Hoendarto et al. Energy Informatics (2025) 8:87
https://doi.org/10.1186/s42162-025-00540-6

Energy Informatics

RESEARCH

Open Access

Bridging IoT devices and machine learning for predicting power consumption: case study universitas Widya Dharma Pontianak

Genrawan Hoendarto[1,2], Ahmad Saikhu[1*] and Raden Venantius Hari Ginardi[3]

*Correspondence:
Ahmad Saikhu
saikhu@if.its.ac.id
[1]Faculty of Intelligent Electrical and Informatics Technology, Department of Computer Science, Institute Technology of Sepuluh Nopember Surabaya, Surabaya 60111, Indonesia
[2]Faculty of Information Technology, Universitas Widya Dharma Pontianak, Pontianak 78117, Indonesia
[3]Faculty of Intelligent Electrical and Informatics Technology, Department of Information Technology, Institute Technology of Sepuluh Nopember Surabaya, Surabaya 60111, Indonesia

Abstract
Multiple methods have been developed and implemented to reduce dependence on fossil fuels and conserve electricity. However, accurately predicting electricity consumption is essential before reducing it. Forecasting building electricity consumption has become increasingly critical, as buildings account for 39% of global electricity consumption. Among these, campus buildings are particularly energy-intensive. In this study, we used Monte Carlo (MC) simulations—trained on each leaf that generated by the regression tree (RT) algorithm—to predict the electricity consumption of Widya Dharma University Pontianak (UWDP)'s campus building. Unlike traditional approaches that rely on the mean of samples within a leaf, our method incorporates their likelihood. Since RT algorithms are prone to overfitting, training each leaf individually is expected to mitigate this issue. The data were collected by measuring hourly electricity consumption on one floor of the UWDP campus building over several months. The proposed MCRT prediction algorithm achieved an accuracy of 91.61%, with a Root Mean Square Error of 3.49 and a Normalized Root Mean Square Error of 0.09.

Introduction

Structural monitoring has become easier in the modern era due to the application of the Internet of Things (IoT), which is capable of capturing various types of data related to structures [1–4]. For example, the use of these devices allows users to track energy consumption in a building at specific instances without incurring high maintenance costs [5–7]. Mathematical models are then derivable from this data to create predictive models that accurately estimate the overall energy consumption of a given building. This approach is commonly known as the data-driven technique [8].

The data-driven technique involves developing a model of a specific structure based on the data obtained from it [9–13]. In contrast to the model-based technique, the data-driven model is built by leveraging the structure's data, thus eliminating the need for a complex understanding of the system. This ultimately lowers the cost of generating the model [14, 15]. One commonly used data-driven technique is the regression tree (RT), which is based on decision tree learning (DTL) [16].

The regression tree (RT) algorithm is a machine learning technique that constructs a binary decision tree to perform regression analysis [17, 18]. The core idea behind the algorithm is to recursively partition a large dataset into smaller, more homogeneous subsets by selecting the optimal split points that minimize the variance within each subset (or leaf node) [19, 20]. By ensuring that each leaf node contains data points with minimal variance, the algorithm enhances the homogeneity of the target variable, leading to more accurate predictions [21]. This approach is particularly useful for handling complex datasets, as it allows the model to capture underlying patterns without requiring an in-depth understanding of the mathematical model [22]. Moreover, by reducing the variance at each step, the RT effectively mitigates the nonlinearity of the dataset [23]. This algorithm has been shown to successfully predict energy consumption in buildings, as demonstrated in [24–31].

Other deep learning techniques were not tested in this study because our suggested approach is more straightforward and adequate for use on devices with daily specifications.

Related works. The authors in [25] utilized the random forest (RF) algorithm, which is a collection of regression trees (RTs), to forecast energy demand in a building located at Pontianak, Indonesia. To enhance the predictive quality of the model, the authors combined the RF model with poly-exponential correction terms. Based on their experiment, they demonstrated that the combination of poly-exponential correction and RF achieved an accuracy of 96% in predicting the building's energy consumption. Another research was conducted to model energy consumption in residential buildings, focusing on the heating load (HL) and cooling load (CL) to optimize the energy consumption of heating, ventilation, and air conditioning (HVAC) systems [26, 32, 33]. In their study, the researchers deployed a decision tree model combined with least squares boosting to model energy consumption. Their results showed that the combined methods effectively decreased the root mean square error (RMSE) by 37–68% for HL and 30–41% for CL, respectively. The authors in [27, 28, 34] researched modeling the daily electrical load from the Electric Generating Authority of Thailand (EGAT) by leveraging RT and deep belief network (DBN). Their findings indicated that the proposed method outperformed other models, including long short-term memory (LSTM), artificial neural network (ANN), support vector machine (SVM), and logistic regression (LR), in modeling the electrical load.

The authors in [28] applied the RT-based approach to study residential energy consumption during the COVID-19 quarantine mandates in Arizona, United States. They found that hourly daytime and weather conditions significantly affected household power consumption. Furthermore, they demonstrated that the RT-based approach performed well in modeling residential power consumption in Arizona. The author in [29] combined the RT-based approach with principal component analysis (PCA)-based feature selection to model buildings' power consumption using weather data. Their research showed that the proposed method was able to successfully model the energy consumption dynamics of four different datasets. Finally, the authors in [30] applied RT-based methods to perform a short-term energy consumption forecast in the city of Agartala, Tripura, India. They found that the RT-based model could accurately predict daily energy consumption and even provide estimations for weekly and monthly energy consumption. In summary, previous research has demonstrated the potential of the RT-based approach to model the energy consumption of buildings. In particular, the RT-based algorithm has been combined with methods such as PCA, DBN, or even a poly-exponential (PE) model to improve the predictive quality of the model. However, none of these studies directly address issues with the base RT model itself, where the RT model may be susceptible to extreme datasets or fail to perform well with scarce datasets [35, 36].

Contributions. As previously shown, the RT-based method performs well in modeling the power consumption of a building. However, it has some limitations. In particular, it is prone to overfitting and does not perform well with extreme or scarce datasets. This paper focuses on enhancing energy consumption modeling in buildings by incorporating IoT devices and the RT model. Specifically, this paper aims to improve the base RT model by integrating a Monte Carlo simulation (MC) into each RT leaf, rather than simply combining it with another method. The proposed methodology is called the MCRT algorithm. In the MCRT algorithm, the prediction of the RT model is no longer taken as the average of the data corresponding to a specific leaf, but as a random process drawn from the data contained within the leaf. This approach makes the output of the RT model more robust than the classical RT model. Furthermore, the model generated by this method will be applied to the building, optimizing its energy consumption based on the power consumption scheme generated by the MCRT algorithm. In this way, it is possible to optimize the energy consumption inside a building. This paper offered 3 contributions as follows:

1. We showed the derivation of the MCRT algorithm based on the power consumption dataset obtained from IoT devices.
2. We compared the model predictive quality of the classical RT algorithm and MCRT algorithm and showed that the MCRT performed better than the RT algorithm.
3. We validated the methodology on the dataset obtained from the IoT devices installed on the 4th floor of the Fransiskus Assisi Building of Universitas Widya Dharma Pontianak. Furthermore, we will show that the electrical energy plan generated by the MCRT algorithm can optimize the usage of electrical energy inside the building.

For the first contribution, we will apply the obtained results to an IoT device to control power consumption. This approach has been explored in several studies, including [37, 38], and [39]. The author in [37] designed an IoT model to enhance the results of the LSTM approach for forecasting electricity consumption using a CT (SCT-013030) sensor and a NodeMCU ESP8266 microcontroller in real-time, with the ThingSpeak cloud platform for data processing. In [38], a model connected to the Internet of Things (IoT) was introduced to reduce electricity costs by increasing solar energy consumption and extending the lifespan of solar panels via optimization. The model uses a lithium-ion battery for energy storage and a thermal cooling system to regulate the panel temperature. Data collected from sensors that monitor residency levels, temperature, humidity, energy, and thermostat settings are sent to a cloud-based platform using IoT. The authors of [39] presented a cutting-edge Android software for all-inclusive smart building management that enables users to use IoT devices to monitor and control data in real-time. By acting as a bridge, the software helps users understand the operational complexities of smart building systems. Five distinct layers make up the suggested system architecture: an Android app for user interaction, a Flask-based Application Programming Interface (API) for data analysis and conversion, JSON files for data storage, Python scripts for sensor data creation, and SQLite for effective data storage. Throughout the smart building management process, this multi-layered architecture guarantees smooth integration and reliable functionality.

