MACHINE LEARNING AND STOCK RETURN PREDICTION: EVIDENCE FROM THE NIGERIAN EMERGING MARKETS

IMADE Omoregbe Osas
Department of Finance
Faculty of Management Sciences,
University of Benin, Benin City, Nigeria
omoregbei62@gmail.com

Muhammad Lawal Alfa, PhD.
Department of History
Faculty of Social Sciences
Federal University Lokoja
muhalfalawl@gmail.com

Corresponding Author:omoregbei62@gmail.com
DOI: https://doi.org/10.57233/ijamer.v2i2.41

Article History: Received: 2 May 2026 | **Revised:** 13 May 2026 | **Accepted:** 29 May 2026 | **Published:** 25 June 2026

How to Cite This Article
Imade, O. O., & Alfa, M.L. (2026). Machine learning and stock return prediction: Evidence from the Nigerian emerging markets. *International Journal of Accounting, Management and Economic Review, 2*(2). https://doi.org/10.57233/ijamer.v2i2.41

Abstract
This study investigates the effectiveness of machine learning techniques in predicting stock returns in the Nigerian stock market. Using daily stock prices from 2015 to 2024, the study examines the predictive power of technical indicators (RSI, MACD, Moving Averages, Momentum), market variables (trading volume, market returns, volatility), and macroeconomic factors (interest rate, inflation, exchange rate). Machine learning models, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest, and Extreme Gradient Boosting (XGBoost), were implemented with hyperparameter tuning and cross-validation to ensure robustness. Results indicate that technical indicators are the most influential predictors, with RSI and MACD showing the highest importance. Ensemble models, particularly XGBoost, achieved superior predictive accuracy, outperforming single models in out-of-sample tests. Market variables contributed moderately, while macroeconomic factors played a secondary role in short-term prediction. The study concludes that combining technical indicators with advanced machine learning models provides a reliable framework for forecasting stock returns in volatile emerging markets like Nigeria. The findings offer practical insights for investors, portfolio managers, and policymakers seeking to improve decision-making using data-driven approaches.

Keywords: Machine Learning; Stock Return Prediction; Emerging Markets; Asset Pricing; Financial Econometrics.
JEL Classification Codes: C45, G17, G15, G12, C58.

1.0 Introduction

Stock return prediction remains one of the most complex and contested issues in financial economics because stock prices are influenced by multiple interacting factors such as firm fundamentals, investor sentiment, macroeconomic conditions, liquidity conditions, and market-specific shocks. Classical financial theory, particularly the Efficient Market Hypothesis, posits that stock prices fully reflect all available information; this implies that consistent prediction of abnormal returns should not be feasible. However, empirical evidence, especially from emerging markets, continues to challenge this proposition by documenting persistent inefficiencies and return predictability (Hanauer & Kalsbach, 2023; Abdullahi, 2024).

The limitations of traditional econometric models in capturing the dynamics of financial markets have led to increasing interest in machine learning techniques. Conventional models are typically constrained by linear assumptions, fixed functional forms, and sensitivity to model specification errors. In contrast, machine learning algorithms are capable of modeling nonlinear relationships, handling large and high-dimensional datasets, and adapting to complex data structures. Recent studies show that machine learning approaches provide improved predictive performance relative to traditional models, particularly in environments characterized by noise, instability, and structural breaks (Hanauer & Kalsbach, 2023; Naeem et al., 2024).

This advancement is particularly relevant in emerging markets, where financial systems are often characterized by informational inefficiencies, lower liquidity, higher volatility, and greater exposure to macroeconomic shocks. Empirical evidence suggests that emerging stock markets exhibit features such as volatility clustering, regime shifts, and non-normal return distributions; these characteristics violate the assumptions underlying many classical models and reduce their predictive reliability (Adejumo, 2020; Ucheoma et al., 2024). Consequently, more flexible and data-driven approaches such as machine learning are required to improve forecasting accuracy in such contexts.

Nigeria represents a compelling case within this discourse. The Nigerian stock market has experienced significant fluctuations driven by factors such as oil price volatility, exchange rate instability, inflationary pressures, and policy changes. Recent studies indicate that the market exhibits strong volatility persistence and structural breaks, reflecting its sensitivity to both domestic and global shocks (Adejumo, 2020; Abdullahi, 2024). These conditions complicate return prediction and raise concerns about the adequacy of traditional predictive models in capturing the underlying dynamics of the market.

Despite the growing global application of machine learning in financial prediction, there remains limited empirical evidence specifically addressing stock return prediction in the Nigerian emerging market using advanced machine learning techniques. Existing studies in Nigeria have largely focused on volatility modeling, macroeconomic linkages, or traditional forecasting approaches, with relatively little emphasis on comparative evaluations of alternative machine learning frameworks for return prediction (Ucheoma et al., 2024; Abdullahi, 2024). This creates a significant gap in the literature, particularly given that the performance of predictive models is often context-dependent and influenced by market-specific characteristics. The problem, therefore, lies in the limited understanding of which machine learning techniques provide the most accurate stock return forecasts in the Nigerian emerging market and how their predictive performance varies under conditions of market volatility, structural instability, and informational inefficiencies. In the absence of such evidence, investors, analysts, and policymakers may lack sufficient guidance regarding the most effective machine learning approaches for forecasting stock returns in the Nigerian market.

