IKR Publishers

IKR Journal of Engineering and Technology (IKRJET)

Journal homepage: https://ikrpublishers.com/ikrjet/

Volume-1, Issue-3 (November-December) 2025

OPEN ACCESS

ISSN: 3107-7331 (Online)

Effect of Data-Driven Perspective on Stock Market Dynamics Using Machine Learning: A Case Study of Nigeria Cement Industries

Zingdul Kenneth Ponnan[1*], Bitrus Emmanuel[2] & Aaron Thomas Dalok[3]

[1,2,3]Department of Computer Science, Federal Polytechnic N'yak, Shendam, Plateau State

***DOI:**10.5281/zenodo.17661802*

ARTICLE INFO

Article history:

Received : 07-11-2025

Accepted : 15-11-2025

Available online : 20-11-2025

Citation: Ponnan, Z. K., Emmanuel, B., & Dalok, A. T. (2025). The Effect of Data-Driven Perspective on Stock Market Dynamics Using Machine Learning: A Case Study of Nigeria Cement Industries.*IKR Journal of Engineering and Technology (IKRJET)*, 1(3), 178-185.

ABSTRACT **Original research paper**

The Nigerian stock market, despite its vast potential, remains underutilized, hindering the country's economic growth. This underutilization stems from the inability of predictive models to forecast market trends with considerable accuracy, largely due to insufficient data and inadequate modelling techniques. The volatile nature of the market exacerbates poor investment decisions, resulting in significant losses for those who invest and discouraging potential investors. This research aims to develop a robust stock market prediction model using data-driven derivative which leverages on machine learning algorithms and statistical methods using the Nigeria's quoted cement companies on the Nigeria Stock Exchange (NSE). By uncovering meaningful patterns and trends, the model will provide valuable insights to inform investment decisions. The study employed a data-driven derivative approach on the utilized historical stock market data to train and test the model. The results demonstrate the model's effectiveness in predicting stock market trends, thereby promoting informed investment decisions. This research contributes to the development of a proven solution for stock market prediction, fostering economic growth, attracting investment, and improving Nigerians' overall well-being. By harnessing the potential of Nigeria's stock market, this study uncovers new opportunities for economic progress and prosperity.

Keywords: Data Science, Machine Learning, Predictive Model, Stock Marketing.

**Corresponding author: Zingdul Kenneth Ponnan*

Department of Computer Science, Federal Polytechnic N'yak, Shendam, Plateau State

Introduction

Stock marketing; the concept of buying and selling of shares as stock on an exchange by investors to make profit. Nigeria stock exchange is underutilized despite the sounding potentials it has, due to fear of permanent loss involve and its volatile nature scared investors from patronage of the stock market. This singular effect is a contributing factor that keep plummeting our economy as a nation. Predictive model in existence with an aim to cushion and reduce this high risk in stock marketing, perform greatly below expectation due to it low predictive power which stem from; machine learning models prone to overfitting due to inefficient training data. Successful machine learning model, require large amounts of data to be effective, but financial data can be limited or expensive to acquire. This limited data, has a significant challenge in using machine learning models for stock price prediction. Most machine learning inclined models lack techniques that can adapt to the dynamic nature of the stock market and continual adjustment in recent prices has dwindle it accuracy in nearest time than expected.

Similarly, Statistical models are also prone to overfitting, which occurs when a model fits too closely to the training data and does not generalize well to new data, []. Models developed using assumptions (generative) data has fail largely in forecasting of future pattern, []. Statistical models, developed based on generated data on the precedence of the underlying data, any violations to its assumptions can lead to inaccurate predictions, making it unreliable [].Statistical models may not always have high predictive power because they rely on historical data solely without further technics to improve it accuracy is not a good indicator of future market trends. Statistical methods are sacrosanct in stock price prediction model development because of it roles which is a base; such as pattern finding and relationship between variables, but they also have several limitations that should be improve.

Model which uncovered meaningful patterns and trends, provide valuable insights to inform investment decisions. This models contribute to economic growth as it reliability is recognised.

As long as models are crucial in stock price prediction, it is necessary to familiarise with core limitations that reduced it predictive power in order to bridge trust gap of prediction. Understanding these limitations, pose the need for a thorough investigation on the dataset to be used in predictive model development for better accuracy that can consistently manage the volatile natureof the stock market.

Literature Review

Accurate prediction of stock market encourages investment in stock marketing which is vital in an economy growth [1]. Nigeria stock exchange overtime have witnessed a low participation which stem from poor decision making which has brought a huge lost to investors due to the inability of model to make accurate prediction [1, 2]. Considering the facts that Stock markets are non-stationary in nature [3], this constant change in the stock market calls for a careful and sensitive analysis before investment. [4] Existing prediction model with low predictive power due to its development on inadequate data, insubstantial data for training and inadequate modelling techniques are some of the major pitfall this study attempt to solve using adequate and evident-based data derivative, leveraging on machine learning technique; [5, 6] a dynamic approach that can be used to adapt to the ever-changing environments which can be applicable to stock market. There exists a strong correlation between reduction of investment risk in stock market investments when there is reduction in forecasting error [7].

