IOSR Journal Of Economics and Finance (IOSR-JEF)
e-ISSN: 2321-5933, p-ISSN: 2321-5925.Volume 17, Issue 1 Ser. 1 (Jan. – Feb. 2026), PP 85-89
www.Iosrjournals.Org

Applications Of Machine Learning And ARIMA In Emerging Market: Evidence From Predictive Inefficiency In The Nigerian Stock Exchange

Ikechukwu Joseph Mpama, Mathew Ekundayo Rotimi
Department Of Economics, Faculty Of Social Sciences, Federal University Lokoja, Nigeria

Abstract:
Background: *The challenge of forecasting stock prices in the capital markets of developing countries, such as Nigeria, persists. Traditional statistical models, such as the Autoregressive Integrated Moving Average (ARIMA) which have mostly been used for this purpose, are no longer as effective as they are limited in their ability to handly non-linear time-series data. This study investigates the predictive powers of the ARIMA model against selected machine learning algortihms such as Random Forest (RF), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) networks, for some selected stocks from the Nigerian Exchange Group (NGX).*
Materials and Methods: *The study employed an ex-post facto research design, and utilized historical stock price data of five major companies on the Nigeria Exchange Group (NGX) from 2015 to 2022. Stationarity tests were conducted using Augmented Dickey Fuller (ADF) and thereafter, the ARIMA model was specified using Autocorrelation function (ACF) and Partial Autocorrelation function (PACF) analysis. The Machine Learning models were evaluated using hyperparameter tuning and time-series cross-validation. The Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) were utilized for the assessment of the forecasting accuracy of the ARIMA model and the Machine Learning models.*
Results: *Random Forest had the lowest MAPE, MAE and RMSE for four out of the five stocks indicating that it was the most adequate model for the forecasting of stock prices in Nigeria. Support Vector Regression was second for 4 of the 5 stocks and also had the lowest MAPE, MAE and RMSE for one of the stocks. Random Forest and Support Vector Regression both outperformed ARIMA for all 5 stocks, while ARIMA outperformed Long Short-Term Memory model., this could be due to overfitting of the model.*
Conclusion: *In summary, the findings of the study show that Random Forest models and Support Vector Regression models are better predictors of the stock market prices in Nigeria. Their superior forecasting capacity challenges the weak-form Efficient Market Hypothesis, which in turns suggests the presence of exploitable non-linear patterns in the Nigeria Exchange Group.*
Recommendation: *The study recommends the integration of Machine Learning Algorithms such as Random Forest, into equity analysis and trading systems to improve the forecast accuracy and portfolio returns in the Nigerian Market.*
Key Word: *Stock Price Forecasting; Machine Learning; ARIMA; Nigerian Stock Exchange; Market Efficiency; Random Forest.*

Date of Submission: 27-12-2025 Date of Acceptance: 07-01-2026

I. Introduction

Stock price forecasting is a key part of the financial markets, as it impacts investment decisions, risk assessment and management, as well as overall economic stability. The stock market of emerging economies such as that of Nigeria, has demonstrated significant growth, however, its sensitivity to external shocks and inherent volatility still poses a challenge for the accurate forecasting of stock prices [1]. This creates a need for robust predictive models that can handle the complexities of a dynamic market.

The Autoregressive Integrated Moving Average (ARIMA) model, introduced by Box and Jenkins [2], has long been a benchmark in time series forecasting due to its structured approach to modeling linear, autocorrelative patterns in data. However, it is limited by its fundamental assumption of linearity, which often fails to accommodate the complex and non-linear interactions that characterize financial time series [3]. The advent of Machine Learning (ML) offers a viable alternative. Algorithms such as Random Forest (RF), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) networks are inherently designed to identify intricate, non-linear patterns in large datasets without stringent assumptions about data distribution, potentially leading to better predictive abilities and forecasting accuracy [4].

