Iosr Journal Of Economics And Finance (Iosr-Jef)
E-Issn: 2321-5933, P-Issn: 2321-5925.Volume 15, Issue 6 Ser. 5 (Nov. – Dec. 2024), Pp 19-31
www.Iosrjournals.Org

Predictive Modeling Of Macroeconomic Trend In Nigeria: An Integrated Approach Of Using Correlation, Cluster-Based Similarity, And Artificial Neural Network

Robert, C.O[1]., Adekunle, J. D[2]., Oyeniran, M. I[2]., Collins, T.P[3]., Sule, H. S[4]., Ayanlowo, E.J[5]., Ogu, C. K[6].,Akinpelu, T. T[4], Alagbe S. A., Ogu, E. M.[7]
Management Information Systems Department, Topdel Engineering Limited, Lagos, Nigeria.
Department Of Mathematics, Federal University Of Agriculture, Abeokuta, Ogun State.
Investment Banking Department, Pathway Advisors Limited, Lagos, Nigeria.
Department Of Statistics, Federal University Of Agriculture, Abeokuta, Ogun State.
Graduate School Of Asia Pacific Studies, Ritsumeikan Asian Pacific University.
Production Department, Medipolis Gmbh, Thuringia, Germany.
21st Century Technologies Limited, Lagos, Nigeria

Abstract
This study analyzed economic indicators, to assess patterns in Nigeria's GDP, poverty rate, and other key variables over a multi-year period, using a combination of predictive modeling, correlation analysis, clustering, and dimensional reduction techniques. The predictive analysis revealed a strong alignment between the observed and predicted values for GDP and poverty rates, with minor error margins and accuracy levels close to 100%, underscoring the model's robustness in forecasting these economic indicators. Correlation analysis highlighted significant positive relationships between GDP and other variables, particularly importation, export, and exchange rates, while inflation and MPR showed moderate associations. Cluster analysis identified three distinct periods with economic similarity: 2014–2018, marked by stable growth; 2019–2022, likely influenced by external shocks or adjustments; and 2023, standing out as unique, potentially due to significant economic changes. The t-Distributed Stochastic Neighbor Embedding (t-SNE) further validated these clusters by revealing distinct separations between the identified periods. The combined results offer insights into Nigeria's economic patterns, revealing periods of stability and adjustment, and highlighting variables that significantly influence GDP and poverty, aiding policymakers in identifying trends and making data-informed decisions.

Keywords: *Economic Forecasting, GDP Prediction, Poverty Rate Analysis, Economic Clustering, Macroeconomic Indicators, Nigeria's Economic Trends*

Date of Submission: 04-12-2024 Date of Acceptance: 14-12-2024

I. Introduction

Nigeria stand-out as the African largest economy, rich with crude oil, Agriculture products and other natural resources and has the population about 200 millions. The economic progress depends heavily on the oil, making its economic prosperity to be at risk with fluctuations in world oil prices. Nigeria witnessed a 2-3% of economic recovery progress after 2016 economic recession which eventually collapsed in 2020 because of COVID- 19 pandemic. Consequently, the economy regain its strength in 2021 and 2022 with growth of 3.4% and 3.1% respectively caused by increased in oil production, rise in oil prices, and restoration in other non-oil production services (World Bank, 2023).

Nevertheless, about 40% of the country's population lives below poverty line irrespective of this economic progress, causing social imbalances among the citizens across the regions (N.B.S, 2023). Poverty is caused by a combination of factors including inadequate food supply, low income and the inability to access essential needs such as shelter, clothing, proper healthcare, quality education, clean water, and sanitation. It also involves lack of physical security and opportunities to improve one circumstances which make individuals exposed to violence and force them to live in neglected and unstable environment (Kuke, et al 2016).

Formulation of policies that guide resources allocation, economic interventions, and long term planning are directly influenced by the quality of the forecast. Therefore, poor predictions can lead to deluded policies that may worsen economic issues such as unemployment, inflation or poverty rate (Leamer, 2010). Karabell (2014), indicated that identifying future trends, risks, and opportunities in economic planning and predicting GDP growth with high accuracy which allows government to design fiscal policies that balance economic growth with inflation control heavily depend on precise forecasts. Furthermore, Policy makers target social welfare programs effectively from accurate poverty level forecast which ensure proper resources allocations in most needed areas (U.N.D.P, 2020).

According to Athey and Imbens (2019), Machine Learning (ML) was said to have gained popularity in forecasting econometric models because of its ability to handle economic situations that might display chaotic properties. Thus, economic situations that are complex, non-linear, sensitive, unpredictable and contain large data (Guo, 2002). Traditional econometric models such as Autoregressive Moving Average (ARIMA) and Vector Autoregression (VAR) often really on linearity and stationarity assumptions which pose limitations in capturing complicated interactions between multiple variables whereby Machine Learning offers a greater flexibility by identifying patterns and trends without underlying assumptions about data distributions. Agu et al, (2020) stated that Machine Learning models such as Artificial Neural Network outperforms in time-series forecasting of economic indicators that evolve over time like GDP growth or inflation because it captures long-time dependencies in sequential data.

Moreover, Machine Learning models have proven to outperform traditional models in both accuracy and efficiency, particularly in settings where data are non-linear. Studies has shown that most world economic systems are non-linear; it is unreasonable to assume a priori that a particular realization of a given time series is generated by linear process (Yassen, 2011). Several non-linear models like Bilinear Model (BM), Threshold Autoregressive Model (TAR) and Autoregressive Conditional Heteroscedastic Model (ARCH) has been developed, but these non-linear models has limitations, as they require specific relationship for the data sets to be assumed. These has been difficult task because there are too many possible non-linear patterns and pre-specified non-linear models may not be generated enough to capture the important factors. Machine Learning models which are non-linear data-driven models are capable of performing non-linear modeling without prior knowledge about the relationship between the input and output variables. Thus, they are generally and more flexible modeling techniques for forecasting econometric variables (Yassen, 2011).

Machine learning is broadly defined as a branch of Artificial intelligence (AI) that focuses on developing algorithms and models that allow computers to learn from and make predictions or decisions based on data without being explicitly programmed. According to Bishop (2006), it is a process that enables computers to improve their performance on a task over time as they gain more data. It involves data-driven approaches to solve complex problems and can be applied across multiple domains, such as image recognition, natural language processing, and economic forecasting.

This research aims to investigate the effectiveness of machine learning models in forecasting Nigeria's real GDP and poverty levels, using data from key economic indicators.

