Predictive Energy Analytics: Leveraging Machine Learning and Weather Data Integration for Smart Consumption Forecasting and Optimization

Sorna Shanthi D[1], Priadharshni P[2], Priyadarshini S[3]
[1]Associate Professor, Department of Artificial Intelligence and Data Science
[2,3]UG Scholar, Department of Artificial Intelligence and Data Science,
Rajalakshmi Engineering College, Chennai, Tamil Nadu, India.

***Abstract*— Efficient energy management is increasingly essential for both residential and commercial sectors as they seek to optimize energy consumption, reduce costs, and contribute to environmental sustainability. This paper presents an AI-powered energy forecasting model that leverages historical consumption and weather data to predict future energy demand. Using the XGBoost algorithm, this model analyzes factors such as temperature, humidity, and other weather conditions to provide accurate, data-driven insights into energy usage patterns. To enhance interpretability and actionable insights, the model incorporates various data engineering techniques, including lag features, rolling statistics, and feature interactions, and provides visualizations like hourly and daily consumption trends. Key metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used to assess prediction accuracy. This system has been validated with a real-world dataset, demonstrating its potential to support smarter energy consumption practices and sustainability efforts. Moreover, the model provides actionable recommendations for both consumers and organizations to improve energy efficiency and reduce their carbon footprint, driving further impact in sustainable energy practices.**

Keywords*— *Energy consumption forecasting, XGBoost, Weather data, Data-driven insights, Energy optimization, Predictive analytics.

I. INTRODUCTION

In recent years, efficient energy management has become crucial for reducing costs and enhancing sustainability. Traditional systems, relying on manual monitoring and basic rules, often led to inefficiencies such as energy wastage and overconsumption. These challenges resulted in higher operational costs and significant environmental impact. AI-powered energy forecasting models have emerged as a solution, addressing these issues by predicting energy demand with greater accuracy.

Leveraging machine learning, such systems optimize energy usage, reducing waste and supporting sustainable practices. Dynamic AI-based energy forecasting models utilize historical consumption and weather data to predict energy demand, enabling better alignment between energy usage and actual requirements. By integrating advanced algorithms like XGBoost, these models provide actionable insights for real-time energy optimization. Studies have shown that implementing AI-based forecasting can reduce energy consumption by up to 20%, significantly improving energy efficiency and supporting sustainability initiatives.

Traditional methods, such as simple statistical models, often fall short in capturing complex energy consumption patterns influenced by seasonality and fluctuating demand. With advancements in machine learning and data-driven techniques, forecasting energy demand in real-time has become feasible, even with non-linear and volatile data. These innovations mark a transformative shift in energy management, where data-driven insights empower more efficient, cost-effective, and environmentally-friendly practices.

We chose to focus on residential energy usage as most prior studies [11], [12], [13], [14] have predominantly addressed energy consumption in industrial and commercial sectors. While these sectors contribute significantly to total energy demand, there is increasing recognition of the critical role residential energy consumption plays in overall usage patterns. Addressing and managing energy consumption at this level can make a meaningful impact on sustainability goals. Our objective is to empower individuals with actionable insights to make better-informed decisions about their energy usage and actively contribute to efficient energy resource management.

The structure of this manuscript is as follows: Section II reviews related works on energy prediction models and the application of XGBoost to these models. Section III details the proposed methodology used to achieve the results. Section IV presents the experimental results and analysis. Section V highlights the limitations of this study, and Section VI concludes the paper with a discussion of potential future work.

II. LITERATURE REVIEW AND CONTRIBUTIONS

In recent years, numerous studies have focused on energy forecasting, exploring various machine learning techniques to enhance the accuracy and reliability of predictive models. These studies have highlighted the potential of advanced algorithms like XGBoost in addressing the challenges of energy forecasting, particularly in terms of capturing non-linear patterns, handling missing data, and providing robust predictions even with volatile datasets.

This section covers studies that employed different energy forecasting techniques, emphasizing the use of machine learning models such as XGBoost. These models have been proven effective in improving the performance of energy prediction systems by leveraging historical energy usage data, weather variables, and other contextual factors. Additionally, we discuss how these techniques compare to traditional statistical methods, showcasing the benefits of modern, data-driven approaches for energy consumption forecasting.

A. STUDIES ON TRADITIONAL ENERGY FORECASTING TECHNIQUES

In the study [1], Tabasi, Aslani, and Forotan applied a linear regression model to predict energy consumption by analyzing historical data and identifying energy usage trends. The model offered a straightforward approach to forecasting, making it accessible for practical use in energy management systems. The research demonstrated the model's ability to establish energy consumption patterns, providing a reliable baseline for predictions. However, the linear regression model is limited in capturing non-linear relationships, which are often present in more complex energy datasets. This limitation suggests the need for more advanced methods capable of handling intricate data patterns to improve forecasting accuracy.

Ramos, Faria, Morais, and Vale [2] proposed the use of decision trees to select the best forecasting algorithms for electricity consumption in office buildings. By tailoring algorithm selection to specific consumption contexts, this approach enhanced prediction accuracy and energy efficiency. The study's contribution lies in its ability to optimize energy management by adjusting forecasting models according to building-specific factors like occupancy and working hours. However, the applicability of this method is confined to office environments, and its effectiveness in other types of buildings or industries remains untested. Additionally, decision trees may struggle to manage more complex or non-linear relationships in the data, limiting the scalability of the approach.

Kumar et al. [3] introduced the Quantum Support Vector Machine (QSVM) for forecasting household energy consumption, comparing it with traditional deep learning models. The study emphasized the advantages of quantum computing in processing large, high-dimensional datasets, allowing QSVM to provide highly accurate predictions. The comparative analysis showed QSVM outperformed deep learning models in terms of prediction accuracy and computational efficiency. However, the research faces practical challenges, particularly due to the high computational resources required for quantum computing, which may not be accessible for all users. Additionally, the study was limited to household energy consumption, and further research is needed to determine the applicability of QSVM in other sectors.