The paper is organized as follows: In Sect."Monte Carlo (MC) simulation on Regression Tree (RT)", we will derive the MCRT model and discuss a specific case where this algorithm is applied to our case study. In Sect."Problem formulation", we will describe the dataset used to validate the methodology presented in this paper, as well as the metrics used to assess the predictive quality of the model. The research findings and outcomes will be presented and discussed in Sect."Experiment setup". We will analyze the results in the context of the conditions of the research object in Sect."Case study: UWDP campus". Finally, Sect."Model predictive quality parameters"will provide the conclusion of this research and outline directions for future studies.

Monte Carlo (MC) simulation on Regression Tree (RT)

In this section, an electrical energy consumption prediction algorithm based on the Monte Carlo (MC) algorithm combined with Regression Tree (RT) is discussed. Specifically, for each leaf of the binary tree formed using the RT algorithm, an MC model is inserted into it, to produce more accurate predictions. Figure 1 shows the proposed system with the objective of research on the campus building of the Widya Dharma University Pontianak (UWDP), Indonesia. Readers are encouraged to read [12] and [36] for the basic notion of the MCRT algorithm, which was designed specifically for the sparse dataset.

Problem formulation

Let $D \in \mathbb{R}^{m\times n}$ be a dataset corresponding to the measurement of the electrical properties of a building. Furthermore, suppose that D consists of 2 parts, such as

$$D = [D_t \, D_p]\,,$$

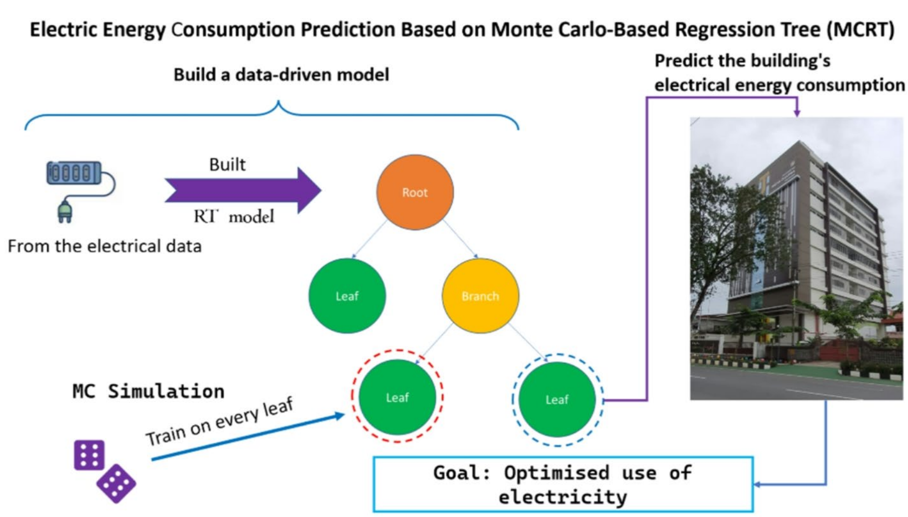


Fig. 1 A schematic of the proposed MCRT algorithm

$$D_t \in \mathbb{Z}^{m_1 \times n};$$

$$D_p \in \mathbb{Z}^{m_2 \times n};$$

$$m_1 + m_2 = m$$

where D_t denotes the time information of the dataset D and D_p is the dataset related to the measurement of the electrical power consumption in a building. The goal of this algorithm is to estimate the power consumption, $p \in \mathbb{R}^n; p \in D_p$, such that it is possible to estimate the power consumption at any time instance, k. This goal is achievable by following these 3 steps: Generate RT models, assign the MC model to each leaf of RT's, and define the upper and lower bounds for the prediction.

Step 1: Generate RT Models. To create an accurate power consumption model, the first step is to create the RT model for each necessary electrical parameter. In this step, according to the research conducted in [36], it is sufficient to build the model according to the 3-phase electrical current's information. Let $\boldsymbol{I}_1, \boldsymbol{I}_2$ and $\boldsymbol{I}_3$ be vectors for measuring the electrical current, where $\boldsymbol{I}_1, \boldsymbol{I}_2$ and $\boldsymbol{I}_3 \in \mathbb{R}^{\boldsymbol{n}}; [\boldsymbol{I}_1\ \boldsymbol{I}_2\ \boldsymbol{I}_3] \in D_p$. Then the RT model for all 3-phase of current can be expressed as

$$\begin{aligned} \widehat{I_1}(k) &= f_{RT}\left(\boldsymbol{x_t}(k), I_1(k-1), I_1(k-2), \ldots I_1(k-p)\right), \\ \widehat{I_2}(k) &= f_{RT}\left(\boldsymbol{x_t}(k), I_2(k-1), I_2(k-2), \ldots I_2(k-p)\right), \\ \widehat{I_3}(k) &= f_{RT}\left(\boldsymbol{x_t}(k), I_3(k-1), , I_3(k-2), \ldots I_3(k-p)\right). \end{aligned} \tag{1}$$

where, p denotes the order of auto-regressive (AR) terms, $\boldsymbol{x_t} = [x_{t1}, x_{t2}, \ldots x_{tm_1}]$ denotes the vector of time measurement at instance k, with x_{ti} is the $i-th$ time information (i.e. day of week, day of month, month, etc.), $\widehat{I_j}(k)$ denotes the prediction of the $j-\mathrm{th}$ electrical current at k.

Given Eq. 1, it is possible to express the model for power consumption can be expressed as follows:

$$\widehat{p}(k) = f_{RT}\left(\boldsymbol{x_t}(k), \widehat{I_1}(k), \widehat{I_2}(k), \widehat{I_3}(k)\right). \tag{2}$$

In particular, the model denoted by (2) is used as the main model for power consumption prediction. In the next step, an MC model is assigned to each leaf of the model to enhance the prediction accuracy.

Step 2: Assign MC models. The goal of this step is to assign an MC model to each leaf of the RT expressed in Eq. 2. Let T_p be the tree of the RT model for the power consumption as in Eq. 2 and $L = [L_1, L_2, \ldots L_l]$ be a collection ofleaves corresponding to the tree T_p. In addition, let $\boldsymbol{p_{L_i}}$ be the vector that contains the value of $\boldsymbol{p}$ contained in the leaf L_i. For each leaf, a random variable $\mathrm{R_L}$ can be assigned to it as in (3).

$$R_{L_i} = \begin{cases} 1 \text{ if } \boldsymbol{p_{L_i}} < \mathrm{q_{L_i}}(1) \\ 2 \text{ if } \mathrm{q_{L_i}}(1) \leq \boldsymbol{p_{L_i}} < \mathrm{q_{L_i}}(2) \\ 3 \text{ if } \mathrm{q_{L_i}}(2) \leq \boldsymbol{p_{L_i}} < \mathrm{q_{L_i}}(3) \\ 4 \text{ if } \boldsymbol{p_{L_i}} \geq \mathrm{q_{L_i}}(3) \end{cases}. \tag{3}$$

