Comprehensive Literature Review on Electricity Tariff Reforms, Social Equity, and Machine Learning in Energy Pricing.

I. Introduction (Economic/ applied approach)

Electricity tariff reforms in Sub-Saharan Africa play a crucial role in ensuring the financial viability of power utilities while maintaining affordability for consumers. However, the transition from subsidized to cost-reflective tariffs often results in disproportionate burdens on low-income households, exacerbating energy poverty and income inequality. This literature review synthesizes the findings of four key studies that examine the impact of electricity tariff reforms on social equity. The review provides a detailed discussion of data sources, key variables, and methodological approaches used in these studies, followed by an analysis of their findings.

II. Data Sources Used in the Literature

The studies reviewed utilized a combination of macro-level national data, household survey data, regulatory reports, and international energy datasets. The most commonly used sources include:

1. World Bank Energy & Development Indicators – Used to assess household electricity affordability, energy poverty rates, and income distribution.
2. Government Tariff Regulations & Utility Reports – Provide historical and current electricity pricing structures.
3. International Energy Agency (IEA) & African Development Bank Reports – Supply regional and national-level electricity access statistics.
4. Household Income and Expenditure Surveys (HIES) – Used to evaluate the proportion of household income allocated to electricity costs.
5. Smart Meter & Real-time Market Data – Employed in some studies to track household energy consumption behavior under different tariff schemes.

These datasets are often complemented with macroeconomic indicators, such as GDP per capita and inflation, to contextualize the socioeconomic effects of tariff adjustments.

III. Key Studies and Methodologies

This section presents a detailed methodological overview of each study, focusing on the models and empirical approaches used to analyze tariff reforms and social equity.

1) Electricity Sector Reforms in Sub-Saharan Africa (Asantewaa, 2023)

In her study titled "Electricity Sector Reforms in Sub-Saharan Africa" (2023), Asantewaa assessed the impact of tariff reforms across 37 Sub-Saharan African countries, focusing on affordability and access disparities. The study utilized a Difference-in-Differences (DID) approach to compare household energy affordability before and after major tariff policy changes.

- **Data Sources**: World Bank energy access data, national tariff regulations, African Development Bank reports.
- **Dependent Variable:** Household electricity affordability index.
- **Independent Variables:** Tariff structures (fixed vs. time-of-use pricing), subsidy levels, and regional GDP.
- **Methodology:**
    - The DID model identified variations in affordability between treatment groups (households affected by tariff changes) and control groups (households with stable tariffs).
    - The study also performed Gini coefficient analysis to measure the inequality in electricity access pre- and post-reform.
    - Regression Discontinuity Analysis (RDA) was employed to determine threshold effects, particularly when subsidy reductions led to an abrupt increase in non-payment rates.

Findings from the study indicate that households in low-income brackets experienced a 15-20% reduction in affordability, with urban consumers being more affected due to higher base electricity costs.

2) **Electricity Access, Human Development Index, and Income Inequality in Sub-Saharan Africa (Sarkodie & Adams, 2020)**

Sarkodie and Adams (2020), in their paper titled "Electricity Access, Human Development Index, and Income Inequality in Sub-Saharan Africa", examined the relationship between electricity pricing policies and socioeconomic development using panel data regression models.

- **Data Sources:** IEA energy statistics, World Bank development indicators, government tariff reports.
- **Dependent Variables**: Energy poverty index, Human Development Index (HDI), household income inequality (Gini coefficient).
- **Independent Variables:** Tariff price levels, governance quality, and national electrification rates.
- **Methodology:**
    - Fixed Effects & Random Effects Panel Models were used to capture variations across countries while controlling for unobserved heterogeneity.
    - Bayesian Model Averaging (BMA) was applied to identify the most statistically significant variables affecting energy access and inequality.
    - A Nonlinear Autoregressive Distributed Lag (NARDL) Model was employed to test the asymmetric effects of tariff increases and decreases over time.

The study found that energy affordability significantly influences human development, with low-income households allocating up to 30% of their income to electricity costs in countries with recent tariff hikes.

3) **Sector Reforms and Institutional Corruption: Evidence from the Electricity Industry in Sub-Saharan Africa (Imam, Jamasb & Llorca, 2019)**

Imam et al. (2019) analyzed the role of institutional corruption and governance in shaping the outcomes of electricity tariff reforms in their study titled "Sector Reforms and Institutional Corruption: Evidence from the Electricity Industry in Sub-Saharan Africa".

- **Data Sources:** Transparency International corruption index, national regulatory tariff reports, World Bank enterprise surveys.
- **Dependent Variable:** Household affordability index, non-payment rates.
- **Independent Variables:** Corruption index, regulatory changes, presence of cost-reflective tariffs.
- **Methodology:**
    - Instrumental Variable (IV) Regression was used to account for endogeneity between corruption levels and tariff reform outcomes.
    - Time-Series Analysis examined trends in tariff adjustments and electricity affordability over a 15-year period (2002-2017).
    - Fixed Effects Model controlled for country-specific governance characteristics.

The study concluded that higher corruption levels correlate with greater inequalities in electricity affordability, as subsidies are often misallocated in corrupt governance structures.

4) A Review of the Impacts of Electricity Tariff Reform in Africa (Klug et al., 2022)

Klug et al. (2022) provided a comprehensive review of electricity tariff reforms in Africa in their paper "A Review of the Impacts of Electricity Tariff Reform in Africa", highlighting common challenges and policy solutions.

- **Data Sources:** World Bank energy studies, national utility company reports, household expenditure surveys.
- **Dependent Variables:** Electricity consumption patterns, proportion of income spent on electricity.
- **Independent Variables:** Fixed charges, subsidy allocations, household income deciles.
- **Methodology**:
    - Comparative Policy Analysis examined different tariff structures, including cost-reflective pricing, subsidized models, and time-of-use pricing.
    - Cost-Benefit Analysis (CBA) assessed the economic trade-offs of tariff hikes.
    - Empirical Review synthesized quantitative findings from multiple case studies.

The study found that progressive block tariffs (charging higher rates for higher consumption levels) were most effective in protecting low-income households while ensuring cost recovery for utilities.

IV. Synthesis of Findings and Policy Implications

Across these four studies, several common themes emerge regarding the social equity impacts of electricity tariff reforms:

1. **Affordability and Energy Poverty**
    - Cost-reflective tariffs often increase the financial burden on low-income households.

    - Subsidy misallocation remains a persistent challenge, leading to inefficient support for vulnerable groups.
2. **Tariff Structure and Household Consumption**
    - Time-of-use tariffs have the potential to encourage demand-side energy management, but they disproportionately affect consumers with limited flexibility in electricity use (e.g., small businesses and informal households).
    - Prepaid metering is increasingly implemented as a tool to improve payment compliance, but it has raised concerns regarding self-disconnection among the poorest users.
3. **Institutional and Governance Factors**
    - Countries with high corruption levels tend to experience less effective tariff reforms, as regulatory decisions are often driven by political interests rather than economic efficiency.
    - Transparency and data-driven policy interventions are essential for improving reform outcomes.
4. **Policy Recommendations**
    - Targeted cash transfers may be more effective than generalized electricity subsidies.
    - Gradual price adjustments can mitigate abrupt financial shocks.
    - Decentralized renewable solutions (such as solar microgrids) can offer affordable alternatives for low-income consumers.