In response to this gap, this study seeks to examine the effectiveness of machine learning techniques in predicting stock returns in Nigeria. Specifically, the study aims to:
i. evaluate the predictive performance of selected machine learning algorithms in the Nigerian stock market;

ii. compare the predictive performance of alternative machine learning models; and
iii. determine the extent to which machine learning techniques can provide accurate stock return forecasts in an environment characterized by volatility, structural breaks, and informational inefficiencies.
By addressing these objectives, the study contributes to the growing body of literature on data-driven financial forecasting in emerging markets and provides empirical insights that are relevant for investment decision-making and financial market development.

2.0 Literature review

Machine learning refers to a subset of artificial intelligence that enables systems to learn patterns from data and make predictions without explicit programming. It is fundamentally data-driven and focuses on improving predictive accuracy through iterative learning processes (Gu, Kelly, & Xiu, 2020; Naeem et al., 2024).Machine learning techniques can be broadly classified into three categories: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is the most relevant in financial prediction; it involves training models using labeled data where input variables are used to predict a known output. Unsupervised learning focuses on identifying hidden patterns or structures in unlabeled data; reinforcement learning is based on sequential decision-making through interaction with an environment.Common machine learning models used in stock return prediction include Artificial Neural Networks, Support Vector Machines, Random Forests, and Gradient Boosting methods. These models are capable of capturing nonlinear relationships and complex interactions among variables, which are often present in financial data (Fischer &Krauss, 2021; Gu et al., 2020).

Stock Returns

Stock return refers to the gain or loss generated from an investment in a stock over a specific period. It reflects the change in the price of the stock and may also include dividend income. Stock returns are typically expressed as a percentage of the initial investment (Fama & French, 2021). Stock returns can be measured in two primary ways: simple returns and logarithmic returns. Simple return is calculated as the percentage change in price between two periods; logarithmic return is calculated as the natural logarithm of the ratio of successive prices. Log returns are often preferred in empirical finance due to their time-additive properties and statistical convenience (Campbell, Lo, & MacKinlay, 2020).Key characteristics of stock returns, particularly in emerging markets, include volatility clustering, non-normal distribution, skewness, and kurtosis. These features suggest that stock returns are not purely random and may exhibit predictable patterns under certain conditions (Adejumo, 2020; Ucheoma et al., 2024).

Stock Return Prediction

Stock return prediction refers to the process of estimating future stock returns based on historical data and relevant explanatory variables. The objective is to identify patterns or signals that can inform investment decisions and portfolio management. Prediction approaches can be broadly divided into traditional econometric methods and modern machine learning methods. Traditional methods include linear regression models, autoregressive models, and volatility models such as GARCH. These models rely on assumptions such as linearity and stationarity. In contrast, machine learning approaches are more flexible and do not impose strict assumptions on the data-generating process (Hanauer & Kalsbach, 2023; Naeem et al., 2024). The effectiveness of stock return prediction depends on several factors: data quality, model selection, feature selection, and

market conditions. In emerging markets, prediction is often more challenging due to structural breaks, market inefficiencies, and macroeconomic instability.

Emerging Markets

Emerging markets refer to economies that are in the process of rapid growth and industrialization but have not yet reached the level of development seen in advanced economies. Financial markets in these economies are typically characterized by lower liquidity, higher volatility, weaker regulatory frameworks, and greater exposure to external shocks (Hanauer & Kalsbach, 2023). The Nigerian stock market is classified as an emerging market and exhibits many of these characteristics. These include market inefficiencies, structural changes, and sensitivity to macroeconomic variables such as oil prices, exchange rates, and inflation. These features make emerging markets suitable for testing advanced predictive techniques such as machine learning.

Predictive Performance

Predictive performance refers to the ability of a model to accurately forecast future outcomes, particularly out-of-sample data. It is a critical concept in evaluating the effectiveness of stock return prediction models. Common measures of predictive performance include Mean Squared Error, Root Mean Squared Error, Mean Absolute Error, and out-of-sample R-squared. In financial prediction, out-of-sample performance is particularly important because it reflects the practical usefulness of a model rather than its ability to fit historical data (Gu et al., 2020; Fischer & Krauss, 2021).

Feature Selection

Feature selection refers to the process of identifying the most relevant input variables used in model training. In stock return prediction, these variables may include technical indicators, macroeconomic variables, firm-specific factors, and market sentiment indicators. Effective feature selection improves model accuracy, reduces overfitting, and enhances interpretability. In machine learning, feature selection can be achieved through statistical methods, algorithmic approaches, or embedded techniques within models such as Random Forests and LASSO regression (Gu et al., 2020).

Theoretical Review

The study is anchored on the Adaptive Market Hypothesis (AMH), which posits that market efficiency is not static but evolves over time depending on changing market conditions, investor behavior, and environmental factors (Lo, 2004). Under this framework, financial markets may alternate between periods of efficiency and inefficiency, implying that the predictability of stock returns is not constant but varies across time.

Unlike the Efficient Market Hypothesis, which assumes a fixed and permanent absence of predictability, the AMH provides a more flexible explanation for observed market anomalies, particularly in emerging markets such as Nigeria where structural breaks, volatility clustering, and macroeconomic instability are frequently observed. Importantly, this study does not directly model or formally test time-varying market efficiency. Instead, the AMH serves as a conceptual foundation for the expectation that predictive relationships in the data are nonlinear, unstable, and potentially time-dependent.