The objectives of this study is to develop a machine learning models that can accurately predict Nigeria's GDP growth and Poverty level and to evaluate the performance of this machine learning algorithms in predicting future GDP trends and poverty level. The predictions and insights generated by this machine learning models are to inform economic policy recommendations and help the Nigerian government address issues related to GDP growth and poverty alleviation.

II. Litrature Review

Gross Domestic Product (GDP) and Poverty Theories in Nigeria

Economic Theories highlight the patterns of GDP and poverty, thereby providing a fundamental knowledge on how a country's economic policies and growth patterns influence poverty levels. In Nigeria, GDP and poverty are intertwined, cutting across several economic theories, which in turn elaborates the country's development challenges and opportunities. In interpreting Nigeria's inconsistent economic performance and the prevalence of poverty in different regions of the country, these theories are found very useful.

Classical Economic Theory

The Classical Theory, basically from the works of Adam Smith (1776) stressed the responsibility of free markets and minimal government interference in economic growth. According to classical economic theory, the economy is self-regulating hence, growth can be achieved by allowing market forces to operate without restrictions. In Nigeria, economic liberalization reforms that facilitate free trade, deregulation and privatization are in consonance with the classical school of thought. These reforms which came into spotlight since the 1980's targeted at stimulating economic growth by enhancing production, trade and investment. While the Nigerian GDP has improved as a result of such policies, poverty remains prevalent, disputing that market-driven growth solely has been insufficient in reducing poverty significantly.

Dependency Theory

Dependency Theory often linked to Andre Gunder Frank (1967) debates that developing countries like Nigeria are structurally disadvantaged within the global capitalist system. This theory proposes that the transfer of resources from poorer countries to wealthier nations encourages the dependence of countries such as Nigeria, on external markets, precisely through their reliance on primary commodities that is, crude oil. As a result, the vulnerability of Nigeria GDP to global oil price fluctuations is on a high. These reliance on oil revenue has led to volatility in Nigeria's GDP growth, resulting to difficulty in achieving a sustainable economic development. Dependency Theory posits the need for diversity in Nigeria's economy and the reduction of its reliance on exports of raw materials for the escape of dependency, thereby achieving more sustainable and inclusive growth.

Endogenous Growth Theory

Endogenous Growth Theory propounded by Economists like Paul Romer (1986) and Robert Lucas (1988), highlights the role of human capital, innovation and knowledge in driving long-term economic growth. This Theory proposes investment in education, technology and research which in turn lead to sustained GDP growth without diminishing returns. In Nigeria, poor investments in education and technology have hampered the country's ability to fully optimize endogenous Growth factors, stemming to a slow transition from an oil-dependent economy to a more expanded one. Irrespective of Nigeria's GDP growth, high level of poverty prevail as a result of inequitable distribution of resources and insufficient opportunities for many in the participation of formal economy.

Kuznets Curve

The Kuznets Curve as proposed by Simon Kuznets (1955), posits that in the initial stages of economic development, disparity and penury arise as countries industrialize. Therefore as GDP increases and development progresses, disparity and penury decline. The occurrence in Nigeria partially aligns with this theory. As rapid municipalization and growth in sectors like oil and telecommunication have aggravated income inequality, specifically between urban and rural areas. However, the poverty-reducing effects of GDP growth have been restricted in Nigeria. The wealth generated by the country's oil-driven economy has not significantly trickled-down to alleviate poverty levels, as proposed by Kuznets hypothesis. Following this, scholars question whether Nigeria's growth is truly inclusive or it benefits only a small section of the population.

Pro-Poor Growth Theory

Pro-poor growth underlines that economic growth should unevenly benefit the poor to efficiently alleviate poverty. According to Ravallion and Chen (2003), pro-poor growth emerges when the income of the poor rise faster than average income levels. In Nigeria, GDP growth has not been pro-poor, as wealth has been fixated within the small exclusive, leaving widespread penury unattended to. For Nigeria to cultivate pro-poor growth focused policies promoting broad-based economic development, precisely in agriculture, small and medium-sizes enterprises (SMEs), and other sectors influencing low-income households are essential. Capitalization in rural infrastructure, education and healthcare could ensure that GDP growth is beneficial to Nigeria's poorer communities, thence overcoming the barrier that exist between national economic performance and poverty mitigation.

Previous Studies on Using Machine Learning in Predicting GDP and Poverty Level

Adewale et al. (2024), applied Machine Learning models to predict the GPD of Nigeria using variables such as healthcare expenditure, net migration rates, population demographics, life expectancy, access to electricity, and internet usage. The analysis reveals that all selected indicators have a strong correlation with GDP.

In the study conducted by Agu et al. (2022), the researchers employed four machine learning techniques to predict GDP based on various macroeconomic indicators. The objective was to determine which key macroeconomic variables most likely impact GDP growth. His results indicated that increases in population, federal government expenditure, import rates, and export rates contribute to GDP growth, whereas foreign direct investment, exchange rates, and oil revenue have a negative impact on GDP.

Ibrahim and Ren (2021) utilized machine learning and deep learning methods to analyze and predict poverty in Nigeria using data from the Living Standard Measurement Study (LSMS). They concluded that although each model has its strengths and weaknesses, both can be combined to serve as robust tools for understanding and forecasting poverty.

In his 2021 study, Jim Kim applied two supervised multiclass machine learning models to predict the poverty status of households in Costa Rica. The aim was to aid both government and business sectors in making informed decisions within a rapidly evolving social and economic context. His study revealed that education has the most significant influence on predicting a country's poverty status. In 2022, Roode utilized hybrid models to forecast Nigeria's GDP growth. The findings indicated that the performance of hybrid model was good, when compared with that of Autoregressive models, demonstrating that combining factor models with boosting improves the accuracy of forecasting Nigerian GDP growth.

Table shows the summary of selected related studies on GDP and Poverty.

