Optimizing Energy Efficiency in Data Centers Through Machine Learning Model

Md Shadman Sakib[1], Md Shorif Hossain[1], Ifterkhar Ahmed Shohan[1], Md Aahadul Islam Fardin[1], *Ahmed Wasif Reza[1], and Mehedi Masud[2]

[1]Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
[2]Computer Science, Taif University, KSA

Emails:
shadmansakib77@gmail.com,shorifhossen248@gmail.com,
popjb303@gmail.com,iaahadul@gmail.com,
wasif@ewubd.edu, mehediku@hotmail.com

*Corresponding Author

Abstract. The power consumption of data centers has risen sharply due to the increasing need for data processing and storage. Addressing this issue requires optimizing the energy efficiency of these assets through effective design and operation. This study aims to solve this problem by creating a machine learning-based method to forecast the energy consumption of future data centers. We collected a dataset of data center energy consumption and performance data. We applied machine learning models to this dataset to predict energy consumption under various operating conditions. Our results demonstrate that machine learning techniques can adequately project data centers and identify key factors affecting energy efficiency. These insights can inform the design and operation of future energy-efficient data centers, lowering environmental impact and energy use.

Keywords: Optimizing, Energy efficiency, Chiller, Machine Learning, PUE, Accuracy.

1 Introduction

A data center is a sophisticated facility that offers shared access to software and data through networks, computing, and storage resources. Standardized practices ensure the construction and maintenance of data centers for high availability and security. Centralizing a company's IT operations in a data center allows efficient storage, processing, and transfer of data. A data center is vital for daily operations and houses a company's valuable assets. In the digital economy, they provide essential infrastructure for services, but their energy-intensive operations contribute to global emissions. Managing rising energy demands and environmental concerns, including excess energy use, emissions, and resource depletion, calls for energy-efficient data center design and operation.

To meet this challenge, we propose a machine-learning method to model data center energy usage. Through analysis of complex datasets, machine learning finds patterns, which make accurate energy forecasts. Our approach uses data center performance data, aiming to predict energy consumption across varied operating conditions. This contributes to energy-efficient design, reducing carbon impact and fostering a sustainable digital economy.

2 Literature Review

In [1], different machine learning models (NN, LightGBM, RNN, Random Forest) were employed to compute PUE, with NN (Google) showing the best performance. In [2], a power model was developed to reduce energy costs and limit data centers' carbon impact. Its accuracy was confirmed, demonstrating effective brown energy reduction during maximum and non-peak hours.

In [3], developed an algorithm by combining the heuristic method and migration control, and evaluated their different algorithms using Cloudsim 3.0.3. In [4], They present a method for developing and deploying performance indicators and measuring instruments for calculating development tools' energy consumption, which should enable engineers to repeatedly measure and monitor their software's power usage while it is being developed.

In [5], He used a machine-learning algorithm to assess and develop efficient energy predictive performance. Because they regularly operate in large data centers, these algorithms consume significant energy in many countries. In [6], they used machine learning to calculate energy usage in multiple buildings to improve energy sustainability and efficiency. When estimating power consumption for houses one point ahead, the RF model surpassed the RT model by approximately 49.21% inside the MAE and 46.93% inside the utter and total error percentage (MAPE).

In [7], machine learning predicted energy-efficient plans for metal matrix modules and software, optimizing multi-core workloads. Core type, voltage, and frequency prediction improved energy efficiency. In [8], cluster computing energy-saving techniques were explored, with potential future avenues for efficiency enhancement discussed in the survey's conclusion.

In [9], they used machine learning to evaluate the economics of energy efficiency and renewable energy technologies. They approached AI, which can be used to overcome various challenges. The proposed approach could boost energy efficiency to 97.32% while growing the number of renewable power sources. In [10], artificial intelligence was applied to model heating and cooling demands for the building's energy-efficient design. SVR models outperformed all others in predicting CL and HL to absolute percentage mean errors of less than 4%. Furthermore, when compared to the previous efforts, the SVR model and ensemble method reduced the root mean square errors by at least 39.0% to 65.9% for CL and HL prediction.

