Machine Learning for Energy Management

Predictive Analysis of Energy Consumption Patterns Using Machine Learning

Sadia Ahsan

***Abstract*— Efficient energy management has become increasingly vital in addressing the global challenges related to resource sustainability and environmental conservation. Robust energy management strategies are essential for achieving sustainability, cost reduction, and infrastructure resilience. The integration of Machine Learning (ML) techniques holds promise for optimizing energy consumption and enhancing sustainability. ML algorithms effectively enhance energy management by learning patterns and insights from historical energy data. This research explores the application of the state-of-the-art XGBoost (Extreme Gradient Boosting) machine learning (ML) algorithm to accurately predict energy consumption. The analysis involves studying the hourly, daily, and weekly energy consumption values of a building, utilizing a rich historical publicly available electricity consumption dataset obtained from individual household smart meters to predict the Global Average Power (GAP). The dataset spans four years and is sourced from the UCI Machine Learning (ML) Repository. To comprehensively assess the model's prediction accuracy, metrics such as mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) are computed. The results demonstrate the efficiency of the XGBoost model in predicting the global average power (GAP), highlighting the potential of machine learning (ML) algorithms in developing efficient energy management solutions for real-time monitoring of physical parameters.**

Keywords—Machine Learning (ML), XGBoost, Energy management, Energy consumption, MAE, RMSE, MSE

I. INTRODUCTION

In recent years, the escalating demand for efficient energy management systems has emerged as a pivotal response to the global challenges of resource sustainability and environmental conservation. Efficient energy management is imperative for mitigating environmental impact and ensuring sustainable operations, cost-effectiveness, and resilience in the face of growing energy demands and fluctuating resources.

Energy management encompasses a diverse set of practices aimed at understanding, planning, and optimizing energy use within an organization [1]. This includes measures to reduce energy usage, enhance efficiency, and adopt sustainable energy sources.

One of the key applications of energy management is in predicting future energy consumption to reduce overall energy usage [2]. Over the last two decades, energy consumption has experienced an overwhelming increase globally, driven by economic developments and a growing population [3]-[4]. Accurate energy consumption prediction is essential for the appropriate implementation of efficient energy management strategies, guiding organizations toward informed decision-making and sustainable practices. Precise energy consumption forecasting enables energy producers to generate and distribute the appropriate amount of power in advance [5].

Energy demand fluctuates throughout the day, week, and even seasonally due to several factors such as weather conditions and industrial activities. Therefore, estimating overall energy consumption at different time scales allows for robust or monitored strategies to optimize resource allocation, minimize costs, and ensure grid stability [6], [7]. This proactive approach helps prevent potential blackouts resulting from insufficient power production compared to consumption [5]. Therefore, the objective of this study is to develop a machine learning-based model that can accurately predict household energy consumption across various time intervals, including hourly, daily, and weekly periods. By leveraging historical energy consumption data and employing advanced machine learning techniques, the aim is to build a predictive model that not only captures the underlying patterns in energy consumption but also provides insights for effective energy management strategies.

II. LITERATURE REVIEW

Previous studies have explored various methodologies, from statistical models to machine learning algorithms, to address the crucial need for effective energy management by predicting energy consumption. Statistical methods such as time series analysis [8] and regression models [9], [10] have been widely utilized for their simplicity and interpretability. However, they may struggle to capture the complex patterns and nonlinear relationships present in energy consumption data.

In recent years, Machine Learning algorithms have emerged as powerful tools for energy consumption prediction, offering the capability to discern complex patterns and make accurate predictions [11]. In a study conducted by Ningning Zheng et al. [12], conducted a study on short-term electricity load forecasting, evaluating multiple ML algorithms including Random Forest, XGBoost, and Support Vector Machine Regression (SVM). Their findings highlighted the superior performance of the XGBoost model in terms of accuracy and efficiency.

Similarly, S. Pokharel and P. Ghimire investigated the application of ML models in predicting total energy consumption in a low-energy house [13]. In this comparative study, they evaluated four ML models: Extreme Gradient Boosting (XGBoost), Random Forest, Decision Tree, and Support Vector Machine (SVM) [13]. Their results indicated that the XGBoost model outperformed all other models in terms of prediction accuracy. With compelling metrics such as a coefficient of determination (R2) of 61%, RMSE of 65.28, MAE of 29.81, and MAPE of 28.55 on the testing set, the XGBoost model exhibited superior performance compared to others [13].

