**Control of Direct Air Free Cooled Data Centers in Sub-Tropical Zones Using Machine Learning Decision Support System**

Dissertation Submitted in part fulfilment of the requirements for the degree of MSc in Data Analytics

From



By

**Ravi Teja Mannuru**

Student ID: 10540730

MSc. in Data Analytics 2020 -2021

**Declaration**

I, Ravi Teja Mannuru, declare that this is my original work and the same has not been submitted to any other educational institution for a Degree award. I have cited all the sources and provided the corresponding references regarding literature used in my work and can confirm that this work is complaint with the Dublin Business School’s academic policies.

Signature: Ravi Teja Mannuru

Date: 11/01/2021

**Acknowledgement**

I would like to express my sincere gratitude to my Dissertation supervisor Professor Muhammad Farooq for guiding me at each stage through out the process. Professor Muhammad provided critical feedback on my work and helped me ensure the quality in thesis planning, formulation, artefact creation, referencing and documentation. Professor Muhammad conducted weekly online review meetings and additional meetings to help me address any unexpected issues or concerns I faced in the course of this dissertation. Also, Professor Muhammad provided valuable suggestions regarding the various tools, techniques, and technologies that I have incorporated in my work.

I would also like to express my thanks to Dr. David Williams whose Applied Research Process concepts and in particular the research methodologies, helped me come up with the research concept and progress accordingly with dissertation. I am very much grateful to Dr. Shazia Afzal, Dr. Shahram Azizi Sazi, Dr. Terri Hoare, Dr. Abhishek Kaushik, and Professor Kunwar Madan for providing me with the necessary knowledge on the technologies, tools and techniques in data storage, data processing, programming, statistical analysis, data mining, machine learning methodologies and data visualization concepts.

I would also like to express my gratitude to MeteoBlue organisation that helped me with the historical meteorological data simulation and collection for the selected geographical locations. Also, I am very much grateful to DBS for supporting and enabling students to carryout their academic activities seamlessly during the Covid-19 pandemic times.

Finally, I would like to show appreciation to my spouse and family for supporting me through out the study programme and the dissertation process.

Thank you.

Ravi Teja Mannuru.

**Abstract:**

Numerous green initiatives are being used for the greening of the datacenter (DC) creation and maintenance. Cooling is one of the major power consuming areas of a DC and this paper concentrates on reducing the carbon footprint associated with this component by employing natural direct air free cooling economizer cycle that operates in coordination with the DC’s conventional cooling systems with help of an novel control system. As a result, the cooling associated greenhouse effect will be reduced. The applicability and reliability of airside free cooling is dependent on the location of the DC and the climatic conditions associated with it. This requires an effective Decision Support System (DSS) to ensure seamless switching between the free and conventional cooling solutions for the optimized and reliable operation of the DC. The meteorological data for past 20 years will be collected and ML algorithms will be used to produce a DSS model to control the multi-mode economizer based cooling system of the DC. The DSS will ensure a proactive and seamless switching between different cooling modes eliminating the transition times involved between these modes to the maximum possible extent. The effectiveness of the proposed system will be validated and hypothetical quantification of the reduction in power consumption and associated reduction in carbon footprint is ascertained in the study.

**Key Words:**

Data Centers, Decision Trees, Decision Support System, Machine Learning, Random Forest Regression, Visualizations, Green cloud Computing, Multi target regression

**Special Terms:**

DC – Data Center

IT – Information Technology Equipment

PUE – Power Usage Efficiency.

DASFC – Direct Air Side Free Cooling.

IDASFC – Indirect Air Side Free Cooling.

AI – Artificial Intelligence

ML – Machine Learning

DL – Deep Learning

DRL – Deep Reinforcement Learning

NN – Neural Networks

ASHRAE – American Society of Heating, Refrigerating and Air-Conditioning Engineering

TES – Thermal Energy Storage

SLA – Service Level Agreement

DBT – Dry Bulb Temperature

WBT – Wet Bulb Temperature

DPT – Dew Point Temperature

DX – Direct Expansion

AHU – Air Handling Unit

GDBT – Gradient Boosted Trees

DT – Decision Tree

RF – Random Forest

MSE – Mean Squared Error

RMSE – Root Mean Squared Error

MAE – Mean Absolute Error

MAPE – Mean Absolute Percentage Error

**Contents**

[**Chapter 1 – Introduction** 11](#_Toc61123832)

[1.1 Background 11](#_Toc61123833)

[1.1.1 Air Side Economizer: 15](#_Toc61123834)

[1.1.2 Choice between DASFC and IDASFC 17](#_Toc61123835)

[1.2 Problem domain and Solution Concept 18](#_Toc61123836)

[1.2.1 Research Questions 20](#_Toc61123837)

[1.2.2 Research Objectives 21](#_Toc61123838)

[1.3 Key Factors 21](#_Toc61123839)

[1.4 Dissertation Structure 23](#_Toc61123840)

[**Chapter 2: Literature Review** 25](#_Toc61123841)

[2.1 Green cloud computing in DC 25](#_Toc61123842)

[2.2 Conventional energy efficient practices in DCs 28](#_Toc61123843)

[2.3 Existing techniques and emerging technologies for cooling DCs 31](#_Toc61123844)

[2.3.1 Running the DCs hot 37](#_Toc61123845)

[2.4 Machine Learning for Green Computing in DCs 38](#_Toc61123846)

[2.5 Machine Learning Algorithms used in the Study 39](#_Toc61123847)

[2.5.1 Decision Tree Algorithm 39](#_Toc61123848)

[2.5.2 Random Forest Algorithm 41](#_Toc61123849)

[2.5.3 Gradient Boosted Trees 44](#_Toc61123850)

[2.5.4 Other Models 46](#_Toc61123851)

[**Chapter 3 – Methodology and Artefact Design** 47](#_Toc61123852)

[3.1 Introduction 47](#_Toc61123853)

[3.2 ML model for Multi Target Prediction 47](#_Toc61123854)

[3.2.1 Data Collection 47](#_Toc61123855)

[3.2.2 Technology Stack used for the creation of the Model 48](#_Toc61123856)

[3.2.3 Data Exploration and Pre-processing 48](#_Toc61123857)

[3.2.4 Selection of the ML model 52](#_Toc61123858)

[3.3 DSS system for cooling control 55](#_Toc61123859)

[**Chapter 4 – Results and Analysis** 59](#_Toc61123860)

[4.1 RF Model Performance Evaluation 59](#_Toc61123861)

[4.2 Performance evaluation of other algorithms considered 60](#_Toc61123862)

[4.2.1 Scikit Regression Models – Linear, KNN, SVR, DT and GBT 60](#_Toc61123863)

[4.2.2 Gradient Boosted Trees – LightGBM 67](#_Toc61123864)

[4.3 Evaluation of the effectiveness of the control system 68](#_Toc61123865)

[**Chapter 5 – Conclusion and Future Work** 70](#_Toc61123866)

[**References** 72](#_Toc61123867)

[**APPENDIX** 77](#_Toc61123868)

[PUE Calculation for Conventional Cooling System 77](#_Toc61123869)

[PUE Calculation for Novel DSS based Controller 77](#_Toc61123870)

**List of Figures**

[**Figure 1.** Power Consumption Distribution in DC 11](#_Toc61097540)

[**Figure 2.** Power Usage Among Conventional Cooling System Components 12](#_Toc61097541)

[**Figure 3.** Direct Air Free Cooling Airflow 14](#_Toc61097542)

[**Figure 4.** Indirect Airside Free Cooling Schematic 15](#_Toc61097543)

[**Figure 5.** Techniques for establishing a Green DC 25](#_Toc61097544)

[**Figure 6.** Environmental classes from ASHRAE for DC economizers 30](#_Toc61097545)

[**Figure 7.** Worldwide Climatic Zone Classification 31](#_Toc61097546)

[**Figure 8.** Percentage Energy Savings with DSAFC systems in worldwide climate zones 32](#_Toc61097547)

[**Figure 9.** Airside Free Cooling Implementation Schemes 33](#_Toc61097548)

[**Figure 10.** Impact of Raising Inlet Temperatures in DC 35](#_Toc61097549)

[**Figure 11.** Decision Tree Hyperparameters 38](#_Toc61097550)

[**Figure 12.** Decision Tree Algorithm Example 38](#_Toc61097551)

[**Figure 13.** Random Forest Hyperparameters 40](#_Toc61097552)

[**Figure 14.** Random Forest Algorithm Structure 41](#_Toc61097553)

[**Figure 15.** Extending Multi Target Support to Gradient Boosted Trees (LightGBM Model) 43](#_Toc61097554)

[**Figure 16.** Extending Multi Target Support to Gradient Boosted Trees (Scikit Learn Model) 43](#_Toc61097555)

[**Figure 17.** Level (Depth) wise growth of tree 44](#_Toc61097556)

[**Figure 18.** Leaf wise growth of tree 44](#_Toc61097557)

[**Figure 19.** Dataset column list 47](#_Toc61097558)

[**Figure 20.** Correlation Heat Map 48](#_Toc61097559)

[**Figure 21.** Feature Importance Estimation 49](#_Toc61097560)

[**Figure 22.** Temperature Prediction MAPE 51](#_Toc61097561)

[**Figure 23.** RH Prediction MAPE 52](#_Toc61097562)