Author	Objective	Methodology	Results
Adewale et al., 2024.	To enhance the prediction accuracy of Nigeria's GDP using a diverse range of socio-economic indicators.	Applied Machine Learning models such as Random Forest Regressor, XGBoost Regressor, and Linear Regression to predict the GPD of Nigeria using variables such as healthcare expenditure, net migration rates, population demographics, life expectancy, and access to electricity, and internet usage.	The Random Forest Regressor emerges as the most robust model, boasting an R-square score of 0.96 and a Mean Absolute Error (MAE) of 24.29.
Agu et al., 2022.	To determine which key macroeconomic variables most likely impact GDP growth of Nigeria.	Employed four machine learning techniques—Principal Component Regression (PCR), Ridge Regression (RR), Lasso Regression (LR), and Ordinary Least Squares (OLS)—to predict GDP based on various macroeconomic indicators.	Among the methods, PCR emerged as the most accurate, achieving an 89% accuracy rate and producing a mean square error of -7.552007365635066e+21, outperforming the other models in terms of prediction accuracy.
Ibrahim and Ren, 2021	To conduct a multidimensional poverty analysis and predict poverty using different Machine Learning approaches.	Logistics Regression, Decision Tree, Random Forest and Convolutional Neural Network models.	Logistics Regression had 99% specificity, 99% sensitivity and 98% accuracy. Decision Tree had 69% sensitivity, 94% specificity and 95% accuracy. Random Forest had 99% specificity and 99% sensitivity and 98% accuracy while Convolutional Neural Network had 99% sensitivity and 100% accuracy.
Roode, 2022	To investigates whether a hybrid forecasting model is useful in forecasting the gross domestic product (GDP) growth of Nigeria.	Utilized hybrid models that incorporated boosting, Principal Component Analysis, and Sparse Principal Component Analysis to forecast Nigeria's GDP growth.	The hybrid models outperformed the benchmark AR (1) model in terms of mean squared forecast error (MSFE), of 0.501 and 0.840 respectively.
Jim Kim, 2021	To aid both government and business sectors in making informed decisions within a rapidly evolving social and economic context.	Applied Random Forest and Gradient Boosted Trees machine learning models to predict the poverty status of households in Costa Rica.	Random Forest and Gradient Boosted Trees, achieved F1 scores of 64.9% and 68.4%, respectively.
Abdirizak et al., 2024.	To address the gaps of limited prediction capability of regression by using advanced Machine Learning methods in predicting poverty in Somali.	Used four ML models of Random Forest, Decision Tree, Support Vector Machine and Logistic Regression.	The accuracy of prediction of the models ranges from 67.21% and 98.36 with Random Forest model demonstrating the best performance.
Proposed Models	To analyze and predict Real GDP and Poverty level of Nigeria from Historical data of selected economic Variables.	Applied combination of predictive modeling, correlation analysis, clustering, and dimensional reduction techniques.	ANN showed 99.85% and 100% for the both models.

III. Methodology

Data collection

We conducted a secondary data analysis within the quantitative and observational research design by focusing on numeric data of selected economic indicators. The historical data used in this study was collected from government agencies and organizations. The organizations are National Bureau of Statistics (NBS), Central Bank of Nigeria (CBN) and International Monetary Fund (IMF), and the data was verified through data triangulation to avoid bias.

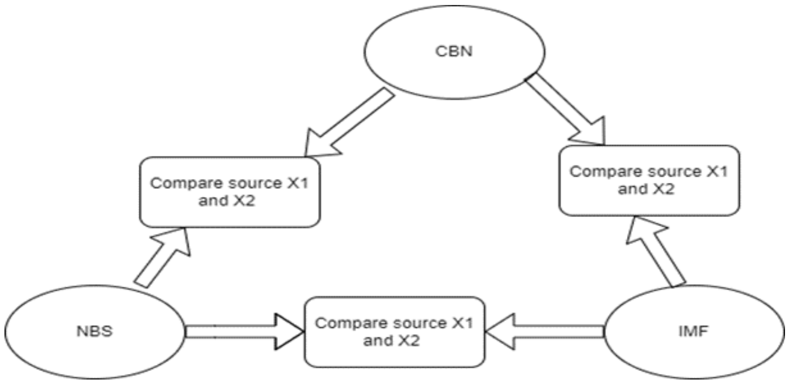

Figure 1: Data triangulation process

Economic indicators

Table 2: Key variables used in forecasting models were

Variable	Description
Real GDP	A measure of the total value of goods and services produced by an economy in a specific year, adjusted for inflation to reflect the true purchasing power and economic output during that period. It removes the effects of price changes, offering a clearer picture of an economy's performance.
Poverty level	Percentage of the population living below the poverty line, as defined by the World Bank.
Inflation Rate	Annual percentage change in the Consumer Price Index (CPI).
Unemployment Rate	The unemployment rate is the percentage of the total unemployed labor force actively seeking employment and willing to work.
Exchange Rate	Value of the Nigerian Naira against the US Dollar.
Monetary Policy Rate (MPR)	Key interest rate set by CBN to guide lending and borrowing rates within the economy.
Total Importation	The total value of goods and services purchased from other countries during a specific time period, typically measured annually.
Total Exportation	Total value of goods and services sold by a country to other nations over a specific period, often calculated annually.
Total Government Expenditure	The sum of all government spending on goods and services, including consumption, investment, and transfer payments such as pensions and welfare in all levels of government in a given year.

Data Processing

Before feeding the data into the ML models, the raw data was processed to ensure quality and consistency. Normalization was employed to standardize the values between 1 and 0, since the economic indicators were of different scales ensuring the model treats all variables with equal importance (Adekunle et al 2024). There was no missing and duplicate values.

$$X_{scaled} = \frac{X_i - Min(x)}{Max(x) - Min(x)} ----------(1)$$

Machine Learning Models

Model selection

We trained both supervised and unsupervised model to predict and to recognize the patters between the years. An artificial neural network was used to predict GDP and poverty rate. In a similar manner, we trained Hierarchical Clustering, K-means clustering, t-Distributed Stochastic Neighbor embedding (t-SNE), and DBSCAN Clustering.

$$\begin{bmatrix} GDP \\ Poverty\ rate \end{bmatrix} = \sum_{j=1}^{m} W_{kj} + bk ------(2)$$

Supervised Artificial Neural Network

A multilayer Layer Feed-forward Neural Network trained with back-propagation algorithm of Artificial Neural Network models was used in this study based on its ability to handle time series data and complex interactions between variables. The general mathematical structure of this model with an input of X and output of Y as seen below:

$$y_k = \sum_{j=1}^{m} W_{kj} + bk ---------(3)$$

Where:

$$x_1, x_2, \dots\dots, x_m = set\ of\ inputs$$
$$w_{k1}, w_{k2}, \dots\dots, w_{km} = The\ set\ of\ snaptic\ weight$$
$$b_k = Is\ the\ bias\ with\ weights\ (w_{ko})$$
$$y_k = Is\ the\ output.$$

Being a multilayer feed forward neural network models which allow unidirectional forward connections of inputs and outputs, it is mathematically represented as:

$$Y_k(x) = \varphi(w_o + \sum_{j=1}^{J} w_j \varphi(w_{oj} + \sum_{j=1}^{m} w_{ij} x_i)) ------(4)$$
$$= \varphi(w_o + \sum_{j=1}^{J} w_j \varphi(w_{oj} + w_j^T x$$

Where:

$w_O =$ The intercept of the output neuron.

$w_{oj} =$ The intercept of the *j*th hidden neuron.

