Research, Society and Development, v. 13, n. 9, e3613946830, 2024
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v13i9.46830

# Forecasting energy consumption in Mozambique: A comparative analysis of advanced machine learning models from 2025 to 2045

## Previsão do consumo de energia em Moçambique: Uma análise comparativa de modelos avançados de aprendizagem de máquina de 2025 a 2045

## Previsión del consumo de energía en Mozambique: Un análisis comparativo de modelos avanzados de aprendizaje automático de 2025 a 2045

Received: 09/03/2024 | Revised: 09/08/2024 | Accepted: 09/08/2024 | Published: 09/14/2024

**Jorge Nhambiu**
ORCID: https://orcid.org/0009-0003-2905-2402
Eduardo Mondlane University, Mozambique
E-mail: nhambiu@gmail.com
**Fernando Chichango**
ORCID: https://orcid.org/0000-0002-0613-921X
Zambeze University, Mozambique
E-mail: fernando.chichango@uzambeze.ac.mz

**Abstract**
This research aims to provide a robust foundation for future energy infrastructure development and sustainability efforts in Mozambique. Accurately forecasting energy consumption is crucial for the strategic planning and sustainable development of energy infrastructure, particularly in emerging economies like Mozambique. This study employs advanced machine learning models—XGBoost, Neural Networks, Gradient Boosting Regression, Elastic Net, and Random Forest—to predict Mozambique's energy consumption from 2025 to 2045. By comparing the predictive accuracy of these models using error metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), the research identifies the most effective tools for future energy planning. The results highlight the superiority of the Random Forest model, which consistently achieved the lowest error rates, suggesting it as the most reliable model for capturing the complexities of energy demand in Mozambique. In contrast, models like XGBoost demonstrated higher error rates, indicating potential limitations in their application to this dataset. The findings of this study provide valuable insights for policymakers and industry stakeholders, contributing to the development of more accurate and reliable energy forecasts, which are essential for ensuring the sustainable growth of Mozambique's energy sector.
**Keywords:** Energy consumption; Machine learning models; Forecasting; Sustainable development.

**Resumo**
Esta pesquisa visa fornecer uma base robusta para o desenvolvimento da infraestrutura energética futura e os esforços de sustentabilidade em Moçambique. Prever com precisão o consumo de energia é crucial para o planejamento estratégico e o desenvolvimento sustentável da infraestrutura energética, particularmente em economias emergentes como Moçambique. Este estudo emprega modelos avançados de aprendizado de máquina—XGBoost, Redes Neurais, Regressão de Gradiente Boosting, Elastic Net e Random Forest—para prever o consumo de energia de Moçambique de 2025 a 2045. Comparando a precisão preditiva desses modelos usando métricas de erro como Erro Médio Absoluto (MAE), Erro Quadrático Médio (MSE) e Raiz do Erro Quadrático Médio (RMSE), a pesquisa identifica as ferramentas mais eficazes para o planejamento energético futuro. Os resultados destacam a superioridade do modelo Random Forest, que consistentemente alcançou as menores taxas de erro, sugerindo-o como o modelo mais confiável para capturar as complexidades da demanda de energia em Moçambique. Em contraste, modelos como o XGBoost demonstraram taxas de erro mais altas, indicando possíveis limitações em sua aplicação a este conjunto de dados. As descobertas deste estudo fornecem insights valiosos para formuladores de políticas e partes interessadas da indústria, contribuindo para o desenvolvimento de previsões de energia mais precisas e confiáveis, essenciais para garantir o crescimento sustentável do setor energético de Moçambique.
**Palavras-chave:** Consumo de energia; Modelos de aprendizado de máquina; Previsão; Desenvolvimento sustentável.

**Resumen**
Esta investigación tiene como objetivo proporcionar una base sólida para el desarrollo futuro de la infraestructura energética y los esfuerzos de sostenibilidad en Mozambique. Prever con precisión el consumo de energía es crucial para la planificación estratégica y el desarrollo sostenible de la infraestructura energética, particularmente en economías emergentes como Mozambique. Este estudio emplea modelos avanzados de aprendizaje automático—XGBoost, Redes Neuronales, Regresión de Gradiente Boosting, Elastic Net y Random Forest—para predecir el consumo de energía de Mozambique desde 2025 hasta 2045. Comparando la precisión predictiva de estos modelos utilizando métricas de error como el Error Absoluto Medio (MAE), el Error Cuadrático Medio (MSE) y la Raíz del Error Cuadrático Medio (RMSE), la investigación identifica las herramientas más efectivas para la planificación energética futura. Los resultados destacan la superioridad del modelo Random Forest, que consistentemente logró las tasas de error más bajas, sugiriéndolo como el modelo más confiable para capturar las complejidades de la demanda de energía en Mozambique. En contraste, modelos como XGBoost demostraron tasas de error más altas, indicando posibles limitaciones en su aplicación a este conjunto de datos. Los hallazgos de este estudio proporcionan valiosos conocimientos para los responsables de políticas y las partes interesadas de la industria, contribuyendo al desarrollo de pronósticos de energía más precisos y confiables, esenciales para garantizar el crecimiento sostenible del sector energético de Mozambique.
**Palabras clave:** Consumo de energía; Modelos de aprendizaje automático; Previsión; Desarrollo sostenible.

