Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predictive Modeling of Urban Residential Power Consumption: An Explainable AI Approach to the Kenyan Prepaid Grid

Paul Obasanjo

Paul.obasanjo2@gmail.com

Technical University of Kenya

Annette Mutinda

Technical University of Kenya

Adero Awuor

Technical University of Kenya

Research Article

Keywords: Residential electricity demand forecasting, prepaid electricity systems, explainable artificial intelligence (XAI), XGBoost, SHAP, autoregressive modelling, load forecasting, energy analytics, Kenya

Posted Date: July 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10335415/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Predictive Modeling of Urban Residential Power Consumption: An Explainable AI Approach to the Kenyan Prepaid Grid

Paul Obasanjo[1,*], Annette Mutinda[1], Adero Awuor[1]

[1]Technical University of Kenya,

Department of Physics, Earth and Environmental Sciences

Nairobi, Kenya

[*]Corresponding Author

Email: paul.obasanjo2@gmail.com

Abstract

Accurate electricity demand forecasting is essential for reliable grid operation, revenue planning, and energy management, particularly within rapidly urbanizing regions where prepaid metering has become the dominant electricity billing system. However, many contemporary load forecasting models are developed under assumptions of strong weather sensitivity and price-responsive consumer behaviour that may not fully reflect the realities of residential electricity consumption in sub-Saharan Africa. This study develops an Explainable Artificial Intelligence (XAI) framework to investigate the drivers of residential electricity demand within the Kenyan prepaid electricity sector and to evaluate the forecasting performance of machine-learning and statistical approaches.A longitudinal dataset spanning January 2017 to April 2026 was analysed using historical consumption records, meteorological variables, and tariff-related information. An Extreme Gradient Boosting (XGBoost) model was benchmarked against an Ordinary Least Squares (OLS) regression model using a strict chronological validation strategy. Model interpretability was achieved through SHapley Additive exPlanations (SHAP), while additional

statistical diagnostics, including autocorrelation analysis, seasonal decomposition, and ablation testing, were employed to examine the temporal structure of electricity demand.The results indicate that historical consumption behaviour is the dominant determinant of future residential electricity demand. Lag-based autoregressive variables consistently exhibited the highest predictive influence, whereas weather and tariff-related variables contributed comparatively limited additional forecasting information. Despite its greater complexity, XGBoost achieved forecasting performance broadly comparable to that of the OLS model, suggesting that much of the predictive information within the dataset can be captured using relatively simple autoregressive relationships. SHAP analysis further revealed that short-term consumption persistence exerts substantially greater influence on future demand than meteorological variability or tariff adjustments.These findings demonstrate that residential electricity demand within the Kenyan prepaid sector is governed primarily by behavioural persistence and recurring consumption patterns. The study highlights the value of interpretable forecasting frameworks and suggests that computationally efficient, transparent forecasting models can provide reliable demand estimates for utility planning and operational decision-making in prepaid electricity environments.

Keywords: Residential electricity demand forecasting; prepaid electricity systems; explainable artificial intelligence (XAI); XGBoost; SHAP; autoregressive modelling; load forecasting; energy analytics; Kenya.

1 Introduction

The rapid urbanization of developing economies has intensified pressure on electrical power infrastructure, making accurate electricity load forecasting an increasingly important component of modern utility planning and grid management. Reliable forecasts support a wide range of operational activities, including generation scheduling, demand-side management, infrastructure expansion, tariff design, and long-term energy policy formulation. As urban populations continue to grow and electricity demand becomes increasingly dynamic, utility providers require forecasting frameworks capable of delivering both predictive accuracy and operational insight. Therefore, electricity demand forecasting has evolved into a critical research area at the intersection of energy systems, economics, and data science.

Within sub-Saharan Africa, the forecasting challenge is further complicated by the widespread adoption of prepaid electricity metering. Unlike conventional postpaid billing systems, prepaid architectures fundamentally alter the relationship between consumers and electricity consumption by requiring users to purchase electricity credits in advance. This structure introduces distinct behavioral, economic, and social dynamics that directly influence demand patterns. Utility providers operating within prepaid environments must therefore balance multiple objectives simultaneously, including maintaining grid stability, ensuring financial sustainability, improving energy accessibility, and expanding electrification coverage [7].

Recent studies across the African continent have demonstrated that prepaid metering substantially modifies household electricity consumption behavior relative to traditional billing systems. Evidence from Ethiopia [11] and Ghana [10] indicates that prepaid technologies influence both consumption patterns and utility revenue collection mechanisms. In Nairobi, prepaid electricity systems have emerged as a central instrument for urban grid expansion and service delivery, effectively becoming a new paradigm through which electricity access is managed and regulated [16]. However, while prepaid systems have improved billing efficiency and reduced revenue losses, they have also introduced complex socioeconomic interactions between utilities and consumers.

These dynamics are particularly evident among low-income urban households. Across several East African settings, vulnerable consumers often report a stronger concern about accumulating debt than about temporary loss of electrical service, leading them to tightly

regulate and ration their electricity usage according to the available prepaid credit balance [22]. Consequently, observed electricity demand may reflect not only physical energy requirements but also economic constraints and behavioral adaptation strategies. Furthermore, researchers have noted that the mere existence of a grid connection does not necessarily imply adequate energy access. Within many urban prepaid communities, significant levels of hidden energy poverty persist, with households deliberately suppressing consumption in order to manage electricity expenditures [15]. These realities suggest that electricity demand in prepaid systems may be governed by fundamentally different drivers than those commonly observed in conventional electricity markets.

Simultaneously, advances in computational intelligence have transformed the field of electricity demand forecasting. Recent literature increasingly promotes machine learning and artificial intelligence techniques as state-of-the-art solutions for capturing complex consumption dynamics and improving predictive performance [2]. Algorithms such as Random Forests, Support Vector Regression, Artificial Neural Networks, and Gradient Boosting frameworks have demonstrated substantial success in modeling non-linear relationships between electricity demand and exogenous variables such as temperature, humidity, occupancy patterns, and economic activity [18]. Within residential energy forecasting specifically, tree-based ensemble methods have emerged as particularly powerful predictive tools due to their ability to learn intricate interactions among multiple explanatory variables [20, 21].

