Engineering and Technology Journal e-ISSN: 2456-3358

Volume 10 Issue 02 February-2025, Page No.- 3942-3952

DOI: 10.47191/etj/v10i02.24, I.F. – 8.482

Long-Term Load Forecasting Incorporating GDP and Population Dynamics

Marcel Uche Ezeh[1], Emmanuel U. Ubeku[2], Smith O. Otuagoma[3]

[1,2,3]Electrical and Electronic Engineering Department, Delta State University, Abraka, Nigeria.

ABSTRACT: This research paper seeks to create a dependable long-term load forecasting (LTLF) model for Nigeria that includes socio-economic indicators like Gross Domestic Product (GDP) and population trends. Accurate electricity demand forecasts help policymakers plan infrastructure expansion to meet demand increases due to population growth and socio-economic development. A multi-model approach integrating Linear Regression with Support Vector Machines and Artificial Neural Networks was used to reach this objective. The performance of these models was assessed through implementation and comparison on MATLAB and Python platforms to determine predictive capabilities. The Artificial Neural Network model achieved the best results among its counterparts by having the lowest RMSE and MAPE values. By employing a multi-model approach, the test results showed that this approach gave an RMSE less than 1000GWh compared to all forecast methods giving an RMSE less than 9000GWh and greater than 1500GWh thereby establishing it as the most dependable forecast technique with superior error metrics. The findings encompass electricity load predictions from 2024 through 2035 for every model.

KEYWORDS: Linear Regression, Support Vector Machine, Artificial Neural Network

1. INTRODUCTION

In recent years, the role of reliable and affordable electricity in fostering economic, social, and political growth has become increasingly critical [1]. Electricity is "prime mover that propels the socioeconomic development of every country and any nation that does not pay attention or ignores its power sector does so at its peril." [2]. It means that any nation whose energy need is insufficient and inadequate in supply; prolong their development and risk of losing potential investors [3]. Within this context, accurate load prediction has gained more significance.

Load forecasting is "the process of predicting how much power will be needed at a given time and how that demand will affect the utility grid. It is used to ensure that enough power is available to meet consumption needs while avoiding waste and inefficiency." [4]. It is a critical function in the planning and operation of power systems, especially in rapidly growing economies like Nigeria. It is extremely important for energy suppliers, financial institutions and other users in electric energy generation, transmission, distribution and market. Accurate load prediction will assist the utilities and policymakers to optimize grid operation, plan infrastructure investment, reduce cost, and integrate renewable energy sources which will in turn support economic growth and improve quality of life.

Nigeria contends with a continuous power crisis which manifests through insufficient electricity supply and unpredictable power availability along with regular blackouts and load shedding that creates an energy shortfall. Nigeria's economic growth faces major obstacles because of inadequate power supply that affects multiple economic sectors and slows industrialization while negatively impacting citizens' quality of life.

With a 2022 population of approximately 216.8 million with anticipated increase rate of 2.41% from 2021 [5], making it the most populous country in Africa and the sixth most populous in the world. The country needs increasing energy needs to support its large and growing populace. Economic growth, measured by Gross Domestic Product (GDP), also intensifies electricity consumption as industries expand and consumer wealth rises [6]. However, despite abundant natural resources like oil, natural gas, minerals, and a favorable climate for agriculture, Nigeria's energy sector struggles to meet demand. Electricity generation which supposed to propel economic activities in Nigeria is still low [7]. In 2020, around 35.7 thousand Gigawatt hours of electricity were generated. This was low in comparison to the level of electricity demand, which exceeded 29 Terawatt hours in the same year. Integrating GDP and population dynamics into load prediction models could improve forecasting accuracy and help address Nigeria's energy challenges [8].

Population dynamics and economic growth are two key drivers of electricity demand and their integration into load prediction models is essential in accurate and reliable predicting. They significantly influence the long-term trends in energy consumption, particularly in a fast-growing and developing countries like Nigeria. Understanding their roles allows for better anticipation of future energy needs and more effective planning in the power sector.

The growing worldwide energy demand due to economic development and population expansion shows how essential reliable long-term load prediction models become. Modern research indicates that the intricate connection between these factors and long-term load prediction remains largely unknown [9]. The existing models often fail to integrate key socio-economic factors particularly GDP and population dynamics, leading to inaccurate prediction and ineffective energy planning. These models typically focus on historical consumption trends, technical developments, and environmental concerns, overlooking the complex interdependencies between demographics and economics variables that impact energy demand [10].

This study therefore aims at developing an improved long-term load prediction model for Nigeria, incorporating socio-economic factors to enhance accuracy. By creating sophisticated models that include demographic and economic trends; this research can contribute to more resilient energy planning, supporting sustainability development and economic security in Nigeria.