In [11], the study investigates server virtualization's impact on energy use and throughput in data centers. While it consolidates applications for energy efficiency, its potential energy overhead and throughput reduction remain unclear. The research discovers a trade-off between energy efficiency from convergence and the harmful impacts of virtualization.

In [12], Leo Breiman's "Random Forests" introduces the algorithm, combining estimations from multiple decision trees in a machine-learning model. It discusses model generalization error and the convergence as tree count increases. The paper highlights feature importance measurement, error monitoring, and correlation

analysis. Applications of random forests across machine learning tasks are also explored.
In [13], Peter Bodik's "Automating Datacenter Operations Using Machine Learning" focuses on maintaining data center performance and availability. The paper emphasizes understanding user workloads, and promptly identifying and addressing performance issues. Statistical machine learning (SML) is proposed to efficiently process extensive monitoring data, detect patterns, and create accurate performance models for enhanced data center operations. In [14], The paper "Early predicting cooling loads for energy-efficient design in office buildings by machine learning" by Ngoc-Tri Ngo describes how to create structures that consume less energy. In order to forecast cooling loads for buildings with few common factors during the design phase, this study developed an alternative approach utilizing machine learning (ML). This paper also describes the entire power simulation based on physics.
In [15], machine learning was used to optimize the Routing Protocol for underwater sensor networks, aiming to prolong network lifetime with equitable energy distribution among nodes. This research focuses on node residual energy and its role in objective function calculation during routing. In [16], Cesar Benavente-Peces and Nisrine Ibadah utilized machine learning classifiers to analyze and classify building energy efficiency. The paper showcases tools and methodologies to assess building efficiency using ICT and data science, detailing machine learning's role in categorizing buildings based on energy efficiency.

3 Energy Efficiency of Data Centers

Data centers are massive energy purchasers, accounting for as much as 3% of all electricity used in the United States. As such, growing data center energy efficiency has become a significant focus for businesses and organizations responsible for their operation and maintenance. Many organizations have implemented various energy-saving strategies to reduce energy costs and environmental impact, including improved hardware and software design, improved cooling methods, and the use of renewable energy sources. Data centers can be made more efficient at the hardware level by using more energy-efficient processors, power supplies, and other components. Designing hardware for greater efficiency can reduce energy requirements and extend the life of the equipment.
Additionally, virtualization technologies can reduce energy usage by consolidating multiple applications and workloads onto a smaller number of servers. Data centers use significant amounts of energy to cool the equipment. However, improved cooling methods can help reduce this energy drain. For example, hot-aisle/cold-aisle configurations can be implemented to reduce cooling needs and eliminate hot spots within the data center. Additionally, natural cooling methods such as evaporative and liquid can reduce the energy required to cool the equipment.
The Power Usage Effectiveness (PUE) metric measures data center energy efficiency. Calculated by dividing the total electricity used by the IT equipment power, lower values signify higher efficiency, with a PUE of 1.0 denoting perfection. Aimed at 1.2 or below, PUE showcases operational efficiency and avenues for improvement, enabling operators to discern power-intensive segments

and compare with industry benchmarks. Cooling systems, accounting for 30 to 40 percent of energy consumption in data centers, can be optimized. Large enterprise centers target PUE values of 1.13 to 1.60, while mid-tier centers range from 1.79 to 1.88. The data center scale profoundly influences cooling energy use, and as the industry advances, continuous efforts are necessary to balance performance with energy efficiency.

4 Materials and Methods

4.1 Data Collection and Dataset

Measuring PUE requires identifying relevant features. Factors like chiller water flow, outside air, and chilled water temperatures are linked to PUE. However, getting the actual dataset to predict the PUE was complicated. That is why we measure those data on our own. Before data gathering, we established feature-PUE relationships. For example, higher air-wet bulb temps correlate with increased PUE. Python and its functions are used in data collection, using range and building libraries to generate desired random values. Data was then compiled into a CSV file, comprising 5 columns and 1000 rows. Our dataset includes 5000 cells with integer and continuous values.

In our dataset (Fig. 1), we've analyzed four key features directly affecting a data center's Power Usage Effectiveness (PUE). These include chiller water flow, cold water output temperature, and total IT load, all in relation to outside air bulb temperature. Chiller performance impacts data center temperature, so water flow and output temperature were considered. External air bulb temperature influences cooling, while total IT load reflects energy use. Analyzing these features provides insights into energy efficiency and highlights areas for enhancement.