Among these approaches, Extreme Gradient Boosting (XGBoost) has become one of the most widely adopted algorithms for structured load forecasting applications. Its ability to model non-linear dependencies, capture temporal interactions, handle heterogeneous feature spaces, and maintain strong predictive performance across diverse forecasting environments has established it as a benchmark methodology in contemporary energy analytics [14]. Consequently, many recent studies implicitly assume that increasingly sophisticated machine learning architectures will consistently outperform simpler statistical approaches.

However, a critical limitation exists within much of the current forecasting literature. The overwhelming majority of machine learning models have been developed and validated using datasets originating from North America, Europe, and other regions characterized by temperate climates, widespread HVAC utilization, and relatively high

household income levels [1]. As a result, many forecasting frameworks are built upon assumptions that may not hold universally across different socioeconomic and climatic environments. Two assumptions are particularly prevalent. First, residential electricity demand is often assumed to be highly sensitive to environmental temperature fluctuations because heating and cooling systems constitute a substantial fraction of household electricity consumption. Second, consumers are frequently assumed to exhibit measurable short-term price responsiveness, resulting in observable demand adjustments following tariff changes.

Whether these assumptions remain valid within urban prepaid electricity systems in developing economies remains insufficiently explored. In many Kenyan households, electricity consumption is dominated by essential baseload activities such as lighting, refrigeration, communication devices, and basic household appliances rather than energy-intensive climate control systems. Consequently, the relative influence of meteorological variability on electricity demand may differ substantially from that observed in Western forecasting studies. Similarly, households operating under prepaid constraints may exhibit consumption patterns that are driven more strongly by habitual usage requirements and credit availability than by immediate responses to tariff restructuring. These distinctions raise important questions regarding the actual predictive drivers of electricity demand within prepaid environments and whether highly complex machine learning architectures provide meaningful advantages over simpler forecasting approaches.

An additional challenge concerns model interpretability. Although advanced machine learning algorithms often achieve impressive predictive performance, they frequently operate as opaque "black-box" systems that provide limited insight into the mechanisms driving their predictions [3]. For utility planners, regulators, and policymakers, predictive accuracy alone is rarely sufficient. Effective decision-making requires an understanding of why a model generates a particular forecast, which variables exert the greatest influence on demand, and whether those influences align with physical, economic, and social realities. Without such transparency, stakeholders may struggle to diagnose abnormal predictions, evaluate policy interventions, or translate model outputs into actionable operational strategies [17].

These challenges reveal a methodological gap within the existing literature. While substantial effort has been devoted to improving forecasting accuracy through increasingly

sophisticated machine learning architectures, comparatively little attention has been directed toward understanding whether the underlying assumptions of these models remain valid within prepaid electricity systems in developing economies. Furthermore, limited research has investigated the relative importance of historical consumption behavior, environmental variability, and tariff structures within the Kenyan residential prepaid context using explainable analytical frameworks.

To address this gap, the present study develops an Explainable Artificial Intelligence (XAI)-based framework for forecasting residential electricity demand within the Kenyan prepaid grid. Recent systematic reviews emphasize that explainability is becoming an essential requirement for the practical deployment of AI systems in energy management, enabling greater transparency, stakeholder trust, and operational accountability [12, 13]. This study therefore combines predictive modeling with interpretable machine learning to investigate not only how accurately residential demand can be forecasted, but also why particular forecasting outcomes emerge.

Specifically, the study benchmarks an XGBoost architecture against a traditional Ordinary Least Squares (OLS) regression framework to evaluate whether the predictive complexity offered by advanced machine learning translates into meaningful forecasting improvements within a prepaid residential environment. Furthermore, by integrating SHapley Additive exPlanations (SHAP), a game-theoretic interpretability methodology capable of quantifying feature contributions at both local and global scales, the study systematically examines the relative influence of historical consumption behavior, meteorological variability, and gazetted tariff restructuring by the Energy and Petroleum Regulatory Authority (EPRA) on residential electricity demand.

The central contribution of this research is therefore twofold. First, it provides one of the few comprehensive examinations of residential electricity forecasting within the Kenyan prepaid grid using both advanced machine learning and explainable AI methodologies. Second, it evaluates whether the dominant forecasting assumptions derived from Western electricity markets remain applicable within a prepaid urban African context. By identifying the true drivers of residential electricity demand and quantifying their relative importance, this study contributes both methodological and policy-relevant insights for the design of transparent, scalable, and operationally effective forecasting systems in emerging electricity markets.

2 Methodology

2.1 Data Acquisition and Scope

This study utilized a longitudinal dataset of residential prepaid electricity consumption spanning the period from January 2017 to April 2026. The extended temporal coverage was selected to capture long-term consumption behaviour, seasonal variability, regulatory tariff adjustments, and evolving household electricity usage patterns within the Kenyan prepaid electricity sector.

To account for factors that may influence electricity demand beyond historical consumption, the primary consumption dataset was supplemented with both tariff and meteorological information. Historical tariff data were obtained from official gazette notices published by the Energy and Petroleum Regulatory Authority (EPRA). The extracted tariff variables included the Lifeline Rate, Ordinary Rate, and Lifeline Limit thresholds applicable during different billing periods. These tariff parameters were mapped to the corresponding consumption records according to their effective implementation dates.

Meteorological data were obtained from the Open-Meteo Historical Climate API. Monthly average temperature values were extracted and used to compute Cooling Degree Days (CDD), a widely used indicator of cumulative thermal conditions. CDD values were calculated using a baseline temperature of 18 °C and were incorporated into the forecasting framework to evaluate the influence of temperature-related variability on residential electricity consumption.

The resulting dataset integrated electricity consumption records, tariff information, and meteorological observations into a unified analytical framework. This approach enabled a systematic assessment of the relative importance of historical consumption behaviour, environmental variability, and tariff-related factors in shaping residential electricity demand.

2.2 Feature Engineering and Temporal Dynamics

Residential electricity consumption exhibits strong temporal dependence, whereby current demand is often influenced by previous consumption behaviour and recurring seasonal patterns. To capture these dynamics, a series of autoregressive features were engineered from the original time series.

Three lag variables were generated for each household meter: `Lag_1_kWh`, `Lag_2_kWh`, and `Lag_12_kWh`. The first- and second-order lag variables capture short-term persistence in household electricity consumption, while the twelfth-order lag captures annual seasonal recurrence by incorporating information from the corresponding month of the previous year. These autoregressive features were selected based on established time-series forecasting principles and were subsequently validated using autocorrelation and partial autocorrelation diagnostics.

