Kiran Thatikonda, et. al.International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 14, Issue 2, February, 2024, pp: 45-52

RESEARCH ARTICLE OPEN ACCESS

Novel Power Demand Forecasting using Cutting Edge Machine Learning Models

Kiran Thatikonda *, Dr M V Kamal**
**(Industry Principal Director - Electric & Gas Utilities, Accenture)*
*** (Department of Computer Science & Engineering, MRCET Autonomous Campus, Hyderabad)*

ABSTRACT
Power consumption has an important impact on energy management, energy distribution costs, and environmental consequences. Traditional methods for forecasting power use, characterized by limited accuracy and scalability, are now surpassed by the advancements in machine learning techniques. Consequently, accurate predictions of power usage based on historical data have become feasible. However, given the crucial role of short-term forecasts in key functions performed by power operators like load dispatch, it is imperative for forecasting algorithms to deliver high accuracy predictions. A consumption estimating model is presented in this study with the goal of providing load estimates for a week or a month as well as an accurate forecast of the load for the upcoming 24 hours. This research also describes how to significantly improve prediction accuracy using an ensemble machine learning procedure. The study presents the results of the Random Forest and XGBoost models individually and collectively. The enhanced accuracy obtained by the RF and XGBoost ensemble ranged from 33% to 42%. Interesting findings on energy use are also obtained from these analyses.
Keywords - Energy distribution, Voltage, Power, Random Forest, XGBoost, and Machine learning

---

Date of Submission: 09-02-2024 Date of acceptance: 24-02-2024

I. INTRODUCTION

Electricity is expected to overtake other energy sources and become the standard choice for transportation, business, and residential needs in the near future [1]. This underscores the critical importance of accurately forecasting power consumption, given its substantial impact on various operational and business processes. In the field of electrical engineering, electricity demand is commonly denoted as load, and both terms will be used interchangeably here. Electricity has evolved into an integral component of daily life, transitioning into a fundamental necessity in contemporary society. The rapid surge in power consumption is attributed to the proliferation of intelligent devices, making energy an indispensable resource shaped by lifestyle changes. Consequently, the demand for electricity continues to grow in distinctive ways. To harness renewable energy sources, substantial investments are made in large-scale manufacturing facilities and energy plants. Following an uninterrupted energy flow, electricity serves as a cornerstone for advancing Machine Learning (ML) approaches. The widespread deployment of sensors and smart meters across the grid creates an environment that is conducive to the best possible use of these technologies [2].

Accurate estimation of the demand for power is essential in the field of intelligent energy management systems [3]. It is essential for both long-term planning for new generation and transmission facilities as well as short-term load allocation. Accurate projections facilitate educated choices on expenses and energy conservation. The growing prevalence of load forecasting applications underscores the significance of electricity demand prediction[4]. Some studies suggest that variables such as temperature, humidity, precipitation, or season might affect energy use [5], whereas other studies base their forecasts on demographic and socioeconomic variables [6]. While it is important to anticipate power usage using broad characteristics, improving the prediction model can yield better results.

Energy supply forecasting is critical to the electrical business because it informs choices about power system design and operation. Electrical companies use a variety of methodologies that are appropriate for short medium-, or long term projections for estimating power demand. Conventional forecasting methods are inadequate in this dynamic setting, requiring the use of more sophisticated approaches [7]. The goal is to identify the underlying issues and conduct a thorough analysis of every situation that makes change necessary. Even yet, it might be challenging to analyze a variety of social and private elements. A comprehensive review requires rich data and a variety of prediction techniques[8].

Extreme Gradient Boosting (XGBoost) and Random Forest (RF) machine learning ensembles are used to create the model in this work. This article's main goal is to classify and assess the most relevant material. This study's primary goal is to determine the best strategy, as well as the best input variables and parameter combinations, for various scenarios of electricity demand [9]. The study also explores other important areas related to machine learning, such as data pre-processing techniques, training and validation set selection, hyper-parameter modification for models, graphical displays, and results presenting.

II. LITERATURE STUDY

A statistical study of time series sequences is what forecasting involves, and it requires looking at a lot of different things, including past and future data. Since the load signal is a time series, a forecasting tool is needed to estimate how it will evolve in the future using past observations and predictor factors that affect its trajectory. The goal of this review of the literature is to provide a thorough overview of machine learning methods used in load forecasting to estimate energy usage. The writers present the idea of load forecasting and discuss its importance in controlling the supply and demand of electricity in the opening paragraph. The research compares the performance of each approach in terms of precision, durability, processing complexity, and data needs, doing a thorough examination of each technique's benefits and drawbacks. The authors stress how these techniques might be used to predict electricity consumption. Additionally, the research examines current developments in load forecasting, such as the use of big data and cloud computing, the use of hybrid models that combine different machine learning techniques, and the integration of meteorological and climatic data [10].