	Condenser water flow of chillers	Outside air wet bulb temperature	Chilled water output temperature of chillers	Total server IT load	PUE
0	400.12	9.04	12.00	5999	1.2003
1	400.23	9.04	12.00	5997	1.2006
2	400.48	9.05	12.00	5996	1.2007
3	400.80	9.06	12.00	5994	1.2011
4	400.82	9.07	12.00	5991	1.2014
...	...	...	...	...	...
995	599.61	24.94	11.01	4007	1.3990
996	599.90	24.96	11.01	4006	1.3994
997	599.91	24.97	11.00	4006	1.3994
998	599.91	24.97	11.00	4004	1.3994
999	599.99	24.99	11.00	4004	1.3998

1000 rows × 5 columns

Fig.1. Dataset of our model.

4.2 Design and Progress Flowchart

We recommend optimizing the PUE model, as shown in Fig. 2. However, obtaining a physical or numerical model for the system is challenging, which makes variable selection and alteration complex. To optimize PUE, we employ machine learning methods. A data-driven optimization framework is created using extensive system data. Initially, data collection and feature engineering are pivotal. Variable identification and connection to PUE are done through feature engineering. Prediction models are built via data hyperparameters, cleaning, and validation. A simulation model predicts PUE based on input features, identifying adjustable variables. The system implements the optimal solution, effectiveness is assessed, and actions are adjusted accordingly. The optimization framework's process is shown in Fig.2.

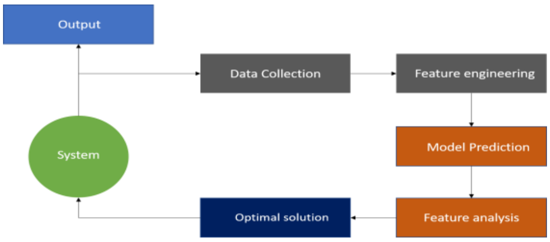


Fig.2. PUE optimization framework.

4.3 Prediction Model

Our study introduced various machine learning techniques, including neural networks, LightGBM, and recurrent neural networks (RNN), to predict data center PUE. Accurate PUE prediction is crucial for model performance analysis and sensitivity assessment. Neural networks and LightGBM models predict PUE for the next minute based on the current minute's features, while RNN considers current and past time steps, capturing temporal dependencies. IT load is an input but not the prediction target; PUE is. Our method isn't real-time and uses historical data for offline model creation. Thus it is suitable for data centers, aiding optimization techniques.

4.3.1 Random Forests

Random Regression Forest (RRF) is a learning technique that uses a collection of trees called decision or "forests" to make predictions. It works by building many decision trees using random subsets of data and features and then averaging the predictions of all trees to derive a final prediction, which helps reduce overfitting and increase model accuracy.

The equation for a decision tree in a Random Forest can be represented as:

$$f(x) = \Sigma\ (wj * hj(x) \qquad (1)$$

Here,

f(x) means the final prediction, wj is the weight of each decision tree, hj(x) is the decision tree function for the jth decision tree.

The decision tree function hj(x) is a function that takes an input x and outputs a class label (in classification) or a continuous value (in regression). The weights can be determined by various methods such as Gini impurity, information gain, or by fitting the model to the training data.

In a random forest, a random forest model will randomly choose a sample from the given dataset; after that, a decision tree will be created for every training data. After that, a voting decision tree will be chosen, and the most votes will be chosen. For a better understanding, Fig.3 is given below.

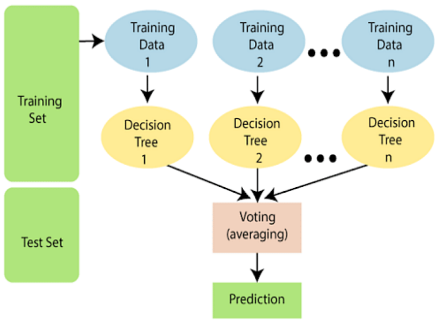


Fig.3. The working steps of Random Forest from Simplilearn.