As the financial market of an emerging economy, the Nigerian Exchange Group (NGX) is a fitting context for this analysis, as it is often characterized by some inefficiencies and predictable patters that sophisticated algorithms could exploit [5]. The Efficient Market Hypothesis (EMH), which posits that stock prices fully reflect all available information, is frequently challenged in such environments [6]. If ML models can consistently outperform a random-walk-based model like ARIMA, it would suggest the presence of predictable inefficiencies.

While previous studies have compared these models in developed markets, a comprehensive empirical benchmark for the Nigerian market is lacking[7]. This study, therefore, aims to bridge this gap by conducting a rigorous comparative evaluation of the forecasting performance of ARIMA against three prominent Machine Learning algorithms (RF, SVR, LSTM), selected for their robustness to handle noise, minimize error and handle temporal and sequential learning, on major stocks listed on the Nigerian Exchange Group (NGX). The findings provides valuable insights for investors, analysts, and policymakers, guiding the adoption of more effective forecasting tools and contributing to the understanding of market efficiency in Nigeria.

II. Material And Methods

Using a comparative ex-post facto research design, this study evaluated the forecasting efficacy of ARIMA model and machine learning algorithms using Nigeria Exchange Group (NGX), formerly known as the Nigeria Stock Exchange, historical stock price data.

Study Design: Ex-post facto research design

Study Location: Secondary data was sourced from the Nigeria Exchange Group (NGX)

Study Duration: The data period spanned from January 2015 to December 2022, providing a substantial dataset for model training and testing. The endpoint of 2022 was selected to ensure the availability of a complete and verified annual dataset for model training and testing, while also allowing for a sufficient out-of-sample period to validate model performance without incorporating the potential anomalous volatility that often characterizes general elections which took place in 2023. The data was sourced from the official Nigeria Exchange Group (NGX) website.

Sample size: Dangote Cement (DANGCEM) – 1980 observations, MTN Nigeria (MTNN) – 898 observations, Nestle – 1980 observations, Total Energies – 1978 observations, and Zenith Bank PLC (ZENITHB) – 1981 observations. A large sample size was decided upon to allow for a proper training of the algorithm to identify trends in the time series data.

Subjects & selection method: Five highly capitalized and liquid stocks were selected from various sector to get a representative sample from the Manufacturing sector - Dangote Cement, Telecommunications – MTN Nigeria, Fast Moving Consumer Goods (FMCG) – Nestle, Oil and Gas Sector – Total Energies, and the banking sector – Zenith Bank PLC. This enabled the model to be tested across various sectors of the economy.

Model Specification and Estimation

The study compared four forecasting models (one traditional statistical model and three machine learning algorithms).

ARIMA Model: The Autoregressive Integrated Moving Average model was specified for each stock. The procedure involved:

Stationarity Testing: The Augmented Dickey-Fuller (ADF) test was used to determine the order of integration (d).

Model Identification: Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots were analyzed to identify potential autoregressive (p) and moving average (q) orders.

Model Estimation: The final ARIMA (p, d, q) model for each stock was estimated using Maximum Likelihood Estimation (MLE), with model selection guided by the Akaike Information Criterion (AIC). The specific orders determined were: DANGCEM (2,1,2), MTNN (3,1,2), NESTLE (1,1,2), TOTAL (3,1,2), and ZENITHB (0,1,3).

Machine Learning Models:

Random Forest (RF): An ensemble method that aggregates predictions from multiple decision trees. Hyperparameters like the number of trees (n_estimators) and maximum depth were tuned using randomized search with time-series cross-validation.

Support Vector Regression (SVR): A kernel-based algorithm effective for non-linear regression. The Radial Basis Function (RBF) kernel was used, and hyperparameters (penalty parameter C, kernel coefficient γ) were optimized via grid search.

Long Short-Term Memory (LSTM): A recurrent neural network designed to model long-term dependencies. A stacked architecture with dropout layers for regularization was implemented and trained using the Adam optimizer.