1.1. Methods for Integrating Population and GDP Dynamics

A. Econometric Models

Econometric models are a common approach to integrating population and GDP dynamics into load forecasting. These models use statistical methods to quantify the relationship between electricity demand and its driving factors, such as population and GDP. Key econometric techniques include:

Regression Analysis

This technique models the relationship between electricity demands (dependent variable) and independent variables such as population and GDP. For instance, a simple linear regression model may take the form:

$$E_t = \alpha + \beta_1 P_t + \beta_2 GDP_t + \epsilon_t \quad (1)$$

Where E_t is the electricity demand at time t, α is the intercept (the constant term), β_1 is the coefficient for population, β_2 is the coefficient for GDP, P_t is the population at time t, GDP_t is the gross domestic product at time t, and ϵ_t is the error term at time t.

Since $E_t = \alpha + \beta_1 P_t + \beta_2 GDP_t + \epsilon_t$ is the actual Model.
The Estimated model will be $\hat{E}_t = \hat{\alpha} + \hat{\beta}_1 P_t + \hat{\beta}_2 GDP_t + \hat{\epsilon}_t$ (2)

$$E_t = \hat{E}_t + \hat{\epsilon}_t \quad (3)$$

$$\hat{\epsilon}_t = E_t - \hat{E}_t \quad (4)$$

$$\hat{\epsilon}_t = E_t - \hat{\alpha} - \hat{\beta}_1 P_t - \hat{\beta}_2 GDP_t \quad (5)$$

Using the method of OLS, $\sum \hat{\epsilon}_t^{\,2} = \sum (E_t - \hat{\alpha} - \hat{\beta}_1 P_t - \hat{\beta}_2 GDP_t)^2$ (6)

Differentiating $\sum \hat{\epsilon}_t^{\,2}$ W.R.T. $\hat{\alpha}$, $\hat{\beta}_1$ and $\hat{\beta}_2$

$$\frac{\partial(\sum \hat{\epsilon}_t^{\,2})}{\partial \hat{\alpha}} = \sum E_t = n\,\hat{\alpha} + \hat{\beta}_1 \sum P_t + \hat{\beta}_2 \sum GDP_t \quad (7)$$

$$\frac{\partial(\sum \hat{\epsilon}_t^{\,2})}{\partial \hat{\beta}_1} = \sum E_t P_t = \hat{\alpha} \sum P_t + \hat{\beta}_1 \sum P_t^{\,2} + \hat{\beta}_2 \sum P_t . GDP_t \quad (8)$$

$$\frac{\partial(\sum \hat{\epsilon}_t^{\,2})}{\partial \hat{\beta}_2} = \sum E_t GDP_t = \hat{\alpha} \sum GDP_t + \hat{\beta}_1 \sum P_t . GDP_t + \hat{\beta}_2 GDP_t^{\,2} \quad (9)$$

Solve simultaneously to obtain the estimates of $\hat{\alpha}$, $\hat{\beta}_1$, and $\hat{\beta}_2$.

B. Machine Learning and Artificial Intelligence

The latest breakthroughs in machine learning along with artificial intelligence have created robust tools to predict future energy demand accurately. Artificial Neural networks together with Support Vector Machines and ensemble methods enable effective management of large datasets to reveal intricate patterns. Population and GDP dynamics become part of these models through historical data training which enables understanding of nonlinear relationships between input variables and electricity demand.

Artificial intelligence constitutes a computer science discipline focused on creating intelligent computer behaviors while Machine learning involves developing algorithms that learn from experience and has remained a fundamental component of AI research from its beginning. Machine learning investigates algorithmic development which enables learning from data to produce predictions. These algorithms create models from example inputs to enable data-driven predictions or decisions instead of following predefined static program instructions.

Support Vector Machine

Support Vector Machine is a machine learning algorithm that uses supervised learning models to solve complex classification, regression, and outlier detention problems by performing optimal data transformations that determine boundaries between data points based on predefined classes, labels or outputs. The main objective of the SVM algorithm is to find the optimal hyperplane in an N-dimensional space that can separate the data points in different classes in the feature space [11]. The dimension of the hyperplane depends upon the number of features. For instance, if the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. In mathematical context, an SVM uses kernel methods to transform data features by employing kernel functions [12].

Artificial Neural Network

Engineers have been utilizing artificial neural network (ANN), one of the most effective and adaptable tools offered by artificial intelligence, for many years in a variety of applications. A simple mathematical model of brain functions is provided by ANNs, which are computational tools. They can be used for tasks like modeling, categorization and prediction when combined with raw data and a learning system. In the majority of cases, an ANN is an adaptive system that modifies its structure in response to input coming from the outside or inside the network during the learning period [13].

The term "Artificial neural network" refers to a network of nodes that mimics the biological neural networks. Simple artificial nodes, also known as "neurons", "neurodes" , "processing elements"(PEs), or "units" are connected together to create an artificial neural network. The biological network and this share some similarities. Currently, neural network models used in statistics, cognitive psychology, and artificial intelligence are most often referred to as artificial neural networks (ANNs).Theoretical and computational neurosciences are concerned with neural network models created with simulation of the central nervous system (CNS) in mind.