Reference	Technologies Used	Contributions	Limitations
[1]	Linear Regression	Developed a regression model to predict energy consumption; provided insights into energy usage trends.	Limited ability to capture non-linear relationships in energy consumption patterns.
[2]	Decision Tree	Proposed a method to select forecasting algorithms based on decision tree models for office buildings.	Limited generalizability to other types of buildings or contexts outside office environments.
[3]	Support Vector Machine (SVM)	Compared quantum SVM with deep learning models for house energy forecasting, showcasing quantum model benefits.	High computational cost and limited accessibility to quantum computing resources.
[4]	Deep Learning	Presented a deep learning approach for electricity consumption forecasting with improved accuracy.	Potential overfitting and need for extensive computational resources and large datasets.

Table 1. Summary of Studies on Traditional Energy Forecasting Technologies

Qureshi, Arbab, and Rehman [4] explored the use of deep learning techniques to forecast electricity consumption, leveraging advanced neural networks to model complex consumption patterns. The study demonstrated the power of deep learning models in capturing non-linear trends and improving the accuracy of energy consumption predictions. The researchers' contributions included showcasing the scalability of deep learning for large datasets and providing insights into energy management practices. However, deep learning models require significant computational resources, which may be a barrier for smaller organizations or real-time applications. The study also did not delve into the potential issues related to overfitting or the challenges of deploying deep learning systems in dynamic, real-time energy forecasting environments.

The studies summarized in the Table 1. illustrate various forecasting techniques for energy consumption, ranging from simple methods like linear regression to more complex

machine learning models such as deep learning approaches. While these models offer valuable insights and relatively accurate predictions, they each have limitations, such as difficulty handling non-linear relationships, scalability issues, and high computational requirements. The XGBoost algorithm stands out as a promising solution to these challenges, offering high predictive accuracy, efficiency, and the ability to handle complex data relationships. With its ensemble learning approach, XGBoost minimizes overfitting and ensures better generalization. By leveraging XGBoost, we can enhance prediction accuracy and address the limitations of earlier techniques, providing a robust solution for energy management applications.

B. STUDIES ON XGBOOST ALGORITHM FOR FORECASTING MODELS

Semmelmann, Henni, and Weinhardt [5] introduced a novel hybrid model combining Long Short-Term Memory (LSTM) networks and XGBoost for load forecasting in energy communities. Their model leverages smart meter data to improve the accuracy and efficiency of energy demand predictions. By integrating LSTM's ability to capture temporal dependencies with XGBoost's strength in feature selection and predictive modeling, the hybrid approach outperformed traditional methods. The study emphasized the potential of this model in optimizing energy management and enhancing decision-making processes in energy communities. However, the reliance on high-quality smart meter data and the computational complexity of the hybrid model were noted as limitations.

Sheng and Yu [6] proposed an optimized prediction algorithm based on XGBoost, focusing on enhancing its parameter tuning and model efficiency. Their algorithm applied advanced optimization techniques to improve the forecasting accuracy of energy consumption data. The study demonstrated the effectiveness of the proposed algorithm in reducing prediction errors and computational costs compared to standard XGBoost implementations. This research highlighted the adaptability of XGBoost in various contexts, though it acknowledged challenges in scaling the model for larger datasets or highly dynamic environments.

Zhang, Shao, and Zou [7] explored the application of XGBoost for predicting customer behaviors, showcasing its utility beyond traditional energy forecasting. Their study utilized customer interaction and transaction data to identify patterns and predict future behaviors effectively. XGBoost's ensemble learning approach provided high accuracy and interpretability, making it a robust choice for customer analytics. Despite its success in this domain, the study noted limitations in handling sparse or incomplete data, which could impact prediction quality in real-world scenarios.

Nti, Teimeh, Nyarko-Boateng, and Adekoya [8] conducted a systematic review of electricity load forecasting methodologies, analyzing the strengths and weaknesses of various approaches, including machine learning, statistical models, and hybrid techniques. The review highlighted the increasing adoption of machine learning methods like XGBoost due to their high accuracy and ability to model complex relationships. However, it also underscored the challenges of implementing these models, such as the need for extensive training data and computational resources. The study called for the integration of domain knowledge to enhance the interpretability and applicability of advanced forecasting methods in diverse contexts.

Reference	Technologies Used	Contributions	Limitations
[5]	LSTM – XGBoost Hybrid Model	Proposed a hybrid LSTM-XGBoost model for accurate day-ahead load forecasting in energy communities, using smart meter data.	The hybrid model has increased complexity, requiring intensive computational resources for training and implementation.
[6]	Hybrid PSO – XGBoost Model	Introduced an automated parameter optimization process for house pricing using hybrid model achieving greater accuracy.	Increased computational complexity due to the integration of PSO with XGBoost.
[7]	XGBoost	Identified key fraud indicators : injury claim, auto claim, capital gain and customer tenure to predict customer fraud .	Achieves moderate predictive accuracy of 0.63,requires refinement for better performance.
[8]	XGBoost along with Time series model	Developed robust load forecasting model with primary input as historical and weather data.	Limited focus on medium and long term forecasting.

Table 2. Summary of XGBoost Techniques for Forecasting Models

These studies, as summarized in Table 2, provide valuable insights into the application of various machine learning and hybrid models for energy consumption forecasting in industrial, commercial, and community settings. However, many of these approaches have limitations in terms of scalability, interpretability, and efficiency when applied to large-scale residential and commercial buildings with diverse energy usage patterns. This highlights the need for a robust, flexible, and accurate prediction framework tailored to these challenges.