The authors of the literature study, "Machine Learning Techniques for Electricity Consumption Prediction: A Review," hope to provide readers with a thorough understanding of machine learning methods for energy consumption forecasting. The first part of the review explains what predicted power consumption is and why it matters for energy management. The writers investigate several machine learning techniques to forecast electricity usage, including artificial neural networks (ANNs), decision trees, regression analysis, and support vector machines (SVMs). A detailed analysis of the advantages and disadvantages of each method is provided, along with a performance comparison based on accuracy, robustness, computational complexity, and data needs. The paper also covers the issues of predicting power consumption and how machine learning techniques may be used to help overcome them. It explores new breakthroughs in the field of power demand forecasting, such as the adoption of hybrid models that mix various machine learning techniques, the use of big data and cloud computing, and the integration of meteorological and climatic data [11].

"A Comprehensive Review of Machine Learning Techniques for Short-Term Load Forecasting," the title of the literature study, provides a thorough analysis of machine learning techniques used in short-term load forecasting. It presents the idea of short-term load forecasting and emphasizes how crucial it is to energy management. The authors examine many machine learning techniques used in short-term load forecasting, including artificial neural networks (ANNs), decision trees, deep learning, support vector regression, and support vector machines (SVRs). A comparative examination of various tactics is provided by the review, which takes into account

variables including data needs, processing complexity, accuracy, and robustness.

Scholars have created machine learning (ML) algorithms to increase forecast accuracy in comparison to conventional energy consumption methodologies [12], [13]. One popular method for estimating electricity use is regression analysis, which is based on the idea that there is a relationship between energy use and climatic factors including temperature, precipitation, air density, humidity, and wind speeds. In this section, two popular machine learning methods for energy consumption forecasting—RF and XGB—as well as their ensembles are assessed.

Random-forest model

In machine learning (ML), Random Forest (RF) models are thought of as different approaches to bagging techniques. Using decision trees as basis estimators improves the bagging estimator technique in RF. A random sample is taken from the training set using this procedure. RF trains each tree for optimal splits using only a subset of data, as opposed to bagging, where all trees have access to the whole set of features. When compared to bagging, the prediction effectiveness is enhanced because to the trees' independence. Furthermore, the procedure becomes more efficient since just a subset of characteristics are used to train each tree.

In contrast, bagged decision trees follow a greedy approach to split variables, aiming to minimize errors. Consequently, even in bagging, several structural similarities are retained, leading to strong correlations in predictions. An ensemble's performance is optimized when the sub-model predictions exhibit low or no correlation. In order to address this problem, Random Forest (RF) modifies the algorithm to train sub-trees and decrease correlation between their predictions. The learning method is limited to evaluating a random sample of features when selecting a split point. In order to produce a model decision tree for each subset, this procedure comprises bootstrapping, or producing a random subset within the sample, and picking a random feature set for the best split at each decision tree node. By aggregating projections from each decision tree and computing their average, the final forecasting is completed.

XGBoost

One well-known gradient boosting method for regression applications is called XGBoost. The method builds a series of decision trees and integrates their forecasts to minimize the discrepancy between the anticipated and observed values [14]. To counteract overfitting, the algorithm integrates regularization techniques and furnishes a metric for gauging the significance of features. The XGBoost procedure involves data splitting, model initialization, model training and evaluation, hyperparameter tuning, and prediction generation for new data. The implementation of XGBoost via the gradient boost decision tree algorithm showcases its remarkable efficiency as a machine learning method. Notably, XGBoost exhibits high prediction accuracy, and its runtime is approximately 10 times faster than traditional gradient-boost methods.

An novel model that assesses the mistakes of the prior model is developed by the XGBoost ensemble. It uses a gradient descent method in conjunction with these mistakes to reduce loss and produce an overall forecast. XGBoost further demonstrates versatility in forecasting variable instability, offering advantages such as multithreaded parallelism for swift predictions, especially with substantial time-series data, surpassing the speed of other prevalent ensembles. Additionally, the availability of L1 and L2 regularization functions, coupled with the convenience of not requiring data normalization in the tree structure model, enhances XGBoost's proficiency. The algorithm also exhibits adeptness in handling missing data.