V. Conclusion

Electricity tariff reforms in Sub-Saharan Africa are a double-edged sword, balancing economic efficiency with social equity considerations. The studies reviewed demonstrate that poorly implemented reforms exacerbate income inequality and energy poverty, while well-structured, evidence-based policies can improve both financial sustainability and affordability. Future research should leverage real-time energy consumption data and predictive models to enhance policy targeting and reform design.

1.Introduction (Machine Learning)

Access to reliable and affordable electricity remains a significant challenge, particularly in Sub-Saharan Africa, where electrification rates are among the lowest globally. Electricity tariff structures, including cost-reflective pricing, block tariffs, and subsidies, play a crucial role in affordability and equity but require careful economic evaluation to balance financial sustainability and social impact.
Advancements in machine learning (ML) and artificial intelligence (AI) offer potential for improving electricity pricing and policy decisions, yet their integration into regulatory frameworks remains in its early stages. This review explores how AI and ML can enhance tariff structuring, policy optimization, and energy access modeling.
Key methodologies such as welfare analysis, difference-in-differences (DiD), and empirical policy evaluation help assess the effects of tariff reforms on consumers and utilities. Essential variables include tariff structures, income levels, subsidy allocations, and equity measures such as the Gini coefficient.
To develop a robust analytical framework, data from regulatory reports, household surveys, macroeconomic indicators, and electricity consumption records are consolidated. This review synthesizes insights from diverse studies, identifies research gaps, and explores AI-driven approaches to creating sustainable electricity pricing mechanisms.

2.Economic Models and Empirical Evidence

Economic models play a crucial role in understanding electricity pricing, tariff structures, and access policies. The literature reviewed spans diverse methodologies, including welfare analysis, difference-in-differences (DiD), machine learning-based forecasting, and empirical evaluations of power sector reforms. These approaches provide insights into how tariff structures impact affordability, energy poverty, and utility financial sustainability, particularly in Sub-Saharan Africa.

2.1. Welfare Analysis and Tariff Design

Welfare economics helps assess how different tariff structures influence consumer well-being and equity. Studies on electricity pricing, such as *A Review of Electricity Tariffs and Enabling Solutions for Optimal Energy Management*, examine cost-reflective tariffs, block pricing, and subsidized electricity models. The trade-offs between affordability and cost recovery are central to policy debates, as seen in *Why Are Connection Charges So High? An Analysis of the Electricity Sector in Sub-Saharan Africa*, which highlights the financial burdens imposed by high upfront connection costs. These studies emphasize the need for equitable pricing mechanisms that balance consumer affordability with utility sustainability.

2.2. Difference-in-Differences (DiD) and Policy Impact Evaluations

Empirical studies often employ DiD models to evaluate policy reforms, such as subsidy reductions or tariff restructuring. *Learning from Power Sector Reform: The Cases of Kenya and Uganda* provide case studies on the effectiveness of market liberalization, unbundling, and regulatory interventions. These reforms were expected to improve efficiency and increase electrification rates, but empirical evidence reveals mixed results. While some reforms led to improved cost recovery and private sector participation, others resulted in high consumer prices, underscoring the challenges of balancing financial viability with accessibility.

2.3. Machine Learning and Predictive Modeling

Advancements in ML provide new tools for analyzing electricity markets and consumer behavior. *Predictive Modeling of Energy Poverty with Machine Learning Ensembles* demonstrates how ML algorithms can enhance the identification of energy-poor households and predict consumption patterns. Similarly, *Forecasting Day-Ahead Electricity Prices: A Comparison of Time Series and Neural Network Models* compares traditional econometric models with ML-based forecasts, highlighting the potential of AI in improving price predictions. However, the role of ML in tariff design and regulatory decision-making remains underexplored, warranting further research.

2.4. Energy Access and Electrification Challenges

Electrification efforts are closely linked to economic development, yet the impact of grid expansion remains uncertain. *Electrification: When Does Electrification Work? Evidence from Sub-Saharan Africa* reveals stark disparities in adoption rates, where grid access does not always translate to household connections. Factors such as income levels, productive electricity use, and connection costs significantly influence outcomes. This aligns with findings from *Why Are Connection Charges So High?* which suggests that utilities set high upfront fees to compensate for low per-unit consumption, thereby limiting access for low-income households.

2.5. The Role of Synthetic Data and AI in Trustworthy Energy Policy

A critical gap in energy policy research is the availability of high-quality, granular data. *A Survey on the Use of Synthetic Data for Enhancing Key Aspects of Trustworthy AI in the Energy Domain* discusses the potential of synthetic data in mitigating privacy concerns and enhancing energy policy simulations.

AI-driven approaches could improve demand forecasting, optimize subsidy allocation, and refine equity-based tariff models, but their integration into regulatory frameworks requires further exploration.

3.Data Preprocessing Methods

3.1. Handling Missing Data

- **Mean/Median Imputation**: Missing household electricity consumption data in survey-based studies were imputed using the median consumption of similar income groups. (*Predictive Modeling of Energy Poverty with Machine Learning Ensembles*)
- **Mode Imputation for Categorical Variables**: Missing tariff structure classifications (fixed vs. time-of-use) in regulatory reports were filled using mode imputation. (*A Review of Electricity Tariffs and Enabling Solutions for Optimal Energy Management*)
- **K-Nearest Neighbors (KNN) Imputation**: Missing socioeconomic variables, such as income and energy poverty indicators, were estimated based on the nearest households with similar demographic profiles. (*A Survey on the Use of Synthetic Data for Enhancing Key Aspects of Trustworthy AI in the Energy Domain*)
- **Multiple Imputation by Chained Equations (MICE)**: Household survey data missing key responses on electricity expenditures were estimated through chained regressions. (*Learning from Power Sector Reform: The Case of Kenya and Uganda*)
- **Forward and Backward Filling**: Missing electricity price data in time series models were filled using temporal interpolation methods. (*Forecasting Day-Ahead Electricity Prices: A Comparison of Time Series and Neural Network Models*)

3.2. Outlier Detection and Treatment

- **Winsorization**: Extreme electricity prices and energy expenditure values were adjusted at the 1st and 99th percentiles to reduce skewness in tariff impact studies. (*Why Are Connection Charges So High? An Analysis of the Electricity Sector in Sub-Saharan Africa*)
- **Interquartile Range (IQR) Method**: Consumption anomalies in electricity demand datasets were identified and replaced if they exceeded 1.5 times the IQR. (*Predictive Modeling of Energy Poverty with Machine Learning Ensembles*)
- **Rolling Standard Deviation for Price Spikes**: Sudden spikes in electricity price forecasts were smoothed using rolling standard deviation thresholds. (*Forecasting Day-Ahead Electricity Prices*)
- **Isolation Forest Algorithm**: Used for detecting fraudulent electricity usage data and abnormal connection charge values. (*A Survey on the Use of Synthetic Data for Enhancing Key Aspects of Trustworthy AI in the Energy Domain*)

3.3. Feature Engineering

- **Electricity Affordability Ratio**: A new feature was created by dividing monthly electricity expenditures by household income to assess affordability levels. (*Why Are Connection Charges So High?*)
- **Energy Poverty Index (EPI)**: An aggregated measure of electricity consumption, connection status, and household income was developed to assess energy poverty levels. (*Predictive Modeling of Energy Poverty with Machine Learning Ensembles*)
- **Tariff Classification Encoding**: One-hot encoding was applied to categorize different pricing schemes (fixed, block, time-of-use) for predictive tariff modeling. (*A Review of Electricity Tariffs and Enabling Solutions for Optimal Energy Management*)

