for investment strategy optimization and risk management in the era of intelligent and data-driven financial systems.

Literature Review and Related Works

Forecasting stock prices has been a central topic in financial research due to its importance in investment strategy, risk management, and policy decision-making. Traditional time series models such as the Autoregressive Integrated Moving Average (ARIMA) and its seasonal variant, SARIMA, have been extensively used for forecasting due to their simplicity, interpretability, and solid statistical foundation [6], [7]. However, these linear models are limited in handling the nonlinear and non-stationary characteristics of modern financial data, particularly in volatile digital markets [8]. Their performance often deteriorates in the presence of structural breaks, irregular trading patterns, and high-frequency price movements, which are typical of technology sector equities such as NVIDIA.

To overcome these challenges, machine learning and deep learning models have emerged as powerful alternatives for financial time series prediction [9], [10]. Among these, the LSTM network has gained significant attention for its ability to capture long-term dependencies and nonlinear dynamics in sequential data [11]. Comparative studies have shown that LSTM-based architectures outperform traditional methods in predicting stock indices and digital asset prices [12], [13]. Nevertheless, despite their superior performance, deep learning models are often treated as black-box systems and can overfit when trained on limited datasets, leading to reduced interpretability and unstable generalization [14].

To address these limitations, researchers have proposed hybrid forecasting models that integrate statistical and deep learning approaches, leveraging the strengths of both paradigms [15]. In these hybrid architectures, statistical models such as ARIMA or SARIMA capture the linear components of the data, while LSTM or other neural networks model the residual nonlinear patterns [16]. Empirical evidence shows that such hybrid approaches significantly improve prediction accuracy and stability compared to standalone models [17]. For instance, hybrid ARIMA–LSTM frameworks have demonstrated superior performance in forecasting exchange rates, commodity prices, and stock indices by effectively combining interpretability with flexibility [18].

Recent studies have extended hybridization by incorporating exogenous variables and dynamic weighting mechanisms to improve adaptability under varying market conditions. These advancements enable the models to adjust their learning focus based on volatility or structural shifts, enhancing robustness and responsiveness [19]. Despite these improvements, few studies have specifically applied dynamically weighted hybrid models to high-technology equities, which exhibit unique volatility patterns driven by innovation cycles, investor sentiment, and rapid information diffusion. This research gap underscores the need for hybrid frameworks capable of balancing statistical trend modeling with deep learning adaptability in real-time digital market environments.

In light of these developments, this study introduces a Hybrid SARIMAX–LSTM model that integrates the Seasonal ARIMA model with Exogenous Variables (SARIMAX) and the LSTM network into a single adaptive framework. The model incorporates a smart weighting algorithm to dynamically balance the linear and nonlinear components based on predictive performance. Using NVIDIA stock price data as a representative example of high-volatility digital assets, this research aims to evaluate the hybrid model's predictive accuracy, residual behavior, and robustness compared to standalone approaches. By addressing the limitations of existing forecasting methods, the study contributes to the growing body of research on AI-driven financial forecasting, offering a practical and interpretable approach for digital market prediction.

Methodology

This study develops a Hybrid SARIMAX–LSTM model aimed at improving the accuracy and stability of stock price forecasting in digital and high-volatility financial markets. The proposed approach integrates the SARIMAX and the LSTM neural network into a single adaptive forecasting framework. The overall methodology follows a structured sequence consisting of five main stages: data collection, data preprocessing, model development, hybrid integration, and model evaluation. These stages are illustrated in figure 1, which presents the sequential research workflow implemented in this study. The figure outlines the process beginning with data acquisition from the NASDAQ database, followed by data transformation and stationarity testing, development of the SARIMAX and LSTM models, dynamic weighting integration, and final model validation. Each stage contributes to building a robust hybrid forecasting system capable of capturing both linear and nonlinear behaviors in NVIDIA's stock price series.

