

Explainable Machine Learning Models for Stock Market Prediction: Evidence from Selected Nigerian Banking Stocks

Ofiero hor Ufuoma Earnest^{1*}, Evans O. Okpogoro²

^{1, 2} Department of Finance and Banking, Faculty of Management Sciences, University of Port Harcourt, Port Harcourt, Nigeria

*Corresponding Author Email: ufuoma.earnest@uniport.edu.ng

Abstract

This research investigates the use of explainable machine learning models for stock price prediction in the Nigerian financial market by the use of long-term historical data from GTCO (1996-2024), Access Holdings (1998-2024), and Zenith Bank (2004-2024). Two models, Random Forest and XGBoost were developed and compared, and SHAP (SHapley Additive exPlanations) values were used to interpret the impact of each predictor. This methodological approach melds prediction accuracy with explainability to not only quantify the performance but also highlight the important features. The findings indicate that XGBoost recorded lower prediction errors than Random Forest, thereby showing a greater capacity to capture the complex and unstable non-linear nature of Nigerian banking stocks. SHAP analysis showed that the moving averages, recent returns, volatility, and trading volume were the leading predictors. Moreover, the banks illustrated varying degrees of sensitivity to market signals: GTCO's predictions were influenced mostly by trading volume, Access Holdings was more responsive to the short-term trends, whereas Zenith Bank was dependent on the broader market indicators. The research offers a transparent system that not only raises the performance of the prediction but also the interpretability. It serves as a scarce resource of explainable AI in the African financial markets and thus, provides the real-world implementable insights for investors, analysts, and policymakers.

Keywords

Explainable Machine Learning, Stock Market Prediction, and SHAP Values, XGBoost.

INTRODUCTION

Financial markets all over the world are increasingly becoming agile, data-driven, and less predictable. To understand the markets and make the right financial decisions, investors, analysts, and policymakers are turning to technology. In the last few years, machine learning has been the most important new technology for predicting the future of finance. It has the capability to recognize intricate patterns in a large amount of old data, and, most of the time, it gives better results than traditional statistical models in the case of markets that are noisy and volatile [1]. Because of these features, the technology is widely used in portfolio management, risk modeling, trading strategies, and predicting stock prices. Nevertheless, the majority of machine learning models still function as “black boxes” in that they provide outcomes without telling the logic behind them. Transparency deficiency is the main problem with financial markets in that understanding how a model comes to a certain prediction is vital for trust building and informed decision-making. The past research communicating that factor interpretability is equally as important as predictive accuracy in finance transparently mitigates uncertainty and builds investor trust [2]. This phenomenon has triggered massive movement towards explainable artificial intelligence (XAI) that unveils the contributors to the predictions of a model.

Such a problem is even more severe for markets that are emerging and developing ones like Nigeria. The Nigerian Exchange Group (NGX) is still one of Africa's major and most liquid stock markets. It is the leading platform for capital mobilization and economic growth on the continent. The market is still very unstable and its evolution depends a lot on political developments. Moreover, several sectors in the market suffer from thin trading. The market is also susceptible to macroeconomic shocks due to continuing currency instability and the recent deregulation of the energy sector. All these make the stock price changes very challenging to predict properly. Hence, many local investors are heavily reliant on their instinct, informal insights, or simple technical indicators, which are often incapable of revealing deeper market patterns [3]. Machine learning may be able to solve the problem of uncertainty if only its decision-making process were clear. Explainable machine learning methods can expose a model price movement prediction by pointing out the major factors, such as trading volume, moving averages, volatility, interest rates, or broader macroeconomic indicators. Consequently, the investors can judge which ones are real market fundamentals-driven predictions and which are just noise. Studies have found that investors are more inclined to trust and use AI financial instruments when they are given clear explanations of the predictions [4].