III. PROBLEM STATEMENT

Anticipating electricity consumption stands as a pivotal challenge in energy management. In the realm of effective energy management, precise forecasting of electricity consumption holds paramount importance, facilitating energy suppliers in optimizing energy distribution, curbing energy wastage, and preventing power system overloads [15]. The limitations in accuracy and scalability associated with conventional techniques for predicting power usage necessitate the exploration of more reliable and efficient forecasting methods.

Thus, the goal of this research is to provide a machine learning-based strategy for precise and effective power use prediction [16].

The proposed method needs to exhibit capabilities such as handling large datasets, dealing with missing values & outliers, and extracting relevant characteristics from the data. Additionally, the approach should possess the ability to discern the optimal model for accurate power usage anticipation. Evaluating the effectiveness of the suggested approach in forecasting power usage requires the utilization of diverse assessment indicators. This project's main goal is to advance energy management by presenting a comprehensive and effective approach to power consumption prediction.

IV. MODEL INPUTS

The Pearson correlation heatmap serves as a tool to identify the most influential parameters from historical data that impact future energy consumption. Interestingly, there are significant similarities in the amounts of power consumed at the same hour on days prior, especially when it comes to predictions made 24 hours ahead of time. As a result, models for 1-day, 1-week, and 1-month projections for every hour of the day are created. Consequently, the input of the model consists of the consumption loads for each day at a certain hour.

Furthermore, specific parameters, including two peak hours of a day (Hour_x and Hour_y), Day_of_week, Day_of_year, and peak seasons of the year (Day_of_year_x and Day_of_year_y), are chosen as additional inputs. An additional method for guaranteeing independence among variables is the feature significance plot, which is calculated using the F-score. This analysis shows that several factors, like Month and Day_of_week, which have weak correlations, have low F-scores and may be safely ruled out. High-scoring parameters are usually used as model inputs to make the system resistant to unimportant factors. A thorough explanation of the Pearson correlation heatmap and feature significance is given in the following sections.

Selection of model and implementation

Predicting a state's electrical load is crucial for effective power management and waste minimization. Given the inherent challenges posed by noise and unpredictability, achieving accurate load predictions can be a daunting task. In this study, we present a predictive model for electricity load utilizing a machine learning ensemble known as RF-XGBoost. The choice of the XGBoost regressor is motivated by its superior speed and robust performance in supervised learning for electricity usage prediction.

Several alternative approaches, including SVM, neural network, lone RF, and boosting, were considered in the prediction process. However, the comparative analysis revealed that XGBoost consistently outperformed these methods in terms of accuracy. The selection of the XGBoost regressor was further justified by its capability to simultaneously forecast future values for multiple variables and effectively model-nonlinear relationships within the datas structures.

Initially, the dataset is split using a fivefold cross-validation (CV) method into training and test sets. Ensemble techniques frequently use fast algorithms like decision trees. A type of decision tree ensemble known as Random Forests (RF) increases the variance of base models by bagging them with random subspaces (CART), which improves the model's performance. The suggested model builds many RFs using the training set, which are then improved by XGBoost training. Gradient-boosted decision trees, which perform exceptionally well on training data but run the risk of overfitting because of their flexibility, are typically trained using XGBoost. For this reason, XGBoost is used to train RF in this study. In order to reduce overfitting, RF randomly chooses data points for the tree and takes random feature subsets into account while dividing nodes. The best split method and the ensuing loss minimization follow well-established guidelines. To achieve the optimal split, XGBoost is used, which supports the precise greedy algorithm. Certain settings are setup for XGBoost training: booster to 'gbtree','subsample', and 'colsample_bynode', both set to 0.8, and learning rate (eta) to 1. "num_parallel_tree" set to 100 and

"max_depth" set to 5 are preferred. Num_boost_round is fixed at 1 to prevent boosting multiple random forests. Using a tenfold cross-validation (CV) strategy—a widely used method for assessing model performance—the model is further validated. The studies are conducted on a Windows GPU platform with an Intel Core i5-4790 CPU (2.6 GHz, 8 GB RAM) in a Python environment. An overview of the entire prediction model is shown in Figure 1.

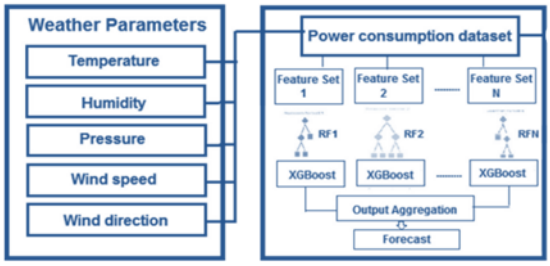

Fig 1. The overall prediction of model.