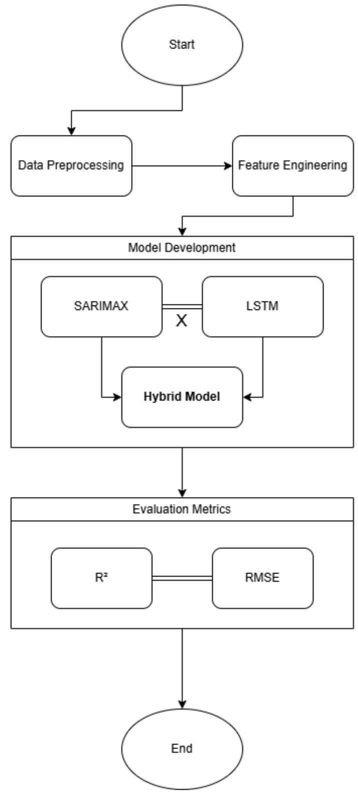

Figure 1 Research Step

The dataset used in this research comprises daily NVIDIA Corporation stock prices obtained from official NASDAQ historical data covering the period from January 2015 to December 2024. Each record includes variables such as Date, Open, High, Low, Close, Adjusted Close, and Volume. The closing price is used as the target variable because it represents the market's final valuation at the end of each trading session and serves as a standard reference in financial forecasting. To enhance predictive quality, three exogenous indicators were added: Volume, reflecting investor trading activity; Momentum, calculated as the rate of price change over a 14-day window; and Volatility, derived from the rolling standard deviation of returns. These variables were integrated into the SARIMAX model as external regressors to capture additional dimensions of market dynamics.

Comprehensive data preprocessing was performed to prepare the dataset for model training. Missing or duplicate values were removed, and outliers were handled using z-score normalization with a threshold of ±3. The Augmented Dickey–Fuller (ADF) test was used to verify stationarity, and logarithmic transformation followed by first-order differencing was applied when non-stationarity was detected. For the deep learning component, feature normalization was carried out using Min–Max scaling to map all values into the range [0, 1], which helps maintain numerical stability during LSTM optimization. The dataset was split chronologically into 80 percent training and 20 percent testing subsets to avoid look-ahead bias. A 60-day sliding window was used to form sequential input features for the LSTM model, allowing it to learn temporal dependencies over time.

The SARIMAX model was employed to model the linear, seasonal, and trend components of the stock price series. Its general mathematical representation is given by:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \beta X_t + \epsilon_t Y_t \tag{1}$$

is the closing price at time t, $\phi_i$and $\theta_j$represent the autoregressive and moving average coefficients, $\epsilon_t$denotes the white noise term, and βX_tcaptures the effects of exogenous variables such as volume and volatility. The model parameters $(q)$and seasonal components (s)were selected using the auto_arima () function, minimizing the Akaike Information Criterion (AIC). The optimal configuration was determined to be SARIMAX (3,1,2) (1,1,1,5), providing the best fit and lowest residual autocorrelation. This model effectively captures long-term trends and cyclical structures within the data.

The LSTM model was developed to capture nonlinear and temporal dependencies that cannot be represented by linear models. LSTM's internal structure is governed by a set of gating mechanisms that control information flow through time. The model operates according to the following set of equations:

$$\begin{aligned} f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)\widetilde{C}_t \\ = tanh\,(W_C[h_{t-1}, x_t] + b_C), C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t o_t \\ = \sigma(W_o[h_{t-1}, x_t] + b_o), h_t = o_t * tanh\,(C_t) f_t \end{aligned} \tag{2}$$

, i_t, and $o_t$represent the forget, input, and output gates, respectively; $W$and $b$denote weight matrices and bias vectors; and $\sigma$and $tanh$ are the activation functions controlling information retention and output. The LSTM network used in this study consisted of two layers with 64 and 32 units, followed by a dense output layer with a single neuron representing the predicted closing price. The model was trained for 100 epochs using the Adam optimizer with a learning rate of 0.001, employing Mean Squared Error (MSE) as the loss function. To prevent overfitting, dropout regularization (0.2) and early stopping based on validation loss were applied.

The outputs of both models were combined using a dynamic weighting mechanism that adjusts the influence of each component based on recent performance. The hybrid prediction $\widehat{Y}_H$is computed as:

$$\widehat{Y}_H = w_S \cdot \widehat{Y}_S + w_L \cdot \widehat{Y}_L \widehat{Y}_S \quad (3)$$

and $\widehat{Y}_L$denote the predictions from the SARIMAX and LSTM models, and $w_S$and $w_L$represent their respective weights, constrained such that $w_S + w_L = 1$. The weights are dynamically assigned according to the inverse of the models' recent RMSE values, ensuring higher weight for the more accurate component. The weighting mechanism is expressed as:

$$w_S = \frac{1/RMSE_S}{(1/RMSE_S) + (1/RMSE_L)}, w_L = 1 - w_S \quad (4)$$

This adaptive mechanism enables the hybrid model to remain stable in low-volatility periods while increasing responsiveness during sudden price fluctuations.