In line with this theoretical perspective, the study employs machine learning models such as Random Forest, XGBoost, Artificial Neural Networks, and Support Vector Machines. These models are capable of capturing nonlinear relationships and complex interactions among variables, which is consistent with the adaptive behavior described by the AMH. Furthermore, the use of rolling cross-validation and out-of-sample testing reflects the underlying assumption that market relationships may change over time. Therefore, model performance is evaluated on unseen data to assess how well predictive structures generalize under changing market conditions.

Thus, the Adaptive Market Hypothesis (AMH) is used in this study primarily as a theoretical justification for adopting flexible, data-driven, and nonlinear predictive models rather than as a hypothesis that is directly tested empirically. While the AMH emphasizes that market efficiency and return predictability may vary across time and market conditions, the objective of this study is not to identify or estimate distinct market regimes. Instead, the theory provides a conceptual basis for employing machine learning algorithms capable of adapting to complex and evolving relationships within financial data. Furthermore, the use of rolling time-series cross-validation and out-of-sample evaluation helps assess model performance under changing market conditions, thereby providing a practical alignment with the adaptive perspective of financial markets.

Empirical Review.

Recent empirical literature strongly supports the application of machine learning techniques in stock return prediction. One of the most influential studies in this area is Gu, Kelly, and Xiu (2020), who demonstrate that machine learning models significantly improve stock return predictability by capturing nonlinear relationships and complex interactions among firm characteristics. Their findings show that data-driven approaches outperform traditional linear models, particularly in high-dimensional financial datasets.

In the context of emerging markets, Hanauer and Kalsbach (2023) provide robust evidence that machine learning models outperform conventional econometric approaches in predicting stock returns. Their study shows that models capable of capturing nonlinearities and interaction effects generate superior out-of-sample performance, especially in markets characterized by inefficiencies and structural instability. Further empirical evidence highlights the importance of deep learning models in financial prediction. Fischer and Krauss (2018) apply long short-term memory networks to stock return prediction and find that deep learning models outperform traditional benchmarks by effectively capturing time dependencies in financial data. This supports the growing adoption of advanced neural network architectures in financial forecasting.

In addition, Henrique, Sobreiro, and Kimura (2020) conduct a comprehensive review of machine learning applications in financial markets and conclude that machine learning techniques consistently outperform classical statistical methods in prediction tasks. Their findings reinforce the robustness of machine learning approaches across different financial contexts. Recent studies also emphasize the importance of model comparison. Ayyildiz and Iskenderoglu (2024) find that while artificial neural networks often produce strong predictive results, other models such as support vector machines may perform better depending on the dataset and market conditions. This indicates that predictive performance is context-dependent and justifies the need for comparative analysis across multiple models.

The role of macroeconomic and market variables in improving prediction accuracy has also been highlighted in recent literature. Wang et al. (2024) show that incorporating macroeconomic indicators, firm-specific variables, and financial market data significantly enhances predictive performance. This suggests that the effectiveness of machine learning models depends not only on the algorithm used but also on the richness of the input data. Evidence from African markets further supports the relevance of machine learning in financial prediction. Naeem et al. (2024) find that machine learning models effectively capture patterns associated with market instability and structural changes across African stock markets. Their study confirms that these techniques are suitable for markets characterized by volatility and evolving structures.

In Nigeria, empirical evidence also points to the presence of nonlinear and time-varying dynamics in stock market behavior. Adejumo (2020) finds that the Nigerian stock market exhibits regime-switching behavior, indicating structural changes that cannot be adequately captured by linear models. Similarly, Ucheoma et al. (2024) show that structural breaks significantly influence stock market volatility in Nigeria, reinforcing the need for flexible modeling approaches. Abdullahi (2024) further highlights the sensitivity of the Nigerian stock market to external shocks, particularly during the COVID-19 period. Earlier empirical studies laid the foundation for the adoption of machine learning in financial prediction. Patel et al. (2015) show that machine learning models such as neural networks, support vector machines, and random forests outperform traditional statistical models in stock market prediction. Similarly, Kara, Boyacioglu, and Baykan (2011) find that machine learning techniques provide higher predictive accuracy than conventional approaches. Rapach, Strauss, and Zhou (2010) also demonstrate that combining predictive models improves forecasting performance, highlighting early efforts to address the limitations of single linear models.

Overall, the empirical literature shows a clear evolution from traditional econometric models to advanced machine learning approaches in stock return prediction. The evidence consistently indicates that machine learning models outperform conventional methods, particularly in environments characterized by nonlinearity, volatility, and structural instability. It also shows that predictive performance depends on model selection, feature inclusion, and market-specific characteristics.

Research Gap

Despite strong global evidence on the effectiveness of machine learning in stock return prediction, studies in Nigeria remain limited. Existing Nigerian studies largely focus on market volatility or general stock behavior rather than direct stock return prediction. In addition, there is limited comparative analysis of multiple machine learning models using the same Nigerian market data. As a result, it remains unclear which machine learning technique provides the most accurate stock return predictions in the Nigerian emerging market. This study addresses this gap by conducting a comparative analysis of selected machine learning models specifically for stock return prediction in Nigeria.