Given the robust performance demonstrated by the XGBoost algorithm in previous studies [12]-[13], this study employs the XGBoost ML model to develop an efficient energy management solution.

II. METHODOLOGY

A. *Design*

This section outlines the procedural steps employed in developing a machine learning-based model for predicting global energy consumption. Fig. 2 illustrates the design methodology adopted in this study. First, a historical electric power consumption dataset is collected from individual household smart meters [14]. Then Initial preprocessing steps, including handling of missing values, resampling, and normalization, are applied to a publicly available Individual Household Electric Power Consumption Dataset [14].

Following preprocessing, the dataset is partitioned into training and test sets using a 70-30 ratio, where 70% is allocated for training the machine learning model, and the remaining 30% is reserved for testing. Subsequently, the state-of-the-art XGBoost (Extreme Gradient Boosting) ML model is trained on the prepared training data.

In the final stage, the performance of the trained model is evaluated using standard metrics such as Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE) on the held-out test data. These metrics provide a comprehensive assessment of the model's accuracy and predictive capabilities.

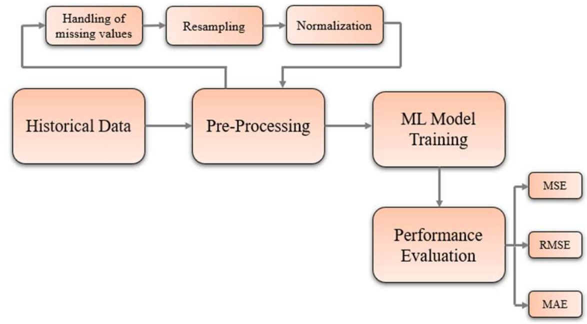

Fig. 1. Design methodology

B. *XGBoost – Machine Learning Model*

XGBoost (Extreme Gradient Boosting) is a machine learning algorithm that utilizes an optimized and efficient implementation of gradient boosting [15]. This technique is employed as an alternative modeling approach to assess its effectiveness in predicting energy consumption patterns. Unlike traditional machine learning methods, XGBoost constructs an ensemble of weak learners, typically decision trees, and iteratively enhances their predictive performance [16].

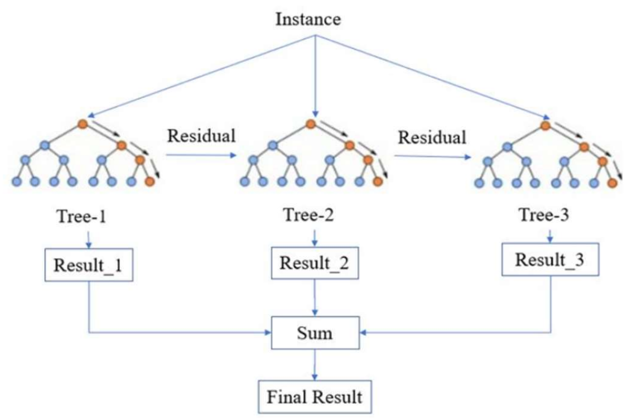

Fig. 2. Illustration of XGBoost algorithm [18]

1. *Training Phase:*

During training, XGBoost sequentially builds multiple decision trees, as depicted in Figure 2. Each tree aims to correct errors made by its predecessors. XGBoost optimizes its training by focusing on minimizing residuals, representing the differences between actual and predicted values. These residuals are intelligently fed into the next tree, enabling each subsequent tree to target and reduce the remaining errors not captured by the ensemble.

2. *Prediction Phase:*

After constructing an ensemble of decision trees during training, XGBoost makes predictions efficiently on new data. Each instance is passed through every tree and the individual predictions from each tree are combined to yield the final prediction [17]. The prediction for a specific instance is generated by summing the predictions from each tree, with each tree's contribution weighted by a factor that reflects its performance on the training data [16].