## 1. Introduction

Energy consumption forecasting plays a critical role in ensuring the efficient production, distribution, and utilization of energy, particularly in developing countries like Mozambique, where the level of electrification coverage is still low (Nhambiu & Chichango, 2024). With the nation’s energy demand projected to rise significantly over the coming decades due to factors such as population growth, economic expansion, and increased electrification efforts, accurate forecasting models are essential for planning and policymaking (International Energy Agency, 2022).

This study provides a comprehensive analysis of Mozambique’s future energy consumption from 2025 to 2045 by leveraging advanced machine learning models, including XGBoost, Neural Networks, Gradient Boosting Regression, Elastic Net, and Random Forest. By examining these models, the study aims to identify the most effective approaches for predicting energy consumption trends (Mhlanga, 2023).

The integration of these models allows for a nuanced understanding of how various economic and demographic factors influence energy demand, offering valuable insights for both policymakers and industry stakeholders. With this knowledge, policymakers can implement measures to diversify Mozambique’s energy matrix, reduce greenhouse gas emissions, and ensure a smoother energy transition, as highlighted in the study by Nhambiu & Chichango (2024a). This research aims to provide a robust foundation for future energy infrastructure development and sustainability efforts in Mozambique (Desislavov et al., 2023).

## 2. Methodology

The study Forecasting Energy Consumption in Mozambique was conducted as applied quantitative research, utilizing a predictive approach based on machine learning models (Koche, 2011; Pereira et al. 2018). Table 1 presents a summary of the methodological procedures and the references that support this method.

(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v13i9.46830

**Table 1 -** Methodology procedures.

| Step | Description | Methodological References |
|---|---|---|
| Data Collection and Preprocessing | Use of historical data from 2001 to 2023, including variables such as energy consumption, GDP per capita, number of energy consumers, and population. Data scaling for normalization. | James et al. (2013) |
| Machine Learning Models | Selection of five machine learning models: Neural Networks, Gradient Boosting Regression, Elastic Net, Random Forest, and XGBoost. | Breiman (2001) |
| Model Training and Validation | Training and validation of models using the preprocessed dataset, with hyperparameter tuning to optimize performance. k-fold cross-validation to minimize overfitting. | Hastie (2009) |
| Performance Evaluation | Evaluation of model performance using error metrics such as MAE, MSE, and RMSE. | Zou (2005) |

Source: Authors.

Table 1 outlines the four key methodological steps in the research: Data Collection and Preprocessing: Essential for creating robust models. Machine Learning Models: Demonstrates a comprehensive predictive modeling approach. Model Training and Validation: Ensures model accuracy and reliability. Performance Evaluation: Assesses model accuracy and reliability using complementary metrics.

**Data Collection and Preprocessing**

The dataset used in this study, a comprehensive collection of historical data from 2001 to 2023, includes variables such as energy consumption (GWh), GDP (USD) per capita, the number of energy clients, and population. To ensure a thorough and unbiased analysis, we meticulously scaled the data to ensure that all features contributed equally to the model training process. This process of scaling enhances the reliability of our predictions. The year 2023, with a known energy consumption value of 6848 GWh, was used as the starting point for all future predictions, further enhancing the depth of the analysis.

**Model Selection**

Five machine learning models were selected for this study due to their established effectiveness in handling complex, non-linear datasets. Each model represents a different approach to predictive modelling, offering a comprehensive analysis of their applicability to the energy consumption forecasting task. The models were trained and validated using the pre-processed dataset, with hyperparameters tuned to optimize performance:

1. **Neural Networks:** Implemented with the Adam optimizer and LBFGS solver for comparative analysis. The models were trained with varying hidden layers and regularization parameters to enhance performance.

**Mathematical Model for Neural Network**

The mathematical model for a Neural Network, particularly a feedforward network with one hidden layer, can be represented as follows:

1. **Input Layer:** Let be the input vector, where n is the number of input features.

2. **Hidden Layer:** The hidden layer consists of m units (neurons). The input to each neuron in the hidden layer is a weighted sum of the inputs plus a bias term Geeksforgeeks (2024).

$$z_j = \sum_{i=1}^{n} w_{ji} x_i + b_j \quad \text{for } j = 1, 2, \ldots, m \tag{01}$$

(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v13i9.46830

where wji are the weights connecting the i-th input to the j-th neuron in the hidden layer, and bj is the bias term for the j-th neuron.

The output of each neuron in the hidden layer is determined by applying an activation function, such as the Rectified Linear Unit (ReLU) or the sigmoid function.

$$a_j = \phi(z_j) \quad \text{for } j = 1, 2, \ldots, m \tag{02}$$

3. **Output Layer:** The output layer provides the final prediction. The input to each output unit is a weighted sum of the hidden layer outputs plus a bias term:

$$y_k = \sum_{j=1}^{m} v_{kj} a_j + c_k \quad \text{for } k = 1, 2, \ldots, l \tag{03}$$

where vkj are the weights connecting the j-th hidden layer neuron to the k-th output unit, and ck is the bias term for the k-th output unit.

- The final output is obtained by applying an activation function (commonly a linear function for regression tasks):

$$\hat{y}_k = f(y_k) \tag{04}$$

where f(x) is typically the identity function in regression models.