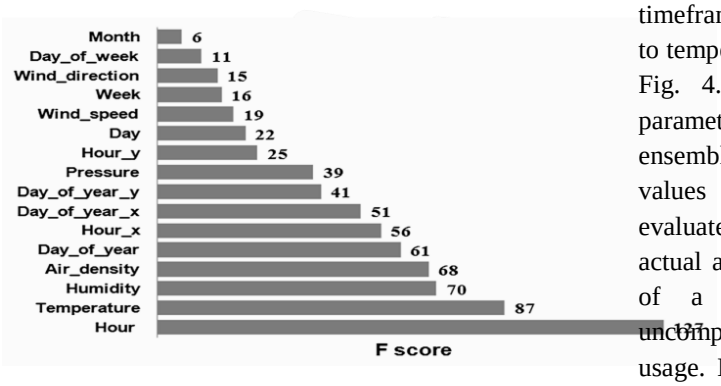

Fig. 2 Parameter feature significance graph with f-score.

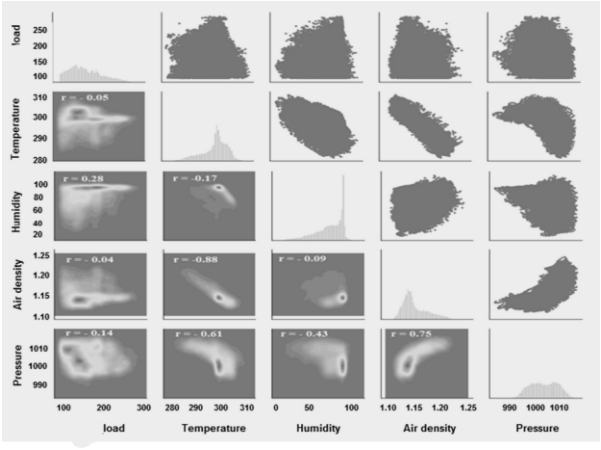

Fig 3. Bivariate density matrix.

V. RESULTS AND DISCUSSIONS

By giving input parameters in a prediction model scores that represent the relative importance of each variable for prediction, a technique known as the feature importance graph is created. As shown in Fig. 2, this feature significance is essential for finding pertinent information in the dataset and assisting in the identification of the most important features. Based on the presented heatmap and feature significance graph, attributes that have little effect on energy use are not included in the training process. Load is chosen as an attribute in a bivariate correlation that is created by the correlation heatmap, which shows relationships between several parameters. Fig. 3 shows the bivariate matrix for the dataset's primary characteristics. A strong linear link between energy consumption & temperature, humidity, air density, as well as pressure is predicted by this correlation diagram. Looking at the most linked aspects, load is shown over several timeframes (mean hourly and weekly) in connection to temperature, humidity, air density, and pressure in Fig. 4. The model is trained on the chosen parameters using the suggested RF-XGBoost ensemble and the previously described parameter values once the variable correlations have been evaluated. Figure 5 shows the contrast of the model's actual and anticipated values throughout the course of a single day's 24-hour projection. An uncomplicated method is utilized to forecast energy usage. In order to identify significant relationships shown in the graph, the target variable, "load," was extracted from the characteristic's matrix, which included Temperature, Humidity, Air density, Pressure, Hour, etc. After that, the model was subjected to a fivefold cross-validation process in order to forecast future energy use at various time horizons. Figure 6 shows a visual comparison of the actual and expected load values across three different time horizons: one day, one week, and one month.

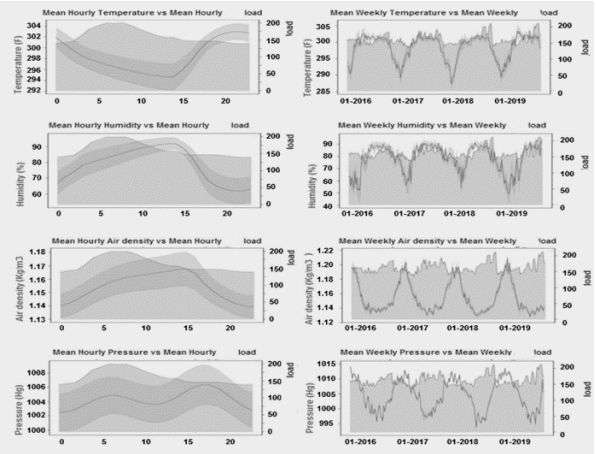

Fig. 4 Load graph on different timelines with temperature, humidity, air density, & pressure.