Also, statistical models developed on the basis on assumptions about an underlying data, any violations of these assumptions can lead to inaccurate predictions in future which is certain due to the natural volatility of the stock market. For example, a recent study [] found that the normality assumption of the GARCH model was violated in predicting the Indian stock market. However, this calls for a better opinion of model development which will warrant an efficient prediction.

Machine learning models trained on efficient can counter the common cases of overfitting generally, as verified [9], overfitting occurs when a model is trained on a limited dataset and fails to generalize on a new data [10]. This is particularly problematic for stock market data, which is complex and non-linear, making it easy for models to overfit and gives inaccurate prediction. Nevertheless, the application of dynamic approach in stock market prediction, which can be used for dynamic optimization, where an agent learns through interaction with its environment can serve the problem of overfitting [6,11]. This is evident also as a recent study by [6], used RL to optimize portfolio allocation and achieved better results than traditional statistical portfolio optimization techniques.

In consideration of this research as a data science related, many models build solely using the statistical method several limitations have been identified in its application on the Nigeria stock exchange as exposed by [12], found that statistical models, including ARIMA and Generalized Auto Regressive Conditional Heteroskedasticity (GARCH), had limited predictive power for the Nigerian stock market. However, statistical methods can't be avoided because of its values in stock price prediction such as computation to finds out relationship existing between variables [13].

Also, statistical models reliance on historical data alone for prediction reduces the predictive power of the model which is not a good pointer for future market trends [12]. However, model can further be improved by including additional variables and by using more advanced statistical techniques, hence this research rallies round concrete data to improve prediction accuracy [6].

ML is becoming increasingly important in stock prediction as it can adapt to the complex nature of the stock marketing environments, optimize portfolio allocation, and learn from complex market data. This research would utilize evidence-based data to trained and validate the model, leverages the strengths of machine learning ML to improve accuracy of prediction [14].

In view of the challenges embedded in the stock market predictions using many existing model and considering the application of data-driven derivative that leverages on ML in modelling of the stock market as viewed through it potentials it will be promising in building a stock predictive model with higher accuracy which will restored trust and confidence amount investors and stockbrokers in stock predictive model.

Methodology

This section presents the methods that will be followed in achieving this highly sensitive model in predicting the stock market with higher accuracy.

1. Historical Data

Actual dataset was gotten from the competitive Nigeria Cement Industries; Bua Cement PLC, Dangote Cement PLC, and Lafarge Africa PLC which are quoted on the Nigeria Stock Exchange (NSE), the are used in training and testing of the model to avoid the anomaly of assumption in training, validating and prediction of the model on **objective(iv).** The dataset used in training and validating of a stock prediction model is pertinent in determining of its accuracy, []. The research historical data was obtained for a period of three to four years and does not account for events that spans beyond these periods.

2. Data pre-processing

The dataset of the selected stocks will be pre-processed using statistical tools, which is necessary to have a well ordered data for analysis in order to avoid conflicting results. Data pre-processing involves cleaning, transforming, and preparing raw data into a format that can be easily analysed. The goal of pre-processing the historic data is to ensure that the data is accurate, complete, consistent, and relevant to the task at hand.

3. Data Exploration and Visualization

The pre-processed data will further undergo data exploration and visualization which is critical in any data analysis task, it presents the pictorial view of the data for a better understanding and further analysis. This steps involves understanding and exploring the data to identify patterns, relationships, which gives insights that can be used to guide further analysis.
The steps involved;

a) Understanding the data: Involves familiarity with the basic of the data, such as; its size, structure, and format.
b) Descriptive statistics: This includes computation of descriptive statistical measure i.e. mean, median, standard deviation, and other summary statistics to gain insight into the data.
c) Data visualization: Presentation of visuals charts of the stock, for instance scatter plots, histograms, and box plots, to identify patterns and relationships between stocks.
d) Correlation analysis: This is the analyses of the relationship between different variables for the purpose of identifying the interactions and dependencies between the stocks, using regression analysis. It is commonly used to estimate the effect of one or more explanatory variables on a particular outcome or response variable, [15].

Regression analysis used a given linear function for predicting continuous values:

$$y = \beta 0 + \beta 1x + \varepsilon \qquad (1)$$

where:

- y is the dependent variable
- x is the known independent variable
- $\beta 0$ is the intercept
- $\beta 1$ is the coefficient of the independent variable x
- ε is the error term

4. OLS Regression Model

Developed the model using the actual data of the four stock cement companies. According to [16] OLS regression involves calculating the difference between the predicted values of y and the actual values of y for each data point. The sum of the squared differences is then minimized to find the best estimates of $\beta 0$ and $\beta 1$. The estimates of $\beta 0$ and $\beta 1$ can be used to predict the value of y for a given value of x. developed the OLS model with the actual (historical) data of the stock train and test the model to visualizes it predictive power and to further obtain the predicted value for the training of the data-driven model in order to discover and compare their predictive power for conclusion.