[**Figure 24.** Flowchart for Cooling Control Scheme 54](#_Toc61097563)

[**Figure 25.** Flowchart for DSS Logic 55](#_Toc61097564)

[**Figure 26.** Monthly Average Temperature and Relative Humidity Levels 57](#_Toc61097565)

[**Figure 27.** Predicted Vs Test Scatter Plots for Temperature and RH (Random Forest Model) 58](#_Toc61097566)

[**Figure 28.** Prediction Performance of the model for Temperature and RH (RF Model**)** 59](#_Toc61097567)

[**Figure 29.** Predicted Vs Test Scatter Plots for Temperature and RH (Linear Regression Model) 60](#_Toc61097568)

[**Figure 30.** Prediction Performance of the model for Temperature and RH (Linear Regression Model) 60](#_Toc61097569)

[**Figure 31.** Predicted Vs Test Scatter Plots for Temperature and RH (KNN Regression Model) 61](#_Toc61097570)

[**Figure 32.** Prediction Performance of the model for Temperature and RH (KNN Regression Model) 61](#_Toc61097571)

[**Figure 33.** Predicted Vs Test Scatter Plots for Temperature and RH (SVR- Multioutput Wrapper) 62](#_Toc61097572)

[**Figure 34.** Prediction Performance of the model for Temp and RH (SVR- Multioutput Wrapper) 62](#_Toc61097573)

[**Figure 35.** Predicted Vs Test Scatter Plots for Temperature and RH (SVR- Regressor Chain) 63](#_Toc61097574)

[**Figure 36.** Prediction Performance of the model for Temperature and RH (SVR – Regressor Chain) 63](#_Toc61097575)

[**Figure 37.** Predicted Vs Test Scatter Plots for Temperature and RH (DT) 64](#_Toc61097576)

[**Figure 38.** Prediction Performance of the model for Temperature and RH (DT) 64](#_Toc61097577)

[**Figure 39.** Predicted Vs Test Scatter Plots for Temperature and RH (GBT) 65](#_Toc61097578)

[**Figure 40.** Prediction Performance of the model for Temperature and RH (GBT) 65](#_Toc61097579)

[**Figure 41.** Predicted Vs Test Scatter Plots for Temperature and RH (GBT - LightGBM) 66](#_Toc61097580)

[**Figure 42.** Prediction Performance of the model for Temperature and RH (GBT – LightGBM) 66](#_Toc61097581)

**List of Tables**

[**Table 1.** Local and Global Impacts of DCs on Environment 26](#_Toc61098851)

[**Table 2.** ASHRAE environmental operating classes for DCs 33](#_Toc61098852)

[**Table 3.** External air temperature and enthalpy measurements in worldwide climate zones 34](#_Toc61098853)

[**Table 4.** Metrics for Model comparison 52](#_Toc61098854)

[**Table 5.** Sample Predictions Using different models 54](#_Toc61098855)

[**Table 6.** Power Ratings of DASFC components 68](#_Toc61098856)

[**Table 7.** Hourly distribution of the supporting system involvement over 365 days 69](#_Toc61098857)

[**Table 8.** Min-Max values predicted for T and RH over 365 days 69](#_Toc61098858)

# **Chapter 1 – Introduction**

The general introduction to the problem domain, research question, objectives and other research considerations are outlined in this chapter.

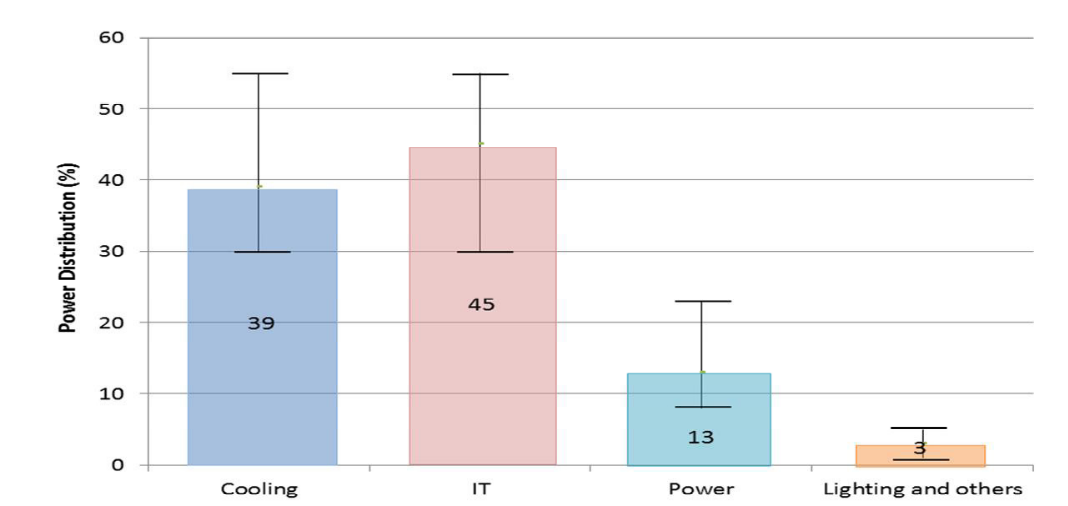
## **1.1 Background**

Cloud computing offerings are evolving on an enormous scale supported by the rapid innovation in the field. The increasing demand for computing resources and the data storage as well as processing needs at various levels has been attracting massive customer base for the cloud computing platforms ranging from large-scale industries to an individual user. This is because of the flexibility, reliability, scalability, security, and cost structure involved in the cloud platform. The multitude of cloud services have DC at their core. The ever-growing technology, services and devices in turn uses and generates more data which creates an exponentially increasing demand for the upgradation of the existing DCs and establishment of the new DCs. These DCs serve the live business models that drive the modern world and are expected to have the highest possible levels of reliability and efficiency all the while maintaining the least amount of downtime in line with the stringent service level agreements of the respective industry or customer segments they are serving (Maurya and Sinha, 2013).

The substantial increase in the demand for the new DCs aggravates the effect on the environment (Uddin, Shah and Memon, 2014). To ensure a sustainable cloud computing future, the environmental effects caused by the DCs must be controlled to the maximum possible extent. The eco-friendly DC infrastructure establishment is encouraged by Green cloud computing practices (Goyal, Arya and Nagpal, 2016). An analysis on the climate change and associated effects from the business and IT sector perspectives in (Agarwal and Nath, 2011) talks about the need to implement the best practices in the Data Center expansions.

As part of the green cloud computing initiatives, several improvement areas are identified through research in academia for reducing the power consumption of the IT equipment thereby reducing the effect on the environment proportionately (Maurya and Sinha, 2013) (Agarwal and Nath, 2011) (Uddin, Shah and Memon, 2014) (Nixon, Francis and Devaraj, 2015) (Buyya, Beloglazov and Abawajy, 2010) (Suresh, 2015). While this can be effective, an even greater result can be achieved by reducing the cooling related power needs of the DCs in addition to the optimization of the DC equipment and Load Management. The improvements in the software and hardware coupled with the changes in the designs of the data center cooling systems will help in the reduction of the power consumption and ultimately the associated emissions (Woods, 2010).

The approximate distribution of the power consumption across the major segments in a data center is illustrated in fig. 1 (Song, Zhang and Eriksson, 2015). It can be seen that the cooling needs of the data center amounts to around 40 percent of the total power consumption due to the usage of conventional cooling systems. This value could reach higher numbers in case of data centers located in climatic zones prone to higher temperatures which include the tropical and sub-tropical regions along the equator (Van Le *et al.*, 2019)(Lee and Chen, 2013).

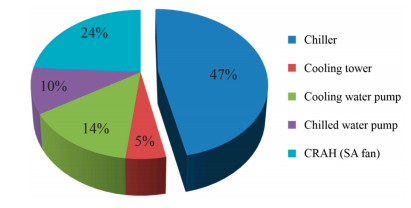


**Figure 1. Power Consumption Distribution in DC** (Song, Zhang and Eriksson, 2015)

The excess of power consumption by the conventional cooling systems is attributed to the following factors (Zhang *et al.*, 2014) :

1. The continuous running of the conventional mechanical cooling equipment irrespective of the meteorological conditions surrounding the DC.
2. The power consumption corresponding to the auxiliary cooling support systems which includes the fans, pumps and the coolant supply or return systems.
3. Design issues leading to the leakage or cooling containment failure. As a result, the cooling demand is further increased to compensate for the said losses which increases power consumption.

Among the factors discussed above, the mixing of the cold-hot streams and the losses in the piping system can be addressed through design and maintenance considerations. But the major power consumption part of the conventional system attributed to the chillers, air compressors, pumps and related control equipment must be addressed to achieve maximum reduction in PUE of the DC. The power consumption among the traditional cooling system components is shown in fig. 2.