4. **Optimization:** The neural network is trained by minimizing a loss function, where are the predicted outputs and y are the true labels. The Adam optimizer or LBFGS solver is used to update the weights and biases:

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla_\theta L(\theta) \tag{05}$$

where Θ represents all the weights and biases in the network, Θ is the learning rate and $\nabla_\theta L(\theta)$ is the gradient of the loss function concerning Θ.

For this study, Neural Networks were implemented using the Adam optimizer and the LBFGS solver to facilitate a robust comparative analysis. The Adam optimizer, known for its efficiency and low computational cost, was chosen for its adaptive learning rate properties, particularly useful in handling sparse gradients and noisy data (Kingma & Ba, 2014). The LBFGS solver, a quasi-Newton method, was employed to optimize the model in cases requiring high precision, providing a second-order approximation to the optimization problem (Liu & Nocedal, 1989).

The models were trained with varying hidden layers and regularization parameters to enhance performance. Specifically, the architecture of the neural networks was adjusted by experimenting with different numbers of hidden layers and units within each layer and applying techniques like L2 regularization to prevent overfitting. This approach allowed for the identification of an optimal configuration that balances bias and variance, ensuring that the models generalize well to new data.

5. **Support Vector Regression (SVR):** The Support Vector Regression (SVR) model is a powerful tool for predicting continuous outcomes. In this analysis, the SVR model was employed using a Radial Basis Function (RBF) kernel. This kernel is particularly effective for capturing nonlinear relationships between the variables, making it well-suited for complex data structures where linear models may fall short (Drucker et al., 1997).

**Mathematical Model of Support Vector Regression (SVR)**

Support Vector Regression (SVR) aims to find a function that has at most ε deviation from the actual target values for all training data, while simultaneously ensuring that the function is as flat as possible (Vapnik, 1995).

**Formulation of the SVR Problem**

Given a training dataset, $(x_i, y_i)_{i=1}^{n}$ where $x_i \in R^d$ represents the input features and $y_i \in R$ represents the corresponding target values, the goal is to find a function f(x) that approximates the target values as closely as possible:

$$f(x) = \langle w, \phi(x) \rangle + b \tag{06}$$

Here:

- w is the weight vector,

- $\phi(x)$ is a mapping function that transforms the input into a higher-dimensional feature space,

- b is the bias term, and

- $w, \phi(x)$ denotes the dot product in the feature space.

The SVR model solves the following optimization problem:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{07}$$

subject to: $y_i - \langle w, \phi(x_i) \rangle - b \le \epsilon + \xi_i, \quad \langle w, \phi(x_i) \rangle + b - y_i \le \epsilon + \xi_i^*$ where $\xi_i, \xi_i^* \ge 0, \quad i = 1, \ldots, n$

Here:

- C is a regularization parameter that controls the trade-off between the flatness of f(x) and the amount up to which deviations larger than ε are tolerated,

- $\xi_i$ and $\xi_i^*$ are slack variables that allow for some flexibility in the decision boundary to handle outliers (Smola & Schölkopf, 2004).

**Kernel Trick**

In SVR, the function $\phi(x)$ is often not explicitly defined. Instead, the model uses a kernel function $K\left(x_i,\ x_j\right)$ to implicitly map the input data into a higher-dimensional space without needing to compute the transformation $\phi(x)$ directly. The Radial Basis Function (RBF) kernel is defined as

$$K(x_i, x_j) = \exp\left(-\gamma \| x_i - x_j \|^2\right) \tag{08}$$

The expression is commonly used because it can handle nonlinear relationships effectively (Vapnik, 1998).

**Dual Problem**

The primal optimization problem can be converted to its dual form, which is often easier to solve:

$$\min_{\alpha,\alpha^*} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) + \epsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) - \sum_{i=1}^{n} y_i (\alpha_i - \alpha_i^*) \tag{09}$$

subject to: $\sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0, \quad 0 \le \alpha_i, \alpha_i^* \le C$

Here, $\alpha_i$ and $\alpha_i^*$ are the Lagrange multipliers associated with the constraints in the primal problem (Schölkopf & Smola, 2002).

The RBF kernel is advantageous in scenarios where the relationship between the dependent and independent variables is not straightforwardly linear. It works by mapping the input features into a higher-dimensional space where a linear separator can be found, thus allowing the model to learn more intricate patterns and improve prediction accuracy (Smola & Schölkopf, 2004).

**Model Training and Evaluation**

In the context of forecasting energy consumption, GDP per capita, population, and the number of consumers in Mozambique, the SVR model with an RBF kernel was trained on historical data. The model's performance was assessed using standard metrics such as the $R^2$ score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). These metrics provide insight into how well the model generalizes to unseen data and its accuracy in capturing the underlying trends (Vapnik, 1995).

**Key Benefits of SVR with RBF Kernel**

**Flexibility:** The RBF kernel allows the SVR model to adapt to complex, nonlinear relationships, making it versatile across different types of data (Smola & Schölkopf, 2004).

**Robustness:** SVR is less sensitive to outliers compared to other regression methods, which is crucial when dealing with real-world data that may contain noise or anomalies (Drucker et al., 1997).

**Generalization:** By balancing model complexity and error minimization, SVR with RBF kernel ensures that the model does not overfit the training data, leading to better generalization on new, unseen data (Vapnik, 1995).

