Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00056-4

Original Article

Residual-Based Hybrid Models for Forecasting Stock Prices of Nigerian Banks

Oluwatayofunmi Durodola [1], Aderonke Adegbenjo [1], Ayoade Iyiade [1], Oluwaseun Ilori [1]

1. Computer Science, Babcock University, Ilisan-Remo, NGA

Received: February 27, 2026 | Review began: March 05, 2026 | Review ended: March 24, 2026 | Published: April 10, 2026

Abstract

Introduction

The high volatility and non-linear dynamics of emerging financial markets, such as the Nigerian Stock Exchange (NSE), pose significant challenges to traditional linear forecasting models. This study developed and evaluated a residual-based hybrid forecasting framework designed to predict short-term stock prices for a stratified sample of three Nigerian banks. The primary objective was to enhance predictive accuracy by integrating the linear efficiency of Auto-Regressive Integrated Moving Average (ARIMA) with the non-linear learning capabilities of Support Vector Regression (SVR) and Long Short-Term Memory (LSTM) networks.

Methods

The study utilized daily stock data and key macroeconomic indicators spanning a 10-year period (April 2015-April 2025). A Walk-Forward Validation protocol was implemented to simulate real-world trading conditions, strictly preventing look-ahead bias. The framework operated on an additive decomposition principle, where ARIMA modeled the linear trend and the machine learning algorithms modeled the complex residuals. Predictive performance was assessed using Mean Absolute Error and Root Mean Square Error. Furthermore, the study applied SHapley Additive exPlanations (SHAP) to deconstruct the "black box" models and identify the primary drivers of volatility.

Results

Empirical results demonstrated that the ARIMA-LSTM hybrid consistently outperformed both the standalone ARIMA and the ARIMA-SVR models across all bank tiers. For the Tier 1 institution, the ARIMA-LSTM model achieved a Mean Absolute Error of 2.35, effectively capturing market trends despite high trading volumes. SHAP analysis revealed a distinct hierarchy of influence: large-cap stocks were predominantly driven by global macroeconomic factors, specifically the USD/NGN exchange rate and crude oil prices, while mid-cap stocks showed higher sensitivity to domestic fixed-income signals, such as bond yields and technical momentum indicators.

Conclusions

The study concludes that hybrid deep learning architectures significantly reduce forecasting errors in emerging markets and provides a transparent mechanism for investors to identify tier-specific economic triggers.

Categories: Explainable AI, Machine Learning (ML), Deep Learning

Keywords: hybrid forecasting, nigeria banks, shap, stock price prediction, deep learning, machine learning, arima, explainable ai, lstm, svr

Introduction

Investing in the stock market is a powerful strategy for building long-term wealth, offering higher returns than traditional savings when diversified to manage risks [1]. In Nigeria, the performance of the stock market serves as an important barometer of economic health, showing investor sentiment and systemic liquidity [2]. The Nigerian Stock Exchange (NSE), one of Africa's largest capital markets [3], reached a market capitalization of ₦112.58 trillion in January 2025, representing a substantial increase from ₦102.35 trillion in December 2024 [4]. With this impressive market size, retail investor participation has gradually moved from 28 percent in 2022 to 35 percent as of September 2023 (calculation from Nigerian Exchange (NGX) Domestic & Foreign Portfolio Investment monthly reports) [5], with recent acceleration through fintech adoption, where retail trading volumes surged to ₦516.50 billion in July 2025, an 88.07 percent month-on-month increase from June 2025 [6].

Stock forecasting represents a critical analytical tool for enhancing investment decision-making at both individual and institutional levels, with particular significance for emerging markets like Nigeria. Accurate stock price predictions enable investors to allocate capital more effectively, thereby facilitating wealth creation and portfolio optimization [7]. At the macroeconomic level, stock market forecasting serves as a leading indicator of economic performance; rising stock prices signal investor confidence and stimulate corporate investment, while accurate forecasts help anticipate and mitigate risks associated with adverse price movements [8]. In the Nigerian context specifically, stock forecasting becomes increasingly vital given the NSE critical role in capital formation and economic development [9,10]. The NSE plays a foundational intermediation function, channeling savings from surplus economic units to deficit units, yet the market has historically underperformed relative to its potential and comparable African capital markets [10]. Robust forecasting models enable Nigerian policymakers, institutional investors, and individual stakeholders to make better decisions relating to capital allocation, further strengthening the linkage between market performance and real economic growth [11]. Furthermore, given the high volatility characteristic of Nigerian stock returns and the behavioral finance factors that influence investor decision-making in the NSE [11], precise forecasting models using advanced methods like machine learning algorithms and artificial neural networks provide essential decision-support mechanisms that can enhance market efficiency and reduce information asymmetry [8,11]. Stock forecasting, by improving predictive accuracy through quantitative methodologies, ultimately contributes to deepening Nigeria's capital markets, facilitating sustainable capital formation, and supporting the industrial production necessary for long-term economic development [10].

Popular forecasting methods, like Seasonal ARIMA, Auto-Regressive Integrated Moving Average (ARIMA), and Generalized Autoregressive Conditional Heteroskedasticity, have traditionally been widely used to predict stock prices. However, these linear statistical models face significant limitations when applied to volatile markets like Nigeria's. Financial time series data in such emerging economies are characterized by non-linearity, high variability, and non-stationarity, often exhibiting structural breaks caused by sudden policy shifts or external shocks. Consequently, classical models struggle to incorporate these dynamic irregularities and the influence of exogenous macroeconomic indicators, leading to suboptimal predictive performance in unstable environments [12,13]. For instance, studies on the NSE and other markets, such as the New York Stock Exchange and Tesla stock, show that ARIMA is effective for short-term trends but struggles with complex, non-linear behaviors [8,14,15]. Other analytical approaches present additional methodological limitations: fundamental analysis, which assesses company valuations and macroeconomic fundamentals, proves more effective for long-term investment horizons, and exhibits diminished utility for short-term tactical forecasting, while technical analysis relying solely on historical prices and trading volumes cannot adequately address the chaotic and non-linear nature of emerging market dynamics [3,12,13].

Recent developments in artificial intelligence, machine learning, and deep learning offer potential solutions to these challenges. Models like Artificial Neural Networks and Long Short-Term Memory (LSTM) networks, a type of recurrent neural network, excel at identifying non-linear patterns and long-term dependencies in time series data [6,9]. Research on the NSE, Egyptian markets, S&P 500, and Forex markets demonstrates LSTM's superior ability to handle complex financial data, with variants like Bi-LSTM and stacked LSTM showing even greater accuracy [16-18]. Ensemble learning, combining models like Random Forest, XGBoost, and LSTM, has also improved forecasting reliability across markets, including India's NSE and multiple global exchanges [19,20]. Some studies have enhanced predictions by incorporating sentiment analysis or online data, though macroeconomic factors like inflation and exchange rates are rarely integrated [19,20].