- **Time-Based Features**: Daily, weekly, and seasonal trends were extracted to enhance electricity demand forecasting accuracy. (*Forecasting Day-Ahead Electricity Prices*)

3.4. Data Normalization and Scaling

- **Min-Max Scaling**: Household electricity consumption, income, and tariff variables were scaled to the range [0,1] for better comparability across different datasets. (*Predictive Modeling of Energy Poverty with Machine Learning Ensembles*)
- **Z-Score Standardization**: Standardized all continuous numerical variables to have a mean of zero and a standard deviation of one to maintain consistency in regression models. (*A Review of Electricity Tariffs and Enabling Solutions for Optimal Energy Management*)
- **Log Transformation**: Applied to household income and electricity expenditure variables to handle skewness in tariff impact studies. (*Electrification: When Does Electrification Work? Evidence from Sub-Saharan Africa*)
- **Quantile Normalization**: Used in synthetic data generation to align distributions of socioeconomic variables across different datasets. (*A Survey on the Use of Synthetic Data for Enhancing Key Aspects of Trustworthy AI in the Energy Domain*)

3.5. Data Aggregation and Consolidation

- **Merging Multi-Source Data**: Combined regulatory reports, macroeconomic indicators, and household survey data to construct a comprehensive dataset for tariff analysis. (*A Review of Electricity Tariffs and Enabling Solutions for Optimal Energy Management*)
- **Spatial Data Aggregation**: Electrification data were grouped at the district level to analyze regional variations in energy access. (*Electrification: When Does Electrification Work?*)
- **Rolling Mean Smoothing**: Applied to electricity price time series to reduce volatility and enhance forecasting accuracy. (*Forecasting Day-Ahead Electricity Prices*)

4. Model Estimation Methods

4.1. Table of Regression Models Used in Each Paper

Paper Title	OLS	IV (2SLS)	DiD	FE/RE Panel	Quantile	Probit/Logit	Multinomial Logit	Tobit
Predictive Modeling of Energy Poverty	**Yes**	-	-	-	**Yes**	**Yes**	-	**Yes**
A User-Centric View of Demand Side Mgmt	-	-	-	-	-	-	-	-
A Survey on Synthetic Data for AI in Energy	-	-	-	-	-	-	-	-
Learning from Power Sector Reform (Kenya)	-	**Yes**	**Yes**	-	-	-	-	-
Learning from Power Sector Reform (Uganda)	-	**Yes**	**Yes**	-	-	-	-	-

Forecasting Day-Ahead Electricity Prices	**-**	**Yes**	**-**	**-**	**-**	**-**	**-**	**-**
Electrification: When Does It Work?	**Yes**	**-**	**Yes**	**Yes**	**-**	**Yes**	**Yes**	**-**
A Review of Electricity Tariffs	**Yes**	**-**	**-**	**Yes**	**Yes**	**-**	**Yes**	**-**
Why Are Connection Charges So High?	**Yes**	**Yes**	**Yes**	**Yes**	**-**	**Yes**	**-**	**-**

Justification for Regression-Based Models Used in the Papers:

4.1.1. Ordinary Least Squares (OLS) Regression

Why was OLS used?

- **To estimate relationships between electricity pricing, energy consumption, and socioeconomic factors.**
- **To assess the impact of tariff structures on affordability and consumer behavior.**

Papers Using OLS and Their Justifications

- **A Review of Electricity Tariffs and Enabling Solutions for Optimal Energy Management →** Used to estimate how different tariff structures (fixed charges, block pricing, and time-of-use pricing) influence household electricity consumption.
- **Why Are Connection Charges So High? →** Assesses the effect of high upfront costs on electrification rates and household willingness to connect to the grid.
- **Electrification: When Does Electrification Work? →** Examines the correlation between access to electricity and economic development indicators such as income growth and employment.

4.1.2. Instrumental Variable (IV) Regression (Two-Stage Least Squares - 2SLS)

Why was IV used?

- **To address endogeneity in electricity pricing and policy analysis.**
- **To correct for biases arising from external factors influencing electricity adoption and affordability.**

Papers Using IV and Their Justifications

- **Forecasting Day-Ahead Electricity Prices →** Uses external shocks (hydroelectric generation fluctuations) as an instrument to correct for endogeneity in price forecasting models.
- **Learning from Power Sector Reform (Kenya & Uganda) →** Uses policy-driven changes in the electricity sector as an exogenous instrument to evaluate the impact of deregulation.
- **Why Are Connection Charges So High? →** Uses external price factors (e.g., fuel price fluctuations) as an instrument to assess how high connection costs affect electrification.

4.1.3. Difference-in-Differences (DiD) Estimation

Why was DiD used?

- **To evaluate the causal impact of electrification programs, tariff reforms, and power sector deregulation.**
- **To compare pre- and post-policy effects on electricity access, affordability, and household welfare.**

Papers Using DiD and Their Justifications

- **Learning from Power Sector Reform (Kenya & Uganda)** → Measures how market liberalization and tariff changes influenced electricity access over time.
- **Why Are Connection Charges So High?** → Evaluates the effect of subsidized vs. non-subsidized connection fees by comparing adoption rates before and after subsidy introduction.
- **Electrification: When Does Electrification Work?** → Examines the impact of rural electrification programs by comparing electrified and non-electrified households before and after grid expansion.

4.1.4. Fixed Effects (FE) & Random Effects (RE) Panel Models

Why were FE & RE models used?

- **To control for unobserved heterogeneity in electricity pricing, consumption, and policy evaluation.**
- **To analyze variations in electricity access across different regions and time periods.**

Papers Using FE/RE and Their Justifications

- **A Review of Electricity Tariffs** → Uses FE models to remove unobserved factors affecting electricity pricing across different locations.
- **Why Are Connection Charges So High?** → Uses RE models to account for country-level variations in electrification rates.
- **Electrification: When Does Electrification Work?** → Applies FE models to study the long-term effects of electrification on economic growth.

4.1.5. Quantile Regression

Why was Quantile Regression used?

- **To examine how electricity tariffs impact different income groups, especially low-income households.**
- **To assess affordability disparities in energy access.**

Papers Using Quantile Regression and Their Justifications

- **Predictive Modeling of Energy Poverty** → Evaluates how the impact of electricity costs differs across income percentiles.
- **A Review of Electricity Tariffs** → Measures how low-income vs. high-income households respond to pricing changes.

4.1.6. Probit & Logit Regression Models

Why were Probit & Logit Models used?

- **To predict the likelihood of households adopting electricity connections.**
- **To model binary decisions related to electricity tariffs and energy affordability.**

Papers Using Probit & Logit and Their Justifications

- **Electrification: When Does Electrification Work?** → Uses **Probit Regression** to estimate the probability of a household adopting grid electricity based on socioeconomic factors.
- **Why Are Connection Charges So High?** → Uses **Logit Regression** to model how high connection charges influence household decisions on whether to connect to the grid.
- **Predictive Modeling of Energy Poverty** → Uses **Logit Regression** to classify households based on electricity affordability and consumption patterns.

4.1.7. Multinomial Logistic Regression

Why was Multinomial Logistic Regression used?