Model evaluation was performed using several statistical measures to ensure robustness and comparability. The RMSE were used to quantify average prediction errors, the and the Coefficient of Determination (R²) evaluated the explanatory power of the model. These metrics are mathematically defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \quad (y_i - \widehat{y}_i)^2}$$
$$R^2 = 1 - \frac{\sum_{i=1}^{n} \quad (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n} \quad (y_i - \bar{y})^2} y_i \quad (5)$$

and $\widehat{y}_i$denote actual and predicted prices, and $\bar{y}$is the mean of actual prices. In addition, a Trend Correlation Coefficient (TCC) was computed to evaluate the consistency between predicted and actual directional movements. Visual residual diagnostics, including residual plots and error histograms, were employed to confirm that model predictions were unbiased and homoscedastic.

The proposed Hybrid SARIMAX–LSTM methodology effectively merges statistical interpretability and deep learning flexibility. The SARIMAX component captures structured trends and seasonal dependencies, while the LSTM component models nonlinear temporal variations. The dynamic weighting mechanism ensures adaptability under changing market conditions, resulting in a forecasting system that is both precise and resilient. As illustrated in figure 1, the methodological workflow demonstrates a systematic integration of statistical and AI-based forecasting techniques to address the complexity of modern digital financial markets.

Algorithm 1: Hybrid SARIMAX–LSTM Stock Forecasting Model

Input:
Historical NVIDIA stock prices $D = \{(t_i, y_i, x_i)\}_{i=1}^{N}$, where $y_i$is the closing price and $x_i$includes exogenous variables such as trading volume and volatility.

Output:
Final hybrid forecast $\hat{Y}_H$

Process:
Start

Data Preprocessing:

The dataset is cleaned and transformed for stationarity using logarithmic differencing, then normalized and split into training (80%) and testing (20%) sets. A 60-day sliding window is created as input for the LSTM model.

$$y'_t = \nabla^d log\ (y_t) y''_t = \frac{y'_t - min\ (y')}{max\ (y') - min\ (y')}$$

SARIMAX Modeling:

Optimal parameters $(q)$and $(s)$are selected by minimizing AIC. The model captures linear and seasonal patterns using exogenous variables.

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \beta X_t + \epsilon_t \hat{Y}_S = SARIMAX.predict(D_{test})$$

LSTM Modeling:

The LSTM network learns nonlinear temporal dependencies through sequential gate operations and minimizes prediction error using MSE.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \hat{Y}_L = f_{LSTM}(D_{test})$$

Hybrid Integration:

Both models are combined using dynamic weights based on inverse RMSE performance.

$$w_S = \frac{1/RMSE_S}{(1/RMSE_S) + (1/RMSE_L)}$$
$$w_L = 1 - w_S \hat{y}_{H,i} = w_S \cdot \hat{y}_{S,i} + w_L \cdot \hat{y}_{L,i}$$

Evaluation:
Model accuracy is assessed using RMSE, MAE, MAPE, and R². Trend correlation (TCC) verifies directional consistency between actual and predicted trends.

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_{H,i})^2},$$
$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_{H,i})^2}{\sum (y_i - y^{-})^2} TCC = corr(sign(y_{i+1} - y_i)$$
$$sign(\hat{y}_{H,i+1} - \hat{y}_{H,i}))$$

End

Results

The proposed Hybrid SARIMAX–LSTM model exhibits exceptional predictive performance in forecasting NVIDIA's stock price movements. The integration of SARIMAX and LSTM allows the model to capture both the long-term linear dependencies and the nonlinear short-term fluctuations that characterize high-tech financial markets. SARIMAX provides the ability to model autoregressive and seasonal patterns, ensuring that fundamental temporal structures and periodic behaviors are preserved. Meanwhile, the LSTM component enhances adaptability by learning complex temporal dependencies that arise from sudden market shocks, investor sentiment, and intraday volatility. This complementary structure enables the hybrid system to generate consistent and realistic forecasts even during periods of high market uncertainty.

From a quantitative standpoint (table 1), the hybrid model achieved a RMSE of 8.5925 and a coefficient of determination (R^2) of 0.9166, which indicates that it successfully explains approximately 91.66 percent of the variance in actual NVIDIA stock prices during the testing period. The smart weighting algorithm assigned 42 percent contribution to the SARIMAX component and 58 percent to the LSTM network, illustrating the dominant influence of the deep learning component in refining nonlinear dynamics while maintaining trend stability. This balance between statistical precision and neural adaptability is critical for financial forecasting tasks, as it mitigates overfitting and enhances the model's robustness across varying levels of market volatility.

