Power Usage Efficiency (PUE) Optimization with Counterpointing Machine Learning Techniques for Data Center Temperatures

Rajendra Kumar
Amity Institute of Information Technology,
Amity University, Noida, Uttar Pradesh, India.
Corresponding author: rajendra.kumar1077@gmail.com

Sunil Kumar Khatri
Amity University Tashkent, Uzbekistan.
E-mail: skkhatri@amity.edu, sunilkkhatri@gmail.com

Mario José Diván
Data Science Research Group, Economy School,
National University of La Pampa Santa Rosa, La Pampa, Argentina.
E-mail: mjdivan@eco.unlpam.edu.ar

Abstract
The rapid increase in the IT infrastructure has led to demands in more Data Center Space & Power to fulfil the Information and Communication Technology (ICT) services hosting requirements. Due to this, more electrical power is being consumed in Data Centers therefore Data Center power & cooling management has become quite an important and challenging task. Direct impacting aspects affecting the power energy of data centers are power and commensurate cooling losses. It is difficult to optimise the Power Usage Efficiency (PUE) of the Data Center using conventional methods which essentially need knowledge of each Data Center facility and specific equipment and its working. Hence, a novel optimization approach is necessary to optimise the power and cooling in the data center. This research work is performed by varying the temperature in the data center through a machine learning-based linear regression optimization technique. From the research, the ideal temperature is identified with high accuracy based on the prediction technique evolved out of the available data. With the proposed model, the PUE of the data center can be easily analysed and predicted based on temperature changes maintained in the Data Center. As the temperature is raised from 19.73 °C to 21.17 °C, then the cooling load is decreased in the range 607 KW to 414 KW. From the result, maintaining the temperature at the optimum value significantly improves the Data Center PUE and same time saves power within the permissible limits.

Keywords- Data, Data center, Energy efficiency, Power losses, Temperature, Relative humidity, Machine learning.

1. Introduction
To fulfil the futuristic requirement of information technology, the Data Center industry is receptive to build more centers across the world (Stephen & Lucy, 2020). It is because of an ever-increasing demand in IT equipment hosting & cloud systems since IT companies are moving from traditional in-house IT Data Centers to large-scale Data Center services providers (Crane-Droesch, 2018).

Due to various social initiatives around energy sustainability programs, there is always pressure on the Data Center industry to evaluate and maintain better energy efficiencies (Yuventi & Mehdizadeh, 2013).

Data Centers are experiencing major growth in the number of clouds & hosting requirements which

need a large number of Servers, Storage Devices including network equipment with higher bandwidth and major Data Center services providers & OEM expanding their Data Centers & building new Data Centers across the globe (Uzaman et al., 2019).

For many decades, many industries are trying to identify challenges of Data Center Efficiency including organisations like Green Grid and Uptime Institute, US Department of Energy who are guiding with various suggestions to have energy efficiencies by implementing best practices and developing various standards. Their initial focus has been on the Data Center facilities to work on the cooling infrastructure which requires double the time of energy than IT equipment hosted in the data center (Delforge & Whitney, 2014).

In half a decade, the projected Data Center power consumption usage is going to be around 33% of total global electricity consumption which may be 20% of the entire energy consumed around the world (Taheri et al., 2020). It has become a necessity & demand that the industry should work to improve the efficiency of power usage and in this paper, it has been tried how machine learning techniques can help in PUE prediction based on the changes suggested for maintaining the required temperatures to reduce PUE for better efficiency. As it is known that optimising the variables will improve the overall PUE, it is necessary to understand the relationship between the variables, and their effect on the other variables. Hence, in this research, the temperature and the Data load are varied to show how they affect the PUE, in order to lessen the energy consumed and save costs. The implementation of the linear regression and the optimal identification of the cooling forms the novelty of the paper.

This section has given a brief background to the data centers, their load, and the PUE. Section 2 discusses the literature review where the concept is expanded, and related work is discussed. Section 3 discusses the cooling systems, and their technologies used for cooling the data center. Section 4 discusses the conventional approaches utilised for optimising the parameters in the data center. Section 5 studies the temperature parameters and their relation to the energy efficiency of the data centers. Section 6 discusses the methodology in detail and is subdivided into individual subsections. Section 7 discusses the obtained results and numerically details how the parameters affect one another. Section 8 concludes the paper, and recommendations are specified for the future direction of this research.

2. Literature Review and Related Work

The Literature Review is conducted on the basis of standard sources & databases of IEEE, Scopus, and Springer and from reputed journals on the topic of ML techniques, Data Center, Energy Systems, Power & Cooling, and Refrigeration Systems. The diverse studies have been taken into consideration based on the abstract & title of the manuscripts. The best method of analysing the Data Center is Power Usage Efficiency (PUE) (Sharma et al., 2015).

PUE is calculated as the ratio of total Data Center (DC) load and IT load, Total DC Load includes Data Center Cooling Load and IT Load. If the PUE is higher for any Data Center, then it is considered as with lower efficiency. The ideal PUE value is one (1.0) and this cannot be narrowed down to the prerequisite value by handling some of the DC facility components that are not achievable due to power loss at UPS, PDU, Transformers, Auxiliary Pumps, distribution wiring, auxiliary equipment's related to IT Loads (Sharma et al., 2015).

For effective PUE, it has to be adjusted basis the DC component which is to be monitored over

time. This can only be done by estimating the PUE for a longer duration. Continuous analysis on PUE & DC components like power & Cooling components over a period of time help to improve energy efficiency (Yuventi & Mehdizadeh, 2013). Recently, some efforts are made to focus & improve Data Center efficiency by implementing machine learning techniques. It indicates that ML model is an effective practice to utilise the Data Center Sensor data to predicate required PUE which further help reducing significant cost (Gao, 2014).

