International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V70I5P236

Volume 70 Issue 5, 328-338, May 2022

Original Article

Performance Analysis of Machine Learning Regression Techniques to Predict Data Center Power Usage Efficiency

Rajendra Kumar[1], Sunil Kumar Khatri[2], Mario José Diván[3]

[1]*Amity Institute of Information Technology, Amity University, Noida, Uttar Pradesh, India*
[2]*Director of Campus, Amity University Tashkent, Tashkent, Uzbekistan*
[3]*Data Science Research Group, Economy School, National University of La Pampa Santa Rosa, La Pampa, Argentina*

[1]rajendra.kumar1077@gmail.com

Received: 19 March 2022 Revised: 16 May 2022 Accepted: 25 May 2022 Published: 31 May 2022

Abstract - *Data Centers & cloud hosting services are critical for IT workload. Datacenter organizations need to equip them with the latest technologies to estimate the power usage efficiency (PUE) to cater to their hosting customers' requirements. Power usage efficiency is one of the major metrics to check how efficiently Data Center consumes their power. To better understand whether machine learning technology can forecast PUE with more accuracy, we have used multiple machine learning regression methods to predict the PUE in a data center and compared their accuracy. The research's originality resides in the fact that no previous research has examined the regression methods for PUE prediction in data centers. Once the accuracies are identified, future researchers can use the algorithm for effective PUE prediction. The experimental result shows that DT and KNN work effectively with the data center's PUE data in the research scope. Further, the analysis clearly shows that the Decision tree and KNN predict the PUE with 97% & 98% accuracy, respectively, compared with other regression techniques.*

Keywords - *Cooling, data center, Machine learning, Optimisation, Power usage efficiency*.

1. Introduction

Cloud computing and Data Center hosting has grown over recent years as a significant technology supporting most enterprises' organizations that need their data hosting requirement. Due to the growth of data stored in data centers due to application needs, new bigger data centers are required to satisfy user expectations. In data centers, a lot of energy is consumed. Data Centers are expected to account for one-third of the world's total energy output and one-fifth of its total consumption by 2025, with a carbon footprint of 5.5% of the total energy output. [1].

Because of the high demand for cloud services, Data Centers have to operate to provide application availability to their end-users continuously. As a result, they use a significant quantity of energy and usually have an adverse effect on the local power system. Additionally, energy costs have become a significant part of Data Center running expenditures [2].

Because of numerous variables such as the hardware specifications, workload, cooling needs, application kinds, etc., the DC power consumption pattern cannot be accurately determined. Therefore, accurately modeling a DC's power consumption behavior is not as simple as one would think [3]. Data centers are expected to be the center of all Internet activity. Additional data centers will be needed in the future to sustain the global computer penetration expansion. On the other side, data centers consume a significant amount of energy due to their role as information infrastructure. For example, a typical data center may use the same amount of energy as 25,000 families and consume around 150 times the energy of a typical office of equal size [4]. Since power is now a major cost in modern data centers, the cost of running a standard data center increases each year. [5]. On the other hand, data center electricity use has many environmental issues. In 2010, data center energy usage was 1.5% of total worldwide electricity consumption and around 2% of total electricity consumption in the USA [6]. Data centers will become the biggest energy consumers globally, increasing to 4.5% in 2025 [7]. Energy usage has become one of the most significant considerations when deciding where to locate a data center.

Moreover, because the monitoring infrastructure integrates an overhead system, it is impossible to carry out comprehensive energy consumption assessments of all existing components [3], [8]. Therefore, power forecasting techniques are designed which could estimate a data center's electricity consumption for a certain workload. Power Usage Efficiency (PUE) is a measuring tool that defines how effectively energy is utilized by a computer data center, especially how much energy is consumed by equipment. To improve data center energy efficiency, SPUE has been instrumental. Fig. 1 and 2 show the data center's historical and average PUE. [8], [9].

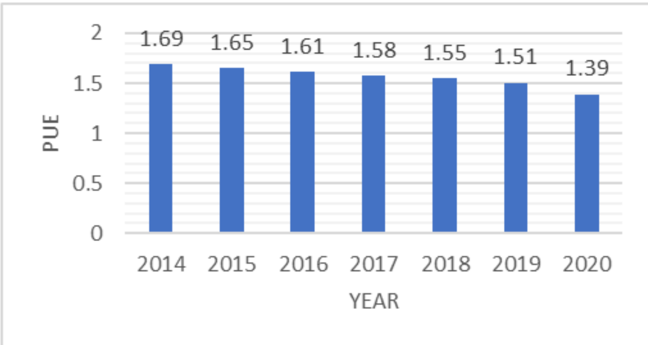

Fig. 1 Historical PUE data of a data center

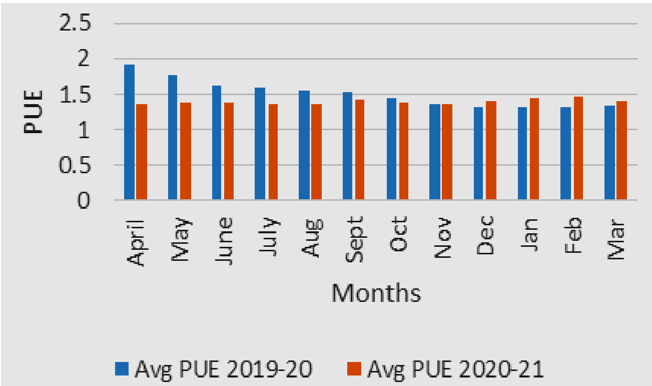

Fig. 2 Average PUE

2. Pue of Data Centers

Storing equipment, information management systems, climate control devices, electrical equipment, and other auxiliary equipment are the primary components of data centers [10]. Fig. 3 illustrates the energy consumption breakdown in a data center.