5. Data-Driven Model

Data-driven model, leveraging on machine learning ML technique in which policy optimization will be used to optimized decisions by training the model with the derivative of the predictive values of the OLS model. The purpose is to obtain a model with higher predictive power than the later. In this study, data-driven approaches will be use to improve the predictability of an Ordinary Least Squares (OLS) model.

6. Develop A GARCH Predictive Model Using Generative Data

GARCH models is used to analyse the volatility of financial time-series data. GARCH stands for Generalized Autoregressive Conditional Heteroscedasticity. The models are designed to capture the conditional volatility of a financial asset or a portfolio of assets. Use generative dataset to developed a model, train and validate the model with it, and make visual chart of the model. Make comparison using like dataset for another testing, make visual chart and draw conclusions on it findings. To discovered accuracy in the volatility test using the GARCH model with the normality assumption in predicting the stock market.

7. Model Evaluation

Evaluating a machine learning model is critical for determining its performance on previously unseen data. This will be carryout on the two separate model. This approach identifies whether the model is overfitting or under fitting and assesses its generalizability. This study evaluates models using three main metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These metrics offer a thorough view of the model's performance and forecast accuracy.

Results and Discussion

The findings and discussion of the results of the developed model which helps in drawing conclusion and recommendation goes here:

Correlation of Stocks; Bua Cement, Dangote Cement, and Lafarge Africa

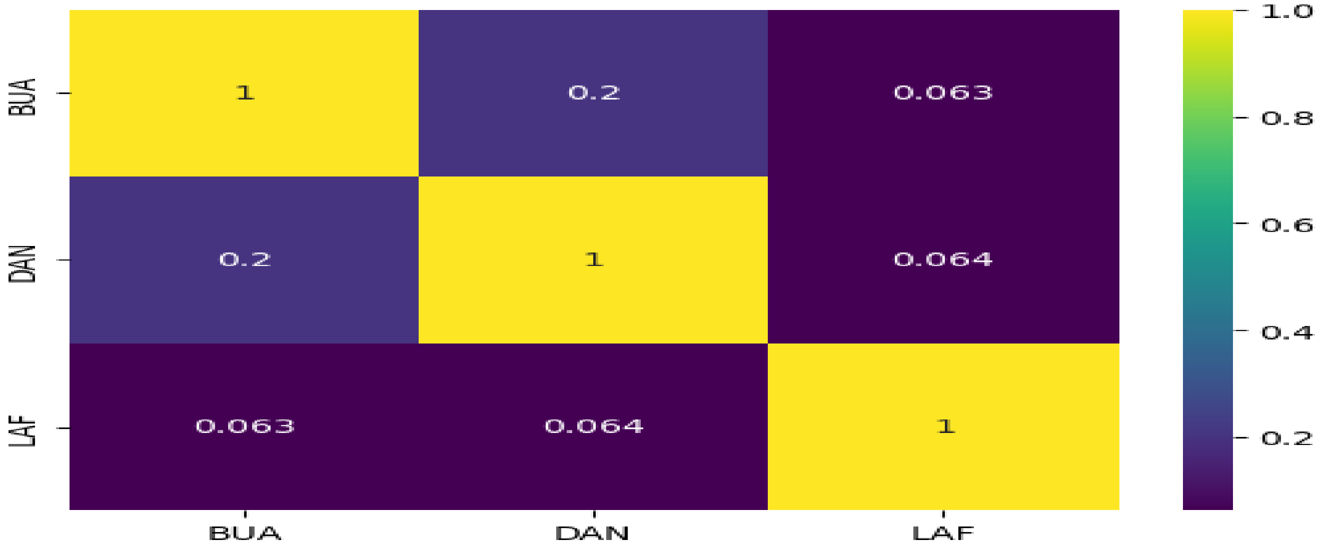

Figure 1: Correlation heat map for daily returns

Figure 1 shows the correlation heat map of daily returns for three equities. The findings show weak positive associations between BUA Cement and DANGOTE Cement (0.20), the association between BUA Cement and LAFARGE AFRICA Cement (0.062) is very weak positive correlation, and DANGOTE Cement and LAFARGE AFRICA Cement (0.064) was equally a weak positive correlation. These low correlation coefficients point to a modest association between the equities. Further investigation may reveal hidden patterns in the data.

Daily Return Check for Stocks

Chart showing the volatile nature of the stock market.