Azimzadeh & Tabrizi (2015) has reviewed the different power issues in data centers. The hardware face problems such as increasing energy usage, high infrastructure costs, and maintenance costs. The study has indicated that computers and servers use unnecessary heat and the evaporative cooling mechanism is necessary to sustain the heat in the data centers. The article has proposed a novel model that includes the scaling of adjustable voltage frequency and the dynamic methods for power management. The energy usage has been reduced significantly using a MapReduce optimization strategy. Other issues faced by the data centers are outdated or malfunctioning equipment, thermal control, and compatibility errors (Elmasry et al., 2019).

Baccour et al. (2019) has studied the designs of data centers and have suggested quality of service based improvements. The proposed approach has been made to offset electricity demand and power-compatible routing algorithms from renewable energy sources. Some problems have been identified like idle networks, insufficient management of cooling infrastructure, and the use of brown. To reduce energy use, data centers can use equipment such as optical infrastructure, commodity networks and cellular infrastructure. Four different parameters are considered which are adaptive link rate, power algorithms, virtualisation, dynamic voltage/frequency scaling.

A novel approach of thermo-siphoning in conjunction with conventional mechanical refrigeration would be able to save energy by 47.30% (Zhang et al., 2008). The key necessity and subject of research in the data centers is also the cooling performance of the effective airflow distribution. The modules can be optimised effectively using machine learning techniques. A hybrid cooling solution utilising predictive control measures and thermal energy storage systems will save up to 18 percent of electricity (Wang et al., 2018).

Schmidt & Iyengar (2007) have discussed the various issues that might take place in data centers in the future, and the best practices to follow to solve these issues. The heat control of the data centers, which results in maximum energy conservation, is necessary to sustain the temperature. Ni & Bai (2017) have conducted a review of various research on refrigerating systems. It is inferred that more than 50% of data centers did not function properly and wasted a lot of energy resources. The development of energy-efficient strategies such as regulation of humidity, temperature control, airflow optimization, and economic cycles has been proposed to further improve energy quality. (Oró et al., 2015) have compared different energy-saving strategies by integration into waste heat recovery through absorption refrigeration or by a direct generation of electricity and indirect generation of electricity. Various models and innovative technologies in data centers were checked for comprehension.

Fulpagare & Bhargav (2015) have analysed the optimization of the cooling effectiveness. It is concluded that the new laws on input temperatures may be implemented easily by taking into account other electrical devices such as fans and blowers. But for improved and efficient device activity, physics-based principles related to the thermal environment are needed. Daraghmeh & Wang (2017) have also researched free cooling methods in Data Center applications, where other

operations, such as solar systems, absorption, evaporative cooling, and geothermal cooling have been integrated into air-, water- and heat pipe applications. Rong et al. (2016) has reviewed the development of different methods of energy conservation from different angles of technology. Various innovations have been suggested for optimising data center performance, such as network optimization, servers, and processors. When paired with powerful temperature control along with other energy conservation strategies, the other contemplations, such as energy use, price, and the atmosphere, have to be utilised efficiently to recover data centers energy performance. The air movement is more focused on work and provides helpful guidance. The engineers for the design and construction of modern data centers with effective ventilation and cooling systems will obey these directions.

To reduce Data Center PUE with optimization for energy efficiency, it is possible to predict PUE for different range values like higher temperatures and other parameters. Data Center cooling systems can be set to a higher Data Center temperature (which saves energy & increases cooling capacity) and still supply the load with safe operating temperatures. Further, this also can be maintained using air containment with hot-aisle and cold-aisle management. An efficient machine learning model can enable the Data Center services Providers to evaluate PUE to Data Center Operational various parameters (Singh & Guo, 2016).

3. Cooling System and Technologies Used the Data Center
The Data Center cooling systems use the following technologies (Kaylie, 2020):

  (i)   Calibrated Vectored Cooling (CVC) - It is air-cooled technology that optimises the airflow inside the equipment. This technology increases the effectiveness and eliminates the heat.
  (ii)   Chilled Water System (CWS) – Mid Level & Large Scale of Data Centers uses chilled water to cool air by air handlers (CRAHs). Computer room air handler (CRAH) when working with a chiller water system then this combination is called a chilled water system. In this system, refrigeration cycles are replaced with Chilled water in place of the air conditioning system.
  (iii)   Cold Aisle /Hot Aisle Design - Data Center uses cold aisles and hot aisle systems for effective cooling. Cold aisles take cold air intake from the front side of the rack and the hot aisle exhausts hot air from the backside of the rack. Hot Air expelled from the backside of the rack goes to Air cooling system to get chilled then comes back to be vented from the cold aisle.
  (iv)   Computer Room Air Conditioner (CRAC) – These AC Units are more common Air Conditioning systems that are being used in Data Centers. It uses a compressor that draws air across a refrigerant. These units are inefficient in terms of expected energy efficiencies but are less expensive compared to other efficient systems.
  (v)   Computer room air handler (CRAH) – CRAH units use chiller water to cool down the cooling coil which comes in contact with return air from racks. These units work with a water chiller system installed generally in the premises or top of the Data Center roof and these systems pull outside air which works well when outside temperature or climate is colder.
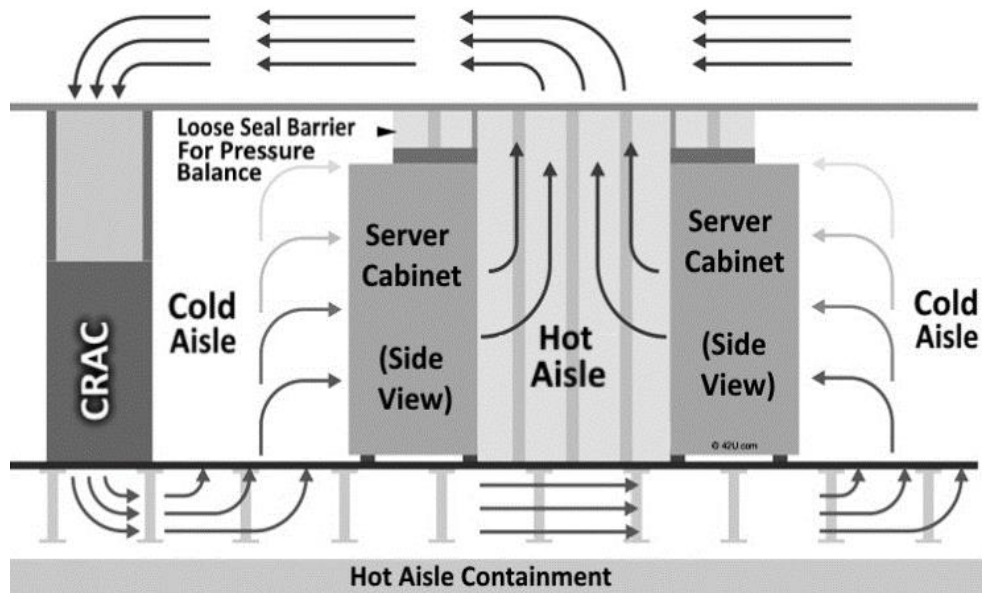
Figure 1. Hot & cold aisle data center design.