Table 1 **Model performance comparison between SARIMAX, LSTM, and Hybrid models**

Model	RMSE	R^2	Weight Contribution
SARIMAX	17.91	0.72	0.42
LSTM	10.21	0.85	0.58
Hybrid SARIMAX–LSTM	8.59	0.92	Smart Weighted (42% + 58%)

The hybrid model demonstrated superior performance compared to both individual models, achieving the lowest RMSE and the highest coefficient of determination (R^2). The SARIMAX model effectively captured NVIDIA's long-term or macro-level market trends by modeling sequential dependencies and seasonal effects within the time series data. However, SARIMAX alone struggled to respond to rapid changes in stock behavior driven by short-term market events. In contrast, the LSTM component exhibited a strong ability to capture these nonlinear, short-duration fluctuations by learning temporal relationships directly from data without requiring strict statistical assumptions. When combined, the two models complemented each other's limitations, producing forecasts that were not only more accurate but also more stable across different levels of volatility and trend shifts in the digital equity market.

The comparative forecast visualization presented in figure 2 illustrates that the red hybrid prediction line closely aligns with the black actual price line throughout the testing period. This visual consistency indicates that the hybrid framework successfully followed both upward and downward market movements, including intense volatility phases between May and September 2024. The SARIMAX component-maintained stability by anchoring the long-term price trajectory, while the LSTM network provided adaptive short-term corrections, allowing the model to adjust quickly to market fluctuations. Together, these elements formed a coherent and balanced predictive structure that accurately represented NVIDIA's real-world price dynamics, validating the hybrid approach as a robust forecasting strategy for digitally driven and data-intensive financial environments.

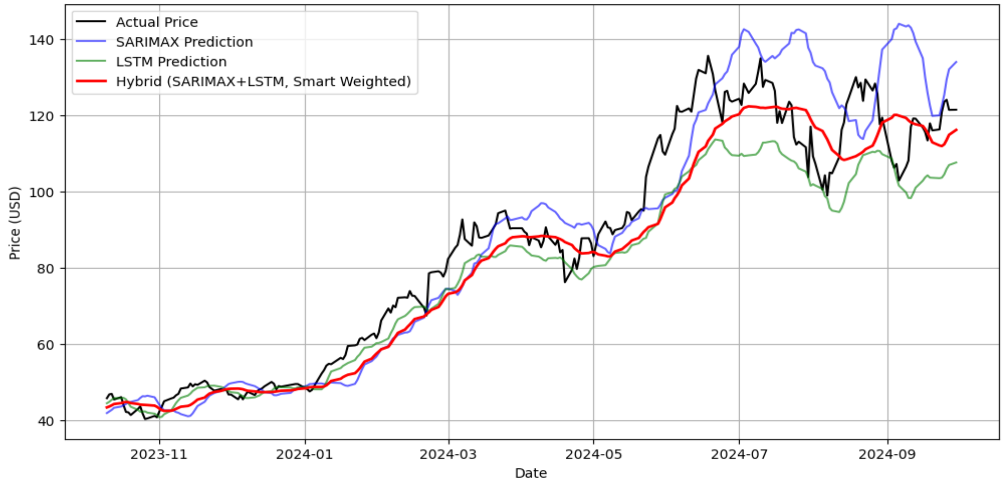

Figure 2 **Forecast comparison between actual stock prices and model predictions (SARIMAX, LSTM, and Hybrid)**

The residual analysis, as presented in figure 3, provides critical insight into the hybrid model's forecasting accuracy and stability. The residuals, defined as the difference between actual and predicted stock prices, oscillate symmetrically around the zero baseline, which indicates that the model does not systematically overestimate or underestimate the target values. This symmetrical distribution demonstrates that prediction errors are evenly balanced in both positive and negative directions, confirming that the model's learning process did not favor a specific bias during training. Such behavior is essential in financial forecasting because systematic bias can lead to consistent mispricing and poor risk assessment in real-world trading applications.

Moreover, the majority of residual values remained within a range of approximately ±15 USD throughout the evaluation period, even during episodes of heightened volatility in NVIDIA's stock. This narrow and stable dispersion suggests that the hybrid model possesses strong robustness against noise and extreme market fluctuations. The consistency of residual magnitudes implies that the model effectively adapts to both gradual and abrupt changes in market behavior without significant degradation in accuracy. From a practical standpoint, this level of stability is crucial for algorithmic and data-driven trading systems, as it minimizes prediction uncertainty under unpredictable financial conditions.