In addition to the lag variables, meteorological and tariff-related predictors were incorporated into the modelling framework. Weather-related variables included monthly average temperature and Cooling Degree Days (CDD), while tariff-related variables included the Lifeline Rate and Lifeline Limit applicable during each billing period. Together, these predictors enabled the forecasting models to evaluate the relative contributions of historical consumption behaviour, environmental conditions, and electricity pricing structures.

Prior to model development, all engineered features were examined for completeness and consistency. Observations containing insufficient historical information for lag generation were excluded to ensure that all models were trained using complete predictor sets.

2.3 Predictive Modeling Framework

A comparative forecasting framework was adopted to evaluate the suitability of both traditional statistical approaches and modern machine-learning techniques for residential electricity demand prediction.

The primary machine-learning model employed in this study was Extreme Gradient Boosting (XGBoost), a gradient-boosted decision tree algorithm recognized for its ability to model non-linear relationships and complex interactions among predictor variables [4]. Owing to its predictive accuracy, computational efficiency, and robustness, XGBoost has become one of the most widely used algorithms in electricity load forecasting applications [9, 14].

To improve model generalization and reduce the risk of overfitting, hyperparameter optimization was performed using a grid-search procedure. Multiple combinations of model parameters were evaluated, and the configuration producing the best predictive performance was selected for subsequent analysis.

To provide an interpretable benchmark against which the machine-learning model could be evaluated, an Ordinary Least Squares (OLS) regression model was also developed. OLS was selected because of its transparency, computational simplicity, and extensive use in forecasting studies. Comparing OLS and XGBoost enabled an assessment of whether the additional complexity of machine-learning approaches yields meaningful forecasting improvements within the Kenyan prepaid electricity context.

2.4 Time-Series Validation Strategy

Electricity consumption data are inherently sequential, and preserving temporal order during model evaluation is essential to avoid information leakage. Randomly partitioning time-series data can allow future observations to influence model training, resulting in overly optimistic performance estimates that do not reflect real-world forecasting conditions.

To avoid this issue, a strict chronological validation strategy was implemented. Observations from January 2017 to December 2024 were used for model training and hyperparameter optimization, while observations from January 2025 to April 2026 were reserved exclusively for out-of-sample testing.

This approach replicates practical forecasting conditions in which future observations are unavailable during model development. By maintaining temporal integrity throughout the modelling process, the resulting performance metrics provide a realistic assessment of forecasting capability while eliminating look-ahead bias.

Following feature engineering and data preparation, the final dataset consisted of 97,354 training observations and 24,350 testing observations. Model performance was evaluated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). MAE measures the average magnitude of forecasting errors, while RMSE places greater emphasis on larger deviations and therefore provides additional insight into model behaviour during periods of unusual electricity consumption.

2.5 Model Diagnostics and Statistical Evaluation

To assess the statistical validity of the forecasting framework, several diagnostic procedures were conducted.

First, multicollinearity among predictor variables was evaluated using the Variance Inflation Factor (VIF). This analysis was performed to ensure that excessive correlations among predictors did not compromise model stability or interpretation. VIF values below commonly accepted thresholds were considered indicative of acceptable predictor independence.

Second, autocorrelation and partial autocorrelation analyses were performed to examine the temporal structure of residential electricity consumption and to validate the selection of autoregressive lag variables. These diagnostics provided statistical evidence regarding the persistence and seasonal characteristics of the consumption series.

Finally, seasonal decomposition was applied to separate the time series into trend, seasonal, and residual components. This procedure enabled a clearer understanding of the relative importance of long-term trends, recurring seasonal behaviour, and unexplained variability within the residential electricity consumption record.

2.6 Explainable Artificial Intelligence (XAI)

To improve model transparency and identify the primary drivers of residential electricity demand, Explainable Artificial Intelligence (XAI) techniques were incorporated into the analytical framework. Specifically, SHapley Additive exPlanations (SHAP) were employed to quantify the contribution of individual predictors to model forecasts.

SHAP is based on cooperative game theory and assigns each feature a contribution value according to its marginal influence on the final model output [5]. For the XGBoost model, the SHAP TreeExplainer was used to compute feature-attribution values and quantify the relative importance of each predictor.

The SHAP framework was applied to evaluate the influence of autoregressive consumption variables, meteorological indicators, and tariff-related parameters on forecasting performance. By decomposing model predictions into additive feature contributions, SHAP enabled a transparent interpretation of the factors driving residential electricity demand. Consequently, the forecasting framework not only generated accurate predictions but also provided interpretable insights into the behavioural, environmental, and regulatory influences shaping electricity consumption within the Kenyan prepaid sector.

3 Results and Discussion

3.1 Exploratory Analysis of Residential Electricity Consumption

Understanding the statistical structure of the dataset is an important first step in interpreting the forecasting results. Figure 1 presents the correlation matrix of all variables included in the modelling framework.