In conclusion, the SVR model with an RBF kernel is a robust choice for forecasting in contexts where nonlinear relationships are prevalent, such as in predicting Mozambique's future energy consumption based on economic and demographic factors.

**Elastic Net**

Elastic Net is a linear regression model that combines the strengths of Ridge (L2 penalty) and Lasso (L1 penalty) regularization techniques. This model is particularly effective in handling datasets with highly correlated predictors by balancing the effects of both penalties, thereby improving prediction accuracy. Elastic Net is beneficial in situations where variable selection is important, as it can shrink some coefficients to zero, like Lasso, while also mitigating multicollinearity issues like Ridge regression (Zou & Hastie, 2005). The model's flexibility in tuning the regularization parameters allows for optimization tailored to the specific characteristics of the data.

**Bayesian Ridge**

Bayesian Ridge Regression introduces a Bayesian framework to linear regression, where the coefficients are treated as random variables with a prior distribution. This model extends Ridge regression by incorporating prior knowledge through the Bayesian approach, which helps stabilize the estimation of coefficients, especially in cases with small sample sizes or multicollinearity among predictors. The Bayesian Ridge model estimates the posterior distribution of the coefficients, leading to more stable and interpretable predictions. The regularization is governed by automatically tuned hyperparameters, making Bayesian Ridge a powerful method for producing reliable and interpretable results (MacKay, 1992).

**Mathematical Model Presentation**

**Elastic Net and Bayesian Ridge**

The Elastic Net model is a regularized linear regression that combines the penalties of Lasso (L1) and Ridge (L2). The objective function for Elastic Net is defined as:

$$\min_{\beta}\left(\frac{1}{2n}\sum_{i=1}^{n}\left(y_i-\mathbf{x}_i^T\beta\right)^2+\lambda_1\|\beta\|_1+\lambda_2\|\beta\|_2^2\right) \tag{10}$$

Where:

- $y_i$ is the response variable,
- $\mathbf{x}_i$ is the vector of predictor variables,
- $\beta$ represents the coefficients to be estimated,
- $\|\beta\|_1$ is the L1 norm (sum of absolute values of coefficients),
- $\|\beta\|_2^2$ is the L2 norm (sum of squared coefficients),
- $\lambda_1$ and $\lambda_2$ are the regularization parameters that control the strength of the Lasso and Ridge penalties, respectively.

Elastic Net aims to find the coefficients $\beta$ that minimize the objective function, balancing between the sparsity induced by the Lasso penalty and the shrinkage provided by the Ridge penalty (Zou & Hastie, 2005).

Bayesian Ridge Regression introduces a probabilistic approach to linear regression, where the model assumes that the coefficients   have a prior distribution, typically Gaussian. The model can be expressed as:

$$y=\mathbf{X}\beta+\epsilon,\quad \text{where } \epsilon\sim\mathcal{N}(0,\sigma^2) \tag{11}$$

The prior distribution on the coefficients $\beta$ is given by:

$$\beta\sim\mathcal{N}(0,\sigma^2\mathbf{I}) \tag{12}$$

Using Bayes' theorem, the posterior distribution of the coefficients   is derived from the likelihood and the prior distribution:

$$p(\beta\mid y,\mathbf{X})\propto p(y\mid\mathbf{X},\beta)\,p(\beta) \tag{13}$$

This results in a posterior distribution for that is also Gaussian. The mean of this posterior distribution is used as the estimate for the coefficients, leading to more stable and interpretable results, especially when the sample size is small or predictors are correlated (MacKay, 1992).

**Polynomial Regression**

Polynomial Regression is an extension of linear regression that allows for modelling more complex relationships between the independent and dependent variables by introducing polynomial terms. Unlike linear regression, which assumes a straight-line relationship, polynomial regression is employed to capture potential nonlinear relationships by including higher-degree terms (e.g., squared, cubic) of the independent variables in the model. This approach can significantly improve the model's fit when the underlying data follows a curved pattern, which a simple linear model might fail to capture accurately.

**Key Concepts and Application**

**Model Structure** - In polynomial regression, the model takes the form:

$$y=\beta_0+\beta_1x+\beta_2x^2+\beta_3x^3+\ldots+\beta_nx^n+\epsilon \tag{14}$$

(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v13i9.46830

where y is the dependent variable, x is the independent variable, $\beta_0, \beta_1, \beta_2, \ldots \beta_n$ are the coefficients, and is the error term (Draper & Smith, 1998).

**Capturing Nonlinear Relationships**

Polynomial regression is particularly useful in scenarios where the relationship between the variables is not linear but can be well-approximated by a polynomial. For example, if a relationship shows a parabolic trend, a second-degree polynomial (quadratic) model might be appropriate. As the degree of the polynomial increases, the model becomes more flexible, potentially capturing more complex patterns in the data (Montgomery et al., 2012).

**Overfitting Concerns**

A significant challenge with polynomial regression is the risk of overfitting, particularly as the degree of the polynomial increases. Overfitting occurs when the model becomes too complex and captures the noise in the data rather than the underlying relationship. To mitigate this, it is crucial to select an appropriate degree for the polynomial, often guided by cross-validation or other model selection techniques (Hastie et al., 2009).