**Figure 2. Power Usage Among Conventional Cooling System Components** (Park and Seo, 2018)

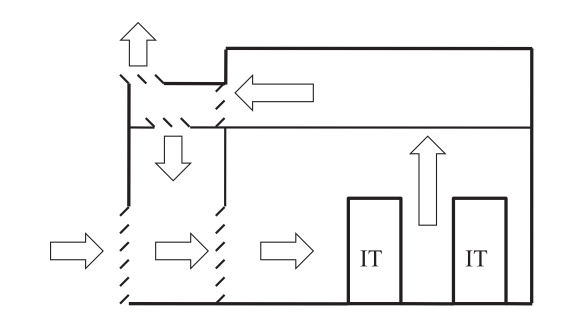
The power consumption of the conventional cooling equipment can be reduced by using Natural Free Cooling solutions as economizer cycles in the cooling scheme of a DC. Depending on the location, load on the DC and the meteorological factors around the DC, the economizer cycles can be run independently or in coordination with the existing refrigerant based mechanical cooling options to achieve the desired cooling effect. Free Cooling is a relatively new area that has great potential to reduce the data center’s cooling related green house effect on the environment. The economizer cycles comprising of the free cooling solutions help in reducing the cooling related power consumption in the DCs by making use of the ambient cooling of the natural coolant medium which includes water and air. As part of this study, air side economizer cycles are considered for improving the PUE while simultaneously reducing the environmental impact of the DCs.

The cooling demand for the DCs with conventional cooling systems operating in tropical regions is comparatively higher than that of the DCs in colder regions leading to higher carbon emissions. So, implementation of economizer cycles in the DC establishments in these regions is a viable solution to reduce the emissions.

### **1.1.1 Air Side Economizer:**

Air side economizer cycle involves the pumping of cold outside air into the DC through a series of dampers, filters, and humidifiers. The lower enthalpy of the colder air helps in cooling the IT equipment of the DC. This economizer is implemented in two variants as follows

#### **Direct Air Side Free Cooling (DASFC)**

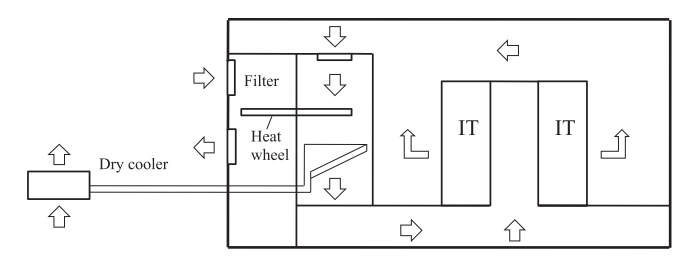


**Figure 3. Direct Air Free Cooling Airflow** (Zhang et al., 2014)

DASFC system involves the flow of surrounding air directly into the DC through a system of fans, filters, humidity controllers, and pressure regulators when the temperature, RH, and other criteria are within the setpoints for DC operation. DASFC system exposes the DC equipment to contaminants, particulate matter and humidity variations imposed by the outside air humidity levels. But the efficiency of cooling is also high in this case as the ambient temperature is least which ensures more heat transfer from the server equipment into the coolant. Also, the contamination and humidity factors can be regulated using appropriate filters and humidity controlling equipment.

A schematic for a typical DASFC system is shown in fig 3. This configuration introduces the cold outside air into the DC through the filtering, temperature, and humidity control systems. The ambient temperature of the inlet air is used to cool the IT equipment which is then directed outside the DC through the exhaust system. In this case a part of the exhaust air is considered for recirculation which helps in maintaining the DC setpoints especially when lower temperatures and higher RH levels are involved.

#### **Indirect Air side Free Cooling (IDASFC)**



**Figure 4. Indirect Airside Free Cooling Schematic** (Zhang et al., 2014)

The indirect airside free cooling mechanism is shown in fig. 4. This method also uses the ambient temperature of the outside air to cool the DC equipment but in this case the DC environment is completely isolated from the external influences like humidity variations, dust, particulate matter, and other contaminants. A heat exchanger (like a heat wheel as shown in fig. 4.) will be used to cool the air circulating in the DC. This does involve more mechanical components but eliminates the need of additional filters or humidity controllers needed to condition the air at the inlet as in the case of direct air side cooling.

### **1.1.2 Choice between DASFC and IDASFC**

There are several factors that dictate the choice of direct or indirect air cooled economizers in the DC (Lin, Niemann and Long, 2015). These primarily include initial capital and maintenance expenses, location of the DC, availability risks and overall benefits of the intended system. The implementation and maintenance of DASFC economizers generally involves lower cost and easier maintenance in contrast to the IDASFC. Although, the indirect air side cooling makes it easier to maintain a stable environment inside the DC, the heat transfer efficiency is considerably lower compared to that of the DASFC. This is the driving factor for considering the DASFC based economizer solution in this dissertation. Heat transfer losses at any scale will have severe implications in the case of DCs in the selected sub tropical regions where the temperature fluctuations are expected more often than usual compared to other regions. In addition to this, the design considerations, the ability of the existing DCs to be retrofitted with new economizer modes, and the ability of the IT equipment in the DC to withstand the external factors are also deciding factors in selecting the required economizer mode.

In the recent times the manufacture of server rack components and associated equipment has improved on a large scale and most of the IT equipment is being replaced or retrofitted to be complaint with the ASHRAE and GreenGrid recommendations. The temperature and RH operating conditions for the DC are relieved in the ASHRAE’s 2011 recommendations for DCs

(Zhang *et al.*, 2014). This makes it possible for the free cooling options to be implemented in DCs established in hot and humid areas which was not a feasible option a decade ago. So, DASFC has been selected as the go to option for implementing the economizer mode in the DCs for the purpose of this research.

## **1.2 Problem domain and Solution Concept**

The reliability and performance of the DC equipment depends on the cooling efficiency of the DC. It is observed that the thermal attack has a linear degrading impact on the chip (Gao *et al.*, 2017)(Chung, Kalbarczyk and Iyer, 2018) causing leakage power which affects the efficiency, life-cycle and in worse situations triggers a failure of the components. This is a critical consideration when Natural Free Cooling options are involved in the cooling schema of the DC. The main concern behind this is that the ambient temperatures and relative humidity levels of air around the DC keeps varying depending on the meteorological factors. This makes it very difficult to implement these economizer cycles in selected sub tropical region as the varying levels of temperature and relative humidity may not always satisfy the cooling needs of the DC. So, these economizer cycles should be operated in coordination with a supporting systems and backup cooling mechanism to ensure maximum reliability of the cooling system.

There are several modes of operating a DASFC which may or may not involve standby systems that will need to operate in coordination with the economizer to ensure reliable cooling in the DC (Niemann, Bean and Avelar, 2013). This research considers one such approach where the DASFC economizer is supported by a vapour pressure chamber and off-peak TES (Oró *et al.*, 2015)system for immediate backup. A refrigerant based mechanical cooling option will be used in ca**s**e of persistent backup requirements. The primary concern of incorporating a DASFC economizer in a tropical DC is the control system that is needed to make a smooth transition of the cooling mode from economizer mode to backup or mechanical cooling mode depending on the meteorological conditions around the DC. The control system should also be able to maintain the supporting systems for enabling maximum possible DASFC operational hours.

The traditional control systems use sensors placed inside and outside the DC in order to determine the point of engagement of the required cooling mode including the economizer. These sensors monitor the static pressure, temperature, and humidity level in the DC and at the inlet for DASFC economizers. But the time period involved in switching from one cooling mode to the other depends on the immediate cooling requirement of the DC equipment, especially when a cold start of the mechanical cooling mode is needed in cases where the economizer mode and TES systems are incapable of serving the cooling demand of the DC (Oró *et al.*, 2015). If the necessary cooling is not available during the transition period, it will have damaging effects on the IT equipment.

The cooling system mode transition times therefore play a vital role in the reliability of the DC equipment. The existing sensor driven control schemes are suitable for a single mode of cooling method. But when multiple modes such as economizers, backup and mechanical systems are involved, the sensor based systems would need additional feedback about the future meteorological factors in order to proactively schedule the cooling modes to maintain the reliability of the DCs. These meteorological factors primarily include Temperature and RH which dictate the feasibility of operation of DASFC economizer modes.

Artificial Intelligence can be used to address the concerns in the control schema of the DC when economizers modes are involved. Artificial Intelligence deals with the science of creating machines, computers and processes that can replicate the thinking and decision making capabilities similar to that of a human (Oracle EU, 2020). Machine Learning and Deep Learning are among the subsets of AI. Machine Learning uses least programming or external interference. It uses statistical tools and methods for the learning and training of a computer system or process on a given set of data (IBM Cloud Education, 2020). The models obtained from this training will be used for prediction, classification and producing inferences from the data patterns. Deep Learning is a branch of machine learning that deals with the learning process using a neural network approach which includes artificial neuron like structures that can simulate human like analysis and decision making capabilities. The learning process of deep learning is experience centered (Oracle EU, 2020). The data undergoes non-linear transformation under multiple neural network layers and the algorithm self adjust in order to get the desired prediction or classification model and for clustering the data based on observed patterns.

A multi target regression model will be built using ML techniques to predict the temperature and RH factors from the meteorological features associated with the selected DC location. These predictions will be used in the generation of a scheduling scheme in advance which will be used in the novel control scheme.

### **1.2.1 Research Questions**

What is the effect of the Data Center related cooling power needs on the environment and how can this be reduced?

How can we use a Machine Learning based Decision Supporting System (DSS) to control a Direct Air Free Cooled Data Center?

### **1.2.2 Research Objectives**

The primary aim of this project is to promote the Green Cloud Computing Practices in DC industry operation, maintenance, and expansion. This is achieved by creating an efficient ML model that can be used in the cooling control scheme of a sub-tropical DC to ensure smooth and quick transition times between different cooling modes when DASFC economizer cycles are involved.

