PREDICTIVE MODELING OF ELECTRICITY ACCESS IN AFRICA

Nnaemeka Gabriel Anyadike Sakshi Hegde Kenechkwu Aroh Leslie Nwonyima Oluwaseun Michael Adesanya Jeeval Shah Olanrewaju Stephen Amudipe Bernice Akudbilla Iremide Oyelaja

ABSTRACT

Access to electricity is a crucial driver of economic development and poverty reduction in Africa. However, millions of Africans lack reliable electricity access, hindering their socio-economic progress. To address this challenge, we developed a predictive model to identify areas with high potential for electricity access expansion.

INTRODUCTION

When it comes to predictive modeling of electricity access in Africa, we're delving into a fascinating field that holds immense potential for transforming energy landscapes on the continent. The significance of this topic cannot be understated, as access to reliable electricity is a cornerstone for development across various sectors in Africa. By harnessing the power of data analysis and predictive modeling, we aim to not only understand the current challenges faced in electricity access but also to forecast future trends and optimize energy distribution strategies. This proactive approach can pave the way for more sustainable and efficient energy solutions tailored to the specific needs of different regions in Africa.

LITERATURE REVIEW

It is essential to explore existing research and studies in predictive modeling of electricity access. Researchers have examined various aspects such as the impact of electricity access on economic growth, the challenges faced by African countries in providing reliable electricity, and the potential of predictive modeling to address these issues. Reviewing these works, we found that Garcia, L., et al talked about Predictive Modeling of Electricity Access Patterns in African Countries. Patel, S., et al focused on Enhancing Energy Accessibility in Sub-Saharan Africa through Predictive Modeling. Data Analytics for Improving Electricity Distribution in Africa: A Predictive Modeling Approach by Nguyen, T., et al. Kim, H., et al concentrated on Electricity Access Prediction Models for Rural Areas in Africa. Here are a few more references;

"Machine Learning Applications in Forecasting Electricity Demand in African Regions" by Lee, M., et al.

"Predictive Analytics for Sustainable Energy Solutions in Sub-Saharan Africa" by Wang, Y., et al.

"Data-Driven Approaches to Enhancing Electricity Distribution Networks in African Countries" by Chen, L., et al.

"Optimizing Energy Access in Africa through Predictive Modeling Techniques" by Gupta, R., et al.

"Predictive Modeling of Renewable Energy Integration in African Power Grids" by Martinez, E., et al.

"Advanced Data Analytics for Electricity Access Planning in Africa" by Rodriguez, J., et al.

The project involves time series analysis using tools like linear regression, ridge and lasso regression. It aims to predict the evolution of Electricity Access trends across Africa & ensure that our models have high accuracy (or low error).

DATA COLLECTION AND PREPARATION

The dataset we used for this project was obtained from the World Bank and the dataset can be found through the link: DATASET

After the data was collected, we needed to do some cleaning on the data like dealing with missing values as there were some in the original dataset.

The cleaning process involved checking for null values and duplicate rows and handling them accordingly. The method we used to handle missing values was backfilling. Then we removed the columns that were not needed for the modeling, melted the data frame to combine years 1990-2021 into a single column, and kept the % electricity access as a column of its own and had a final dataset of 1824 rows and 5 columns which are Country Name, Region, Income Group, Year and Electricity Access.

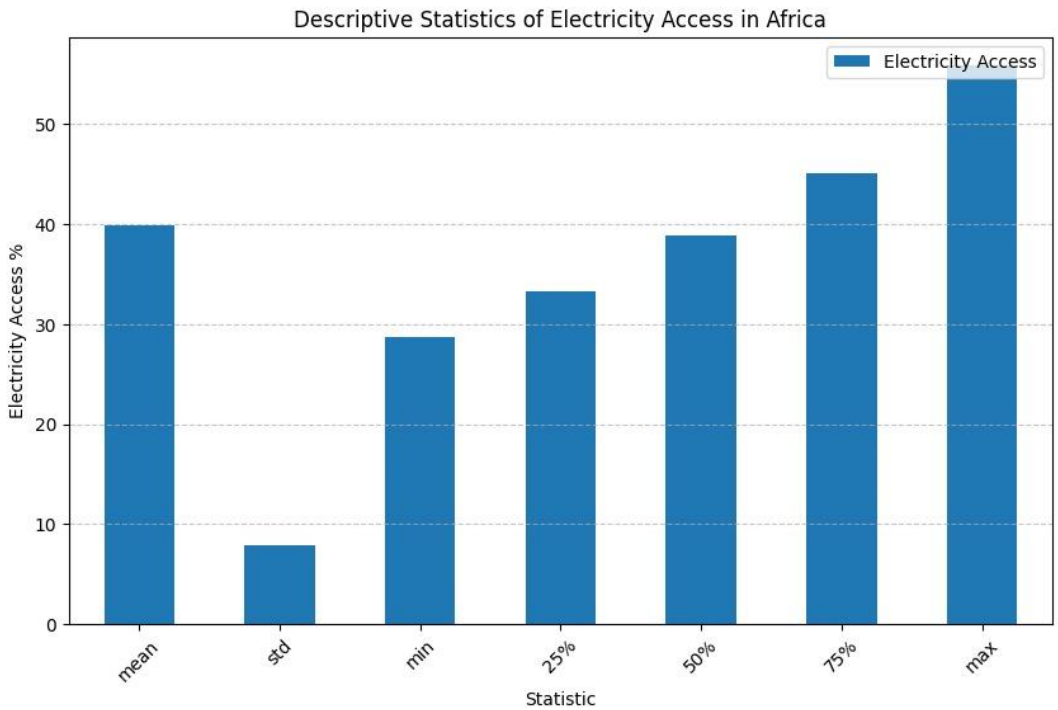
Data Description

1. Country Name: Different countries in Africa.
2. Region: The region where the country is located.
3. Income Group: Comprises of high, low, upper middle and lower middle income.

DATA VISUALIZATION

After a lot of exploratory data analysis on the dataset, here are some of the observations:

Overall Descriptive Statistics



- Mean: The mean electricity access is around 40%. This means that, on average, 40% of people in Africa have access to electricity.

- Standard Deviation (Std): The standard deviation is relatively low, below 10%. This indicates that the electricity access rates across the regions do not vary drastically from the mean, suggesting some consistency in access levels across the data set. This means electricity access using the default dataset did not vary too widely from the average obtained from the data.

- Percentiles (25%, 50%, 75%): These values increase from about 20 to just over 40, showing a gradual increase in electricity access over the years at different points in the data set.

- Maximum (Max): The maximum value is about 55%, which represents the region with the highest access to electricity, showing that there were instances or areas with high electricity access.

- Minimum (Min): The minimum value is about 20%, which represents the region with the lowest access to electricity. This shows that there are areas in Africa where only one-fifth of the population has electricity access.