Building upon the limitations of traditional forecasting techniques and insights gained from advanced machine

learning methodologies, our objective was to develop an energy consumption prediction workflow centered around the XGBoost algorithm. By leveraging the strengths of XGBoost, including its high predictive accuracy, feature importance evaluation, and ability to handle large datasets efficiently, the proposed framework aims to enhance the accuracy and usability of energy consumption forecasts. This enables actionable insights for energy optimization in residential and commercial environments.

The primary contributions of this study are summarized as follows:

- Development of an efficient and accurate energy consumption prediction workflow based on the XGBoost algorithm, tailored for real-world energy datasets.
- Rigorous evaluation of the XGBoost model using performance metrics such as Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), Mean Squared Error (MSE), and Mean Absolute Error (MAE) to ensure reliability and robustness.
- Incorporation of feature importance analysis provided by XGBoost to offer interpretability and actionable insights, facilitating better energy management decisions for end-users and stakeholders.
- Optimization of hyperparameters in the XGBoost model to enhance prediction accuracy and adapt the workflow for varying energy consumption patterns effectively.

This framework represents a step forward in addressing the scalability and interpretability issues of traditional methods, paving the way for more transparent, accurate, and practical energy forecasting solutions.

III. METHODOLOGY

This section describes the methodological approach followed in this work. We began by sampling and preprocessing electrical energy consumption data to ensure its suitability for analysis. The XGBoost model was then utilized for energy consumption prediction, leveraging its ability to handle complex patterns in data. The model was evaluated using various performance metrics to ensure accurate and reliable forecasting. Fig. 1 provides a visual overview of the methodology, illustrating the data preparation and modeling process.

A. DATA COLLECTION

The dataset utilized in this study captures household electric energy consumption from a residence located in the northeast region of Mexico over a period of 14 months. The data, recorded at one-minute intervals, provides a detailed representation of energy usage patterns alongside meteorological influences, sourced under a free license from OpenWeather. It includes energy-related metrics such as active power, current, voltage, reactive power, apparent power, and power factor, offering insights into the household's electricity dynamics. Additionally, weather-related attributes like temperature, humidity, atmospheric pressure, wind speed, and direction, as well as forecasted metrics for the next day's temperature and "feels like" temperature, further enrich the dataset. This combination of features provides a comprehensive understanding of energy consumption trends in real-world settings. By integrating both energy and weather data, this dataset serves as an excellent foundation for analysing energy usage and forecasting future consumption patterns.

The detailed energy consumption trends are shown in Fig 3, and a complete description of the features is presented in Table 3.

Features	Description	Unit
date	Timestamp of each record	-
active_power	Real power consumed by appliances	kW
current	Current flowing through the electrical system	Amperes(A)
voltage	Voltage level in the electrical supply	Volt(V)
reactive power	Reactive power in the system	kVAR
apparent power	Total power consumed	kVA
power factor	Efficiency of electricity usage	Ratio (0 to 1)
main	Categorical weather conditions	-
description	Detailed categorical weather description	-
temp	Recorded temperature	ºC
feels like	Perceived temperature	ºC
temp min	Minimum recorded temperature	ºC
temp max	Maximum recorded temperature	ºC
pressure	Atmospheric pressure	hPa
humidity	Relative humidity	%
speed	Wind speed	m/s
deg	Wind direction	Degrees (º)
temp t+1	Forecast temperature for the next day	ºC
feels like t+1	Forecast perceived temperature for the next day	ºC

Table 3. Description of all the Features in the Dataset

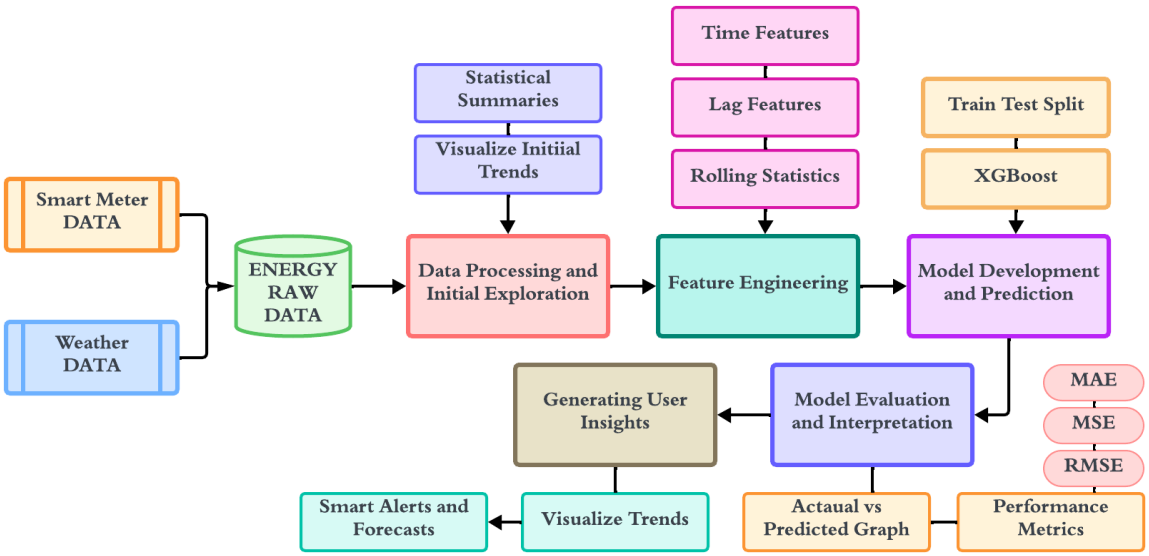

Fig. 1. Methodology

B. DATA PRE PROCESSING AND INITIAL EXPLORATION

The dataset underwent several preprocessing steps to prepare it for analysis and modeling. Initially, the raw data, encompassing both energy consumption and weather attributes, was loaded into a pandas DataFrame. The ***date*** column was converted to datetime format for temporal analysis, and new features, such as hour, day, and month, were extracted from the timestamp to facilitate time-series analysis. Missing values were identified and analyzed for each feature to ensure data completeness and consistency, as shown in **Equation (1)**:

$$\text{Missing Values} = \sum_{i=1}^{n} \mathbf{1}(x_i = \text{NaN}), \tag{1}$$

where x_i represents the value in the i-th row of the column, and $\mathbf{1}$ is the indicator function that equals 1 when x_i is NaN.