In random forest regression, multiple decision trees are fitted on various subsets of our dataset, as shown in Fig. 3, followed by prediction averaging to improve accuracy and prevent overfitting. As decision trees can be built concurrently, this algorithm is efficient. Random forests, an ensemble learning technique, amalgamate predictions from diverse independent models for refined forecasts, with decision trees being the models in this case. Numerous decision trees are trained on distinct bootstrapped sub-samples of the training data. Their predictions are combined for the random forest's overall prediction. Additionally, random forests are relatively insensitive to overfitting, meaning they tend to generalize well to new data. Finally, random forests can be trained quickly because the individual decision trees can be trained in parallel, even on large datasets.

4.3.2 LightGBM

LightGBM is a popular machine learning algorithm that uses gradient boosting to train decision trees and reduce error. With tunable hyperparameters, it is efficient, accurate, and can handle large datasets. LightGBM is often faster and more accurate than gradient-boosting algorithms like XGBoost.

LightGBM is a structure for gradient boosting that employs tree-based learning algorithms. The equation for LightGBM's objective function is as follows:

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} |w_i| \qquad (2)$$

Where y is known as the true label for the i-th sample, ŷ is the predicted label for the i-th sample, and the sum is taken over all samples in the dataset. The "loss" term in the equation represents the loss function being used (such as mean squared error or log loss), and the "regularization" term represents any regularization applied to the model (such as L1 or L2 regularization).

4.3.3 XGBoost Regression

XGBoost, or Extreme Gradient Boosting, is an algorithm commonly used for regression tasks. A gradient-boosting algorithm builds a model by training a series of decision trees and combining their predictions to reduce the model's error. The algorithm is widely known for its effectiveness, fluidity, and scalability. The key feature of XGBoost is the tree-based model that allows it to make predictions based on input data. It also has a number of hyperparameters that can be fine-tuned to

improve performance, such as the learning rate and maximum depth of the trees, making it a powerful and widely used algorithm for regression tasks.
The equation for XGBoost for regression can be represented as:

$$y = f(x) = \Sigma\,(wj * hj\,(\,x\,)) \qquad (3)$$

where y is known as the predicted value, f(x) is known as the final prediction, wj is the weight of each decision tree, hj(x) is the decision tree function for the j-th decision tree.
The decision tree function hj(x) is a function that takes an input x and outputs a continuous value. The weights wj can be determined by various methods, such as least squares or gradient descent, by fitting the model to the training data.
The steps of the XGBoost algorithm are shown in Fig. 4. Firstly, it makes some decision trees based on features and then predicts from multiple decision trees based on the voting system. From bagging boosting algorithms, randomly make a forest or decision trees. Then minimize the error from the previous model or boost the model. After that, we get our final XGBoost model.

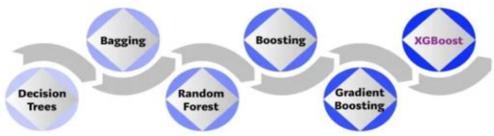


Fig.4. The structure of XGBoost Regression from toward data science

5 Model Performance

LightGBM, XGBoost Regression, and random forests are employed to construct the PUE predictive model. The samples are randomly divided into two groups with a 7:3 ratio, the training and testing sets. Utilizing three cross-validations in the training set, we develop and validate our models, lowering the mean absolute error metric. Finally, The performance of each model in the testing set is shown in Fig.5, with the results being an average of three measurements. The precise prediction can be attributed to the right selection of features and the fine-tuning of the hyperparameters.
Making the right feature choices provides the flexibility of utilizing simpler models with less computational work, which helps to produce more accurate predictions. We used features including water temperature, flow, chiller power, water pressure, and historical data that are directly or indirectly tied to PUE.

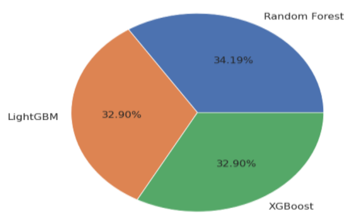


Fig.5. Prediction of success of various models.