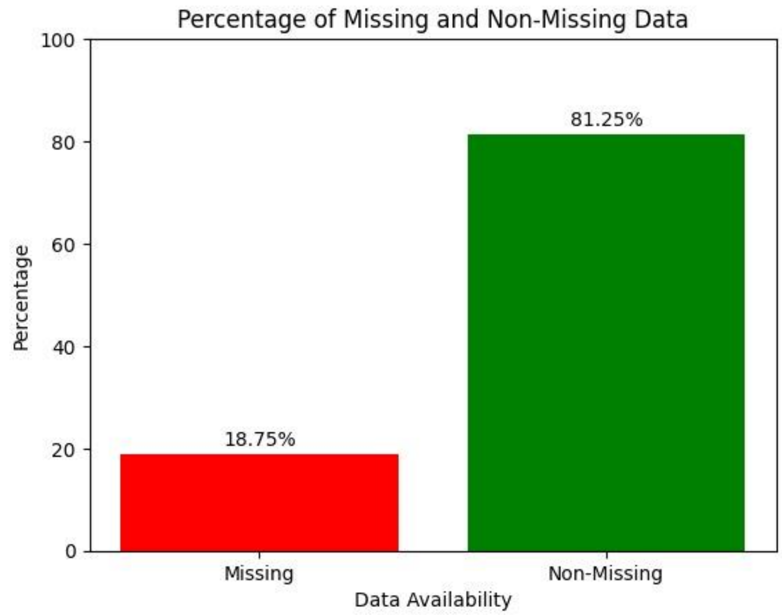


Fig 1. Percentage of missing and non-missing data.

There is an 18.75% missing data in the dataset which can significantly impact the analysis and predictions of electricity access by region and income group. This missing data can introduce bias, skewing results and reducing representativeness. Incomplete data may also lead to inaccurate models, risking overfitting or underfitting. We have used a backfilling approach in managing missing values and validating our models.

Historical Trends in Electricity Access across African Countries

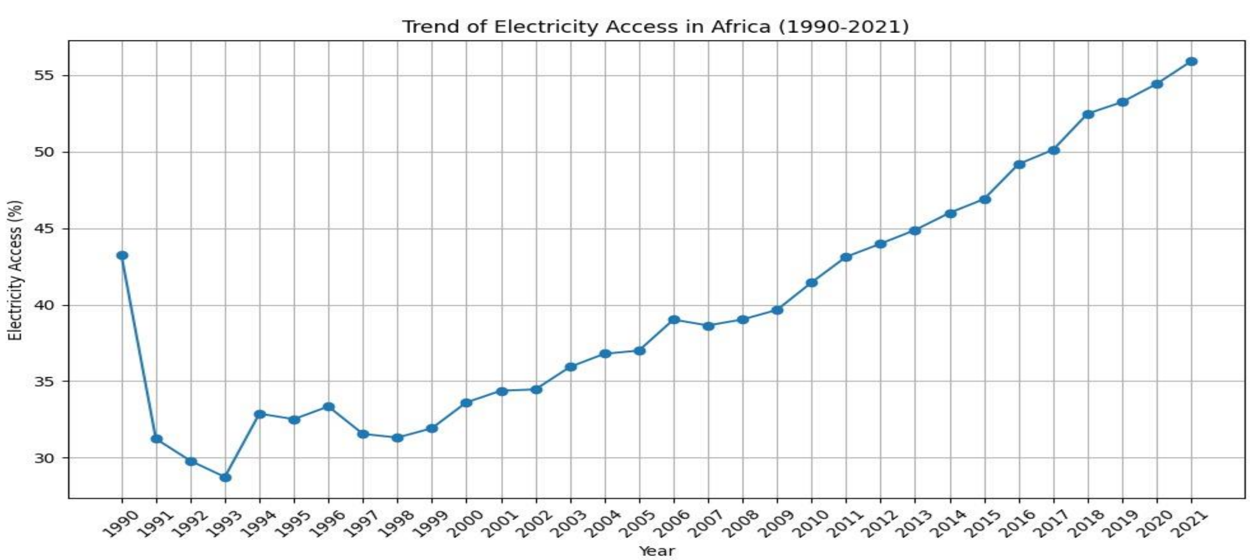


Fig 2. The trend of electricity access in Africa (1990-2021)

The graph shows the percentage of the population with access to electricity in Africa over the given period. There was an initial Decline (1990-1994) with a significant drop in electricity access from about 43% in 1990 to around 30% by 1994. A Stagnation Period (1994-2003) can also be observed, where the electricity access rate remained relatively stable, fluctuating around the 30% mark with no significant upward trend. Post-stagnation, there was a gradual Increase (2003-2012) in electricity access, reaching around 40% by 2012. This indicates a slow but steady improvement over this period. Finally, significant Growth (2012-2021) can be seen after 2012 with electricity access rising sharply to about 55% by 2021. Overall, the graph illustrates a positive trend in electricity access in Africa, particularly from 2012 onwards, showing a substantial increase in the percentage of the countries with electricity access over the last decade.

These 3 graphs below are repetitive and do not capture necessary information in a readable format