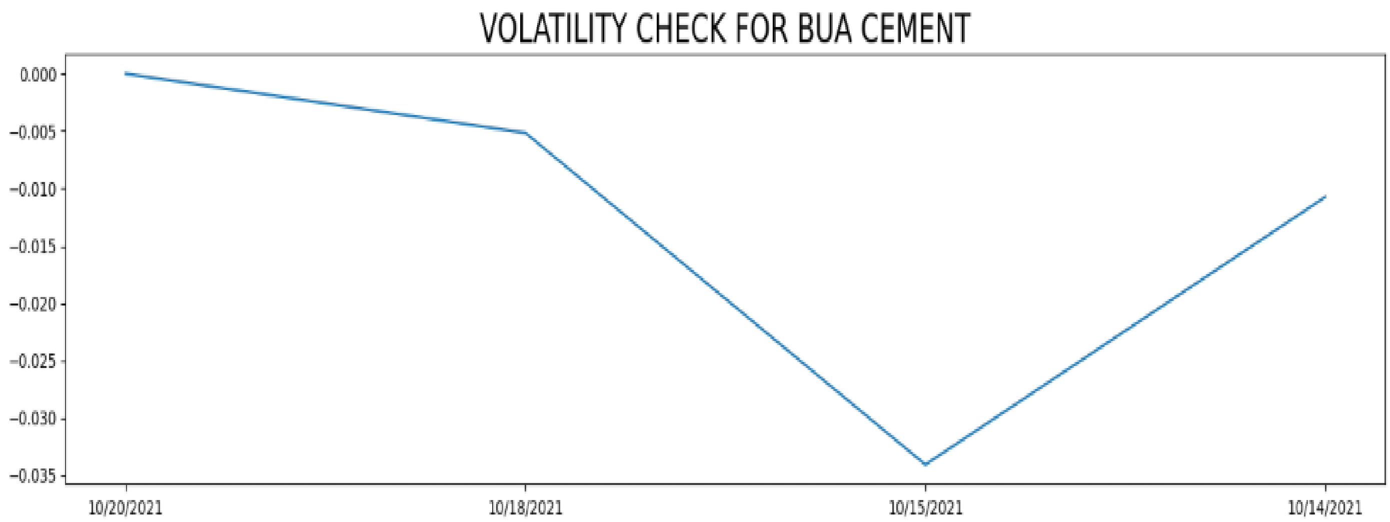

Figure 2: Volatility (Daily Return) of the Stock Market

Figure 2 shows the trend plot for the daily returns for BUA Cement (BUA) within the first 4 days. The visual representation of the market trend corroborated with the assertion against the use of assumption data for predictive model of the stock market considering high level of instability within the stock market and generated dwells better in the design of discreet system. Hence, with the high volatility noted, it is an indicator to perform a deeper analysis on the stock which will implicate positivity of an efficient system in predicting the future price of the stock.

Generative Data Effect

Figure 2, visualizes the trend of the stock of the Bua cement which a high level of instability was observed. Hence, a model built on assumption is will not perform well in prediction of the future stock price. However, we will visualize the behavior of the model developed using the generated data.