A suite of supervised machine learning algorithms will be applied to process the historical meteorological data and train the AI models which will predict the temperature and RH variations throughout a given day in hourly intervals. The best models will be deployed into the control scheme of the DC for the effective control of multimode cooling. Also, a comparative analysis will be performed for ascertaining the power savings and a simultaneous reduction in the carbon footprint. A novel control scheme flow for controlling the switching scheme of the cooling modes in the DC will be proposed and the quantification of the effectiveness of this system will be performed using a hypothetical DC.

## **1.3 Key Factors**

The unprecedented levels of increase in telecommunication technologies, devices, and associated increase in the demand for data processing and storage needs coupled with the advantages of computing services offered by the Cloud Computing platforms resulted in an immense demand for the DC infrastructure expansion. But this comes at the cost of the environment in the form of increased carbon footprint and electronic waste. Cooling is an essential factor of running a datacenter and is a major area of power consumption amounting to around 40 to 50 percent of the total power consumption depending the location and load on the data center. Among several innovations, free cooling is an important and relatively new concept that when implemented can provide disruptive solutions to support the cooling needs of the data centers and reduce the carbon emissions accordingly. By employing free cooling techniques, the ambient temperatures of the air or water surrounding the datacenter can be used for cooling the datacenter equipment to the maximum possible extent there by reducing the power consumption of the conventional power-hungry mechanical cooling mechanisms.

Free Cooling is relatively easier to implement and is more reliable in the case of colder regions. But when it comes to tropical regions, the effectiveness of the free cooling is negated by comparatively higher temperature and humidity levels which are fairly closer or above the temperature or the humidity thresholds of the data center equipment. But recent changes in the manufacturing procedures involved in making of datacenter equipment complaint to ASHRAE recommendations that includes several tropical zones (The Green Grid, 2012) (Lee and Chen, 2013) as feasible locations for implementing Air Side Economizers. These equipment can handle higher temperatures and humidity levels which means the datacenters can be run at higher temperature levels (Zhang *et al.*, 2014) without causing any damage or reduction in the efficiency of the equipment. This provided an avenue for the implementation of the free cooling techniques in the tropical areas as well. In fact, the cooling related power consumption is more in tropical areas than that of other regions as was published in some studies (Niemann, Bean and Avelar, 2013). So, implementation of free cooling technologies in tropical regions will have an even greater benefit future growth of DCs.

DASFC technology uses the ambient temperature of the air surrounding the data center to cool the data center equipment. Fan and filter-based ventilation mechanisms with humidity regulators will be used in the process to circulate the air. This free cooling will be used in conjunction to the existing conventional cooling mechanisms. The primary issue with the free cooling is the uncertainty of the meteorological conditions which needs an effective system that can guide the cooling scheme to operate accordingly for engaging the appropriate cooling mechanism to maintain the efficient and safe operation of the data center.

A ML model will be trained based on the selected DC location’s meteorological data to predict the future temperature and humidity profiles for the upcoming day and an associated switching scheme will be used to engage the different cooling mechanism for proactive cooling of the DC. The primary target of the study is to provide a system that can efficiently switch over to the required cooling mechanism as needed with emphasis on the reduction in carbon emission to the maximum extent possible.

## **1.4 Dissertation Structure**

The subsequent chapters in this dissertation module include Literature Review, Methodology and Artefact Creation, Results – Analysis, Conclusion and Future Work sections in the same order.

The Literature Review section will be concentrated on multiple concepts related to the Data Center Impact on the Environment, green cloud computing initiative research in the academia and industry, air side ecGreonomizer practical considerations and implementations, analysis of worldwide implementations of DASFC systems in different climatic zones, details of machine learning algorithms considered for the research and the various cooling schemes involving economizer operation.

The Methodology chapter will comprise the proposed implementation of the machine learning algorithms for creating the temperature and RH prediction model followed by the implementation of this model in the novel control scheme for controlling the multi mode cooling system with economizers and mechanical backups. Various technologies and techniques will be used for the creation of corresponding artefacts.

Result – Analysis chapter ascertains the effectiveness of the machine learning model implementation in the cooling control scheme of the DC. A detailed review of the reduction in the carbon footprint of the DC due to the economizer implementation will be carried out in this chapter.

Conclusion and Future Work will be the final chapter containing the overall summary, findings of the research and identification of future improvement areas.

# **Chapter 2: Literature Review**

The literature review is divided into several themes that helps at providing a holistic view on the green cloud computing initiatives, increasing the DC energy efficiency, cooling techniques used in the DCs, the effect of increasing the DC operating temperatures, air side economizer mode implementation practices in the DC, the different machine learning processes used for implementing the energy efficient practices in DC and the machine learning models used in the dissertation for predicting the DC temperature and RH inlet levels.

## **2.1 Green cloud computing in DC**

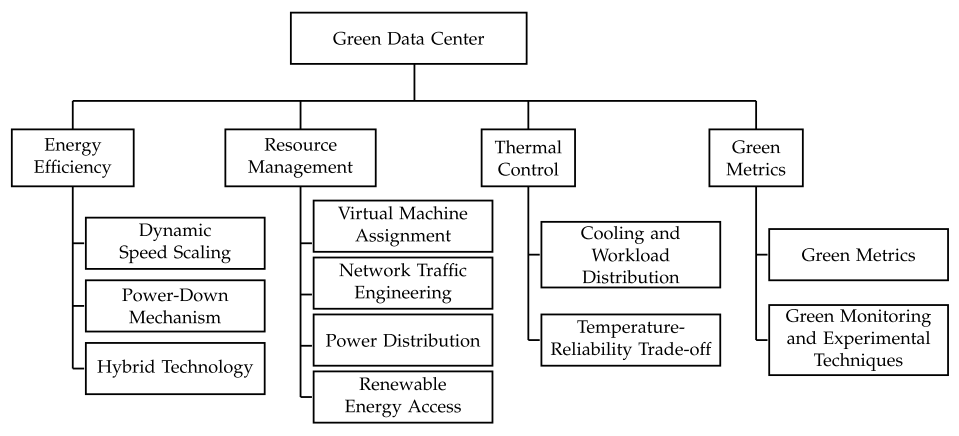
Several research areas are being identified and implemented in order to promote the green cloud computing practices in the DC establishment and operations for sustainable future growth. This section of literature review is targeted on related work.

It is said that the primary responsibility of safeguarding once environment falls on the immediate social animal or human being residing in the surroundings (Nixon, Francis and Devaraj, 2015). Among several initiatives, Green Cloud computing initiatives are being considered for controlling the carbon footprint and the associated effects on the environment caused by the DC operations and expansions. Nixon et al., discusses about the problems caused by the increasing electronic and IT industry on eco-systems followed by a review of the current generation green IT practices. This study also identified the potential research areas for reducing the DC associated impact on the environment. The recommendations cover a wider range of development areas which primarily included hardware, software, and process optimization in addition to the environmentally friendly manufacturing, distribution, and disposal practices. A supporting argument on the need for establishing green practices in the DC industry from both IT and business viewpoints is presented in (Agarwal and Nath, 2011).

|  |  |
| --- | --- |
| **Local Effects** | **Global Effects** |
| E-Waste | Global Warming |
| Risk of Health | Climate Change |
| Air Pollution | Raising of ocean levels |
| Water Pollution | Temperature Increase |
| Land Pollution | Melting of Ice Caps |

**Table 1. Local and Global Impacts of DCs on Environment (Agarwal and Nath, 2011)**

A Carbon Efficient Green Policy(CEGP) is proposed by (S, Niranjan and N, 2014) which talks about an effective system for user specific cloud frameworks. This addresses the carbon footprint reduction techniques that are typically not considered by the cloud providers who usually concentrate on improving energy efficiency based on costs rather than the impact on the environment. Another study (Anwar, 2013) proposes different techniques to reduce the carbon footprint individually at physical server hardware and software levels. A similar argument is presented in (Mohiddin, 2016)(Vikram, 2016). Jin et al., discussed about the different techniques that can be used for greening a DC as shown in fig. 5. This identifies improvement areas and implementation of new techniques in almost all of the essential working components of the DC including the IT equipment, Cooling, and any associated auxiliary equipment in support infrastructure.



**Figure 5. Techniques for establishing a Green DC** (Jin et al., 2016)

The green metrics discussed in the study will be used in this dissertation module for estimating the effectiveness of the novel control scheme based DASFC system implementation. The first and the most commonly used metric among them is PUE. It gives a comprehensive outlook covering IT and auxiliary consumption including the cooling systems involved. Datacenter infrastructure Efficiency (DCiE) is the inverse of the PUE and has been proposed for estimating the energy efficiency of the DC. The average PUE and DCiE values are 1.83 and 0.54, respectively. Data Center Energy Productivity (DCeP) is a metric used to estimate the efficiency of computing in the DC.

From the various green cloud computing propositions in the above review, it is observed that most of the studies concentrated on reducing the carbon footprint of the DCs by concentrating on the conventional energy efficient practices which primarily involved scheduling, resource management and optimization of DC internal as well as the auxiliary components including software. This also included temperature and power aware scheduling, VM migration, consolidation, and load balancing techniques. The literature corresponding to these conventional techniques will be discussed in the next section 2.2 and the following sections will deal with the literature on the cooling in the DCs and the implementation of air side economizers to reduce the carbon footprint.