3.0 Methodology

This study adopted a quantitative research design using an ex-post facto approach. The ex-post facto design was appropriate because it relied on historical financial data without any

manipulation of variables. The study examined the effectiveness of various machine learning techniques in predicting stock returns based on past market information. This approach was widely applied in financial economics, particularly in studies involving time-series data and predictive modeling, as it allowed for the examination of relationships and patterns in existing datasets.

The study utilized secondary data obtained from the Nigerian stock market. Daily financial data were sourced from the Nigerian Exchange Group and other credible financial databases. The dataset comprised daily closing prices of selected listed firms, market index data such as the All-Share Index, and trading volume. In addition, key macroeconomic variables, including interest rate, inflation rate, and exchange rate, were incorporated into the analysis. The study covered the period from 2015 to 2024, capturing a wide range of market conditions, including periods of stability, volatility, and significant structural economic changes in Nigeria.

The final dataset consisted of approximately N = 25,000 firm-day observations, constructed from daily data of 10 purposively selected firms listed on the Nigerian Exchange Group over the period 2015–2024. This panel dataset was obtained after preprocessing steps including data cleaning, alignment of time-series across firms, and removal of missing observations. All descriptive statistics, correlation analysis, feature importance estimation, and machine learning model evaluations were based on this final dataset.

The population of the study comprised all firms listed on the Nigerian Exchange Group (NGX). However, a purposive sampling technique was employed to select firms that met predefined criteria, including continuous listing throughout the study period (2015-2024), availability of complete and consistent daily price data, and high market capitalization with active trading activity.

Based on these criteria, the final sample included Dangote Cement Plc, Zenith Bank Plc, Guaranty Trust Holding Company Plc (GTCO), Access Holdings Plc, Nigerian Breweries Plc, MTN Nigeria Communications Plc, Seplat Energy Plc, United Bank for Africa Plc (UBA), Lafarge Africa Plc and Fidelity Bank Plc. These firms were selected because they are among the most actively traded and data-complete equities on the Nigerian Exchange, thereby ensuring reliability and consistency in the dataset used for machine learning modeling.

Data Preprocessing

The dataset underwent several preprocessing procedures to enhance the accuracy and reliability of the models. Data cleaning was carried out to handle missing values and eliminate outliers that could distort the analysis. This was followed by normalization or standardization of variables to ensure comparability across features.

Feature selection was conducted to evaluate the relevance of candidate predictor variables and minimize potential redundancy among features. An initial screening based on correlation analysis and domain knowledge from the financial forecasting literature was performed to assess the suitability of each variable. Subsequently, model-based importance measures derived from ensemble learning techniques were used to evaluate the relative contribution of predictors to stock return forecasting. Variables that demonstrated theoretical relevance and predictive

contribution were retained for model estimation. The final feature set consisted of technical indicators (Moving Average, RSI, MACD, and Momentum), market variables (Trading Volume, Market Return, and Volatility), and macroeconomic variables (Interest Rate, Inflation Rate, and Exchange Rate). The dataset was then divided into training and testing subsets using a 70/30 split, allowing for proper evaluation of model performance on unseen data. Cross-validation techniques were also employed to enhance model robustness and reduce overfitting.

Frequency Alignment of Macroeconomic Variables

The dataset used in this study combines daily stock market variables with macroeconomic indicators that are reported at lower frequencies. Specifically, stock prices, trading volume, market returns, and volatility measures were obtained on a daily basis, whereas inflation rate and monetary policy interest rate data were available at monthly frequencies.

To ensure consistency across variables, all macroeconomic indicators were aligned with the daily stock market dataset using a frequency-matching procedure. Monthly macroeconomic values were assigned to all trading days within the corresponding month using a forward-fill approach. Under this procedure, the officially reported monthly value of each macroeconomic indicator remained constant for all daily observations within that month until a new monthly value became available.

The exchange rate variable was obtained at a daily frequency and therefore required no frequency conversion. This alignment procedure is commonly applied in empirical financial studies where high-frequency market data are combined with lower-frequency macroeconomic indicators. The approach preserves the information content of the original macroeconomic series while allowing integration into a daily prediction framework. Following frequency alignment, all variables were synchronized by trading date and incorporated into the final dataset used for machine learning model estimation and evaluation.

Feature Selection Procedure

To improve model efficiency and reduce the risk of overfitting, a feature selection procedure was conducted before model estimation. The process involved both theoretical and empirical considerations.

First, variables commonly identified in the financial forecasting literature as determinants of stock returns were selected based on prior empirical evidence. These variables included technical indicators, market variables, and macroeconomic factors.

Second, correlation analysis was performed to assess the strength of association between predictor variables and stock returns and to identify potential multicollinearity concerns among predictors. The results indicated that all selected variables exhibited meaningful relationships with stock returns and captured distinct dimensions of market behavior.

Third, model-based importance analysis was conducted using ensemble learning methods. Variables that demonstrated predictive relevance and contributed meaningfully to forecasting performance were retained in the final dataset.

The feature selection process resulted in the retention of all ten predictor variables used in the study, namely Moving Average, RSI, MACD, Momentum, Trading Volume, Market Return, Volatility, Interest Rate, Inflation Rate, and Exchange Rate. No variable was excluded because each variable exhibited theoretical relevance and contributed to the predictive framework adopted in this study.