3. Mathematical Insight:

The goal of XGBoost is to iteratively optimize an objective function $L(t)$ during the training phase. This objective function is expressed as follows:

$$L(t)=\sum_{i=1}^{n} l(y_i , \hat{y}_i^{(t-1)} + f_t(x_i)) + \sum_{k=1}^{K} \Omega\,(f_t)$$

Where:

- n is the number of training instances.
- $l(y_i ,\hat{y}_i^{(t-1)} + f_t(x_i))$ is the individual loss term for each training instance i at iteration t, where $\hat{y}_i^{(t-1)}$ is the predicted value at iteration t-1 and $f_t(x_i)$ is the contribution of the new tree f_t, for an instance i at iteration t.
- $\Omega(f_t)$ is the regularization term for the tree ft at iteration t.

The algorithm $L(t)$ focuses on minimizing the overall prediction error and incorporates regularization techniques to prevent overfitting [19].

C. Dataset Description

The experiments in this study utilize the Individual Household Electric Power Consumption dataset [14], obtained from the UCI Machine Learning Repository. This dataset [14] meticulously recorded multivariate time series captures the electricity consumption of a single household at a one-minute sampling rate, providing insights over a comprehensive four-year period from December 2006 to November 2010. The dataset [14] containing 2,075,259 measurements, originates from a residence in Sceaux, approximately 7 km from Paris, France, offering a detailed perspective on power consumption dynamics within the household. The recorded measurements represent real-world electrical consumption patterns and contribute to the exploration of efficient energy management strategies.

Key features in the dataset:

The Individual Household Electric Power Consumption dataset [14] contains essential features that provide a detailed insight into various aspects of energy usage:

- **Global Active Power (GAP):** The overall active power averaged over a minute consumed by the household (kilowatts)
- **Global Reactive Power (GRP):** The overall reactive power consumption averaged over a minute consumed by the household (kilowatts).
- **Voltage:** The voltage level averaged over a minute (volts).
- **Global Intensity:** The household's minute-averaged current intensity (Amperes).
- **Sub-Metering 1:** The active energy consumption for specific appliances or areas (watt-hours).
- **Sub-Metering 2:** The active energy consumption for designated areas (watt-hours).
- **Sub-Metering 3:** The active energy consumption for climate control systems (watt-hours).

TABLE I. KEY CHARACTERISTICS OF THE INDIVIDUAL HOUSEHOLD ELECTRIC POWER CONSUMPTION DATASET [14]

Attribute	Description
Building(s)	Residential
Date Range	2006 - 2010
Duration	47 months
Type of Attributes	Real Numbers
Number of Buildings	1
Level of Detail	The whole building, circuit level
Location of Building	France
Sampling Frequency	Minutes

D. Data Pre-Processing

The Individual Household Electric Power Consumption dataset [14], obtained from the UCI Machine Learning Repository, undergoes a series of preprocessing steps to facilitate meaningful analysis and model training. This section outlines the fundamental data processing steps involved in transforming the initial raw data into a format suitable for machine learning tasks.

1. Handling of Missing values

The dataset [14], formatted in .txt, contains around 8.76% missing values in its measurements with each feature containing 25979 missing values. Although all calendar timestamps are available, some specific timestamps lack corresponding measurement values, which can be identified by the absence of a value between consecutive semi-colon attribute separators [14]. It is crucial to emphasize that these missing values cannot be disregarded, and a strategy other than deletion is employed to address this issue effectively [20]. To handle this issue effectively, mean imputation [21] is employed for columns with missing entries. Mean imputation is chosen for its effectiveness in handling missing data by replacing the missing values with the mean of the observed values for the respective variable. This approach provides a balanced treatment in the pre-processing stage without introducing bias [21].

2. Resampling:

To provide a comprehensive analysis of energy consumption trends across various time scales, the temporal resolution of the data undergoes resampling. The original dataset [5], characterized by measurements at a per-minute frequency, is resampled to higher intervals. The data is resampled from minutes to higher temporal resolutions, such as hours, weeks, and days.

By aggregating minute-level data into higher intervals, such as hours, days, and weeks, we effectively reduce the volume of data points while preserving the essential patterns and trends within the dataset [22]. This adjustment allows for a more in-depth understanding of energy consumption patterns over different time scales, enabling a more nuanced examination and providing valuable insights for the prediction model [22].