With $\boldsymbol{q_{L_i}} = [q_{L_i}(1)\ q_{L_i}(2)\ q_{L_i}\ \text{Eq. }(3)\ ]$ denotes a vector that contains the value of 1 st, 2nd (also known as median) and 3rd quartile of $\boldsymbol{p_{L_i}}$. Let,

$$\mathbb{P}_{\boldsymbol{L_i}} = [P_r(R_{L_i} = 1)\ \dots\ P_r(R_{L_i} = 4)] \tag{4}$$

be a vector denoting the probability of the random variable R_{L_i} and

$$\bar{\boldsymbol{p}}_{\boldsymbol{L_i}} = \left[\bar{p}_{L_i}(1)\ \dots\ \bar{p}_{L_i}(4)\right] \tag{5}$$

be a vector containing the average of $\boldsymbol{p}$ for each sample that satisfies $R_{L_i} = 1{,}2, 3$ and 4. Given Eqs. 3–5, it is possible to estimate the value of $\widehat{p}(k)$. For a random vector $x_r \in [0{,}1]^{n_r}$, let

$$g(j) = \begin{cases} \boldsymbol{b_{L_i}}(1)\ \text{if}\ \mathrm{x_r}(\mathrm{j}) < \mathrm{P_r}(\mathrm{R_{L_i}} = 1), \\ \boldsymbol{b_{L_i}}(2)\ \text{if}\ \mathrm{x_r}(\mathrm{j}) \in \left[\mathrm{P_r}(\mathrm{R_{L_i}} = 1), \sum_{\mathrm{i}=1}^{2} \mathrm{P_r}(\mathrm{R_{L_i}} = \mathrm{i})\right], \\ \boldsymbol{b_{L_i}}(3)\ \text{if}\ \mathrm{x_r}(\mathrm{j}) \in \left[\sum_{\mathrm{i}=1}^{2} \mathrm{P_r}(\mathrm{R_{L_i}} = \mathrm{i}), \sum_{\mathrm{i}=1}^{3} \mathrm{P_r}(\mathrm{R_{L_i}} = \mathrm{i})\right], \\ \boldsymbol{b_{L_i}}(4)\ \text{if}\ \mathrm{x_r}(\mathrm{j}) \geq \sum_{\mathrm{i}=1}^{3} \mathrm{P_r}(\mathrm{R_{L_i}} = \mathrm{i}). \end{cases}$$

Then, the estimate for $p(k)$ is defined as

$$\widehat{p}(k) = \tfrac{1}{n_r} \sum_{j=1}^{n_r} g(j). \tag{6}$$

Equation 6 gives the estimated power consumption for each time instance. However, this estimation is not sufficient. In particular, it is possible to assign upper and lower bounds to estimate, the maximum and minimum consumption for that day.

Step 3. Let $\epsilon = p - \widehat{p}; \epsilon \in \mathbb{R}^n$, such that ϵ denotes the residual between the estimate and real values. Let $E(\epsilon)$ be a metric of error measurement that depends on ϵ (i.e., the Root Mean Square Error or the Normalized Root Mean Square Error). The upper and lower bounds of the estimate in (6) can then be expressed as follows:

$$U_b(k) = \widehat{p}(k) + \widehat{p}(k) \cdot E(\epsilon),$$

$$U_b(k) = \widehat{p}(k) - \widehat{p}(k) \cdot E(\epsilon). \tag{7}$$

Figure 2 shows an example of the prediction generated by Eqs. 6 and 7. In this case, the NRMSE is used to represent the value $E(\epsilon)$. In particular, the bounds provide relaxation toward the prediction by, providing a range of values to the prediction.

Experiment setup

In this section, we will discuss the data used in this study, specifically focusing on how the model is generated from the UWDP dataset. Next, we will present the model's predictive quality parameters, which include the root mean square error (RMSE) and the Normalized Root Mean Square Error (NRMSE). Finally, we will present the circuit diagram for the IoT devices, with particular emphasis on how the circuit is installed in the classroom.

Case study: UWDP campus

This study's dataset was gathered by monitoring the power usage of classes on the fourth floor of UWDP's Fransiskus Asisi Building. This floor was chosen as a sample because the classrooms are consistently occupied with lecture sessions, making it ideal for

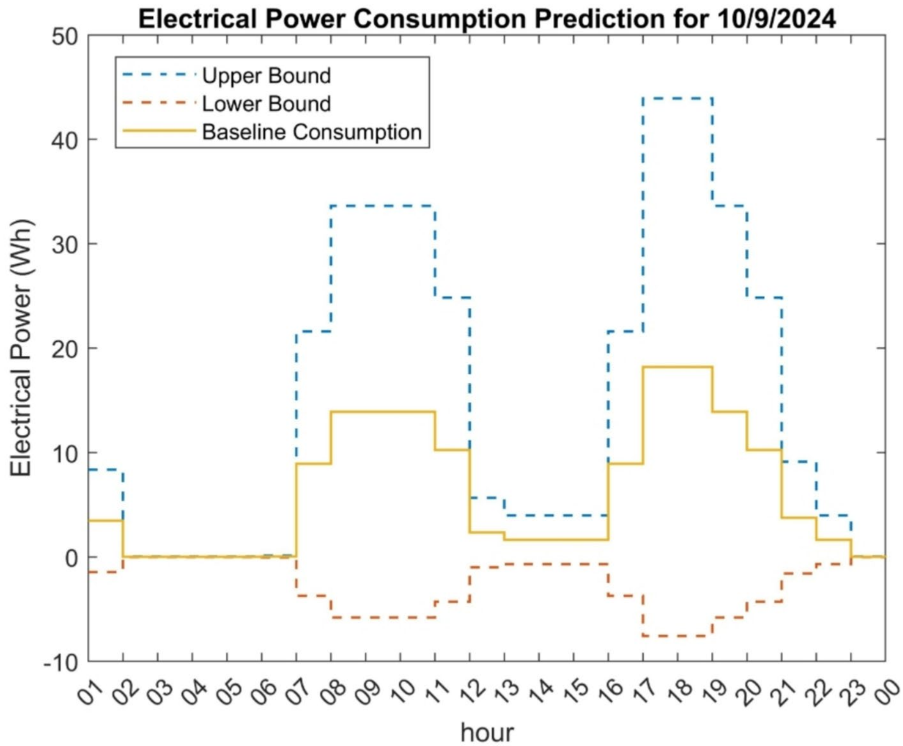


Fig. 2 Prediction of power consumption for 10 th September 2024. The proposed algorithm can generate bounds to provide minimum and maximum electrical power consumption

observation. The hourly power consumption was measured by installing a clamp power energy monitor on the three-phase currents in the control box, as shown in Fig. 3.

The dataset consists of 1,571 hourly samples, spanning from 11 th March 2024 to 15 th May 2024. The data includes five electrical parameters: three phases of currents, voltage, and the total current used in each power instance. In this study, 786 samples (50% of the total data) were used to train the MCRT model, while the remaining samples were used to validate the model's predictive quality. Additionally, the AR order for the model was set to 2. This choice was based on results from trial and error simulations, as well as the consideration that the power consumption of the current instance might depend on previous instances. We will also simulate other algorithms, such as the RT, Monte Carlo (MC) simulation, random forest (RF), and gradient boosting (GB), since it has been demonstrated that these techniques can accurately model a building's energy use in earlier studies.

Model predictive quality parameters

The suggested model's prediction accuracy is evaluated using the RMSE and NRMSE. The RMSE denotes the square root of the quadratic residual. For a vector $\boldsymbol{y} = [y(1), y(2), \ldots\ y(n)]; \boldsymbol{y} \in \mathbb{R}^n$ and its estimate $\widehat{\boldsymbol{y}}$, i.e., $\boldsymbol{y}$ and $\widehat{\boldsymbol{y}}$ denote the actual and the predicted values, the RMSE for $\boldsymbol{y}$ and $\widehat{\boldsymbol{y}}$ is shown in (8).

$$RMSE(\boldsymbol{y}, \widehat{\boldsymbol{y}}) = \sqrt{\sum_{i=1}^{n} \frac{\left(y(i) - \widehat{y}(i)\right)^2}{n}}. \tag{8}$$