The financial services sector in Nigeria, which encapsulates both the money and capital markets, presents an ideal domain for predictive analysis. Within this broad sector, listed banking equities are particularly suitable for this study due to their high trading volumes and systemic economic significance. Notably, as of late December 2025, the NGX Banking Index had recorded a strong year-to-date gain, while the NGX All-Share Index returned over 50% in 2025, signaling renewed investor confidence in Nigerian equities [21]. However, prior studies often focus solely on technical indicators, overlooking macroeconomic variables like oil prices and exchange rates, which heavily influence Nigeria's market [22]. Nigeria's status as a mono-product, oil-dependent economy creates a unique environment where crude oil price fluctuations and exchange rate movements directly amplify stock market volatility. Empirical research confirms that oil prices impact the market through multiple transmission channels. Research has shown that oil price volatility constitutes a critical barrier to economic growth due to its destabilizing effects on macroeconomic variables, including GDP, interest rates, and exchange rates [23]. Furthermore, one study examined the nexus among exchange rate, oil price, and stock market performance in Nigeria using vector auto-regression techniques, revealing a unidirectional relationship from crude oil prices to stock shares and bidirectional relationships between crude oil prices and exchange rates [24].

Recent deep learning research specifically focused on African stock markets has demonstrated the effectiveness of these approaches in emerging market contexts. For example, one assessment of the forecasting potential of four deep learning models (simple LSTM, gated recurrent unit, weighted LSTM, and LSTM with attention) for stock prices in five African countries including Nigeria found that the gated recurrent unit model produced the lowest error rates and highest accuracy in Nigeria, with Mean Squared Error (MSE) of 0.90, Mean Absolute Error (MAE) of 0.89, Root Mean Square Error (RMSE) of 0.92, and R^2 of 0.93 [25]. Nevertheless, existing predictive studies demonstrate a critical deficiency: they predominantly focus on technical indicators in isolation, systematically overlooking macroeconomic variables including crude oil prices, exchange rates, and inflation, all of which heavily influence Nigerian financial market dynamics given the nation's oil-dependent economy. Furthermore, while a 2025 hybrid autoregressive-LSTM study on Airtel's stock demonstrated that the hybrid AR-LSTM model outperformed the standalone ARIMA model across all reported metrics, achieving 0.79 accuracy compared to 0.72 for ARIMA [26], such research has not been systematically applied to Nigerian banking stocks with macroeconomic integration. Combining statistical models like ARIMA with advanced techniques such as LSTM and Support Vector Regression (SVR) can yield more robust predictions, especially when including macroeconomic data [15,22].

Despite global research demonstrating the individual strengths of ARIMA, SVR, and LSTM models for financial forecasting, few studies in Nigeria have systematically evaluated these three models as hybrid approaches in comparative analysis while incorporating macroeconomic variables with technical indicators, particularly within the banking sector context where such comparative insights can deliver immediate practical value to retail and institutional investors. This research gap is particularly acute given Nigeria's unique macroeconomic structure, characterized by oil dependency, exchange rate volatility, and structural market rigidities that differentiate the NSE from developed market contexts. This study addresses this critical gap by strategically applying ARIMA, LSTM, and SVR hybrid models to forecast Nigerian banking stock prices.

Materials And Methods

Design of hybrid forecasting models

The first step was to design robust hybrid forecasting models capable of processing the unique volatility of the Nigerian financial market. This design was based on an additive decomposition principle, where a time series Y_t is viewed as a combination of a linear component L_t and a non-linear residual component N_t given as:

$$Y_t = L_t + N_t$$

In this architecture, the linear component was modeled using the econometric precision of ARIMA, while the non-linear residuals, representing the variance that linear models fail to capture, were assigned to SVR and LSTM models. This layered design allowed each algorithm to specialize in the patterns it is mathematically best suited to identify.

Figure 1 presents an overview of the system architecture for the study.

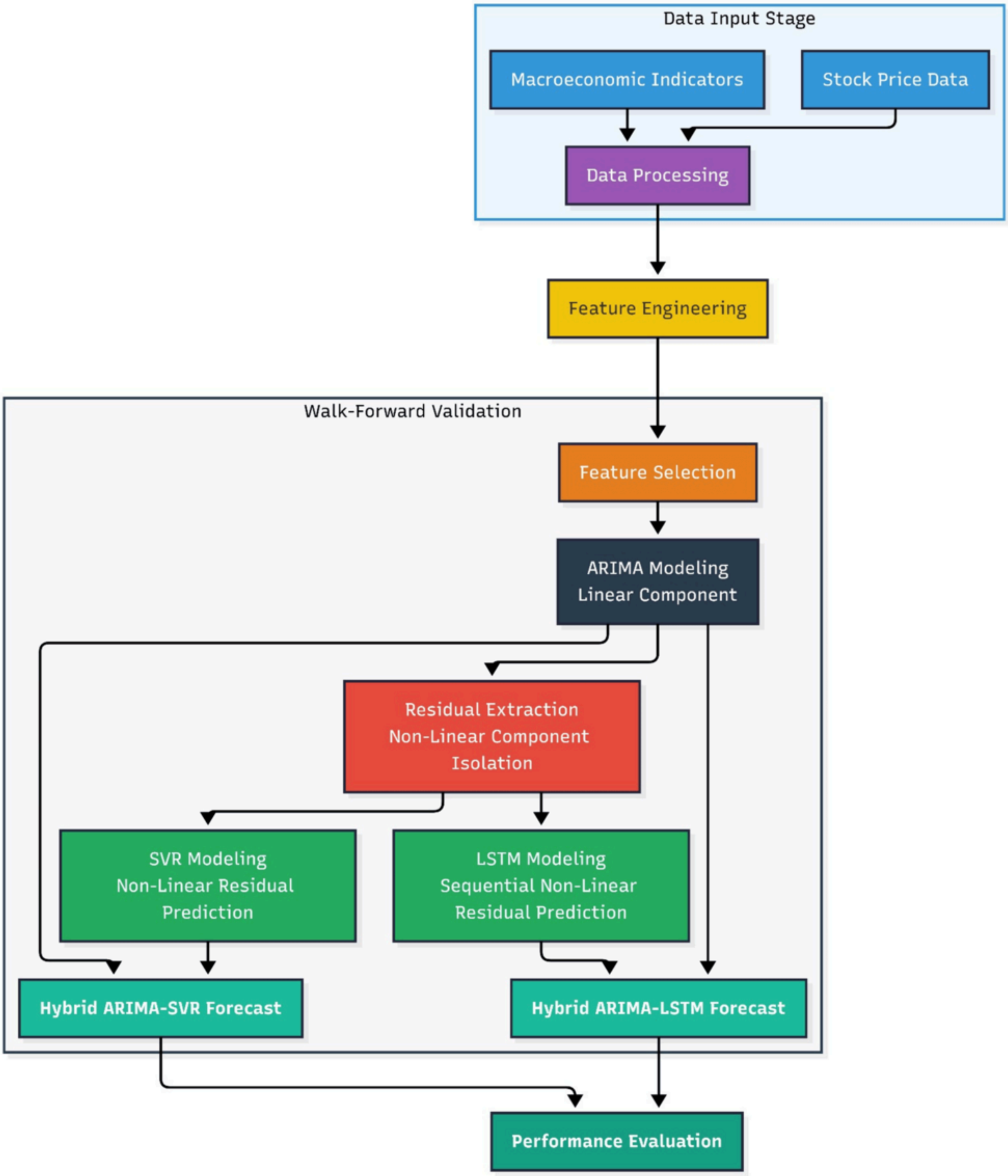

FIGURE 1: System Architecture

ARIMA, Auto-Regressive Integrated Moving Average; LSTM, Long Short-Term Memory; SVR, Support Vector Regression

ARIMA Model Design

The ARIMA model was designed as a classical time series tool to handle the linear trends and autocorrelations in univariate stock data. The architecture followed the ARIMA p,d,q structure, where p represents autoregressive terms, d represents the degree of differencing required to achieve stationarity, and q represents moving average terms. The design utilized the Augmented Dickey-Fuller test as a gateway to determine the differencing order, followed by an automated search using the Akaike Information Criterion (AIC) to establish the optimal p,q configuration.