3. Normalization

To ensure uniformity in the scale of features, normalization is applied using Min-Max scaling. This process transforms the values of each feature to a standardized range, typically between 0 and 1 [23]. The following formula [24] calculates Min-Max scaling, where X is the original feature value, min(X) is the minimum value of the feature, and max(X) is the maximum value of the feature.

$$X' = \frac{X - m \quad (X)}{\max(X) - \min(X)}$$

This process prevents any particular feature from dominating the learning process due to differences in scale, promoting effective model training and result interpretation.

E. Data splitting

After the preprocessing step, the dataset [14] was partitioned into two sets i.e. train and test. In this study, a 75-25 split ratio is employed, with 75% of the data allocated for training and the remaining 25% for testing.

F. XGBoost Model Training:

XGBoost (Extreme Gradient Boosting) is chosen as the machine learning algorithm due to its proven effectiveness in handling complex relationships in time series data within features and ensuring accurate predictions of total power utilization.

1. Features and Target variable:

For training the XGBoost ML model, the selection of features and target variables is crucial to ensure accurate prediction of energy consumption patterns. Features such as Global Reactive Power (GRP), Voltage, Global Intensity, Sub-Metering 1, Sub-Metering 2, and Sub-Metering 3 are selected as inputs to capture the inherent dynamics of energy consumption. These features have been carefully selected to encompass the complex dynamics of energy consumption within the household. By incorporating these features, the model can learn patterns essential for predicting total power consumption accurately. The target variable selected for prediction is the Global Active Power (GAP) which reflects the overall energy consumed by the household. By focusing on Global Active Power (GAP) as a target variable, the model aims to provide valuable insights into total energy demand, facilitating informed decision-making regarding energy management strategies and resource allocation. Global Active Power (GAP) is pivotal for gaining valuable insights into total energy demand and accurately predicting overall energy consumption patterns, thus enabling effective energy management [25].

To comprehensively analyze energy consumption trends across different time scales, the dataset [14] undergoes resampling. Initially captured at the minute level, it is transformed into varying intervals, including hours, weeks, and days. This transformation provides a more detailed understanding of energy consumption patterns.

2. Hyperparameters:

The hyperparameters of the XGBoost model are tuned to achieve optimal performance while preventing overfitting.

- n_estimators: The number of trees in the XGBoost model is set to 150 which balances the computational efficiency and model complexity.
- max_depth: A maximum depth of the tree is set to 3 to capture the more complex patterns in the training data while preventing overfitting.
- Learning rate: The learning rate is set to 0.1 to optimize the training process.
- subsample: The Subsample is set to 1.0 to utilize the entire training set and to consider all samples during the construction of each tree.
- colsample_bytree: The fraction of features used for building each tree is set to 1.0 to consider all features for splitting at each node.

These hyperparameter settings ensure the XGBoost model's robustness and effectiveness in capturing the intricate patterns present in the energy consumption dataset [14].

G. Evaluation Matrix

The performance of the XGBoost for energy consumption prediction for efficient energy management was evaluated based on the Mean Squared Error (MSE) [26], Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Error (MAE) [27].

1. Mean Squared Error (MSE)

Mean Squared Error (MSE) [26] quantifies the average of the squared differences between predicted yi and actual y_i energy consumption values. It signifies the overall magnitude of errors, by describing how closely the predictions align with the true values [26].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2$$

2. Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) is the square root of the Mean Squared Error, providing a measure of the average magnitude of prediction errors. A lower RMSE indicates a closer alignment of predictions with actual values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

3. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) [27] represents the average of the absolute differences between predicted and actual values.

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

H. RESULTS

Table 2 showcases the predictive performance of the XGBoost Model for Global Active Power (GAP) across three distinct time scales: Hourly, Daily, and Weekly. The performance is evaluated based on the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

TABLE II.

PERFORMANCE EVALUATION METRICS FOR GLOBAL ACTIVE POWER (GAP) PREDICTION AT DIFFERENT TIME SCALES I.E HOURLY, DAILY, AND WEEKLY

Time Scale	**MSE**	**RMSE**	**MAE**
Hourly	2.704e-05	0.0052	0.0023
Daily	6.735e-05	0.0082	0.0037
Weekly	1.0803e-4	0.0103	0.0056

The presented Performance Evaluation Metrics for Global Active Power (GAP) Prediction reveal insight into overall energy consumption patterns across different time scales i.e. Hourly, Daily, and Weekly.