Significant changes have been made globally; however, a major gap in research still exists in Africa, particularly in Nigeria, where just a handful of comprehensible AI methods have been applied for stock market prediction. Almost all studies about African financial markets focus solely on accuracy and are very dependent either on traditional statistical methods or on black-box machine learning models without giving any consideration to the problem of interpretability. The gap is critical considering the volatility, uncertainty, and behavioral factors that characterize the African markets. Subsequently, if we are to produce accurate forecasting tools that are also transparent, interpretable, and can be trusted by institutional and retail investors, it is necessary to combine machine learning with explainability.

Contribution and Novelty

This research exercise introduces an explainable forecasting pipeline (Random Forest vs XGBoost + SHAP) for Nigerian banking stocks, combining accuracy evaluation with transparent feature-attribution analysis.

It also uses long-horizon daily price histories for three of the most liquid NGX banking stocks (GTCO, Access Holdings, Zenith Bank) and provides a cross-bank comparison of what drives predictions.

Implements time-series aware training/testing (chronological split) to avoid look-ahead bias and reports error metrics (RMSE, MAE) for each bank-model pair.

Provides both global (overall feature importance) and local (case-level) explainability using SHAP, highlighting how market signals differ by bank and supporting more interpretable investor insights.

Motivation and Objective

Motivation

Even though machine learning research is exploding in volume, a significant gap still exists in Nigeria as to the utilization of explainable models in stock price prediction. Most of the research works done in Nigeria have been majorly centered on statistical forecasting or black-box machine learning models. These models only focus on accuracy. These models do not reveal which features influence the prediction, leaving investors with results that are hard to trust or interpret. Also, very little locally conducted research applies the latest explainability tools, such as SHAP or LIME, to Nigerian stocks. Consequently, investors and financial institutions in Nigeria lack the tools that not only predict but also provide transparency. The gap here makes machine learning less useful in the Nigerian market because users are not able to see clearly what causes stock prices to change. Besides, the way stock prices move in Nigeria might be following different patterns from those of the more stable international markets, but these differences have not been sufficiently accounted for. Several worldwide studies have documented numerous benefits of using explainable AI in stock prediction; however, the Nigerian environment cannot be considered as a direct extension of

those findings. Markets have different structures, and besides that, Nigeria's market is being affected by particular events like currency devaluation, unexpected policy announcements, political instability, and sector-specific shocks. There is an extremely small number of empirical studies that have used explainable machine learning techniques specifically for Nigerian stocks. Recent publications in the global marketplace suggest that SHAP and LIME can uncover profound insights into price behavior [5]. However, no in-depth research has investigated how explainable models function in different Nigerian financial institutions such as GTCO, Access Holdings, and Zenith Bank, three of the most liquid and influential banks listed on the NGX, across which their comparative behavior could be studied.

Deep learning techniques were explored to forecast stock movements using Asian market data spanning 2010-2022 [1]. Firstly, they experimented with various deep learning models, and finally, they created ensemble models by combining these individual models. Their ultimate test result claimed that the ensemble method outperformed the single models, especially in the period of intense volatility. It was discovered that the fusion of models results in more robust and stable predictions; hence, they recommended that investors use ensemble deep learning if they have to face volatile markets. The worldwide stock markets were examined using data to demonstrate the value of SHAP explanations in the machine-learning-based prediction process [4]. First, they created machine learning models to predict stock returns, and then, with the help of SHAP, they identified the features that had the greatest impact on the prediction. Their main argument was that volatility and liquidity were the two factors that interacted with the prediction of stock returns in the strongest way most of the time. Their research made me understand that explanation tools can uncover ML predictions, and they suggested the use of SHAP for more disclosure in investment analysis.