$W_j =$ The synaptic weights corresponding to the weights starting at the *j*th hidden neuron and leading to the output neuron.

$w_j = (w_{1j}, \dots, w_{mj}) =$ The vector of all synaptic weights corresponding to the synapses leading to the *j*th hidden neuron,

$X = (x_1, \dots, x_m) =$ The vector of all explanatory variables.

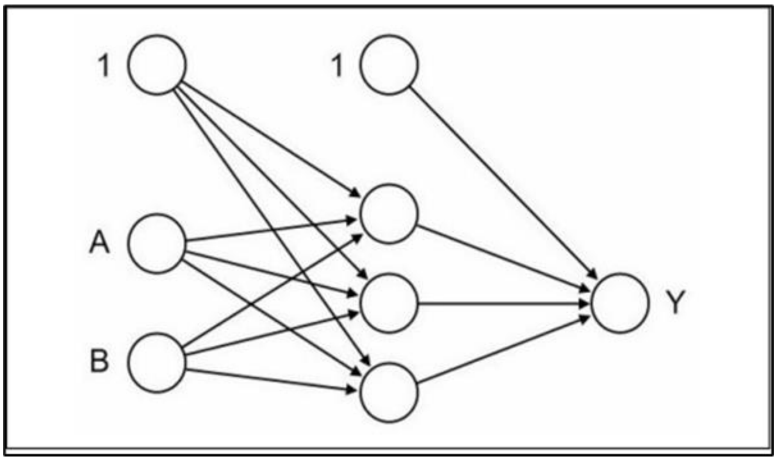

Figure 2: shows the Multi-layer Feed-forward Neural Network.

The Activation Function

The output of the models were obtained by applying sigmoid activation function which is commonly used because it is continuously differentiable. The logistic function whose domain is [0, 1], can be represented as:

$$\varphi(v) = \frac{1}{1 + \exp(-av)} = \begin{cases} a > 0 \\ -\infty < v < \infty, \end{cases} ------- (5)$$

Evaluation Of The Accuracy Of Measurement

The error function of Artificial Neural Network is given as:

$$E_k = T_k - Y_k --------- (6)$$

Where:

T_k is the observed output

Y_k is the predicted output

Therefore the Mean square error is expressed as:

$$MSE = \frac{1}{N}\sum_{k=1}^{N}(T_k - Y_k)^2 ------(7)$$

And $RMSE = \sqrt{MSE} ----------(8)$

In application of the back propagation, if E is the value of the cost function, then the rate of change in E with respect to the weights θ is given by:

$$\nabla\mathrm{E}(\theta) = \frac{\partial E}{\partial \theta_k} ------(9)$$

Where

∇ Is the gradient operator

$$\nabla = \left[\frac{\partial}{\partial\theta_1}, \frac{\partial}{\partial\theta_2}, \ldots\ldots, \frac{\partial}{\partial\theta_k}\right]^T$$

$\nabla E(\theta)$ is the gradient of the cost function

$$\nabla\mathrm{E}(\theta) = \left[\frac{\partial E}{\partial\theta_1}, \frac{\partial E}{\partial\theta_2}, \ldots\ldots, \frac{\partial E}{\partial\theta_k}\right]^T$$

θ_k Is the vector of all weights of the network at t^{th} iteration.

The network weights are determined by:

$$\theta_{k+1} = \theta_k + \Delta(\theta_k) ----- (10)$$

Where:

$\theta_k =$ Network weights of k^{th} iteration

θ_{k+1} = Parameters of $(k+1)^{th}$ iteration

$\Delta(\theta)_k$ = Learning process or the weights correction.

Unsupervised Machine Learning

Hierarchical Clustering, K-means clustering, t-Distributed Stochastic Neighbor embedding (t-SNE), and DBSCAN Clustering are said to be prominent techniques in unsupervised learning, each with unique characteristics suited to various types of data and clustering requirements. Each of this model were used for validity and unique purposes. The hierarchical clustering and DBSCAN was used to handle complex shapes for the patterns between the respective years, Hierarchical Clustering was used to organize data points into hierarchy by divisive (top-down) approaches. It builds clusters based on the distance or similarity between data points by year, which was visualized through a dendrogram to determine an optimal cut-off for clusters.

$$Distance(X,Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2} ----(11)$$

K-means Clustering, by contrast, is a partition-based approach that aims to divide data into a predefined number of clusters k, based on the proximity of points to the nearest cluster mean or centroid. Known for its speed and efficiency, K-means is commonly applied in customer segmentation and image compression, though it can be sensitive to outliers and requires spherical clusters of similar sizes to perform well. K-means is spherical clusters, and t-SNE for data visualization in high-dimensional spaces.

$$Objective: \sum_{i=1}^{k}\sum\nolimits_{x\epsilon C_i}\|x - \mu_i\|^2 ----- (12)$$

Where C_i represent each cluster, and μ_i is the centroid of C_i

On the other hand, t-SNE, a powerful dimensionality reduction technique, is primarily used for visualizing high-dimensional data in lower-dimensional spaces (often 2D or 3D) by preserving local data structures. It's especially helpful in making complex, high-dimensional relationships interpretable.

$$KL(P\rrbracket Q) = \sum_{i\neq i} P_{ij} log\frac{P_{ij}}{Q_{ij}} ----- (13)$$

Where P_{ij} and Q_{ij} represent pairwise similarities in the original and embedded spaces

Lastly, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a density-based method that groups dense regions of data points while identifying noise in sparser areas. Unlike K-means and hierarchical clustering, DBSCAN does not require the number of clusters to be specified beforehand and can identify clusters of arbitrary shapes, making it ideal for spatial data analysis and anomaly detection.

$$Core\ Points: |N_{\epsilon}(p)| \geq minPts \ldots\ldots\ldots\ldots(14)$$

IV. Results

Table 3: Summary Statistics

	Mean	SD	Min	Median	Max	IQR	CV
GDP	7.075192e+13	3.052713e+12	6.715279e+13	6.990715e+13	7.668494e+13	3.517993e+12	0
Importation	1.497019e+13	1.02856e+13	8.81756e+11	1.293304e+13	3.591762e+13	1.194324e+13	0.7
Export	1.799367e+13	8.26786e+12	8.52743e+12	1.741804e+13	3.596239e+13	6.32954e+12	0.5
Inflation	14.6	4.9	8.1	14.4	24.5	5.3	0.3
Exchange	334.7	137.1	157	306.5	644.3	122.2	0.4
MPR	13.5	1.9	11.5	13.2	18.4	1.4	0.1
Gov_Exp	7.437606e+12	5.558159e+12	971222352634	6.157629e+12	1.980844e+13	4.939908e+12	0.7
Unemployment	4.8	0.7	3.8	5	5.6	1	0.1
Poverty	39.4	0.7	38.3	39.5	40	11	0