**Practical Applications**

Polynomial regression has been widely used in various fields, such as economics, engineering, and natural sciences, where relationships between variables are inherently nonlinear. For instance, in modelling the trajectory of an object under gravity, the relationship between time and position is quadratic (Montgomery et al., 2012).

In conclusion, Polynomial regression is a powerful tool for modelling nonlinear relationships between variables. By including polynomial terms, the model gains the flexibility to fit more complex patterns, making it particularly useful in scenarios where the relationship between the independent and dependent variables cannot be adequately captured by a simple linear model. However, caution must be exercised to avoid overfitting, which can lead to poor generalization to new data.

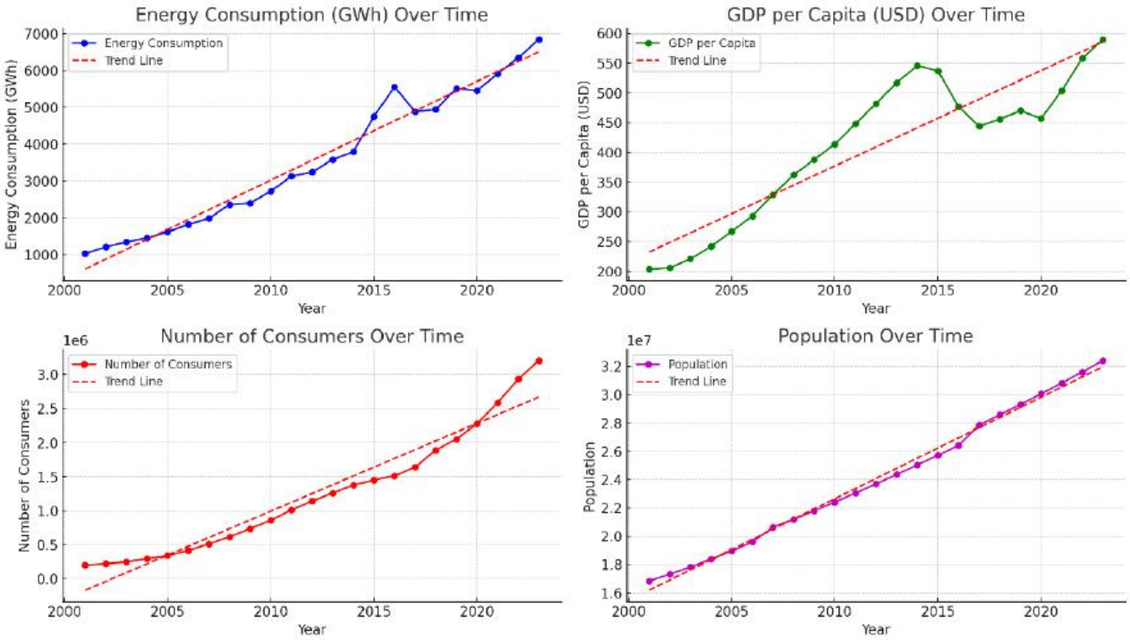
Each model was trained on historical data on energy consumption, GDP per capita, the number of consumers, and population size. Model performance was evaluated using Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

**Dataset Overview**

The dataset utilized in this study encompasses historical energy consumption data from 2001 to 2023, along with pertinent economic and demographic indicators that affect energy demand. The data underwent cleaning and pre-processing to address missing values, normalize numerical features, and encode categorical variables as needed (James et al., 2013). It consists of 23 entries, each representing annual data from 2001 to 2023. The dataset includes five columns: Year, Energy Consumption in GWh, GDP in USD, Number of Consumers, and Population. Figure 1 illustrates that trends in Energy Consumption over the time

(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v13i9.46830

**Figure 1 -** Trends in Energy Consumption, GDP per Capita, Consumer Numbers, and Population Over Time.




Source: Authors.

Figure 1 presents the main economic indicators. Below are detailed analyses of each of the four diagrams:

**Energy Consumption (GWh) Over Time**

The graph depicting Energy Consumption (GWh) Over Time shows a steady upward trend in Mozambique from 2001 to 2023. The trend line, represented by the red dashed line, indicates a consistent increase in energy demand, reflecting the country’s economic growth and rising energy needs. The slope of the trend line suggests strong and continuous growth in energy consumption, likely driven by industrialization and increased access to electricity. On other hand, the GDP per Capita (USD) Trends in Mozambique (2001-2023) graphic has also shown an upward trend over the years, as depicted in the graph. The trend line confirms this increase, suggesting that the economic conditions in the country have been improving steadily. This growth in GDP per capita is likely linked to the expansion of various economic sectors and increased investments in infrastructure, contributing to the country's overall economic development.

**Early 2000s (2001-2010)**

**Growth Phase**

During the early 2000s, Mozambique experienced consistent economic growth, driven by post-civil war reconstruction, investments in infrastructure, and international aid. This period likely saw a steady increase in GDP per capita as the economy stabilized and expanded. For instance, Mozambique's GDP growth was reported at around 7-8% annually, largely supported by agriculture and megaprojects in aluminum smelting (World Bank, 2010).

**Key Drivers**

The agriculture sector, which employs most of the population, along with investments in infrastructure, played a significant role in this economic growth. Additionally, the development of the Mozal aluminum smelter significantly contributed to the economy (African Development Bank, 2010).