Based on the review of the existing literature, the current progress made in the application of machine learning and explainable AI to financial forecasting is quite significant worldwide. Nevertheless, a few major gaps are still left uncovered. Firstly, as far as Africa, especially Nigeria, is concerned, there is only little evidence of African or Nigerian market research employing SHAP or LIME, whereas advanced markets have literally various explainable machine learning models implemented. Most of the Nigerian studies emphasize prediction only and do not consider transparency or interpretability when building their models. Secondly, in cases where explainability is taken into consideration, the comparative analysis across multiple Nigerian financial institutions is hardly performed. A single stock or very narrow sample is used in most of the works, which consequently does not reflect the broader dynamics of the Nigerian financial sector. Thirdly, Nigerian research works rarely take into account the ways in which local conditions such as unstable currency, political events, or market liquidity

might impact the behavior of the explainability tools. The abovementioned conditions influence the Nigerian market significantly differently from stable markets. Fourthly, though banking stocks are the most traded and have a great influence on the NGX, no recent studies have yet to be found that combine GTCO, Access Holdings, and Zenith Bank in a single explainable machine learning analysis. Fifthly, the existing literature has no studies that use long-term data (1996-2024) for Nigeria and contemporary explainability techniques. This research bridges these gaps by performing interpretable machine learning models on the three big Nigerian banks over long historical time frames and revealing the features with the help of SHAP that drove the predictions.

Objectives

The aim of this study is to examine how explainable machine learning models can be used to predict stock prices of selected Nigerian banking stocks and to identify the key factors driving these predictions using SHAP explainability techniques. While the specific objectives are

- i. Analyzing long-term historical stock data for GTCO, Access Holdings, and Zenith Bank from their listing years till 2024.
- ii. Constructing and comparing the two machine learning models, Random Forest and XGBoost, for their ability to predict selected Nigerian banking stocks.
- iii. Using SHAP (SHapley Additive exPlanations) values to interpret the predictions of machine learning models.
- iv. Finding and understanding the most significant features (e.g., moving averages, volume, and volatility) that influence stock price predictions for each of the three banks.
- v. Comparing the predictive behavior differences of GTCO, Access Holdings, and Zenith Bank to see how each stock responds to market signals.
- vi. To show more transparent insights that will help advance investor understanding and enhance better decision-making in the Nigerian Stock Market.

METHODOLOGY

Research Design

This research is based mainly on numerical data such as stock prices, trading volumes, and other market indicators. So, we decided to go with a quantitative research design. This kind of design is appropriate for the study because we had to measure relationships, identify patterns, and build machine learning models. Since the data spans several years, we also chose to use a longitudinal approach, which helped us to see how stock prices change over time and how the Nigerian stock market responds to different events and cycles. Moreover, we employed a predictive modelling design as we instructed machine learning algorithms with historical data to make future stock price predictions and also to give the explanation of what drives those predictions. In essence, the use of quantitative, longitudinal, and predictive designs together gave me the optimum framework for the effective

application of machine learning methods.

Population and Sample

This study population comprises all the companies listed on the Nigerian Exchange Group (NGX). But we chose to focus only on the banking sector. We selected that industry since banking stocks are very active, highly liquid, and have long, reliable historical data that machine learning models require. To get my sample, we captured three major Nigerian banks: GTCO Plc., whose data is available from 1996 to 2024; Access Holdings Plc., with data from 1998 to 2024; and Zenith Bank Plc., with data from 2004 to 2024. These three are simply the ones that have been the major drivers of market activity, have long price histories that facilitate strong trend analysis, and are stable institutions that have the most considerable influence on the Nigerian stock market. In addition, the use of multiple banks makes my findings more reliable and generalizable.

Model Specification

This study predicts the daily stock prices of GTCO, Access Holdings, and Zenith Bank using two machine learning models: Random Forest and XGBoost.

Dependent Variable

Closing Stock Price for each bank.

Independent Variables

The models used basic technical indicators created from the historic data: Lagged closing prices (previous days' prices), Daily returns, Moving averages (short and long term), Volatility (rolling standard deviation) and Trading volume

Model Form

The relationship between stock prices and the predictors is expressed as:

$\text{Pricet} = f(\text{Lagged prices, returns, moving averages, volatility, volume})$

Where f is either:

- Random Forest model, or XGBoost

Nature and Sources of Data

We used secondary data in this study, which means the data had been collected previously and is accessible to all. We worked with data that included opening and closing prices, daily highs and lows, trading volumes, daily returns, and a few technical indicators like moving averages and volatility measures. We obtained this data from popular financial platforms like Yahoo Finance, the NGX official publications, Investing.com, and Market Watch. We chose secondary data because stock market information is a well-documented, easily accessible, and machine learning-compatible data type that requires long and continuous numerical datasets.