## **2.2 Conventional energy efficient practices in DCs**

Many approaches considered in section 2.1 points towards the improvement of the energy efficiency in DC industry as a means to reduce the carbon footprint. This section of the literature review will concentrate on such conventional works in the academia.

The research in (Buyya, Beloglazov and Abawajy, 2010) presents the concepts, associated issues and architectural concerns for reducing the power consumption levels in DCs. This study proposes a requirement for a holistic control of all the elements of the DC including the IT equipment, power supply and cooling infrastructures for improving the DCs performance levels and overall energy efficiency. This is achieved by concentrating on the energy efficient allocation of the DC resources using policies and dynamic provisioning algorithms. This also proposes an outline of a possible software solution that can be used for power aware management of cloud infrastructure taking the SLAs and quality of service delivery into consideration. In another study (Berral *et al.*, 2010), intelligent consolidation of the DC infrastructure to reduce the power consumption is proposed. This uses machine learning practices to optimize the DC resource allocation by predicting the power consumption levels and server loads thereby adjusting the scheduling according to the SLAs of the heterogenous workloads.

Similar views are presented in (Rezai and Speily, 2017) and (Elgelany and Nada, 2013) where different VM migration and hybrid power conserving scheduling models are discussed. These are further refined in the following years of research (Suresh, 2015) where the virtualization techniques are considered for the consolidation of the physical nodes into virtual machines to reduce the overall active physical nodes operating at any given time leveraging the fine grained fluctuations in the DC workload. This supported the efficient use of hardware as well as software components to achieve the desired power saving through virtualization.

Adaptive migration of VMs using dynamic allocation of the resources and load balancing was discussed in (Maurya and Sinha, 2013). This research considered the degree of load balancing. SLA violations levels and power consumption as the primary parameters of thresholds for a VM migration. Similar research was done in (Thakur, 2016) for the energy conscious task allocation at each VM or group of VM levels for the tasks in queue in a DC. This introduced the concepts of space and time sharing while minimizing the SLA violations involved. Comparative analysis of different power consumption algorithms in DCs was performed in (Nagpal, 2018) that gave an in depth view of the various factors that causes excessive power draw in the DC infrastructure along with associated control or preventive measures.

Goyal, Arya and Nagpal (Goyal, Arya and Nagpal, 2016) explained a hybrid technique that meets both the requirement of energy conservation as well as SLA maintenance. Another notion of increasing energy efficiency through the optimization of software and hardware techniques was reviewed in (Appasami, College and K, 2011). Additional research on reducing the DC power demand using VM migration and related concepts was found in (Glasser *et al.*, 2016) and (Tafani *et al.*, 2013).

From the above review, the conventional energy conservation practices in DCs can be broadly classified into

* Operational parameter control.
* Load aware, Power aware and temperature aware adaptive scheduling of the IT resources.
* Optimizing the existing mechanical cooling control systems.
* VM consolidation and migration techniques.

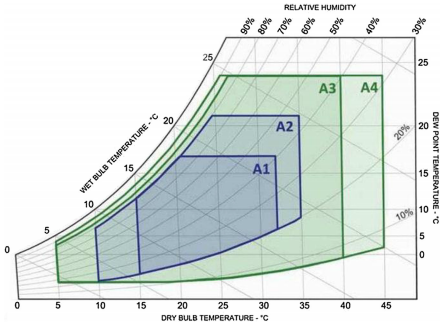
If the natural free cooling economizer is implemented in addition to these conventional practices, it would result in much better power savings and associated reduction in the greenhouse effect of the DCs. The various cooling techniques including the free cooling practices are reviewed in the next section.

## **2.3 Existing techniques and emerging technologies for cooling DCs**

The conventional cooling systems involves the use of mechanical systems in majority of cases to provision the cooling needs of the DC. This contains a primary mechanical system (Compression systems, Chiller towers), heat rejection infrastructure like a cooling tower and the terminal cooling infrastructure that carries the coolant medium to the required locations of the DC (Capozzoli and Primiceri, 2015). Capozzoli and Primiceri also presented different cooling configurations involving liquid cooling and enhanced air side economizers as potential methods to reduce the carbon footprint in a DC. They also presented Facebook’s Open Compute Project (OCP) with an achieved PUE levels of 1.08 that has an effective DASFC economizer based cooling implementation in DC. The server racks of the OCP were developed in line with the higher allowable ranges for temperature and RH as prescribed by ASHRAE. A compelling case of incorporating the renewable resources and TES systems into the overall cooling schemes of the DC was also supported in this discussion.

A similar view is proposed in (Woods, 2010) which provides a detailed overview of the different airflow configuration in the server rooms of the DC which uses an air to air refrigerant scheme. This also presents a critical discussion about the cooling efficiencies and the associating factors influencing it in each of the configurations. To add to this, a leakage aware system is proposed in in (Zapater *et al.*, 2015) which optimizes the coolant circulation based on the analysis of hot-cold mixing or leakage with emphasis on controlling and predicting the leakage factors.

The changes to the ASHRAE environmental classes for DCs in 2011 increased the thresholds for the DC equipment in terms of the operating temperatures, humidity and the ability to withstand particulate matter deposits among other considerations(Zhang *et al.*, 2014)(Van Le *et al.*, 2019). This provision made the previously unsuitable climatic zones acceptable for establishing free cooling and in particular DASFC system implementation in DCs around the tropics.



**Figure 6. Environmental classes from ASHRAE for DC economizers** (Daraghmeh and Wang, 2017)

ASHRAE environmental classes for the DCs is depicted in fig. 6. It contains the recent inclusions of the classes A3 and A4 which supports wider operating ranges for temperature and RH for a DC operation. The inclusion of these new environmental classes resulted in the design of advanced servers and associated IT equipment that can handle large temperature variations. The approximate operating thresholds for the classes A1, A2, A3 and A4 are shown in table 2. The DBT is typically considered to be the temperature of the air where as the WBT corresponds to the saturation of the temperature in the presence of water vapour at constant temperature and pressure around the monitoring equipment. The DPT is the temperature at which complete saturation of the air is observed which causes the condensation of the water vapour in air (*Dry Bulb, Wet Bulb and Dew Point Temperatures*, 2020)

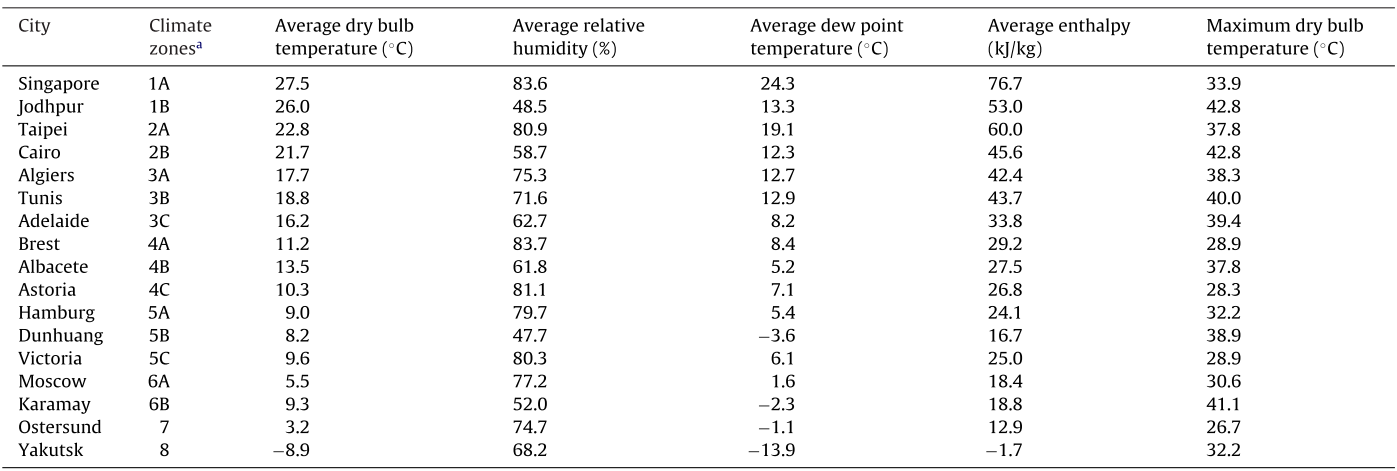
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Element** | **Class A1** | **Class A2** | **Class A3** | **Class A4** |
| Dry Bulb Temperature (in °C) | 15 – 32 | 10 – 35 | 5 – 40 | 5 – 45 |
| Wet Bulb Temperature (in °C) | 13 – 18 | 10 – 22 | 5 – 25 | 5 – 25 |
| Dew Point Temperature (in °C) | 6 – 17 | 8 – 21 | 0 – 24 | 2 – 24 |
| Relative Humidity (in %) | 74 – 80 | 63 – 80 | 48 – 85 | 37 – 90 |

**Table 2. ASHRAE environmental operating classes for DCs**

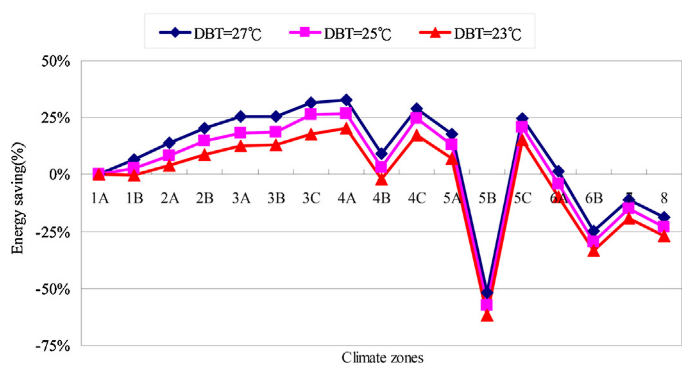
Based on the changes in the ASHRAE’s allowable ranges research was done in (Lee and Chen, 2013) to present the DASFC economizer based system to save energy in DCs established in different climatic zones in the world. The worldwide climatic zone classification is shown in fig. 7. The average temperature and enthalpy measurements for these zones is given in table3 and the corresponding savings using DASFC systems in these zones is depicted in the fig. 8.