For the past 10yrs [2014 to 2023], Nigeria incurred an average GDP of 70.75 trillion [SD: 3.05 trillion, Range: 67.15 trillion to 76.68 trillion, IQR: 3.52 trillion, CV: 0] showing that the GDP has been relatively consistent. The Importation metric has an average of 14.97 trillion [SD: 10.29 trillion, Range: 0.88 trillion to 35.92 trillion, CV: 0.7] showing significant fluctuation and high variability in import volumes. In contrast, the average exports was 17.99 trillion [CV; 0.5] indicating that exports are more stable than imports. The average Inflation was 14.6% [SD: 4.9%, Range: 8.1% to 24.5%, IQR: 5.3%, CV: 0.3] for the 10years indicating that inflation has been fairly volatile. The average exchange rate was 334.7[SD: 137.1, Range: 157 to 644.3. CV: 0.4], suggesting notable fluctuations. The average Monetary Policy Rate (MPR) was 13.5% [SD of 1.9% and CV: 0.1], while the average government expenditure was 7.44 trillion [SD: 5.56 trillion, CV: 0.7]. This indicates significant swings in spending over time. The average unemployment was 4.8% [SD: 0.7%, CV: 0.1]. Finally, the average poverty was 39.4% [SD: 0.7%, CV: 0], suggesting that poverty rates have remained consistently high over the period.

Table 3: Correlation Matrix

	Real Gdp (N)	Importation (N)	Export (N)	Inflation (%)	Exchange Rate(N)	Mpr	Government Expenditure(N)	Unemployment Rate(N)	Poverty Rate
Real Gdp (N)	1	0.97	0.90	0.78	0.92	0.65	0.48	0.28	0.81
Importation (N)	0.97	1	0.95	0.74	0.91	0.67	0.53	0.32	0.77
Export (N)	0.90	0.95	1	0.71	0.85	0.76	0.47	0.17	0.58
Inflation (%)	0.79	0.74	0.71	1	0.91	0.69	0.62	0.33	0.53
Exchange Rate(N)	0.92	0.91	0.85	0.91	1	0.75	0.71	0.51	0.74
Mpr	0.65	0.67	0.76	0.69	0.75	1	0.56	0.29	0.28
Government Expenditure(N)	0.46	0.53	0.47	0.63	0.71	0.56	1	0.61	0.30
Unemployment Rate(N)	0.28	0.32	0.17	0.33	0.51	0.29	0.60	1	0.54
Poverty Rate	0.81	0.77	0.58	0.53	0.74	0.28	0.30	0.54	1

This correlation result (Table 3) shows that real GDP is highly correlated with Importation (0.97) and Exchange Rate (0.92), Exports (0.90), Poverty Rate (0.81) and Inflation (0.78). Importation shows a similarly high correlation with Exports (0.95) and Exchange Rate (0.91), Inflation (0.74) and Poverty Rate (0.77). Export has a strong correlation with the exchange rate (0.85), MPR (0.76), Exchange Rate (0.91), and Government Expenditure (0.62). Exchange Rate itself has moderate to high correlations with most variables, especially with MPR (0.75) and Government Expenditure (0.71). The Monetary Policy Rate (MPR) is notably correlated with Export (0.76) and the Exchange Rate (0.75). Government Expenditure shows moderate correlations with Exchange Rate (0.71) and Inflation (0.63), Unemployment Rate (0.60). The Unemployment Rate has relatively low correlations with most variables but shows a moderate association with Poverty Rate (0.54) and Government Expenditure (0.60). Finally, Poverty Rate has moderate to strong correlations with several variables, including Exchange Rate (0.74), Inflation (0.53), and Unemployment Rate (0.54).

V. Artificial Neural Networks

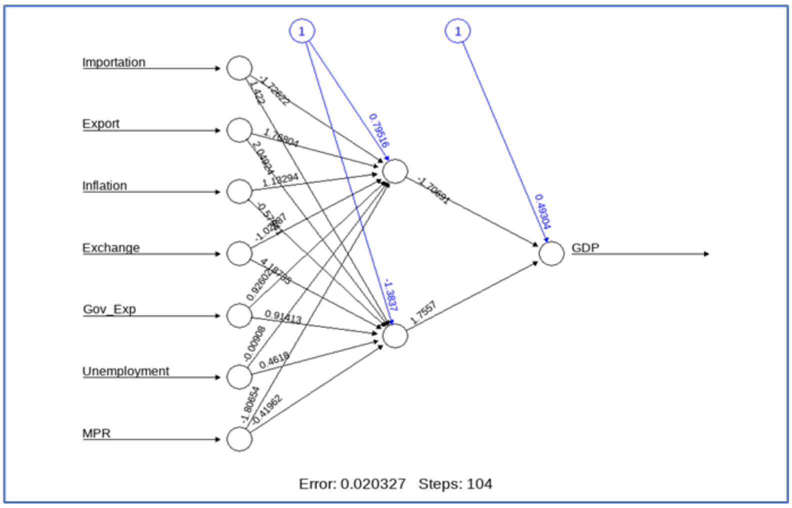

Figure 3: Real GDP as the dependent variable (y1)

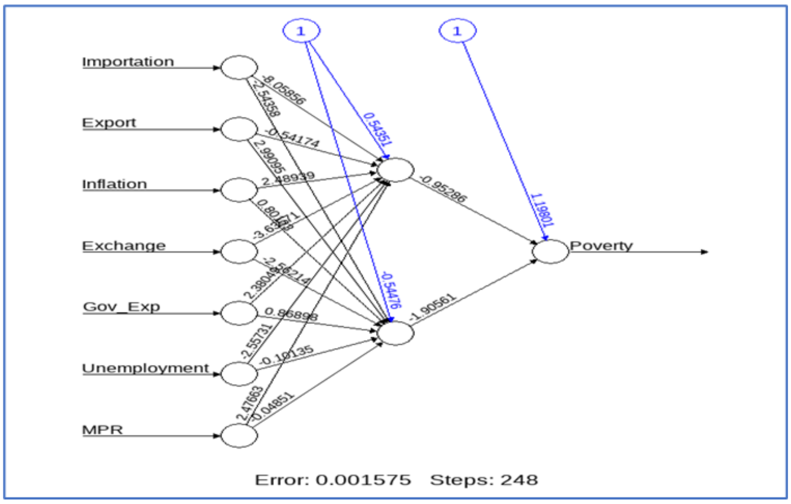

Figure 4: Poverty level as the dependent variable (y2)