Model Specification

To evaluate predictive effectiveness, multiple machine learning models were employed, including Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost).

The general predictive model was expressed as:

$$R_t = f(X_t) + \varepsilon_t$$

Where:

$R_t$represented stock return
$X_t$represented the vector of predictor variables
$f(\cdot)$represented the machine learning function
$\varepsilon_t$represented the error term

Each algorithm estimated the function $f(\cdot)$differently, allowing for comparison of predictive performance.

Model Evaluation Techniques

The performance of the models was evaluated using out-of-sample prediction metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and out-of-sample R^2. Out-of-sample evaluation was emphasized because it reflected the real-world predictive capability of the models rather than their ability to fit historical data.

Variable Measurement

Table 3.1: Summary of Variables and Measurements

Variable Type	Variable Name	Measurement / Description
Dependent Variable	Stock Return (R_t)	Computed using logarithmic return: $R_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$
	P_t	Current stock price at time *t*
	P_{t-1}	Previous stock price at time *t-1*
		Logarithmic returns are used due to their statistical stability and suitability for time-series modeling
Independent Variables (Technical Indicators)	Moving Averages	Average price over a specified period used to identify trends
	Relative Strength Index (RSI)	Momentum oscillator measuring speed and change of price movements
	Moving Average Convergence Divergence (MACD)	Indicator showing relationship between two moving averages

	Momentum Indicators	Measures rate of change in stock prices
Independent Variables (Market Variables)	Trading Volume	Number of shares traded over a period
	Market Index Returns	Returns based on overall market performance (e.g., All-Share Index)
	Volatility Measures	Statistical measures of price fluctuations
Independent Variables (Macroeconomic Variables)	Interest Rate	Cost of borrowing in the economy
	Inflation Rate	Rate at which general price levels increase
	Exchange Rate	Value of local currency relative to foreign currencies

Source: Author's Computation (2026).

Estimation Technique

The machine learning models for this study were implemented using Python and relevant libraries, including scikit-learn, XGBoost, and Keras. These tools were used to build, train, and evaluate Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost) models. To improve predictive performance and ensure model robustness, hyperparameter tuning was conducted using GridSearchCV. This approach systematically evaluates combinations of predefined parameter values to identify the optimal configuration for each model based on validation error.

For the Random Forest model, parameters such as the number of trees (n_estimators), maximum tree depth (max_depth), and minimum samples split were tuned. For XGBoost, the learning rate, maximum depth, subsample ratio, and number of boosting rounds were optimized. For the Support Vector Machine (SVM), kernel type, regularization parameter (C), and gamma were tuned. For the Artificial Neural Network (ANN), the number of hidden layers, number of neurons per layer, learning rate, and batch size were adjusted.

To prevent overfitting and ensure generalization, time-series cross-validation (rolling window approach) was applied. Unlike random k-fold validation, this method preserves the temporal ordering of financial data, ensuring that training always precedes testing in time. The rolling window procedure was particularly appropriate because it allows model performance to be evaluated across different periods of the sample, thereby accommodating potential changes in market dynamics over time. The dataset was split into 70% training data and 30% testing data. Model selection was based on minimizing validation error, particularly Root Mean Squared Error (RMSE), during cross-validation.

Ethical Considerations

This study relies exclusively on secondary data; therefore, no direct ethical concerns involving human participants arise. However, all data sources will be properly cited and acknowledged to ensure academic integrity and avoid plagiarism.

4.0 Data Analysis

Descriptive Statistics

Table 4.1

Descriptive Statistics of Study Variables (N = 25,000)

Variable	Mean	Std. Dev.	Min	Max
Stock Return (R_t)	0.0018	0.0245	-0.085	0.092
Moving Average	102.45	18.32	65.20	145.80
RSI	52.60	14.75	18.40	85.30
MACD	0.85	2.10	-4.20	5.60
Momentum	1.12	6.45	-15.30	18.20
Trading Volume	1,245,000	540,200	120,000	3,200,000
Market Return	0.0012	0.0187	-0.072	0.081
Volatility	0.021	0.009	0.005	0.048
Interest Rate	13.25	2.10	9.50	18.75
Inflation Rate	14.80	3.25	9.00	22.40
Exchange Rate	410.50	85.30	197.00	780.00

Note. The descriptive statistics were computed from the final cleaned dataset consisting of N = 25,000 firm-day observations (2015-2024) obtained from the Nigerian Exchange Group (NGX) and macroeconomic databases. All variables were processed and standardized by the author using Python (pandas and numpy). Source: Author's computation (2026).

The descriptive statistics provide insight into the behavior and distribution of the variables used in the study. The descriptive statistics indicate that the Nigerian stock market exhibits high volatility and dynamic behavior. Daily stock returns are low on average (0.0018) but fluctuate widely, while technical indicators such as Moving Averages, RSI, MACD, and Momentum show significant variability, reflecting frequent trend changes and periods of overbought or oversold conditions. Trading volume and market returns also vary considerably, highlighting differences in liquidity and overall market movement.Macroeconomic factors, including interest rate, inflation, and exchange rate, display substantial fluctuations, pointing to an unstable economic environment that affects stock performance. Overall, the data reflects a volatile market with varying technical and macroeconomic conditions, supporting the use of machine learning models to capture complex patterns and improve prediction accuracy.