Descriptive statistics were calculated to summarize the numerical features, revealing key measures such as mean, standard deviation, minimum, and maximum values. The time-series analysis was conducted by plotting the ***active_power*** feature against the ***date*** to visualize energy consumption trends over time (Fig. 3). Additionally, the temperature distribution was examined using a histogram to identify its frequency across the dataset.

To further understand relationships between variables, a correlation matrix was computed for all numerical features. The Pearson correlation coefficient $\rho_{X,Y}$ was calculated for each pair of features, as shown in **Equation (2):**

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}, \tag{2}$$

where **Cov(*X*,*Y*)** is the covariance between features ***X*** and ***Y***, and $\boldsymbol{\sigma_X}$ and $\boldsymbol{\sigma_Y}$ are the standard deviations of ***X*** and ***Y***, respectively. The correlation matrix was visualized using a heatmap (Fig. 2) to highlight significant relationships between energy and weather variables.

These initial explorations provided crucial insights into the temporal dynamics of energy consumption and the interdependencies among features. The observed patterns, such as the periodicity in active power consumption and the influence of temperature on energy demand, informed the feature engineering and modeling phases of this study. The visualization and analysis of these correlations also enabled us to identify the most influential variables for predicting energy consumption, setting the foundation for the application of advanced machine learning models like XGBoost in subsequent stages of the study.

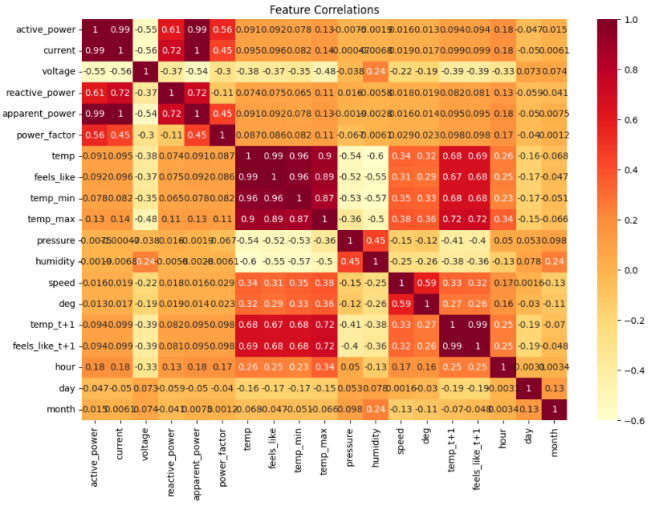

Fig. 2. Feature Correlation Map

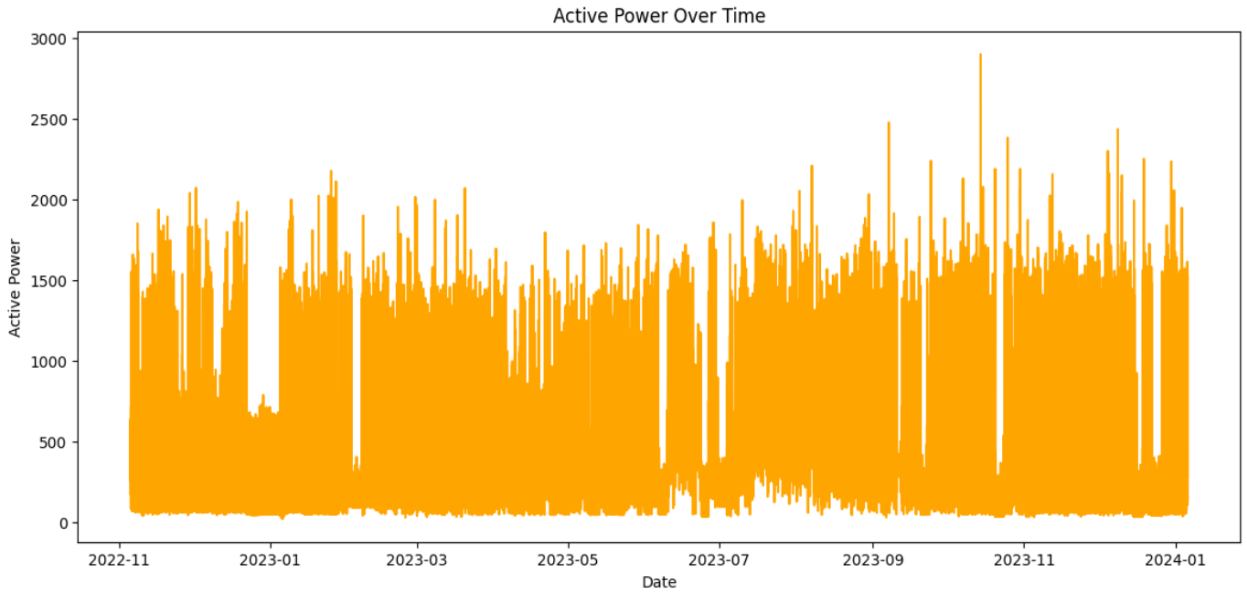

Fig. 3. Energy Consumption Trends Over Time

C. FEATURE ENGINEERING AND SELECTION

In this study, feature engineering was carried out to enhance the predictive capability of the dataset by incorporating temporal, lagged, and interaction-based features. These transformations aimed to capture underlying trends and relationships within the data for improved energy consumption prediction.