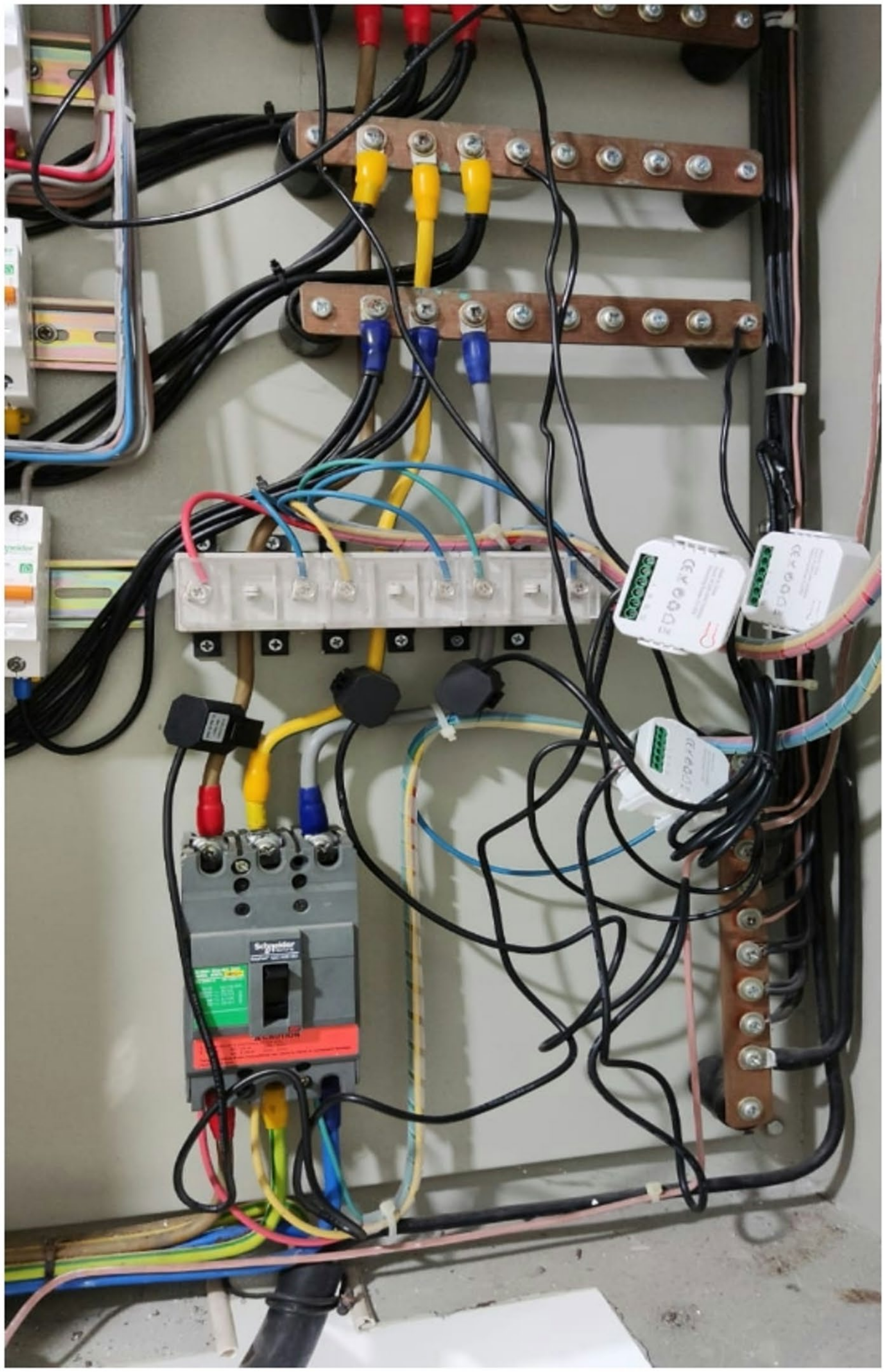


Fig. 3 The clamp power energy monitors installed in the 4 th floor control box

Specifically, the smaller the RMSE, the more similar the estimate is to the true value. In contrast, NRMSE is the normalized version of RMSE. Suppose that $\bar{\boldsymbol{y}} \approx 0$. Then, the NRMSE for $\boldsymbol{y}$ and $\widehat{\boldsymbol{y}}$ can be expressed as:

$$NRMSE\left(\boldsymbol{y}, \widehat{\boldsymbol{y}}\right) = \frac{1}{\max(y) - \min(y)} \times \ RMSE\left(\boldsymbol{y}, \widehat{\boldsymbol{y}}\right). \quad (9)$$

Normally, the value of NRMSE ranges from 0 to 1 (in the case of a poor model, it is possible to exceed 1). Specifically, the closer it is to 0, the better the model. Thus, it is possible to define the accuracy of the model to be

$$\%A\left(\boldsymbol{y},\widehat{\boldsymbol{y}}\right)=\left(1-NRMSE\left(\boldsymbol{y},\widehat{\boldsymbol{y}}\right)\right)\times\ 100\%\,. \tag{10}$$

In particular, these metrics were chosen the following needs: the RMSE was considered to measure how far the prediction generated by the MCRT model deviates from the reality. While the NRMSE was used to understand the normalized error of the overall process. The $\%A$ instead was used as the opposite of the NRMSE, which explain the overall accuracy of the model. Specifically, the higher $\%A$ is, the better the model in predicting the consumption of a building. In this way, we will be able to understand how the model perform in general.

To enhance the analysis, we also plot the parameter importance estimate. This estimate is based on the sum of the node risk due to the split for each parameter divided by the number of branch nodes. For example, a parent node splits into 2 child nodes, based on the Gini Diversity Index (GDI), which is defined as follows:

$$G\left(i\right)=1-\sum_{j=1}^{n}p_i^2\left(j\right) \tag{11}$$

where j denotes the $j-th$ class inside the tree T, n is the total number of classes, and p_i is the probability of 2 different samples that are chosen randomly representing the same class.

The Gini index as in Eq. 11 was used to measure how often a randomly chosen sample woulf be incorrectly classified in a dataset (i.e., the lower the Gini index, the better the classification). From Eq. 11, the parameter importance estimate can then be calculated as the average of the node risk, i.e.,

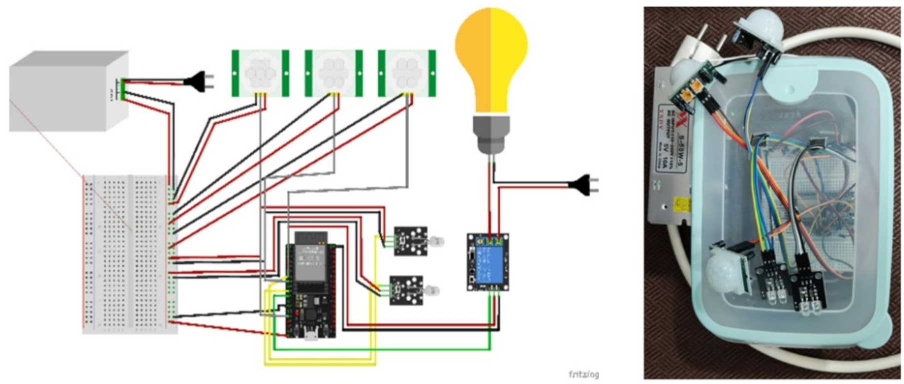
$$I=\frac{1}{N_b}\sum_{k=1}^{3}G\left(k\right) \tag{12}$$

where N_b indicates the number of branch nodes. In particular, the higher the importance estimate, the more crucial a parameter is. Thus, by calculating this estimate, it is possible to check whether a parameter affects the model's predictive quality. For example, it is possible to drop the non significant features from the model to make a simpler yet precise model.

Installation of automatic controlling devices

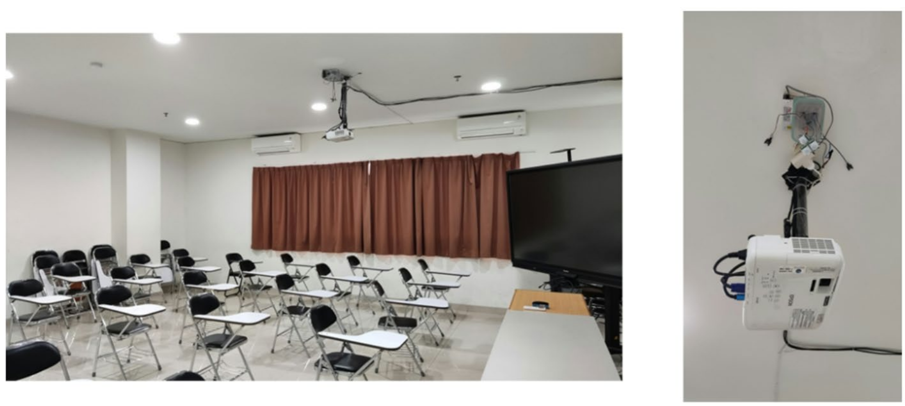
After the model derivation and validation process, the next step is to install an IoT device that will optimize the power consumption plan, as shown in Fig. 4. This device will control electrical appliances in the classroom, such as air conditioners and lights. Figure 4 illustrates the schematic of the automatic control devices installed in each room on the 4th floor of the Fransiskus Assisi Building. The device includes passive infrared (PIR) sensors to detect human presence in the room and infrared (IR) sensors to control the air conditioning. Specifically, this device is installed on top of the projector to minimize disruptions, such as those caused by human movement or distractions, to the sensors.