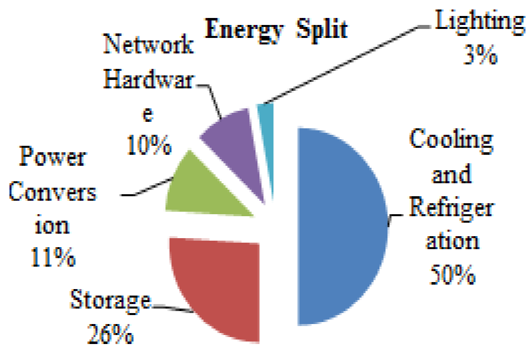

Fig. 3 Energy consumption of a typical data center [9]

It can quantify the amount of power used by data center equipment using the Power Usage Effectiveness (PUE) metric [11]. Power consumption in a data center is measured as a percentage of total power consumption for all of the gear in the data center.

$$PUE = \frac{Total\,power}{IT\,Equipment\,power} = \frac{P_{Total}}{P_{ITEqp}} \qquad (1)$$

The amount of electricity devoted only to the data center is the total power. IT equipment power refers to the power IT equipment uses to process, analyze, maintain, or transmit data inside the compute area.

$$PUE = \frac{P_{cooling} + Pp_{ITEqp} + P_{ElecLoss} + P_{Misc}}{P_{ITEqp}} \qquad (2)$$

The PUE value is also defined by three major factors cooling, power loss, and miscellaneous power (other electricity consumptions). The additional value is divided by the value of IT equipment. This is a simple and accurate way of gauging a data center's total power and energy usage. If you want to reduce the PUE value of a data center, you need to address all of the variables that affect electricity consumption.

The PUE values are influenced by the cooling of the infrastructure (PCooling), installed IT equipment (PITEqp), power loss through switches (PElec.Loss), and other electricity consumptions (PMisc) [12], [13]. Current data center refrigeration infrastructure management and control solutions do not address the real behaviors or offer control-side coupling of the targeted fast-computing systems in the data centers. The cooling infrastructure is usually built for maximum power demands from deployed computer systems to respond poorly to the dynamic changes in the computer systems installed.

3. Related Work

Shoukourian *et al.* [12] utilized a neural network-based method to predict the performance ratio of a high-performance data center. The Coefficient of Performance has been predicted for the data centers. Basha et al. [14] studied that the assessment is parametrized by outcomes related to the performance assessment indicators, which reduces the difference between the genuine evaluation and the machine learning prediction. The regression method aims to create a feasible plane whose equation yields a more accurate result. Balanici *et al.* [15] have utilized server traffic flow to enhance the PUE of the data center. The auto-regressive neural network approach has been utilized to forecast the flow of traffic in the server.

Moreover, optimizing the cooling system's control policy can decrease the data center's energy consumption. The cooling system of a data center might benefit from a reinforcement-learning method, according to Li et al. [16]. The new model has a higher power efficiency with an 11 percent drop in cooling expenses. The energy consumption of huge data centers has also been decreased by Haghshenas et al. [17], who used a variety of agent reinforcement-learning approaches. Kumar et al. [18] used a supervised machine learning method, linear regression, to model the data. The PUE falls when the temperature rises within the Uptime Institute and ASHRAE's allowed limits and standards. Apart from regulating the greater denomination of temperatures within the limitations as stated and intended for the IT Equipment, other ways can cover IT loads and dig further into the cooling infrastructure.

Liu et al. [19] discussed that the polynomial fitting technique, which is dependent on the Romonet simulation method and the global data center traffic, which combines the PUE for the global dynamic and the PUE for the high latitude area, is used to estimate global data center energy consumption and CO2 emissions in various scenarios. In the future, data centers in the Pan-Arctic area will be able to handle climate and energy issues successfully. Gao [20] put in a lot of effort to forecast the PUE of a Google data center. This study aimed to demonstrate that current sensor data may be used to use machine learning to estimate data center effectiveness and improve energy efficiency. The model has been implemented in Google's data centers. The dataset included twenty variables. This customized model has reduced the overall cooling by 40% and power usage by 15%, thereby improving the PUE. However, the sensor data contained various types of uncertainty due to faults and old sensors. While there are many complex AI-based approaches, plenty of simpler regression-based algorithms may be effective for optimizing the PUE. Hence regression-based algorithms will be implemented and compared in this work. A comparison summary of the above-related research work is specified in Table 1 with adopted methods, goals, and weaknesses.

Table 1. Comparison of related work

Ref.	Method	Goal	Weakness
[12]	ANN	To predict the performance of the data center	The model only explores the refrigeration point of view
[15]	Auto-regressive neural-networks	A server's traffic flow is monitored to optimize the traffic flow.	The results were not very effective.
[16]	Reinforcement Learning algorithm	To reduce cooling cost	The demand response signals are not sufficient
[14]	Regression Model	To forecast the cost imposed on the consumer based on their consumption	The efficiency of the results can be improved
[17]	Reinforcement Learning algorithm	To mitigate the energy consumed by large data centers.	VM migration is used, which is a complicated process.
[18]	Regression Model	Shows how the temperature and the Data load affect the PUE to save cost and reduce energy usage	The results were not very effective.
[19]	Regression Model	Predicting global data center traffic growth	The efficiency of the results can be improved
[20]	ANN	To enhance the PUE of data centers	This has only been tested on Google Datacenter

4. Regression Algorithms Tested

There are many algorithms used for regression. Some of them can also be used for classification. Some of the regression algorithms are discussed in this section.

4.1. Linear Regression

This type of regression is a statistical approach for predictive modeling that is most elementary and often utilized [21]. It gives us an expression where the chosen characteristics are separate variables and rely on the target variables.

$$Y = \theta_1 X_1 + \theta_2 X_2 \ldots + \theta_n X_n \quad (3)$$

y and x are dependent and independent variables, respectively, and the rest are the coefficients [22]. If just one independent variable is present, then it is labeled as simple linear regression, and if it has multiple independent variables, then it is considered multiple regressions.

4.2. Ridge Regression

Ridge Regression is a multicolor regression data analysis technique. When multi-linearity occurs, smaller square estimates are unbiased, but their discrepancies are substantial and thus far from real value [23]. The inverse ridge reduces normal inaccuracies when the reverse estimates add a degree of partiality. The net effect is expected to be more accurate. The existence of almost linear links between the separate variables is termed multi-linearity. This link results in a zero partition during regression calculations, leading to a computational abortion [24]. There is no zero-division and computational abortion if the link is incorrect. The division still makes very few changes to the results. Therefore, one of the first phases in this approach is to check if multi-co-linearity is a problem.