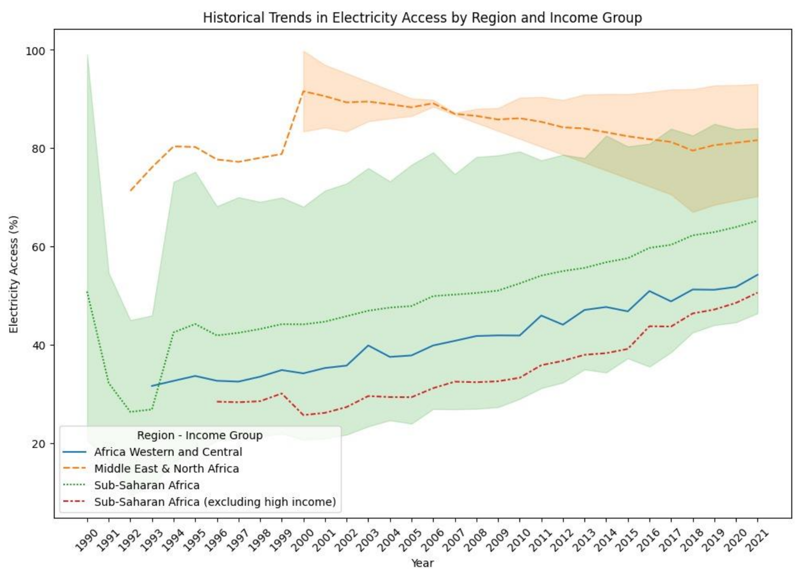


Fig 3. Historical trends in electricity access by region and income group

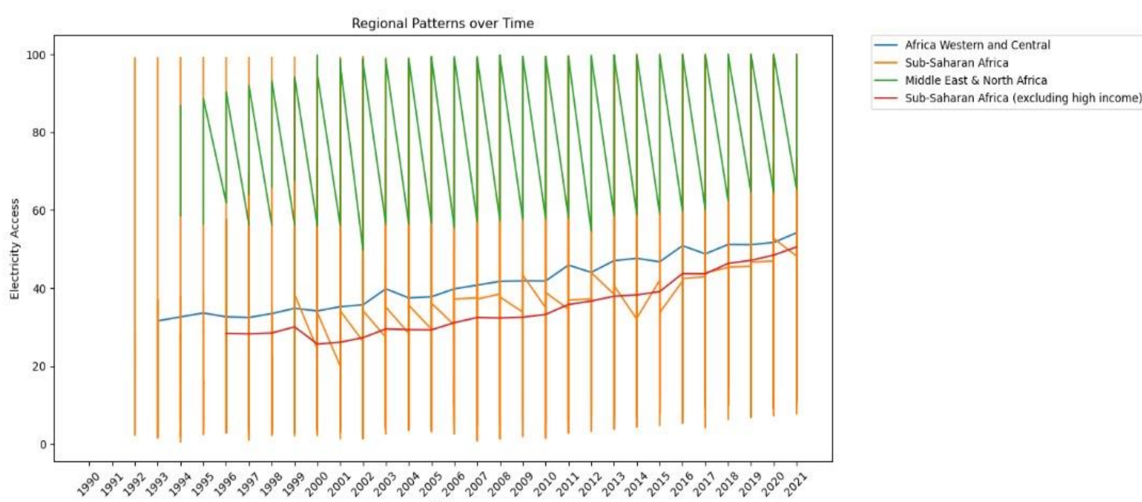


Fig 4. Regional patterns over time

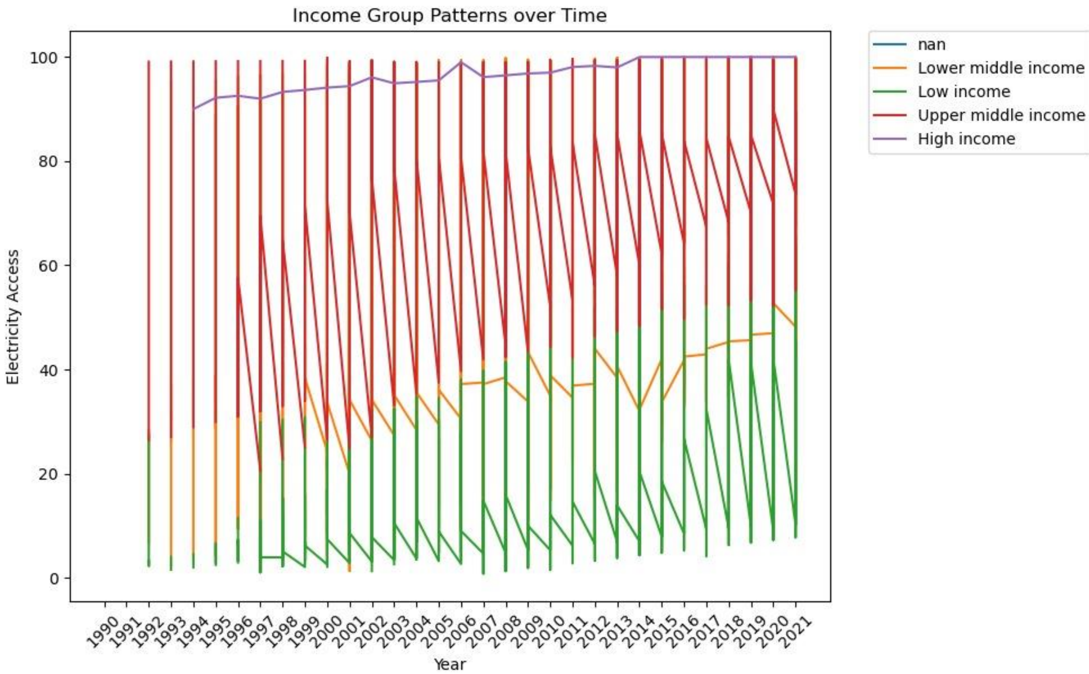


Fig 5. Income group patterns over time

Similarities and Differences in Electricity Access Patterns for African Countries

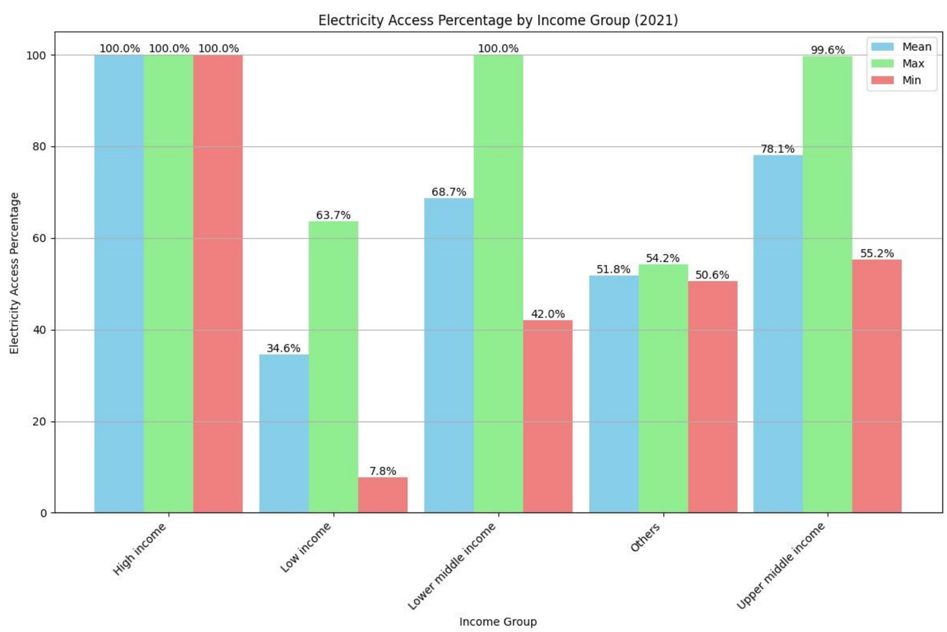


Fig 6. Electricity access percentage by income group (2021)

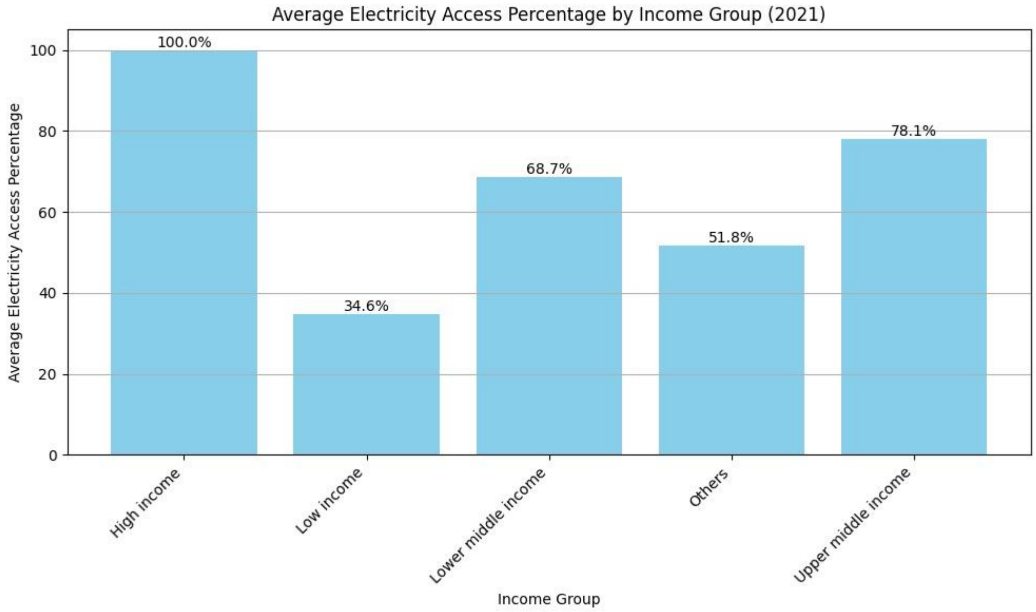


Fig 7. Average electricity access percentage by income group (2021)

The graphs above emphasize the differences between the percentage of electricity access and varied income levels. At 100%, the high-income countries have the best electricity access within this income bracket. A significant majority of individuals in upper-middle-income countries have access to electricity (78.1%), but there is still a portion (21.9%) without access. Over two-thirds of the lower-middle-income group have access to electricity, but access is less prevalent compared to the upper-middle-income countries (68.7%). Only about one-third of the low-income group has access to electricity. This highlights a substantial lack of access among lower-income countries. Overall, high-income countries have universal access, whereas lower-income individuals face significant challenges. Targeted efforts to improve electricity access could particularly focus on the low-income and lower-middle-income groups, where access is significantly lower.