SVR Design

To address the non-linearities in the extracted residuals, the SVR was designed using a Radial Basis Function (RBF) kernel. This kernel choice was critical as it allowed the model to map input indicators into a higher-dimensional space where complex relationships between residuals and macroeconomic variables could be identified. The architectural parameters were designed to balance flexibility and stability, utilizing a regularization constant C of 100, a kernel coefficient γ of 0.1, and an epsilon-insensitive tube ε of 0.1.

LSTM Design

The LSTM was designed as a deep recurrent neural network specifically suited for capturing long-term temporal dependencies. The network architecture consisted of four primary layers:

1. Input Layer: Configured to accept a 10-day lookback window, allowing the model to "remember" the immediate historical context of market volatility.

2. Hidden LSTM Layers: Two stacked LSTM layers with 50 units each were designed to extract high-level temporal features from the residuals.

3. Dropout Layer: A dropout rate of 0.2 was integrated after each LSTM layer to prevent overfitting by randomly muting neurons during training.

4. Dense Output Layer: A single-unit fully connected layer was designed to produce the final continuous prediction for the residual value.

Data sampling and acquisition

This study employs a stratified sampling approach to select three banks from Nigeria's banking sector. Stratified sampling divides a large population into smaller, similar groups called strata, then selects samples from each group [27]. This method is better than random sampling because it ensures that each important subgroup is represented in the research [28]. In Nigeria's banking system, there are different types of banks based on size and importance. Bank 1 represents the Tier-1 group, which includes the largest and most important banks.

According to [29], Zenith Bank is Nigeria's number one bank by Tier-1 capital, holding this position for 16 years in a row. It has ₦2.59 trillion in market value and trading volumes of over 18 million shares per day [29]. Fidelity Bank and Wema Bank represent the Tier-2 group, which are medium-sized banks. As of early March 2026, Wema Bank has shown strong YTD performance on the NGX, with the share price rising to NGN 27.90, marking a 36.8% gain since the start of the year [30]. Using these three banks gives a clear picture of the Nigerian banking market across different sizes. This approach is widely used in banking research because these banks have enough historical data, high trading activity, and good data quality needed for time series models like ARIMA, LSTM, and SVR. This sampling design allows this study to capture both systemic importance and diversity while keeping the research practical and focused.

Data acquisition (as seen in Table *1*) prioritized the retrieval of reliable, high-frequency daily data for the selected banks, alongside critical macroeconomic indicators, to facilitate accurate stock price forecasting. Daily stock market data, specifically the closing price (target variable), opening price, high, low, and trading volume, was sourced from Investing.com [31], a platform recognized for providing comprehensive charts and technical metrics for the NSE. Macroeconomic data includes inflation, monetary policy rate, treasury bill yield, prime lending rate, exchange rate (NGN/USD), Brent crude oil price, and Nigerian 10-year bond yield, sourced from the Central Bank of Nigeria [32].

Variable(s)	Description/Role	Source	Frequency
Closing Price (Target variable)	Daily closing price of the selected bank's stock (primary dependent variable for forecasting)	Investing.com	Daily
Opening Price	Price at market open	Investing.com	Daily
High Price	Highest price during the trading day	Investing.com	Daily
Low Price	Lowest price during the trading day	Investing.com	Daily
Trading Volume	Number of shares traded during the day	Investing.com	Daily
Inflation (general and food sub-indices)	Consumer Price Index-based inflation rates	Nigeria Bureau of Statistics	Monthly
Monetary Policy Rate	Central bank's benchmark interest rate	Central Bank of Nigeria	Monthly
Treasury Bill Yield	Yield on short-term government securities (e.g., 91-day, 182-day)	Central Bank of Nigeria	Monthly
Prime Lending Rate	Average rate at which commercial banks lend to prime customers	Central Bank of Nigeria	Monthly
Exchange Rate (NGN/USD)	Official/interbank Naira to US Dollar rate	Investing.com	Monthly
Brent Crude Oil Price	International benchmark crude oil price (USD per barrel)	Investing.com	Monthly
Nigerian 10-Year Bond Yield	Yield on long-term government bonds	Central Bank of Nigeria	Monthly

TABLE 1: Data Description

These raw variables were utilized to compute derived technical indicators, specifically: Simple Moving Averages (SMA 5, 10, 20), Exponential Moving Averages (EMA 5, 10, 20), the Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD). These indicators serve distinct predictive functions: SMAs smooth out short-term price noise to reveal underlying trends; the RSI identifies potential overbought or oversold market conditions; and the MACD signals shifts in price momentum, all of which are critical for capturing the specific volatility patterns of Nigerian banking stocks.

Preprocessing and feature engineering

Data preprocessing was implemented to ensure consistency and compatibility with the ARIMA, SVR, and LSTM models. First, all datasets were aligned to a daily frequency spanning from April 9, 2015, to April 9, 2025 (3,635 observations). Stock data was cleaned by removing rows with missing values to ensure data integrity, rather than imputing them, resulting in a robust dataset of valid trading days. Macroeconomic data, originally monthly, was integrated with the daily stock data using an inner join operation on the date index. This process effectively propagated monthly values (e.g., inflation, monetary policy rate) across daily records, resulting in a step-wise alignment where macroeconomic indicators remained constant throughout the month while stock prices fluctuated daily.

Feature engineering was conducted to derive technical indicators, including SMA, EMA, RSI, and MACD. Additionally, temporal dependencies were captured by creating lagged versions t_1, t_2, t_3 of the closing price and macroeconomic variables. To address the sensitivity of the SVR and LSTM models to feature magnitude, all numerical features were standardized using StandardScaler (Z-score normalization) from Python’s sklearn.preprocessing, which centers the data around a mean of 0 with a standard deviation of 1. Three separate datasets were created, one for each bank, merging bank-specific technical indicators with shared macroeconomic features using pandas.merge. Data quality was validated by confirming the absence of NaN values post-processing. The preprocessing steps were executed using Python 3.8+ in a Jupyter Notebook environment.

Evaluation

Statistical Accuracy Metrics

Model performance was evaluated using multiple statistical accuracy metrics. The primary evaluation metric was RMSE, which is given as $\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}}$, where:

$\sum$ represents the sum of all values;

P_i is the predicted value for the i-th observation;

O_i is the observed value for the i-th observation;

and *n* is the total number of observations.