By applying a penalty on the magnitude of the coefficients, Ridge regression overcomes some of the issues with Ordinary Least Squares. The ridge coefficients are used to minimize the penalty residue sum of squares:

$$\min_{w} \left| \left| Xw - y \right| \right|_2^2 + \alpha \left| \left| w \right| \right|_2^2 \tag{4}$$

The degree of shrinkage is controlled by the complexity parameter α≥0: the bigger the value, the more shrinkage there is, and therefore the parameters become more resilient to co-linearity.

4.3. Lasso Regression

This regression is a kind of linear regression that utilizes shrinkage. Shrinkage is when data values are reduced, like the mean, to a core point. The lasso process promotes basic, sparse models which do not contain many parameters [25]. This kind of regression is ideal for approaches with higher multi-collinearity levels or when automating certain portions of selecting the models, like selecting the variables or removing the parameters [26].

$$\min_{w} \frac{1}{2n_{samples}} \left| \left| Xw - y \right| \right|_2^2 + \alpha \left| \left| w \right| \right|_2^2 \tag{5}$$

Thus the lasso estimation minimizes the least-squares penalties with α||w||1 added, where α subjects as constant and ||w||1 is represented as ℓ1-norm of the coefficient vector. The Lasso implementation employs coordinated descent as the algorithm to fit the coefficients.

4.4. Decision Tree

This approach is one of the supervised learning approaches that include algorithms like C4.5, C5.0, CHAID, and CART algorithms. It works for both regression and classification applications. The algorithm is used to analyze the data gathered and draw patterns from the current shape of the tree. The module can be used to forecast the appropriate value type based on the requirement [27]. The decision tree is trained by transmitting root-node-to-leaf information. The data is constantly split by predictor variables to clean nodes for children. The data are predictor-sensitive. The root node begins with all the training information. The decision tree will split progressively into groups by selecting the one variable predictor to form the root partition. Now three children's nodes will be created. One with black casings is known as a node with leaves. Two more branches are split to create another four branches. Either every leaf has just one resulting class, or it is too small to split. A number of potential split sites are identified for each node for each predictor variable. The method estimates the increase in the data purity that each dividing point would produce [28]. The split with the greatest substantial improvement is chosen to partition data and generate children's nodes.

Determining the characteristic of the root at each level of the Decision Tree is a major task. Attribute selection is the name given to this procedure. For deciding on attributes, there are two popular methods.:

1. Information Gain
2. Gini Index

For information, Gain:

The set of all potential values of A is called Values (A), and it is a subset of the larger set Sv, with A equal to v., then

$$Gain(S, A) = Entopy(s) - \sum_{veValues(A)} \frac{|S_p|}{|s|}.Entropy(S_v) \tag{6}$$

The result is mostly represented as either yes orno in the decision tree. Entropy's formulae are given as follows:

$$E(S) = -P_{(+)} - P_{(+)} \log P_{(+)} - P_{(-)} \log p_{(-)} \tag{7}$$

Here P_+ represents the probability of a positive class

P is denoted as the probability of a negative class

S is denoted as the subset of the training example

For Gini Index :

The Gini Index is calculated using the formula listed below.

$$Gini\ Index = 1 - \sum_j p_j^2 \tag{8}$$

4.5. Random Forest

This algorithm consists of many individual decision trees, which act as a group. When a tree randomly spreads a class prediction, the one with the most votes is used to make the model's prediction. [29]. A simple yet powerful idea behind random forest is the knowledge value. The random forest approach works effectively since many substantially uncorrelated decision trees that work as a group will outperform each component's model.

The little correlation of models is the greatest priority. Just as investments with low links come together to create more than the sum of their parts, uncorrelated models are more accurate than any of the individual forecasts to produce set predictions [30]. The explanation is that the trees shield one another from each other's mistakes. Although some trees may be mistaken, many others will be accurate so that the trees can travel in the proper direction.

In the Random Forest, the average of all trees is derived by dividing the total number of trees by a relevant feature's value.:

$$RFfi_i = \frac{\sum_{j \in all\,trees} normfi_{ij}}{T} \tag{9}$$

- RFfi sub(i)= the relevance of feature i is estimated from all of the trees in the Random Forest model.
- normfi sub(ij)= the normalized feature importance for i in tree j
- T is denoted as the total number of trees

4.6. K-Nearest Neighbour

K-Nearest Neighbour is a method of supervised learning. The new case/data assumes the comparability with the existing cases and places the new case in the most comparable category to the current categories. It groups and predicts the accessible data based on similarity [31]. This implies that the K-NN method can be readily categorized into a well-suited category if fresh data is present. It may be used for regression and grading but is mainly utilized for grading issues. It is a non-parametric method, such that the underlying facts are not assumed [32].

KNN regression is used to obtain the K nearest neighbor's quantitative output average. Another option is to use an inverse distance normalized average of the K nearest neighbors. In KNN regression, the distance functions are the same.

Distance Function

$$Euclidean = \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2} \tag{10}$$

$$Manhattan = \sqrt{\sum_{i=1}^{k} (x_i - y_i)} \tag{11}$$

$$Minkowski = \left(\sum_{i=1}^{k} (|x_i - y_i|)^q \right)^{1/q} \tag{12}$$

4.7. Support Vector Machine

The SVM is a technique used for both regression and classification. Its concept is the limit fitting with the area of points, and they all belong to the same class. The insertion of additional points is categorized after the limit has been set on the sample (training). I must verify whether it falls inside or outside the border before categorization. Once the border is fixed, the training data are redundant. An essential set of points is needed to identify and define the border. Suppliers support a data point called a vector, which describes the data points supporting the border [33].