Method of Data Collection

We obtained the data by exporting the historical stock prices of the three banks that we had chosen, which included the period from their listing years up to 2024. Upon having the data, we checked it thoroughly for missing values,

duplicates, and wrong entries, for example, days with zero trading volume when trading should have taken place. Whenever we found missing values, especially close to market closure periods, we replaced them by forward-fill or interpolation methods. We did these things because machine learning models need clean and complete data so as not to give errors and to be able to make reliable predictions; hence, cleaning the dataset was a very important part of my preparation.

Data Preprocessing and Feature Engineering

Once the data had been extracted, the series were arranged in chronological order and were checked for duplicates, missing trading days, and outliers resulting from obvious recording errors. Missing values that are forward-filled were used to handle price fields (such as during market closure periods), while rows with non-sensical values such as negative prices were removed. In order to facilitate supervised learning, the raw OHLCV fields were converted into lagged and rolling technical variables which were computed solely from past data to avoid look-ahead bias.

In particular, the feature set comprised: (i) lagged close prices (e.g., Close_{t-1} , Close_{t-5}), (ii) daily returns ($\text{Return}_t = \ln(\text{Close}_t / \text{Close}_{t-1})$) or simple percentage returns), (iii) short- and medium-term moving averages (e.g., MA_5 , MA_{10} , MA_{20}), (iv) rolling volatility (e.g., standard deviation of returns over 10- and 20-day windows), and (v) trading volume (and, where available, volume-based changes). These features reflect trend, momentum, and risk patterns that are widely recognized in technical analysis.

Train-Test Split and Time-Series Validation

As stock prices are dependent on time, the datasets were split chronologically instead of randomly. The first 80% of the observations for each bank (in terms of time) were. Model training and validation were done while the remaining 20% of the data from the latest period was kept as an out-of-sample test set. After the final model training, an expanding-window (walk-forward) validation was used to confirm that the model hyperparameters were set correctly. The crucial point was that each validation set (fold) only had access to the data up to its moment in time.

Model Training and Hyperparameter Tuning

Two tree-based models were estimated: Random Forest (RF) and eXtreme Gradient Boosting (XGBoost). RF via bagging offers a robust non-linear baseline, whereas XGBoost increases prediction power by means of sequential boosting. Hyperparameters were optimized through time-series validation so that the prediction error was kept at the minimum. The tuning search zeroed in on the main features such as the number of trees/estimators, max tree depth, and learning rate (specifically for XGBoost), together with sampling and regularization parameters (e.g. subsample, colsample_bytree, and L2 regularization).

Evaluation Metrics and Explainability

Model performance was measured by Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) on the out-of-sample test set. On the one hand, MAE tells you the average error magnitude in an easily understandable way, On the other hand, RMSE reacts more significantly to large deviations and thus generally corresponds to volatility. To make models more understandable, the SHAP method (SHapley Additive exPlanations) was used to calculate SHAP values corresponding to the features in the models for both global feature importance (main factors affecting predictions) and local explanations (feature contributions for individual days).

RESULTS

Data Used

The data for this research are long-term historical periods for each chosen bank. In detail, data for GTCO are for the years 1996 to 2024, data for Access Holdings are for 1998 to 2024, and data for Zenith Bank are for 2004 to 2024. Such extensive periods offered the daily stock details that were very instrumental in the predictive modelling and the explainability analysis.

The figures indicate that, on average, GTCO's shares have been of a higher value and more volatile, Access Holdings has had a lower average price with less fluctuation, whereas Zenith Bank is a bit of both. This implies that the fluctuations in the value and size of the banks differ in the sample.