RMSE is chosen due to its sensitivity to large errors, making it a suitable measure for financial forecasting where large deviations are costly [22]. Other evaluation metric includes MAE, which calculates the average difference between the calculated values and actual values and also known as scale-dependent accuracy as it calculates error in observations taken on the same scale used to predict the accuracy of the machine learning model. Furthermore, Mean Absolute Percentage Error and the R-squared (R^2) coefficient were utilized to evaluate the goodness-of-fit and the proportion of variance explained by the hybrid models across different price scales of the three stratified banks.

Walk-Forward Validation Protocol

A central component of the evaluation strategy was the implementation of a Walk-Forward Validation protocol, which simulated real-world investment scenarios more accurately than traditional k-fold cross-validation. The protocol utilized an initial training window of 80% of the data (approximately 2,908 observations), followed by a rolling test window of 30 days. In each iteration, the models were retrained on an expanding window of historical data to forecast the subsequent 30-day period. This iterative process effectively evaluated the "system efficiency" by ensuring that the models remained updated with the most recent market information while strictly preventing look-ahead bias, as no future data was ever visible to the models during the training phase.

Model Interpretability and Predictor Identification (SHAP)

To address the "black box" nature of the SVR and LSTM components, the evaluation phase integrated SHAP (SHapley Additive exPlanations). This method was employed to quantify the "utility" of individual features by calculating their marginal contributions to the final forecast. The implementation involved initializing a shap.KernelExplainer using a representative background sample of 100 observations from the standardized training set to establish a baseline expectation.

SHAP values were then computed for the test set observations to identify which technical and macroeconomic features most significantly influenced the model's residual predictions. This allowed the research to move from simple forecasting to "explainable AI," identifying whether momentum indicators like RSI or external shocks like NGN/USD exchange rate fluctuations were the primary drivers of stock price movements across the different banking tiers. By visualizing these contributions through SHAP summary plots, the study successfully identified the top predictors, ensuring the hybrid system provided actionable insights into the underlying mechanics of the Nigerian financial market.

Results And Discussion

Experimental setup

To ensure the reproducibility and rigorous validation of the hybrid forecasting framework, all computational experiments were conducted using a Python 3.10+ runtime environment hosted on Google Colab Pro, utilizing an NVIDIA T4 GPU for accelerated deep learning training. The core implementation relied on the following open-source libraries: Pandas (v2.0) and NumPy for data manipulation; Matplotlib and Seaborn for visualization; Pmdarima (v2.0.4) for automated ARIMA order selection; Scikit-learn (v1.3) for SVR implementation, feature scaling, and error metrics; TensorFlow/Keras (v2.15) for constructing and training the LSTM networks; and SHAP (v0.44) for model interpretability.

The dataset for each bank (Zenith, Fidelity, Wema) spanned a 10-year period from April 9, 2015, to April 9, 2025, resulting in approximately 3,635 daily observations per institution after cleaning missing values. A strict chronological split was applied to prevent data leakage: the first 80% of the data (approximately 2,908 observations covering 2015-2023) was utilized for training, while the remaining 20% (covering 2023-2025) was reserved for testing. To maintain a uniform feature scale essential for the convergence of the SVR and LSTM algorithms, all input variables were normalized using StandardScaler (Z-score normalization) based strictly on training set statistics.

Model configurations were standardized as follows:

ARIMA: Optimal p,d,q parameters were selected automatically using *pmdarima.auto_arima* based on the minimization of the AIC, utilizing a stepwise search algorithm.

SVR: The SVR model employed an RBF kernel with fixed hyperparameters derived from preliminary grid search: regularization parameter $C = 100$, kernel coefficient $\gamma = 0.1$, and epsilon $\epsilon = 0.1$.

LSTM: The recurrent neural network architecture consisted of two stacked LSTM layers (50 units each) with a tanh activation function, followed by a dropout layer (rate=0.2) to prevent overfitting. The model was compiled using the Adam optimizer (learning rate=0.001) and an MSE loss function. Training was executed with a batch size of 32 for a maximum of 100 epochs, utilizing an early stopping callback (patience = 10) to halt training when validation loss plateaued.

Validation: To simulate real-world forecasting, a Walk-Forward Validation protocol was implemented. This involved iteratively retraining the models on a rolling basis to forecast a 30-day horizon, ensuring the models were continuously exposed to the most recent market regime.

It is important to note that the hyperparameters for the SVR and LSTM models were held constant across all walk-forward iterations. This approach was deliberately chosen to establish a stable performance baseline and manage the substantial computational expense of training deep learning networks across a decade of daily data with expanding windows.

ARIMA

A critical element of the model design was the identification of the optimal linear baseline for each bank tier. Through an automated iterative search using the AIC, the study established the "linear blueprint" p,d,q for each institution, as seen in Table 2. These parameters are essential design choices, as they define the extent to which the model relies on historical prices (auto-regression) versus historical errors (moving averages).

Bank	Optimal Order (p,d,q)	AIC Values
Bank A	(0,1,1)	-1245.31
Bank B	(0,1,2)	-8625.31
Bank C	(2,1,4)	-9525.67

TABLE 2: Optimal ARIMA Parameter Specifications Across Banks

AIC, Akaike Information Criterion; ARIMA, Auto-Regressive Integrated Moving Average

It can be derived that the divergent ARIMA specifications in Table 2 reflect the heterogeneous "linear memory" of the different bank tiers. For Zenith Bank, the optimal design of (0, 1, 1) suggests a pure moving average process where the current price is a function of the immediate preceding shock, reflecting a high-liquidity environment where information is absorbed rapidly. Fidelity Bank followed a similar (0, 1, 2) structure, implying that market shocks in the mid-cap tier linger slightly longer, influencing the price for up to 48 hours. In contrast, the design for Wema Bank was significantly more complex at (2, 1, 4). This indicates a "long-memory" process where the linear price trajectory is influenced by both the past two days of trading (AR2) and a four-day window of market shocks (MA4), characteristic of the higher volatility and thinner trading volumes.

The implementation results demonstrate a significant performance gap between ARIMA, and the two hybrid architectures. Firstly, the trained ARIMA models generated one-step-ahead forecasts on the 727-observation holdout test set. Prediction accuracy metrics shown in Table 3 reveal substantial differences across institutions.

Bank	Mean Absolute Error	Root Mean Squared Error
Zenith Bank	12.63	14.06
Fidelity Bank	6.18	7.32
Wema Bank	3.14	3.87

TABLE 3: ARIMA prediction results

ARIMA, Auto-Regressive Integrated Moving Average

As seen above, Wema Bank achieved the lowest ARIMA prediction error (MAE = 3.14), suggesting that the ARIMA (2, 1, 4) specification effectively captured the price dynamics of this institution's stock. The higher error magnitude for Zenith Bank (MAE = 12.63) indicates that pure moving average specifications may inadequately capture the nonlinear relationships present in this stock's returns. Notably, RMSE values exceed MAE across all three banks, revealing that the models occasionally generate substantial prediction errors that disproportionately influence squared-error metrics. This pattern suggests that while ARIMA performs reasonably for near-mean forecasts, it struggles with unusual market movements or structural shifts not captured by the linear autoregressive framework.

Standalone SVR

Because ARIMA captures only linear temporal dependencies, the analysis employed SVR with an RBF kernel to predict ARIMA residuals, the non-linear component that ARIMA fails to explain. The results are presented in Table *4*.