It is the updated version of the KNN from a memory-based learning system to a real learning method [34]. Like the method k-Nearest Neighbour, SVM posits that the algorithm depends on the seen data given to the algorithm, while new invisible data are predicted. However, SVMs take further precautions by dividing the learning material into partitions. These partitions are constructed of more than three dimensions, using hyperplanes or lines. Use vectors are calculated to generate hyperplanes from learning data for each predictor variable. Comparable to other learning algorithms, such as neural networks, SVMs make predictions based on the position of fresh data; this is similar to the nearest k method.

A training data set with the corresponding observed values yn could describe xn as a multivariate collection of N observations.

Find f(x) with the lowest norm value (β′β) to check the linear function and ensure it is as flat as possible. This can be written as a convex optimization problem to reduce subject to all residuals with a value less than ε or as an equation:

f(x)=x′β+b, (13)

J(β)=12β′β (14)

∀n:↓yn−(xn′β+b)↓≤ε . (15)

No other function f(x) may exist to meet all of these requirements. Introduce slack variables ξn and ξ*n for each point to deal with the unsolvable constraints.

4.8. MLP

Multilayer perceptron (MLP) is a type of deep learning algorithm which consists of many perceptrons. They consist of an input layer for receiving the signal, a layer of output for deciding or predicting the inputs, and a random number of hidden layers in between, which are the actual working of the MLP [35]. MLPs with a hidden layer may estimate any permanent function. An MLP regression model is used in this study. Three convolution layers are used with activation functions 'ReLU' and Tanh (The hyperbolic tangent activation) on the hidden layers and softmax on the output layer. The Optimizer function used are 'adam (Adaptive Moment Estimation) & lbfgs (Limited-memory Broyden–Fletcher–Goldfarb–Shanno). The early stopping technique is used to get better accuracy in a short period of time. The algorithm is frequently used to control supervised learning issues. The training set learns to predict the connection between these inputs and outcomes. It includes modifying the model's parameters or weights and biases to reduce the mistakes. Backpropagation is used for weighing and biasing the error, and error may be assessed by a wide range of methods such as Root Mean Square Error (RMSE) [36].

Hidden layer cells produce their output yj as a function f [13] of the original input multiplied by the weight wji with the threshold applied and calculated using the activation function.

$$Yj = f\left(\sum w_{ji} x_i\right) \quad (16)$$

There is a transfer functions such as linear, log-sigmoid, and tan-sigmoid. The activation function must be able to differentiate.

$$\text{Linear: } f(x) = x \quad (17)$$

$$\text{Log - Sigmoid: } f(x) = \frac{2}{1+e^{-x}} \quad (18)$$

$$\text{Tan - Sigmoid: } f(x) = \frac{2}{1+e^{-2x}} - 1 \quad (19)$$

5. Results and Analysis

This research aims to demonstrate the cause-and-effect relationships between the variables. Since changes in one variable are directly responsible for changes in another variable, the optimization of certain parameters would assist in optimizing the final power and costs. In this research, the optimum PUE values are predicted using various regression algorithms to check the accuracy of each algorithm. Many of the regression methods outlined in the preceding section have been implemented, including linear regression, ridge regression; lasso regression; random forest; decision tree; SVM, KNN, and MLP.

An Indian Tier-4 data center provides the data for the study. Tier 4 data centers are considered free of fault, and there is very less chance of data failure since the flow of data is maintained effectively. Despite this, there are still issues in the power efficiency which require further improvement. The data is directly obtained and collected from the data centers. Various power consumptions and losses, such as the cooling power required by the installed IT equipment, power loss through switches, and other miscellaneous electricity consumptions, are included in the gathered data. The variables in the dataset consist mainly of Total Load, IT Load, and PUE. The other variables considered for the analysis are HT Power, LT Power, UPS Power, DG Power Consumption, Power Distribution Units (PDU), HVAC consumption, PAC Consumption, and Chiller plants. The HT and LT power lines supply power to the data centers from the grid, while the DG power denotes the power from the diesel generator at the premises. UPS power is the power stored in case of power failure for a short period of time. The PDUs convert the power to the allocated voltage and current requirements and supply the power to each piece of equipment. PAC is the power consumed by Precision Air Conditioners in the datacenters, while HVAC represents power consumed by all refrigeration units.

The data is analyzed, and the PUE is predicted for varying loads through regression approaches. The code for each algorithm is written and executed using Python Programming. The dataset is imported into the environment and then analyzed to predict the PUE by the model. The algorithms used in the comparison are linear regression, lasso regression, ridge regression, DT, SVM, KNN, random forest, and MLP. K-fold validation is performed for each of the regression approaches to attain accuracy. The predicted PUE is compared to each other to identify the algorithm with the highest accuracy.

The scatter diagrams for the collected data are shown in Fig. 4 for different loads and PUE. In all scatter diagrams, it can be seen that the density of data is higher for the mid-level of Load. The density of data is less for lower and higher loads and PUE. This is suitable since, most of the time, the loads in actual data centers are neither low nor high.

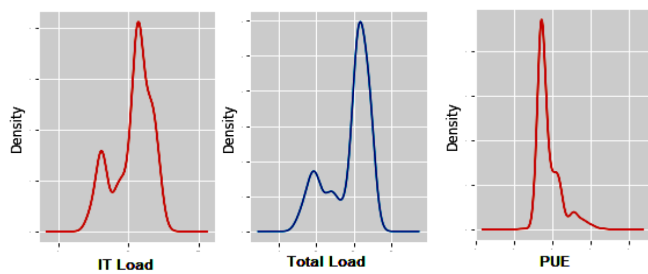

Fig. 4 Data Collected Scatter Diagrams

The relation between the IT load and PUE is shown in Fig. 5. The PUE is higher at lower loads and decreases as the load increases. The PUE is the highest at 2.18 when the Load is at 10,500, the lowest above 50,000.