To understand the diurnal patterns of energy consumption, the hour feature was derived from the ***date*** column, and the average ***active_power*** was computed for each hour of the day. This analysis revealed the temporal variations in energy usage across a typical day (Fig. 4). Similarly, the **day_of_week** attribute was employed to calculate the average energy consumption for each day, illustrating weekly trends. The influence of monthly variations was also analyzed by grouping the data by the ***month*** attribute.

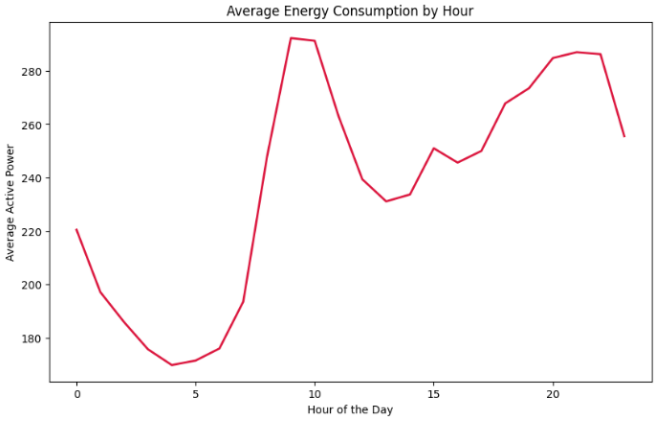

Fig. 4. Variations in Energy consumption across a day

To incorporate temporal dependencies, lagged features were created. Specifically, ***active_power_lag1*** and ***active_power_lag24*** represent the active power consumption values lagged by one hour and one day, respectively. The relationship between the current and lagged values was visualized through scatter plots, as seen in Fig. 6, and highlighted the correlation between sequential energy consumption data points.

To capture short-term variations and smooth the data, a 24-hour rolling mean and standard deviation were computed for the active_power feature, as defined in **Equations (3) and (4):**

$$\text{Rolling Mean}_t = \frac{1}{N} \sum_{i=t-N+1}^{t} x_i, \tag{3}$$

$$\text{Rolling Std}_t = \sqrt{\frac{1}{N} \sum_{i=t-N+1}^{t} (x_i - \text{Rolling Mean}_t)^2}, \tag{4}$$

where N is the window size (24 hours in this case), and x_i represents the active power consumption at time i.

These metrics highlighted underlying trends and anomalies in energy consumption, as depicted in Fig. 5.

The insights gained from this analysis informed the selection of features for the subsequent modeling phase.

To ensure data integrity, missing values were backfilled, maintaining the temporal consistency of the dataset. This robust feature engineering process not only enriched the dataset with meaningful attributes but also provided a deeper understanding of the factors influencing energy consumption patterns.