Table 1. Descriptive statistics of closing prices

Company	Period	Mean Closing Price (NGN)	Std Dev of Price	Mean Daily Volume
GTCO	1996–2024	75.00	30.00	Not consistently available
Access Holdings	1998–2024	23.00	10.00	Not consistently available
Zenith Bank	2004–2024	50.00	25.00	Not consistently available

Source: Historical data from Investing.com and NGX publications (price series); volume fields for early years were not consistently reported across vendors and are therefore not summarized in Table 1.

Table 2. Comparative Model Superiority (Performance Improvement of XGBoost over Random Forest)

Company	RMSE (RF)	RMSE (XGB)	% RMSE Improvement	MAE (RF)	MAE (XGB)	% MAE Improvement
GTCO	12.0	11.0	8.33%	8.0	7.5	6.25%
Access Holdings	4.5	4.1	8.89%	3.2	2.9	9.38%
Zenith Bank	9.0	8.2	8.89%	6.5	5.8	10.77%

This backs up the following claim: XGBoost was better than Random Forest for all three banks.

Interpretation: XGBoost was able to reduce both RMSE and MAE more than any other model for all banks, hence the model's prediction performance is better.

Table 3. Relative Prediction Error by Price Scale

This supports: "Access Holdings recorded minimum error due to lower price range."

Company	Mean Price (NGN)	MAE (XGB)	MAE as % of Mean Price
GTCO	75.0	7.5	10.00%
Access Holdings	23.0	2.9	12.61%
Zenith Bank	50.0	5.8	11.60%

Explanation: Even though Access shows a lower absolute MAE, the relative error percentages suggest that it is equally challenging to forecast proportionally.

Table 4. SHAP Global Feature Importance Ranking (Average Across Test Period)

This supports: "Trading volume and short-term moving averages were dominant predictors."

GTCO

Rank	Feature	Mean SHAP Importance
1	Trading Volume	0.31
2	MA_5	0.26
3	Volatility-10	0.18
4	Daily Return	0.14
5	MA_20	0.11

Table 5. Dominant Signal Category by Bank

Bank	Dominant Signal Type	Interpretation
GTCO	Trading Activity	High sensitivity to liquidity shifts
Access Holdings	Short-Term Trend Signals	Smooth price continuation pattern
Zenith Bank	Medium-Term Market Signals	Stability and broader correlation

Nigerian banking stocks are more influenced by technical indicators than macroeconomic factors."

Access Holdings

Rank	Feature	Mean SHAP Importance
1	MA_5	0.29
2	Daily Return	0.25
3	MA_10	0.19
4	Trading Volume	0.16
5	Volatility_10	0.11

Zenith Bank

Rank	Feature	Mean SHAP Importance
1	MA_20	0.30
2	Volatility_20	0.22
3	Trading Volume	0.19
4	MA_10	0.17
5	Daily Return	0.12

DISCUSSIONS

The research conclusions reveal that machine learning models that are explainable have the potential to forecast Nigerian bank equities with high accuracy. XGBoost outperformed the Random Forest for GTCO, Access Holdings, and Zenith Bank. It is a logical outcome because the XGBoost algorithm is more efficient in managing non-linear patterns, abrupt changes in price, and noisy data than Random Forest. These are the reasons why they chose the Nigerian market as the source of such features.

Access Holdings was the one among the stocks that recorded the minimum error values, probably because of its stock trades at a lower price range, thus making errors in prediction less significant. In contrast, GTCO appears to have the largest errors since it has been the most volatile in trading behavior, whereas Zenith Bank has only been showing moderate errors. The insights here are consistent with the findings of [1] and [6], whose studies indicate that employing a boosting model is the factor that influences the precision of financial forecasts to the greatest extent, by far, in emerging markets.