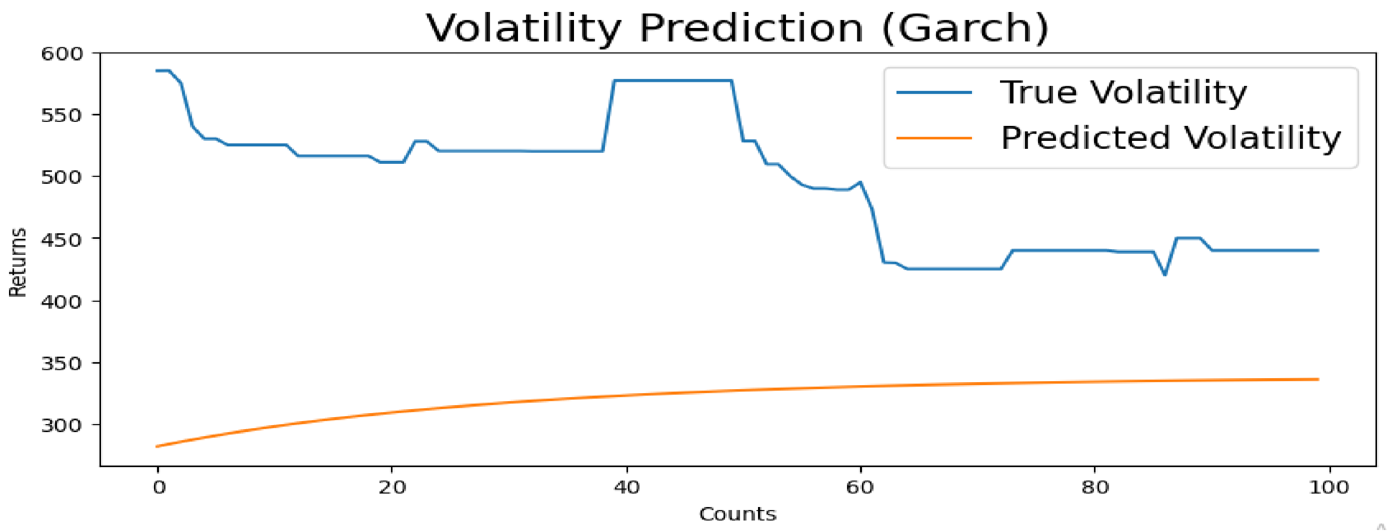

Figure 3: Prediction using Generated data

The finding revealed that generative data (synthetic) are good predictions parameters, however they are not efficient and reliable enough for future prediction in models who are built to predict volatility, this uncertainty can hamper its underlying principle, leading to inaccurate prediction. However, its core benefit of being a cheap data source, with good labels but the aim of it to bring perfect prediction cannot be achieved. Hence, generative (synthetic) data as observed and found in many statistical models which based on assumptions about the underlying data, any violations of the assumptions can lead to inaccurate predictions. As discovered in the volatility test using the GARCH model reveals that the normality assumption of the GARCH model was violated in predicting the stock market with the nature of the system.

Actual Data (Historical data) Effect

This study used Ordinary Least Squares (OLS) regression analysis to model the relationship between variables, specifically calculating the link between a stock's past and future prices. Using the returns of three related stocks as a training set, the OLS model produced reasonable forecasts for the future returns of the chosen stock, LAF. The findings indicate the effectiveness of actual data usage in development of predictive model, it further clarified that OLS can be a useful tool for stock prediction, especially when combined with appropriate historical

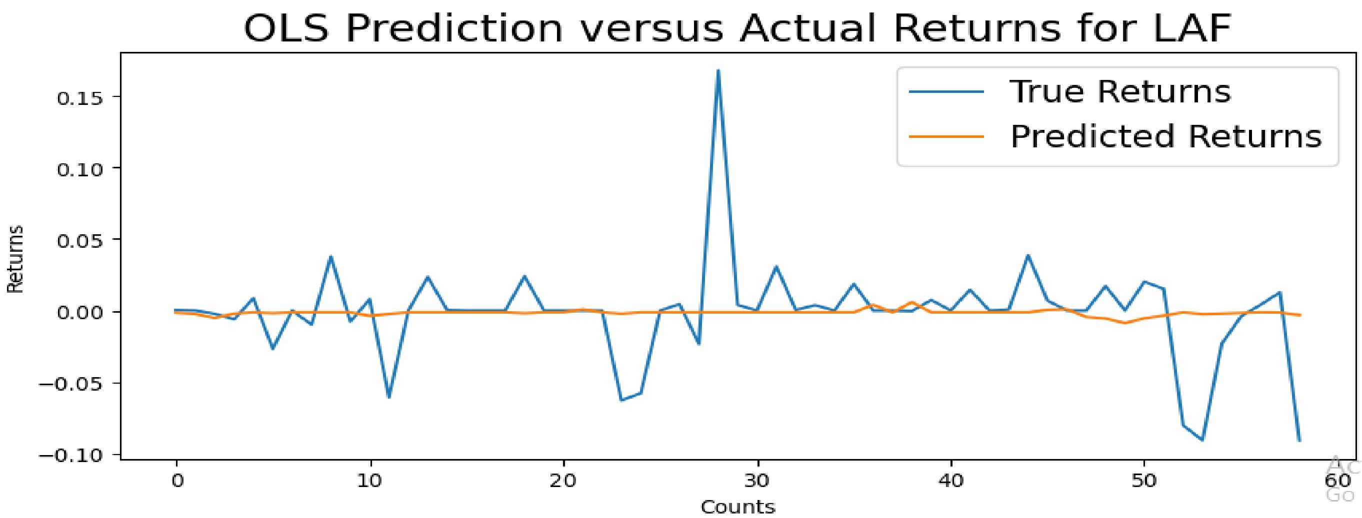

Figure 4: OLS Predicted Returns vs Actual Returns

Figure 4 shows the predicted returns against actual return. From the visual chart it can be deduce that the model performance is better when compares to figure 3 which is the GARCH model developed on assumption data. Hence, we can conclude that the models developed for prediction using concrete data as training dataset gives a high predictive power compare to models developed using generated data. The consistency in patterns between actual and expected returns demonstrates the value of employing evidence-based data to train and develop stock predictive models, indicating a significant improvement in statistical modeling.

Thus a statistical evaluation metric will be used to substantiate the conclusion on the model. Figure 5 shows, the result of the evaluation metric for statistical analysis; the Mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) which are useful in evaluating the performance of regression models.