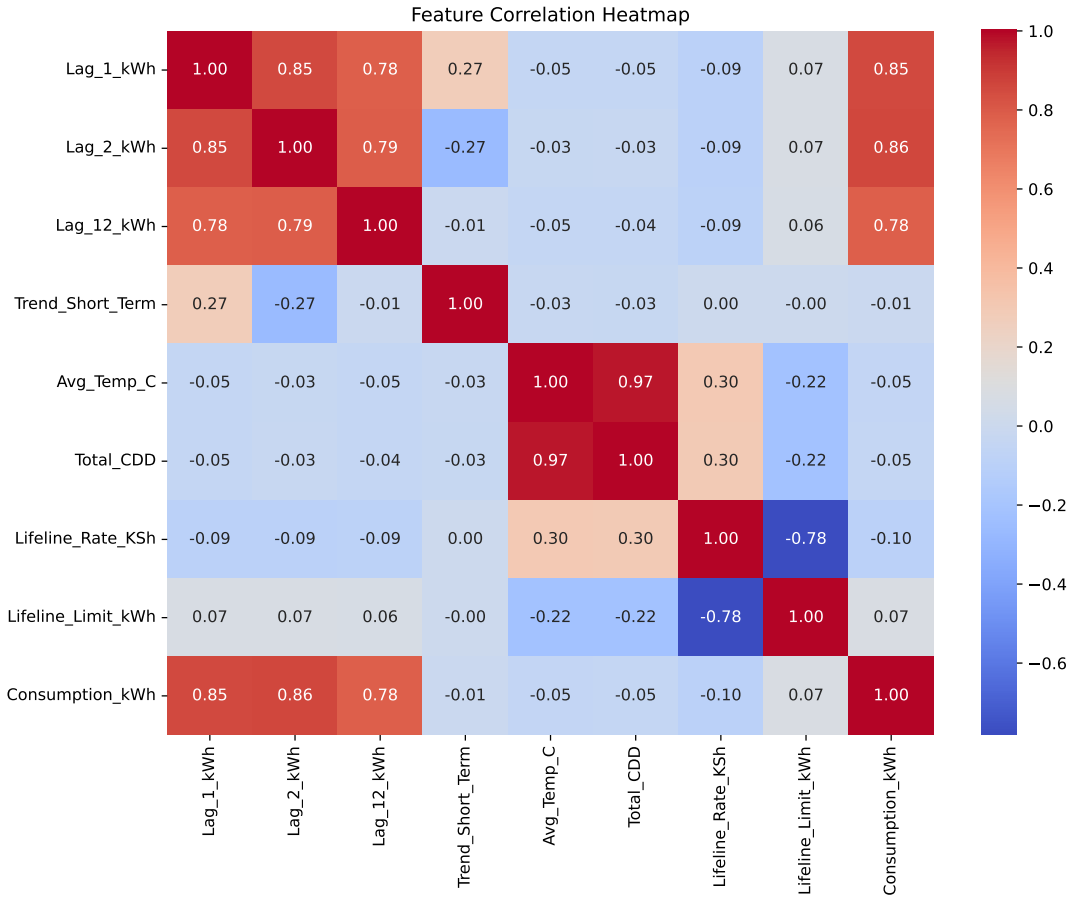


Figure 1: Correlation matrix of consumption, meteorological, tariff, and autoregressive variables.

The correlation analysis reveals that historical electricity consumption variables exhibit the strongest relationships with current demand. In particular, the first-, second-, and twelfth-order lag variables display a stronger associations with present consumption than either meteorological or tariff-related predictors. This pattern provides preliminary evidence that residential electricity demand possesses a strong autoregressive structure,

whereby future consumption is largely conditioned by past household behaviour.

By contrast, weather-related and tariff-related variables exhibit comparatively weaker correlations with electricity consumption. Although Total Cooling Degree Days (CDD) displays a measurable association with demand, its relationship remains considerably smaller than those observed for the lag variables. Similarly, tariff indicators show only modest correlations with electricity use. These findings suggest that historical consumption behaviour is likely to be the dominant source of predictive information within the dataset.

3.2 Multicollinearity Assessment

Prior to model interpretation, predictor independence was evaluated using the Variance Inflation Factor (VIF). Excessive multicollinearity can destabilize regression coefficients and reduce model interpretability. The VIF diagnostics for the final forecasting framework are presented in Table 1.

Table 1: Variance Inflation Factor (VIF) Diagnostics

Feature	**VIF**
Lag_1_kWh	7.92
Lag_2_kWh	8.12
Lag_12_kWh	5.62
Total_CDD	2.70
Lifeline_Rate_KSh	2.14
Lifeline_Limit_kWh	2.20

All predictors exhibit VIF values below the commonly accepted threshold of 10, indicating the absence of severe multicollinearity. The highest values occur among the lag variables, which is expected given the temporal persistence inherent in electricity consumption data. Nevertheless, the observed values remain within acceptable limits, preserving the stability and interpretability of the OLS framework. The relatively low VIF values associated with weather and tariff variables further indicate that these predictors contribute distinct information to the forecasting process.

3.3 Comparative Forecasting Performance

The predictive capability of the proposed XGBoost framework was evaluated against a conventional Ordinary Least Squares (OLS) regression model using a strict chronological training-testing strategy. Data spanning 2017–2024 were used for model training, while observations from 2025–2026 were reserved exclusively for out-of-sample evaluation.

Table 2 summarizes the forecasting performance of both models.

Table 2: Forecasting Performance Comparison

Model	**MAE (kWh)**	**RMSE (kWh)**
OLS Regression	20.64	38.31
XGBoost	20.43	39.28

The two forecasting frameworks produced remarkably similar results. XGBoost achieved a marginally lower Mean Absolute Error (MAE), whereas OLS yielded a slightly lower Root Mean Square Error (RMSE). Given the scale and variability of the dataset, these differences are small and suggest that both approaches capture the underlying consumption dynamics with comparable effectiveness.

This outcome indicates that much of the predictive information contained within the dataset can be captured using relatively simple autoregressive relationships. Although XGBoost offers greater flexibility and the ability to model complex non-linear interactions, these capabilities provide only limited additional forecasting benefit within the prepaid residential electricity context.

Figure 2 compares observed and predicted electricity consumption during the evaluation period.

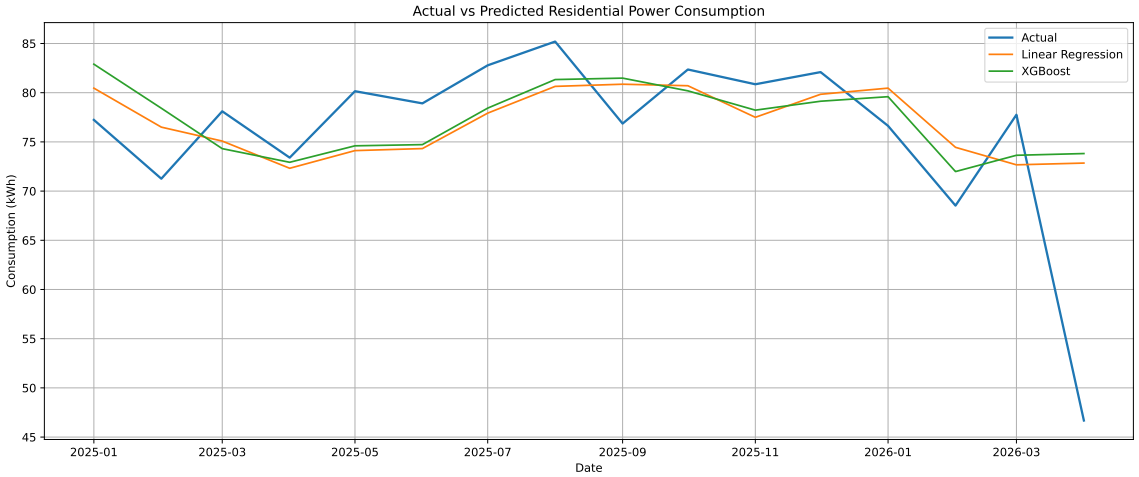


Figure 2: Observed and predicted residential electricity consumption during the independent testing period.