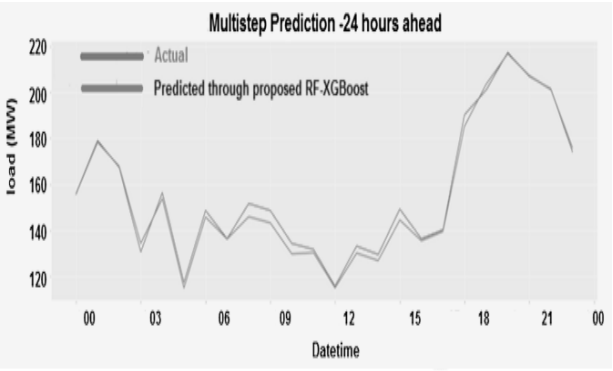

Fig. 5 Forecast estimate for a single day.

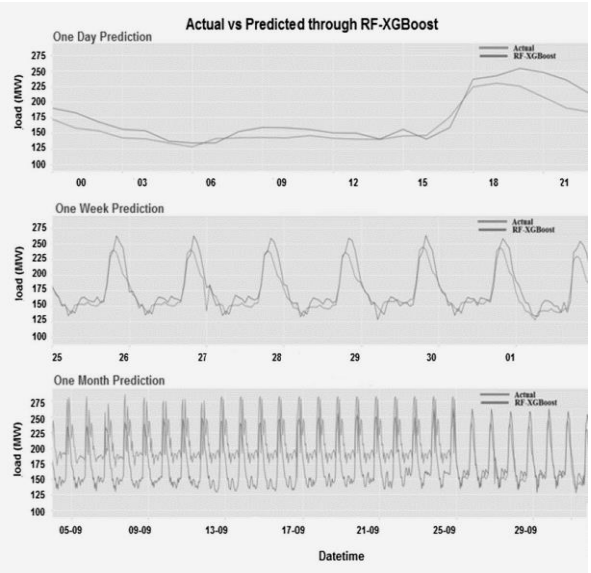

Fig. 6 Actual and expected values for one day, one week, and one month are compared.

The two most important measures are the root mean square error (RMSE) and the coefficient of determination (R2) that are used to evaluate the performance of the suggested model (RF_XGBoost) [48–50]. The percentage variance of forecast values, which fall between 0 and 1, is denoted by the symbol R2. It is provided by,

$$R^2 = 1 - \frac{\sum(\text{Actual} - \text{Predicted})^2}{\sum(\text{Actual} - \text{Mean\_actual})^2}$$

RMSE is the name of the variability test used to compare the actual and expected values of the model. It is defined as the square root of the mean squared error, given by,

$$\text{RMSE} = \sqrt{\sum \frac{(\text{Predicted} - \text{Actual})^2}{N}}.$$

Table 1 compares the various models for 24 hour, one week, and one month prediction in terms of R2 and RMSE. The comparative study revealed that the recommended RF-XGBoost ensembles outperformed other models for both the short-term (1 day–1 week) & long-term (1 month) power consumption projection. As a result, it seems that the recommended RF-XGBoost ensemble is a preferable option for estimations of electrical energy consumption over the short and long terms.

Table.1 Evaluation of the RF-XGBoost model's predictive power for electrical energy consumption.

Model	1 month prediction		1 week prediction	
	R2-score	RMSE	R2-score	RMSE
Random forest	0.456	3.308	0.47	3.202
XGBoost	0.622	2.803	0.602	2.775
RF-XGBoost (proposed)	0.703	1.886	0.768	1.221

Model	24-h prediction	
	R2-score	RMSE
Random forest	0.477	3.283
XGBoost	0.651	2.352
RF-XGBoost (proposed)	0.974	0.844

VI. CONCLUSION

This work introduces a novel ensemble model that combines the well-known machine learning techniques of XGBoost and RF to anticipate total electrical energy usage. Important characteristics that directly affect the amount of power consumed, such as temperature, humidity, air density, and pressure, are extensively investigated by the model. This demonstrates the usefulness and promise of the model in the current estimation of total electrical energy usage. Furthermore, the study highlights the significant improvement in accuracy achieved by analyzing data based on the time of day, as evidenced by the test results. The research pioneers the innovative use of RF in conjunction with XGBoost, deviating from conventional decision trees. With an impressive R2 score of 0.974, indicating high accuracy, this approach proves successful in providing valuable insights into power consumption prediction. The findings draw several key conclusions about the RF-XGBoost ensemble: (1) Its efficiency surpasses that of individual structures or similar methods. (2) The model may be used to create forecasts for a range of time periods, including the short, medium, and long terms. Ongoing investigations involve the integration of renewable power sources and data accumulation to enhance accuracy. Additionally, exploration of alternative machine learning ensembles is underway, with a focus on the mentioned areas of interest.