Table 4: GDP as the dependent variable (y1)

Predicted	Observed	Error	Percent
6.75E+13	6.72E+13	-3.4E+11	100.5137
6.89E+13	6.90E+13	1.03E+11	99.85085
6.75E+13	6.79E+13	4.41E+11	99.35129
6.87E+13	6.85E+13	-2.2E+11	100.3234
6.98E+13	6.98E+13	4.01E+10	99.94262
7.14E+13	7.14E+13	-3.2E+10	100.0445
7E+13	7.00E+13	-6.6E+09	100.0094
7.24E+13	7.24E+13	-4.1E+10	100.0566
7.46E+13	7.46E+13	2.55E+10	99.96584
7.67E+13	7.67E+13	3.19E+10	99.95839

Table 5: Poverty Rate as dependent variable (y2)

Predicted	Observed	Error	Percent
38.30229	38.3	-0.00229	100.006
38.90221	38.9	-0.00221	100.0057
38.57396	38.6	0.026043	99.93253
38.92488	38.9	-0.02488	100.064
38.99992	39	8.15E-05	99.99979
40.00077	40.01	0.009228	99.97694
40.01188	40.01	-0.00188	100.0047
40.01655	40.01	-0.00655	100.0164
40.01504	40.01	-0.00504	100.0126
40.00206	40.01	0.007936	99.98017

These tables show predicted vs. observed values for two dependent variables—GDP (Table 4) and Poverty Rate (Table 5)—along with associated error values and the percentage accuracy for each prediction. The predicted GDP values closely match the observed values, with minimal error and percentage deviations hovering around 100%, indicating high prediction accuracy. A prediction of 6.75E+13 against an observed 6.72E+13 yields a minor error of -3.4E+11 and a percentage of 100.5137, showing slight overestimation, while 6.89E+13 (observed 6.90E+13), show a slight underestimation at 99.85%. Similarly, the Poverty Rate predictions are highly accurate, with errors as low as ±0.00229 and percentage accuracy near 100%. The predicted 38.30229 (observed 38.3) with an accuracy of 100.006%, reflecting an almost exact match, further underscoring the model's precision in poverty rate predictions

Table 6: Cluster Means

GDP	Importation	Export	Inflation	Exchange	MPR	Gov_Exp	Unemployment	Poverty
0.4444961	0.3940358	0.1646372	0.09386171	0.2706677	- 0.3782319	- 0.2198615	0.3028608	0.9118852
1.9435249	2.0365788	2.1733222	2.00744367	2.2583021	2.5611243	2.2257068	0.9366407	0.9118852
- 0.7443019	-0.7225444	- 0.5663742	- 0.47657810	- 0.6681946	- 0.2096394	- 0.2692522	-0.4296168	- 0.9118852
Within cluster sum of squares by cluster								
1st Cluster	2nd Cluster	3rd Cluster						
13.919135	0.000000	9.759192						

K-means clustering with 3 clusters of sizes 4, 1, 5 (between_SS / total_SS = 70.8 %)

The table summarizes the results of a K-means clustering analysis with three clusters, highlighting the average values of key economic indicators. Cluster 1 shows moderate economic levels across most indicators, with a notable mean for GDP (0.4445) and a high poverty rate (0.9119), Importation(0.3940) and Export levels(0.1646) suggest a mid-range economic status, while the lower Inflation (0.0939) and MPR (-0.3782) values hint at relatively mild economic pressures and policy rates. Cluster 2 stands out with the highest mean values across all variables, reflecting a group with high economic performance -- GDP (1.9435), Importation (2.0366), and Export (2.1733), representing a high-income group. Government expenditure (2.2257), MPR (2.5611), and poverty (0.9119). Cluster 3 has negative means for most indicators (e.g., GDP -0.7443, Importation -0.7225, and Exchange Rate -0.6682), representing a lower economic status group with distinctive negative poverty levels (-0.9119) compared to other clusters. The within-cluster sum of squares (WCSS) values indicate variance within each cluster, with Clusters 1 and 3 showing the largest variance (13.9191 and 9.7592, respectively). Cluster 2, being a single data point, has a WCSS of zero. The clustering solution explains 70.8% of the total variability (between_SS / total_SS), demonstrating that these clusters account for a significant portion of the data's overall variation.

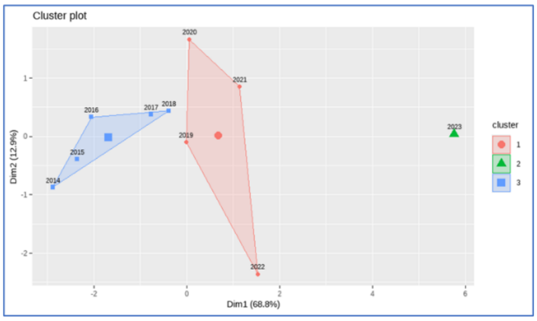

Figure 5: K means clustering

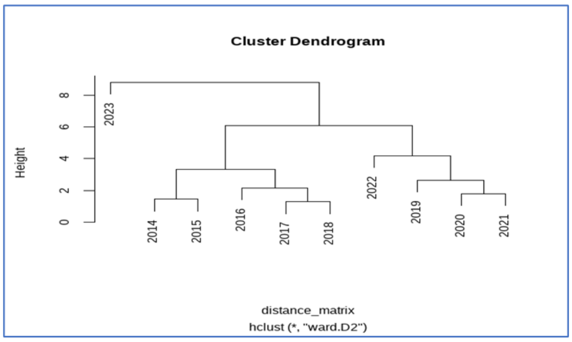

Figure 6

Table 7: Euclidean distance matrix and hierarchical clustering

1	2	3	4	5	6	7	8	9
1.444								
2.350	1.753							
2.938	2.308	1.479						
2.977	2.421	2.361	1.291					
3.958	3.157	3.521	2.602	2.030				
4.335	3.400	3.232	2.425	2.062	2.196			
4.861	4.117	3.982	3.066	2.687	2.698	1.797		
5.123	4.608	4.842	3.990	3.845	3.169	4.327	3.403	
8.750	8.240	7.937	6.639	6.253	6.453	6.349	5.390	5.421

Table 8: t-Distributed Stochastic Neighbor Embedding (t-SNE)

Dim1	Dim2	Cost	Cluster	Year
249.0004	-72.0755	0.0004235926	3	2014
220.6182	-23.4995	-0.001411715	3	2015
141.3915	-18.6747	0.0057425756	3	2016
72.80388	-6.28507	0.0077393198	3	2017
14.27075	6.054544	0.0210900252	3	2018
-61.5776	48.28622	0.0134484923	1	2019
-99.7059	-41.3805	0.0081321322	1	2020
-154.195	-23.4941	0.0100240508	1	2021
-146.508	89.93022	0.0103441799	1	2022
-236.098	41.13848	0.0055759932	2	2023

Table 7 shows varying distances between pairs of observations. The distance between the first and second data points is 1.4436, whereas the distance between the first and last data point is 8.7496, highlighting the variation in the proximity of these observations. The distances suggest that some observations are more similar to each other than others, which influence the clustering structure and identification of groups. Table 8 displays the t-SNE results, which map high-dimensional data into two dimensions for visualization. The observations from the years 2014 to 2018 belong to Cluster 3, suggesting these data points share similar patterns. However, starting from 2019, the observations shift into Clusters 1 and 2, indicating a change in the economic conditions represented by these years.