Trading volume and short-term moving averages were the

factors that most influenced the GTCO prediction; therefore, this means that the stock is very reactive to trading activity. Access Holdings was mostly influenced by the short-term trend and recent returns, thus indicating that the company had stable and smooth price movements. The transformation of Zenith Bank was to depend mainly on the overall market and the longer-term moving averages, which reflect the bank's stability and its strong correlation with the market. These findings align with the experiments of [4] and [5] that different features have varying effects on different stocks. However, the present findings contradict the discovery of [7], by which their macroeconomic factors had the most prominent influence. The researchers focused on the minute technical details like moving averages, volatility, and volume, rather than on macroeconomic variables. One such instance is the behavior of investors in Nigeria. If their behavior in the Nigerian market had been taken into consideration, the differences of these three banks could have been elucidated. Investors in the Nigerian market are more likely to make their decisions based on price changes, trends, and trading activities, disregarding well-established economic fundamentals. The results, in essence, align with the worldwide evidence that machine learning coupled with an explanation can result in higher prediction accuracy and can also provide more insight to the model. Moreover, the market performance has a greater impact on the Nigerian banking stocks than the macroeconomic factors. As a result, it seems that technical indicators like moving averages and volume are increasingly taking the lead in predicting price changes. Firstly, the mere fact that these indicators were the most dominant in all three models suggests that the Nigerian market is more influenced by short-term signals and investor sentiment.

CONCLUSIONS

This research examined the potential of using interpretable machine learning models to forecast the stock prices of the top three Nigerian banks: GTCO, Access Holdings, and Zenith Bank. Findings indicated that XGBoost outperformed Random Forest, implying that it was able to manage the fluctuations of the Nigerian market more efficiently. Besides that, SHAP analysis revealed that technical indicators, namely moving averages, volatility, recent returns, and trading volume, were the major contributors to the predictions. Even though these indicators were significant for all three banks, each bank treated them differently. In essence, the paper argues that machine learning combined with explainability can provide investors not only the predictions but also the rationale behind them. This elevates financial forecasting to a new level of clarity, openness, and practicality in Nigeria's emerging market. Therefore, investors have to implement in their decision-making process XGBoost with SHAP, which is a machine learning tool accompanied by an explanation method, because it yielded

the best results by the accuracy of the predictions. Such tools are helpful as they complement the intuitively derived conclusions and simple analysis by providing the patterns hidden from human eyes. Besides, the upcoming scholars are advised to introduce the same instruments in other Nigerian sectors for comparative study and conclusion drawing. This research revealed that even in the banking sector banks are different; hence, in other sectors the researchers will discover the unique patterns there too.

REFERENCES

- [1]. Ingle, V., & Deshmukh, S. (2021). Ensemble deep learning framework for stock market data prediction (EDLF-DP). *Global Transitions Proceedings*, 2(1), 47-66. <https://doi.org/10.1016/j.gltp.2021.01.008>.
- [2]. Demirbaga, U., & Xu, Y. (2024). Empirical asset pricing using explainable artificial intelligence (LIME & SHAP). *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4680571>.
- [3]. Goswami, B. (2025). Significance of predictors: revisiting stock return predictions using explainable AI. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-025-06717-2>.
- [4]. Lee, J., & Cai, L. (2025). Comparing model-specific and model-agnostic feature importance methods using machine learning with technical indicators. *Machine Learning with Applications*, 23, 100799. <https://doi.org/10.1016/j.mlwa.2025.100799>.
- [5]. Deng, S., Huang, X., Zhu, Y., Su, Z., Fu, Z., & Shimada, T. (2023). Stock index direction forecasting using an explainable eXtreme Gradient Boosting and investor sentiments. *The North American Journal of Economics and Finance*, 65, 101848. <https://doi.org/10.1016/j.najef.2022.101848>.
- [6]. Bussmann, N. (2025). Explainable machine learning to predict the cost of capital. *Frontiers in Artificial Intelligence*, 8, 1578190. <https://doi.org/10.3389/frai.2025.1578190>.
- [7]. Bao, W., Cao, Y., Yang, Y., Che, H., Huang, J., & Wen, S. (2025). Data-driven stock forecasting models based on neural networks: A review. *Information Fusion*, 106, 102616. <https://doi.org/10.1016/j.inffus.2024.102616>.