The ANN consists of interconnected processing units called node and this node is the same with the neuron of the human brain. The connecting link between these artificial nodes is similar to the axon synapse dendrite connections of the human brain. In mathematical terms, the performance of the neuron k can be illustrated as follows:

$$Y_k = \sum_{k=0}^{n} \left(W_{kj} X_j + b_k\right)^{\varphi} (.) \tag{10}$$

WhereX_1, X_2, ….X_n Are the input signals (soma); W_{k1}, W_{k2}, … W_{kn} are the synaptic weights (dendrites); b_k is the bias; $^{\varphi}$ (.)is the activation function, and Y_k is the output (axon) of the neuron. In figure 1, a neuron k receives more than one input signals X_1, X_2, X_3 and multiplies each of the input signals by synaptic weights W_{k1}, W_{k2}, W_{k3} respectively, sums them and adds a bias b_k, and then applies an activation function $^{\varphi}$ (.) to produce an output Y_k.

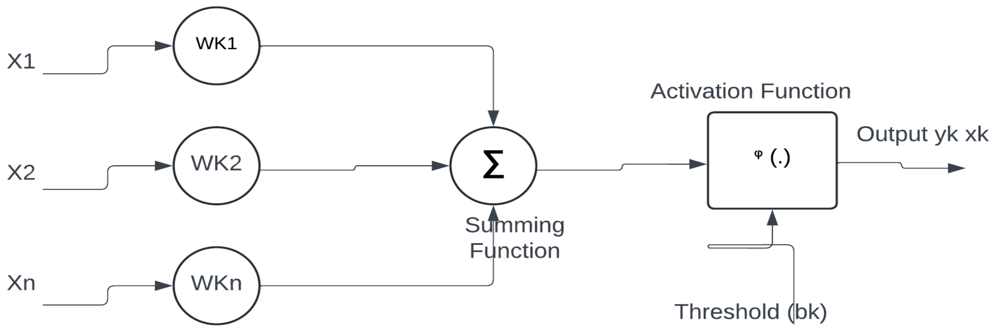

Figure 1: Mathematical Model of Artificial Neural Network

2. METHODOLOGY

The Research methodology is grounded in a data-driven approach, combining historical data analysis with statistical and machine learning models. The study utilized annual data from 2010 to 2023, sourced from the National Bureau of Statistics and United Nations World population Prospects. The dependent variable, Electricity load, was modeled using GDP and population as independent predictors.

2.1 Data Collection

Load demand data which is the dependent Variable deployed for the entire study is the yearly electricity consumption of Nigeria, which spans from 2010 to 2023, a total period of 13 years. This data was obtained from the National Bureau of statistics website (https://nigerianstat.gov.ng/)

Population and Gross domestic product (GDP) which is the independent variables are also included in the dataset collection process. This data was obtained from the United Nations-world population prospect website (https://population.un.org/). Table 1 shows a proper array of the original data.

Table 1: Electricity Load Consumption, GDP and Population Data for Nigeria

YEAR	Load Data (GWH)	Load data (GW)	GDP (BS)	Population (N)
2010	32913.1	1371.4	366.9904171	160,952,853
2011	30955.3	1289.8	414.4666768	165,463,745
2012	32087.4	1336.97	463.9710182	170,075,932
2013	37782.6	1574.3	520.1171803	174,726,123
2014	42142.8	1755.95	574.1837634	179,379,016
2015	31592.97	1316.4	493.0266828	183,995,785
2016	28675.9	1194.8	404.6491253	188,666,931
2017	27398.2	1141.6	375.7457311	193,495,907
2018	28983	1207.6	421.7392515	198,387,623
2019	60939.6	2539.1	474.5174911	203,304,492
2020	35720.3	1488.3	432.1988982	208,327,405
2021	36397.92	1516.6	440.8389922	213,401,323
2022	35972.2	1337.5	472.6245974	218,541,212
2023	36768	1375	503.65	223,804,632

Sources: Nigeria National Bureau of Statistics (*https://nigerianstat.gov.ng/*) and United Nations –world Population prospect (*https://population.un.org/*)

2.2 Data Pre-Processing

Data Pre-Processing operations include normalization, Feature Engineering and Correlation. The raw data was first imported as a spreadsheet into MATLAB's Workspace using the MATLAB Import Wizard and as a comma-separated values (CSV) file into Python's environment using Pandas libraries. Feature Engineering is then applied to create new input features or modify existing ones to improve the performance of the models; this can be done both programmatically in Python or using the Regression Learner app in MATLAB.

2.3 Toolboxes and Software

MATLAB: Used for its advanced statistical functions and Neural Network toolbox.

Python: Selected for its flexibility, with libraries like Scikit-learn for regression and SVM and TensorfFlow for ANN implementation.

2.4 Model Selection and Training

The core of the methodology involves selecting and training predictive models. The chosen models are both Statistical and Machine learning-based to leverage their strengths in different aspects of prediction. The following subsections outline the procedures for using the selected toolboxes to model the real-world yearly electricity consumption scenario, with Nigeria as the case study.

Linear Regression Model

In this research, the use of linear regression was explored for long-term load prediction, utilizing two distinct programming environments: Python and MATLAB. Annual data points across several years were utilized, incorporating GDP and Population as independent variables (predictors), while the load is the dependent variable (response). The model was initialized using "LinearRegression( )" and trained using the "fit( )" function, after the training, the "predict( )" function was used to make predictions on the test data and future load values based on GDP and Population forecasts.