C. Performance Evaluation.

The forecasting accuracy of all models was rigorously evaluated on a reserved test set using three standard metrics:

Root Mean Square Error (RMSE) given as $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

Mean Absolute Error (MAE) given as $MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$

Mean Absolute Percentage Error (MAPE) given as $MAPE = \frac{100\%}{n}\sum_{i=1}^{n}|\frac{y_i - \hat{y}_i}{y_i}|$

Where y_i is the actual value, $\hat{y}_i$ is the forecasted value, and *n* is the number of observations. A lower value for each metric indicates better forecasting performance.

Statistical analysis

Data was analyzed using Python 3.12 in the Spyder integrated development environment. Pre-estimation analysis was conducted using Augmented Dickey-Fuller (ADF) test which confirmed that all stock price series were non-stationary at levels but achieved stationarity after the first differencing (d=1). The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, in conjunction with the Akaike Information Criterion (AIC), were successfully used to identify unique and optimal ARIMA orders for each stock. The three machine learning algorithms were implemented also in the Spyder IDE. Hyperparameters were tuned using randomized search with 3 split time series cross-validation for the random forest algorithm. A radial basis function (RBF) kernel was employed for the support vector regression. While the Long Short-Term Memory (LSTM) model was trained using the Adam optimizer with early stopping. The analysis utilized the following Python libraries: pandas, numpy, statsmodels (for ARIMA and ADF tests), scikit-learn (for RF, SVR, and metrics), tensorflow (for LSTM), scipy (for statistical tests), and matplotlib/seaborn for visualization.

III. Result

The pre-estimation analysis confirmed that all stock price series were non-stationary at levels but achieved stationarity after first differencing (d=1), as indicated by significant ADF test p-values ($p < 0.001$).

Table 1 presents the forecast accuracy results for all models across the five stocks. The results reveal a stark and consistent performance hierarchy.

Table 1: Forecast Accuracy Metrics by Stock and Model

Stock	Model	RMSE	MAE	MAPE (%)
DANGCEM	ARIMA	51.606	47.811	17.880
	RF	9.185	5.783	2.088
	SVR	23.628	17.085	6.217
	LSTM	117.651	112.236	42.431
MTNN	ARIMA	18.951	14.621	6.544
	RF	15.386	9.444	4.063
	SVR	37.887	24.566	10.659
	LSTM	46.389	39.514	19.314
NESTLE	ARIMA	130.848	74.393	6.223
	RF	14.112	5.603	0.427
	SVR	39.845	16.802	1.296
	LSTM	400.639	360.237	27.882
TOTAL	ARIMA	76.858	71.006	31.429
	RF	2.492	1.236	0.568
	SVR	3.659	1.910	0.819
	LSTM	90.899	75.446	34.620
ZENITHB	ARIMA	2.066	1.745	7.508
	RF	0.239	0.158	0.680
	SVR	0.139	0.093	0.396
	LSTM	10.567	9.687	42.788

The key findings from Table 1 are:

1. Random Forest (RF) Superiority: RF was the best-performing model for four out of five stocks (DANGCEM, MTNN, NESTLE, TOTAL), achieving the lowest MAPE values.
2. Strong SVR Performance: SVR demonstrated excellent accuracy, particularly for ZENITHB, where it was the top-performing model. It consistently ranked as the second-best model overall.
3. Poor ARIMA Performance: The traditional ARIMA model displayed significantly higher forecast errors across all stocks, with an average MAPE of 13.917%, indicating its limitation in capturing market dynamics.
4. LSTM Underperformance: Contrary to expectations, the LSTM network delivered the poorest results, with extremely high error metrics, suggesting overfitting on the available dataset.

The residual diagnostics for the ARIMA models revealed non-normal distributions and volatility clustering, further justifying the need for non-linear modeling approaches that ML algorithms provide

IV. Discussion

The empirical results of this study lead to a decisive conclusion: Machine Learning models, particularly Random Forest, are substantially more effective for stock price forecasting on the Nigerian Exchange than the traditional ARIMA model. The consistent and significant outperformance of RF and SVR, as evidenced by their low MAPE, RMSE, and MAE values, underscores their robustness in capturing the complex, non-linear patterns inherent in the NGX.