- **To analyze categorical choices between different electricity pricing plans and tariff structures.**

Papers Using Multinomial Logistic Regression and Their Justifications

- **A Review of Electricity Tariffs** → Classifies household choices among **fixed, time-of-use, and block pricing tariffs**.
- **Electrification: When Does Electrification Work?** → Categorizes **electrification adoption patterns** based on infrastructure access and economic indicators.

4.1.8. Tobit Regression

Why was Tobit used?

- **To model censored data where electricity consumption is limited by affordability constraints.**

Papers Using Tobit and Their Justifications

- **Predictive Modeling of Energy Poverty** → Models **household electricity expenditures**, considering affordability constraints.

4.2. Efficiency and Productivity Estimation models

Table for Efficiency & Productivity Estimation Models

Paper Title	Total Factor Productivity (TFP) Estimation	Stochastic Frontier Analysis (SFA)	Data Envelopment Analysis (DEA)
Predictive Modeling of Energy Poverty	Yes	-	-
A User-Centric View of Demand Side Mgmt	-	-	-
A Survey on Synthetic Data for AI in Energy	-	-	-
Learning from Power Sector Reform (Kenya)	Yes	-	-
Learning from Power Sector Reform (Uganda)	Yes	-	-
Forecasting Day-Ahead Electricity Prices	-	-	-
Electrification: When Does Electrification Work?	-	Yes	-
A Review of Electricity Tariffs	-	-	Yes
Why Are Connection Charges So High?	-	Yes	-

4.2.1. Total Factor Productivity (TFP) Estimation

Why was TFP used?

- **To measure the impact of electricity access and power sector reforms on economic productivity.**
- **To analyze how changes in electricity supply and tariffs influence firm- and household-level efficiency.**
- **To assess whether electrification leads to improvements in labor productivity and economic growth.**

Papers Using TFP and Their Justifications

- **Predictive Modeling of Energy Poverty** → Used TFP estimation to measure how electricity access influences overall economic productivity at the household level, particularly in rural areas.
- **Learning from Power Sector Reform (Kenya & Uganda)** → Applied TFP models to evaluate how energy sector reforms impacted labor and capital efficiency in industrial and commercial sectors.

4.2.2. Stochastic Frontier Analysis (SFA)

Why was SFA used?

- **To assess inefficiencies in electricity consumption and grid expansion.**
- **To determine whether pricing structures, subsidies, or connection charges create inefficiencies in the electricity market.**

- **To measure how effectively electricity providers convert inputs (tariffs, subsidies, connection infrastructure) into outputs (electrified households, revenue recovery).**

Papers Using SFA and Their Justifications

- **Electrification: When Does Electrification Work?** → Used SFA to analyze inefficiencies in electricity consumption due to economic barriers, unreliable supply, or lack of productive use of electricity.
- **Why Are Connection Charges So High?** → Applied SFA to assess whether high connection charges create inefficiencies by discouraging low-income households from accessing the grid, leading to suboptimal electrification rates.

4.2.3. Data Envelopment Analysis (DEA)

Why was the DEA used?

- **To evaluate the cost-effectiveness of electricity pricing models.**
- **To compare the efficiency of different tariff structures (fixed pricing, time-of-use pricing, block pricing) in ensuring affordability while maintaining financial sustainability.**
- **To measure the performance of electricity distribution companies in achieving economic and policy goals.**

Papers Using DEA and Their Justifications

- **A Review of Electricity Tariffs** → Used DEA to compare the efficiency of different tariff structures in balancing cost recovery for utilities and affordability for consumers.

4.3. Machine Learning-Based Models

Paper Title	**K-Means Clustering**	**Random Forest Regression**	**Gradient Boosting (GBM, XGBoost, LightGBM)**	**Support Vector Regression (SVR)**	**Neural Networks (MLP, LSTM, Transformers)**
Predictive Modeling of Energy Poverty	**Yes**	**Yes**	**Yes**	**-**	**-**
A User-Centric View of Demand Side Mgmt	**-**	**-**	**-**	**-**	**-**
A Survey on Synthetic Data for AI in Energy	**-**	**-**	**Yes**	**-**	**Yes**
Learning from Power Sector Reform (Kenya)	**-**	**-**	**-**	**-**	**-**

Learning from Power Sector Reform (Uganda)	**-**	**-**	**-**	**-**	**-**
Forecasting Day-Ahead Electricity Prices	**-**	**Yes**	**Yes**	**Yes**	**Yes**
Electrification: When Does It Work?	**-**	**-**	**-**	**-**	**-**
A Review of Electricity Tariffs	**-**	**-**	**-**	**-**	**-**
Why Are Connection Charges So High?	**-**	**-**	**-**	**-**	**-**

Justification for Machine Learning-Based Models Used in the Papers

4.3.1. K-Means Clustering

Why was K-Means used?

- **To group households based on energy affordability and consumption behavior.**
- **To identify patterns of energy poverty and categorize users based on economic and geographic factors.**

Papers Using K-Means and Their Justifications

- **Predictive Modeling of Energy Poverty** → Used K-Means clustering to group households based on electricity affordability, access patterns, and socioeconomic indicators.

4.3.2. Random Forest Regression

Why was Random Forest used?

- **To predict electricity consumption based on historical and socioeconomic data.**
- **To improve model accuracy and robustness by reducing overfitting.**

Papers Using Random Forest and Their Justifications

- **Predictive Modeling of Energy Poverty** → Used Random Forest to predict electricity consumption patterns and classify households based on affordability.
- **Forecasting Day-Ahead Electricity Prices** → Applied Random Forest regression to improve electricity price predictions by incorporating time-series market data.

4.3.3. Gradient Boosting Models (GBM, XGBoost, LightGBM)

Why was Gradient Boosting used?

- **To enhance predictive accuracy in energy forecasting and household energy consumption classification.**
- **To capture complex relationships between variables with improved efficiency.**

Papers Using Gradient Boosting and Their Justifications

- **Predictive Modeling of Energy Poverty** → Used GBM models for high-accuracy predictions in classifying energy poverty levels.
- **A Survey on Synthetic Data for AI in Energy** → Used XGBoost to validate synthetic data reliability for energy policy research.
- **Forecasting Day-Ahead Electricity Prices** → Applied LightGBM and XGBoost for improving price forecasting with large-scale datasets.

4.3.4. Support Vector Regression (SVR)

Why was SVR used?

- **To handle non-linear relationships in electricity pricing and consumption forecasting.**
- **To improve performance in datasets with high variance and complex features.**

Papers Using SVR and Their Justifications

- **Forecasting Day-Ahead Electricity Prices** → Used SVR to capture non-linear trends in electricity price movements and improve long-term forecasting accuracy.

4.3.5. Neural Networks (MLP, LSTM, Transformers)

Why were Neural Networks used?

- **To improve time-series forecasting by capturing long-term dependencies in electricity price and consumption data.**
- **To enhance predictive performance in non-linear, high-dimensional datasets.**

Papers Using Neural Networks and Their Justifications

- **A Survey on Synthetic Data for AI in Energy** → Used neural networks to improve the quality of synthetic electricity consumption data.
- **Forecasting Day-Ahead Electricity Prices** → Applied **LSTM networks** for time-series forecasting and **transformer models** to predict price fluctuations with greater accuracy.