a. Raised Floor Architecture: Data center raised floor built above the concrete floor which keeps 2-3 feet space for free flow of cold air. To have better & effective cooling nowadays Data centers keep other network & power cables from the top of racks so that cold air can flow efficiently. A design of the hot & cold aisle data center is shown in Figure 1.

Despite the various options of cooling design, different factors impact the cooling (Jin et al., 2019):

(i)   Server Room Design,
(ii)   Ceiling Height,
(iii)   The gap between the equipment,
(iv)   Airflow speed,
(v)   Inlet & outlet air temperature
(vi)   Air leakage.

Therefore, it is identified that if these factors may be controlled then there is a possibility of improving the energy efficiencies.

4. Conventional Method Adopted for Energy Efficiency

Data Center services providers and many Data Center related Governance & Standards Body suggest many approaches which help in the reduction of power energy consumption which can be broadly categorised as being hardware & software-centric. They are collectively responsible for power consumption and power efficiency for which Coefficient of Performance (COP is a metric for measuring the efficiency of cooling infrastructure). COP value can be arrived at by the fraction between cold amounts from cooling infrastructure & consumed energy by cooling infrastructure (Shoukourian et al., 2017);

$$COP = Cooling\ Circuits\ /\ Cooling\ Circuits \tag{1}$$

Power Usage Effectiveness (PUE) is significant in the Data Center to find out the Data Center efficiency measurement that determines how energy is being utilised in the Data Center. It is defined as the ratio of the total power used by the Data Center to the total power consumed by the IT equipment.

$$PUE = PTotal\ /\ PIT = (Pcooling + PIT + Pelectrical\ Losses + Pmisc\ )/PIT \tag{2}$$

Heat reduction methods play an important role with respect to hardware because this keeps control of the frequency and voltage on the processing units of the equipment's which control & help in the reduction of energy consumption (Fouladi et al., 2017).

With consolidation/segregation of multiple disks to one place or by optimising the total storage utilisation can help in the reduction of the power usage of the disk (Ganesh et al., 2007). Similarly, the use of software, servers can be used for effective utilisation to improve efficiency. Generally, virtualisation software's are used for optimising the usage of the server and shutting down the unnecessary running IT and networking equipment (GeSI, 2020).

Type of air-conditioning methods like heat pipe (Capozzoli & Primiceri, 2015) and on-chip cooling method practices help to reduce the power utilisation. When compared to the general air cooling method, the heat pipe method helps in the reduction of power consumption by 38.9%, while another end on-chip cooling method reduces power consumption by 60% (Marcinichen et al., 2012). Despite this approach, an efficient cooling method reduces heat loss from the data centers as every electronic equipment delivers heat loss. The best efficient cooling infrastructure should achieve high effectiveness and consistency at a lower cost. The efficiency index should not be higher than 2.5 (Ni & Bai, 2017).

Normally, a Data Center uses 50% of installed power energy capacity. Effective cooling methods help in the reduction of power consumption and costs (Koronen et al., 2020). Despite this, the Cooling efficiency including airflow distribution has been the focus area of research for the Data Centers Machine learning technologies can help in the optimization of cooling components effectively. A hybrid approach using predictive control methods can help in power saving up to 18% (Wang et al., 2018).

Power optimization of the cooling systems includes methods & approaches of effective cooling systems. The Air Condition system works as per the temperature setting so optimization is difficult without knowledge of the refrigeration & cooling infrastructure. There are different types of technologies that use electrical, mechanical & thermal principal. The above methods and approaches adopted are not adequate enough for the Data Center cooling optimization & energy efficiency. Therefore, Machine learning-based techniques are emerging as an alternative to adopted conventional approaches. Machine learning techniques learn from historical data which is collected over a while on energy utilisation related to electrical & cooling infrastructure (Taheri et al., 2020).

5. Temperature & its Effect on Energy Efficiency of the Data Center

According to the American Society of Heating, Refrigerating, and Air Conditioning Engineers (ASHRAE) advise that the inlet temperature which goes inside the rack & server should be between 18 and 27 degrees Celsius (or 64.4 to 80.6 degrees Fahrenheit) with relative humidity (RH) from 20 to 80 percent. It is on the higher side; however, the Uptime Institute suggests an upper limit of

25 degrees celsius (77 degrees Fahrenheit).

Utility Infrastructure and IT equipment must be calculated to find out the total impact on energy consumption. Following relationship to be considered while analysing the multiple components for efficiency.

$$\text{Power} \propto f(T) \tag{3}$$

Every component in the Data Center which uses energy must be analysed for its impact on a temperature change as well as for any impending interrelationships (Patterson, 2008).

Hot Aisle Containment System (HACS) can help to save 43% in the yearly energy costs for the cooling infrastructure, corresponding to a 15% reduction in the annualised PUE compared to Cold -Aisle Containment System (CACS) while holding the uncontained Data Center area to 24°C/75°F (Niemann et al., 2011).