The control device will automatically turn on and off the air conditioner and lights based on the power consumption plan generated by the MCRT model. In particular, the microcontroller will pull the next hour power consumption from the database. If the



(a) (b)

Fig. 4 The circuit scheme of the proposed automatic control device: **(a)** circuit diagram **(b)** realization of the circuit



(a) (b)

Fig. 5 One of the classrooms located at the Fransiskus Asisi Building: **(a)** side view of the classroom **(b)** location where the device is installed (above the projector)

microcontroller detected that the power consumption at the next hour will be much higher than the current power consumption, then the microcontroller can turn on the air conditioner 15 min before the specific hour in order to maintain the comfortability inside the room. Additionally, when no one is present in the room, the PIR sensor will override the power consumption plan and turn off any lights or air conditioning that are currently on. To ensure accurate readings, three PIR sensors were assigned to each classroom; if any of the sensors detected movement, the electrical appliances in that classroom would remain on.

This IoT control device also allows for the manual, remote control of temperature and wind speed settings to ensure maximum comfort for students while studying. In particular, the administrator can control the wind speed of all air conditioner manually for that specific floor without the need of visiting the classroom one by one. If the room needs to be used for extra classes outside the predefined power consumption plan, the automatic control system can be bypassed by manual control. Figure 5 shows the classroom on the 4th floor of the Fransiskus Assisi Building at UWDP.

In the application created for this purpose, there is an option to adjust the wind speed and temperature as per the remote. The settings will depend on the predefined preferences. we thank you for the great feedback to be able to set it according to the external weather conditions so that it will make the control system feel more dynamic.

Results

In this section, the numerical simulation results are presented. Specifically, we compare the predictive accuracy of the proposed algorithm with four different algorithms: the classical MC simulation, RT, RF, and GB. These methods were chosen because of 2 reasons: first, these methods have been proven can predict energy consumption accurately. In particular, the MC Simulation was considered since it is a stochastic approach, which is good for modeling randomness or uncertainty, (e.g., sudden usage of energy inside the classroom due to unscheduled class). The RT algorithm was chosen due to its easy interpretation. We considered the RF algorithm which is a collection of DT to generate a more robust model. Finally, the GB was chosen since it is an ensemble method that can handle non-linear relationships effectively. The second reason is since the MCRT algorithm is built from RT and MC, we need to understand how far the model improves by assigning the MC model to the leaf, compared to the classical model. In this study, 100 tree depths and 10 learning rates were used, as they showed the desired results. Next, we present the power consumption prediction we for April, when the IoT device had not yet been installed, and for May, after the device was installed in the classrooms at UWDP. The installation of IoT-based automatic control devices is expected to reduce wasteful electricity consumption by turning off lights and air conditioner when the room is not in use.

Model forecasting accuracy

Table 1 presents the model predictive accuracy of the MCRT, MC, RT, RF and GB algorithms in forecasting the power consumption of UWDP buildings. The results show that the MCRT algorithm outperforms the other methods, improving accuracy by approximately 4%. This demonstrates the MCRT algorithm's ability to forecast power consumption in a building. Moreover, it highlights that modifying how the output is generated from a decision tree enhances the forecasting accuracy of the tree-based algorithm. It is important to note that the classical RT approach uses the average of the leaf samples as the output, while in MCRT, the output is generated based on the probability of the samples in the leaf, leading to a more accurate model. The details of these findings will be discussed in the next section.

Figure 6 presents the parameter importance estimates based on model risk, specifically the impurity reduction in the Gini Index, and the number of branches within the tree. In particular, the higher the importance level, the more significant the parameter. From Fig. 6, it can be observed that the power consumption at the previous time step is the most important parameter for deriving this model, followed by the hour of the day

Table 1 Model forecasting accuracy of the MC, RT, RF-10, RF-100, GB, and MCRT algorithms

Algorithm	RMSE	NRMSE	Accuracy (%)
MC	6.08	0.14	85.39
RT	6.54	0.16	84.29
RF-10	8.05	0.19	80.66
RF-100	8.18	0.20	80.37
GB	6.63	0.16	84.07
MCRT	3.49	0.09	91.61

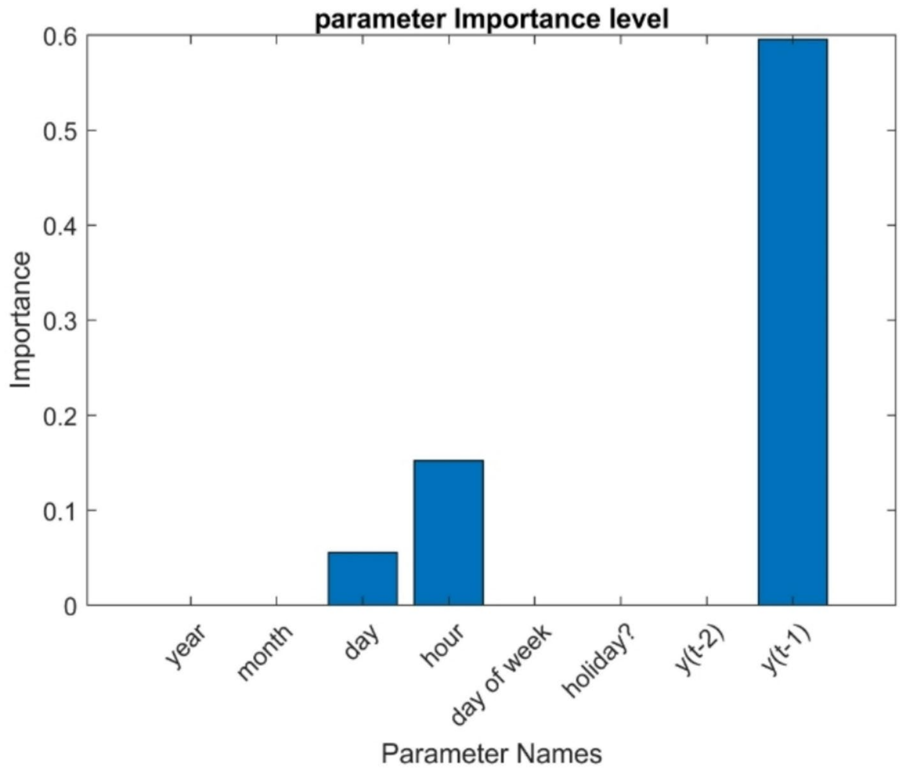


Fig. 6 Parameter importance level according to impurity reduction in Gini Index The power consumption at the previous moment strongly affected the model

Table 2 The two-sample t test with the assumption of unequal variances

	Before installation	**After installation**
Mean	1,914,558	1,696,441
Variance	4,500,544	1.3,435,412
Observations	26	17
Hypothesized mean difference	0	
df	23	
t Stat	2.2222	
P (T ≤ t) one-tail	0.0182	
t Critical one-tail	1.7139	
P (T ≤ t) two-tail	0.0364	
t Critical two-tail	2.0687	

and the day of the week. This result is expected, as human activities inside a building tend to vary daily. For instance, minimal activity occurs during midnight, while higher activity levels are observed during lecture sessions, from 8 a.m. to 12 p.m. and from 5 p.m. to 9 p.m., when the last session ends. The details of these findings are discussed in the next section.

The authors use the two-sample t test with the assumption of unequal variances is carried out to determine the statistical difference between the averages of the two groups, namely the groups before and after the installation of the automatic control device as shown in the Table 2.

With a significance level of 5% (α = 0.05), the two-way p value is 0.0364 which is smaller than 0.05. In addition, the calculated t value (2.222) is greater than the two-way critical t value (2.0687). Therefore, we reject the null hypothesis (H0) and accept the alternative hypothesis (H1). Thus, it can be concluded that there is a statistically significant difference between the group means before and after the installation of the automatic control device. Confidence interval (CI) of the difference of two means in the two-sample t-test with unequal variances (Welch's t-test) with 95% confidence level, the mean before treatment was greater than after treatment by between 1.5066 and 42.1168 units. Since the confidence interval does not include zero, this difference in means is statistically significant.