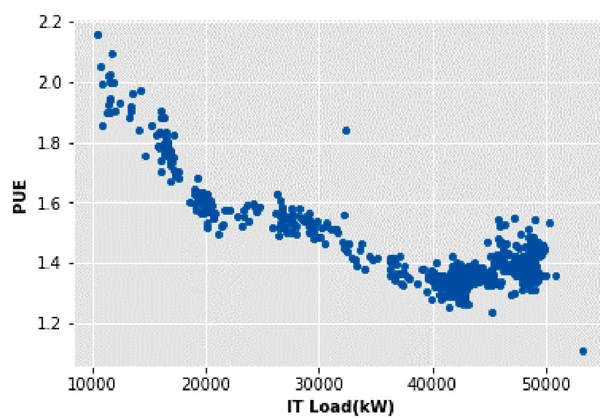

Fig. 5. Relation between IT Load and PUE

The relation between the Total Load and PUE is shown in Fig. 6. The PUE is higher at lower loads and decreases as the total load increases. The PUE is the highest at 2.19 when the Load is at 22,500, and it is the lowest at 59,000.

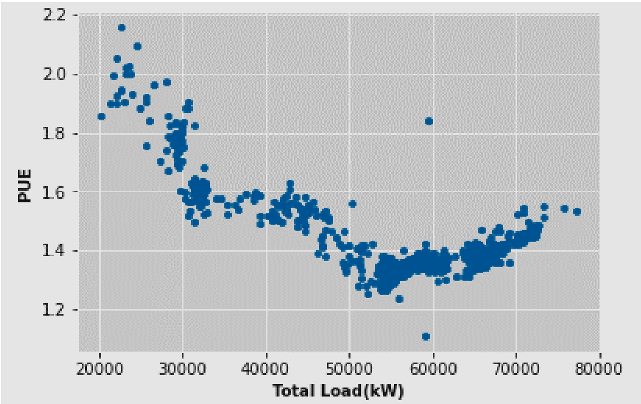

Fig. 6 Relation between Total Load and PUE

The relation between the IT load and Total Load is shown in Fig. 7. The relation is mostly linear, and there are interdependent on one another. When the IT load is at 10,000, the total Load is 20,000. This increases, and when the IT load corresponds to 50,000, the total Load corresponds to 78,000.

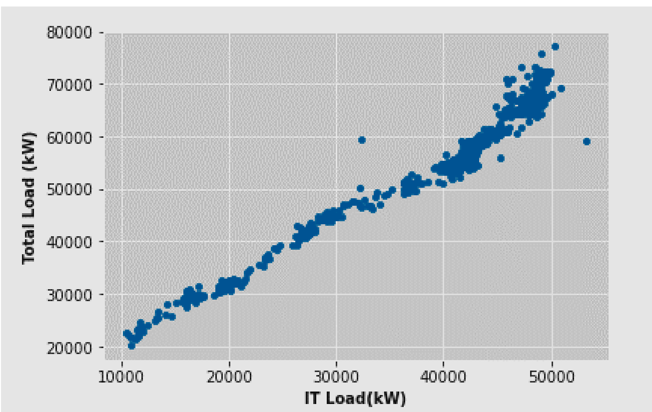

Fig. 7 Relation between IT Load and Total Load

Various algorithms are used for regression, and the results are obtained concerning PUE. The PUE is predicted using the algorithm and compared with the measured values. The relation for linear regression is shown in Fig. 8. The predicted value coincides with the measured value at lower values, but it does not coincide much at higher values. The accuracy obtained for this algorithm is 89.41%

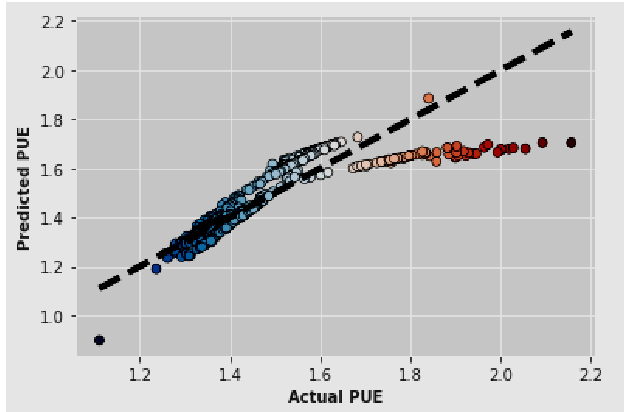

Fig. 8 Linear Regression

The relation for ridge regression is shown in Fig. 9. The predicted value for the ridge regression coincides with the measured value at lower values, but it does not coincide much at higher values. The results are similar for linear regression and ridge regression and also had a similar accuracy of 89.41%

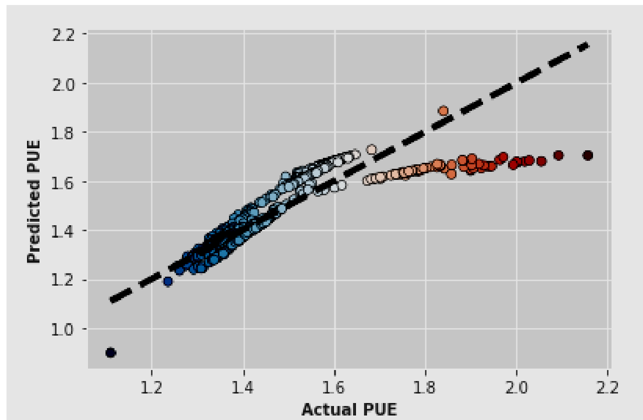

Fig. 9 Ridge Regression

The relation for lasso regression is shown in Fig. 10. The predicted value for the lasso regression coincides with the measured value at lower values to some extent, while most of the other lower values are very much accurate. The values at higher PUE are also skewed. The accuracy obtained for this algorithm is 80.85%.