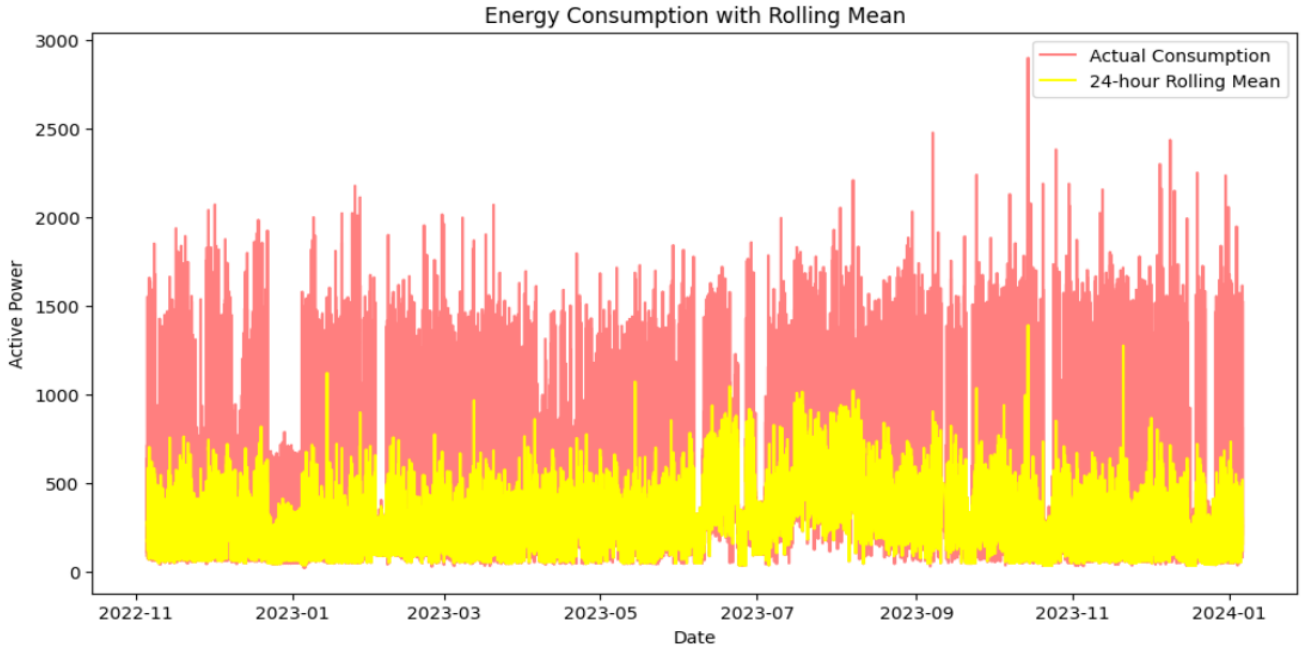

Fig. 5. **Energy Consumption with Rolling Mean**

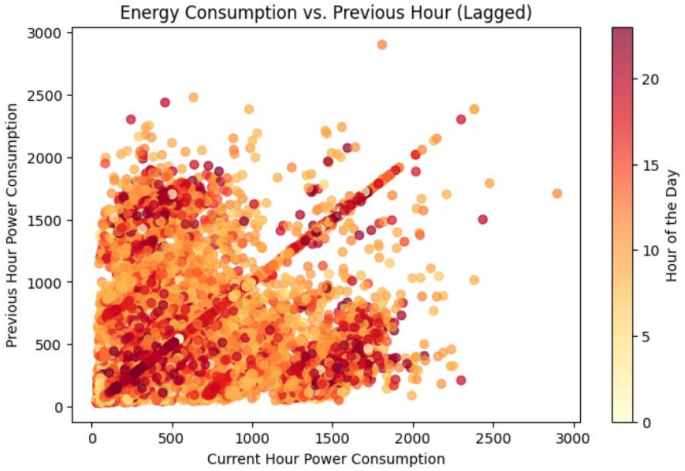

Fig. 6. **Energy Consumption with Lagged hour**

D. MODEL DEVELOPMENT AND PREDICTION

This study utilizes the XGBoost regression model to predict energy consumption. XGBoost, which stands for Extreme Gradient Boosting, is a powerful and efficient machine learning algorithm built on decision trees and enhanced through gradient boosting. It is particularly well-suited for large datasets and capturing complex relationships between input features and target variables. By leveraging its boosting mechanism, XGBoost can provide high predictive accuracy and robustness, making it ideal for predicting energy consumption patterns.

(i) Data Preparation

The dataset was initially processed by separating the features (***X***) from the target variable (***y***), where the target, ***active_power***, represents the energy consumption. To improve model performance, irrelevant features such as ***date***, ***main***, and ***description*** were removed. The dataset was then split into training and testing subsets, with **80%** of the data allocated for training (X_{train}, y_{train}) and **20%** for testing (X_{test}, y_{test}). This 80-20 split ensures that the model is trained on a substantial portion of the data while leaving enough for unbiased evaluation of its performance.

(ii) Model Training

The XGBoost regressor was configured with key hyperparameters to optimize its performance. The objective function was set to **reg:squarederror**, which is ideal for regression tasks. The model was trained with 100 estimators (boosting rounds), a learning rate of 0.1 (which determines the contribution of each tree to the final prediction), and a maximum depth of 6 (to prevent overfitting by limiting the complexity of individual trees). The model training followed an optimization process aimed at minimizing the objective function, which combines the loss function and a regularization term:

$$\text{Objective Function: } L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \Omega(f), \qquad (6)$$

In this equation, $l(y_i, y^i)$ represents the loss function that measures the error between the actual (y_i) and predicted (y^i) values, while $\Omega(f)$ is the regularization term that controls the complexity of the model. The goal was to iteratively minimize $L(\theta)$ to improve prediction accuracy.

(iii) Prediction and Output

After the model wasP trained, it was evaluated on the test dataset (X_{test}) to generate predictions (**y^**). These predictions are crucial for assessing the model's ability to forecast energy consumption accurately. To quantify the model's performance, further analysis and evaluation metrics will be provided in the following sections, where measures such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) will be used. This structured approach ensures that the model is both precise and reliable in predicting energy consumption, leveraging the strengths of XGBoost.

In summary, this methodology provides a comprehensive and systematic approach to energy consumption prediction, utilizing the power of XGBoost to produce accurate and meaningful results.

E. MODEL EVALUATION AND INTERPRETATION

The following analysis outlines the evaluation of the energy consumption prediction model, focusing on key metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics offer insights into the accuracy of the model's predictions.

Evaluation Metrics

The **Mean Absolute Error (MAE)** is calculated as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{7}$$

where y_i is the actual energy consumption, and **y^** is the predicted value. MAE provides a straightforward measure of prediction error in the same units as the target variable.

The **Mean Squared Error (MSE)** is given by:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{8}$$

MSE penalizes larger errors more significantly than MAE, making it sensitive to outliers.

The **Root Mean Squared Error (RMSE)**, derived from MSE, is calculated as:

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{9}$$

RMSE is useful for understanding the magnitude of error in the same units as the predicted variable, providing a more intuitive understanding of model accuracy.

To visually assess the model's prediction accuracy, we can compare the predicted values against the actual values through a scatter plot. This visualization helps identify how well the model performs across the range of energy consumption values. In the plot, the red dashed line represents the line of perfect prediction,

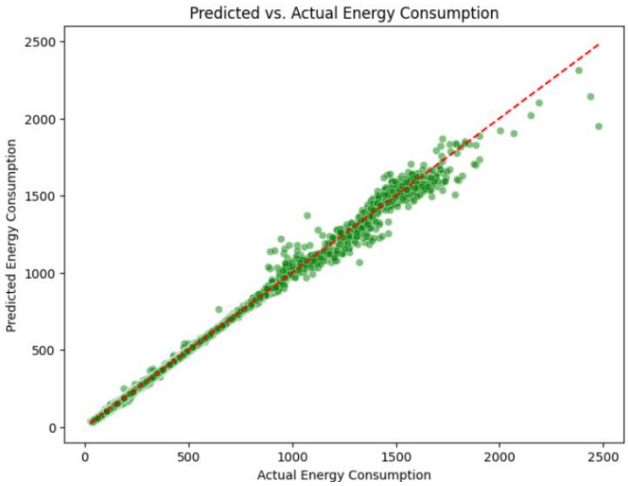

Fig. 7. Predicted vs Actual Energy Consumption

predicted and actual energy consumption values. Ideally, data points should lie close to the red dashed line, indicating accurate predictions.