In our proposed method for reducing power consumption efficiency, we have selected the random forest algorithm as the primary method, and it has demonstrated superior performance in comparison to all other supervised learning algorithms for PUE in Fig. 5.

6 Model Accuracy

Fig.6 was simulated via the Random Forest model as a graphical representation of the model's performance. The red line represents the measured values, while the green line represents the actual value. It describes how accurately our model predicts values. The mean square error (MSE) is 0.000162, and our r2 score is 95%.

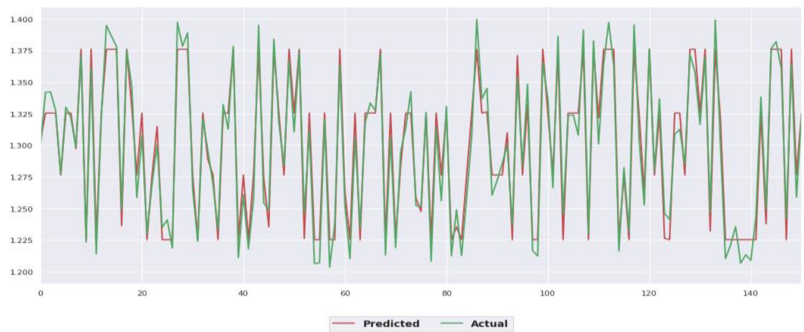
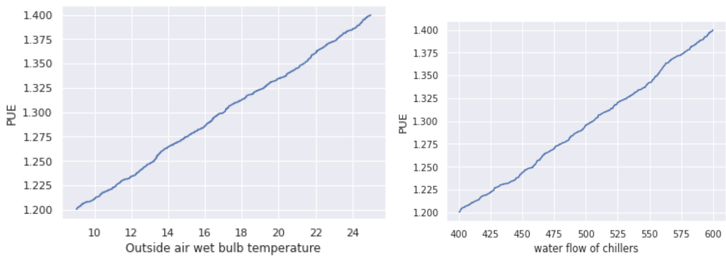


Fig.6. Accuracy of our proposed model.

7 Feature Analysis

Feature analysis in this study aimed to identify key factors impacting Power Usage Effectiveness (PUE) in data center cooling. While our method cannot establish a causal link between the features and PUE, the method offers insights into vital cooling aspects for lowering PUE. Feature importance scores interpret PUE prediction and highlight influential factors, aiding future research and optimization in cooling systems. Fig.7 illustrates PUE fluctuations with increased chiller condenser water flow, showcasing linear relationships with chilled water and outside air temperatures, plus server load.



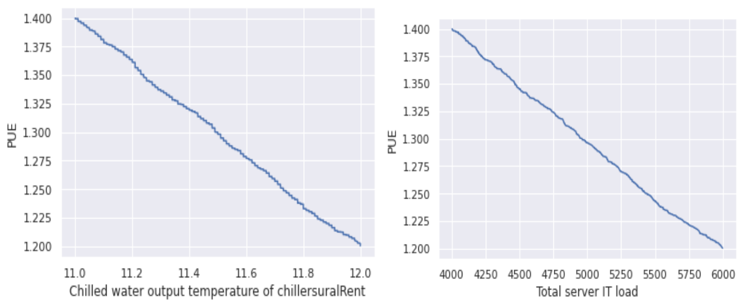


Fig.7. Some features were subjected to sensitivity testing.

So, we can optimize the PUE of a data center by reducing the water flow of the chiller and adjusting the total server IT load.

8 Effectiveness of PUE Optimization

Power Usage Effectiveness (PUE) is a key metric for assessing data center energy efficiency. Calculated by dividing total data center energy by IT equipment power, a PUE of 1.0 indicates optimal computing power usage. Higher PUE values indicate greater energy diversion to non-computing functions like cooling and lighting. Various strategies exist to optimize PUE. Enhancing cooling system efficiency via advanced equipment, insulation, and airflow management can lower the energy needed for temperature control. Optimizing power distribution and cable management reduces energy loss during transmission. Efficient IT equipment, including servers, storage, and networking devices, can reduce overall energy consumption. Virtualization, simulating hardware functions, enables multitasking within virtualized systems. PUE optimization effectiveness depends on specific measures and existing systems' efficiency. Implementing such strategies enhances data center energy utilization and overall efficiency.