Improvements in server architecture helped to decrease the overall temperature needed in Data Centers. Earlier optimal temperatures required to be maintained around 18 degrees celsius. Due to these new technological advances, Data Center Server Rooms can be kept at high temperatures up to 35 degrees celsius. The recommended temperature and relative humidity suggested by ASHRAE are shown in Table 1.

Table 1. Temp & RH suggested by ASHRAE (TC, 2016).

Classes(a)	Equipment Environmental Specifications							
	Product Operation (b) (c)					Product Power Off (c) (d)		
	Dry-Bulb Temperature (*F) (e) (g)	Humidity Range, non-condensing (h) (i)	Maximum Dew Point (*F)	Maximum Elevation (f)	Maximum Rate of Change (*F/hr) (f)	Dry-Bulb Temperature (*F)	Relative Humidity (%)	Maximum Dew Point (*F)
Recommended (Applies to all A Classes; individual data centers can choose to expand this range based upon the analysis described in this document)								
A1 to A4	64.4 to 80.6	41.9*F DP to 60% RH and 59*F DP						
Allowable								
A1	59 to 89.6	20 to 80% RH	62.6	10,000	9/36	41 to 113	8 to 80	80.6
A2	50 to 95	20 to 80% RH	69.8	10,000	9/36	41 to 113	8 to 80	80.6
A3	41 to 104	10.4*F DP & 8% RH to 85% RH	75.2	10,000	9/36	41 to 113	8 to 85	80.6
A4	41 to 113	10.4*F DP & 8% RH to 90% RH	75.2	10,000	9/36	41 to 113	8 to 90	80.6
B	41 to 95	8% RH to 80% RH	82.4	10,000	NA	41 to 113	8 to 80	84.2
C	41 to 104	8% RH to 80% RH	82.4	10,000	NA	41 to 113	8 to 80	84.2

Unfortunately, there are many Data Centers that are yet to adapt to these new technology advancements in IT equipment and it is still unusual to see temperatures in Data Centers be maintained at 21 or 22 degrees celsius (Linesight, 2019). This is understood that cooling infrastructure and temperature play a large role in the Data Center's energy efficiency. It is also well known that both are interrelated, to the extent that if a slightly higher temperature than usual is maintained then it saves cooling power which in turn reduces the overall power consumption for the vital PUE optimization of the Data Center.

6. Machine Learning Approach & Model

A Machine learning model to work effectively, training data is desirable for a six-month duration of all Data Center data-related components of the following stored in comma-separated value (csv) local file.

   (i)   PUE Data – Power Usage Efficiency is intended as a ratio of total Data Center (DC) load and IT load, total DC Load includes Data Center Cooling Load and IT Load. The value of PUE is stored in this column.
   (ii)   UPS Units – Uninterrupted Power Supply units ensure that the servers and other apparatus do not lose power. The cooling components are necessary for efficient working, and UPS units make sure that there is no uninterrupted power. The total number of UPS units connected is recorded in this column.
   (iii)   Server Room's Temperature – The temperature is varied manually or automatically due to external factors. This factor is stored.
   (iv)   PAHU Units - Primary Air Handling Unit collects the outside air and dehumidifies it before sending it in. The number of PAHU units is recorded in this column.
   (v)   Chiller Units - The number of chiller units is recorded in this column.
   (vi)   Cooling Towers - The number of cooling towers is recorded in this column.
   (vii)   PDU Units – The number of power distribution units are recorded in this column.
   (viii) Electrical Panels - The number of electrical panels connected to the system and DC is recorded in this column.
   (ix)   Energy Meters – The amount of energy consumed is recorded in this column. The data is collected from digital energy meters, which have the capability of storing dynamic power.
   (x)   DG Units – The number of Diesel generators and their power capacities are stored in this column.
   (xi)   Transformers – The total number of transformers are stored in this column. This value may not change often since new transformers are added or removed only occasionally.
   (xii)   MDBU – This is the Main Distribution Board. This may not change often, however it is also collected as a part of data collection.

To work on the research data, the PUE, Cooling load & temperature data are first targeted, so that the relationships between the data can be assessed to see how dependent variables are impacted by independent variables. Following Machine Learning logical steps are followed during research as shown in Figure 2.
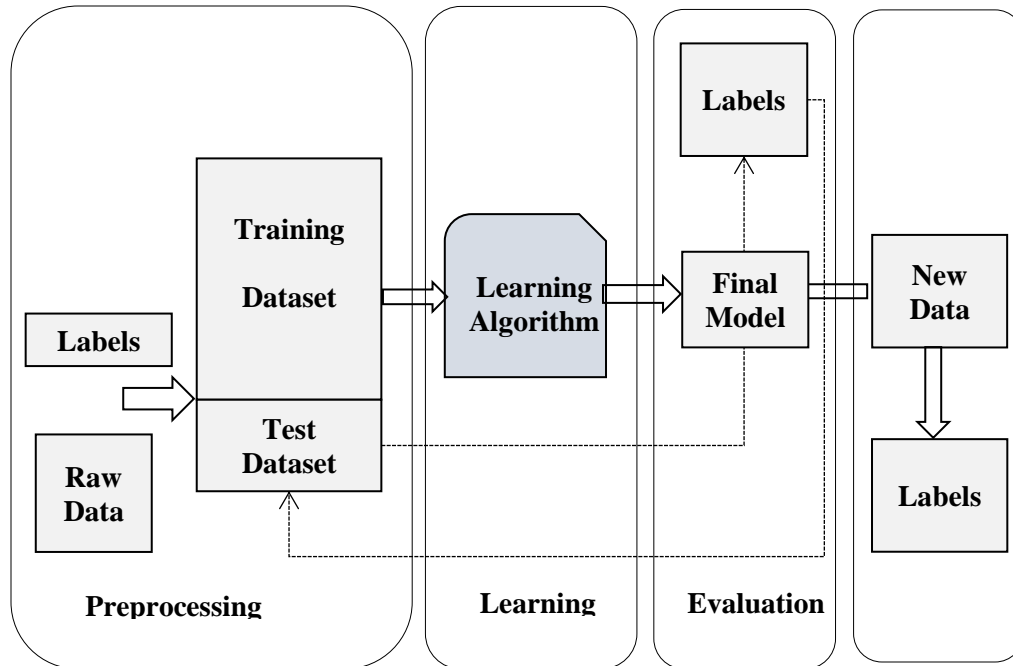
Figure 2. Machine learning logical step.