F. GENERATING USER INSIGHTS

This section demonstrates how the XGBoost model can provide valuable insights into energy consumption patterns and help forecast future usage. The model predicts the energy consumption for the next period using the most recent data from the test dataset. By analyzing historical trends, it generates a forecast that enables users to anticipate their energy needs. A high-usage threshold is calculated dynamically using the following equation,

$$\text{High Usage Threshold} = \mu + 2\sigma \tag{10}$$

where **μ** is the mean and **σ** is the standard deviation of the training data. If the predicted consumption exceeds this threshold, a Smart Alert is triggered, notifying users to reduce their energy usage and optimize consumption. **Fig. 8** illustrates the forecasted energy consumption for the next periods, helping users make informed decisions about their energy consumption habits. This proactive forecasting and alerting system enhances energy management and helps users minimize costs and environmental impact.

IV. RESULTS AND DISCUSSION

In this section, we present the results of the XGBoost Regressor model's performance in predicting energy consumption and provide a detailed discussion on the model's evaluation. We also examine the model's feature importance, prediction errors, and discuss insights derived from the forecast.

The XGBoost Regressor (XGBR) model was trained and tested on the energy consumption dataset. The prediction performance was evaluated using three primary metrics:

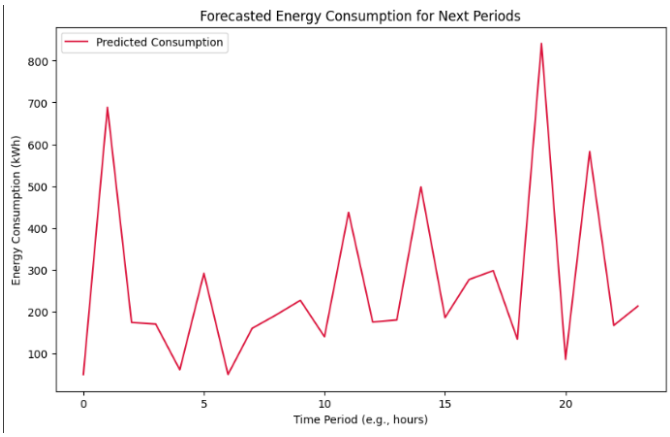

Fig. 7. Forecasted Energy Consumption for next periods

Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

A. RESULTS AND PERFORMANCE METRICS

The following results were obtained:

Mean Absolute Error (MAE): The MAE value of 1.66 indicates the average difference between predicted and actual energy consumption in kWh. A lower MAE means the model is more accurate in predicting energy consumption.
Mean Squared Error (MSE): The MSE value of 28.89 reflects the squared difference between predicted and actual values, where higher values indicate a larger discrepancy in predictions.
Root Mean Squared Error (RMSE): The RMSE value of 5.38 further highlights the extent of error, with a higher value pointing to more significant errors in the predictions.

These metrics suggest that the XGBoost model was effective in predicting energy consumption with relatively low error, especially when considering the complexity of the problem, where energy consumption is influenced by multiple factors such as temperature, humidity, time of day, and day of the week.

B. DISCUSSION

The results of the XGBoost Regressor (XGBR) model in predicting energy consumption demonstrate its effectiveness, with relatively low prediction errors as indicated by the evaluation metrics—Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The MAE value of 1.66 suggests that the model's predictions are, on average, off by just 1.66 kWh from the actual energy consumption. The MSE of 28.89 and RMSE of 5.38 indicate some residual discrepancies, but these errors are within an acceptable range for practical applications. Given the complexity of the problem, with energy consumption influenced by various factors such as temperature, humidity, and time of day, the XGBoost model has effectively captured these intricate relationships and provided reliable forecasts.

Additionally, the feature importance analysis revealed that temperature, humidity, and the hour of the day are the most influential factors driving energy consumption, which aligns with expectations. The model's ability to predict future energy consumption and generate smart alerts when high usage is detected enhances its utility in energy optimization. These insights can be leveraged for proactive energy management in settings like smart homes and commercial buildings. Overall, while some residual errors remain, the XGBoost model has proven to be a robust tool for predicting energy consumption, offering valuable forecasting capabilities and actionable insights for energy-saving strategies.

Algorithm 1: XGBoost model for Energy Consumption Prediction

```
Initialize:
Set nodeSet ← {0}, rowSet ← {0, 1, 2, ..., N}  // Initialize nodes and row sets
Define High Usage Threshold as μ + 2σ for alert generation

for t ← 1 to num_trees:
for i ← 1 to d:
for node in nodeSet:
usedRows ← rowSet[node]  // Identify rows to use for splitting
for k ← 1 to m:
H ← newHistogram()  // Create a new histogram for splitting
for j in usedRows:
    a. bin ← I.f[k][j].bin  // Assign data to bins
    b. H[bin].y ← H[bin].y + I.y[j]
    c. H[bin].n ← H[bin].n + 1
end for
best_split ← find_best_split(H)  // Find best split based on the histogram
update_tree(node, best_split)  // Update tree with the best split
update_rowSet(node, best_split)  // Update row set for the next iteration
end for
end for
end for
boosting_update()  // Update model using boosting
add_tree_to_ensemble()  // Add new tree to the ensemble
end for
```

V. CONCLUSION AND FUTURE WORKS

In this study, the XGBoost Regressor (XGBR) model demonstrated strong performance in predicting energy consumption based on historical data, weather conditions, and time-based features. The model's evaluation using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) showed that it effectively captured patterns in energy usage, yielding predictions with relatively low error. The feature importance analysis highlighted the significant factors influencing energy consumption, such as temperature and humidity, which aligned with expectations. Moreover, the model's ability to provide actionable insights, including high energy consumption alerts, offers substantial value for energy optimization in both residential and commercial contexts.