Both models reproduce the dominant temporal evolution of residential electricity demand during the testing period and exhibit highly similar forecasting trajectories. Although a small number of abrupt deviations remain under-predicted, the overall agreement between observations and forecasts indicates that most of the predictive information can be extracted from historical consumption behaviour. The limited performance advantage achieved by XGBoost suggests that complex non-linear interactions contribute comparatively little additional forecasting skill beyond that already provided by autoregressive consumption patterns.

3.4 Residual Behaviour and Forecast Reliability

To evaluate model reliability, the distribution of forecasting residuals was examined. Figure 3 presents the residual distribution obtained from the forecasting framework.

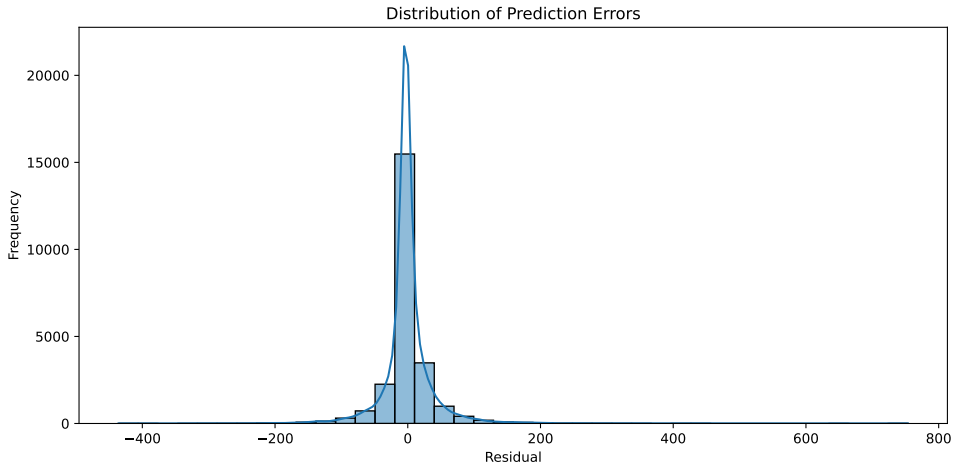


Figure 3: Distribution of forecasting residuals.

The residuals remain centred around zero and show no evidence of systematic bias toward either overprediction or underprediction. This indicates that the forecasting framework produces balanced estimates across the evaluation period. While the distribution exhibits heavier tails than an ideal Gaussian distribution, such behaviour is expected in large residential datasets where occasional high-consumption households introduce extreme observations. Overall, the residual structure suggests that the models provide stable and reliable forecasts.

3.5 Explainable AI and the Drivers of Electricity Demand

A major objective of this study was not only to forecast electricity demand but also to identify the factors driving model predictions. To achieve this, SHAP (SHapley Additive exPlanations) analysis was performed on the trained XGBoost model.

Figure 4 presents the SHAP summary plot.

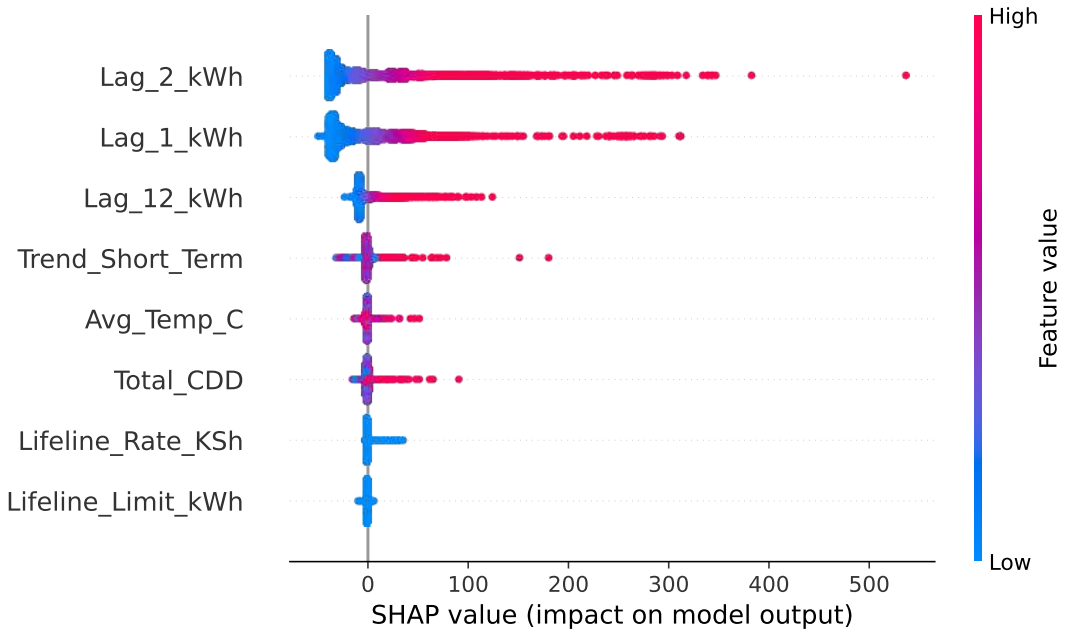


Figure 4: SHAP summary plot showing the contribution of each predictor to model forecasts.

The SHAP analysis reveals a clear hierarchy of predictor importance. Lag_2_kWh emerges as the most influential predictor, followed by Lag_1_kWh and Lag_12_kWh. Together, these autoregressive variables account for the overwhelming majority of model explanatory power and consistently generate the largest SHAP magnitudes across the dataset. High values of these predictors are generally associated with positive SHAP contributions, indicating that households with higher recent electricity consumption are more likely to exhibit elevated future demand. This finding highlights the strong persistence inherent in residential electricity consumption and suggests that historical usage patterns contain most of the predictive information required for accurate load forecasting.

In contrast, weather and tariff-related variables contribute substantially smaller SHAP values. Although Total Cooling Degree Days (CDD) is the most influential exogenous predictor, its contribution remains modest relative to the lag variables. Likewise, the Lifeline Rate and Lifeline Limit exhibit comparatively weak influence on forecast outcomes.

These findings suggest that household electricity consumption within the prepaid residential sector is governed primarily by behavioural persistence. Future demand is strongly conditioned by recent consumption patterns, indicating that households tend to maintain relatively stable electricity usage habits over time. The contribution of Lag_12_kWh further points to the existence of recurring annual consumption behaviour,

reflecting the influence of seasonal and cyclical household activities. Nevertheless, these longer-term effects remain secondary to the dominant influence of short-term consumption history, highlighting the importance of recent usage patterns in shaping future electricity demand.