Support Vector Machine Model

This model's approach for load prediction using both programming environment focuses on the linear kernel for both implementations.

The "SVC( )" function with the kernel = 'linear' parameter was used to build the model and the hyperparameters, such as regularization parameter C, were tuned through cross-validation to enhance the accuracy of the model.

Artificial Neural Network Model

The ANN model provided uses a backpropagation algorithm for training. The forward pass computes the outputs, and the backward pass adjusts the weights to minimize the error. This process is efficiently handled by TensorFlow/ Keras through the compile and fit methods or the Levenberg-Marquardt (trainlm) algorithm. ANN model input will include the past system load data from the year, Population and GDP. In the ANN training stage, GDP and Population (N) are input independent variables or (Predictors), while the corresponding load data are the ANN targets or (responses).

Table 2: Designed Network Topology for the ANN Load Forecast.

ELEMENT OF TOPOLOGY	DESCRIPTIVE / VALUE	
	Python	MATLAB
Network	Open loop network for training and prediction.	Open loop network for training and close loop network for prediction.
Input nodes	2	2
Hidden Layer	2	1
Hidden neurons	3	10
Output nodes	1	1
Training function	Adam	Levenberg-Marquardt (trainlm)
Momentum Factor	0.9	0.9
Learning Rate	0.001	0.01

The inputs were fed into the ANN model; and after sufficient training the model is used to predict load-output. Subsequently, forecast GDP and Population data are applied as inputs to the trained ANN model to predict system load for the forecast period in this case, the next 12 years. Since the data of GDP and Population were needed for the forecast years (2024-2035), simple Linear progression equations that were based on historical data were applied to grow GDP and Population data for desired time steps (Years).

2.5 Choice of Toolboxes

The algorithm employed by the several primary classes of forecasting techniques were thoroughly examined, and three approaches were chosen since their model formulation algorithms do not significantly resemble one another. The selection of appropriate toolboxes for implementing linear regression, Support Vector Machines (SVM) and Artificial Neural Networks (ANN) is critical for ensuring the accuracy and efficiency of the models below are the toolboxes selected for each model:

Linear Regression Toolbox

This is one of the simplest and widely used models in load forecasting. The toolboxes used for linear regression implementation are:

1. Scikit-learn (Python): The “LinearRegression” class from the “scikit-learn” library was employed due to its reliability and efficiency in handling large datasets.
2. MATLAB statistics and Machine learning Toolbox: The “fitlm” function was chosen for linear regression, which allows flexibility in modeling, and also easy to integrate additional statistical tools.

Support Vector Machine (SVM) Toolbox

SVM is a powerful tool for classification and regression especially when dealing with non-linear data; it operates on a kernel based method. The toolboxes used for SVM implementation are:

1. Scikit-learn (Python): The SVC for classification or SVR for regression from scikit-learn provides optimized implementation of SVMs. It supports linear, polynomial, RBF, and sigmoid kernels allowing flexibility in kernel choice.
2. LIBSVM: This is integrated in both Python and MATLAB; LIBSVM is a widely used library for SVM. It also offers fast computation.
3. MATLAB’s “fitcsvm” function was used for the linear SVM model.

Artificial Neural Networks (ANN) Toolbox

Artificial Neural Networks are used for modeling complex non-linear relationships in load forecasting. The toolboxes used for ANN implementation are:

1. TensorFlow/ Keras (Python): TensorFlow with its Keras API, was selected due to its ease of model-building, it also has support for multi-layered networks. Keras offers a wide range of activation functions and layers for building robust ANN structures.
2. MATLAB Neural Network Fitting Toolbox: This toolbox offers a pre-built function like the (feedforwardnet and train) that simplifies the design and training of ANN.

All chosen toolboxes provide Cross-Validation support, graphical tools for visualizing network performance and training progress. By using these toolboxes, Measures were taken to ensure that the models are built on efficient platforms that are widely accepted in the research community and can also be useful for future model enhancements.

2.6 Data Splitting and Model Training

Data splitting is part of the data preparation process. It involves dividing the data set into unique subset for different action. The training data is split into training and validation sets using time-based splitting to prevent data leakage. Training data is used to estimate parameters. Validation data is used to select models. Test data is used to confirm the model performance.

This section will focus on the three representative forecasting techniques used, linear regression, SVM and Artificial neural network (ANN). In both linear regression and SVM, the data is split into multiple folds (5-fold) using cross-validation method. While the data for ANN is split into a training set, validation set and a test set using the Holdout validation process. The model is trained, the linear regression model and SVM is then used to calculate the errors on the validation data.

In ANN modeling, the parameters (weights and biases) are estimated iteratively with the objective to enhance the goodness of fit on the training data. Without any stopping criteria, such a training process may go on and on until the ANN perfectly fits the training data (assuming the ANN is large enough). Validation data is used to tell the algorithm when to stop updating the parameters. After each aforementioned update, the ANN is used to predict the validation data. The training stops when the prediction error starts increasing. The parameters corresponding to the lowest prediction error on the validation data are locked as the final ones for the ANN structure being tried.