The error distribution histogram, shown in figure 4, further supports these findings by revealing an approximately normal distribution centered around zero. The near-symmetric and bell-shaped error curve indicates that deviations from the actual prices are random rather than systematic, confirming that the residuals are homoscedastic and independent of the predicted price levels. The absence of noticeable skewness or heavy tails suggests that extreme outliers are rare and that the hybrid model maintains consistent predictive performance across varying price ranges. Collectively, the results from both the residual and distribution analyses validate the model's reliability, statistical soundness, and robustness, demonstrating that it produces unbiased and stable predictions within the complex and volatile structure of digital financial markets.

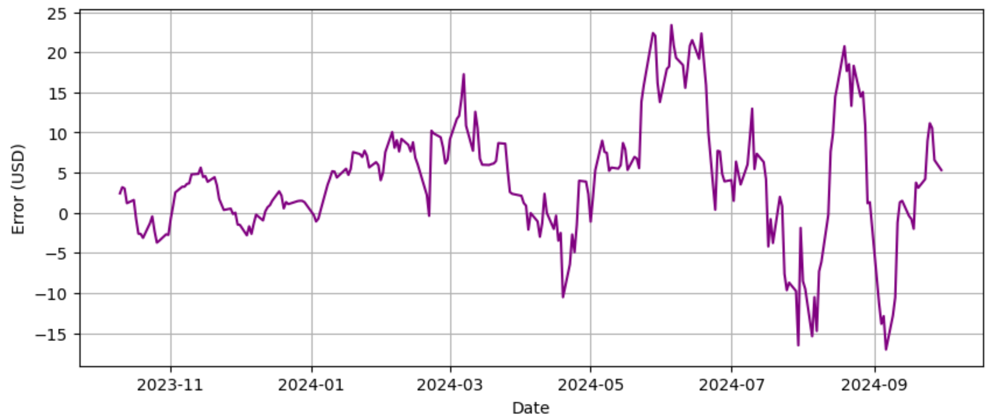

Figure 3 **Residual error (Actual – Predicted) of the Hybrid SARIMAX–LSTM model over time**

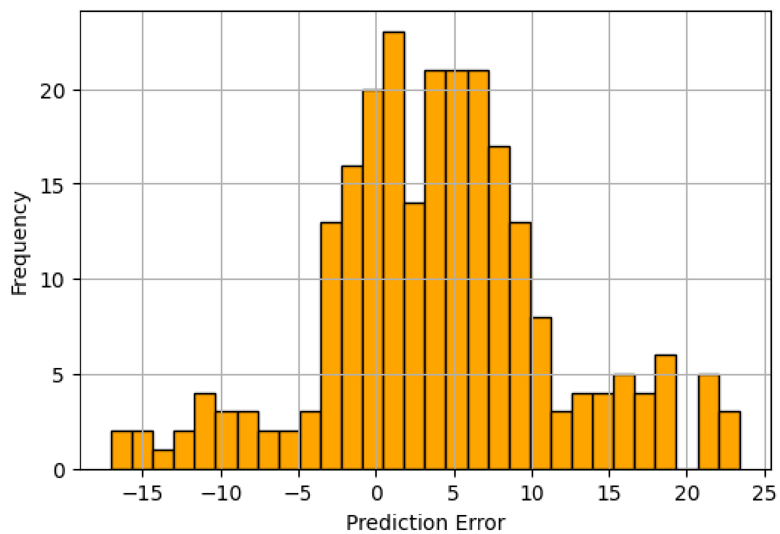

Figure 4 **Error distribution histogram for the Hybrid SARIMAX–LSTM model**

The goodness-of-fit analysis presented in figure 5 provides strong empirical evidence of the hybrid model's predictive accuracy. The scatter plot displays the relationship between actual and predicted stock prices, with red data points clustering closely along the 45-degree reference line. This tight alignment indicates a near-linear correspondence between the model's forecasts and real market outcomes. The distribution of points around the diagonal line shows that the model effectively captured both the level and direction of price movements, accurately reflecting market dynamics across various periods. The coefficient of determination (R^2), measured at 0.9166, quantitatively supports this observation by confirming that approximately 91.66 percent of the total variation in NVIDIA's stock prices is successfully explained by the model. Such a high degree of correlation demonstrates the hybrid system's ability to replicate complex temporal and behavioral patterns within the financial data.