The model exhibits commendable performance in Hourly predictions, as evidenced by the low Mean Squared Error (MSE) of 2.704e-05 and Root Mean Squared Error (RMSE) of 0.0052. The minimal Mean Absolute Error (MAE) of 0.0023 indicates that, on average, predicted values deviate insignificantly from actual values. This precision at an hourly scale is essential for real-time monitoring and control of energy consumption. In daily predictions of overall power consumed, the model maintains its effectiveness with an MSE of 6.735e-05, RMSE of 0.0082, and MAE of 0.0037. The slightly higher error metrics compared to the Hourly predictions are expected due to the aggregation of the dataset [14] over a longer period. However, the model's accuracy remains commendable, making it suitable for broader energy management decisions on a daily basis. For the predictions of time scale over a week, the model continues to exhibit reliability with an MSE of 1.0803e-4, RMSE of 0.0103, and MAE of 0.0056. The model's robustness in capturing overall trends and variations over a week is evident. Although slightly higher error metrics are observed, compared to shorter time scales i.e. the power consumed on an hourly and daily basis. This is anticipated due to data aggregation over a more extended period.

The XGBoost Model's proficiency is evident in its ability to provide accurate predictions across diverse time scales, making it a valuable tool for real-world energy management applications.

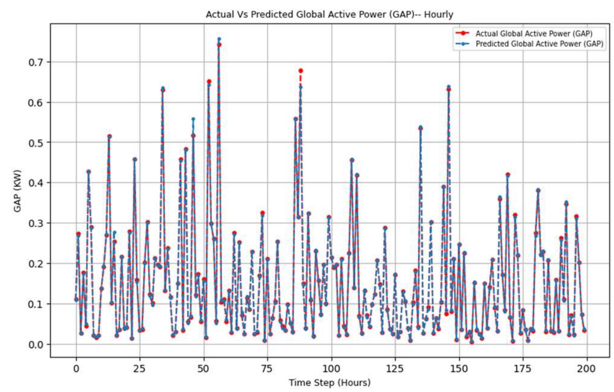

Fig. 3. Actual vs Predicted Global Active Power (GAP) - 200 Test Samples (Hourly)

Fig. 3 illustrates the actual and predicted Global Active Power (GAP) values over an hourly time scale, showcasing the first 200-time steps for clarity. The red dashed line represents the actual Global Active Power (GAP) values, while the blue dashed line represents the predicted Global Active Power (GAP) values generated by the XGBoost model. The closeness of the two lines suggests a high degree of alignment between the actual and predicted values, affirming the model's accuracy in capturing hourly energy consumption patterns. This close correspondence between actual and predicted values is crucial for real-time monitoring and decision-making in energy management.

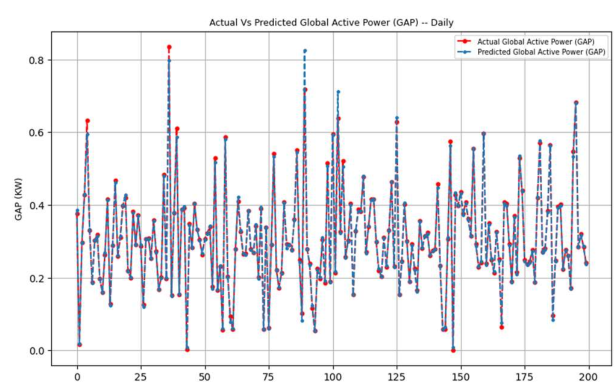

Fig. 4. Actual vs Predicted Global Active Power (GAP) - 200 Test Samples (Daily)

Fig. 4 presents the actual and predicted Global Active Power (GAP) values over a daily time scale, encompassing 200 values of test data. Similar to Figure 3, the red dashed line represents actual values, and the blue dashed line represents predicted values. The consistent alignment between the two lines reinforces the model's capability to accurately predict energy consumption at a more extended time scale.