2.7 Model Analysis and Evaluation

The entire original data was used for analysis so as to obtain more detailed information about the performance strengths/weaknesses of each of the three models. The fourteen input data points were fed as predictors to each developed model in turn, to get three different set of predictions. With the aid of the actual load; the Mean absolute percentage error (MAPE), Root mean squared error (RMSE), Mean absolute error (MAE) and R-Squared of forecast for each model were then computed and analyzed.

2.8 Optimal Model Chart

This is the chart derived from the information gathered from the developed model's analysis, i.e. Table 8 as derived from Tables 5 and Table 6. For each of the three models, there are two sets of predicted values: one from Python and one from MATLAB. The chart is being proposed as a guide to the multi-model forecast technique to ensure that each year of interest has its own optimal model for its forecast and is adequately compensated for any likely errors. This approach is expected to yield more reliable forecast over time compared to the method of using any one of its component models to forecast all the year(s). It comprises two sections the optimal model section and the error compensation section. The MAPE results of the predicted models are used for the Optimal Model while the Mean Absolute Error is used for the error compensation.

3. RESULT AND DISCUSSION

3.1. Performance Metrics Comparison

The Table 3 below summarizes the key performance metrics for each model in Python and MATLAB

Table 3: Comparison Result of Python and MATLAB Models

Model	Platform	MAPE(%)	RMSE(Gwh)	MAE(Gwh)	R-Squared
Linear Regression	Python	13.82	8201.13	5334.79	0.037
Linear Regression	MATLAB	11.4	7635.9	4451.6	0.216
SVM	Python	14.77	8934.06	5795.48	0.23
SVM	MATLAB	12.8	7652.8	5033.3	0.212
ANN	Python	7.16	3662.97	2505.97	0.793
ANN	MATLAB	3.23	1930.91	1117.64	0.943

These results show that both Python and MATLAB produced different outcomes.

Since each of these metrics provides a summary assessment of the models performance, with the MAPE and RMSE values providing a sense of overall error magnitude and R-squared offering insights into the fit and predictive reliability of the models, we can come to a conclusion that the python models showed generally higher MAPE values, indicating that the percentage errors are larger while the MATLAB's MAPE values are lower which implies that the model accuracy percentage wise is better during prediction.

The MATLAB's models also featured lower RMSE and higher R-squared values which indicates better data fitting capabilities and an excellent fit to the variance in the data especially in the ANN model; this makes the MATLAB's model more reliable for predictive tasks compared to its counterpart.

Table 4: Analysis of the developed models

YEAR	MEAN ABSOLUTE ERRORS (MAE) (ACTUAL LOAD- PREDICTED LOAD)						MEAN ABSOLUTE PERCENTAGE ERROR (MAPE) (%)					
	Linear Regression (PYTHON)	Linear Regression (MATLAB)	SVM (PYTHON)	SVM (MATLAB)	ANN (PYTHON)	ANN (MATLAB)	Linear Regression (PYTHON)	Linear Regression (MATLAB)	SVM (PYTHON)	SVM (MATLAB)	ANN (PYTHON)	ANN (MATLAB)
2010	7241.00	7606.50	839.90	4713.80	312.20	2045.10	22.00	23.11	2.55	14.32	0.95	6.21
2011	54.00	872.60	4755.30	2195.20	1363.40	85.70	0.17	2.82	15.36	7.09	4.40	0.28
2012	2768.00	3110.20	3630.80	1731.30	634.60	138.60	8.63	9.69	11.32	5.40	1.98	0.43
2013	1107.50	705.40	3456.30	2073.10	69.30	34.60	2.93	1.87	9.15	5.49	0.18	0.09
2014	859.40	770.00	9209.10	6311.10	135.10	51.80	2.04	1.83	21.85	14.98	0.32	0.12
2015	7206.30	4837.40	4129.50	4835.50	303.30	2.00	22.81	15.31	13.07	15.31	0.96	0.01
2016	4027.00	2503.90	5643.30	2470.90	7100.80	2124.90	14.04	8.73	19.68	8.62	24.76	7.41
2017	4444.40	4499.20	5526.50	6549.90	352.70	15.20	16.22	16.42	20.17	23.91	1.29	0.06
2018	6363.40	6100.70	6732.60	4374.10	1627.90	287.00	21.96	21.05	23.23	15.09	5.62	0.99
2019	26090.30	25305.10	28844.3	24956.7	3652.30	126.60	42.81	41.52	47.33	40.95	5.99	0.21
2020	342.10	474.30	2786.70	3627.60	5874.20	1201.30	0.96	1.33	7.80	10.16	16.44	3.36
2021	3479.30	293.60	4305.00	3645.20	2199.50	2417.00	9.56	0.81	11.83	10.01	6.04	6.64
2022	4664.20	4261.90	242.40	997.10	3098.60	1239.20	12.97	11.85	0.67	2.77	8.61	3.44
2023	6040.00	981.80	1035.30	1985.20	8360.00	5878.00	16.43	2.67	2.82	5.40	22.74	15.99

Table 4 presents the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) value for load data predictions in Nigeria from 2010-2023. The actual load represents the recorded values, while the predicted load was obtained using the Linear Regression, SVM and ANN models. The MAE indicates the average absolute deviation between actual and predicted loads, while the MAPE represents the percentage error relative to the actual load.