5. Model Validation Methods

Paper Title	Cross-Validation (K-Fold, Leave-One-Out)	Train-Test Split	Out-of-Sample Testing	Bootstrapping	Mean Absolute Error (MAE), RMSE, R^2	Holdout Validation
Predictive Modeling of Energy Poverty	**Yes**	**Yes**	**Yes**	**-**	**Yes**	**Yes**
A User-Centric View of Demand Side Mgmt	**-**	**-**	**-**	**-**	**-**	**-**
A Survey on Synthetic Data for AI in Energy	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**-**
Learning from Power Sector Reform (Kenya)	**-**	**-**	**Yes**	**-**	**-**	**-**
Learning from Power Sector Reform (Uganda)	**-**	**-**	**Yes**	**-**	**-**	**-**
Forecasting Day-Ahead Electricity Prices	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Electrification: When Does It Work?	**-**	**-**	**-**	**-**	**-**	**-**
A Review of Electricity Tariffs	**-**	**-**	**-**	**-**	**Yes**	**-**
Why Are Connection Charges So High?	**-**	**-**	**-**	**-**	**-**	**-**

5.1. Cross-Validation (K-Fold, Leave-One-Out)

Why was Cross-Validation used?

- **To prevent overfitting and ensure model generalizability.**
- **To optimize hyperparameters by systematically validating models on multiple subsets of data.**

Papers Using Cross-Validation and Their Justifications

- **Predictive Modeling of Energy Poverty** → Used **K-Fold Cross-Validation** to enhance model robustness in classifying energy-poor households.

- **A Survey on Synthetic Data for AI in Energy** → Applied **Leave-One-Out Cross-Validation (LOOCV)** to validate models on synthetic datasets.
- **Forecasting Day-Ahead Electricity Prices** → Used **Cross-Validation** to tune hyperparameters for machine learning models in price forecasting.

5.2. Train-Test Split

Why was Train-Test Split used?

- **To evaluate model performance on unseen data.**
- **To ensure models are trained on a subset of data and tested separately to check predictive accuracy.**

Papers Using Train-Test Split and Their Justifications

- **Predictive Modeling of Energy Poverty** → Splits data into training and testing sets to assess energy poverty classification performance.
- **A Survey on Synthetic Data for AI in Energy** → Uses **Train-Test Split** to test synthetic data models against real-world datasets.
- **Forecasting Day-Ahead Electricity Prices** → Implements **70-30 Train-Test Split** to train ML models on historical data while validating price forecasts.

5.3. Out-of-Sample Testing

Why was Out-of-Sample Testing used?

- **To validate models using data from different time periods or locations.**
- **To assess model transferability and performance in real-world applications.**

Papers Using Out-of-Sample Testing and Their Justifications

- **Predictive Modeling of Energy Poverty** → Tests energy poverty classification models on different geographic locations.
- **A Survey on Synthetic Data for AI in Energy** → Validates synthetic datasets by applying models to real-world datasets.
- **Learning from Power Sector Reform (Kenya & Uganda)** → Uses out-of-sample validation to assess policy reform impacts in different time periods.
- **Forecasting Day-Ahead Electricity Prices** → Tests models on unseen historical data to evaluate predictive accuracy.

5.4. Bootstrapping

Why was Bootstrapping used?

- **To improve statistical estimation when dealing with small datasets.**
- **To generate multiple samples from existing data to enhance model stability.**

Papers Using Bootstrapping and Their Justifications

- **A Survey on Synthetic Data for AI in Energy** → Uses bootstrapping to simulate additional training data and improve model reliability.
- **Forecasting Day-Ahead Electricity Prices** → Implements bootstrapping for generating confidence intervals in electricity price predictions.

5.5. Mean Absolute Error (MAE), Root Mean Square Error (RMSE), R^2 Score.

Why were these error metrics used?

- **To quantitatively measure model accuracy.**
- **To compare different models and select the most accurate approach.**

Papers Using MAE, RMSE, and R^2 and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses **MAE and RMSE** to evaluate the accuracy of energy affordability predictions.
- **A Survey on Synthetic Data for AI in Energy** → Compares real and synthetic datasets using **R^2 scores**.
- **Forecasting Day-Ahead Electricity Prices** → Applies **RMSE and MAE** to assess price forecasting model accuracy.
- **A Review of Electricity Tariffs** → Uses **MAE and RMSE** to measure tariff impact model accuracy.

5.6. Holdout Validation

Why was Holdout Validation used?

- **To test final models on a separate dataset not used during training or hyperparameter tuning.**
- **To ensure models are not overfitted to specific data points.**

Papers Using Holdout Validation and Their Justifications

- **Predictive Modeling of Energy Poverty** → Reserves a portion of data exclusively for final model evaluation.
- **Forecasting Day-Ahead Electricity Prices** → Uses holdout data from different time periods to validate electricity price forecasting models.

6. Performance Evaluation Metrics

Paper Title	Mean Absolute Error (MAE)	Root Mean Square Error (RMSE)	R^2 Score	F1-Score, Precision, Recall	Mean Squared Error (MSE)	Accuracy Score	Mean Absolute Percentage Error (MAPE)

Predictive Modeling of Energy Poverty	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
A User-Centric View of Demand Side Mgmt	**-**	**-**	**-**	**-**	**-**	**-**	**-**
A Survey on Synthetic Data for AI in Energy	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**-**
Learning from Power Sector Reform (Kenya)	**-**	**-**	**-**	**-**	**-**	**-**	**-**
Learning from Power Sector Reform (Uganda)	**-**	**-**	**-**	**-**	**-**	**-**	**-**
Forecasting Day-Ahead Electricity Prices	**Yes**	**Yes**	**Yes**	**-**	**Yes**	**-**	**Yes**
Electrification: When Does It Work?	**-**	**-**	**-**	**-**	**-**	**-**	**-**
A Review of Electricity Tariffs	**Yes**	**Yes**	**Yes**	**-**	**-**	**-**	**-**
Why Are Connection Charges So High?	**-**	**-**	**-**	**-**	**-**	**-**	**-**

6.1. Mean Absolute Error (MAE)

Why was MAE used?

- **To measure the absolute difference between actual and predicted values.**
- **To evaluate forecasting accuracy while being less sensitive to large errors compared to RMSE.**

Papers Using MAE and Their Justifications

- **Predictive Modeling of Energy Poverty** → Used MAE to measure errors in energy consumption and affordability predictions.
- **A Survey on Synthetic Data for AI in Energy** → Applied MAE to validate synthetic datasets against real-world energy data.
- **Forecasting Day-Ahead Electricity Prices** → Used MAE to quantify deviations between predicted and actual electricity prices.
- **A Review of Electricity Tariffs** → Employed MAE to assess how well tariff models predict consumer expenditures.

6.2. Root Mean Square Error (RMSE)

Why was RMSE used?

- **To penalize larger errors more heavily than MAE.**
- **To measure prediction accuracy in regression-based and machine learning models.**

Papers Using RMSE and Their Justifications

- **Predictive Modeling of Energy Poverty** → Used RMSE to evaluate the precision of machine learning-based electricity affordability models.
- **A Survey on Synthetic Data for AI in Energy** → Applied RMSE to compare synthetic energy consumption patterns with real data.
- **Forecasting Day-Ahead Electricity Prices** → Used RMSE to measure forecasting errors in energy price prediction models.
- **A Review of Electricity Tariffs** → Implemented RMSE to assess the accuracy of electricity tariff modeling.

6.3. R^2 Score (Coefficient of Determination)

Why was R^2 Score used?