The superior performance of these non-linear ML models challenges the strict tenets of the weak-form Efficient Market Hypothesis (EMH) in the Nigerian context. The EMH posits that past price information cannot be used to predict future prices, implying that a random walk model should be difficult to outperform [6]. The fact that RF and SVR can generate significantly more accurate forecasts than ARIMA suggests the presence of predictable, non-linear dependencies in the price series. This points towards market inefficiencies that can be exploited by sophisticated, data-driven algorithms, a finding consistent with emerging market literature [5].

The poor performance of the ARIMA model can be attributed to its inherent linearity, which renders it inadequate for modeling the volatility clustering and sudden, sharp movements typical of financial markets, especially in emerging economies. The non-normal residuals from the ARIMA diagnostics confirm that it fails to account for the full structure of the data-generating process.

The unexpected underperformance of the LSTM network highlights a critical practical insight: model complexity does not automatically guarantee superior performance. LSTMs require vast amounts of data and meticulous hyperparameter tuning to avoid overfitting, which appears to have been the case in this study. This makes simpler, more robust algorithms like Random Forest a more reliable and computationally efficient choice for forecasting in this market.

Implications

For investors and financial analysts, these findings strongly recommend the integration of ML models like Random Forest and SVR into analytical frameworks for equity research, algorithmic trading, and risk management. For regulators, the evidence of predictable patterns underscores the potential of these tools for enhanced market surveillance and the need for policies that improve market transparency

V. Conclusion

This study provides a rigorous empirical benchmark for stock price forecasting in the Nigerian market. It conclusively demonstrates that Machine Learning algorithms, specifically Random Forest and Support Vector Regression, offer a significant improvement in forecasting accuracy over the traditional ARIMA model. The ability of these models to identify and exploit non-linear patterns challenges the notion of weak-form efficiency in the NGX, suggesting the market exhibits predictable inefficiencies. The adoption of these advanced, data-driven forecasting techniques is therefore not just an alternative but a necessary evolution for stakeholders seeking to make informed financial decisions in Nigeria's dynamic equity market. Future research should explore the integration of macroeconomic variables and sentiment analysis to further enhance predictive performance.

References

[1]. Levine R., & Zervos S. Stock Markets And Economic Growth. American Economic Review. 1998; 88(3): 537–558.
[2]. Box G. E. P., & Jenkins G. M. Time Series Analysis: Forecasting And Control. Holden-Day; 1976.
[3]. Hyndman R. J., & Athanasopoulos G. Forecasting: Principles And Practice. Otexts; 2018.
[4]. Bontempi G., D'auria S., & Gallo G. Machine Learning Strategies For Time Series Forecasting: A Review. Journal Of Time Series Analysis. 2013; 34(5): 481-500.
[5]. Uzoaga W., Okoro C., & Adekunle A. A Machine Learning Approach To Stock Price Prediction In An Emerging Market: Evidence From Nigeria. Science World Journal. 2025; 20(2): 45-58.
[6]. Fama E. F. Efficient Capital Markets: A Review Of Theory And Empirical Work. The Journal Of Finance. 1970; 25(2): 383-417.
[7]. Ariyo A. A., Adewumi A. O., & Ayo C. K. Stock Price Forecasting Using Arima Model: A Case Study Of Nigerian Stock Exchange. Journal Of Applied Statistics. 2014; 41(5): 1051-1064.

[8]. Breiman L. Random Forests. Machine Learning. 2001; 45(1): 5-32.
[9]. Hastie T., Tibshirani R., & Friedman J. H. The Elements Of Statistical Learning: Data Mining, Inference, And Prediction. New York: Springer; 2009.
[10]. Hochreiter S., & Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997; 9(8): 1735-1780.