Bank	MAE (Residuals)	RMSE (Residuals)
Zenith Bank	12.87	14.28
Fidelity Bank	6.14	7.28
Wema Bank	3.11	3.85

TABLE 4: SVR Residual Prediction Result

MAE, Mean Absolute Error; RMSE, Root Mean Squared Error; SVR, Support Vector Regression

The SVR model's performance closely parallels the ARIMA baseline across all institutions, with near-identical MAE and RMSE values. This convergence suggests that the residual component exhibits limited exploitable nonlinear structure that SVR can effectively capture using the selected features. The failure of SVR to substantially reduce ARIMA residuals implies either (1) that ARIMA's residuals are dominated by true stochastic noise rather than systematic nonlinear patterns, or (2) that the feature set, while comprehensive, lacks the specific leading indicators necessary to predict the nonlinear component of price movements.

Standalone LSTM

LSTM networks were deployed to capture nonlinear temporal dependencies that both ARIMA and SVR overlook. The architecture comprised two stacked LSTM layers (50 units each) with dropout regularization (0.2) and a single-unit dense output layer. A 10-timestep lookback window was employed, enabling the model to learn patterns from 10 preceding trading days. The results are presented in Table *5*.

Bank	MAE (Residuals)	RMSE (Residuals)
Zenith Bank	12.84	14.20
Fidelity Bank	6.22	7.36
Wema Bank	3.13	3.87

TABLE 5: LSTM Residual Prediction Performance

MAE, Mean Absolute Error; RMSE, Root Mean Squared Error; LSTM, Long Short-Term Memory

Like SVR, LSTM's holdout test performance closely mirrors ARIMA, with no material improvement in prediction accuracy. This parity across three distinct algorithmic approaches, linear ARIMA, kernel-based SVR, and recurrent neural networks, suggests that the linear (ARIMA) component dominates price dynamics, and that residual components contain limited exploitable structure. However, the critical distinction emerges in the hybrid model combination framework, where the superior performance of the ARIMA-LSTM model demonstrates that the framework's success is not driven by the LSTM's inherent superiority as a standalone algorithm, but rather by the additive decomposition structure itself. By systematically separating the linear trend (ARIMA) from the non-linear shocks (LSTM), the hybrid architecture successfully captures information that neither algorithm can accurately model in isolation.

Hybrid results

The core innovation of this research centers on hybrid models that combine linear (ARIMA) and nonlinear (SVR, LSTM) components. The combination strategy follows this principle: the ARIMA model forecasts the base price trajectory, and the SVR or LSTM model predicts the residual deviation from this trajectory. The final forecast equation is

$$HybridForecast_t = ARIMAForecast_t + ResidualForecast_t$$

This decomposition leverages the complementary strengths of each method: ARIMA efficiently captures autoregressive dynamics, while SVR and LSTM specialize in capturing nonlinear features and temporal dependencies that ARIMA misses. The result is presented in Table *6*.

Bank	Model	MAE	RMSE	Improvement vs. ARIMA
Zenith Bank	ARIMA-SVR	2.73	3.63	78.4%
Zenith Bank	ARIMA-LSTM	2.35	3.24	81.4%
Fidelity Bank	ARIMA-SVR	1.07	1.52	82.7%
Fidelity Bank	ARIMA-LSTM	0.74	1.11	88.0%
Wema Bank	ARIMA-SVR	0.90	1.25	71.3%
Wema Bank	ARIMA-LSTM	0.64	0.94	79.6%

TABLE 6: Walk-Forward Validation Results

ARIMA, Auto-Regressive Integrated Moving Average; MAE, Mean Absolute Error; RMSE, Root Mean Squared Error; LSTM, Long Short-Term Memory; SVR, Support Vector Regression

The figures below further presents a comparison of the actual vs. predicted hybrid forecasts for the three banks.

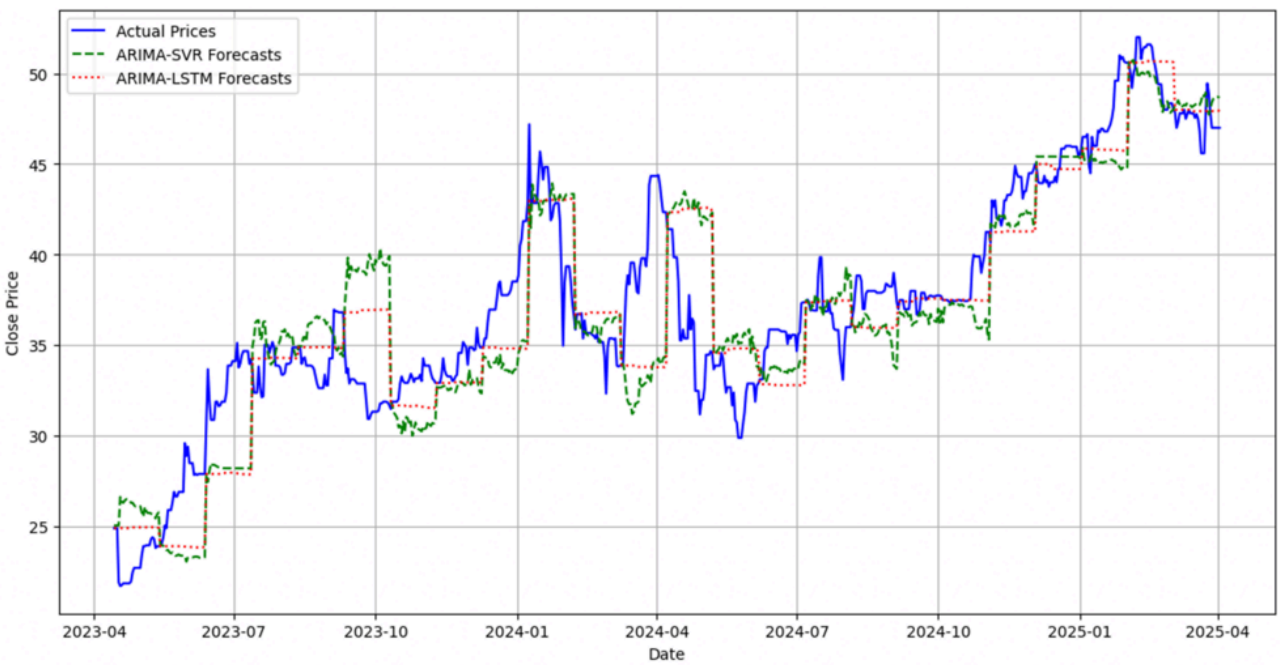

FIGURE 2: Actual vs. Hybrid Model Forecasts for Zenith Bank

ARIMA, Auto-Regressive Integrated Moving Average; LSTM, Long Short-Term Memory; SVR, Support Vector Regression

Figure 2 illustrates the comparative forecasting performance of the developed hybrid models against the actual daily closing prices of Zenith Bank over the testing period (April 2023-April 2025). The blue solid line represents the actual market price, which exhibits a significant bullish trend, rising from approximately ₦22.00 in early 2023 to peaks exceeding ₦50.00 in 2024. This period is characterized by high volatility, evidenced by sharp vertical rallies and sudden corrections, particularly between January 2024 and July 2024.