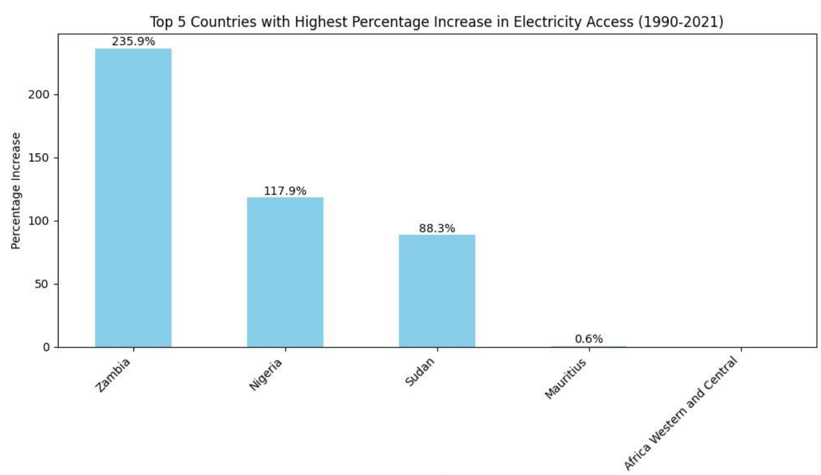


Fig 9. Top 5 countries with significant improvements in electricity access (1990-2021)

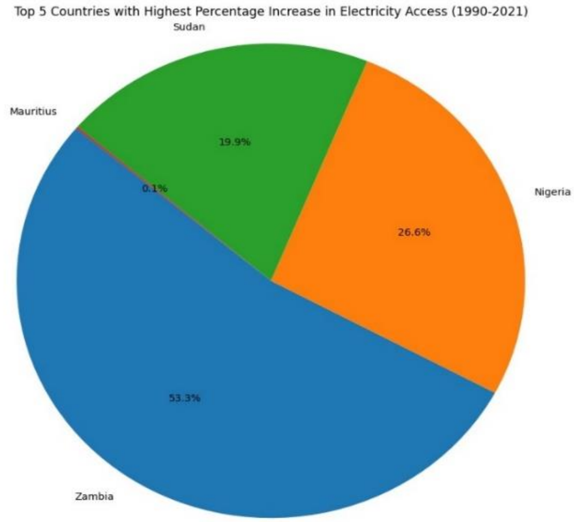


Fig 10. Top 5 countries with the highest percentage increase in electricity access (1990-2021)

This chart displays the top 5 countries in Africa with the highest percentage increase in electricity access from 1990 to 2021. Each segment represents a different country and the proportion of the total increase attributed to that country.

Zambia with 53.3%, has experienced the highest percentage increase in electricity access among the top 5 countries during this period. Nigeria has the second-largest segment, showing a significant increase in electricity access, though substantially less than Zambia with 26.6%. Sudan (19.9%), follows Nigeria with the third-largest segment, contributing a notable increase in electricity access. Mauritius shows a minimal increase of 0.1%, suggesting that while it has improved, its progress is not as pronounced as the other countries listed. Overall, this chart highlights the varying degrees of progress in electricity access among different African countries, with Zambia leading significantly. It showcases the disparities in development and access to electricity infrastructure across the continent.

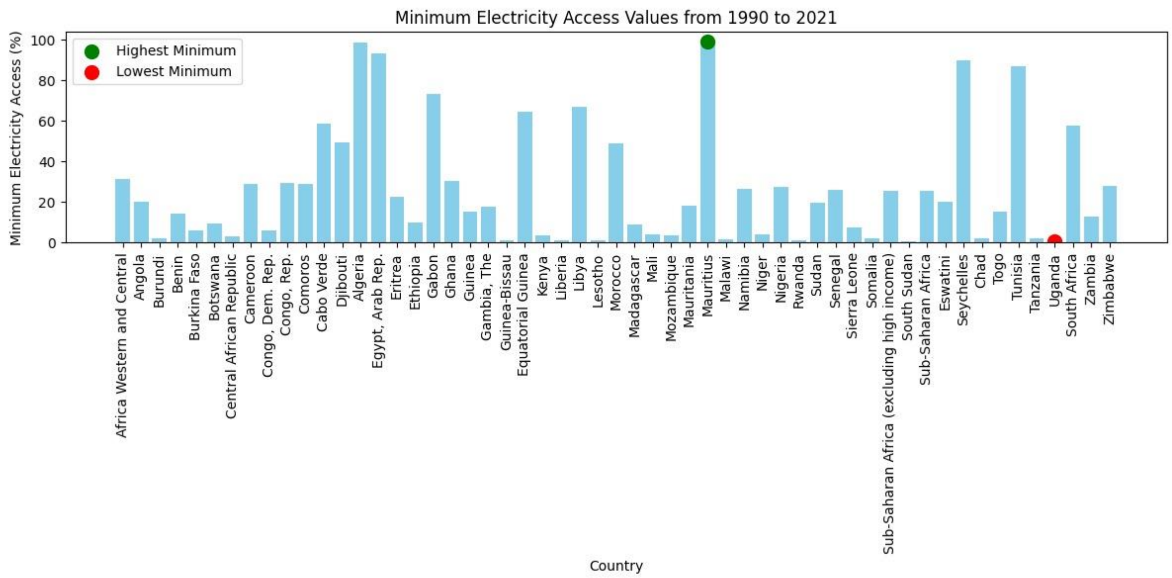


Fig 11. Minimum electricity access values from 1990-2021

The graph presents the minimum electricity access values for various countries and regions from 1990 to 2021. Several countries and regions have very low minimum electricity access values, often below 20%. For example, Chad and Malawi show particularly low values. Zimbabwe is highlighted with a red dot, indicating it has the lowest minimum electricity access value among all the listed countries and regions. Some countries have significantly higher minimum access values, with Mauritius marked by a green dot, indicating it has the highest minimum electricity access among the listed countries and regions. Other countries with relatively high minimum values include Egypt, Algeria, and South Africa. There is a wide variation in minimum electricity access across the different countries and regions. This indicates differing levels of infrastructure development and investment in electricity access over the years.