Monthly power consumption

From Sect."Monte Carlo (MC) simulation on Regression Tree (RT)", the MCRT algorithm outperforms both the MC and RT algorithms. Therefore, in this section, only the MCRT model will be used to estimate the power consumption UWDP building. Figure 7 illustrates the power consumption prediction from 13 th April 2024 until 15 th May 2024 for the MCRT and RT algorithm. Where from Fig. 7, we can see that the plot generated by the MCRT model closely resembles the observed value, which is evidence of the good prediction of power consumption. The advantage of the MCRT model over the RT model is that it is more robust.

Figure 8 depicts the power consumption plan for 4 th April 2024. To complete the information, we also added the actual power consumption for that specific date. We considered the data on this date because the schedule for all classrooms on the 4 th floor was fully occupied on this day. Therefore, from this figure, it is possible to identify how the actual power consumption deviates from the plan. From Fig. 8, it is evident that the power consumption stays within the bounds defined by the prediction. This demonstrates that the proposed algorithm can accurately predict the power consumption in one of the UWDP campus buildings. Notably, during the lunch break (afternoon), the power consumption was minimized, which is attributed to the fact that lectures at UWDP ran from 7:30 a.m. to 12 p.m., then resumed at 5 p.m. and continued until the last session ended at 9 p.m. Figure 8 shows that the proposed algorithm provides a reliable estimate of the building's activity patterns.

Figure 8 also compares the actual and predicted electricity consumption. It can be seen that there are some spikes in the actual electricity consumption, which occur when electrical equipment is turned on for the first time. This is followed by a pattern of low

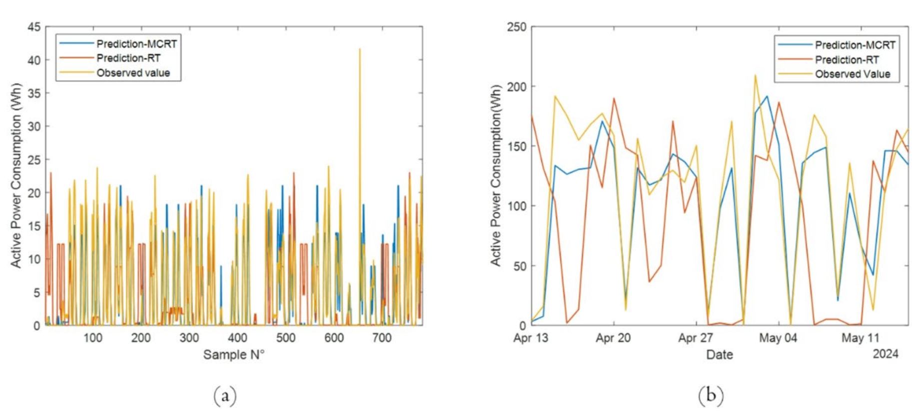


Fig. 7 Power consumption prediction for the Fransiskus Asisi Building during 13 th April until 15 th May 2024: (**a**) hourly (**b**) daily

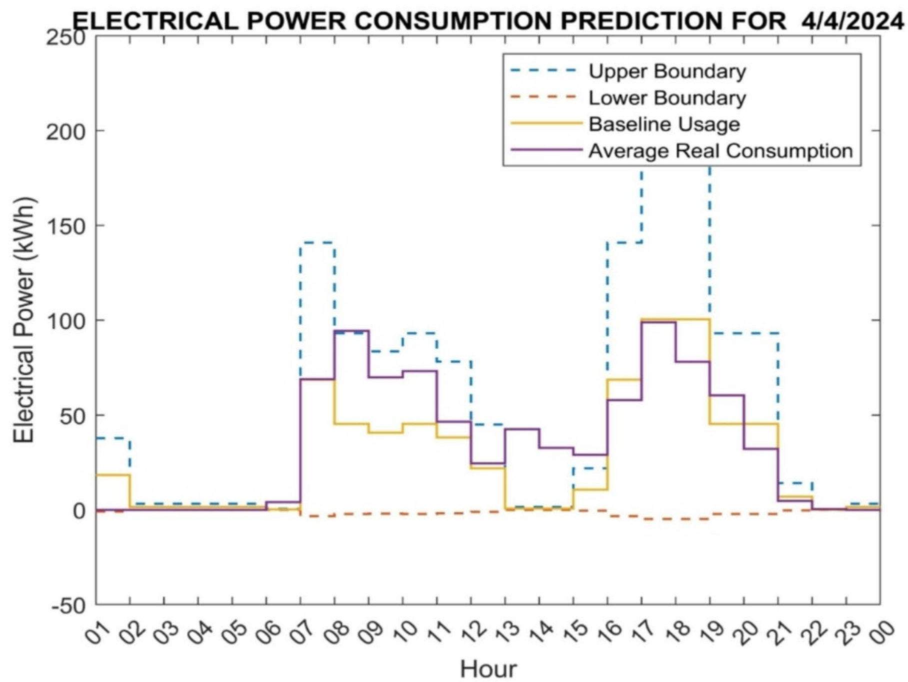


Fig. 8 Prediction of power consumption for 04/04/2024. The algorithm predicted the power consumption precisely as the real value is contained inside the bounds

and stable electricity consumption during hours when no lectures are held. Based on this data, we calibrated the IoT control device. Specifically, we programmed the device to automatically turn off all electrical appliances at 12 p.m. Furthermore, to ensure classroom comfort, the IoT devices were set to turn on the air conditioning 30 min before each lecture session. This schedule can be overridden if the PIR sensor detects that at least one person is present in the room.

Figure 9 shows a plot comparing the observed power consumption with the predictions made by the MCRT model. From Fig. 9, it is evident that the actual power consumption closely follows the pattern generated by the prediction, especially during the noon break when there is no excessive energy waste in the classrooms that are not in use. This result highlights the potential of combining IoT devices with machine learning to optimize a building's electric power consumption.

Discussion

As observed in the previous subsection, the MCRT algorithm significantly outperforms both the RT and MC models. This result emphasizes the potential benefits of integrating a Monte Carlo model into each leaf of the RT structure. Specifically, combining the RT and MC algorithms has substantially enhanced the model's predictive quality. This improvement is expected because assigning probabilities to the dataset enables the model to more effectively capture the relative importance of each data point. In particular, assigning a probability to each sample accounts for its likelihood of occurrence, which helps reduce the impact of abrupt changes caused by outliers. Without proper handling, outliers can disproportionately skew the results. By addressing this issue, the MCRT algorithm ensures that no single sample dominates the outcome, leading to

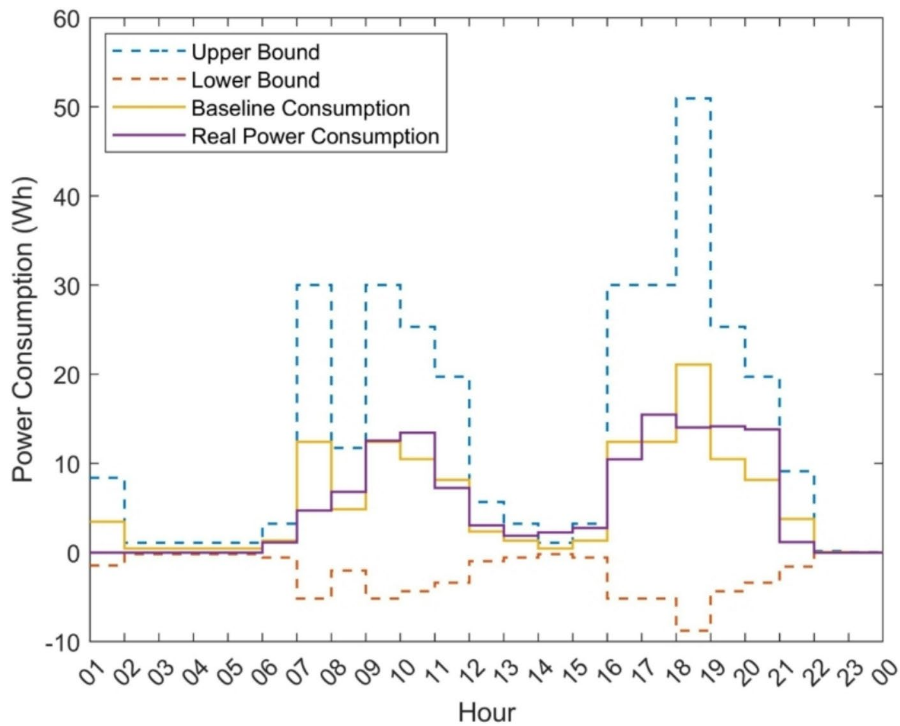


Fig. 9 Prediction of power consumption for 06/05/2024. The IOT control devices were able to optimize the power consumption according to the prediction

a more accurate, balanced, and robust predictive model. This makes MCRT especially valuable when dealing with complex data distributions or datasets prone to variability.