The ARIMA-LSTM hybrid model (red dotted line) demonstrates superior tracking capability, achieving an MAE of 2.35 compared to 2.73 for the ARIMA-SVR model. Given that Zenith Bank trades at a significantly higher price point (approximately ₦47.30) compared to the Tier 2 banks, an average error of ₦2.35 represents a relative deviation of roughly 5%. This indicates that while the absolute error is numerically higher than that of smaller banks, the model remains highly effective at capturing the trend and volatility of a large-cap stock.

FIGURE 3: Actual vs. Hybrid Model Forecasts for Fidelity Bank

ARIMA, Auto-Regressive Integrated Moving Average; LSTM, Long Short-Term Memory; SVR, Support Vector Regression

In Figure 3, for Fidelity Bank, the hybrid framework delivered robust predictive results, with the ARIMA-LSTM model achieving an MAE of 0.74 and an RMSE of 1.11. This performance significantly outperformed the ARIMA-SVR model (MAE 1.07), highlighting the LSTM's superior ability to navigate the moderate volatility associated with Tier 2 banking stocks. Given the stock's trading range during the test period (approximately ₦17.00-₦20.00), this low error margin confirms the model's precision in forecasting mid-cap assets.

FIGURE 4: Actual vs. Hybrid Model Forecasts for Wema Bank

ARIMA, Auto-Regressive Integrated Moving Average; LSTM, Long Short-Term Memory; SVR, Support Vector Regression

In Figure 4, for Wema Bank, the results demonstrate the highest level of predictive accuracy among the three tiers, with the ARIMA-LSTM hybrid model achieving an MAE of 0.72 and an RMSE of 1.04. This superior performance, compared to the ARIMA-SVR model (MAE 0.90), is partly attributable to the stock's lower price range (trading between ₦4.00 and ₦12.00). However, the extremely low error margin indicates that the hybrid framework is exceptionally efficient at modeling the price dynamics of mid-cap stocks, effectively capturing the subtle volatility shifts that characterize this segment of the Nigerian market.

The third objective of this research was to evaluate the developed hybrid models using SHAP to identify the most influential technical and macroeconomic predictors of stock price residuals. While statistical metrics confirmed the accuracy of the hybrid framework, SHAP analysis provided a transparent mechanism to decode the "black box" decisions of the machine learning components. This process quantified the marginal contribution of each variable, allowing for a comparative analysis of what drives volatility in different segments of the Nigerian banking sector. The SHAP summary plots for the three institutions presented in Table 7 reveals a distinct hierarchy of influence that characterizes the different tiers of the Nigerian banking sector.

Rank	Zenith Bank	Fidelity Bank	Wema Bank
1	Close_lag1	Close_lag1	Close_lag3
2	usd_ngn_price	bond_price_lag3	bond_price
3	Close_lag2	market_treasuryBill_lag3	usd_ngn_price_lag1
4	inflation_allItems_lag2	bond_price	inflation_allItems_lag1
5	inflation_foodYearOn	market_primeLending_lag2	inflation_allItems

TABLE 7: Top 5 Influential Predictors by Bank (Based on SHAP)

SHAP, SHapley Additive exPlanations

For Bank A, Close_lag1 and Close_lag2 are the dominant predictors, indicating a high degree of momentum-driven behavior and price continuity in large-cap stocks.

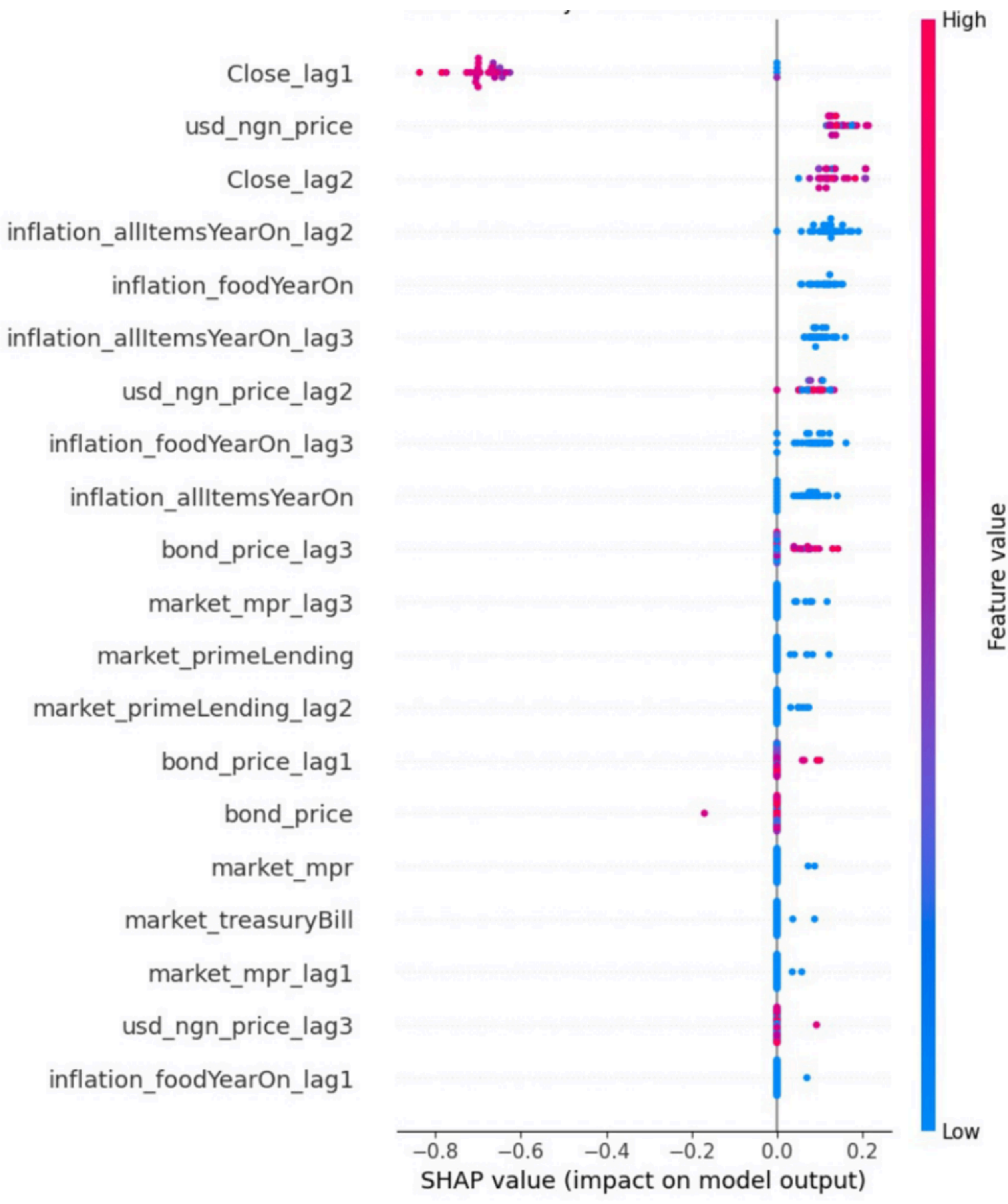

FIGURE 5: SHAP Summary Plot for Zenith Bank

SHAP, SHapley Additive exPlanations

Interestingly, as seen in Figure *5*, the USD/NGN Exchange Rate is the most significant macroeconomic driver for Zenith Bank. Higher exchange rates (represented by pink dots) show a positive impact on the residual prediction, suggesting that as a major international player, Zenith's stock volatility is closely tied to currency market stability.