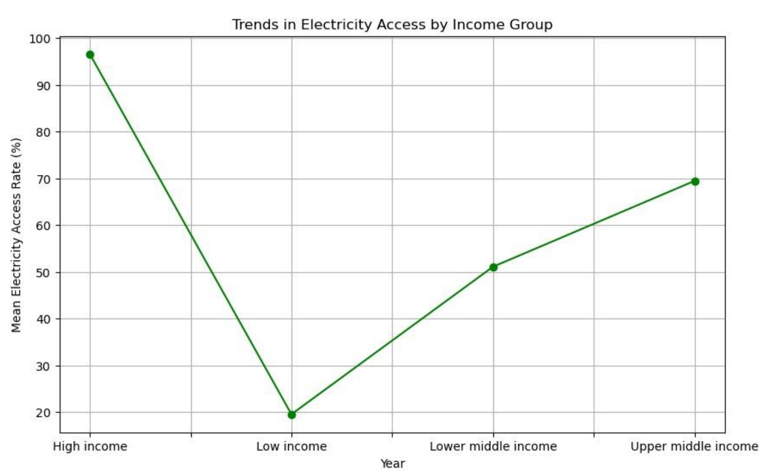


Fig 12. Trends in electricity access by income group

This graph reveals a clear correlation between income level and electricity access: higher-income groups have higher electricity access rates. High-income countries enjoy near-universal access, while low-income countries face significant challenges in providing electricity to their populations. Middle-income countries show progressive improvements, with upper-middle-income countries having a higher access rate than lower-middle-income countries.

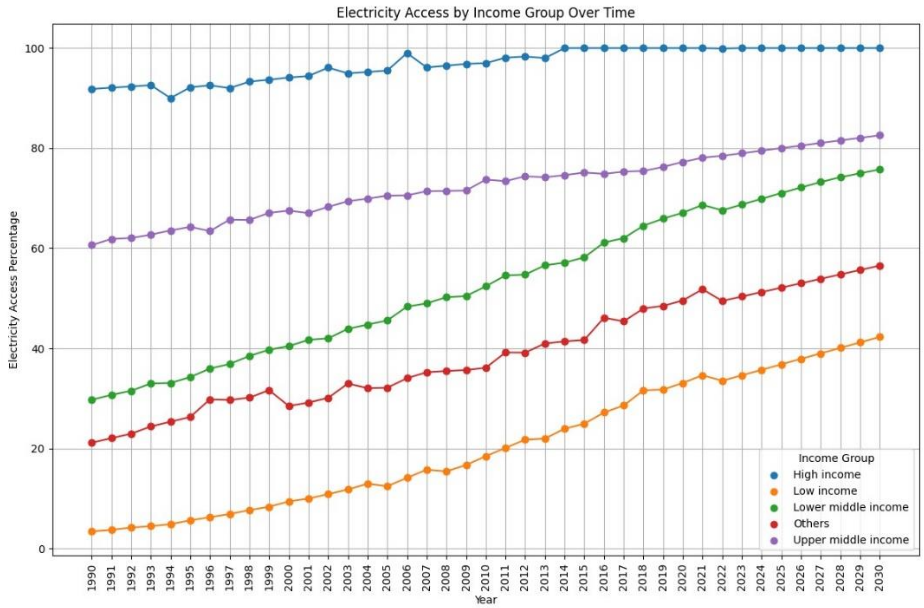


Fig 13. Electricity access by income group over time

The graph illustrates the trend of electricity access over time (from 1990 to 2020) for different income groups. High-income countries have consistently high electricity access, nearing 100% throughout the period. Although minor fluctuations are observed overall access remains widespread. For Upper Middle-Income countries, electricity access starts at around 80% in 1990 and steadily increases to nearly 100% by 2020. This shows significant improvements over the years, reflecting ongoing development and electricity infrastructure investment. Lower Middle-Income country's electricity access begins at around 50% in 1990 and steadily rises to approximately 85% by 2020. A consistent upward trend indicates continuous improvements and efforts to expand electricity access. Low-income countries began with very low access (around 10%) in 1990 and increased to about 50% by 2020. While the trend is positive, the rate of increase is slower compared to higher-income groups. This reflects challenges in infrastructure development in low-income countries. The category "Others" represents areas not specifically classified within the defined income groups. These countries show substantial progress but highlight that access remains relatively low.

We can conclude that income is inversely proportional to electricity access, which has seen steady growth over the years. According to Sarkodoe and Adams (2020), income disparity was found to negatively impact per capita electricity consumption. However, the relationship between electricity access and income inequality depends on the level of development.

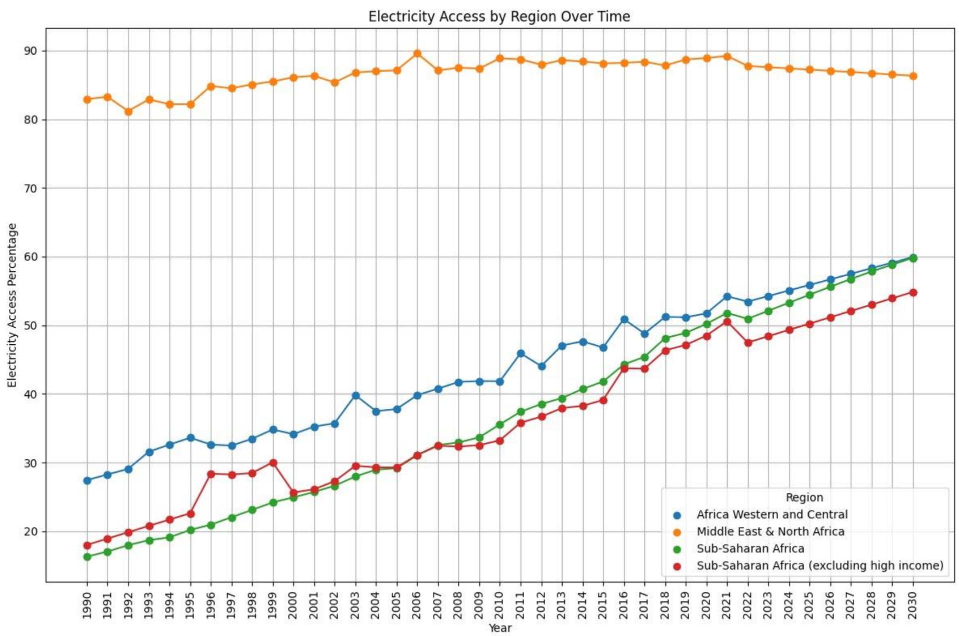


Fig 14. Electricity access by region over time

The graph shows the percentage of the countries with access to electricity in various regions of Africa from 1990 to 2020. The Middle East & North Africa region has the highest and most stable electricity access over the entire period, electricity access starts at around 86% in 1990 and remains consistently high, reaching about 98% by 2020.

For Africa Western and Central, there is a gradual but steady increase in electricity access. Starting at over 20% in 1990 and by 2020, access has reached approximately 55%, showing significant improvement over the 30 years.

Sub-Saharan Africa and Sub-Saharan Africa (excluding high income) also show a similar steady increase over time. Starting at around 20% in 1990, with a steady increase, reaching approximately 45% by 2020 but still lower overall access compared to other regions.