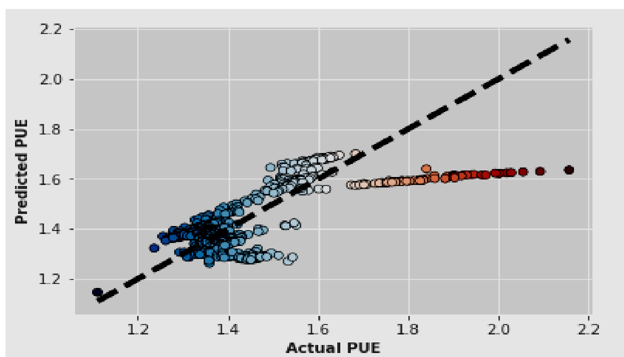

Fig. 10 Lasso Regression

The relation for the random forest is shown in Fig. 11. The predicted value for the random forest regression does not coincide with the measured value at both lower and higher values. The algorithm is comparatively higher than the general regression techniques at 91.07%

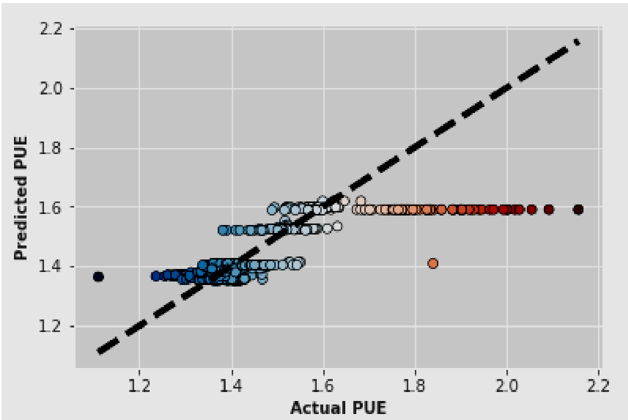

Fig. 11 Random forest regression

The relation for the decision tree is shown in Fig. 12. It is comparatively better than the predicted value for the random forest regression. The predicted PUE does not coincide with the measured value at higher values but with lower values. The algorithm predicts the PUE well and has a very high accuracy of 97.67%.

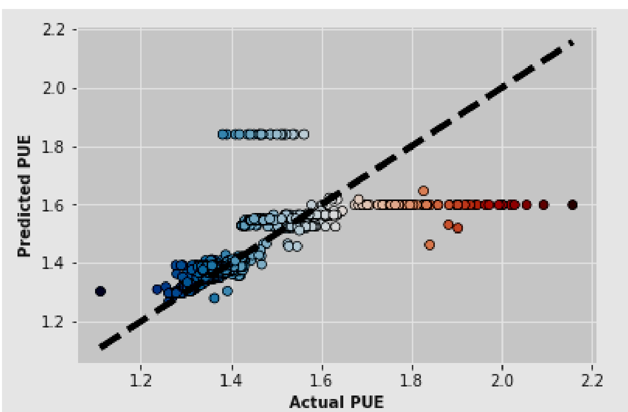

Fig. 12 Decision Tree regression

The relation for regression using SVM is shown in Fig. 13. The predicted value for the SVM regression coincides with the measured value at lower values, but it does not coincide much at higher values. The results are similar to the results of linear regression and ridge regression and also had a similar accuracy of 89.47%

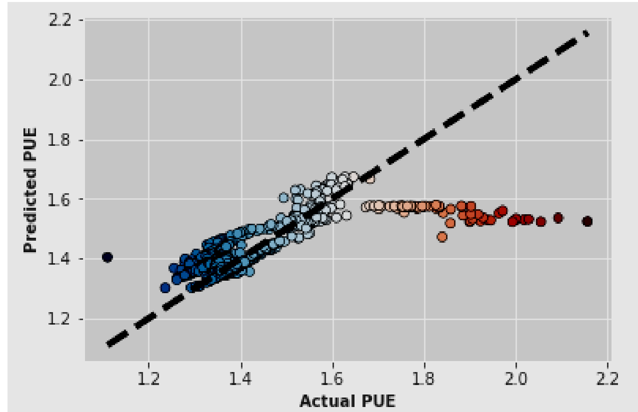

Fig. 13 SVM regression

The relation for regression using KNN is shown in Fig. 14. The predicted value for the KNN regression coincides with the measured value at lower values and slightly at higher values. Findings are similar to the results of the decision tree algorithm and have a very high accuracy rate of 98.29%.

The relation for regression using MLP is shown in Fig. 15. The predicted value for the MLP regression coincides with the measured value at lower values and slightly at higher values. It is evenly scattered throughout the predicted line; however, the accuracy is the lowest among the other compared techniques. The obtained accuracy is 85.32%. Also, while calculating the accuracy with all types of ML regression techniques, few parameters are used, as shown in Table 2.

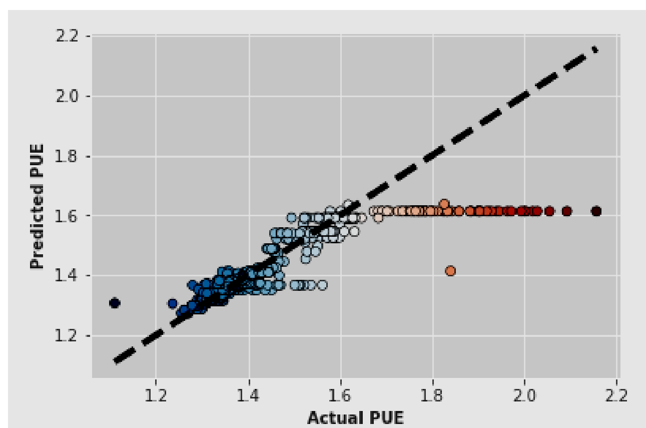

Fig. 14 KNN regression

Table 2. Algorithm Parameters Comparison

Algorithms	Parameters
Linear	fit_intercept=True,positive=False
Ridge	alpha=50, fit_intercept=True,solver='auto'
Lasso	fit_intercept=True,alpha=50, max_iter=10e5,selection='cyclic'
RF	n_estimators = 100, *, criterion='squared_error',max_depth=2, random_state=0
Decision Tree	splitter='best', random_state=0,criterion='squared_error'
SVM	degree=3,kernel='rbf',gamma='scale'
KNN	n_neighbors=2, weights = 'uniform', algorithm='auto', size=30
MLP	activation='relu', *, solver='adam',learning_rate='constant',random_state=1

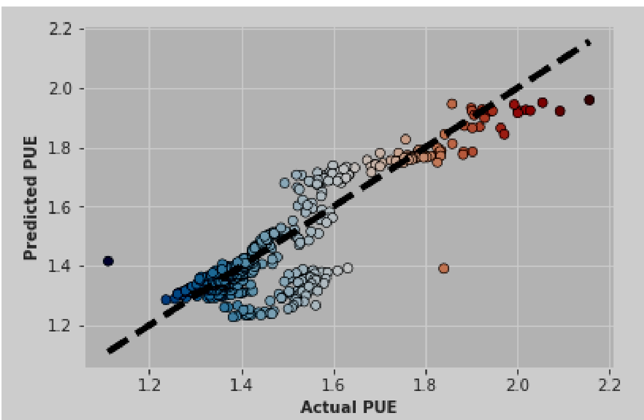

Fig. 15 MLP regression

Table 3 and Fig. 16 show the relative levels of precision achieved by the two methods under consideration. The KNN algorithm has the best accuracy, whereas the MLP approach has the lowest accuracy.