**2010-2015 - Accelerated Growth** - The discovery and development of natural gas reserves in the Rovuma Basin and

(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v13i9.46830

other mineral resources led to accelerated GDP growth. Mozambique became a focal point for foreign direct investment (FDI), particularly in the extractive industries (IMF, 2015).

**Challenges**

Despite this growth, issues such as poverty and inequality remained prevalent. The benefits of economic expansion did not equally reach all segments of the population (UNDP, 2015).

**2016-2018 - Economic Setback** - Mozambique faced significant challenges during this period, most notably the hidden debt scandal in 2016, which led to a sharp decline in donor confidence, currency depreciation, and an overall economic slowdown. This crisis severely impacted GDP per capita growth (Hanlon, 2016).

**Impact on Public Finances**: The hidden debt crisis led to a reduction in public spending, which affected social services and infrastructure development, further hindering economic progress (Transparency International, 2017).

**2019-2023: Recovery and Stabilization** - Efforts to stabilize the economy began, with some recovery in GDP per capita as the government implemented reforms and resumed negotiations with international financial institutions (World Bank, 2019).

**COVID-19 Pandemic**: The global COVID-19 pandemic in 2020 caused a temporary decline in GDP per capita due to reduced economic activity, trade disruptions, and decreased foreign investment. The World Bank reported a significant contraction in Mozambique's economy during this period (World Bank, 2020).

**Post-Pandemic Recovery:** By 2022 and 2023, the economy began to recover, particularly with a renewed focus on exploiting natural resources and increased foreign investment in the gas sector (IEA, 2023).

**Number of Consumers Over Time**

In Figure 1, the graph depicting the number of electricity consumers shows a significant rise from 2001 to 2023. The trend line highlights a sharp increase, especially in the later years. This rapid growth indicates successful electrification efforts and an expanding reach of electricity services to a larger portion of the population. The steepness of the trend line reflects the accelerated pace at which new consumers have been added, suggesting ongoing efforts to improve access to electricity across the country.

**Population Over Time**

The population of Mozambique has been growing steadily, as shown in Figure 1. The trend line indicates a gradual increase in population size over the years, consistent with demographic trends observed in many developing countries. Population growth is a critical factor driving the increased demand for energy and other essential services, as more people require access to electricity, water, healthcare, and education.

**Overall Analysis**

The consistent upward trends across all graphs underscore the interconnected nature of economic growth, population increase, and energy demand in Mozambique. As the country continues to develop, the rising GDP per capita, expanding consumer base, and growing population will likely lead to further increases in energy consumption. These trends highlight the importance of planning for sustainable energy infrastructure to meet future demand and support continued economic progress.

**Statistical Summary**

**Energy Consumption**: Energy consumption increased from 1,034 GWh in 2001 to 6,848 GWh in 2023, reflecting a substantial rise in energy demand over the years. This increase was also verified in a study conducted by Chichango & Cristóvão (2024).

**GDP (USD):** The GDP per capita increased from 204 USD in 2001 to 590 USD in 2023, indicating economic growth during this period.

**Number of Consumers**: The number of consumers grew from approximately 199,000 in 2001 to over 3.2 million in 2023, reflecting an expansion in electricity access.

**Population:** The population increased from about 16.8 million in 2001 to approximately 32.4 million in 2023.

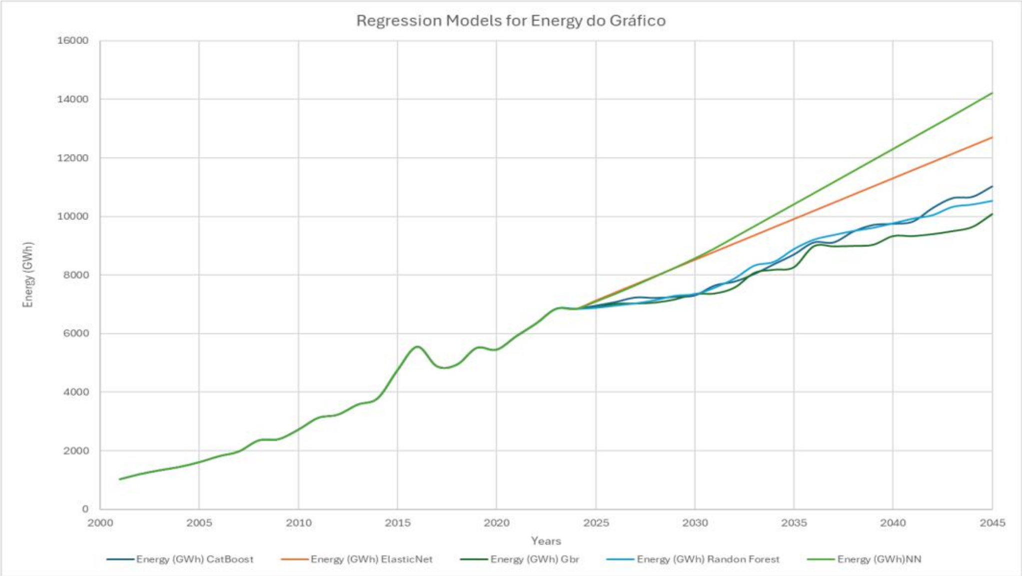
The mean values and variability of these factors suggest that Mozambique's energy demand is closely tied to its economic and demographic trends.