Overall, the graph demonstrates substantial progress in increasing electricity access across all regions, with the Middle East & North Africa achieving near-universal access and the other regions showing steady improvements over the 30 years. However, significant regional disparities remain, highlighting ongoing challenges in achieving widespread electricity access in Sub-Saharan Africa.

CORRELATION PLOT BETWEEN INCOME AND REGION

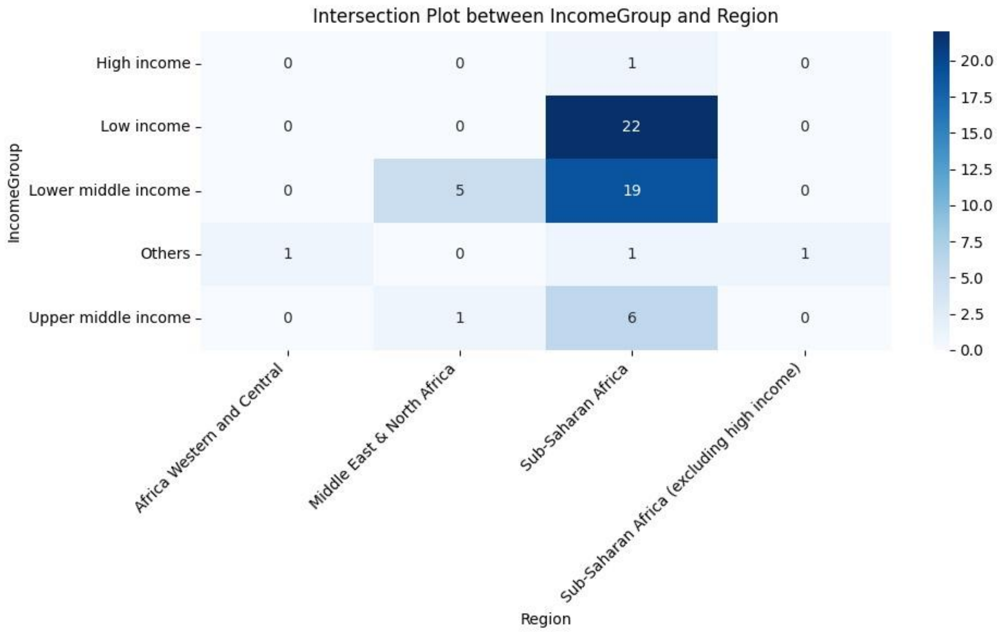


Fig 15. Intersection plot between income group and region

This chart is a heat map showing the intersection between income groups and regions. It visualises the number of occurrences of each income group within each region, with the colour intensity representing the count. The darker shades indicate higher counts, while lighter shades indicate lower counts. High income has very few occurrences, with 1 instance in Sub-Saharan Africa (excluding high income). Low income is concentrated mainly in Sub-Saharan Africa, with 22 occurrences and no occurrences in other regions. Lower middle-income countries are predominantly found in Sub-Saharan Africa with 19 occurrences. However, they are also present in the Middle East & North Africa with 5 occurrences. Upper middle-income countries are found mostly in Sub-Saharan Africa with 6 occurrences and a smaller presence in the Middle East & North Africa with 1 occurrence. Other categories have scattered occurrences, with 1 instance each in Africa Western and Central, Sub-Saharan Africa, and Sub-Saharan Africa (excluding high income). This heat map effectively shows the distribution and concentration of various income groups across different regions, highlighting where low and lower-middle-income groups are most prevalent.

Predictive models to forecast future electricity access levels based on regional and income data.

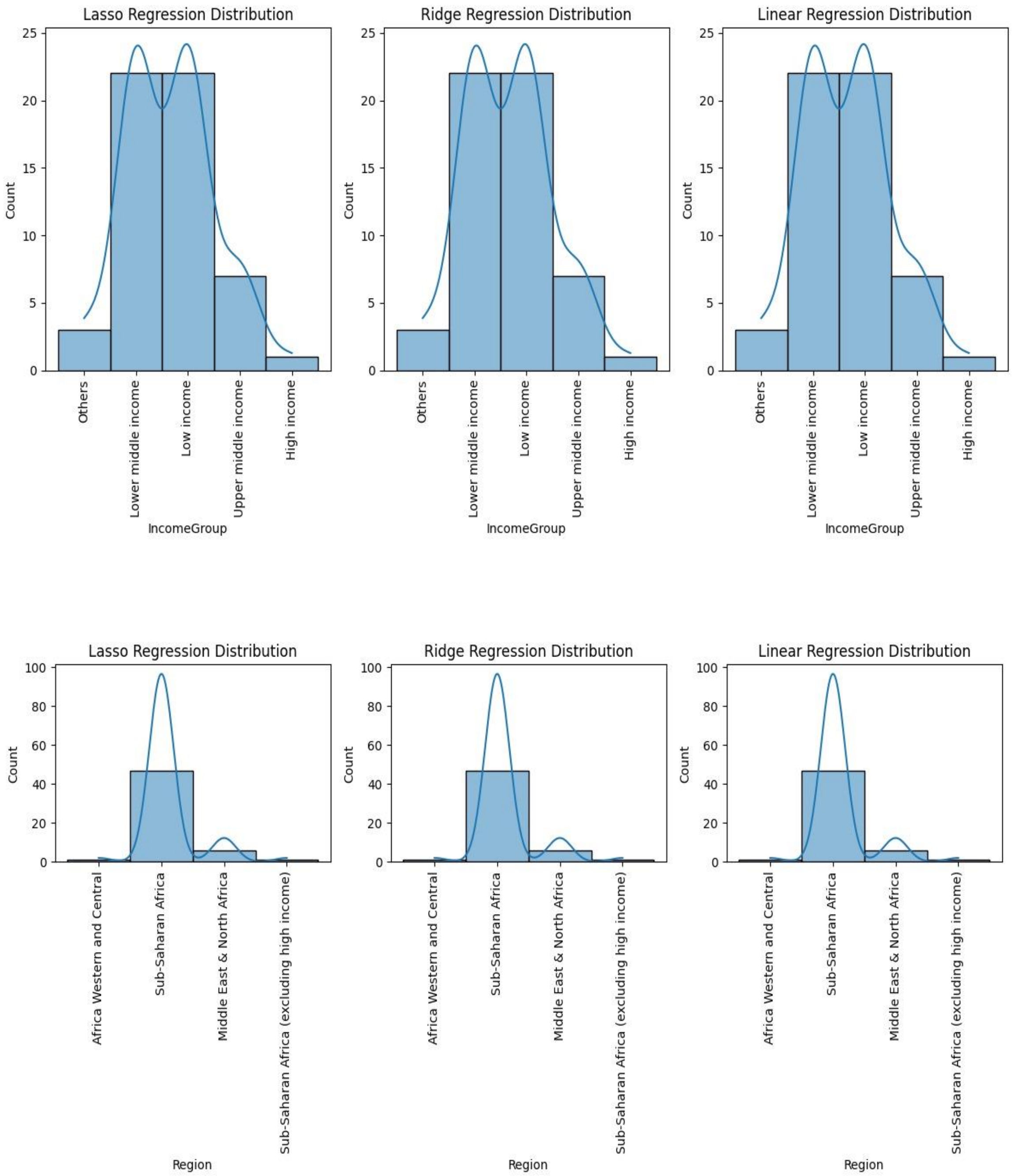


Fig 16. Predictive models to forecast future electricity access levels based on regional and income data.