Table 3. Comparison of accuracies

Algorithms used	Accuracy
Linear	0.8941287802093469
Ridge	0.894128780828702
Lasso	0.808549692228578
RF	0.9107096822759171
Decision Tree	0.9766873873835856
SVM	0.8947861651310054
KNN	0.9828547415799562
MLP	0.8531681574753306

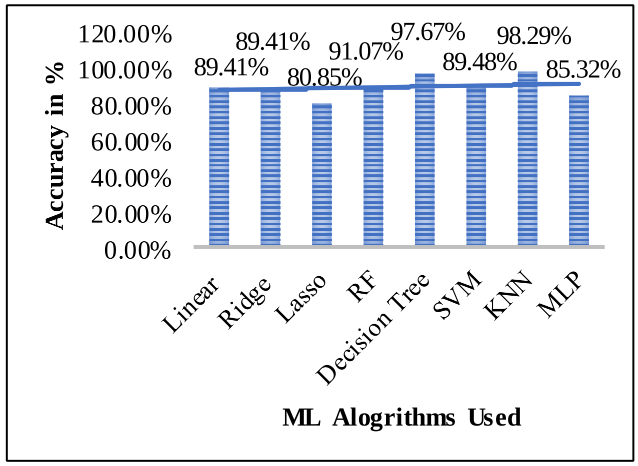

Fig. 16 Accuracy results of various ML regression techniques

6. Conclusions and Future Scope

One of India's Tier-4 data centers has tested machine learning algorithms and discovered that Decision Tree and KNN perform much better than other regression approaches, with 97 and 98 percent accuracy, respectively.

The work has focused on the PUE calculation; however, it is not the only performance metric; other metrics are also involved. PUE only involves active Load and not reactive loads. Although reactive power does not do any real work, it must provide inductive or capacitive loads to ensure network voltage stability. Also, the author has advised this based on the data obtained utilizing the data center in scope. Still, it has to be worked on a variety of data across the different geo-located data centers to derive a generalization of regression algorithm, which would be better and can be done in the future.

Typical inductive loads comprise cooling fans and computer server power units in a data center, whereas capacitive loads include computer server power supply units. If reactive power is not handled promptly at the Load that consumes it, it may result in significant network losses. It is also worth noting that non-linear loads like variable speed drives (VSDs), LED lights, UPS, and servers with SMPS use reactive power. How they consume electricity may lead it to be distorted. Harmonic is a reactive current component that exists alongside active current. Hence, other performance metrics must also be studied to calculate energy efficiency better. Hence, in the future, we can implement the above regression algorithms for other units and compare them.

Also, future work would involve checking if both the suggested models can be checked in combination with different indirect variables. The data size will also be increased by considering more data centers to conduct more effective research.

References

[1] Eltjo Hofstee, The future of data centers is green, (2019). https://www.datacenterdynamics.com/en/opinions/future-data-centers-green/

[2] Dublin, Global Green Data Center Market Report industry Trends, Share, Size, Growth, Opportunity and Forecast 2014-2025, (2020). https://www.globenewswire.com/en/news-release/2020/12/28/2150724/28124/en/Global-Green-Data-Center-Market-Report-2020-Industry-Trends-Share-Size-Growth-Opportunity-and-Forecast-2014-2025.html

[3] S. Taheri, M. Goudarzi, and O. Yoshie, Learning-based power prediction for geo-distributed Data Centers: weather parameter analysis, J. Big Data, 7(1) (2020) 8. doi: 10.1186/s40537-020-0284-2.

[4] H. M. M. Ali, T. E. H. El-Gorashi, A. Q. Lawey, and J. M. H. Elmirghani, Future energy efficient data centers with disaggregated servers, J. Light. Technol., 35(24) (2017) 5361–5380. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=jlt-35-24-5361

[5] A. Shehabi et al., United States Data Center Energy Usage Report, Berkeley, CA (United States), Jun. (2016). doi: 10.2172/1372902.

[6] J. Morley, K. Widdicks, and M. Hazas, Digitalisation, energy and data demand: The impact of Internet traffic on overall and peak electricity consumption, Energy Res. Soc. Sci., 38 (2018) 128–137. doi: 10.1016/j.erss.2018.01.018.

[7] IRENA, Global Energy Transformation: A roadmap to 2050, International Renewable Energy Agency, (2018). [Online]. Available: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2018/Apr/IRENA_Report_GET_2018.pdf

[8] M. S and D. S. L, A Study on Consolidation of Data Servers in Virtualized Cloud Atmosphere, Int. J. Comput. Sci. Eng., 6(11) (2019) 47–50. doi: 10.14445/23488387/IJCSE-V6I11P110.

[9] M. Dayarathna, Y. Wen, and R. Fan, Data Center Energy Consumption Modeling: A Survey, IEEE Commun. Surv. Tutorials, 18(1) (2016) 732–794.doi: 10.1109/COMST.2015.2481183.

[10] J. Dai, M. M. Ohadi, D. Das, and M. G. Pecht, Optimum Cooling of Data Centers. New York, NY: Springer New York, (2014). doi: 10.1007/978-1-4614-5602-5.

[11] E. Jaureguialzo, PUE: The Green Grid metric for evaluating the energy efficiency in DC (Data Center). Measurement method using the power demand, in 2011 IEEE 33rd International Telecommunications Energy Conference (INTELEC), (2011) 1–8. doi: 10.1109/INTLEC.2011.6099718.