6.1 Input Data Pre-Processing

The data is first collected from the appropriate source. In this research, the data is obtained from a Tier 4 Data Center in India. The data is collected using a tool called Data Center Infrastructure Management (DCIM), which monitors the data continuously from different sensors and collected them into the database.

Outlier Detection: Data is gathered for a training model generated out of the DCIM software implemented in the Data Center. It was expected that it may have various outliners, and this may impact our prediction values because outliners can result in misleading information on PUE prediction & relationships between variables so it is necessary to remove the outliers from training data prior to modeling.

Missing Values: There are some instances, where some data might be missing. This may be useful for some cases, but it may create a bias during the implementation since the number of data may be different for different features due to the missing values. Hence, the missing values will be removed or skew will be created wherever necessary. Some of the undesirable missing values are considered as data noise and they are removed from the data through interpolation. Given the range of volume of the data, it was normalised to provide better evaluation & comparison. Alternatively, to clear the adjustment between the precision and handling overhead, the redundant features which do not contribute effectively to the training model are removed since this may lead to overfitting.

6.2 Regression Model

In this research, linear regression (LR) (supervised Machine Learning Approach) to build the model and predict the PUE to see how the PUE can be optimised based on temperature changes. Previous

research has been studied to see the impact of each variable on the prediction consequence since LR has rapid calculation speeds and prediction consistency which yields better results post the fine-tuning iterations (Thiangchanta & Chaichana, 2020). The LR Model has been initiated keeping in view PUE as a dependent variable.

The process of building a model typically includes two stages - Training & Testing data. Training of the model is dependent upon the finding of the linear regression parameters required for mapping input data into output data.

This model is also used to check multivariate linear regression to predict variables in determining PUE basis, on the independent variable like Temperature, Cooling Load, Total Load. Further extended to comprehend the significance of some other variables like IT Load & Relative Humidity, Outside Temperature and various other parameters and practice models, suggestions, guidance from various Data Center Governing Bodies to save energy.

PUE as Dependant variable depends on Temperature Independent variable which is shown as below:
$$PUE = f * (Temperature) + \varepsilon \tag{4}$$

PUE arrives at the formula given in section IV so it is greatly impacted in change to total load and IT load.

Total Data Center Load Consist of Cooling Load and IT load, Cooling Load major depends upon the power consumption by Cooling of the Data Center. Cooling Load may be considered as Dependant value while to see the impact of temperature as an independent value.
$$Cooling\ Load = f * (Temperature) + \varepsilon \tag{5}$$

Total Data Center Load consists of Cooling load & IT Load & as well and while IT load is not considered in this data which is part of future study in research that how this variable also impacts the PUE along with variables considered in scope.

Further to assess relationships and to ensure that how these variables are dependent on each other, the PUE variable & Total Data Center Load relationship can be used further to see how the cumulative Data Center Load impacts the PUE.
$$PUE = f * (Data\ Center\ Load) + \varepsilon \tag{6}$$

6.3 Evaluation of Regression Model
The model is run on the collected data and out of the PUE and temperature which show the following result.
$$PUE = -0.014700315\ (Temperature) + 2.00914952849852 \tag{7}$$

The result shows that the correlation between PUE and Temperature is Negative, which means if the temperature is increased, then PUE also gets reduced. To prevent overfitting, the model's output must be cross-validated. In this analysis, 10-fold cross-validation is used with 10 different sets of data. Since cross-validation is utilised, the accuracy of the proposed approach is effectively improved as shown in Figure 3. The output of cross-validation is tabulated in Table 2.

Table 2. The output result of cross-validation.

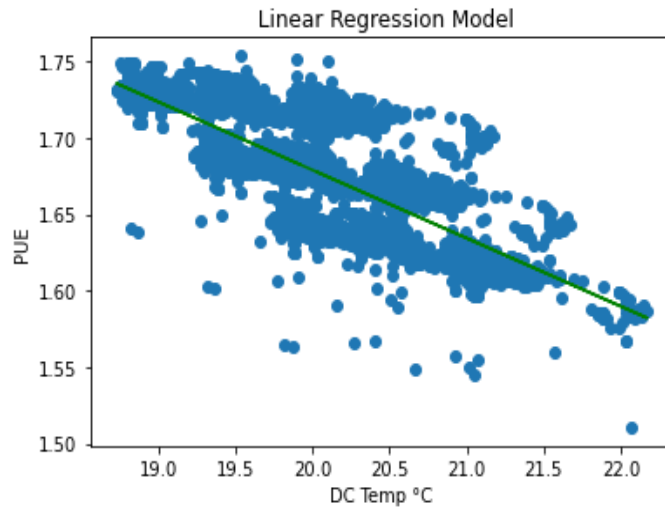| CV Metrics	Value
R²	0.507
Adjusted R²	0.507
MSE	0.000
RMSE	0.011

Figure 3. PUE vs DC temp- scatter plot.

Further, in reference to the model related to Eq(5), the model is run on the Cooling Load and temperature and the following results are obtained.

$$\text{Cooling Load} = -80.58732116 \,(\text{Temperature}) + 2078.0235406 \qquad (8)$$

From Figure 4, it is seen that the Data Center Cooling Load and Temperature is having a negative correlation. It means if the temperature is increased, then the data center Cooling Load also gets reduced.

Figure 4. DC cooling load vs Temp- scatter plot.