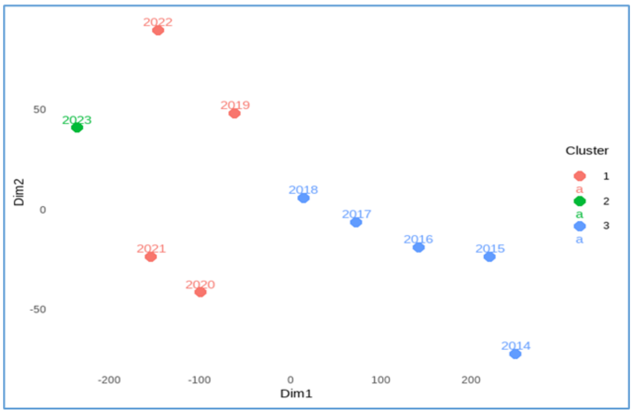

Figure 7: t-Distributed Stochastic Neighbor Embedding (t-SNE)

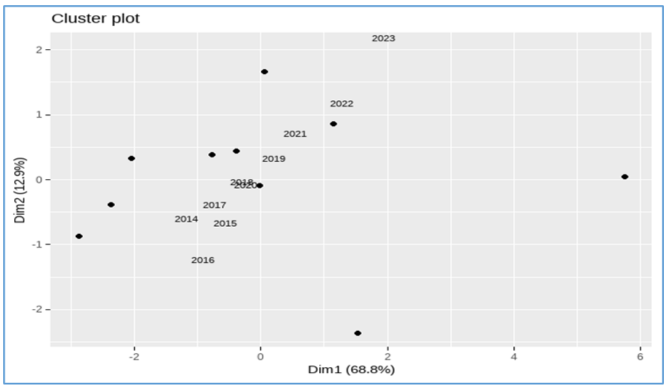

Figure 8: DBSCAN Clustering

VI. Discussion

The results from the data analysis provide a comprehensive overview of the relationships between key economic indicators and their accuracy in predicting GDP and Poverty Rate values. Each results offers distinct insights into correlations, prediction accuracy, and clustering, revealing patterns and interdependencies between economic variables for ten years. The high mean values and low variation for GDP indicate stable economic growth with minimal volatility over time, while high variability in Importation and Export suggests trade fluctuations, likely due to shifts in global demand or exchange rate impacts. Government Expenditure's high variability points to fluctuating fiscal policies, possibly responding to infrastructure or welfare needs. The unpredictable nature of government spending calls for better planning through multi-year budgets, focusing on critical areas like infrastructure and education. Unemployment's low variability suggests a stable labor market, while Poverty's zero variability indicates entrenched poverty rates that might result from structural economic challenges. This persistent poverty levels underline the need for targeted programs that use predictive data to direct resources where they are needed most. Expanding access to education, job creation, skills training and social protection can help to break the cycle of poverty. A strong positive correlation between GDP and Importation (0.97) suggests that economic growth drives import demand, likely for raw materials or consumer goods, while the strong link between GDP and Export (0.90) points to a competitive export sector supporting GDP growth. This strong link between GDP and trade highlights the need to reduce reliance on imports by boosting local production and diversifying exports. Supporting non-oil industries with incentives like tax breaks can make exports more competitive. Additionally, a high correlation between Inflation and Exchange Rate (0.91) suggests that currency devaluation is associated with rising inflation, impacting both import costs and domestic prices. To manage inflation, stabilizing exchange rates is key, as currency devaluation drives up costs. Prediction accuracy for GDP and Poverty Rate demonstrates a robust model, with predicted values aligning closely with observed values, indicating effective modeling of factors driving economic and poverty trends. A GDP prediction of 6.75E+13 against an observed 6.72E+13, and percentage accuracy near 100%, demonstrates the model's reliability in projecting economic growth, supporting planning and investment decisions. Similarly, for Poverty Rate predictions, the model consistently approximates real values (e.g., prediction of 38.30229 versus observed 38.3), helping policymakers target resources effectively to address poverty. Minimal deviations from observed values reflect the model's accuracy and utility in social planning. In the short term, Nigeria can use the predictive models to improve budgets and welfare programs. Over time, refining fiscal policies and stabilizing trade can build a stronger economy. Long-term goals include better economic planning, diversification, and updating the model regularly to adapt to changing realities. These steps can support sustainable growth and informed policymaking. K-means clustering groups economic data into clusters based on similarity, illustrating different economic phases. Cluster 1, with averages close to zero, represents a balanced economic state with stable growth, inflation, and unemployment, likely typical years. Cluster 2, characterized by high values, suggests periods of economic expansion with increased GDP, trade volumes, and spending, possibly due to stimulative fiscal policies. Cluster 3, with lower values, reflects downturns or recessions, marked by reduced economic activity potentially driven by external shocks or restrictive policies. The grouping done by Euclidean distance matrix and hierarchical clustering group highlights unique years, pointing to major policy shifts, global events, or economic crises that caused these differences. Such insights allow analysts to explore years that diverged economically, aiding future policy planning. Finally, t-Distributed Stochastic Neighbor embedding (t-SNE) offers a visual of economic clusters, with most years grouped closely, suggesting consistent economic conditions. However, we found out that, considering the economic indicators, 2014–2018 are similar, and 2019–2022 are similar, but 2023 is distinct from the other years. This unique economic trends in 2023 suggest a need for better preparation for unexpected changes, such as global shocks or policy shifts. Furthermore, t-SNE and other clustering techniques should be consistently use to detect the patterns and groupings in the economic data. This will help to know the economic policies best fit for each of the groups or patterns.

References

[1] Adekule, J.D., Oyeniran, M.I., Dada, A.M., Robert, C.O. Sule, H.S., And Akinpelu, T. T. (2024). A Smart Heart Diagnostic Machine: Developing An Effective Predictive Classification Algorithm. Global Scientific Journals Vol 12 (9) Pg. 804-824.