[12] H. Shoukourian, T. Wilde, D. Labrenz, and A. Bode, Using Machine Learning for Data Center Cooling Infrastructure Efficiency Prediction, in 2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), (2017) 954–963. doi: 10.1109/IPDPSW.2017.25.

[13] T. Wilde, A. Auweter, and H. Shoukourian, The 4 Pillar Framework for energy efficient HPC data centers, Comput. Sci. - Res. Dev., v29(3-4) (2014) 241–251. doi: 10.1007/s00450-013-0244-6.

[14] R. F. K. Basha, M. . Bharathi, and K. Venusamy, Dynamic prediction of energy and power usage cost using linear regression-machine learning analysis, J. Phys. Conf. Ser.,1921(1) (2021) 012067. doi: 10.1088/1742-6596/1921/1/012067.

[15] M. Balanici; and S. Pachnicke, Server Traffic Prediction Using Machine Learning for Optical Circuit Switching Scheduling, (2019).

[16] Y. Li, X. Wang, P. Luo, and Q. Pan, Thermal-Aware Hybrid Workload Management in a Green Datacenter towards Renewable Energy Utilization, Energies, 12(8) (2019) 1494. doi: 10.3390/en12081494.

[17] K. Haghshenas, A. Pahlevan, M. Zapater, S. Mohammadi, and D. Atienza, MAGNETIC: Multi-Agent Machine Learning-Based Approach for Energy Efficient Dynamic Consolidation in Data Centers, IEEE Trans. Serv. Comput., (2019) 1–1, 2019, doi: 10.1109/TSC.2019.2919555.

[18] R. Kumar, S. K. Khatri, and M. J. Diván, Power Usage Efficiency (PUE) Optimization with Counterpointing Machine Learning Techniques for Data Center Temperatures, Int. J. Math. Eng. Manag. Sci., 6(6) (2021) 1594–1611. doi: 10.33889/IJMEMS.2021.6.6.095.

[19] Y. Liu, X. Wei, J. Xiao, Z. Liu, Y. Xu, and Y. Tian, Energy consumption and emission mitigation prediction based on data center traffic and PUE for global data centers, Glob. Energy Interconnect., 3(3) (2020) 272–282. doi: 10.1016/j.gloei.2020.07.008.

[20] J. Gao, Machine learning applications for data center optimization, (2014). [Online]. Available: https://research.google/pubs/pub42542/

[21] K. Bibas, Y. Fogel, and M. Feder, A New Look at an Old Problem: A Universal Learning Approach to Linear Regression, in 2019 IEEE International Symposium on Information Theory (ISIT), (2019) 2304–2308. doi: 10.1109/ISIT.2019.8849398.

[22] A. F. Schmidt and C. Finan, Linear regression and the normality assumption, J. Clin. Epidemiol., 98 (2018) 146–151. doi: 10.1016/j.jclinepi.2017.12.006.

[23] S. Liu and E. Dobriban, Ridge Regression: Structure, Cross-Validation, and Sketching, Oct. (2019). [Online]. Available: http://arxiv.org/abs/1910.02373

[24] Y.-R. Chen, A. Rezapour, and W.-G. Tzeng, Privacy-preserving ridge regression on distributed data, Inf. Sci. (Ny)., 451–452 (2018) 34–49, Jul. 2018, doi: 10.1016/j.ins.2018.03.061.

[25] J. Ranstam and J. A. Cook, LASSO regression, Br. J. Surg., 105(10) (2018) 1348–1348. doi: 10.1002/bjs.10895.

[26] R. Alhamzawi and H. T. M. Ali, The Bayesian adaptive lasso regression, Math. Biosci., 303 (2018) 75–82. doi: 10.1016/j.mbs.2018.06.004.

[27] D. Moon, H. Im, I. Kim, and J. H. Park, DTB-IDS: an intrusion detection system based on decision tree using behavior analysis for preventing APT attacks, J. Supercomput., 73(7) (2017) 2881–2895. doi: 10.1007/s11227-015-1604-8.

[28] W. Mao and F.-Y. Wang, Cultural Modeling for Behavior Analysis and Prediction, in Advances in Intelligence and Security Informatics, Elsevier, (2012) 91–102. doi: 10.1016/B978-0-12-397200-2.00008-7.

[29] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, A comparison of random forest variable selection methods for classification prediction modeling, Expert Syst. Appl., 134 (2019) 93–101. doi: 10.1016/j.eswa.2019.05.028.

[30] S. J. Rigatti, Random Forest, J. Insur. Med., 47(1) (2017) 31–39. doi: 10.17849/insm-47-01-31-39.1.

[31] Y. Song, J. Liang, J. Lu, and X. Zhao, An efficient instance selection algorithm for k nearest neighbor regression, Neurocomputing, 251 (2017) 26–34. doi: 10.1016/j.neucom.2017.04.018.

[32] F. Martínez, M. P. Frías, M. D. Pérez, and A. J. Rivera, A methodology for applying k-nearest neighbor to time series forecasting, Artif. Intell. Rev.,52(3) (2019) 2019–2037., doi: 10.1007/s10462-017-9593-z.

[33] E. Shams, A. Rizaner, and A. H. Ulusoy, Trust aware support vector machine intrusion detection and prevention system in vehicular ad hoc networks, Comput. Secur., 78 (2018) 245–254. doi: 10.1016/j.cose.2018.06.008.

[34] L. A. Winters-Miner et al., Prediction in Medicine – The Data Mining Algorithms of Predictive Analytics, in Practical Predictive Analytics and Decisioning Systems for Medicine, Elsevier,(2015) 239–259. doi: 10.1016/B978-0-12-411643-6.00015-6.

[35] H. Taud and J. F. Mas, Multilayer Perceptron (MLP), (2018) 451–455. doi: 10.1007/978-3-319-60801-3_27.

[36] A. Belay Adege, Y. Yayeh, G. Berie, H. Lin, L. Yen, and Y. R. Li, Indoor localization using K-nearest neighbor and artificial neural network back propagation algorithms, in 2018 27th Wireless and Optical Communication Conference (WOCC), (2018) 1–2. doi: 10.1109/WOCC.2018.8372704.