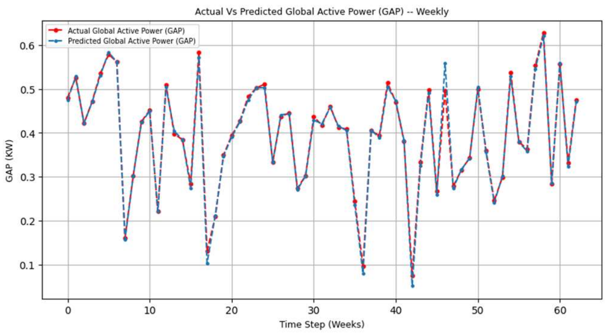

Fig. 5. Actual vs Predicted Global Active Power (GAP) - 200 Test Samples (Weekly)

Fig. 5 highlights the actual and predicted Global Active Power (GAP) values over a weekly time scale, covering 62 weeks of test data. Unlike the predictions on hourly and daily time scales, the results show a higher difference between the actual and predicted Global Active Power (GAP) values. This discrepancy is expected as weekly predictions involve capturing trends and variations over a more extended period. Factors such as changing energy consumption patterns on different days of the week and external influences become more pronounced, contributing to the observed differences. Despite the increased variability, the model maintains a reasonable level of accuracy in capturing overall trends over weekly intervals.

I. Discussion and conclusion

In this study, the application of the XGBoost ML algorithm is investigated for predicting energy consumption patterns across various time scales, ranging from hourly to weekly intervals, aiming at enhancing energy management efficiency. The model exhibited commendable performance, as evidenced by low error metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), across different prediction horizons. Notably, the model exhibits remarkable precision at the hourly scale, facilitating real-time monitoring and control of energy usage. As the prediction time scale expands to daily and weekly intervals, slight increases in error metrics are observed due to data aggregation over longer periods. However, the model's performance remains commendable, underscoring its suitability for broader energy management decisions. The precision of these predictions holds significant implications for improved energy demand planning, mitigating the risks associated with under or overproduction of energy, thereby optimizing resource allocation and minimizing costs. Furthermore, the integration of machine learning techniques such as XGBoost enables the implementation of proactive energy management strategies, contributing to sustainability objectives and bolstering infrastructure resilience. Continuous refinement of ML algorithms and exploration of alternative modeling techniques are crucial for further enhancing performance in energy management applications. Addressing data quality issues and ensuring the reliability of the data used to train and test ML models are critical steps in achieving reliable results and thereby providing success to energy management systems based on machine learning.

Adopting ML based driven energy management solutions represents a significant stride towards attaining energy efficiency objectives, ultimately leading to more sustainable and resilient energy systems.

References

[1] A. A. Ajmi, N. S. Mahmood, K. R. Jamaludin, H. H. A. Talib, S. Sarip, and H. M. Kaidi, "Performance shaping factors for future sustainable energy management: A new integrated approach," *Heliyon*, vol. 9, no. 8, p. e18438, Aug. 2023, doi: 10.1016/J.HELIYON.2023.E18438.

[2] A. De Paola, M. Ortolani, G. Lo Re, G. Anastasi, and S. K. Das, "Intelligent management systems for energy efficiency in buildings: A survey," *ACM Computing Surveys*, vol. 47, no. 1. Association for Computing Machinery, 2014. doi: 10.1145/2611779.

[3] M. Sajjad *et al.*, "A Novel CNN-GRU-Based Hybrid Approach for Short-Term Residential Load Forecasting," *IEEE Access*, vol. 8, pp. 143759–143768, 2020, doi: 10.1109/ACCESS.2020.3009537.

[4] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy Build*, vol. 40, no. 3, pp. 394–398, Jan. 2008, doi: 10.1016/J.ENBUILD.2007.03.007.

[5] J. Hyeon, H. Lee, B. Ko, and H. Choi, "Deep learning-based household electric energy consumption forecasting," *The Journal of Engineering*, vol. 2020, no. 13, pp. 639–642, Jul. 2020, doi: 10.1049/joe.2019.1219.

[6] R. Jinjin, W. Shihuai, W. Zhengyang, and M. Yuting, "Residential Energy Use Prediction across different Time Scales with Advanced Machine Learning Techniques," *2019 2nd Asia Conference on Energy and Environment Engineering (ACEEE)*, no. IEEE, pp. 20–24, Jul. 2019, doi: 10.1109/ACEEE.2019.8816955.