Similarly, the model was further run on the PUE and Data Center total Load based on PUE formulation calculation, it is understood that PUE depends upon the DC facility and further checked the relationship between the Data Center Total Load and PUE. When the modeling results were compared between these two variables, it is again generalised that PUE decreases when DC Total Load decreases based on the temperature.

$$PUE = 0.0000657904368 \,(DC\ Total\ Load) + 1.623689142 \tag{9}$$

These two variables PUE (Dependent Variable) and DC Total Load (Independent Variable) show a positive correlation and validate the PUE formula that if the total load is decreased, then the PUE also decreases as shown in Figure 5.

Figure 5. PUE vs DC total load - scatter plot.

The above model has proved that PUE decreases if the cooling is controlled with a small change in temperature thereby optimising the power usage efficiency.

6.4 Prediction & Forecasting
Once the prediction model gets trained with sample data, the future PUE and its optimization strategy could be forecasted based on the various variables of the data gathered from the Data Center DCIM software output of the sensor and metered devices.

7. Research Results and Analysis
In this section –the obtained results are analysed after executing the model on the training & test data. In relation to PUE and temperature of the data center, it is seen that the predicted values are working as per expected results, where the temperature was increased from 1.5 to 2.5 ºC in temperature variation scenario to optimise the PUE. When the temperature was increased, a decreasing PUE was observed in both actual and predicted data for the comparison of temperature and PUE as shown in Figure 6 and 7.
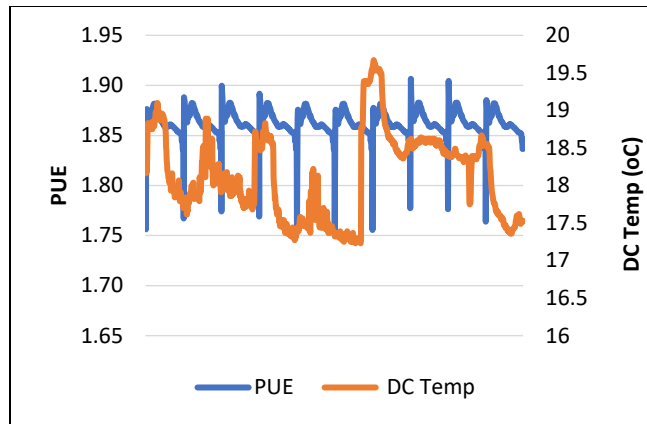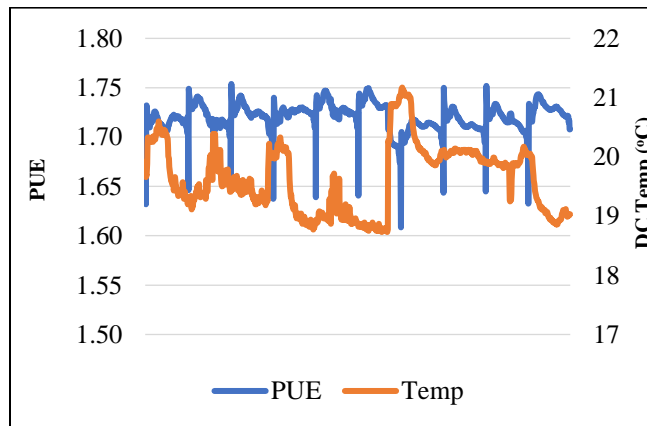
Figure 6. Actual data DC PUE vs temp.

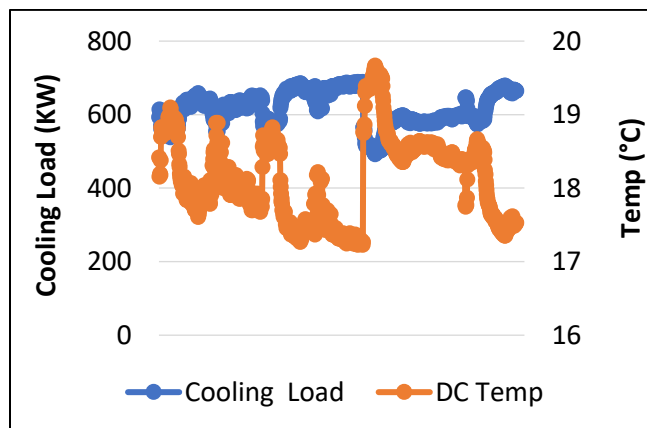**Figure 7.** Predicted data DC PUE vs temp.

Figure 8. Actual data DC cooling load vs temp.

Figure 8 & 9 is the representation of data to have a comparison of actual & prediction data of Data Center Cooling Load and Temperature. This also predicts that DC Cooling Load is also dependent upon the temperature variation. It can be observed that the Data Center Cooling load is decreasing when the temperature to be maintained in the Data Center is increased.
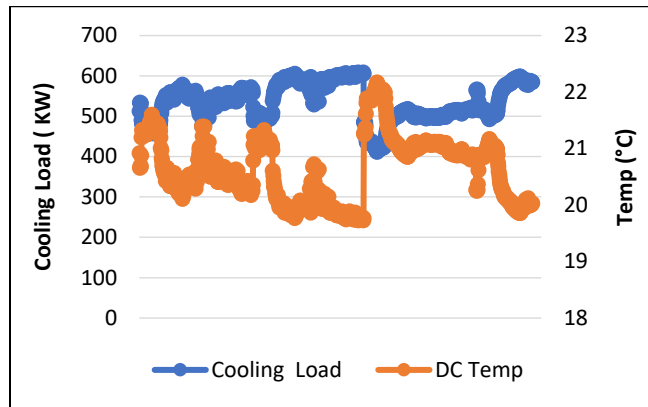

Figure 9. Predicted data DC cooling load vs temp.