3.6 Incremental Contribution of Weather and Tariff Variables

To quantify the contribution of different predictor groups, an ablation analysis was conducted. The results are presented in Figure 5.

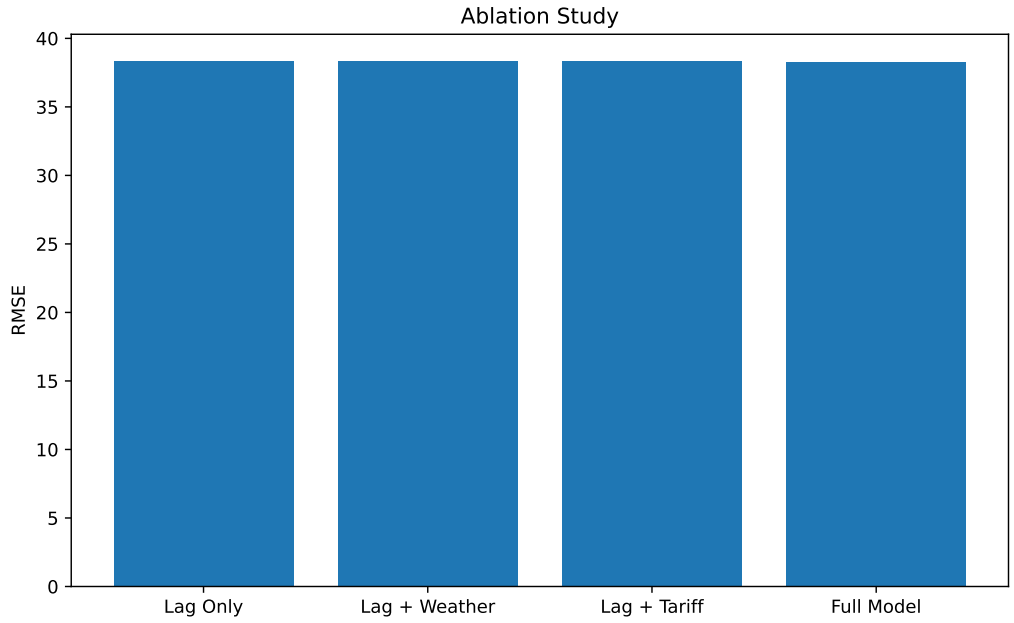


Figure 5: Ablation study illustrating the contribution of different predictor groups to forecasting performance.

The ablation analysis demonstrates that historical consumption variables provide nearly all of the forecasting skill achieved by the model. The inclusion of meteorological information and tariff-related predictors results in only marginal improvements in forecasting accuracy, indicating that these variables contribute comparatively little additional predictive information beyond that already contained within the autoregressive consumption history.

This finding provides further insight into the similar performance observed between OLS and XGBoost. Since the dominant predictive information originates from historical consumption behaviour, there is limited opportunity for more sophisticated machine-

learning algorithms to exploit additional non-linear relationships.

3.7 Temporal Persistence and Seasonal Behaviour

The temporal structure of residential electricity consumption was further investigated using autocorrelation diagnostics. Figure 6 presents the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

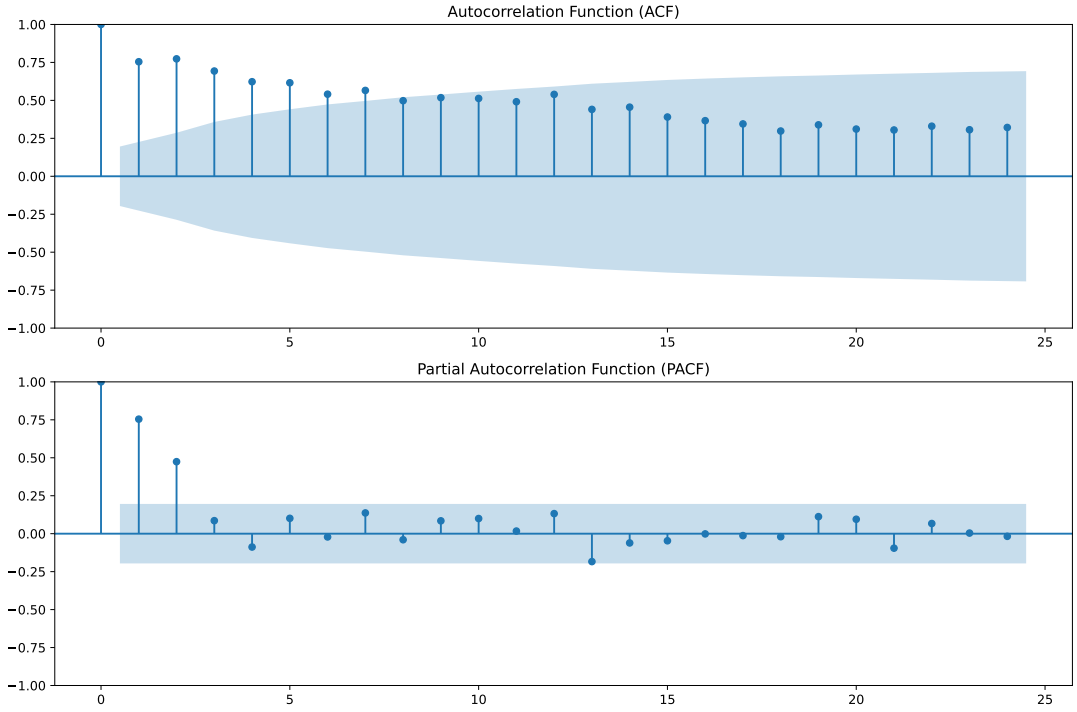


Figure 6: Autocorrelation and partial autocorrelation diagnostics for residential electricity consumption.

The ACF exhibits a gradual decay across successive lags, indicating substantial temporal persistence within the residential electricity consumption series. Significant autocorrelation remains present over multiple lag intervals, suggesting that current electricity demand is strongly influenced by previous observations. Meanwhile, the PACF displays pronounced first-order and second-order spikes, providing statistical support for the inclusion of Lag_1 and Lag_2 as autoregressive predictors within the forecasting framework. These autocorrelation patterns indicate that a considerable proportion of future consumption behaviour can be explained by recent electricity usage, thereby justifying the adoption of lag-based features within the forecasting models.