- **To measure how well the independent variables explain the variance in the dependent variable.**
- **To evaluate the goodness-of-fit for predictive models.**

Papers Using R^2 Score and Their Justifications

- **Predictive Modeling of Energy Poverty** → Used R^2 to assess the performance of models predicting electricity affordability.
- **A Survey on Synthetic Data for AI in Energy** → Applied R^2 to compare how well synthetic data mimics actual energy usage trends.
- **Forecasting Day-Ahead Electricity Prices** → Used R^2 to determine how well predictive models capture price movements.
- **A Review of Electricity Tariffs** → Applied R^2 to validate electricity pricing models.

6.4. F1-Score, Precision, and Recall

Why were F1-Score, Precision, and Recall used?

- **To measure classification model performance, especially for imbalanced datasets.**
- **To evaluate how well models detect energy-poor households or tariff adoption behaviors.**

Papers Using F1-Score, Precision, and Recall and Their Justifications

- **Predictive Modeling of Energy Poverty** → Used F1-Score and Precision-Recall metrics to assess the classification of households into energy-poor vs. non-energy-poor categories.
- **A Survey on Synthetic Data for AI in Energy** → Applied these metrics to validate classification-based synthetic data models.

6.5. Mean Squared Error (MSE)
Why was MSE used?

- **To measure the average squared difference between predicted and actual values.**
- **To heavily penalize larger prediction errors.**

Papers Using MSE and Their Justifications

- **Predictive Modeling of Energy Poverty** → Used MSE to evaluate predictive accuracy in energy affordability modeling.
- **A Survey on Synthetic Data for AI in Energy** → Applied MSE to quantify prediction errors in synthetic data applications.
- **Forecasting Day-Ahead Electricity Prices** → Used MSE to compare different energy price forecasting models.

6.6. Accuracy Score
Why was the Accuracy Score used?

- **To evaluate classification models by measuring the proportion of correctly predicted instances.**
- **To validate tariff adoption and electricity affordability classification models.**

Papers Using Accuracy Score and Their Justifications

- **Predictive Modeling of Energy Poverty** → Used Accuracy Score to validate household electricity access classification.
- **A Survey on Synthetic Data for AI in Energy** → Applied Accuracy Score to compare classification models trained on real vs. synthetic energy data.

6.7. Mean Absolute Percentage Error (MAPE)
Why was MAPE used?

- **To measure the percentage deviation of predicted values from actual values.**
- **To standardize errors across datasets with different scales.**

Papers Using MAPE and Their Justifications

- **Predictive Modeling of Energy Poverty** → Used MAPE to measure deviations in energy poverty classification.
- **Forecasting Day-Ahead Electricity Prices** → Applied MAPE to evaluate the percentage error in electricity price predictions.

7. Exact Data Sources Used

Paper Title	Household Surveys	Regulatory Reports	Macroeconomic Indicators (GDP, Inflation)	Electricity Consumption Records	Energy Pricing Data	Satellite Data & GIS
Predictive Modeling of Energy Poverty	Yes	Yes	Yes	Yes	Yes	-
A User-Centric View of Demand Side Mgmt	-	-	-	-	-	-
A Survey on Synthetic Data for AI in Energy	-	-	-	-	-	-
Learning from Power Sector Reform (Kenya)	Yes	Yes	Yes	-	-	Yes
Learning from Power Sector Reform (Uganda)	Yes	Yes	Yes	-	-	Yes
Forecasting Day-Ahead Electricity Prices	-	-	-	Yes	Yes	-
Electrification: When Does It Work?	Yes	-	Yes	Yes	Yes	Yes
A Review of Electricity Tariffs	Yes	Yes	Yes	Yes	Yes	-
Why Are Connection Charges So High?	Yes	Yes	Yes	Yes	Yes	-

7.1. Household Surveys

Why were Household Surveys used?

- **To gather first-hand data on electricity access, affordability, and consumption patterns.**
- **To analyze socioeconomic factors influencing household electrification and tariff adoption.**

Papers Using Household Surveys and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses household survey data to classify energy poverty levels.
- **Learning from Power Sector Reform (Kenya & Uganda)** → Analyzes how energy policy reforms impact household electricity access.
- **Electrification: When Does It Work?** → Uses survey responses to measure electricity adoption rates.

- **A Review of Electricity Tariffs** → Examines how different tariff structures impact consumers.
- **Why Are Connection Charges So High?** → Uses household surveys to assess affordability constraints.

7.2. Regulatory Reports

Why were Regulatory Reports used?

- **To analyze government policies, energy tariffs, and connection charges.**
- **To examine how pricing regulations impact energy access.**

Papers Using Regulatory Reports and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses regulatory reports to analyze energy policy.
- **Learning from Power Sector Reform (Kenya & Uganda)** → Uses policy documents to track reforms in electricity markets.
- **A Review of Electricity Tariffs** → Examines how electricity pricing regulations impact affordability.
- **Why Are Connection Charges So High?** → Uses reports to assess the cost structure of electricity connections.

7.3. Macroeconomic Indicators (GDP, Inflation)

Why were Macroeconomic Indicators used?

- **To link electricity affordability with economic conditions.**
- **To analyze how inflation and GDP influence energy pricing.**

Papers Using Macroeconomic Indicators and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses GDP and inflation to study energy affordability.
- **Learning from Power Sector Reform (Kenya & Uganda)** → Examines how macroeconomic trends influence electricity policy.
- **Electrification: When Does It Work?** → Uses economic data to assess how GDP growth affects electrification rates.
- **A Review of Electricity Tariffs** → Examines how inflation affects electricity pricing.
- **Why Are Connection Charges So High?** → Uses macroeconomic data to analyze affordability constraints.

7.4. Electricity Consumption Records

Why were Electricity Consumption Records used?

- **To track historical energy use patterns.**
- **To validate machine learning models predicting electricity demand.**

Papers Using Electricity Consumption Records and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses consumption data to predict energy poverty levels.
- **Forecasting Day-Ahead Electricity Prices** → Uses historical consumption trends to forecast prices.
- **Electrification: When Does It Work?** → Tracks electricity usage trends before and after grid expansions.
- **A Review of Electricity Tariffs** → Uses data to measure how tariff changes impact consumption.
- **Why Are Connection Charges So High?** → Uses electricity consumption records to assess affordability.

7.5. Energy Pricing Data

Why was Energy Pricing Data used?

- **To evaluate tariff structures, price elasticity, and affordability.**
- **To analyze how electricity pricing influences energy consumption and connection rates.**

Papers Using Energy Pricing Data and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses pricing data to analyze energy affordability.
- **Forecasting Day-Ahead Electricity Prices** → Uses price trends to predict future market movements.
- **Electrification: When Does It Work?** → Uses tariff data to evaluate adoption patterns.
- **A Review of Electricity Tariffs** → Studies different tariff models and their economic impacts.
- **Why Are Connection Charges So High?** → Examines how high upfront connection charges affect consumer choices.

7.6. Satellite Data & GIS

Why was Satellite Data & GIS used?

- **To analyze geographic disparities in electricity access.**
- **To correlate electrification patterns with demographic and economic conditions.**

Papers Using Satellite Data & GIS and Their Justifications

- **Learning from Power Sector Reform (Kenya & Uganda)** → Uses GIS data to track the expansion of electricity infrastructure.
- **Electrification: When Does It Work?** → Uses satellite data to study electricity access trends in rural vs. urban areas.