Seasonal Decomposition of Electricity Access

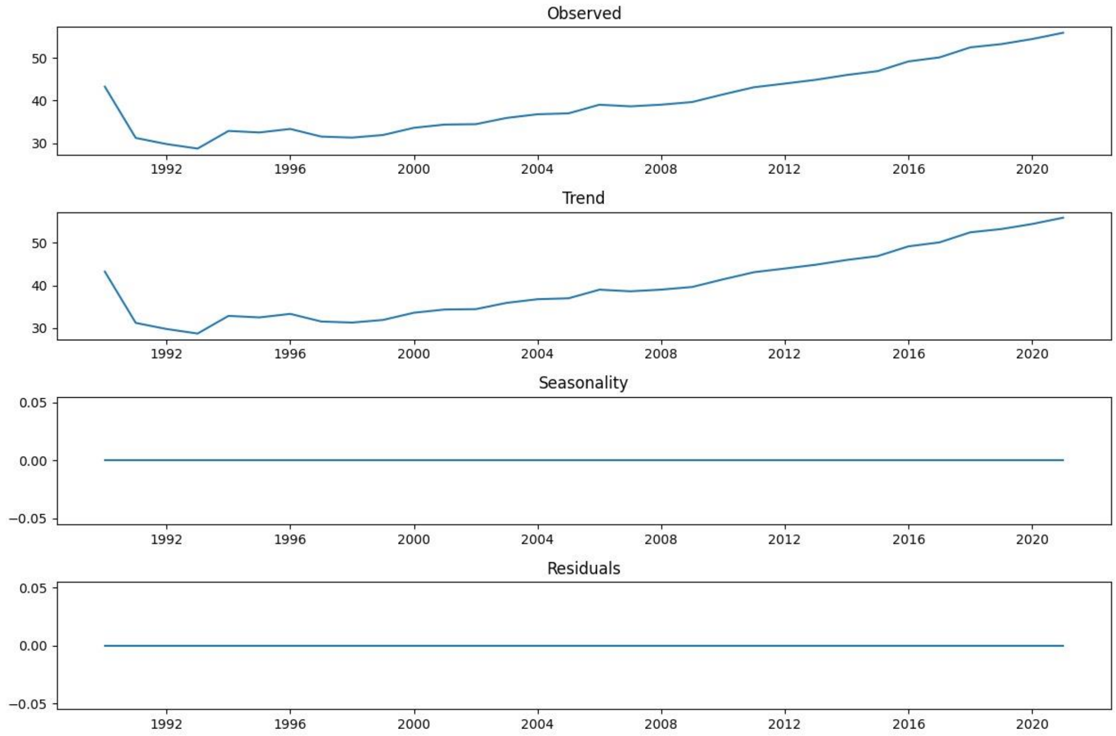


Fig 17. Seasonal decomposition of electricity access

The graph is a decomposition of the time series data for electricity access in Africa from 1990 to 2020. The decomposition breaks down the data into four components: Observed, Trend, Seasonality, and Residuals.

The Observed Plot shows the actual observed data for electricity access over time where there's an initial decline around 1992, followed by a general increase with some fluctuations.

From about 2004 onward, the increase becomes more pronounced, continuing steadily upward through 2020.

The Trend Plot isolates the long-term trend component from the observed data. It shows a smoother, more continuous increase in electricity access over the entire period. The initial dip around 1992 is less pronounced, and the overall upward trend is clearer and more consistent from around 2000 onwards.

The Seasonality Plot shows the seasonal component of the data. In this case, the seasonality component is essentially flat, indicating that there is no significant seasonal pattern in the data. This makes sense as electricity access data is not typically seasonal. The flat line at zero suggests that seasonality does not play a role in this particular time series.

The residual plot shows the residual component, which represents the noise or random variation left after removing the trend and seasonality. It is also depicted as flat and centred on zero, indicating that the model's trend and seasonality components have captured most of the patterns in the data. The lack of significant residuals suggests a good fit of the trend component to the observed data.

Overall, the decomposition shows that the primary pattern in the electricity access data is the long-term upward trend, with no significant seasonal effects or large residual variations.

The upward trend indicates consistent improvement in electricity access in Africa over the 30 years. The flat seasonality and residuals plots imply that the trend component is the dominant feature in the data, capturing the majority of the observed changes in electricity access

Valuable analysis for Targeted interventions in Electricity access

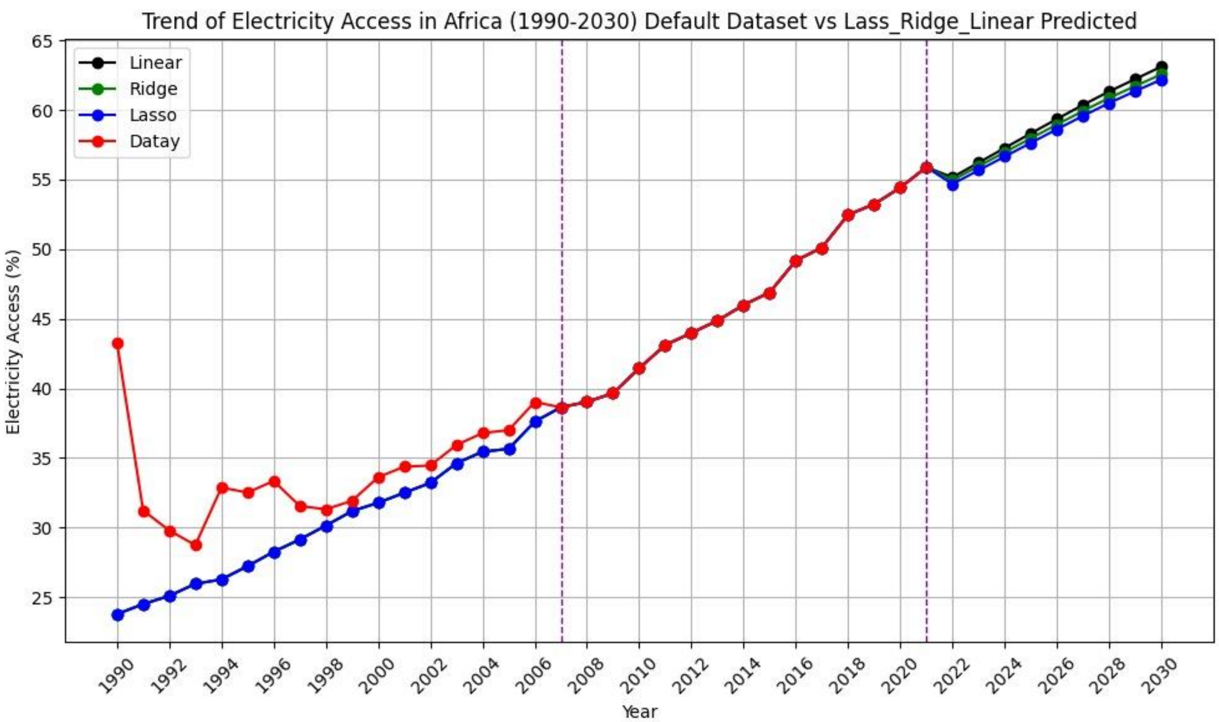


Fig 18. Trend of electricity access in Africa (1990-2030) default dataset vs lasso_ridge_linear reg predicted