OLS Prediction - MSE: 0.0012026978141227002, RMSE: 0.03467993388290555, MAE: 0.01832890551530034

Figure 5: OLS Model Standard Error of Mean

Data-Driven Effect on Predictive Model

Figure 6, visualizes the predicted and actual returns of the stocks on the model developed using the data-driven derivative. It can be deduced from the trend path of the actual and the predicted returns that are in similar patterns. This indicate the high predictive power of the data-driven effect on the model. We will further fleshy the finding with the statistical evaluation metric as shown in figure 7 to have valid argument for a conclusion. These findings show that the machine learning model, which was derived from the Ordinary Least Squares (OLS) model and retrained with evidence-based data, has greater predictive potential.

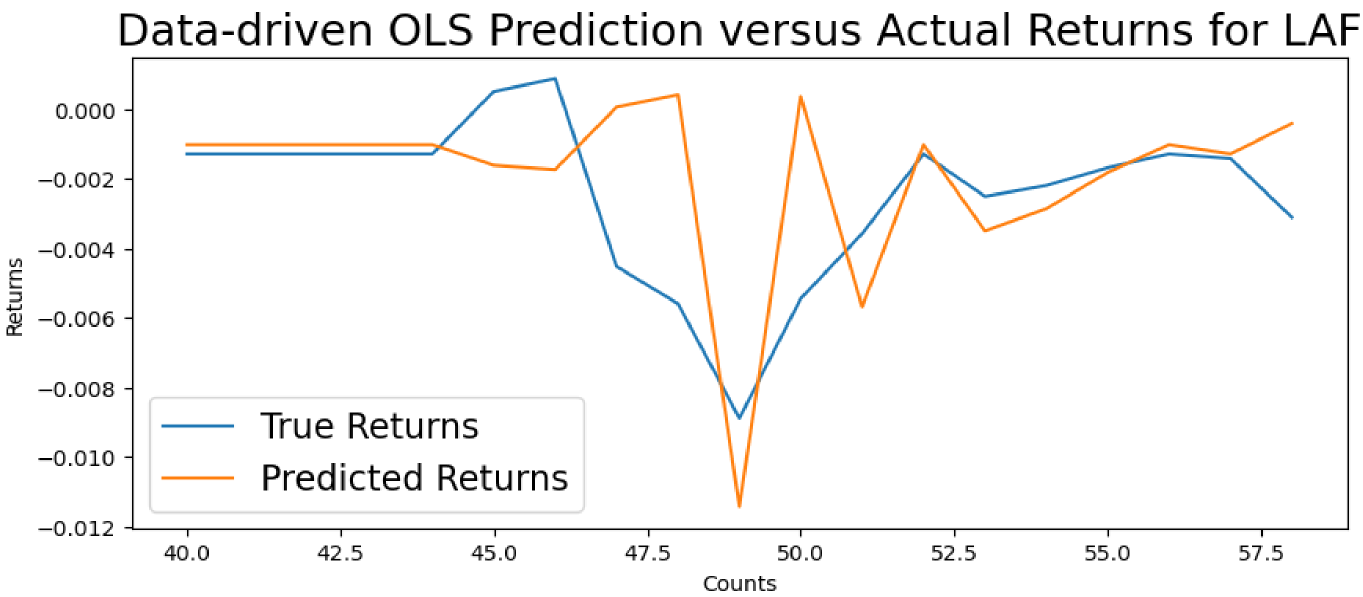

Figure 6: Data-driven OLS Model

Data-driven Prediction - MSE: 6.441496412274758e-06, RMSE: 0.0025380103254862376, MAE: 0.0017008898874623405

Figure 7: Standard Error of Mean

The findings revealed that efficiency and resilient nature of using of data-driven derivative which adds significantly to the predictive power of the model when visualizes through the testing process. An alike pattern of the predicted values and that of the actual values was observed. This promulgation was further strengthen through, the evaluation metrics using the statistical tool which was significantly improved in values compared to that of the model built on actual data alone as shown in figure 8. Hence, it can be concluded that data-driven effect on predictive power of model is more efficient than just relying on actual data (historical data) which are not often available in quantity for model training and testing which contributes to model overfitting. These challenges faced by developers of machine learning solutions for stock price predictions, for instance, machine learning algorithms require a large amount of historical data to learn from. Since, stock market data is limited by the number of years of historical data that is available. This makes it difficult for machine learning models to accurately predict future stock prices.