3.2. Optimal Model Chart Design

The Optimal Model

The yearly optimal model is derived as follows with the aid of the Mean Absolute Percentage Errors (MAPE) gotten from Table 4

Rule 1: For a given year, the MAPE of the three models are calculated over the entire sample and the model that produced the minimum MAPE for that year was adopted as the optimal model for that year.

Thus for 2010, we have:

Linear Regression: $\frac{1}{n}\sum_{i=1}^{n}\frac{|y_i-\hat{y}_i|}{y_i}X\ 100$ = 22% (Python) & 23.1% (MATLAB)

SVM: $\frac{1}{n}\sum_{i=1}^{n}\frac{|y_i-\hat{y}_i|}{y_i}X\ 100$ = 2.55% (Python) & 14.32% (MATLAB)

ANN: $\frac{1}{n}\sum_{i=1}^{n}\frac{|y_i-\hat{y}_i|}{y_i}X\ 100$ = 0.95% (Python) & 6.21% (MATLAB)

The minimum MAPE for the year 2010 is 0.95% as produced by the ANN (Python) model prediction. This model (ANN) is therefore adopted as the optimal model to forecast for the year 2010.

The optimal model for other consecutive years (2011, …., 2023) are similarly derived using rule 1 as shown in Table 5.

Table 5: Deriving the Yearly Optimal Models

	MAPE(%) of Model Forecasts (2010-2023)						Minimum MAPE	Model With the minimum MAPE
	Linear Regression (Python)	Linear Regression (MATLAB)	SVM (Python)	SVM (MATLAB)	ANN (Python)	ANN (MATLAB)		
2010	22	23.11	2.55	14.32	0.95	6.21	0.95	ANN (Python)
2011	0.17	2.82	15.36	7.09	4.40	0.28	0.17	Linear reg. (Python)
2012	8.63	9.69	11.32	5.40	1.98	0.43	0.43	ANN (MATLAB)
2013	2.93	1.87	9.15	5.49	0.18	0.09	0.09	ANN (MATLAB)
2014	2.04	1.83	21.85	14.98	0.32	0.12	0.12	ANN (MATLAB)
2015	22.81	15.31	13.07	15.31	0.96	0.01	0.01	ANN (MATLAB)
2016	14.04	8.73	19.68	8.62	24.76	7.41	7.41	ANN (MATLAB)
2017	16.22	16.42	20.17	23.91	1.29	0.06	0.06	ANN (MATLAB)
2018	21.96	21.05	23.23	15.09	5.62	0.99	0.99	ANN (MATLAB)
2019	42.81	41.52	47.33	40.95	5.99	0.21	0.21	ANN (MATLAB)
2020	0.96	1.33	7.80	10.16	16.44	3.36	0.96	Linear reg. (Python)
2021	9.56	0.81	11.83	10.01	6.04	6.64	0.81	Linear reg. (MATLAB)
2022	12.97	11.85	0.67	2.77	8.61	3.44	0.67	SVM (Python)
2023	16.43	2.67	2.82	5.40	22.74	15.99	2.67	Linear reg. (MATLAB)

3.3. Error Compensation

It is derived based on the Mean Absolute Error (MAE) and is used to adjust the predicted load value to improve accuracy. The error compensation necessary for each optimal model is derived as follows:

Compensation Rule Application

The magnitude of the error compensation for the optimal model forecast of a given year was taken as the arithmetic mean of the Mean absolute errors of prediction for that year taken over the whole sample.

1. If the error direction is positive (Predicted load > actual load), compensation is subtracted from the actual load to adjust the predicted load towards accuracy.
2. If the error direction is negative (Predicted load < actual load), compensation is added to the actual load to adjust the predicted load towards accuracy.

Table 6: Deriving the Error Compensation for the Yearly Optimal Models

Year	Optimal Model	MAPE (%)	Actual Load data (GWH)	Predicted load for the Optimal Model (GWH)	MAE (GWH)	Compensation Action
2010	ANN (Python)	0.95	32913.1	32600.9	312.2	Add
2011	Linear reg. (Python)	0.17	30955.3	30901.3	54	Add
2012	ANN (MATLAB)	0.43	32087.4	32226	138.60	Subtract
2013	ANN (MATLAB)	0.09	37782.6	37748	34.6	Add
2014	ANN (MATLAB)	0.12	42142.8	42091	51.8	Add
2015	ANN (MATLAB)	0.01	31593	31591	2	Add
2016	ANN (MATLAB)	7.41	28675.9	26551	2124.9	Add
2017	ANN (MATLAB)	0.06	27398.2	27383	15.2	Add
2018	ANN (MATLAB)	0.99	28983	29270	287	Subtract
2019	ANN (MATLAB)	0.21	60939.6	60813	126.6	Add
2020	Linear reg. (Python)	0.96	35720.3	36062.4	342.1	Subtract
2021	Linear reg. (MATLAB)	0.81	36398	36104.4	293.6	Add
2022	SVM (Python)	0.67	35972.2	35729.8	242.4	Add
2023	Linear reg. (MATLAB)	2.67	36768	37749.8	981.8	Subtract