These autocorrelation diagnostics provide independent statistical justification for the autoregressive feature engineering strategy adopted in this study and confirm the presence of strong temporal dependence within residential electricity consumption.

Seasonal decomposition provides additional insight into the temporal characteristics of the series. Figure 7 separates the data into trend, seasonal, and residual components.

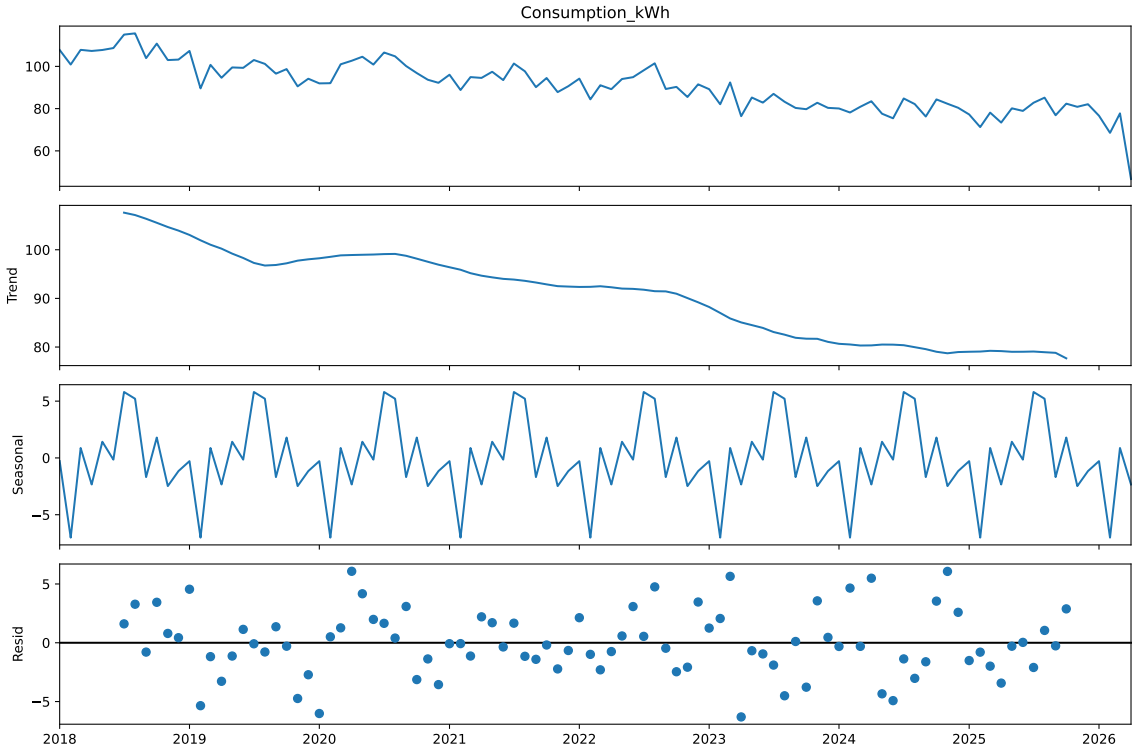


Figure 7: Seasonal decomposition of residential electricity consumption.

The seasonal decomposition reveals the presence of recurring seasonal behaviour within residential electricity consumption. However, the magnitude of the seasonal component remains considerably smaller than that of the long-term trend and the strong autoregressive structure identified throughout the analysis. This observation is consistent with the SHAP results, where Lag_12_kWh contributes meaningfully to forecasting performance by capturing annual consumption recurrence but remains less influential than the short-term predictors Lag_1_kWh and Lag_2_kWh. Consequently, while seasonal effects are evident within the prepaid residential sector, short-term consumption persistence exerts substantially greater influence on future electricity demand. These findings suggest that recent consumption behaviour provides more predictive information than recurring annual patterns, reinforcing the dominant role of behavioural persistence in

residential electricity usage.

3.8 Implications for Utility Planning and Energy Policy

The collective evidence from the forecasting, explainability, and temporal analyses points toward a consistent conclusion. Residential electricity demand within the Kenyan prepaid sector is governed primarily by behavioural persistence. Although weather variability and tariff structures contribute measurable predictive information, their influence remains substantially weaker than that of historical consumption behaviour.

From an operational perspective, these findings suggest that maintaining accurate historical consumption records may be more valuable than deploying increasingly complex forecasting architectures. The near-identical performance of OLS and XGBoost demonstrates that computationally efficient and highly interpretable statistical models can achieve forecasting accuracy comparable to more sophisticated machine-learning approaches.

From a policy perspective, the relatively weak influence of tariff variables suggests that short-term residential electricity demand remains largely inelastic within the prepaid environment. Households appear to prioritise essential electricity services and maintain relatively stable consumption patterns despite moderate tariff adjustments. Consequently, policymakers should not necessarily expect immediate reductions in residential electricity demand following tariff revisions, particularly among lower- and middle-income households.

Overall, the results demonstrate that historical consumption behaviour remains the dominant determinant of future residential electricity demand. Weather variability and tariff structures contribute useful but comparatively limited predictive information, explaining why advanced machine-learning models provide only marginal forecasting improvements over simpler autoregressive approaches.

4 Conclusion and Recommendations

This study investigated the drivers of residential electricity demand within the Kenyan prepaid electricity sector using a forecasting framework that combined machine learning and Explainable Artificial Intelligence (XAI) techniques. By integrating historical

consumption records, meteorological variables, and tariff-related information, the study sought not only to predict future electricity demand but also to identify the factors that most strongly influence consumption behaviour.

The results consistently demonstrated that historical electricity consumption is the dominant predictor of future residential demand. Correlation analysis, SHAP explainability results, autocorrelation diagnostics, and seasonal decomposition all indicated that recent consumption history contains the majority of the predictive information required for accurate load forecasting. While weather variability and tariff-related variables contributed measurable information to the forecasting framework, their influence remained substantially weaker than that of the autoregressive consumption features.

A comparative evaluation of forecasting models showed that the XGBoost and Ordinary Least Squares (OLS) frameworks achieved broadly similar levels of predictive accuracy when evaluated using a strict chronological validation strategy. Although XGBoost achieved a marginally lower Mean Absolute Error, OLS produced a slightly lower Root Mean Square Error, indicating that the additional complexity of the machine-learning model resulted in only limited forecasting gains for this particular application. These findings suggest that residential electricity demand within the prepaid sector is characterised by strong behavioural persistence, allowing relatively simple forecasting approaches to capture much of the underlying demand structure.