OLS Prediction - MSE: 0.0012026978141227002, RMSE: 0.03467993388290555, MAE: 0.01832890551530034

Data-driven Prediction - MSE: 6.441496412274758e-06, RMSE: 0.0025380103254862376, MAE: 0.0017008898874623405

Figure 8: Standard Error of Mean Comparison of OLS Model and Data-driven Model

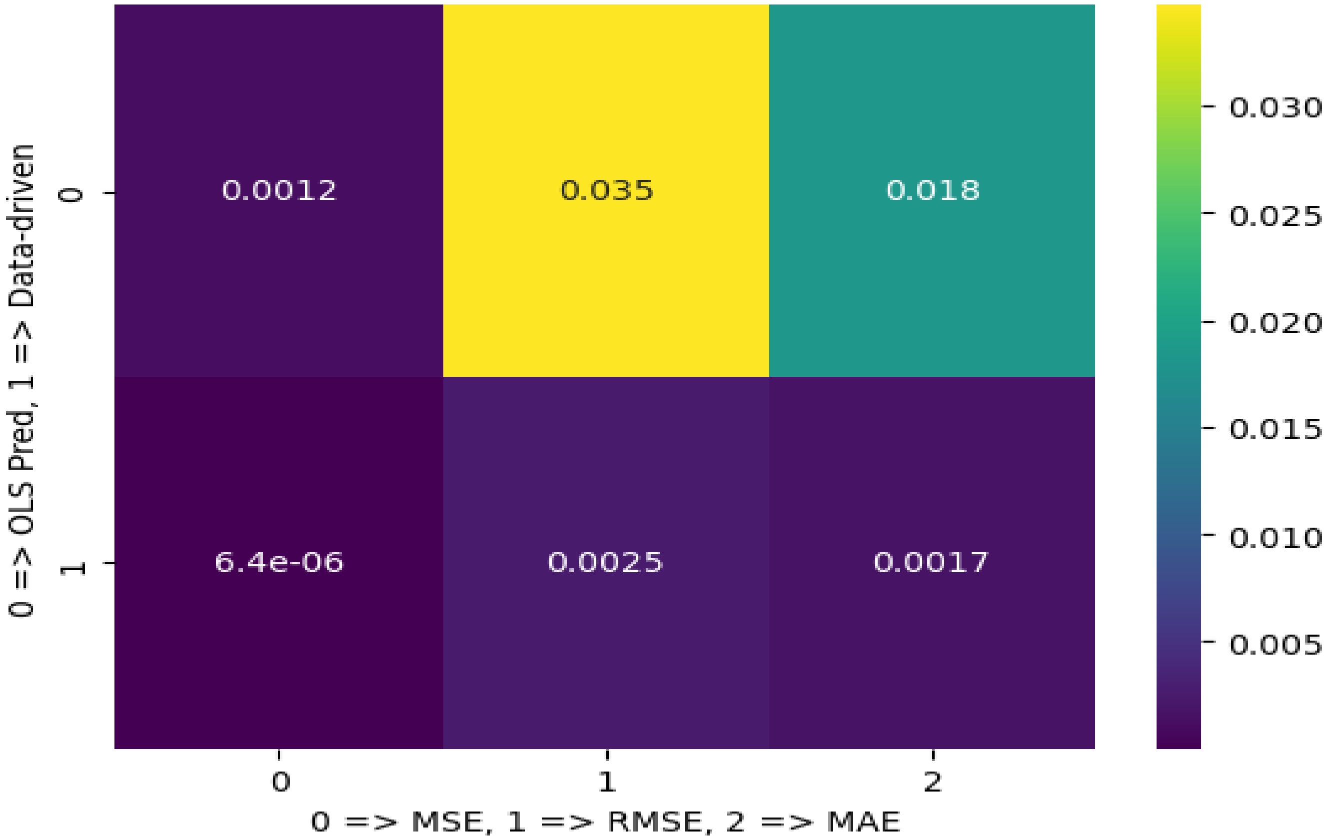

Figure 9: Heatmap of Standard Error of Mean Comparison of OLS Model and Data-driven Model

Conclusion

This research explores the hidden potential of data-driven derivatives which leverages on machine learning techniques to adequately find and predict pattern of the future price of the stock market which is a nonlinear and complex system with opportunities that is mouthwatering but risky to invest in when the know-how or proper guidance is not certain. The association between the stock's previous prices and its future prices was estimated using OLS in the context of stock prediction for this study. The study's findings demonstrate that, when the returns of the three related stocks are used as the training set, the OLS model gave decent predictions for the future returns for the chosen LAFARGE CEMENT stock. By adding an attribute to the training set, the data-driven derivative approach which was applied to the OLS model, which was built using the actual data only in this study, and the results demonstrate a considerable improvement in the Data-Driven model's performance. As a result, it is recommended that the approach be adopted in the real world.

We offer a novel method to decision-making for investors and stockbrokers by combining machine learning technique, statistical tools, and evidence-based data. Which will lead to higher stock market investment, lower risk, and, eventually increases economic growth. The findings emphasize the potential of data-driven tactics to alter the stock market and guide investing decisions through higher predictive power it offers.