Further, the same pattern that is expected out of the relationship between Total Data Center Load and PUE is seen. Data Center total load is a combination of DC Cooling Load & IT load which was compared earlier with DC Cooling Load vs Temperature with an increase in change in temperature. This also results in a change in PUE because it directly impacts the Data Center Total load. Both the results have been compared which also validates our finding that Data Center load decreases on the basis of a decrease in Data Center Cooling Load. Figure 10 & 11 gives the data representation of Actual & predicted data relationship between PUE and DC Total Load.
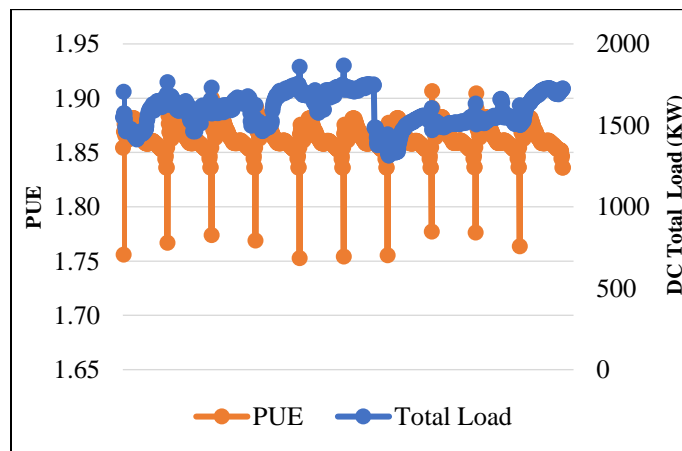

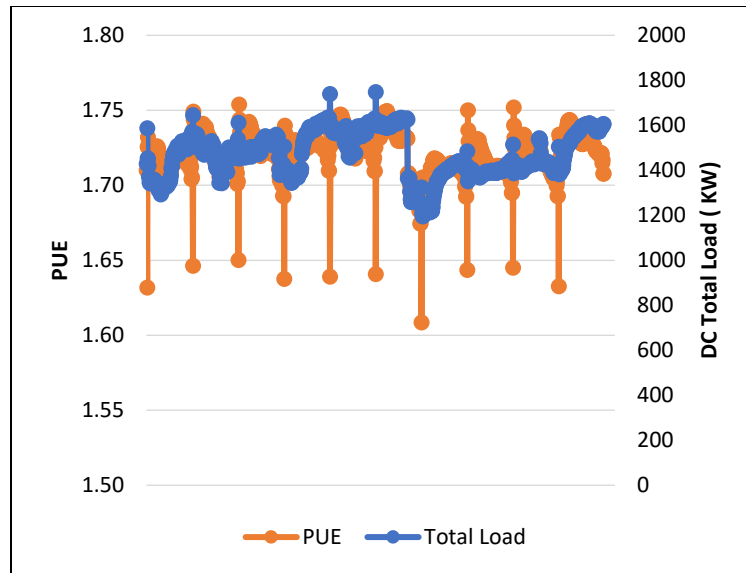**Figure 10.** Actual data PUE vs DC total load.

Figure 11. Predicted data PUE vs DC total load.

Basis the above data analysis, it is inferred that the PUE of the Data Center depends on the two major components which are IT Load and Cooling Load. Data Center Cooling Load depends upon the cooling load of the infrastructure which is directly and indirectly coupled with Data Center Temperature as maintained within the Data Center. Out of the ensuing research and machine learning model, the Data Center efficiency ratios were identified and validated based on the Data Center PUE change factors. However, this research has focused primarily on temperature, where the PUE was reduced with the marginal increase in temperature.

8. Conclusions
PUE is one of the important metrics for assessing the total energy efficiency of a Data Center. In this paper, the temperature of the Data Centers was focussed on, and an archetype was created to have a slight increase from 1.5 deg. to 2.5 deg. above the existing temperature maintained in Data Center. This was modelled with the help of linear regression, which is a supervised machine learning approach. When the heat increases within the allowable limits and guidelines of the Uptime Institute and ASHRAE, the PUE decreases. As the temperature is raised in the range of 19.73 °C to 21.17 °C then Cooling Load is decreased in the range from 607 KW to 414 KW.
This not only helps in optimising the energy efficiency of the Data Center, but it also in turn can help to save the substantial cost of the power since it directly helps in the reduction of power consumed by the cooling systems. Moreover, the study can be expanded in the future for a broader range of historical data in a geo-based Data Center to further evolve into a new adaptive approach in power & cooling management. It shall encompass IT loads and delve deeper into the cooling infrastructure apart from managing the higher denomination of temperature within the limits as specified and desired for the IT Equipment.

Conflict of Interest
The authors confirm that there is no conflict of interest for this publication.

References

Azimzadeh, A., & Tabrizi, N. (2015). A taxonomy and survey of green data centers. In *2015 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 128-131). IEEE. Las Vegas, NV, USA. DOI: 10.1109/CSCI.2015.70.

Baccour, E., Foufou, S., Hamila, R. & Erbad, A. (2019). Green data center networks: a holistic survey and design guidelines. In *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*. June 2019, IEEE, Tangier, Morocco, 1108-1114. DOI: 10.1109/IWCMC.2019.8766705

Capozzoli, A., & Primiceri, G. (2015). Cooling systems in data centers: state of art and emerging technologies. *Energy Procedia*, *83*, 484-493. DOI: 10.1016/j.egypro.2015.12.168.

Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, *13*(11), 114003.

Daraghmeh, H.M., & Wang, C.C. (2017). A review of current status of free cooling in datacenters. *Applied Thermal Engineering*, *114*, 1224-1239. DOI: 10.1016/j.applthermaleng.2016.10.093.

Delforge, P., & Whitney, J. (2014). Data center efficiency assessment. *Natural Resources Defense Council (NRDC)*, (August), 35.

Elmasry, H.M., Khedr, A.E., & Nasr, M.M. (2019). An adaptive technique for cost reduction in cloud data center environment. *International Journal of Grid and Utility Computing*, *10*(5), 448-464.

Fouladi, K., Wemhoff, A.P., Silva-Llanca, L., Abbasi, K., & Ortega, A. (2017). Optimization of data center cooling efficiency using reduced order flow modeling within a flow network modeling approach. *Applied Thermal Engineering*, *124*, 929-939. DOI: 10.1016/j.applthermaleng.2017.06.057.