Table 7: Multi-Model Approach Performance

Year	Actual Load data (GWH)	Multi-Model Approach (Uncompensated)	Multi-Model Approach (Compensated)	Error Compensation Applied
2010	32913.1	32600.9	33225.3	Add 312.2 GWH
2011	30955.3	30901.3	31009.3	Add 54 GWH
2012	32087.4	32226	31948.8	Subtract 138.6 GWH
2013	37782.6	37748	37817.2	Add 34.60 GWH
2014	42142.8	42091	42194.6	Add 51.8 GWH
2015	31593	31591	31595	Add 2 GWH
2016	28675.9	26551	30800.8	Add 2124.9 GWH
2017	27398.2	27383	27413.4	Add 15.2 GWH
2018	28983	29270	28696	Subtract 287 GWH
2019	60939.6	60813	61066.2	Add 126.6 GWH
2020	35720.3	36062.4	35378.2	Subtract 342.1 GWH
2021	36398	36104.4	36691.6	Add 293.6 GWH
2022	35972.2	35729.8	36214.6	Add 242.4 GWH
2023	36768	37749.8	35786.2	Subtract 981.8 GWH

Therefore the optimal model chart is as drafted in Table 8.

Table 8: The Optimal Model Chart

Optimal Model Chart														
Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Optimal Model	ANN (Python)	Linear reg. (Python)	ANN (MATLAB)	ANN (MATLAB)	ANN (MATLAB)	ANN (MATLAB)	ANN (MATLAB)	ANN (MATLAB)	ANN (MATLAB)	ANN (MATLAB)	Linear reg. (Python)	Linear reg. (MATLAB)	SVM (Python)	Linear reg. (MATLAB)
Error Compensation	Add 312.2 GWH	Add 54 GWH	Subtract 138.6 GWH	Add 34.60 GWH	Add 51.8 GWH	Add 2 GWH	Add 2124.9 GWH	Add 15.2 GWH	Subtract 287 GWH	Add 126.6 GWH	Subtract 342.1 GWH	Add 293.6 GWH	Add 242.4 GWH	Subtract 981.8 GWH

The Multi-model approach can be gotten from the optimal model chart which can be split into

1. Multi-Model Approach (Compensated): Here multiple predictive models have been used to forecast a target variable which is the load, and then applying error compensation techniques to adjust the predictions to obtain higher accuracy. It takes into account the individual Mean Absolute Errors by adding or subtracting the error metric.
2. Multi-Model Approach (Uncompensated): This provides a baseline of how each model performs on its own without attempting to correct or modify them based on error metrics. In this case the Optimal Model values will be the Multi-Model Approach (Uncompensated), see the Table 7 for the values.

3.4. Future Outlook

Table 9 presents an Electricity consumption (GWH) outlook for the next 12 years for all the forecast methodologies covered in this study.

The outputs of the Linear Regression, SVM and ANN models were obtained using Python and MATLAB scripts by inputting the future projections of the independent variables gotten by simple Linear progression equations that were applied to grow GDP and Population data.

In order to calculate the Mean Absolute Percentage Error (MAPE) of the various predictive models in this analysis, a unique approach was used due to the identical outcomes of the linear regression models developed using both Python and MATLAB. Recognizing the consistency in these results, the Linear Regression model was used as a proxy for the actual load values. This methodological decision made it easy to calculate the MAPE for the other models by comparing their predictions against these proxy actual values. This approach provides a basis to assess the relative accuracy of each model under the assumption that the Linear Regression predictions are an accurate representation of true load values.

Table 9: Forecasted Load and Error Results

Year	GDP (B$)	Population (N)	Forecasted Load (GWH)					MAPE (%)			
			L. Regression (Python & MATLAB)	SVM (Python)	SVM (MATLAB)	ANN (Python)	ANN (MATLAB)	SVM (Python)	SVM (MATLAB)	ANN (Python)	ANN (MATLAB)
2024	514.1623	228639384.2	42142	34353	38368	45658	46830	18.48	8.96	8.34	11.12
2025	524.6746	233474136.5	43169	34357	39147	45479	51350	20.41	9.32	5.35	18.95
2026	535.1868	238308888.7	44197	34361	39926	45392	54974	22.25	9.66	2.7	24.38
2027	545.6991	243143640.9	45225	34364	40705	45744	57485	24.02	9.99	1.15	27.11
2028	556.2114	247978393.2	46252	34368	41484	46869	59053	25.69	10.31	1.33	27.68
2029	566.7237	252813145.4	47280	34372	42263	49126	59950	27.3	10.61	3.9	26.8
2030	577.2359	257647897.6	48308	34375	43043	52291	60414	28.84	10.9	8.25	25.06
2031	587.7482	262482649.8	49335	34379	43822	55224	60618	30.32	11.17	11.94	22.87
2032	598.2605	267317402.1	50363	34383	44601	57043	60671	31.73	11.44	13.26	20.47
2033	608.7728	272152154.3	51391	34386	45380	57841	60642	33.09	11.7	12.55	18
2034	619.285	276986906.5	52419	34390	46159	58049	60572	34.39	11.94	10.74	15.55
2035	629.7973	281821658.8	53446	34394	46939	57971	60483	35.65	12.17	8.47	13.17