**Figure 7. Worldwide Climatic Zone Classification** (Lee and Chen, 2013)



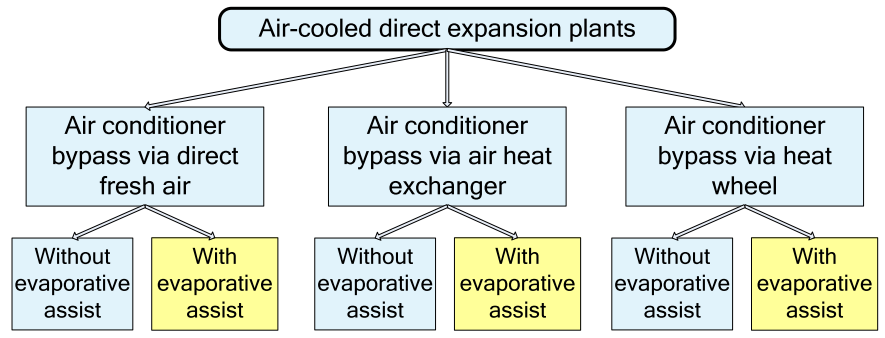
**Table 3. External air temperature and enthalpy measurements in worldwide climate zones (Lee and Chen, 2013)**



**Figure 8. Percentage Energy Savings with DSAFC systems in worldwide climate zones**(Lee and Chen, 2013)

An Additional research on the impact of updated ASHRAE’s allowable threshold for air free cooled DCs (The Green Grid, 2012) includes virtually most of North America and Europe in addition to many tropical zones as viable climatic zones for establishing DASFC. Zhang et al., provided a comprehensive review of different kinds of free cooling technologies used in the DCs which included different configurations of airside, waterside and heat pipe systems working in coordination with the existing mechanical or DX systems. In this review, the potential of the DASFC systems in reducing the power consumption and associated carbon footprint is evaluated.

The various factors that affect the decision making process of opting for a DASFC or IDASFC economizers modes in the data center are critically examined in (Lin, Niemann and Long, 2015). This also included the different configurations of these economizers, the supporting systems needed for each configuration and the effects of the particulate matter in DASFC system. As part of this, the standards for particulate matter allowable ranges, filtering and other auxiliary systems were also discussed. In a related study(Niemann, Bean and Avelar, 2013) the different economizer modes for air and water free cooled systems are illustrated. Although using evaporative assist results in a decrease of the temperature of air by almost 19° Celsius, it results in high humidity because the air will be passed through a wet mesh like material in this method.



**Figure 9. Airside Free Cooling Implementation Schemes**

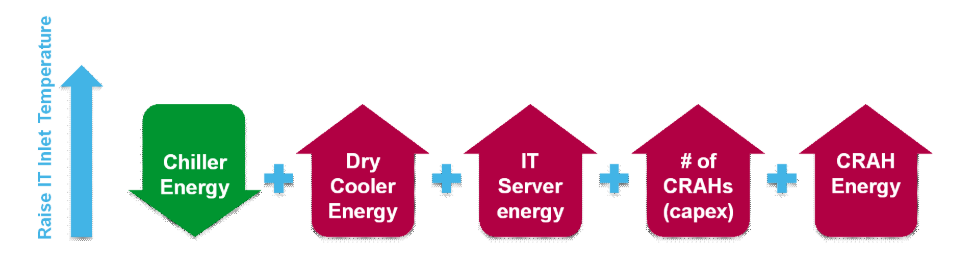
The effectiveness of recirculating the exhaust air in the DASFC system for further reducing the cooling demand and thus the cooling related power consumption related to the draft fans, filters and vapour assists chambers is illustrated in (Park and Seo, 2018). Adding to this, the importance of isolating the hot and cold air streams in air free cooled DCs (Niemann, Brown and Avelar, 2020) is said to be a serious concern especially if the location has higher average inlet temperatures. The leakage or mixing of these streams will negate the cooling effect provided by the DASFC economizer and will cause a threat to the reliability of the DC when running closer to the upper temperature thresholds.

Zhou and Turner (Zhou and Turner, 2008) worked on different activation strategies for enabling or disabling the functioning modes of the DASFC economizers. Their work reviewed the temperature and enthalpy based activation strategies with associated advantages and shortcoming in the respective categories. To overcome any disadvantages in former activation strategies, an optimized activation strategy was proposed that has a multi layer decision making capability.

So, for the purpose of this dissertation, DASFC with evaporative assist will be used. The operation of the evaporative assistance system will be limited to the situations only when the outside temperatures are far above the DC setpoints. This is to reduce the impact of the humidity on the DC equipment as the DC location (Orlando, Florida) selected has higher humidity levels in the air which adds to the humidity accumulated during evaporative assist process. The primary assumption of the server room infrastructure is that the hot and cold aisles are completely isolated. A temperature based activation strategy will be used in the hypothetic model being used in the dissertation.

### **2.3.1 Running the DCs hot**

The primary assumption in the case of implementing the free cooling solution in the tropics is that the DCs are capable of running at higher temperatures (Manousakis *et al.*, 2015)(Song, Zhang and Eriksson, 2015). But there are certain issues and concerns when running the DCs at increased threshold levels. There is always the concern of thermal attack on the IT and supporting equipment which reduces the life span of the devices(Chung, Kalbarczyk and Iyer, 2018)(Gao *et al.*, 2017). The need to evaluate the energy consumption of the DC as a whole instead of just the IT equipment energy consumption is very important. This is because raising the operating temperature of the IT equipment will increase the economizer operating hours but at the same time it will increase the power consumption of the auxiliary components in the DC. Examples include increase in speed of the server fans, increase in power consumption of the servers due to decreased performance levels at higher temperatures, increase in the capital and operating expenditures of the air handling units. If the power consumption of these components crosses that of the savings from the economizer mode, the implementation of the system becomes impractical. A cost based analysis of this view for a test DC was proposed in the work (Torell, Brown and Avelar, 2016)



**Figure 10. Impact of Raising Inlet Temperatures in DC** (Torell, Brown and Avelar, 2016)

## **2.4 Machine Learning for Green Computing in DCs**

Several ML techniques are being applied for implementing green cloud computing practices in DC industry(Karthik *et al.*, 2016). Some of these techniques involved the optimization of the DC resources to reduce the energy consumption. This included smart migration of VMs, predictive maintenance, temperature aware scheduling, power aware scheduling (Berral *et al.*, 2010), load balancing(Ari *et al.*, 2017), proactive control of operational parameters (Hossain *et al.*, 2017), control, and monitoring of cooling systems as wells as auxiliary systems.

A deep learning based control strategy for controlling the free cooling in tropical areas was proposed in (Van Le *et al.*, 2019) which leverages the mixing of hot exhaust air into the incoming stream of fresh air to maintain constant DC setpoints of temperature and RH. The proposal was primarily based on the assumption of the servers being able to withstand higher temperatures as per the modified ASHRAE thresholds. The increase in the temperature of the server racks is used to address the increased RH levels faces in some scenarios. Another DRL based system was designed in (Li *et al.*, 2020) to achieve end to end control of the conventional cooling system components in the DC. Recurrent Neural Networks were used in (Zou, Yu and Ergan, 2020) for optimizing the control scheme of the DX expansion systems and AHUs. Scheduling and consolidation of the VMs for conserving energy was also performed using DRL in (Farahnakian, Liljeberg and Plosila, 2014). A multi agent study using DRL was considered in (Chi *et al.*, 2020) for improving the energy efficiency of the IT and cooling systems together in the DC.

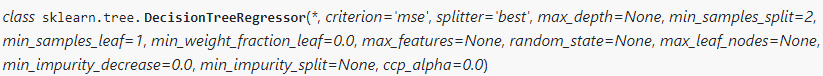
A decision support system was proposed in (Pawlish and Varde, 2010) which is used to arrive at the best sequence of decisions to operate the DC as green as possible considering all the capital costs, operational cost, feasibility of implementation and the practicality of the solution. Apart from the carbon footprint, the study also concentrated on VMs and thermal profiles of the DC. Random Forest Model was used for the statistical analysis and temperature prediction in (Pang *et al.*, 2017).

## **2.5 Machine Learning Algorithms used in the Study**

A suite of machine learning algorithms from various Python libraries will be used in the analysis of the meteorological data collected for the DC location to predict the temperature and RH at the inlet. Since temperature and RH are highly correlated factors (*Humidity | North Carolina Climate Office*, 2020), a multi target prediction approach (*Decision Trees — scikit-learn 0.24.0 documentation*, 2020) has been considered to enable better insights into the prediction behaviour and results. The ML algorithms used for multi target prediction are illustrated in this segment of the literature review.