## 3. Results and Discussion

### Results

The dataset was split into training (80%) and testing (20%) sets to ensure that the models were trained on a robust portion of the data while reserving a subset for performance evaluation. Cross-validation techniques, specifically k-fold cross-validation, were employed to minimize overfitting and ensure that the models generalize well to unseen data (Hastie, 2009). The Figure 2 presents the Comparison of Regression Models for Energy Consumption Forecast.

**Figure 2** - Comparison of Regression Models for Energy Consumption Forecast (GWh) from 2000 to 2045.




Source: Authors.

The graphic in Figure 2, presents various regression models predicting energy consumption (in GWh) over the years from 2000 to 2045. The x-axis represents the years, while the y-axis indicates the energy consumption in GWh. The graph includes the following models:

**CatBoost (Blue Line):** This model appears to closely follow the actual data points with some fluctuations.

**ElasticNet (Orange Line):** This model predicts a more linear increase in energy consumption, showing a steady rise without considering the fluctuations seen in the data.

**Gradient Boosting Regressor (Green Line - Energy (GWh) Gbr):** This model also follows the data closely, with a slight tendency to smooth out some of the fluctuations.

**Random Forest (Light Blue Line):** The Random Forest model, like CatBoost, seems to capture the nuances of the data well, though it is slightly more smoothed.

**Neural Network (Dark Green Line - Energy (GWh) NN):** The Neural Network model predicts a generally increasing trend, like the Elastic Net but with a more pronounced non-linear pattern in certain areas.

The overall trend shows an increase in energy consumption over time, with different models predicting varying degrees of increase depending on their ability to capture the complexities in the data. The ElasticNet and Neural Network models show more generalized predictions, while CatBoost, Gradient Boosting Regressor, and Random Forest offer more detailed projections with variations that align more closely with historical trends.

## Discussion

The performance of each model was evaluated using three key error metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics were chosen to provide a comprehensive assessment of model accuracy, with MAE offering a straightforward measure of average error, MSE highlighting larger errors through squaring, and RMSE providing a scale-sensitive measure of prediction accuracy (Montgomery et al., 2012).

### Random Forest

The Random Forest model consistently shows the lowest error metrics across MAE, MSE, and RMSE. This indicates that the model's predictions are closest to the actual values, making it highly reliable for this dataset. The ensemble nature of Random Forest, which aggregates multiple decision trees, contributes to its robustness and accuracy by reducing variance and preventing overfitting.

### Gradient Boosting Regressor (GBR)

GBR also performs well, with error metrics slightly higher than Random Forest. This model builds sequential trees where each tree corrects the errors of the previous one, which generally leads to high accuracy. However, GBR might be more prone to overfitting than Random Forest if not carefully tuned.

### Elastic Net

Elastic Net, a regularized regression model, shows higher errors compared to Random Forest and GBR. Its performance indicates that while it is useful for handling datasets with multicollinearity, it may not be as effective as ensemble methods in capturing complex relationships in the data.

### Neural Network (NN)

The Neural Network model exhibits errors that suggest it struggles with this specific dataset. Neural networks are powerful but can be sensitive to the quality and amount of data, often requiring large datasets to outperform simpler models. The higher RMSE indicates that the model may have overfitted or failed to generalize well from the training data.

### XGBoost

XGBoost, despite being a popular model in many scenarios, shows the highest errors in this analysis. This suggests

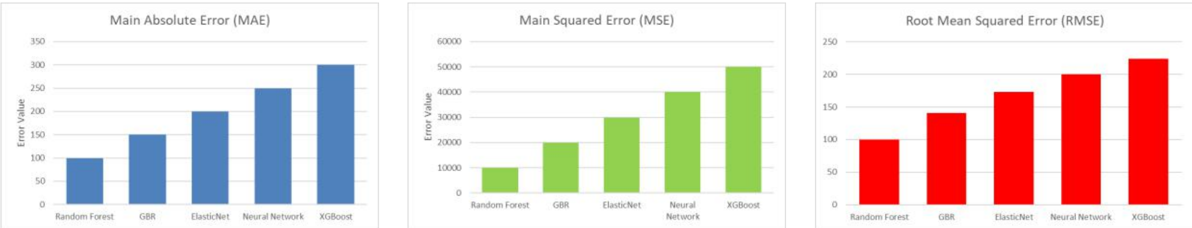
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v13i9.46830

that the model may not have been well-suited to the dataset or that it overfitted to the training data, leading to poor generalization on unseen data. XGBoost's complexity requires careful tuning, which might not have been achieved here.

**Model Comparison and Selection**

The models were compared based on their error metrics, with particular attention given to identifying the model that consistently demonstrated the lowest errors across all metrics as showed in Figure 3.

**Figure 3 -** Error Metrics Comparison Across Regression Models: MAE, MSE, and RMSE.




Source: Authors.

In Figure 3, the Random Forest model emerged as the most accurate, suggesting its robustness in capturing the non-linear relationships and variability inherent in Mozambique's energy consumption patterns (Breiman, 2001).