[7] G. Li, W. Tian, H. Zhang, and B. Chen, "Building Energy Models at Different Time Scales Based on Multi-Output Machine Learning," *Buildings*, vol. 12, no. 12, Dec. 2022, doi: 10.3390/buildings12122109.

[8] Nataraja.C, M.B.Gorawar, and S. HarshaJ, "Short Term Load Forecasting Using Time Series Analysis," *Certified International Journal of*

Engineering Science and Innovative Technology (IJESIT), vol. 9001, no. 2, 2008.

[9] V. Veeramsetty, A. Mohnot, G. Singal, and S. R. Salkuti, "Short term active power load prediction on A 33/11 kV substation using regression models," *Energies (Basel)*, vol. 14, no. 11, Jun. 2021, doi: 10.3390/en14112981.

[10] S. Tabasi, A. Aslani, and H. Forotan, "Prediction of Energy Consumption by Using Regression Model," *Computational Research Progress in Applied Science & Engineering ©PEARL publication*, vol. 02, no. 03, pp. 110–115, 2016.

[11] A. Bakurova, O. Yuskiv, D. Shyrokorad, A. Riabenko, and E. Tereschenko, "NEURAL NETWORK FORECASTING OF ENERGY CONSUMPTION OF A METALLURGICAL ENTERPRISE," *Innovative Technologies and Scientific Solutions for Industries*, no. 1 (15), pp. 14–22, Mar. 2021, doi: 10.30837/itssi.2021.15.014.

[12] N. Zheng, Y. Shao, S. Zou, and Z. Ma, "Short-term load forecasting based on wavelet decomposition and XGBoost," in *WCSE 2020: 2020 10th International Workshop on Computer Science and Engineering*, International Workshop on Computer Science and Engineering (WCSE), 2020, pp. 403–410. doi: 10.18178/wcse.2020.06.060.

[13] S. Pokharel and P. Ghimire, "Data-driven MLmodels for accurate prediction of energy consumption in a low-energy house: A comparative study of XGBoost, Random Forest, Decision Tree, and Support Vector Machine," *Journal of Innovations in Engineering Education*, vol. 6, no. 1, pp. 12–20, Dec. 2023, doi: 10.3126/jiee.v6i1.54965.

[14] G. and B. A. Hebrail, "Individual household electric power consumption. ," UCI Machine Learning Repository. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption

[15] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Mar. 2016, doi: 10.1145/2939672.2939785.

[16] Dishant kharkar, "Unravelling the Power of XGBoost: Boosting Performance with Extreme Gradient Boosting." [Online]. Available: https://medium.com/@dishantkharkar9/unravelling-the-power-of-xgboost-boosting-performance-with-extreme-gradient-boosting-302e1c00e555

[17] *sagemaker-dg*.

[18] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Mar. 2016, doi: 10.1145/2939672.2939785.

[19] Jiri Kaiser, "Dealing with Missing Values in Data," *Journal of Systems Integration* , Jan. 2014.

[20] M. Jamshidian and M. Mata, "Advances in Analysis of Mean and Covariance Structure when Data are Incomplete," *Handbook of Latent Variable and Related Models*, pp. 21–44, Jan. 2007, doi: 10.1016/B978-044452044-9/50005-7.

[21] J. Granderson, S. Fernandes, E. Crowe, M. Sharma, D. Jump, and D. Johnson, "Accuracy of hourly energy predictions for demand flexibility applications," *Energy Build*, vol. 295, p. 113297, Sep. 2023, doi: 10.1016/J.ENBUILD.2023.113297.

[22] S. Gopal, K. Patro, and K. Kumar Sahu, "Normalization: A Preprocessing Stage." [Online]. Available: www.kiplinger.com,

[23] "sklearn.preprocessing.MinMaxScaler." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html

[24] T. Y. Kim and S. B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, Sep. 2019, doi: 10.1016/J.ENERGY.2019.05.230.

[25] Claude Sammut and Geoffrey I. Webb, "Mean Squared Error," *Encyclopedia of Machine Learning. Springer, Boston, MA.*, 2011.

[26] Claude Sammut and Geoffrey I. Webb, "Mean Absolute Error," *Encyclopedia of Machine Learning. Springer, Boston, MA.*, 2011.