When considering parameter importance estimation, it is essential to contextualize the data used in this analysis. Dataset collected from campus buildings that are located in West Kalimantan Province, Indonesia, which has a unique geographical and climatic profile. Indonesia generally experiences two primary seasons: the rainy season and the dry season. While seasonal variations can often influence power consumption patterns in buildings, the situation in Pontianak, the capital city of West Kalimantan, is somewhat different. Pontianak lies directly on the equator, resulting in consistently warm temperatures throughout the year. As a result, the city is notably warmer than other regions of Indonesia, with an average annual temperature of around 29.6 °C.

The constant warmth in Pontianak necessitates the use of air conditioning systems year-round to maintain a comfortable indoor temperature, especially in environments like educational institutions, where a stable climate is crucial for productivity and comfort. Unlike regions with more distinct seasonal variations, where power consumption patterns typically align with weather fluctuations, Pontianak's equatorial climate results in a relatively uniform cooling demand throughout the year. As a consequence, there is no clear or predictable monthly power consumption pattern, as seen in Fig. 3. This uniformity underscores the need for predictive models capable of handling such steady yet dynamic environmental conditions. Understanding these contextual factors is key to interpreting parameter importance estimates and enhancing the model's applicability to real-world scenarios. Additionally, since there are no classes on weekends (i.e., Saturdays and Sundays), the day of the week becomes an important parameter in developing the model.

In Fig. 8, the estimated power consumption in the UWDP building is shown. Notably, the building experiences significantly low power usage (almost negligible, close to zero) during certain periods: from midnight to 7 a.m., during the lunch break between 12 p.m. and 4 p.m., and after 9 p.m. These fluctuations in power consumption are directly influenced by the building's usage patterns and the lecturing schedule. The first class at UWDP begins at 7:30 a.m. and runs until noon, after which classes resume in the evening from 5 p.m. to 9 p.m. Parameter importance estimates revealed that time of day and classroom occupancy were the most significant factors in predicting power consumption, underscoring the sensitivity of the model to human activity patterns and environmental conditions, especially the ambient temperature of campus buildings.

Figure 9 demonstrates that the predictions generated by the proposed algorithm remain accurate across different months. In this case, the predicted values closely align with the actual observed values, thanks to the IoT control devices. It is worth noting that students in Indonesia often arrive earlier than the scheduled class time as part of their daily routine. Although the first class officially starts at 7:30 a.m., many students arrive well before that, which triggers the IoT devices to turn on the air conditioning. This early use of air conditioning significantly contributes to the initial spike in power consumption in the early morning.

Although power consumption remained relatively low between 12 p.m. and 4 p.m., a closer inspection reveals that it was slightly higher than the consumption observed at midnight. This discrepancy can be attributed to students who stay in the building after morning classes. These students often use classrooms for self-study or participate in group activities, leading to moderate use of energy-consuming devices such as air conditioners and lights. While power consumption during the 12 p.m. to 4 p.m. period is slightly higher than at midnight, the system dynamically adjusts to minimize this excess, ensuring that energy usage stays within the predicted bounds.

Furthermore, during evening classes, the increased demand for artificial lighting correlates with higher power consumption. As natural light fades after sunset, classrooms and common areas depend on electrical lighting to maintain visibility and provide a conducive environment for learning. The combined energy demand, along with the continued use of air conditioners in the evening, results in a noticeable increase in power consumption compared to other times of the day.

Understanding these power consumption patterns underscores the strong correlation between human activity, environmental needs, and energy use within buildings. By analyzing these trends, we can identify opportunities to optimize energy usage, such as promoting energy-saving habits among students or implementing smart building technologies to regulate the use of high-energy devices like air conditioners and lights.

Conclusion and future research

This paper introduces a novel regression tree (RT)-based model, the MCRT model, which enhances the classical RT framework by incorporating a Monte Carlo (MC) model into each leaf of the RT structure. Experimental simulations conducted on a dataset from a university building in Indonesia demonstrated that the MCRT model improves the forecasting accuracy of traditional RT and MC models by up to 4%. This enhancement showcases the potential of the proposed algorithm in accurately forecasting power consumption.

The experimental results further highlighted that the most influential parameters for generating accurate predictions were the day of the week, the hour of the day, and the power consumption recorded one hour prior. These findings emphasize the MCRT model's ability to effectively capture temporal patterns and dependencies in power usage, making it a valuable tool for energy management in buildings.

Moreover, this result suggests the feasibility of using the MCRT model as a control mechanism to manage power consumption. For instance, the model could be integrated into a Supervisory Control and Data Acquisition (SCADA) system to optimize and regulate energy usage within a building. Additionally, the MCRT model could be employed to estimate monthly electricity bills with improved accuracy, offering practical applications in energy budgeting and cost management.

In future research, predicting electricity consumption will play a key role in reducing energy usage. The authors plans to conduct further studies by implementing Internet of Things (IoT) technology with a Wi-Fi-based microcontroller to regulate and optimize the electrical power consumption of air conditioning and lighting systems. This approach aims to improve energy efficiency and effectively lower electricity bills.

Acknowledgements
The authors declare that this research is not funded by any institutional or parties.

Author contributions
Genrawan Hoendarto: Prepare test data, Developed and evaluated the proposed method, Create automatic controller application and hardware prototype, Testing the efficiency after applying the automatic controller application software. Achmad Saikhu: Provide direction and supervision to research members, Set research targets, Supervise and evaluate the research performance of research members.Hari Ginardi: Provide supervision to research members, Provide technical assistance and input for the smooth running of the research.

Funding
The authors declare that this research is not funded by any institutional or parties.

Data availability
Readers or researchers interested in obtaining the dataset used in this study are welcome to get in touch with genrawan@widyadharma.ac.id.

Declarations

Ethics and consent to participate
Not applicable.

Consent to publish
The authors auhorize Springer Nature to publish this article widely should this article is accepted in the Energy Informatics Journal.

Competing interests
The authors declare no competing interests.

Received: 7 April 2025 / Accepted: 27 May 2025
Published online: 18 June 2025

References

1. Shah AS, Nasir H, Fayaz M, Lajis A, Shah A (2019) A review on energy consumption optimization techniques in IoT based smart Building environments. Information 10(3):108. https://doi.org/10.3390/info10030108
2. Goudarzi S, Anisi MH, Soleymani SA, Ayob M, Zeadally S (2021) An IoT-Based prediction technique for efficient energy consumption in buildings. IEEE Trans Green Commun Netw 5(4):2076–2088. https://doi.org/10.1109/TGCN.2021.3091388
3. Alijoyo FA (2024) AI-powered deep learning for sustainable industry 4.0 and internet of things: Enhancing energy management in smart buildings, *Alexandria Eng J* 104: 409–422. https://doi.org/10.1016/j.aej.2024.07.110
4. Chatlatanagulchai W, Chantrapornchai C (2024) Energy consumption data collection: case study on data center in a Thai university. Springer Nature. https://doi.org/10.1186/s42162-024-00327-1
5. Javed A, Larijani H, Wixted A (2018) Improving energy consumption of a commercial Building with IoT and machine learning. IT Prof 20(5):30–38. https://doi.org/10.1109/MITP.2018.053891335
6. Imran N, Iqbal, Kim DH (2022) IoT task management mechanism based on predictive optimization for efficient energy consumption in smart residential buildings. Energy Build 257:111762. https://doi.org/10.1016/j.enbuild.2021.111762