Recommendations

As indicated from the findings; it is necessary to recommend that:

I. Data-driven derivative significantly influences the predictive power of models, hence this approach of model development is suitable for application in real-life system of future forecast due to its high predictive power and considerable accuracy.

II. Models built using assumption data (generative data) are not suitable for future forecast, for it does not predict accurately, hence such approach should be applied to application than can operate under specified conditions.

III. Actual data (historical data) strongly influence the predictive power of models and it should be adopted for model development.

IV. Historical data, should be make available for IT personnel's, scholars and all interested parties with meaningful idea of development.

Acknowledgement

I am grateful to Federal Polyechnic N'yak, Shendam for providing the needed support in conducting the Research.

References

1. Oyewole, A. T., Adeoye, O. B., Addy, W. A., Okoye, C. C., Ofodile, O. C., & Ugochukwu, C. E. (2024). Predicting stock market movements using neural networks: a review and application study. *Computer Science & IT Research Journal*, *5*(3), 651-670.
2. Adams, S. O., Asemota, O. J., & Ibrahim, A. A. (2024). Asymmetric GARCH Type Models and LSTM for Volatility Characteristics Analysis of Nigeria Stock Exchange Returns. *American Journal of Mathematics and Statistics*, *14*(2), 17-32.
3. Al Fakih, B., & Abdel-Salam, A. S. G. (2025). On the Gaussian Process for Stationary and Non-stationary Time Series Prediction for the Qatar Stock Market. In *Innovative and Intelligent Digital Technologies; Towards an Increased Efficiency: Volume 2* (pp. 63-75).
4. Cham: Springer Nature Switzerland. Lee, S., & Chun, J. (2025). Development of Machine-Learning-Based and Deep-Learning-Based Models for Predicting Korean Adolescents' Overweight and Obesity. *Journal of the Society for Social Work and Research*, *16*(1), 000-000.
5. Ngwu, C., Liu, Y., & Wu, R. (2025). Reinforcement learning in dynamic job shop scheduling: a comprehensive review of AI-driven approaches in modern manufacturing. *Journal of Intelligent Manufacturing*, 1-16.
6. Zingdul, P. K., Blamah, N., & Oyerinde, Y. (2024). Improving Stock Market Prediction through Reinforcement Learning: A Case Study of Nigerian Telecommunication Companies. International Journal of Technopreneurship and Innovation, 1(1), 104–118. Retrieved from https://journals.unijos.edu.ng/ijti/article/view/305
7. Wood, M., Ogliari, E., & Leva, S. (2024, September). The Value of Forecasting: The Effect of Building Load Forecast Errors on the Performance of an Optimal Energy Management System. In *2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI)* (pp. 42-47). IEEE.
8. Akbar, S., & Shah, S. R. (2024). Mathematical modeling of blood flow dynamics in the cardiovascular system: Assumptions, considerations, and simulation results. *Journal of Current Medical Research and Opinion*, *7*(04), 2216-2225.
9. Aliferis, C., & Simon, G. (2024). Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and AI. *Artificial intelligence and machine learning in health care and medical sciences: Best practices and pitfalls*, 477-524.
10. Adewole, Ayoade I. "On the Hybrid of Arima and Garch Model in Modelling Volatilities in Nigeria Stock Exchange." *Bima Journal of Science and Technology* 8, no. 2A (2024): 169-180.
11. Osho, G. S., & Oloyede, B. (2024). A Generalized Autoregressive Conditional Heteroscedasticity GARCH for Forecasting and Modeling Crude Oil Price Volatility. *The Journal of Applied Business and Economics*, *26*(6), 39-53.
12. Adams, S. O., Asemota, O. J., & Ibrahim, A. A. (2024). Asymmetric GARCH Type Models and LSTM for Volatility Characteristics Analysis of Nigeria Stock Exchange Returns. *American Journal of Mathematics and Statistics*, *14*(2), 17-32.
13. Chung, Y. L., Wu, Z. L., & Pichappan, P. (2024). Application of deep learning and statistical methods in predicting Taiwanese stock trends. *Journal of Computational Methods in Science and Engineering*, *24*(3), 2017-2035.
14. Ngwu, C., Liu, Y., & Wu, R. (2025). Reinforcement learning in dynamic job shop scheduling: a comprehensive review of AI-driven approaches in modern manufacturing. *Journal of Intelligent Manufacturing*, 1-16.
15. Shinkafi, N. I. (2025). Analysis Of Financial Data Based On The Linear Estimator Of The Arch, Garch Model International Journal of Innovative Mathematics, Statistics & Energy Policies 13(1):31-59, Jan.-Mar., 2025.
16. Marquez, B. Y., Realyvásquez-Vargas, A., Lopez-Esparza, N., & Ramos, C. E. (2024). Application of ordinary least squares regression and neural networks in predicting employee turnover in the industry. *Archives of Advanced Engineering Science*, *2*(1), 30-36.