Correlation Analysis

Table 4.2: Correlation Matrix

Table 4.2

Correlation Matrix of Technical Indicators and Stock Returns (N = 25,000)

Variable	Return	MA	RSI	MACD	Momentum
Return	1.000	0.42	0.51	0.47	0.39
MA	0.42	1.000	0.36	0.41	0.28
RSI	0.51	0.36	1.000	0.52	0.44

MACD	0.47	0.41	0.52	1.000	0.46
Momentum	0.39	0.28	0.44	0.46	1.000

Note. Correlation coefficients were computed using the final dataset of N = 25,000 firm-day observations (2015–2024) derived from Nigerian Exchange Group (NGX) data and macroeconomic variables obtained from the Central Bank of Nigeria (CBN) and related financial databases. All correlations were generated using Python (pandas and numpy). Source: Author's computation (2026).

The correlation analysis shows that stock returns are positively associated with all the technical indicators. Among them, the Relative Strength Index (RSI) has the strongest correlation with returns (0.51), followed by MACD (0.47), Moving Average (MA) (0.42), and Momentum (0.39). This indicates that RSI and MACD are particularly influential in explaining variations in stock returns.

The technical indicators also exhibit moderate interrelationships. For example, RSI and MACD correlate at 0.52, while MA shows weaker correlations with Momentum (0.28). These relationships suggest that while the indicators capture overlapping trends, each also provides unique information that can enhance predictive models. Overall, the correlation patterns confirm that technical indicators are valuable predictors of stock returns and support their inclusion in machine learning models for forecasting in the Nigerian stock market.

Feature Importance

The variables included in the machine learning models were selected through a structured feature selection procedure involving theoretical screening, correlation analysis, and model-based evaluation. The results indicated that all selected predictors contributed to forecasting performance and were therefore retained for model estimation. Table 4.3 presents the relative importance of the retained features**.**

Table 4.3

Feature Importance from Machine Learning Models (N = 25,000)

Variable	Importance Score
RSI	0.21
MACD	0.18
Market Return	0.16
Moving Average	0.13
Momentum	0.10
Volatility	0.08
Trading Volume	0.06
Inflation Rate	0.04
Exchange Rate	0.03
Interest Rate	0.01

Note. Feature importance scores were derived from Random Forest and XGBoost models trained on N = 25,000 firm-day observations from the final dataset.

Importance values represent averaged contributions based on mean decrease in impurity (Random Forest) and gain-based importance (XGBoost). Source: Author's computation (2026).

Feature importance analysis was conducted to identify the most influential predictors of stock returns. Since interpretability varies across machine learning models, feature importance was derived using tree-based ensemble methods only, specifically Random Forest and XGBoost.

In the Random Forest model, feature importance was computed using the mean decrease in impurity (Gini importance), which measures how much each variable reduces uncertainty in prediction across decision trees. In the XGBoost model, importance was calculated based on the average gain, which reflects the contribution of each variable to improving model accuracy at each split.

To ensure robustness, feature importance scores from both models were averaged to generate a unified ranking of predictors. The results show that technical indicators dominate predictive power. Relative Strength Index (RSI) recorded the highest importance score (0.21), followed by Moving Average Convergence Divergence (MACD) (0.18). Market return (0.16) and Moving Averages (0.13) also contribute meaningfully to prediction accuracy. Momentum (0.10) and volatility (0.08) provide moderate explanatory power, while trading volume and macroeconomic variables such as inflation, exchange rate, and interest rate exhibit relatively lower importance.

It is important to note that Artificial Neural Networks (ANN) and Support Vector Machines (SVM) do not directly provide interpretable feature importance measures. Therefore, interpretability in this study is based on ensemble learning models, which are more suitable for variable importance analysis in nonlinear financial datasets.

Overall, the results confirm that RSI and MACD are the most influential predictors of stock returns in the Nigerian stock market.

Model Performance

Table 4.4

Performance of Machine Learning Models for Stock Return Prediction (N = 25,000)

Model	MSE	RMSE	MAE	Out-of-Sample R^2
ANN	0.00049	0.0221	0.0170	0.44
SVM	0.00045	0.0212	0.0161	0.48
Random Forest	0.00028	0.0167	0.0129	0.65
XGBoost	0.00024	0.0155	0.0118	0.71

Note. Models were evaluated using out-of-sample test data derived from a final dataset of N = 25,000 firm-day observations (2015–2024). The dataset was obtained from the Nigerian Exchange Group (NGX) and complemented with macroeconomic data from the Central Bank of Nigeria (CBN) and other publicly available financial databases. The sample was split into 70% training and 30% testing sets. All performance metrics were computed using Python (scikit-learn, XGBoost, and Keras). Source: Author's computation (2026).

The model evaluation indicates that all four machine learning techniques are capable of predicting stock returns, but their effectiveness varies. Among the models, XGBoost performs the best, achieving the lowest errors (MSE = 0.00024, RMSE = 0.0155, MAE = 0.0118) and the highest out-of-sample $R^2$of 0.71. This demonstrates its superior ability to capture complex, nonlinear relationships in the data.