9 Conclusion

This study delivered a machine learning-dependent approach to predict and enhance data center energy efficiency. Various models were developed and tested on consumption data, highlighting crucial factors and enabling power consumption prediction. Our findings show that machine learning techniques can reduce data center power efficiency and inform future planning and operation of energy-efficient data centers. While the results are promising, the scope for extended research and development remains. Further optimization of machine learning models and expansion to different data center types and conditions is possible. Findings can aid in energy-efficiency policies, promote greener practices, and underscore machine learning's role in advancing data center power efficiency for a low-carbon digital economy.

References

1. Yang Z, Du J, Lin Y, Du Z, Xia L, Zhao Q, Guan X. Increasing the energy efficiency of a data center based on machine learning. J Ind Ecol. 2022;26:323–335. (2021)
2. Bhattacharya, Tathagata and Xiao Qin. “Modeling Energy Efficiency of Future Green Data centers.” 2020 11th International Green and Sustainable Computing Workshops (IGSC) (2020): 1-3. 2020.
3. Monil, Mohammad Alaul Haque et al. “Incorporating Migration Control in VM Selection Strategies to Enhance Performance.” Int. J. Web Appl. 6 (2014): 135-151.(2015)
4. T. Johann, M. Dick, S. Naumann, and E. Kern, “How to measure energy-efficiency of software: Metrics and measurement results,” in Proceedings of the First International Workshop on Green and Sustainable Software, pp. 51–54, IEEE Press.(2012)
5. E. García Martín, ‘Energy Efficiency in Machine Learning: A position paper’, in 30th Annual Workshop of the Swedish Artificial Intelligence Society SAIS 2017, May 15–16, 2017, Karlskrona, Sweden, 2017, vol. 137, pp. 68–72.(2017)
6. Anh-Duc Pham, Ngoc-Tri Ngo, Thi Thu Ha Truong, Nhat-To Huynh, Ngoc-Son Truoung Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability,Journal of Cleaner Production,Volume 260,2020,121082,ISSN 0959-6526. (2020)
7. H. Sayadi, N. Patel, A. Sasan and H. Homayoun, "Machine Learning-Based Approaches for Energy-Efficiency Prediction and Scheduling in Composite Cores Architectures," 2017 IEEE International Conference on Computer Design (ICCD), 2017, pp. 129-136, doi: 10.1109/ICCD.2017.28. (2017)
8. Valentini, G.L., Lassonde, W., Khan, S.U. et al. An overview of energy efficiency techniques in cluster computing systems. Cluster Comput 16, 3–15.(2013)
9. Cheng Chen, Yuhan Hu, Marimuthu Karuppiah, Priyan Malarvizhi Kumar,Artificial intelligence on economic evaluation of energy efficiency and renewable energy technologies,Sustainable Energy Technologies and Assessments,Volume 47,2021,101358,ISSN 2213-1388.(2021)
10. Jui-Sheng Chou, Dac-Khuong Bui,Modeling heating and cooling loads by artificial intelligence for energy-efficient building design,Energy and Buildings,Volume 82,2014,Pages 437-446,ISSN 0378-7788.(2014)
11. Yichao Jin, Yonggang Wen, and Qinghua Chen. Energy efficiency and server virtualization in data centers: An empirical investigation. In 2012 Proceedings IEEE INFOCOM Workshops, pages 133–138. IEEE.(2012)
12. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.(2001)
13. Bodik, P. (2010). Automating datacenter operations using machine learning. (Ph.D. thesis). University of California.(2010)
14. Ngo, Ngoc-Tri. "Early predicting cooling loads for energy-efficient design in office buildings by machine learning." Energy and Buildings 182 (2019): 264-273. (2019)
15. Hu, Tiansi, and Yunsi Fei. "QELAR: A machine-learning-based adaptive routing protocol for energy-efficient and lifetime-extended underwater sensor networks." *IEEE* transactions on mobile computing, 9.6 (2010): 796-809. (2010)
16. Benavente-Peces, C., & Ibadah, N. (2020). Building energy efficiency analysis and classification using various machine learning techniques *Energies*, *13*(13), 3497. (2020)