FIGURE 6: SHAP Summary Plot for Fidelity Bank

SHAP, SHapley Additive exPlanations

As seen in Figure *6*, Fidelity Bank shows a stronger sensitivity to fixed-income signals, with bond_price_lag3 and market_treasuryBill_lag3 appearing as top predictors. This suggests that mid-cap banking stocks in Nigeria are highly sensitive to "crowding out" effects, where investors shift between equity and debt markets based on interest rate yields.

Close_lag1 remains highly influential, but the spread of SHAP values for macroeconomic variables is wider than in Zenith Bank, indicating more complex, non-linear reactions to economic shifts.

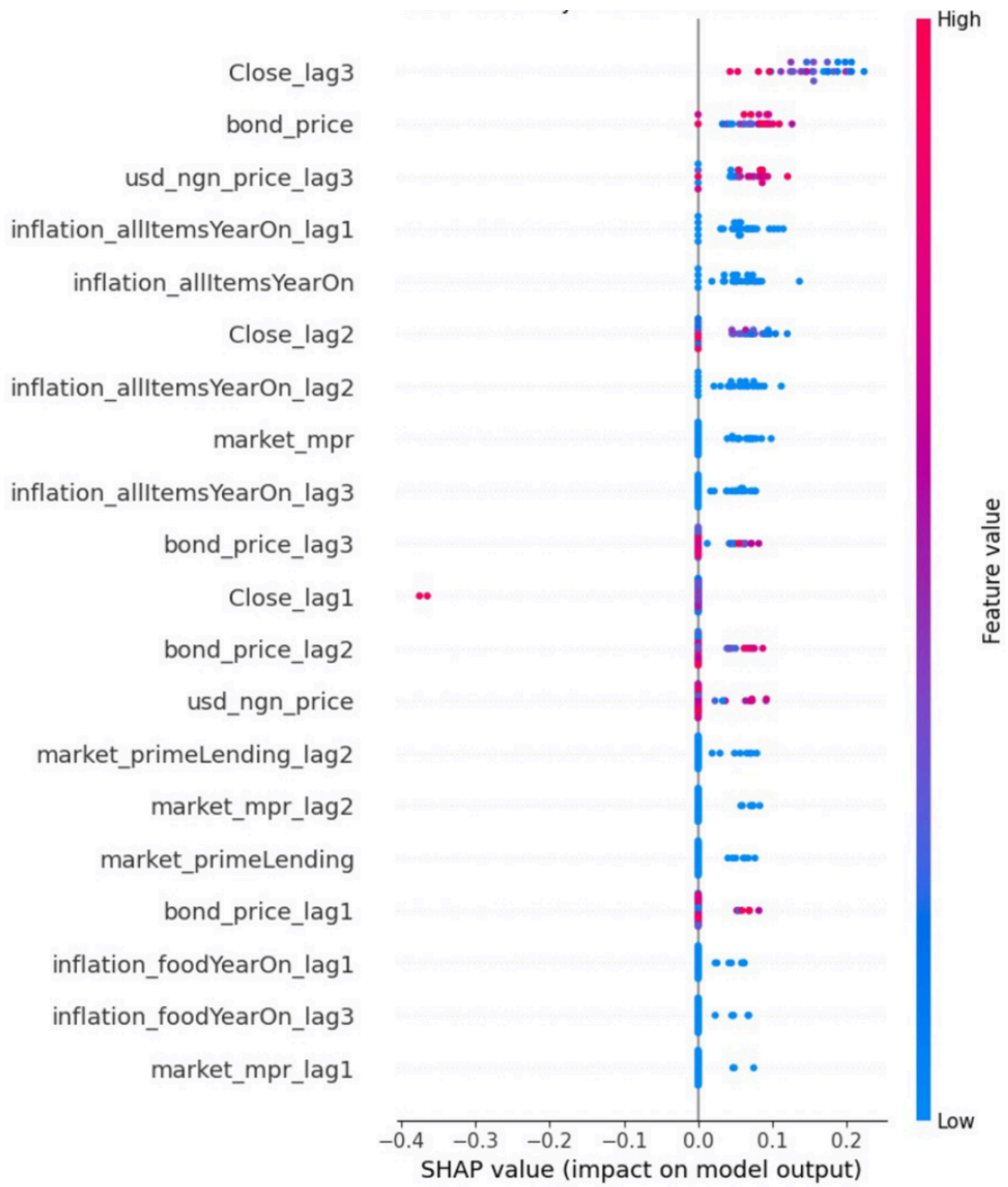

FIGURE 7: SHAP Summary Plot for Wema Bank

SHAP, SHapley Additive exPlanations

In the Wema Bank analysis in Figure 7, Close_lag3 and bond_price emerge as the most critical drivers. Unlike Zenith bank, which responds immediately to yesterday's price, Wema bank shows a "longer memory" for shocks from three days prior. The high significance of bond_price and USD/NGN price lags confirms that smaller-cap banks are more vulnerable to sudden fiscal policy changes and liquidity shocks in the Nigerian economy.

The successful application of SHAP confirms that the hybrid forecasting framework is not just a predictive tool but a diagnostic one. The analysis reveals that while Price Memory (Lags) provides the primary signal for short-term movements, Macroeconomic Shocks (Exchange Rates and Bond Yields) are the definitive drivers of the non-linear "surprises" that standalone models like ARIMA typically miss.

It is worthy of note that, while the hybrid models demonstrate improved predictive performance, the findings must be interpreted with caution. The sample size is limited to three specific banks, which, while stratified to represent different market tiers, restricts the statistical strength and external validity of the study. Consequently, these results are strongly indicative of the selected tiers but may not fully represent the broader, heterogeneous dynamics of the entire NSE or other emerging markets.

Conclusions

This study makes three key contributions to financial forecasting in emerging markets. First, it proves that the additive decomposition structure of combining ARIMA and LSTM models predicts stock prices significantly better than using single models, as neither algorithm achieved superior accuracy in isolation. Second, the research establishes that a bank's size fundamentally determines its vulnerability to economic shocks. Large-cap banks act as macroeconomic proxies, reacting to global factors like oil prices and currency exchange rates, whereas mid-cap and small-cap banks are driven heavily by domestic fixed-income yields and historical price memory. Finally, by integrating SHAP, the study demonstrates that complex artificial intelligence models can be transparent, explicitly identifying the economic triggers influencing market forecasts. However, these findings are derived from a restricted sample of three institutions. While the residual-based hybrid framework presents a strong architectural strategy, the results should be interpreted as indicative rather than conclusive. Future research must incorporate a larger, heterogeneous sample of financial institutions, deploy dynamic hyperparameter optimization, and utilize higher-frequency macroeconomic data to confirm the generalizability of this framework across the broader Nigerian banking sector.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Oluwatayofunmi Durodola, Aderonke Adegbenjo, Ayoade Iyiade, Oluwaseun Ilori

Acquisition, analysis, or interpretation of data: Oluwatayofunmi Durodola

Drafting of the manuscript: Oluwatayofunmi Durodola

Critical review of the manuscript for important intellectual content: Aderonke Adegbenjo, Ayoade Iyiade, Oluwaseun Ilori

Supervision: Aderonke Adegbenjo

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial**

relationships: All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

Acknowledgements

The authors would like to acknowledge the Department of Computer Science, School of Computing and Engineering Sciences, Babcock University, for providing the enabling environment to conduct this research. The data, analytic methods, and study materials are available upon request to the corresponding author.