Random Forest also performs strongly, with an out-of-sample $R^2$of 0.65 and relatively low error values, highlighting the effectiveness of ensemble methods in improving predictive accuracy. Support Vector Machine (SVM) and Artificial Neural Network (ANN) show moderate performance, with R^2values of 0.48 and 0.44, respectively. While they capture nonlinear patterns, their predictive accuracy is lower than that of ensemble models.Overall, the results suggest that ensemble-based models, particularly XGBoost, are the most effective for predicting stock returns in the Nigerian market, while traditional models like ANN and SVM provide reasonable but less accurate predictions.

Integration of Tables and Figures

The study's results are visually and quantitatively supported by several tables and figures. Descriptive statistics (Table 4.1) illustrate stock return volatility and variations in technical indicators, while trend figures for RSI, MACD, and Moving Averages highlight periods of overbought and oversold conditions, reinforcing their predictive importance. Correlation patterns (Table 4.2) demonstrate the relationships among variables, and feature importance scores (Table 4.3) identify the most influential predictors. Model performance metrics (Table 4.4) provide clear evidence of the superiority of ensemble models, particularly XGBoost, in capturing complex relationships and producing accurate predictions.

Discussion of Findings

The findings of this study reveal several important insights into the effectiveness of machine learning techniques in predicting stock returns in the Nigerian market. Consistent with the methodology outlined in Chapter Three, the results demonstrate that technical indicators specifically Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Moving Averages, and Momentum play a central role in enhancing prediction accuracy. Among these, RSI and MACD emerged as the most influential predictors, reflecting the importance of momentum and trend signals in forecasting stock price movements. Moving Averages and Momentum also contributed meaningfully, highlighting the value of capturing both short-term trends and directional price changes in predictive models.

In addition to technical indicators, market variables such as trading volume, market index returns, and volatility were found to positively influence model performance. These variables provide insights into overall market behavior and liquidity conditions, complementing technical indicators and improving the models' ability to capture the dynamics of stock price movements. Conversely, macroeconomic variables, including interest rate, inflation rate, and exchange rate, were found to have a comparatively smaller impact on short-term stock return prediction. While these variables are important in explaining long-term trends and structural shifts in the market, their influence on daily stock return predictions is limited, likely due to the rapid fluctuations and noise inherent in short-term market data.

A notable finding is that the ensemble machine learning models, Random Forest and XGBoost, achieved better predictive performance than the ANN and SVM models based on the evaluation metrics employed in this study. These ensemble models recorded higher out-of-sample R^2 values and lower prediction errors, indicating a greater ability to capture complex nonlinear relationships in the data. Their strength lies in effectively combining information from multiple

predictors, allowing them to integrate signals from technical indicators, market variables, and macroeconomic factors while reducing overfitting. XGBoost produced the strongest forecasting results among the models evaluated. However, it is important to note that the comparison is based on observed differences in forecasting accuracy measures rather than formal statistical tests of predictive superiority. Consequently, the findings indicate comparatively better predictive performance within the study sample but do not establish whether the observed differences are statistically significant.

Overall, the results underscore that the combination of technical indicators and advanced ensemble machine learning techniques offers the most effective framework for predicting stock returns in the Nigerian market, while macroeconomic factors, though relevant, are secondary in short-term forecasting. These findings provide strong evidence for the practical application of machine learning models in emerging markets characterized by volatility and structural complexity. The ability of the models to maintain predictive performance across different periods of the sample further supports the suitability of flexible machine learning techniques in a market environment characterized by evolving conditions and structural complexity.

To address the relative contribution of macroeconomic variables, it is important to clarify that their predictive relevance was assessed within a multivariate machine learning framework that included both technical indicators, market variables, and macroeconomic factors simultaneously. The feature importance results derived from the Random Forest and XGBoost models indicate that macroeconomic variables (interest rate, inflation rate, and exchange rate) consistently ranked lower in predictive contribution compared to technical indicators such as RSI and MACD, as well as key market variables. This suggests that, within the context of short-term (daily) stock return prediction, macroeconomic variables provide weaker marginal explanatory power when combined with high-frequency technical and market data.

This finding does not imply that macroeconomic variables are irrelevant to stock price formation. Rather, it reflects their lower short-term sensitivity relative to technical indicators, which tend to capture immediate market reactions and investor behavior more effectively in high-frequency data environments. Furthermore, machine learning models used in this study evaluate variable importance in a joint predictive setting, meaning that each variable's contribution is assessed conditional on the presence of all other variables. In such settings, macroeconomic variables may exhibit lower importance due to multicollinearity effects, slower-moving dynamics, and their stronger relevance for medium- to long-term horizons rather than daily prediction tasks.

Although a formal ablation study (separate models with and without macroeconomic variables) was not conducted, the consistency of results across both Random Forest and XGBoost models provides robust evidence that macroeconomic variables play a secondary role in short-term predictive performance within the Nigerian stock market context. Future research could extend this analysis by estimating separate model specifications with and without macroeconomic variables to further isolate their marginal contribution.

While the findings are robust, several limitations should be noted. The study relies exclusively on secondary data, which may contain inconsistencies. The focus on daily returns limits generalization to long-term forecasting, and the purposive sample of firms may not represent

smaller or less actively traded companies. Machine learning models are sensitive to hyperparameter tuning, and overfitting remains a potential risk. Finally, macroeconomic variables' limited effect on short-term returns suggests caution when applying these models beyond the daily horizon.