REFERENCES

[1] Y. Tian, J. Yu, and A. Zhao, "Predictive model of energy consumption for office building by using improved GWO-BP," Energy Reports, vol. 6, pp. 620–627, Nov. 2020, doi: 10.1016/j.egyr.2020.03.003.

[2] G. V. Reddy, L. J. Aitha, Ch. Poojitha, A. N. Shreya, D. K. Reddy, and G. Sai. Meghana, "Electricity Consumption Prediction Using Machine Learning," E3S Web of Conferences, vol. 391, p. 01048, Jun. 2023, doi: 10.1051/e3sconf/202339101048.

[3] S. G. Yoo and H.-Á. Myriam, "Predicting residential electricity consumption using neural networks: A case study," J Phys Conf Ser, vol. 1072, p. 012005, Aug. 2018, doi: 10.1088/1742-6596/1072/1/012005.

[4] H. Cai, S. Shen, Q. Lin, X. Li, and H. Xiao, "Predicting the Energy Consumption of Residential Buildings for Regional Electricity Supply-Side and Demand-Side Management," IEEE Access, vol. 7, pp. 30386–30397, 2019, doi: 10.1109/ACCESS.2019.2901257.

[5] G. K. F. Tso and K. K. W. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," Energy, vol. 32, no. 9, pp. 1761–1768, Sep. 2007, doi: 10.1016/j.energy.2006.11.010.

[6] Y. Liu and J. Wang, "Long Short-Term Memory Based Refined Load Prediction Utilizing Non Intrusive Load Monitoring," in 2021 IEEE Power & Energy Society General Meeting (PESGM), IEEE, Jul. 2021, pp. 1–5. doi: 10.1109/PESGM46819.2021.9637863.

[7] G. V. Reddy, L. J. Aitha, Ch. Poojitha, A. N. Shreya, D. K. Reddy, and G. Sai. Meghana, "Electricity Consumption Prediction Using Machine Learning," E3S Web of Conferences, vol. 391, p. 01048, Jun. 2023, doi: 10.1051/e3sconf/202339101048.

[8] Y. Liu et al., "Power data mining in smart grid environment," Journal of Intelligent & Fuzzy Systems, vol. 40, no. 2, pp. 3169–3175, Feb. 2021, doi: 10.3233/JIFS-189355.

[9] Y.-W. Su, "Residential electricity demand in Taiwan: Consumption behavior and rebound effect," Energy Policy, vol. 124, pp. 36–45, Jan. 2019, doi: 10.1016/j.enpol.2018.09.009.

[10] P. C. Albuquerque, D. O. Cajueiro, and M. D. C. Rossi, "Machine learning models for forecasting power electricity consumption using a high dimensional dataset," Expert Syst Appl, vol. 187, p. 115917, Jan. 2022, doi: 10.1016/j.eswa.2021.115917.

[11] J.-S. Chou and D.-S. Tran, "Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders," Energy, vol. 165, pp.

709–726, Dec. 2018, doi: 10.1016/j.energy.2018.09.144.

[12] A. Mosavi and A. Bahmani, "Energy consumption prediction using machine learning; a review," 2019, doi: 10.20944/preprints201903.0131.v1.

[13] E. García-Martín, C. F. Rodrigues, G. Riley, and H. Grahn, "Estimation of energy consumption in machine learning," J Parallel Distrib Comput, vol. 134, pp. 75–88, Dec. 2019, doi: 10.1016/j.jpdc.2019.07.007.

[14] T. Chen and C. Guestrin, "XGBoost," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[15] N. Amral, C. S. Ozveren, and D. King, "Short term load forecasting using Multiple Linear Regression," in 2007 42nd International Universities Power Engineering Conference, IEEE, Sep. 2007, pp. 1192–1198. doi: 10.1109/UPEC.2007.4469121.

[16] G. Gross and F. D. Galiana, "Short-term load forecasting," Proceedings of the IEEE, vol. 75, no. 12, pp. 1558–1573, 1987, doi: 10.1109/PROC.1987.13927.