8. Dependent Variables (Firm Performance Indicators)

Dependent Variable	Purpose

Electricity Consumption (kWh)	Measures household and firm-level energy usage.
Electricity Expenditure ($/month)	Assesses affordability of electricity and household financial burden.
Connection Rate (%)	Evaluates success of electrification programs and grid expansion.
Tariff Affordability Index	Measures whether electricity prices are affordable relative to income.
Revenue of Power Companies	Tracks financial performance of energy utilities.
Firm Productivity (TFP)	Evaluates whether electricity access improves economic efficiency.
Electricity Price Forecast ($/MWh)	Predicts future electricity prices based on market trends.

8.1. Electricity Consumption (kWh)

Why was Electricity Consumption used?

- **To measure household and firm-level electricity usage patterns.**
- **To assess the impact of tariff structures and affordability on energy demand.**

Papers Using Electricity Consumption and Their Justifications

- **Predictive Modeling of Energy Poverty** → Measures household electricity consumption to identify energy-poor households.
- **Forecasting Day-Ahead Electricity Prices** → Uses past electricity consumption data for price prediction.
- **Electrification: When Does It Work?** → Examines whether increased electrification leads to higher consumption.
- **A Review of Electricity Tariffs** → Analyzes how different pricing schemes affect electricity demand.
- **Why Are Connection Charges So High?** → Evaluates whether high connection costs lead to reduced electricity usage.

8.2. Electricity Expenditure ($/month)

Why was Electricity Expenditure used?

- **To assess the financial burden of electricity costs on households.**
- **To measure the affordability of different tariff structures.**

Papers Using Electricity Expenditure and Their Justifications

- **Predictive Modeling of Energy Poverty** → Measures how much households spend on electricity relative to their income.
- **Electrification: When Does It Work?** → Uses expenditure data to evaluate household adoption of electricity.
- **A Review of Electricity Tariffs** → Examines how pricing policies affect household electricity bills.
- **Why Are Connection Charges So High?** → Analyzes how upfront connection fees impact long-term electricity expenditure.

8.3. Connection Rate (%)
Why was Connection Rate used?

- **To measure electrification success and energy access improvements.**
- **To analyze the impact of government policies and subsidies on new electricity connections.**

Papers Using Connection Rate and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses connection rate as an indicator of household electricity access.
- **Learning from Power Sector Reform (Kenya & Uganda)** → Measures whether policy changes led to higher grid connection rates.
- **Electrification: When Does It Work?** → Evaluates the effectiveness of rural electrification projects.
- **Why Are Connection Charges So High?** → Examines whether high upfront costs reduce new connections.

8.4. Tariff Affordability Index
Why was the Tariff Affordability Index used?

- **To measure the burden of electricity costs relative to household income.**
- **To analyze whether subsidies improve affordability for low-income consumers.**

Papers Using Tariff Affordability Index and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses affordability index to classify households at risk of energy poverty.
- **Electrification: When Does It Work?** → Evaluates whether tariff affordability influences electricity adoption.
- **A Review of Electricity Tariffs** → Examines whether cost-reflective pricing models maintain affordability.
- **Why Are Connection Charges So High?** → Measures whether high connection costs worsen affordability.

8.5. Revenue of Power Companies
Why was Revenue of Power Companies used?

- **To measure the financial impact of energy sector reforms and pricing policies.**
- **To evaluate whether privatization or subsidy reductions affect utility revenue.**

Papers Using Revenue of Power Companies and Their Justifications

- **Learning from Power Sector Reform (Kenya & Uganda)** → Assesses whether reforms led to higher revenue collection.
- **Why Are Connection Charges So High?** → Examines whether high connection fees compensate for low per-unit electricity prices.

8.6. Firm Productivity (Total Factor Productivity - TFP)

Why was Firm Productivity used?

- **To assess whether electricity access leads to higher firm-level efficiency and economic growth.**
- **To analyze whether power sector reforms improve productivity.**

Papers Using Firm Productivity and Their Justifications

- **Learning from Power Sector Reform (Kenya & Uganda)** → Uses TFP to measure the economic impact of power sector reforms on businesses.

8.7. Electricity Price Forecast ($/MWh)
Why was the Electricity Price Forecast used?

- **To predict future electricity prices using historical consumption and market trends.**
- **To analyze whether machine learning models improve price forecasting.**

Papers Using Electricity Price Forecast and Their Justifications

- **Forecasting Day-Ahead Electricity Prices** → Uses historical electricity price trends to predict future market fluctuations.

9. Independent Variables

Independent Variable	Purpose
Household Income ($/month)	Assesses affordability and energy poverty classification.
Electricity Tariff ($/kWh)	Measures the impact of pricing on consumption and accessibility.
Subsidy Level ($)	Evaluates how subsidies affect electricity adoption.
Grid Distance (km)	Determines whether proximity to infrastructure influences electrification rates.
GDP Growth (%)	Analyzes whether economic expansion improves energy accessibility.
Inflation Rate (%)	Examines how inflation affects affordability and pricing policies.
Fuel Prices ($/L)	Measures the relationship between fuel costs and electricity pricing.
Weather Conditions	Predicts seasonal variations in electricity demand and pricing.

9.1. Household Income ($/month)
Why was Household Income used?

- **To assess the relationship between income levels and electricity affordability.**
- **To determine whether energy poverty is linked to low household earnings.**

Papers Using Household Income and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses income as a key predictor of energy affordability.

- **Electrification: When Does It Work?** → Examines whether income levels influence electricity adoption.
- **A Review of Electricity Tariffs** → Analyzes how income distribution affects household responses to pricing.
- **Why Are Connection Charges So High?** → Examines whether low-income households are disproportionately impacted by high connection fees.

9.2. Electricity Tariff ($/kWh)

Why was the Electricity Tariff used?

- **To analyze how different pricing structures impact electricity consumption.**
- **To measure the effect of tariff changes on energy affordability.**

Papers Using Electricity Tariff and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses tariff rates to predict energy poverty levels.
- **Forecasting Day-Ahead Electricity Prices** → Uses electricity pricing data for machine learning-based price forecasting.
- **Electrification: When Does It Work?** → Studies whether higher tariffs reduce adoption rates.
- **A Review of Electricity Tariffs** → Compares different pricing models.
- **Why Are Connection Charges So High?** → Analyzes whether high tariffs discourage grid connections.

9.3. Subsidy Level ($)

Why was Subsidy Level used?

- **To assess the effectiveness of government subsidies in improving electricity affordability.**
- **To analyze whether subsidy reductions impact household electricity consumption.**

Papers Using Subsidy Level and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses subsidy data to predict affordability for low-income households.
- **Electrification: When Does It Work?** → Evaluates whether subsidies improve grid adoption rates.
- **A Review of Electricity Tariffs** → Examines whether cost-reflective pricing models maintain affordability.
- **Why Are Connection Charges So High?** → Assesses whether subsidies offset high upfront connection fees.

9.4. Grid Distance (km)

Why was Grid Distance used?

- **To analyze how proximity to electricity infrastructure affects connection rates.**
- **To evaluate the financial feasibility of extending electricity grids to rural areas.**

Papers Using Grid Distance and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses grid distance as a factor affecting electricity access.
- **Learning from Power Sector Reform (Kenya & Uganda)** → Examines how electrification reforms improved grid expansion.
- **Electrification: When Does It Work?** → Uses GIS data to measure grid proximity impact.
- **Why Are Connection Charges So High?** → Studies whether longer grid distances lead to higher connection costs.