Fulpagare, Y., & Bhargav, A. (2015). Advances in data center thermal management. *Renewable and Sustainable Energy Reviews*, *43*, 981-996. DOI: 10.1016/j.rser.2014.11.056.

Ganesh, L., Weatherspoon, H., Balakrishnan, M., & Birman, K. (2007). Optimizing power consumption in large scale storage systems. In: *HotOS*. 2007. https://www.usenix.org/legacy/events/hotos07/tech/full_papers/ganesh/ganesh_html/index.html

Gao, J. (2014). *Machine learning applications for data center optimization*. https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/42542.pdf

GeSI. (2020). *Enabling the low carbon economy in the information age*. 2008. London, UK, 2020.

Jin, C., Bai, X., & Yang, C. (2019). Effects of airflow on the thermal environment and energy efficiency in raised-floor data centers: a review. *Science of the Total Environment*, *695*, 133801. DOI: 10.1016/j.scitotenv.2019.133801.

Kaylie, G. (2020). *What you should know about data center cooling technologies*. 2020. https://www.vxchnge.com/blog/data-center-cooling-technology.

Koronen, C., Åhman, M., & Nilsson, L.J. (2020). Data centers in future European energy systems-energy efficiency, integration and policy". *Energy Efficiency*, *13*(1), 129-144. DOI: 10.1007/s12053-019-09833-8.

Linesight. (2019). *Energy consumption in data centers; the drive to lower PUEs*. https://www.linesight.com/insights/energy-consumption-in-data-centers-the-drive-to-lower-pues/

Marcinichen, J.B., Olivier, J.A., & Thome, J.R. (2012). On-chip two-phase cooling of data centers: cooling system and energy recovery evaluation. *Applied Thermal Engineering*, *41*, 36-51. DOI: 10.1016/j.applthermaleng.2011.12.008.

Ni, J., & Bai, X. (2017). A review of air conditioning energy performance in data centers. *Renewable and Sustainable Energy Reviews*, *67*, 625-640. DOI: 10.1016/j.rser.2016.09.050.

Niemann, J., Brown, K., & Avelar, V. (2011). Impact of hot and cold aisle containment on data center temperature and efficiency. *Schneider Electric Data Center Science Center, White Paper*, *135*, 1-14.

Oró, E., Depoorter, V., Garcia, A., & Salom, J. (2015). Energy efficiency and renewable energy integration in data centers. strategies and modelling review. *Renewable and Sustainable Energy Reviews*, *42*, 429–445. DOI: 10.1016/j.rser.2014.10.035.

Patterson, M.K. (2008). The effect of data center temperature on energy efficiency. In *2008 11th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems* (pp. 1167-1174). IEEE. Orlando, FL, USA. DOI: 10.1109/ITHERM.2008.4544393.

Rong, H., Zhang, H., Xiao, S., Li, C., & Hu, C. (2016). Optimizing energy consumption for data centers. *Renewable and Sustainable Energy Reviews*, *58*, 674-691. DOI: 10.1016/j.rser.2015.12.283.

Schmidt, R.R., & Iyengar, M. (2007). Best practices for data center thermal and energy management- review of literature. *ASHRAE Transactions*, *113*(1), 206-218.

Sharma, M., Arunachalam, K., & Sharma, D. (2015). Analyzing the data center efficiency by using PUE to make data centers more energy efficient by reducing the electrical consumption and exploring new strategies. *Procedia Computer Science*, *48*, 142-148. DOI: 10.1016/j.procs.2015.04.163.

Shoukourian, H., Wilde, T., Labrenz, D., & Bode, A. (2017). Using machine learning for data center cooling infrastructure efficiency prediction. In *2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)* (954-963). IEEE. Lake Buena Vista, FL, USA. DOI: 10.1109/IPDPSW.2017.25.

Singh, V.K., & Guo, J. (2016). Optimization of power usage effectiveness for heterogenous modular data centers using neural network. In *The Sixth International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies*, 27-32. Held on 26-30 June 2016, Lisbon, Portugal. Held at InfoSys 2016.

Stephen, I., & Lucy, B. (2020). Cloud-based approach in data centers". In: *Cloud Control Systems*, (405-444). Elsevier. DOI: 10.1016/B978-0-12-818701-2.00022-6.

Taheri, S., Goudarzi, M., & Yoshie, O. (2020). Learning-based power prediction for geo-distributed data centers: weather parameter analysis. *Journal of Big Data*, *7*(1), 1-16. DOI: 10.1186/s40537-020-0284-2.

TC, A. (2016). Data center power equipment thermal guidelines and best practices. *ASHRAE TC 9.9, ASHRAE, USA*.

Thiangchanta, S., & Chaichana, C. (2020). The multiple linear regression models of heat load for air-conditioned room. *Energy Reports*, *6*, 972-977. DOI: 10.1016/j.egyr.2020.11.090.

Uzaman, S.K., Khan, A.R., Shuja, J., Maqsood, T., Rehman, F., & Mustafa, S. (2019). A systems overview of commercial data centers: initial energy and cost analysis. *International Journal of Information Technology and Web Engineering*, *14* (1), 42-65.

Wang, J., Zhang, Q., Yu, Y., Chen, X., & Yoon, S. (2018). Application of model-based control strategy to hybrid free cooling system with latent heat thermal energy storage for TBSs. *Energy and Buildings*, *167*, 89-105. DOI: 10.1016/j.enbuild.2018.02.036.

Yuventi, J., & Mehdizadeh, R. (2013). A critical analysis of power usage effectiveness and its use in communicating data center energy consumption. *Energy and Buildings*, *64*, 90-94. DOI: 10.1016/j.enbuild.2013.04.015.

Zhang, X., VanGilder, J.W., Iyengar, M., & Schmidt, R.R. (2008). Effect of rack modeling detail on the numerical results of a data center test cell. In 2008 *11th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems* (pp. 1183-1190). May 2008. IEEE. Orlando, FL, USA. DOI: 10.1109/ITHERM.2008.4544395.