References

1. Rizvi DR, Khalid M: Performance analysis of stocks using deep learning models. Procedia Computer Science. 2024, 233:753-762. 10.1016/j.procs.2024.03.264
2. Fattoh IE, El Maghawry Ibrahim M, Mousa FA: Unveiling market dynamics: a machine and deep learning approach to Egyptian stock prediction. Future Business Journal. 2025, 11:18. 10.1186/s43093-025-00421-0
3. Nigerian Stock Exchange (NSE). (2025). Accessed: March 4, 2026: https://www.weforum.org/organizations/nigerian-stock-exchange-nse.
4. Nigeria Nigeria Stock Exchange: Market Capitalization. (2025). Accessed: November 5, 2025: https://www.ceicdata.com/en/nigeria/nigeria-stock-exchange-market-capitalization/market-capitalization-nigeria-stock-....
5. NGX Domestic and Foreign Portfolio Investment Report. (2025). Accessed: March 4, 2026: https://doclib.ngxgroup.com/market_data-site/other-market-information-site/FPI%20Report/NGX%20Domestic%20and%20Foreig....
6. NGX Domestic and Foreign Portfolio Investment Report - July 2025. (2025). Accessed: March 4, 2026: https://doclib.ngxgroup.com/market_data-site/other-market-information-site/FPI%20Report/NGX%20Domestic%20and%20Foreig....
7. Vishwas HN, Usha BA, Sangeetha KN: Artificial intelligence-based stock market price prediction, a review. Mathematical Models in Engineering. November 12, 11:204-209. 10.21595/mme.2025.24861
8. Isenah G, Olubusoye O: Forecasting Nigerian stock market returns using ARIMA and artificial neural network models. CBN Journal of Applied Statistics. 2014, 5:25-48.
9. Olaosebikan O, Olaniyi RB, Ojekunle FJ: Impact of capital market performance on capital formation in Nigeria. FUOYE Journal of Finance and Contemporary Issues. 2022, 3:148-63.
10. Ogochukwu UA, Ishioro BO: Stock market development, capital formation and industrial sector performance in Nigeria. Himalayan Journal of Economics and Business Management. 2021, 2:1-6. 10.47310/hjebm.2023.v04i02.102
11. Edeh MB: Behavioral finance and investors' stock investments decisions in West Africa: evidence from the Nigerian stock market. Journal of Global Economics and Business. 2020, 1:41-58. 10.31039/jgeb.v1i3.76
12. Malki A, Atlam ES, Hassanien AE, Ewis A, Dagnew G, Gad I: SARIMA model-based forecasting required number of COVID-19 vaccines globally and empirical analysis of peoples' view towards the vaccines. Alexandria Engineering Journal. 2022, 61:12091-12110. 10.1016/j.aej.2022.05.051
13. Paliari I, Karanikola A, Kotsiantis S: A comparison of the optimized LSTM, XGBoost and ARIMA in time series forecasting. 2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA), Chania Crete, Greece. 2021, 1-7. 10.1109/IISA52424.2021.9555520
14. Ariyo AA, Adewumi AO, Ayo CK: Stock price prediction using the ARIMA model. 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK. 2014, 106-112. 10.1109/UKSim.2014.67

15. Shi H: Tesla stock price forecast based on ARIMA model and machine learning techniques. Science and Technology of Engineering, Chemistry and Environmental Protection. 2024, 1:1-5. 10.61173/pf98r920

16. Han C, Fu X: Challenge and opportunity: deep learning-based stock price prediction by using bi-directional LSTM model. Frontiers in Business, Economics and Management. 2023, 8:51-54. 10.54097/fbem.v8i2.6616

17. Ayitey Junior M, Appiahene P, Appiah O: Forex market forecasting with two-layer stacked Long Short-Term Memory neural network (LSTM) and correlation analysis. Journal of Electrical Systems and Information Technology. 2022, 9:14. 10.1186/s43067-022-00054-1

18. Bhandari HN, Rimal B, Pokhrel NR, Rimal R, Dahal KR, Khatri RKC: Predicting stock market index using LSTM. Machine Learning with Applications. 2022, 9:100320. 10.1016/j.mlwa.2022.100320

19. Weng B, Lu L, Wang X, Megahed FM, Martinez W: Predicting short-term stock prices using ensemble methods and online data sources. Expert Systems with Applications. 2018, 112:258-273. 10.1016/j.eswa.2018.06.016

20. Pasupulety U, Abdullah Anees A, Anmol S, Mohan BR: Predicting stock prices using ensemble learning and sentiment analysis. 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), Sardinia, Italy. 2019, 215-222. 10.1109/AIKE.2019.00045

21. NGX Group: Steering market to world-beating 51.19% rally in 2025. (2026). Accessed: March 4, 2026: https://ngxgroup.com/ngx-group-steering-market-to-world-beating-51-19-rally-in-2025/.

22. Abdulkareem A, Abdulkareem K: Analysing oil price-macroeconomic volatility in Nigeria. CBN Journal of Applied Statistics. 2016, 7:1-22.

23. Umoru D, Ohiomu S, Akpeke R: The influence of oil price volatility on selected macroeconomic indicators. Acta Universitatis Bohemiae Meridionalis. 2018, 21:1-22.

24. Adenekan AT, Hilili MJ, Okereke AN: Oil price, exchange rate and stock market performance in Nigeria. CBN Economic and Financial Review. 2020, 58:63-103.

25. Oyasor EI: Forecasting global stock prices using gated recurrent unit, long short-term memory weighted-LSTM and LSTM with attention. Malete Journal of Accounting and Finance. 2025, 6:147-161.

26. Arowolo OT, Ekum MI, Okorafor U, Akinyemi JA, Ikegwu EM: A hybrid autoregressive-long short-term memory time series model for forecasting stock prices. Science World Journal. 2025, 20:1191-1198. 10.4314/swj.v20i3.38

27. How stratified random sampling works, with examples. Accessed: November 5, 2025: https://www.investopedia.com/terms/stratified_random_sampling.asp.

28. Stratified random sampling. Accessed: November 5, 2025: https://corporatefinanceinstitute.com/resources/data-science/stratified-random-sampling/.

29. Company profile - Nigerian Exchange Limited. Accessed: November 5, 2025: https://ngxgroup.com/exchange/data/company-profile/?symbol=ZENITHBANK.

30. Wema Bank PLC. (2026). Accessed: March 4, 2026: https://www.tradingview.com/symbols/NSENG-WEMABANK/.

31. Nigeria Share Market. Accessed: February 27, 2025: https://ng.investing.com/equities/nigeria.

32. Data & Statistics. Central Bank of Nigeria. Accessed: November 5, 2025: https://www.cbn.gov.ng/rates/.