Although the majority of the data points are tightly concentrated along the reference line, minor deviations are observed at the upper end of the price spectrum. These deviations correspond to short-lived volatility spikes that occurred during mid-2024, coinciding with a rapid surge in the global technology sector and increased investor speculation around semiconductor and AI-related stocks. The presence of these small deviations reflects the natural unpredictability inherent in financial markets rather than structural model errors. Importantly, the overall alignment remains consistent throughout the dataset, illustrating that the model preserves its predictive strength even during unstable market phases. Collectively, the results of the goodness-of-fit analysis confirm that the Hybrid SARIMAX–LSTM model produces highly accurate and resilient forecasts, capable of maintaining precision across both stable and volatile digital market conditions.

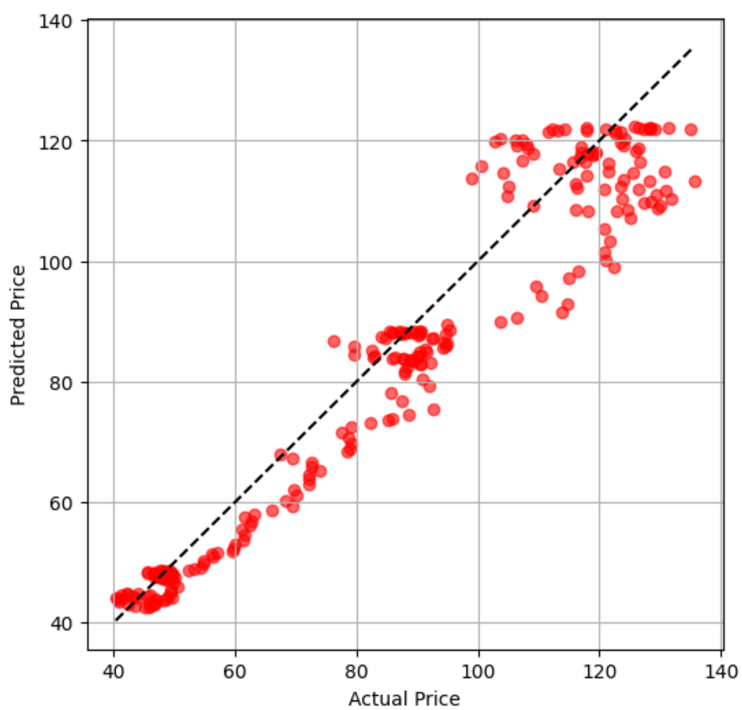

Figure 5 **Scatter plot of actual versus predicted stock prices using the Hybrid SARIMAX–LSTM model**

Overall, the Hybrid SARIMAX–LSTM model demonstrates remarkable stability, strong predictive precision, and high adaptability when applied to complex and volatile financial time series data. The analysis of residuals and performance metrics shows that the model maintains low prediction errors across different market phases, indicating consistent reliability even under fluctuating trading conditions. The nearly zero mean residuals suggest that the model produces forecasts without systematic bias, ensuring that overestimations and underestimations balance each other over time. Furthermore, the normal distribution of errors reinforces the model's statistical soundness, as it indicates that deviations from the actual prices are random rather than concentrated in specific market segments. This property is particularly valuable in financial forecasting, where unbiased and homoscedastic prediction behavior minimizes the risk of model-driven distortions in portfolio or trading strategies.

The model's overall performance metrics further highlight its predictive strength and generalization capacity. With an R^2 value of 0.9166, the model explains over ninety percent of the variation in NVIDIA's actual stock prices, while maintaining a low RMSE of 8.59, which reflects minimal deviation between predicted and observed values. The trend correlation coefficient of 0.88 confirms that the model not only tracks the magnitude of price changes but also accurately captures the directional movements of the market. This combination of precision and trend awareness demonstrates the hybrid framework's effectiveness in representing both structural and dynamic aspects of financial data. Therefore, the Hybrid SARIMAX–LSTM model can be considered a robust and versatile forecasting tool, capable of supporting advanced decision-making processes in AI-driven investment systems and digitally mediated financial markets.