This study evaluates machine learning models based on statistical forecasting accuracy measures such as MSE, RMSE, MAE, and out-of-sample R^2. It does not extend to portfolio optimization, trading strategy development, or profitability analysis. Therefore, while the XGBoost model exhibits superior predictive performance, its effectiveness in generating economically profitable trading strategies was not assessed.Future research could integrate prediction models into a trading framework to evaluate economic value using risk-adjusted returns, Sharpe ratios, and transaction-cost-adjusted back testing.

An additional limitation relates to the comparison of predictive performance across machine learning models. Model evaluation was based on standard forecasting accuracy measures, including MSE, RMSE, MAE, and out-of-sample R^2. Although these metrics indicate differences in predictive performance among the models, formal statistical tests of forecast accuracy differences, such as the Diebold–Mariano test or related predictive accuracy tests, were not conducted. Consequently, the observed performance advantages of certain models should be interpreted as empirical differences in forecasting accuracy rather than definitive evidence of statistically significant superiority.

5.0 Conclusion

This study investigated the effectiveness of machine learning techniques in predicting stock returns in the Nigerian stock market, incorporating a combination of technical indicators, market variables, and macroeconomic factors. The findings indicate that technical indicators specifically Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Moving Averages, and Momentum are the most significant predictors of stock returns. Among these, RSI and MACD exhibited the greatest influence, highlighting the importance of trend-following and momentum signals in short-term stock price forecasting.

Market variables, including trading volume, market returns, and volatility, were also found to meaningfully enhance predictive accuracy, although their contribution was slightly lower compared to that of technical indicators. In contrast, macroeconomic variables such as interest rate, inflation, and exchange rate played a secondary role in short-term prediction. While these factors are important for understanding long-term market trends, their impact on daily or short-term stock returns was limited.

The study further revealed that ensemble machine learning models, particularly XGBoost and Random Forest, achieved better predictive performance than ANN and SVM based on the forecasting accuracy measures employed. XGBoost recorded the lowest prediction errors and the highest out-of-sample R^2 among the models evaluated, indicating strong predictive capability within the sample considered. However, these findings should be interpreted as evidence of comparatively better forecasting performance rather than conclusive proof of statistical superiority, since formal predictive accuracy tests were not conducted.

The superior performance of these ensemble methods can be attributed to their ability to capture nonlinear relationships, effectively integrate multiple predictors, and generalize well to unseen data, resulting in higher predictive accuracy.

Overall, the results demonstrate that combining technical indicators with advanced machine learning models provides a robust and reliable framework for forecasting stock returns in the Nigerian market, which is characterized by high volatility, fluctuating macroeconomic conditions, and dynamic trading behavior.

Recommendations

Based on the findings, the study makes the following recommendations:

1. **For Investors and Traders:** Based on the forecasting accuracy measures reported in this study, particularly the lower error metrics and higher out-of-sample R^2 values, the XGBoost model demonstrated comparatively stronger predictive performance than the other machine learning models evaluated. Consequently, it may be considered a useful analytical tool for stock return forecasting in the Nigerian market. Nevertheless, the findings should be interpreted with caution because formal statistical tests of predictive accuracy differences were not undertaken.
2. **For Financial Analysts and Portfolio Managers:** Incorporate both market variables and technical indicators in predictive models to capture a broader range of market signals. Be cautious in relying solely on macroeconomic variables for short-term forecasting, as their influence is comparatively smaller.
3. **For Regulators and Policy Makers:** Understanding the role of technical indicators and market dynamics can inform policy decisions aimed at enhancing market efficiency and transparency.

References

Abdullahi, S. I. (2024). Impact of COVID-19 pandemic on performance of Nigerian stock market. *CBN Journal of Applied Statistics, 15*(1).

Adejumo, O. A. (2020). Markov regime-switching autoregressive model of stock market price behaviour in Nigeria. *CBN Journal of Applied Statistics, 12*(1).

Ayyildiz, E., & Iskenderoglu, O. (2024). A comparative analysis of machine learning algorithms in stock market prediction. *Heliyon, 10*(1), e2401543. https://doi.org/10.1016/j.heliyon.2024.e2401543

Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications, 42*(20), 7046–7056. https://doi.org/10.1016/j.eswa.2015.05.013

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (2020). *The econometrics of financial markets* (Updated ed.). Princeton University Press.

Charles, A., Darné, O., & Kim, J. H. (2021). Time-varying market efficiency: Evidence from global stock markets. *Economic Modelling, 95*, 199–210. https://doi.org/10.1016/j.econmod.2020.02.010

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*(2), 383-417. https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research, 270*(2), 654-669. https://doi.org/10.1016/j.ejor.2017.11.054

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies, 33*(5), 2223-2273. https://doi.org/10.1093/rfs/hhaa009

Hanauer, M. X., & Kalsbach, B. (2023). Machine learning and the cross-section of emerging market stock returns. *Emerging Markets Review, 55*, 100903. https://doi.org/10.1016/j.ememar.2023.100903

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2020). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications, 150*, 113-129. https://doi.org/10.1016/j.eswa.2020.113129