[2] Adewale, M.D., Ebem, D.U., Awodele, O. A., Sambo-Magaji, E.M., Okechalu, E.A., Donatus, R.E., Olayanju, K.A., Owolabi, A.F., Oju, J.U., Ubadike, O.C., Otu, G.A., Muhammed, U.I., Danjuma, O.R., And Oluyide, O.P. (2024). Predicting Gross Domestic Product Using The Ensemble Machine Learning Method. Systems And Soft Computing 6 200132.

[3] Agu, S.C., Onu, F.U., Ezemagu, U.K. And Oden, D. (2022). Predicting Gross Domestic Product To Macroeconomic Indicators. Intelligent Systems With Applications 14 200082.

[4] Athey, S., And Imbens, G. (2015). Machine Learning Methods For Estimating Heterogeneous Causal Effects. Annual Review Of Economics Vol 11, 575-598.

[5] Bakawu, M. A., Buba, A.T., And Malik, A. (2023). Forecasting The Real Gdp In Nigeria Using Possibilistic Linear Regression Model. Benin Journal Of Statistics Vol. 6, Pp. 59– 69 Issn 2682-5767.

[6] Barhoumi, K., Choi, S.M., Iyer, T., Li, J., Ouattara, F., Tiffin, A., And Yao, J. (2022) "Overcoming Data Sparsity: A Machine Learning Approach To Track The Real-Time Impact Of Covid-19 In Sub-Saharan Africa". International Monetary Fund Publication.

[7] Ben, K. And Patrick V. S. (1996). An Introduction To Neural Network. University Of Amsterdam Publication, Kruislaan 403, Nl1098 Sj Amsterdam Netherlands.

[8] Brett, L. (2015). Machine Learning With R. Second Edition. Packet Publishing, Birmingham Mumbai.

[9] Bishop, C. (2006). Pattern Recognition And Machine Learning. Journal Of Electronic Imaging 16(4): 140-155.

[10] Central Bank Of Nigeria. (2024). Introductory Text For Economic And Financial Review Vol 60 (4) Https://Dc.Cbn.Gov.Ng/Efr/

[11] Frank, A. G. (1967). Capitalism And Underdevelopment In Latin America. Monthly Review Press.

[12] Frauke, G. And Stefan, F. (2015). Neuralnet: Training Of Neural Networks. The R Journals. Vol 2/1 Pp30-38.

[13] Gosasang, V., Chanraprakaikul, W. And Kiattsi, S. (2011). A Comparison Of Traditional And Neural Networks Forecasting Techniques For Container Throughput At Bangkok Port. The Asian Journal Of Shipping And Logistics. Vol 27 Number 3 December 2011 Pp. 463-482.

[14] Guo, S. (2002).Three Enrollment Forecasting Models: Issues In Enrollment Projection For Community Colleges. Presented At The 40thrp Conference, May 1-3 2002, Asilomar Conferencegrounds, And Pacific Grove, California.

[15] Hassan, A.A., Muse, A.H., Chesneau, C. (2024). Machine Learning Study Using 2020 Sdhs Data To Determine Poverty Determinants In Somalia. Nature Scientific Reports. 14:5956

[16] Haykin, S. (1999). Neural Networks: A Comprehensive Foundation. 2nd Edition, Pretence Hall, Upper Saddle River, N.J.

[17] Ibrahim, M. S., And Ren, D. (2021). Nigeria Poverty Analysis, Prediction Using Machine Learning Methods And Deep Learning. Asia Journals Of Social Science, Arts And Humanities. Vol 9 No1 2021 Issn 23113782

[18] International Monetary Fund Reports And Publication (2023). Https://Www.Imf.Org/En

[19] Karabell, Z. (2014). The Leading Indicators: A Short History Of The Numbers That Rule Our World. Simon & Schuster Publishing.

[20] Kim, J. Y. (2021). Using Machine Learning To Predict Poverty Level In Costa Rican Households. Johns Hopkins University, Carey Business School Washington, Dc.

[21] Kuhe, D. A., Chiawa, M.A., And Aboiyar, T. (2016). Time Series Model Of Poverty Indices In Nigeria. Journal Of Scientific And Engineering Research, 3(2):261-283.

[22] Kuznets, S. (1955). Economic Growth And Income Inequality. The American Economic Review 45(1), 1-28.

[23] Leamer, E. E. (2010). Tantalus On The Road To Asymptopia. Journal Of Economic Perspectives, 24(2), 31-46.

[24] Masini, R.P., Medeiros, M.C., And Mendes, E.F. (2021) Machine Learning Advances For Time Series Forecasting. Journal Of Economic Surveys 37:76–111.

[25] Nigeria Bureau Of Statistics (2023). Annual Report. Https://Www.Nigerianstat.Gov.Ng

[26] Okwu, A. T., Oseni, I. O., Aberu, F., And Obiakor, R. T. (2017). Real Gross Domestic Product Of Nigeria (2017-2030): Autoregressive Integrated Moving Average (Arima) Model-Based Forecasts. Babcock Journal Of Economics (Bjoe) Vol 6 Pp 100-113 Issn 2734-2239.

[27] Ravallion, M., And Chen, S. (2003). "Measuring Pro-Poor Growth." Economics Letters, 78(1), 93-99.

[28] Robert, L.E.J. (1988) On The Mechanics Of Economics Development. Journal Of Monetary Economics Development. Vol 22 Issue 1pg 3-42.

[29] Romer, P. M. (1990). Endogenous Technological Change. Journal Of Political Economy, 98(5), S71-S102.

[30] Roode, D.D. (2022) “Forecasting Nigerian Gdp Growth Trough Factor Models And Machine Learning Techniques”. A Bachelor’s Thesis Submitted To Erasmus School Of Economics, Erasmus University Rotterdam.

[31] Smith, A. (1776). The Wealth Of Nations. W. Strahan And T. Cadell, London.

[32] United Nations Development Programme. (2020).The Next Frontier: Human Development And The Anthropocene. Human Development Report 2020

[33] Usmanova, A., Aziz, A., Rakhmonov, D., And Osamy, W. (2022) Utilities Of Artificial Intelligence In Poverty Prediction: A Review. Sustainability, 14, 14238.

[34] World Bank Economic Reviews. Https://Www.Worldbank.Org/En/Country/Nigeria

[35] Yassen, A.A. (2011). “Comparative Study Of Artificial Neural Network And Arima Models For Economic Forecasting”. A Thesis Submitted In Partial Fulfillment Of Requirements For The Degree Of Master Of Statistics, Al- Azhar University Gaza.