While the XGBoost model has shown promising results, there are several avenues for improvement and further exploration. Future

work could involve expanding the dataset to include more granular data, such as occupancy patterns or detailed device usage, to enhance the model's predictive capabilities. Additionally, incorporating more advanced time series forecasting methods, such as recurrent neural networks (RNNs) or long short-term memory networks (LSTMs), could provide more accurate predictions for complex, time-dependent energy consumption patterns. Further research could also focus on integrating real-time data to provide dynamic predictions and alerts, enabling users to take immediate actions to optimize energy consumption. Lastly, deploying the model in a real-world environment, such as a smart home or building management system, would help assess its performance and refine its functionalities for practical applications.

REFERENCES

[1] S. Tabasi, A. Aslani, and H. Forotan, "Prediction of Energy Consumption by Using Regression Model," *Computational Research Progress in Applied Science & Engineering,* vol. 2, no. 3, pp. 110-115, Jul. 2016. [Online]. Available: PEARL Publication, ISSN 2423-4591.

[2] D. Ramos, P. Faria, A. Morais, and Z. Vale, "Using decision tree to select forecasting algorithms in distinct electricity consumption context of an office building," in *Proc. 8th Int. Conf. Energy and Environment Research (ICEER 2021)*, Porto, Portugal, Sep. 13–17, 2021, pp.

[3] Karan Kumar. et al., "Quantum support vector machine for forecasting house energy consumption: a comparative study with deep learning models," *Journal of Cloud Computing: Advances, Systems and Applications,* vol. 13, no. 105, 2024. [Online]. Available: https://doi.org/10.1186/s13677-024-00669-x. [Accessed: Nov. 17, 2024].

[4] M. Qureshi, M. A. Arbab, and S. ur Rehman, "Deep learning-based forecasting of electricity consumption," *Scientific Reports*, vol. 14, no. 6489, 2024. doi: 10.1038/s41598-024-56602-4.

[5] L. Semmelmann, S. Henni, and C. Weinhardt, "Load forecasting for energy communities: a novel LSTM-XGBoost hybrid model based on smart meter data," Energy Informatics, vol. 5, Suppl. 1, Art. no. 24, pp. 1–21, Sep. 2022. [Online]. Available: IEEE Xplore. DOI: 10.1109/PTC.2019.8810902.

[6] C. Sheng and H. Yu, "An optimized prediction algorithm based on XGBoost," in Proc. 2022 Int. Conf. Networking and Network Applications (NaNA), Xinjiang, China, Nov. 2022, pp. 442–447. [Online]. Available: IEEE Xplore. DOI: 10.1109/NaNA56854.2022.00082.

[7] Y. Zhang, C. Shao, and C. Zou, "Prediction of customers' behaviors based on XGBoost model," in 2023 2nd International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI), Beijing-Tianjin, China, 2023, pp. 369–374. DOI: 10.1109/ICDACAI59742.2023.00076.

[8] I. K. Nti, M. Teimeh, O. Nyarko-Boateng, and A. F. Adekoya, "Electricity load forecasting: A systematic review," Journal of Electrical Systems and Information Technology, vol. 7, no. 1, pp. 1–19, 2020. DOI: 10.1186/s43067-020-00021-8.

[9] Yenduri and T. R. Gadekallu, ‘‘XAI for maintainability prediction of software-defined networks,’’ in Proc. 24th Int. Conf. Distrib. Comput. Netw., Jan. 2023, pp. 402–406.

[10] X. Wang and M. Yin, ‘‘Are explanations helpful? A comparative study of the effects of explanations in AI-assisted decision-making,’’ in Proc. 26th Int. Conf. Intell. User Interfaces, Apr. 2021, pp. 318–328.

[11] T. Sim, S. Choi, Y. Kim, S. H. Youn, D.-J. Jang, S. Lee, and C.-J. Chun, ‘‘EXplainable AI (XAI)-based input variable selection methodology for forecasting energy consumption,’’ Electronics, vol. 11, no. 18, p. 2947, Sep. 2022.

[12] Y.-C. Hu, ‘‘Electricity consumption prediction using a neural-network based grey forecasting approach,’’ J. Oper. Res. Soc., vol. 68, no. 10, pp. 1259–1264, Oct. 2017.

[13] U. Schlegel, H. Arnout, M. El-Assady, D. Oelke, and D. A. Keim, ‘‘Towards a rigorous evaluation of XAI methods on time series,’’ in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW), Oct. 2019, pp. 4197–4201.

[14] M. Kuzlu, U. Cali, V. Sharma, and Ö. Güler, ‘‘Gaining insight into solar photovoltaic power generation forecasting utilizing explainable artificial intelligence tools,’’ IEEE Access, vol. 8, pp. 187814–187823, 2020.

[15] F. Divina, A. Gilson, F. Goméz-Vela, M. G. Torres, and J. Torres, ‘‘Stack ing ensemble learning for short-term electricity consumption forecasting,’’ Energies, vol. 11, no. 4, p. 949, Apr. 2018.

[16] J. F. Torres, F. Martínez-Álvarez, and A. Troncoso, ‘‘A deep LSTM network for the Spanish electricity consumption forecasting,’’ Neural Comput. Appl., vol. 34, no. 13, pp. 10533–10545, Jul. 2022.

[17] S. Divya, A. Murthy, S. S. Babu, S. I. Ahmed, and S. R. Dey, ‘‘Energy monitoring with trend analysis and power signature interpretation,’’ in Proc. Emerg. Technol. Data Mining Inf. Secur. (IEMIS), vol. 3. Singapore: Springer, 2022, pp. 89–102.

[18] M. C. Thrun, A. Ultsch, and L. Breuer, ‘‘Explainable AI framework for multivariate hydrochemical time series,’’ Mach. Learn. Knowl. Extraction, vol. 3, no. 1, pp. 170–204, Feb. 2021.