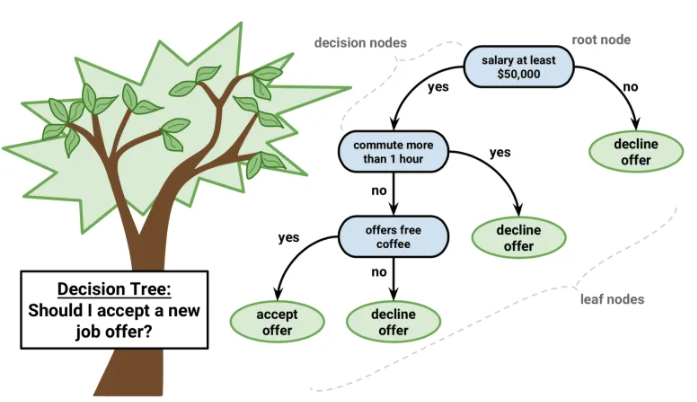
### **2.5.1 Decision Tree Algorithm**

Decision Tree algorithm is a non-parametric supervised learning algorithm which can be used for addressing both classification and regression related problems. This uses a tree like structure in order to resolve the training problem. It contains a series of nodes and leaves that propagate in a tree like structure using probabilistic ‘Yes’ or ‘No’ responses at each node. The learning experience is based on these decision rules inferred from the features. Information Gain and Gini Index are the most commonly used properties for selecting the feature split at each node. The Class specification in the scikit learn library for decision tree regressor (Pedregosa *et al.*, 2011) is shown in fig. 11 which also contains the list of the hyperparameters. This requires very less data preparation and the trees are easy to visualize and understand.



**Figure 11. Decision Tree Hyperparameters** (Pedregosa et al., 2011)

Decision trees are prone to overfitting and need techniques such as sampling, pruning, and setting the maximum depth are needed to avoid this. Also, the outliers have considerable impact on the tree generation and subsequent path followed which can be controlled by using ensemble algorithms like Random Forests and Gradient Boosted Trees. The models trained using these ensemble algorithms generalize well and are reasonably good at extrapolation. An initial model will be created with the help of a Decision Tree algorithm followed by Random Forest and Gradient boosted trees for comparison of the best performing model.



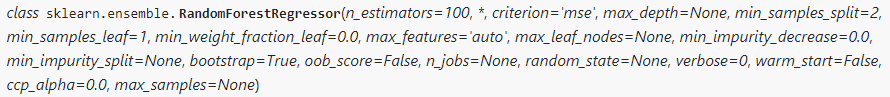
**Figure 12. Decision Tree Algorithm Example** (Saxena, 2017)

Decision Tree Regressor Pseudo Code: (Hu *et al.*, 2012) (Hambali *et al.*, 2019)

|  |
| --- |
| GenerateDecisionTree (Sample D, Attributes A)  Steps:  1. If stopping condition for (D, A) is TRUE then  a. Create LeafNode = createNode ()  c. Return LeafNode  2. root = RootNode = createNode ()  3. test condition for RootNode = BestSpiltFunction (D, A)  4. X= {x | x is a possible outcome for the test\_condition, root.test\_condition}  5. For each value x Є X:  a. Dx= {d | root.test\_conditions = x and d Є D};  b. ChildTree = TreeGrowth ();  c. Create the child tree as the descent of the root.  d. Label the edge {root → child} as x  6. Return root |

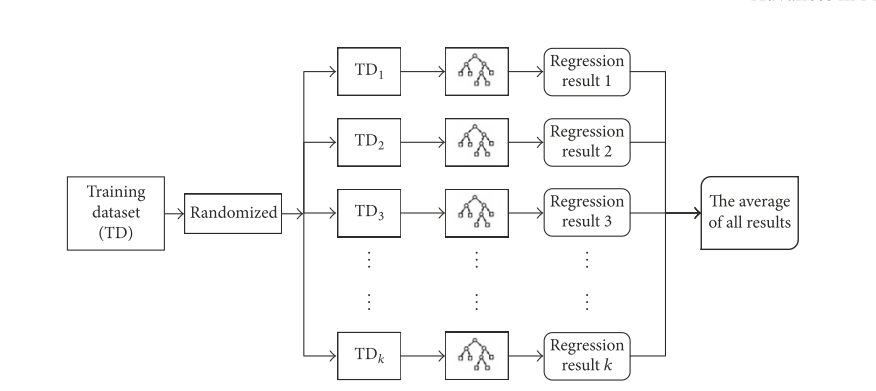
### **2.5.2 Random Forest Algorithm**

The Random Forest Algorithm from Scikit Learn library (Pedregosa *et al.*, 2011) will be used for this purpose. A sample of the Random Forest Regressor library structure and hyperparameter list is shown in fig. 13.



**Figure 13. Random Forest Hyperparameters**

RF algorithm uses an ensemble of DTs for improved accuracy. In this method, unpruned DTs are organized into a forest like structure as shown in the figure-3. This configuration offers improved CART like structure that can be used for both classification and regression problems (Pang *et al.*, 2017). Each of the CART represents a separate regression or classification path which uses a DT for the estimation of the corresponding result using a randomized data subset from the overall dataset as an input. The CART has root nodes branching out into internal nodes and leaves. RF makes use of random feature selection for splitting the nodes in each CART. Also, squared residual minimisation algorithm is employed in the splitting of each node which requires the sum of variance in the resulting nodes to be as minimal as possible thereby reducing the variance related over fitting errors to some extent. The features for splitting each node are also randomized like in the case of the input. Also, RF will be able to handle missing values. The prediction accuracy will be improved by using averaging.



**Figure 14. Random Forest Algorithm Structure** (Pang et al., 2017)

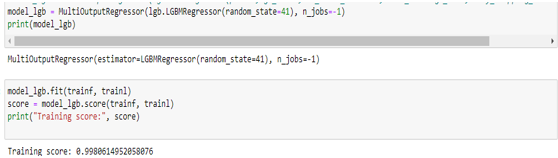
RF algorithm supports both multiclass classification and multi target regression problems inherently. Also, multi target regression models are naturally more stable and provide better overall performance compared to that of a combinational model obtained from two or more single target regressions. Additionally, it has inherent methods for determining the importance of the features and has an OOB (out of box scoring) method for scoring the individual CART performance. It also has native support for multi target regression. It is not sensitive to outliers and handles them using the binning process. The nonlinearity of the feature set does not have much impact of the tree classification or prediction inferences. If bootstrap feature is enabled, the sample size will be equal to that of the max\_samples criteria. In the absence of max samples criteria, entire dataset will be used for building each tree. In this study, the features in the meteorological data will be used to model the behaviour of multiple targets involving temperature and RH levels. So, RF will be a better fitting model for this multi-target regression analysis.

Random Forest Regressor Pseudo Code:

1. Select ‘xi’ Random features from the feature set ‘X’ where xi < X where ‘i’ is the number of features in the selected random subset.
2. Initiate the CART for the selected feature set xi.
3. Calculate the best split feature as the node of the CART
4. Create daughter nodes from the node using best split condition at each node.
5. Steps ‘a’ and ‘b’ will be repeated till ‘i’ nodes are reached
6. Predict outcome of the CART.
7. CART outcome prediction.
8. Build multiple CARTs using step i to iii.
9. Select the targets with highest prediction rating in each CART.
10. Averaging highest prediction from all CARTs to obtain the final prediction.

### **2.5.3 Gradient Boosted Trees**

Gradient boosting is a ML technique which typically uses decision trees in an ensemble where generalization is achieved through the boosting of the differentiable arbitrary loss function(*Ensemble methods — scikit-learn 0.24.0 documentation*, 2020). LightGBM library from Microsoft uses histogram binning and has considerable performance improvement over Scikit Learns Version of GDBT. Also, in LightGBM a leaf wise growth of tree is observed unlike other DT algorithms where level or depth wise tree growth is more prevalent as shown in fig.15 and fig. 16. Additionally, LightGBM supports parallel data processing, GPU enabled training and a wide variety of metrics. But gradient boosting algorithms are not optimized for multi target regression and lacks a native support for it. These needs to be wrapped in multiouputregressor model to enable the multi target support.



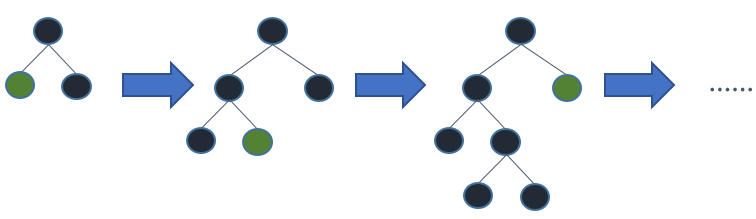
**Figure 15. Extending Multi Target Support to Gradient Boosted Trees (LightGBM Model)**



**Figure 16. Extending Multi Target Support to Gradient Boosted Trees (Scikit Learn Model)**



**Figure 17. Level (Depth) wise growth of tree**



**Figure 18. Leaf wise growth of tree**

### **2.5.4 Other Models**

As the primary requirement is to create a DSS to aid the control of the cooling scheme, decision tree and its ensemble models (Random Forest and Gradient Boosting Trees) are considered for the training of multi target predictive models as explained above. In addition to this several other algorithms have been considered which includes Linear Regression, KNN, Linear SVR and Chained Multi Output Regression (Brownlee, 2020). All the models are discussed in detail in the Methodology section.