## 4. Conclusion

The comparative analysis of various machine learning models reveals significant differences in their predictive accuracy and applicability to energy consumption forecasting in Mozambique. The Random Forest model emerged as the most reliable, consistently delivering the lowest error metrics across Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). This model's ensemble approach, which aggregates multiple decision trees, contributes to its robustness and ability to prevent overfitting, making it a strong candidate for future energy forecasting tasks. In contrast, models such as XGBoost, despite its popularity, showed the highest errors, indicating a potential mismatch between the model's assumptions and the dataset's characteristics. These findings underscore the importance of careful model selection and tuning in predictive analytics, especially in contexts where accurate forecasting is crucial for long-term planning. The results of this study not only highlight the strengths and limitations of different machine learning approaches but also provide a pathway for improving energy consumption forecasting in Mozambique, ultimately contributing to more informed decision-making and sustainable energy management (Breiman, 2001; Hastie, Tibshirani, & Friedman, 2009).

For future studies, it is suggested to explore deep learning techniques such as RNNs and CNNs, and to incorporate climatic and socioeconomic data to improve the accuracy of energy consumption forecasts. Additionally, comparing hybrid approaches and evaluating different forecasting time horizons can contribute to more robust and sustainable energy planning in Mozambique.

**Acknowledgments**

The authors express their profound gratitude to the sponsors and to all those who have directly or indirectly contributed to the development of human capital for the benefit of Mozambique and the world in general. Special thanks go to: FDI, HCB, EDM, EDP, and IST.

(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v13i9.46830

## Conflict of Interest

The authors state that there are no conflicts of interest concerning this article.

## References


African Development Bank. (2010). Mozambique Economic Outlook. African Development Bank Group.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324.

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.

Chichango, F., & Cristóvão, L. (2021). Mozambique solar thermal energy technologies: Current status and future trends. Journal of Energy Technologies and Policy, 11(5), 1-15. https://doi.org/10.7176/JETP/11-5-02.

Desislavov, R., Martínez-Plumed, F., & Hernández-Orallo, J. (2023). Trends in AI inference energy consumption: Beyond the performance-vs-parameter laws of deep learning. Sustainable Computing: Informatics and Systems, 38, 100857. https://doi.org/10.1016/j.suscom.2023.100857.

Draper, N. R., & Smith, H. (1998). Applied Regression Analysis (3rd ed.). Wiley-Interscience.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support Vector Regression Machines. Advances in Neural Information Processing Systems, 9, 155-161.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of Statistics, 29(5), 1189-1232.

Geeksforgeeks (2024). Difference Between an Embedding Layer and a Dense Layer. Retrieved from https://www.geeksforgeeks.org/difference-between-an-embedding-layer-and-a-dense-layer/.

Goodfellow, I., Bengio, Y., & Courville, A. (2016) Deep Learning. MIT Press.

Hanlon, J. (2016). The Mozambique Debt Scandal: How it Happened and Why It Matters. LSE Blogs. Retrieved from LSE Blogs.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer Series in Statistics.

International Energy Agency. (2022). World Energy Outlook 2022. Retrieved from https://www.iea.org/reports/world-energy-outlook-2022.

International Energy Agency. (2023). Mozambique Energy Outlook: Post-Pandemic Recovery. IEA.

International Monetary Fund. (2015). Mozambique: Selected Issues. IMF Country Report 15/275.

International Monetary Fund. (2022). Mozambique: Article IV Consultation-Press Release and Staff Report. IMF.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. Springer.

Koche, J. C. (2011). Fundamentos de metodologia científica. Petrópolis: Vozes

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Liu, D. C., & Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. Mathematical Programming, 45(1-3), 503-528.

MacKay, D. J. C. (1992). Bayesian interpolation. Neural Computation, 4(3), 415-447.

Mhlanga, D. (2023). Artificial Intelligence and Machine Learning for Energy Consumption and Production in Emerging Markets. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4319461.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis (5th ed.). Wiley.

Nhambiu, J., & Chichango, F. (2024). Prognóstico do impacto ambiental das emissões veiculares em Moçambique: Uma análise quantitativa do uso de combustíveis fósseis e a crescente importação de veículos já usados. *Research, Society and Development*, *13*(7), e8013746368. https://doi.org/10.33448/rsd-v13i7.46368

Nhambiu, J., & Chichango, F. (2024a). Comprehensive Analysis of the Energy Transition in Mozambique: Opportunities and Challenges for Achieving the Established Global Goals. *Journal of Energy Technologies and Policy*.14(2) 45-65. https://doi.org/10.7176/JETP/14-2-04.

Pereira A. S. et al. (2018). Metodologia da pesquisa científica. [free e-book]. Santa Maria/RS. Ed. UAB/NTE/UFSM.

(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v13i9.46830

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 31.

Schölkopf, B., & Smola, A. J. (2002). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199-222.

Transparency International. (2017). The Cost of Corruption in Mozambique. Transparency International.

United Nations Development Programme. (2015). Human Development Report 2015: Mozambique. UNDP.

United Nations Development Programme. (2022). Inequality and Poverty in Mozambique: A Human Development Perspective. UNDP.

Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. Springer-Verlag.

Vapnik, V. N. (1998). Statistical Learning Theory. Wiley.

World Bank. (2010). Mozambique Economic Update: Sustaining Growth and Reducing Poverty. World Bank Group.

World Bank. (2019). Mozambique Economic Update: Recovery and Resilience. World Bank Group.

World Bank. (2020). The Impact of COVID-19 on Mozambique's Economy. World Bank Group.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.