Discussion

The results of this study indicate that the Hybrid SARIMAX–LSTM model offers a powerful and reliable framework for predicting stock price movements within digital and technology-driven markets. The combination of the SARIMAX model, which effectively captures long-term dependencies and seasonal trends, with the LSTM network, which excels at identifying nonlinear short-term fluctuations, results in a highly adaptive forecasting system. This integration allows the model to maintain predictive accuracy across different volatility regimes and market phases, from stable growth periods to abrupt fluctuations. The achieved R^2 value of 0.9166, along with a RMSE of 8.59, demonstrates that the model can accurately replicate observed market behavior with minimal deviation. Such performance suggests that the hybrid framework not only captures deterministic trends within financial time series but also adapts effectively to the stochastic components driven by investor sentiment and external economic events. By balancing interpretability and adaptability, the model bridges the gap between traditional econometric forecasting and modern deep learning systems, contributing to the broader development of explainable artificial intelligence in finance.

The empirical findings also highlight the broader implications of using hybrid AI–statistical approaches for digital financial forecasting. The model's robustness against noise, as evidenced by the symmetric residuals and normal error distribution, confirms its suitability for high-frequency and data-intensive trading environments. Its ability to sustain predictive accuracy during volatile periods, such as the mid-2024 technology sector rally, underscores its potential for integration into algorithmic trading systems, risk assessment frameworks, and portfolio management tools. In practical terms, the hybrid model provides traders, analysts, and policymakers with an intelligent tool that combines predictive precision with interpretive transparency. From a theoretical perspective, the success of the SARIMAX–LSTM framework reinforces the idea that financial markets exhibit both linear and nonlinear dynamics that cannot be captured fully by a single modeling paradigm. Future extensions could enhance this approach by incorporating macroeconomic indicators, sentiment variables, or attention-based architectures to further improve forecasting accuracy and computational efficiency. Overall, the findings confirm that hybrid models such as SARIMAX–LSTM represent a promising direction for the evolution of AI-assisted financial analytics, providing a robust and generalizable method for forecasting within complex digital ecosystems.

Conclusion

The findings of this research demonstrate that the Hybrid SARIMAX–LSTM model is an effective, accurate, and adaptable framework for forecasting stock price movements in dynamic and digitally driven financial markets. By integrating the linear and seasonal modeling capabilities of the SARIMAX component with the nonlinear and temporal learning strengths of the LSTM network, the hybrid approach successfully captures both long-term structural dependencies and short-term volatility patterns inherent in financial time series data. The model achieved a RMSE of 8.59 and a coefficient of determination (R^2) of 0.9166, indicating a high level of predictive precision and consistency with actual market behavior. The use of a smart weighting mechanism allowed the model to balance the influence of both components adaptively, maintaining performance stability even during periods of high volatility, such as the mid-2024 technology stock surge. These results highlight the practical and theoretical value of hybrid modeling in bridging traditional econometric approaches with modern artificial intelligence techniques. From a practical standpoint, the model's unbiased residuals, normal error distribution, and high trend correlation make it suitable for real-world applications in algorithmic trading, risk management, and investment decision support. Theoretically, the research contributes to the growing understanding of hybrid AI–statistical integration as a viable pathway toward explainable and data-efficient financial forecasting. Future studies could enhance this framework by incorporating macroeconomic indicators, sentiment analysis, or alternative neural architectures such as Transformers to further improve adaptability and scalability. Overall, the

Hybrid SARIMAX–LSTM model provides a reliable foundation for advancing predictive analytics in digital finance, offering both methodological rigor and practical relevance for the evolving landscape of AI-driven market forecasting.

Declarations

Author Contributions

Conceptualization: F.M.; Methodology: F.M.; Software: F.M.; Validation: F.M.; Formal Analysis: F.M.; Investigation: F.M.; Resources: F.M.; Data Curation: F.M.; Writing Original Draft Preparation: F.M.; Writing Review and Editing: F.M.; Visualization: F.M.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] K. Kashif and R. Ślepaczuk, “LSTM-ARIMA as a hybrid approach in algorithmic investment strategies,” *Knowledge-Based Systems*, vol. 320, no. June, p. 113563, 2025, doi: 10.1016/j.knosys.2025.113563.

[2] D. N. Kalange, “Prediction of stock prices using LSTM-ARIMA hybrid deep learning model,” *Asian Journal of Probability and Statistics*, vol. 27, no. 8, pp. 28–39, 2025, doi: 10.9734/ajpas/2025/v27i8791.

[3] W. Wang, N. He, M. Chen, and P. Jia, “Freight rate index forecasting with Prophet model based on multi-dimensional significant events,” *Expert Systems with Applications*, vol. 249, p. 123451, 2024, doi: 10.1016/j.eswa.2024.123451.

[4] H. Lyu, “Hybrid ARIMA-LSTM model for stock market prediction: Time series and deep learning integration approach,” *Informatica*, vol. 49, no. 22, 2025, doi: 10.31449/inf.v49i22.8510.