Table 9 enables comparison of models based on their yearly accuracy and performance metrics. All models predict a consistent rise in electricity load demand between 2024 and 2035. (SVM Python) predicts the 2024 electricity load to be 34,353 GWh while (ANN MATLAB) estimates it at 46,830 GWh. The electricity demand forecast reaches 34,394 GWh (SVM Python) to 60,483 GWh (ANN MATLAB) by 2035 suggesting growth rates between 0.1% and 29.2% depending on the prediction model. The GDP will rise from $514.16 billion in 2024 to $629.8 billion in 2035 representing a 22.5% increase due to economic development. During this timeframe the population is expected to grow from 228.64 million to 281.82 million with an overall increase of 23.3%. Electricity demand shows a steady increase because of expanding economies and growing populations. Effective planning must be implemented to maintain adequate electricity production to meet these demands as projected load requirements could surpass 60,000 GWh by 2035 during peak scenarios.

4. CONCLUSION

Based on research that has been done by researchers, researchers can draw conclusions as follows: First, the input data used in the estimation of electrical energy consumption using Linear Regression, Support Vector Machine and Artificial neural networks include electrical energy consumption per sector in 2010-2023, Gross Domestic Product (GDP) and Population in Nigeria from the same timeframe. Second, forecasting electricity consumption using artificial neural networks proved to be quite good. The accuracy of the artificial neural network is proven by the resulting error is very small in the learning process. Determination of parameters such as the hidden layer and the number of iterations greatly affects the learning process. Therefore it is necessary to try several times in determining these parameters.

In this research paper, both Multi-Model Approach (Compensated and Uncompensated) results were presented to showcase the effectiveness of the compensation techniques used. This helps in demonstrating the value added by adjusting the predictions and provides a transparent view of the model's performance both before and after the adjustments.

REFERENCES

1. Isaac, S., Tolulope, O., Ayokunle A. & Peter, A (2016). Short-term load forecasting using the time series and artificial neural network methods. *Journal of Electrical and Electronics Engineering*, 11(1), 72-81.
2. Thomas, O. E.& Okafor, E. N. C. (2021). Electrical Energy Demand Forecast in Nigeria Between 2020

- 2040 Using Probabilistic Extrapolation Method. *International Journal of Engineering Science and Application*. 5(3), 71-85.

3. Onochie, U. P., Egware, H. O. & Eyakwanor, T. O. (2015). The Nigeria Electric Power Sector, Opportunities and Challenges. *Journal of Multidisciplinary Engineering Science and Technology (JMEST)*. 2(4), 112-124.
4. McGrath, A., & Jonker, A. (2024). *What is load forecasting?*. IBM Knowledge Center. Retrieved from https://www.ibm.com. Discusses different types of load forecasting and their challenges in modern grids.
5. Bureau of Statistics, (2022), Demographic Statistical Bulletin.Retrieved on September 10, 2024 from https://nigerianstat.gov.ng/elibrary/read/12414.
6. Hammad, M. A., Jereb, B., Rosi, B. & Dragan, D. (2020): Methods and Models for Electricity Load Forecasting: A Comprehensive Review. *Logistics & Sustainable Transport.* 11(1), 51-76.
7. Addeh E. (2023). Nigeria's Power supply woes continue. This day https://thisdaylive.com/index.php/2024/02/11/nigerias-power-supply-woes-continue.
8. Sasu, D. D.(2023) Electricity in Nigeria - statistics & facts.S*tatista*. Retrieved on October 1, 2024 from https://www.statista.com/topics/9470/electricity-in-nigeria/#topicOverview.
9. IEA (2021), *World Energy Outlook 2021*, IEA, Paris https://www.iea.org/reports/world-energy-outlook-2021.
10. Sovacool, B. K., Rayan, S. et al. (2017) Integrating social science in energy research. Energy Research and Social Sciences. 6(1):95-99.
11. Zhanga, X., Wanga, J.,& Zhang, k. (2017). Short-Term Electric Load Forecasting Based on Singular Spectrum Analysis and Support Vector Machine Optimized by Cuckoo Search Algorithm. *Electric Power Systems Research*, 146(1), 270–285.
12. Ben-Hur, Asa; Horn, David; Siegelmann, Hava; Vapnik, Vladimir N. ""Support vector clustering" (2001);". Journal of Machine Learning Research. **2**: 125–137.
13. Buitrago J and Asfour S 2017 Short-Term Forecasting of Electric Loads Using Nonlinear Autoregressive Artificial Neural Networks with Exogenous Vector Inputs 1–24.