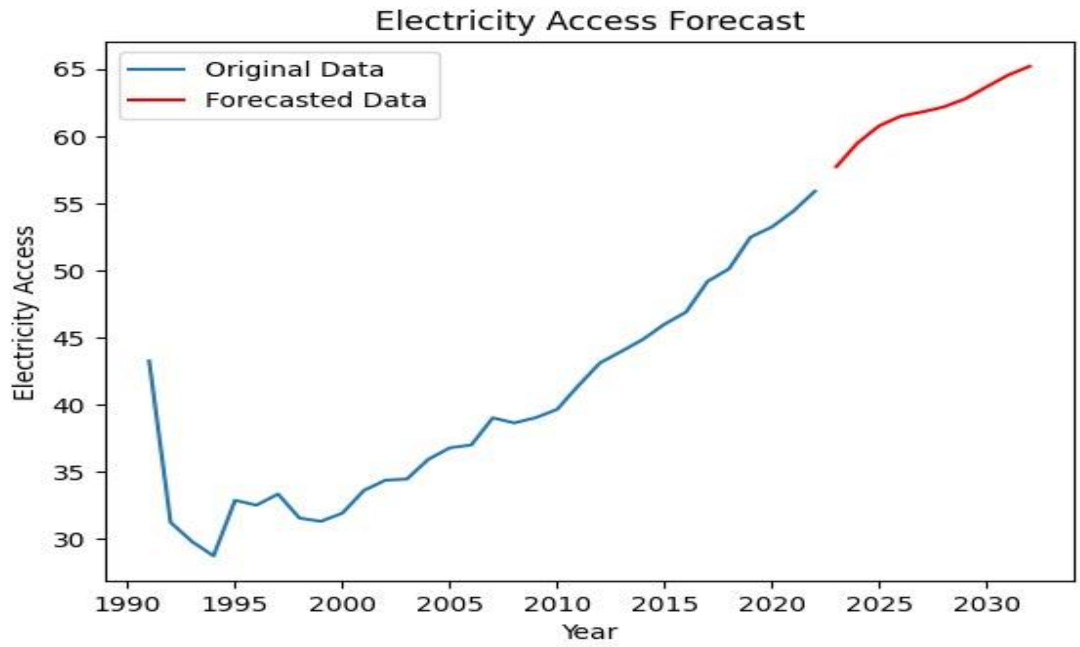


Fig 19. Average electricity access forecast

The graph presents the trend of electricity access in Africa from 1990 to 2030, comparing actual data with predicted values using three different regression models: Linear, Ridge, and Lasso.

The red line with circular markers represents the actual data for electricity access from 1990 to around 2021. There is a noticeable dip in the data around 1992, after which the access rate generally increases steadily. Three regression models are used to predict future electricity access (Linear, Ridge and Lasso) starting from 2022 and extending to 2030. These models show very similar predictions, indicating a consistent expected increase in electricity access over the years 2022 to 2030. These predictions closely follow the trend of the actual data from earlier years.

Between 2000 and 2010, there was a more consistent upward trend. Post-2010, the increase in electricity access became steeper, indicating accelerated progress. The predictions for 2022 to 2030 suggest a continuation of this upward trend, projecting electricity access to reach close to 60% by 2030.

Overall, the graph indicates significant progress in increasing electricity access in Africa over the past few decades, with regression models predicting continued improvement into the future.

RESULTS

After we have trained our model, tested it and evaluated its performance, here is how the models performed.

For the ridge model, here are the performance metrics

A	B	C	D	E	F
	Country Name	MAE	MSE	RMSE	R2
0	Africa Western and Central	0.773917264	1.127725607	1.061944258	0.987969534
1	Angola	0.763671652	3.768396411	1.941235795	0.979563536
2	Burundi	0.38046987	0.351847076	0.593166988	0.979943559
3	Benin	0.744511933	3.03508657	1.742149985	0.977738689
4	Burkina Faso	0.349761026	0.682540112	0.826159859	0.978323215
5	Botswana	0.970504636	2.342260796	1.53044464	0.996773056

The full table can be found here: ridge model results

For the linear regression model, here are the performance metrics

A	B	C	D	E	F
	Country Name	MAE	MSE	RMSE	R2
0	Africa Western and Central	1.092804824	1.594363409	1.26268104	0.966711861
1	Angola	1.420108311	7.02290289	2.650076016	0.88087426
2	Burundi	0.475411836	0.520026253	0.721128458	0.922635622
3	Benin	1.171891124	4.786086414	2.187712599	0.917968985
4	Burkina Faso	0.493410905	0.964965094	0.982326369	0.941055161
5	Botswana	1.28246462	3.097797497	1.760056106	0.992562927

The full table can be found here: linear regression model results

For the lasso model, here are the performance metrics

A	B	C	D	E	F
	Country Name	MAE	MSE	RMSE	R2
0	Africa Western and Central	0.811306705	1.134865724	1.065300767	0.98789075
1	Angola	0.814592759	3.775534814	1.943073548	0.979521686
2	Burundi	0.414035693	0.356345949	0.596947191	0.979428653
3	Benin	0.792470468	3.042225487	1.744197663	0.9776836
4	Burkina Faso	0.396651568	0.68968206	0.830470987	0.978091458
5	Botswana	0.985740801	2.349385543	1.532770545	0.996762844

The full table can be found here: lasso model results

CONCLUSION

Ridge, Lasso, and linear regression predictive models are significant in addressing electricity access challenges in Africa. These models help in predicting electricity demand, optimizing energy distribution, and identifying areas that require infrastructure development. Ridge and Lasso regression techniques are particularly useful for handling multicollinearity and feature selection, which are common issues in energy access data. Linear regression, on the other hand, provides a baseline model for predicting electricity consumption patterns. By utilizing these models, stakeholders can make informed decisions, allocate resources effectively, and improve the overall efficiency of electricity distribution systems in Africa.

Also, our research makes a significant contribution to the field of electricity usage by providing valuable insights into optimizing energy distribution and addressing challenges in electricity access. The use of advanced regression techniques like Ridge, Lasso, and linear regression enhances the accuracy of predictions and aids in efficient resource allocation. Moving forward, potential future research directions could include exploring the integration of renewable energy sources into predictive models, analyzing the impact of policy interventions on electricity access, and investigating the role of smart grid technologies in improving energy distribution in Africa. These avenues can further advance the understanding and implementation of predictive modeling for sustainable energy access in the region.

Predictive modeling plays a crucial role in enhancing electricity access in Africa by enabling better planning and allocation of resources. By using predictive models, policymakers and energy providers can forecast electricity demand, optimize distribution networks, and identify areas in need of infrastructure development. This visionary approach helps in reducing energy shortages, improving reliability, and ensuring more efficient energy delivery to communities. The practical implications of predictive modeling include cost savings through optimized resource allocation, increased energy efficiency, and ultimately, a positive impact on the quality of life for people in Africa by providing more reliable access to electricity.