7. Moudgil V, Hewage K, Hussain SA, Sadiq R (2022) Integration of IoT in building energy infrastructure: A critical review on challenges and solutions, *Renewable and Sustainable Energy Reviews*, vol. 174, p. 113121. https://doi.org/10.1016/j.rser.2022.113121
8. Kathirgamanathan A, De Rosa M, Mangina E, Finn DP (2021) Data-driven predictive control for unlocking Building energy flexibility: A review. Renew Sustain Energy Rev 135:110120. https://doi.org/10.1016/j.rser.2020.110120
9. Bousdekis A, Lepenioti K, Apostolou D, Mentzas G (2021) A review of Data-Driven Decision-Making methods for industry 4.0 maintenance applications. Electron (Basel) 10(7):828. https://doi.org/10.3390/electronics10070828
10. Wan X, Cai X, Dai L (Dec. 2024) Prediction of Building HVAC energy consumption based on least squares support vector machines. Energy Inf 7(1). https://doi.org/10.1186/s42162-024-00417-0
11. Sarker IH, Perspective A (2021) SN Comput Sci 2(5):377. https://doi.org/10.1007/s42979-021-00765-8
12. Liang Z, Chen J (Dec. 2025) Research on Building energy consumption prediction algorithm based on customized deep learning model. Energy Inf 8(1). https://doi.org/10.1186/s42162-025-00483-y
13. Yesilyurt H, Dokuz Y, Dokuz AS (2024) Data-driven energy consumption prediction of a university office Building using machine learning algorithms. Energy 310:133242. https://doi.org/10.1016/j.energy.2024.133242. August
14. Ahmad T, Madonski R, Zhang D, Huang C, Mujeeb A (2022) Data-driven probabilistic machine learning in sustainable smart energy/smart energy systems: key developments, challenges, and future research opportunities in the context of smart grid paradigm. Renew Sustain Energy Rev 160:112128. https://doi.org/10.1016/j.rser.2022.112128
15. Andronie M, Lăzăroiu G, Iatagan M, Hurloiu I, Dijmărescu I (2021) Sustainable Cyber-Physical Production Systems in Big Data-Driven Smart Urban Economy: A Systematic Literature Review. Sustainability 13(2):751. https://doi.org/10.3390/su13020751.
16. Olu-Ajayi R, Alaka H, Sulaimon I, Sunmola F, Ajayi S (2022) Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. J Building Eng 45:103406. https://doi.org/10.1016/j.jobe.2021.103406
17. Breiman L, Friedman JH, Olshen RA, Stone CJ (2017) Classification and regression trees. Routledge. https://doi.org/10.1201/9781315139470
18. Nalenz M, Rodemann J, Augustin T (2024) Learning de-biased regression trees and forests from complex samples. Mach Learn 113(6):3379–3398. https://doi.org/10.1007/s10994-023-06439-1
19. Carrizosa E, Molero-Río C, Morales DR (2021) Mathematical optimization in classification and regression trees. TOP 29(1):5–33. https://doi.org/10.1007/s11750-021-00594-1
20. Klusowski JM, Tian PM (2024) Large scale prediction with decision trees. J Am Stat Assoc 119(545):525–537. https://doi.org/10.1080/01621459.2022.2126782
21. Smarra F, Tjen J, D'Innocenzo A (2022) Learning methods for structural damage detection via entropy-based sensors selection. https://doi.org/10.1002/rnc.6124
22. Bocci L, D'Urso P, Vicari D, Vitale V (2024) A regression Tree-Based analysis of the European regional competitiveness. Soc Indic Res 173(1):137–167. https://doi.org/10.1007/s11205-021-02869-3
23. Zhang H, Zhou A, Qian H, Zhang H (2022) PS-Tree: A piecewise symbolic regression tree. Swarm Evol Comput 71:101061. https://doi.org/10.1016/j.swevo.2022.101061
24. Gong J (2024) Building energy efficiency evaluation based on fusion weight method and grey clustering method. Energy Inf 7(1). https://doi.org/10.1186/s42162-024-00437-w
25. Panjaitan SKSD, Tjen J, Sanjaya BW (2023) A Forecasting Approach for IoT-Based Energy and Power Quality Monitoring in Buildings, *EEE Transactions on Automation Science and Engineering*, vol. 20, no. 2, pp. 892–900. https://ieeexplore.ieee.org/document/9771055
26. Ahn CW (2022) Regression tree ensemble learning-based prediction of the heating and cooling loads of residential buildings, *Build Simul*, vol. 15, pp. 2003–2017. https://doi.org/10.1007/s12273-022-0908-x%0A%0AKeywords
27. Phyo PP, Jeenanunta C (2021) Daily load forecasting based on a combination of classification and regression tree and deep belief network. IEEE Access 9:152226–152242. https://doi.org/10.1109/ACCESS.2021.3127211
28. Ku AL, Qiu Y, Lou J, Nock D, Xing B (2022) Changes in hourly electricity consumption under COVID mandates: A glance to future hourly residential power consumption pattern with remote work in Arizona, *Appl Energy*, vol. 310, p. 118539. https://doi.org/10.1016/j.apenergy.2022.118539
29. Parhizkar T, Rafieipour E, Parhizkar A (2021) Evaluation and improvement of energy consumption prediction models using principal component analysis based feature reduction. J Clean Prod 279:123866. https://doi.org/10.1016/j.jclepro.2020.123866
30. Rita Banik AB, Das P (2021) Prediction of electrical energy consumption based on machine learning technique, *Springer Nature Link*, vol. 103, pp. 909–920. https://link.springer.com/article/10.1007/s00202-020-01126-z
31. Ma Z (2022) Survey data on university students' experience of energy control, indoor comfort, and energy flexibility in campus buildings. Energy Inf 5(Suppl 4):50 https://doi.org/10.1186/s42162-022-00239-y
32. Bogin D, Dorman M, Kissinger M, Erell E (2024) Beyond efficiency: evaluation of the electricity saving potential of green residential buildings. Energy Build 320:114627. https://doi.org/10.1016/j.enbuild.2024.114627
33. Çiçek A (2024) A novel resilience-oriented energy management strategy for hydrogen-based green buildings. J Clean Prod 470:143297. https://doi.org/10.1016/j.jclepro.2024.143297
34. Panjaitan SD, Tjen T, Sanjaya BW, Wigyarianto FTP, Khouw S (2021) Daily Load Forecasting Based on a Combination of Classification and Regression Tree and Deep Belief Network, *IEEE Access*, vol. 15, no. 2, pp. 892–900. https://doi.org/10.1007/s00202-020-01126-z
35. Darmanto T, Tjen J, Hoendarto G (2024) Monte Carlo Simulation-Based Regression Tree Algorithm for Predicting Energy Consumption from Scarce Dataset, *Journal of Data Science and Intelligent Systems*, pp. 1–9. https://doi.org/10.47852/bonviewjdsis42022395
36. Hoendarto G, Saikhu A, Ginardi RVH (2024) Daily power consumption plan derivation via the Monte Carlo-Based regression tree algorithm. ACM Int Conf Proceeding Ser 404–408. https://doi.org/10.1145/3669754.3669816
37. Hasan MW (2024) Measurement: Energy Design of an IoT model for forecasting energy consumption of residential buildings based on improved long short-term memory (LSTM), *Measurement: Energy*, vol. 5, p. 100033, 2025. https://doi.org/10.1016/j.meaene.2024.100033

38. Umaeswari P, Sonia R, Saravanan TR, Poyyamozhi N (2024) IoT-Enabled energy conservation in residential buildings: machine learning models for analyzing annual solar power consumption. Sol Energy vol. 238, pp. 594–601. https://doi.org/10.1016/j.solener.2024.112890
39. Martins P, Ramos A, Pina E, Vaz P, Silva J, Abbasi M (2019) Smart Building Control: An Android Application for Enhanced Monitoring and Management in the Internet of Things Era, *Procedia Comput Sci*, vol. 238, pp. 594–601. https://doi.org/10.1016/j.procs.2024.06.066

Publisher's note