The explainability analysis further revealed that short-term consumption history exerts considerably greater influence on future electricity demand than seasonal, weather-related, or tariff-related factors. The contribution of annual lag effects indicates the presence of recurring seasonal behaviour; however, these effects remain secondary to the dominant influence of recent consumption patterns. Collectively, the results suggest that household electricity usage within the prepaid environment is shaped primarily by established consumption habits and recurring behavioural routines.

From a practical perspective, the findings indicate that utility providers can achieve reliable residential load forecasts using computationally efficient and highly interpretable forecasting frameworks. Maintaining accurate historical consumption records may therefore provide greater forecasting value than adopting increasingly complex modelling architectures, particularly in environments where demand behaviour exhibits strong temporal persistence.

From a policy perspective, the comparatively weak influence of tariff variables suggests that short-term residential electricity demand may not respond strongly to moderate tariff adjustments. While tariff changes remain important for revenue planning and long-term energy policy, their immediate impact on household electricity consumption appears limited relative to the influence of established consumption behaviour. Consequently, policymakers should exercise caution when assuming that short-term changes in electricity prices will necessarily translate into substantial reductions in residential demand.

Future research could extend this work by incorporating higher-frequency smart-meter data, household demographic characteristics, appliance ownership information, and broader socioeconomic indicators. Additional investigations into anomaly detection and behavioural change modelling may also help identify abrupt consumption shifts that are not fully captured by conventional autoregressive forecasting approaches. Such developments would provide a deeper understanding of residential electricity demand dynamics and support more adaptive forecasting strategies for rapidly evolving prepaid electricity systems.

Acknowledgements

The authors acknowledge the Technical University of Kenya for providing an enabling academic environment during the course of this research. We also acknowledge the Energy and Petroleum Regulatory Authority (EPRA) for providing publicly available tariff information and the Open-Meteo project for making historical meteorological data freely accessible. We also acknwoledge Kenya Power for providing us with longitudinal data set for this research.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing Interests

The authors declare that they have no competing interests.

Ethics Declaration

This study did not involve human participants, animals, or any procedures requiring ethical approval..

Consent to Participate

Consent to Participate declaration: not applicable.

Consent to Publish

Consent to Publish declaration: not applicable.

Data Availability

The datasets generated and analysed during the current study are available from the corresponding author upon request.

Author Contributions

Paul Obasanjo: Conceptualization, methodology, software, data curation, formal analysis, visualization, investigation, writing and original draft preparation.Annette Mutinda: Review, validation, writing and editing.Adero Awuor: Supervision, review and Writing and Validation.All authors read and approved the final manuscript.

References

[1] Nabavi, S. A., Aslani, A., Zaidan, M. A., Zandi, M., Mohammadi, S., & Motlagh, N. H. (2020). Machine learning modeling for energy consumption of residential and commercial sectors. *Energies*, 13(19), 5171.

[2] Dada, E. G., Oyewola, D. O., & Yakubu, J. H. (2022). Power consumption prediction in urban areas using machine learning as a strategy towards smart cities. *Arid Zone Journal of Basic and Applied Research*, 1(1), 11-24.

[3] Jena, T. R., Barik, S. S., & Nayak, S. K. (2020). Electricity consumption & prediction using machine learning models. *Mukt Shabd Journal*, 9(6), 2347-3150.

[4] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference* (pp. 785-794).

[5] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.

[6] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3), e0194889.

[7] Jack, B. K., & Smith, G. (2015). Charging ahead: Prepaid metering, electricity use and utility revenue. *American Economic Journal: Applied Economics*, 7(4), 134-168.

[8] Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3), 914-938.

[9] Touzani, S., Granderson, J., & Fernandes, S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, 158, 1533-1543.

[10] Adanu, K., Akpalu, W., & Kuada, E. (2025). Switching from postpaid to prepaid meters and electricity consumption in Ghana. *Utilities Policy*, 95, 101949.

[11] Beyene, A. D., Jeuland, M., Sebsibie, S., et al. (2022). Pre-paid meters and household electricity use behaviors: Evidence from Addis Ababa, Ethiopia. *Energy Policy*, 170.

[12] Arabzadeh, V., et al. (2026). A Four-Dimensional Analysis of Explainable AI in Energy Forecasting: A Domain-Specific Systematic Review. *ResearchGate*.

[13] Alotaibi, S., et al. (2025). Explainable AI Framework Using XGBoost With SHAP and LIME for Multi-Scale Household Energy Forecasting. *IEEE Xplore*, 11141411.

[14] Li, Y., et al. (2025). Short-Term Load Forecasting in Power Systems Based on the Prophet–BO–XGBoost Model. *Energies*, 18(2), 227.

[15] Bloomer, J., & Boateng, D. (2024). Prepaid into poverty? Exploring energy poverty and access to alternative electricity sources among urban poor households in Ghana. *Energy Research & Social Science*, *112*, 103510.

[16] Guma, P. K., Monstadt, J., & Schramm, S. (2022). Post-, pre- and non-payment: Conflicting rationalities in the digitalisation of energy access in Kibera, Nairobi. *Digital Geography and Society*, *3*, 100037.

[17] Mohammed, A. S. (2024). Enhancing smart grid electricity prediction with the fusion of intelligent modeling and XAI integration. *International Journal of Advanced and Applied Sciences*.

[18] Fan, G.-F., Yu, M., Dong, S.-Q., Yeh, Y.-H., & Hong, W.-C. (2021). Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. *Utilities Policy*, *73*, 101284.

[19] Sese, B. M. (2023). *Short-Term Forecast Techniques for Energy Management Systems in Microgrid Applications* (Doctoral dissertation, Nelson Mandela African Institution of Science and Technology).

[20] Moulla, D. K., Attipoe, D., Mnkandla, E., & Abran, A. (2024). Predictive Model of Energy Consumption Using Machine Learning: A Case Study of Residential Buildings in South Africa. *Sustainability*, 16(11), 4365.

[21] Mosetlhe, T., & Yusuff, A. A. (2024). Forecasting of Residential Energy Utilisation Based on Regression Machine Learning Schemes. *Energies*, 17(18), 4681.

[22] Jacome, V., & Ray, I. (2018). The prepaid electric meter: Rights, relationships and reification in Unguja, Tanzania. *World Development*, 105, 262–272.