9.5. GDP Growth (%)

Why was GDP Growth used?

- **To measure the relationship between economic growth and electricity consumption.**
- **To determine whether national economic trends influence energy pricing policies.**

Papers Using GDP Growth and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses GDP growth to assess its impact on energy affordability.
- **Learning from Power Sector Reform (Kenya & Uganda)** → Examines whether economic growth boosted electricity investments.
- **Electrification: When Does It Work?** → Assesses whether economic growth leads to higher electrification rates.
- **A Review of Electricity Tariffs** → Studies whether GDP influences electricity pricing models.
- **Why Are Connection Charges So High?** → Examines whether economic expansion reduces connection costs.

9.6. Inflation Rate (%)

Why was Inflation Rate used?

- **To assess how rising inflation affects electricity affordability and pricing.**
- **To analyze the impact of economic conditions on tariff adjustments.**

Papers Using Inflation Rate and Their Justifications

- **Predictive Modeling of Energy Poverty** → Uses inflation as a factor influencing electricity affordability.
- **Learning from Power Sector Reform (Kenya & Uganda)** → Analyzes whether inflation eroded policy benefits.
- **Electrification: When Does It Work?** → Examines whether inflation affects rural electrification rates.
- **A Review of Electricity Tariffs** → Studies the relationship between inflation and electricity pricing.
- **Why Are Connection Charges So High?** → Uses inflation data to assess whether rising costs impact new connections.

9.7. Fuel Prices ($/L)

Why were Fuel Prices used?

- **To analyze the relationship between fossil fuel costs and electricity prices.**
- **To determine whether high fuel costs drive consumers toward renewable energy sources.**

Papers Using Fuel Prices and Their Justifications

- **Learning from Power Sector Reform (Kenya & Uganda)** → Uses fuel price fluctuations to assess policy effectiveness.
- **Forecasting Day-Ahead Electricity Prices** → Incorporates fuel costs into machine learning-based energy price forecasting.

9.8. Weather Conditions

Why were Weather Conditions used?

- **To assess the impact of temperature, rainfall, and seasonality on electricity demand.**
- **To improve machine learning-based electricity price forecasting.**

Papers Using Weather Conditions and Their Justifications

- **Forecasting Day-Ahead Electricity Prices** → Uses temperature and seasonal variations to enhance predictive accuracy.

10. Imputation Methods for Missing Data

- **Mean/Median Imputation**
  - Fills missing numerical data while preserving distribution.
- **Mode Imputation**
  - Replaces missing categorical data with the most common value.
- **K-Nearest Neighbors (KNN) Imputation**
  - Predicts missing numerical values based on similar data points.
- **Multiple Imputation by Chained Equations (MICE)**
  - Uses iterative regression models to fill in missing values.
- **Forward/Backward Filling**
  - Maintains continuity in time-series datasets by propagating previous/next values.
- **Interpolation**
  - Estimates missing values based on existing trends in time-series data.

11.Policy Recommendations and Future Research Directions

11.1. Predictive Modeling of Energy Poverty

- **Policy Recommendation:** Improve targeted subsidies for energy-poor households.
- **Future Research Direction:** Develop machine learning models for dynamic energy poverty assessment.

11.2. A User-Centric View of Demand Side Management

- **Policy Recommendation:** Enhance demand-side management incentives.

- **Future Research Direction:** Investigate behavioral responses to demand-side management programs.

11.3. A Survey on Synthetic Data for AI in Energy

- **Policy Recommendation:** Develop regulatory frameworks for synthetic energy data use.
- **Future Research Direction:** Expand synthetic data applications for energy market simulations.

11.4. Learning from Power Sector Reform (Kenya)

- **Policy Recommendation:** Ensure policy stability to support private investment in the power sector.
- **Future Research Direction:** Examine long-term effects of power sector liberalization.

11.5. Learning from Power Sector Reform (Uganda)

- **Policy Recommendation:** Strengthen governance and transparency in power sector reforms.
- **Future Research Direction:** Analyze the role of decentralized energy systems in power sector resilience.

11.6. Forecasting Day-Ahead Electricity Prices

- **Policy Recommendation:** Adopt AI-driven forecasting for real-time electricity pricing.
- **Future Research Direction:** Improve AI-based forecasting accuracy using real-time market signals.

11.7. Electrification: When Does It Work?

- **Policy Recommendation:** Prioritize rural electrification with cost-effective strategies.
- **Future Research Direction:** Explore productive electricity use in rural electrification projects.

11.8. A Review of Electricity Tariffs

- **Policy Recommendation:** Implement fair and cost-reflective electricity pricing mechanisms.
- **Future Research Direction:** Study socioeconomic impacts of electricity tariff structures.

11.9. Why Are Connection Charges So High?

- **Policy Recommendation:** Reduce upfront connection charges to increase household grid adoption.
- **Future Research Direction:** Assess microfinance models for financing electricity connections.

12. Conclusion

Reliable and affordable electricity access remains a major challenge, especially in Sub-Saharan Africa, where tariff structures and subsidies impact affordability and equity. While AI and ML offer potential for energy pricing and policy optimization, their integration into regulation is still developing. This review examines key economic methodologies like welfare analysis and difference-in-differences to assess tariff reforms. By consolidating data from regulatory reports, surveys, and macroeconomic indicators, we highlight research gaps and the need for AI-driven policy analysis.

Future research should focus on ML-based tariff modeling, improved forecasting, and policy frameworks that balance financial sustainability with social equity to ensure long-term energy affordability.

References

Godinho, C., & Eberhard, A. (2019). *Learning from power sector reform experiences: The case of Uganda* (Policy Research Working Paper No. 8820). Energy and Extractives Global Practice, World Bank.

Godinho, C., & Eberhard, A. (2019). *Learning from power sector reform experiences: The case of Kenya* (Policy Research Working Paper No. 8819). Energy and Extractives Global Practice, World Bank.

Meiser, M., & Zinnikus, I. (2024). *A survey on the use of synthetic data for enhancing key aspects of trustworthy AI in the energy domain: Challenges and opportunities.* Energies, 17(9), 1992.

Gawusu, S., Jamatutu, S. A., & Ahmed, A. (2024). *Predictive modeling of energy poverty with machine learning ensembles: Strategic insights from socioeconomic determinants for effective policy implementation.* International Journal of Energy Research.

Figueiredo Walter, T., & Moneke, N. (2024). *When does electrification work? Evidence from Sub-Saharan Africa.*

Lehna, M., Scheller, F., & Herwartz, H. (2022). *Forecasting day-ahead electricity prices: A comparison of time series and neural network models taking external regressors into account.* Energy Economics, 106, 105742.

Blimpo, M., McRae, S., & Steinbuks, J. (2018). *Why are connection charges so high? An analysis of the electricity sector in Sub-Saharan Africa* (Policy Research Working Paper No. 8407). Office of the Chief Economist, Africa Region, World Bank

Zaki, D. A., & Hamdy, M. (2022). *A review of electricity tariffs and enabling solutions for optimal energy management.*Energies, 15(22), 8527.

De Vizia, C., Patti, E., Macii, E., & Bottaccioli, L. (2022). *A user-centric view of a demand side management program: From surveys to simulation and analysis.* IEEE Systems Journal, 16(2), 1886–1897.