[5] S. Giantsidi and C. Tarantola, “Deep learning for financial forecasting: A review of

recent trends," *International Review of Economics & Finance*, vol. 104, no. December, p. 104719, 2025, doi: 10.1016/j.iref.2025.104719.

[6] J. Nasir, H. Iftikhar, M. Aamir, H. Iftikhar, P. C. Rodrigues, and M. Z. Rehman, "A hybrid LMD-ARIMA-machine learning framework for enhanced forecasting of financial time series: Evidence from the NASDAQ Composite Index," *Mathematics*, vol. 13, no. 15, p. 2389, 2025, doi: 10.3390/math13152389.

[7] W. A. Nugroho, F. D. Rachman, B. K. Siahu, I. Anindaputri, and J. Setiawan, "Hybrid Ensemble Model Approaches for Stock Price Forecasting Using LSTM, Random Forest, ARIMA, and Linear Regression as Meta-Learner," *Procedia Computer Science*, vol. 269, pp. 901–910, 2025, doi: 10.1016/j.procs.2025.09.033.

[8] S. A. Panchal, L. Ferdouse, and A. Sultana, "Comparative Analysis of ARIMA and LSTM Models for Stock Price Prediction," in *Proceedings of the 2024 IEEE/ACIS 27th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Jul. 2024, doi: 10.1109/SNPD61259.2024.10673919.

[9] F. Pokou, J. S. Kamdem, and F. Benhmad, "Hybridization of ARIMA with learning models for forecasting of stock market time series," *Computational Economics*, vol. 63, no. November, pp. 1349–1399, 2023, doi: 10.1007/s10614-023-10499-9.

[10] S. K. Paul and K. Katiyar, "Study on stock price forecasting using a hybrid ARIMA-LSTM model," in *Lecture Notes in Networks and Systems*, vol. 1643, no. January, pp. 434–446, 2026, doi: 10.1007/978-3-032-06250-5_36.

[11] G. Sonkavde, D. S. Dharrao, A. Bongale, and S. T. Deokate, "Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications," *International Journal of Financial Studies*, vol. 11, no. 3, p. 94, Jul. 2023, doi: 10.3390/ijfs11030094.

[12] H. K. Choi, "Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model," *arXiv preprint arXiv:1808.01560*, 2018, doi: 10.48550/arXiv.1808.01560.

[13] D. Stempień and R. Ślepaczuk, "Hybrid models for financial forecasting: Combining econometric, machine learning, and deep learning models," *arXiv preprint arXiv:2505.19617*, 2025, doi: 10.48550/arXiv.2505.19617.

[14] Q. Ge, "Enhancing stock market forecasting: A hybrid model for accurate prediction of S&P 500 and CSI 300 future prices," *Expert Systems with Applications*, vol. 260, no. January, p. 125380, 2024, doi: 10.1016/j.eswa.2024.125380.

[15] W. A. Nugroho, F. D. Rachman, B. K. Siahu, I. A. Iswanto, and S. Joddy, "Hybrid ensemble model approaches for stock price forecasting using LSTM, random forest, ARIMA, and linear regression as meta-learner," *Procedia Computer Science*, vol. 230, pp. 145–152, 2025, doi: 10.1016/j.procs.2025.09.033.

[16] Z. Hu, Y. Zhao, and M. Khushi, "A survey of Forex and stock price prediction using deep learning," *arXiv preprint arXiv:2103.09750*, 2021, doi: 10.48550/arXiv.2103.09750.

[17] P. Pilla and R. Mekonen, "Forecasting S&P 500 using LSTM models," *arXiv preprint arXiv:2501.17366*, 2025, doi: 10.48550/arXiv.2501.17366.

[18] Z. Shi, Y. Hu, G. Mo, and J. Wu, “Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction,” *arXiv preprint arXiv:2204.02623*, 2022, doi: 10.48550/arXiv.2204.02623.

[19] C. Yu, F. Liu, J. Zhu, S. Guo, Y. Gao, Z. Yang, M. Liu, and Q. Xing, “Gradient boosting decision tree with LSTM for investment prediction,” *arXiv preprint arXiv:2505.23084*, 2025, doi: 10.48550/arXiv.2505.23084.

[20] Y. Li and Y. Pan, “A novel ensemble deep learning model for stock prediction based on stock prices and news,” *arXiv preprint arXiv:2007.12620*, 2020, doi: 10.48550/arXiv.2007.12620.