# **Chapter 3 – Methodology and Artefact Design**

## **3.1 Introduction**

This chapter is primarily divided into two themes. The first theme contains the explanation regarding the selection of data, data exploration techniques and a comparative analysis of different ML algorithms for multi target prediction. The second theme will consist of the proposal for the novel DSS based multi-mode cooling control system.

## **3.2 ML model for Multi Target Prediction**

### **3.2.1 Data Collection**

As part of this research, the meteorological data has been collected for the location with the coordinates, (lat: 28.66654, lon: -81.5625, asl: 25.234) which are close to the city of Orlando in Florida, USA. This is a mild tropical zone with considerable variations in the temperature, RH, and other meteorological factors throughout the year. The particulate matter (average particulate matter) and an average air quality index of 40 (Wilcox, 2020) in the region is generally favourable for the usage of DASFC system. The meteorological data has been obtained from MeteoBlue’s climate modelling system (MeteoBlue-HistoryPlus, 2020) that has historical meteorological data recorded from 1985. In this dissertation, the past 20 years of data starting from January 01st of the year 2000 is considered.

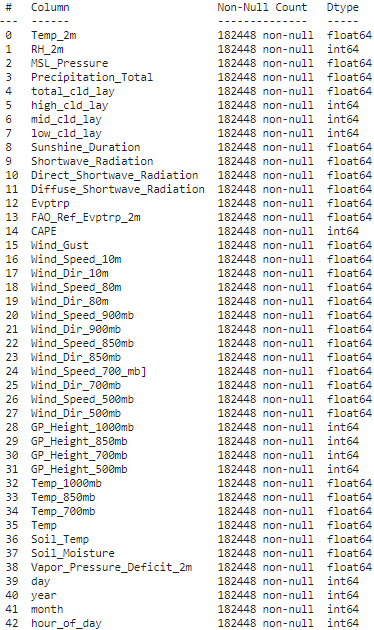
### **3.2.2 Technology Stack used for the creation of the Model**

Google Colab (Google Colab, 2020) and Jupyter notebook (version: 6.1.4) from Anaconda3 (*Anaconda | Individual Edition*, 2020) of application with Python (Version 3.8.5) will be used for creating the machine learning models using multiple ML centered libraries. Tableau Desktop (Version: 2020.2) (*Tableau Desktop*, 2020) is also used to create visualizations from the data. Rapid Miner Studio (*Predictive Analytics Software | RapidMiner Studio*, 2020) (Version 9.7) is used for the initial exploration of the machine learning algorithms and the feasibility of their application.

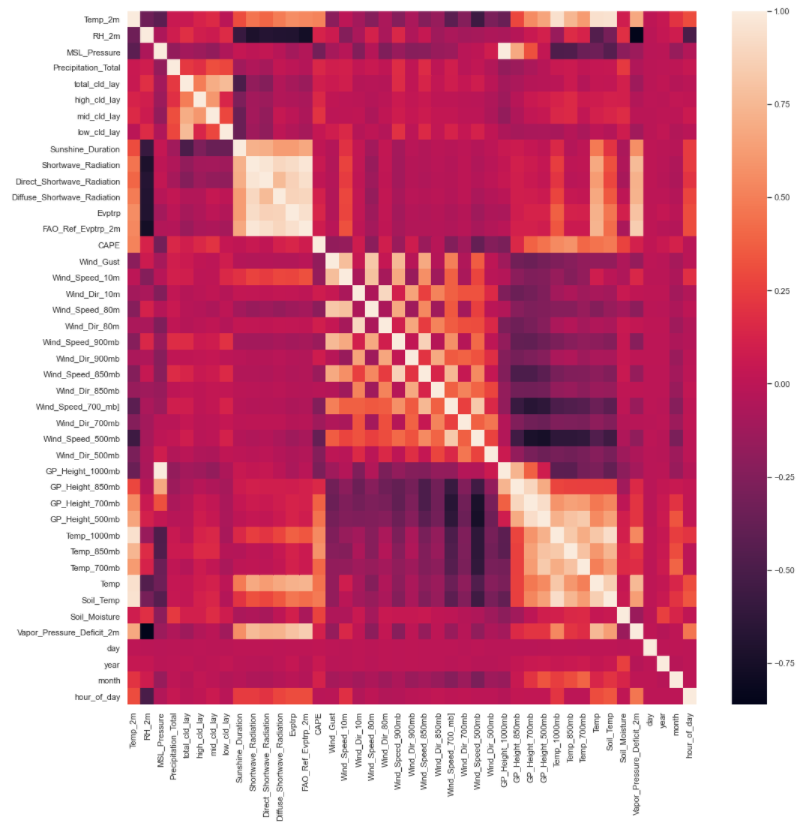
### **3.2.3 Data Exploration and Pre-processing**

The dataset has 182448 rows and 43 columns out of which Temperature and RH at two meters level are considered as the target variables since these are the expected inlet parameters. The remaining columns are considered to be features. The ‘snowfall’ feature has been removed from the feature list as there is no associated data since snowfall phenomenon does not occur in the selected sub-tropical region. Also, none of the rows are duplicated and there are no missing values in any of the remaining features. The timestamp feature has been resolved into multiple fields containing monthly, daily, and hourly data.

There is no categorical data to be resolved further into corresponding numerical data in any of the features or targets. The initial column list is shown in fig. 19. A preliminary correlational analysis is conducted as shown in fig. 20 to identify the correlation between various parameters in the dataset. Although most of the independent features are closely correlated to the targets, it has been observed that the elimination of these features resulted in large deviations in the resultant models. So, based on correlation and causality, all the features have been considered for ascertaining best fit model. A decision tree regressor based feature importance estimation is done to find the relative importance of the features with respect to the targets as shown in fig. 21.



**Figure 19. Dataset column list**



**Figure 20. Correlation Heat Map**



**Figure 21. Feature Importance Estimation**

Splitting the features and targets is done using the test\_train\_split function of the Scikit Learn library assigning 15% of the dataset for validation of the trained model. The resultant training and testing sets are further resolved into numpy arrays for the ease of operation. Both the training and testing feature sets are scaled using Scikit Learn’s standardscaler feature since some of the selected algorithms need all the features on the same scale for optimum performance and results.

### **3.2.4 Selection of the ML model**

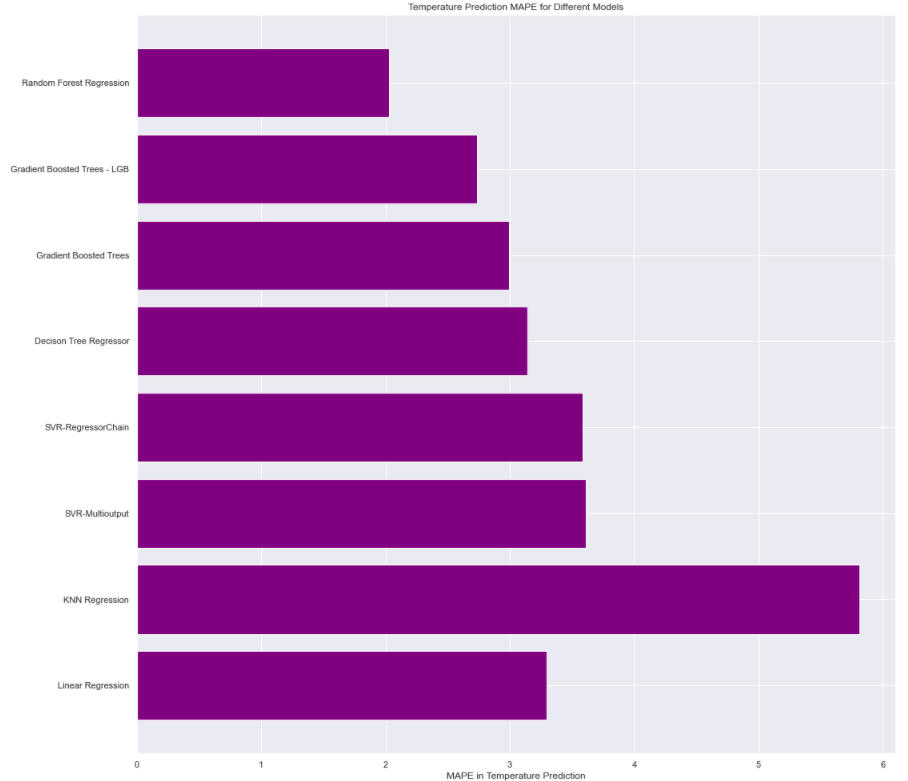
An initial linear regression model is run and the observed R2 score is 0.9575 which indicates that the targets are linearly dependent on the features. As the data is mostly linear, deep learning algorithms are not considered in this analysis. Several ML algorithms have been considered for selecting the most suitable model for multi target prediction. These algorithms are used to train the models that can predict the temperature and RH simultaneously from the meteorological features.

Among these, certain models like Linear Regression, KNN Regression, DTR and Random Forest have native multi target regression support. SVR and GDBT algorithms that does not have this native support achieve the same by using Multioutput Regressor and Chained Multi-Output Regressor functions. Table 4 shows the performance metrics for different models in terms of MSE, RMSE, R2 Score, MAE and MAPE.

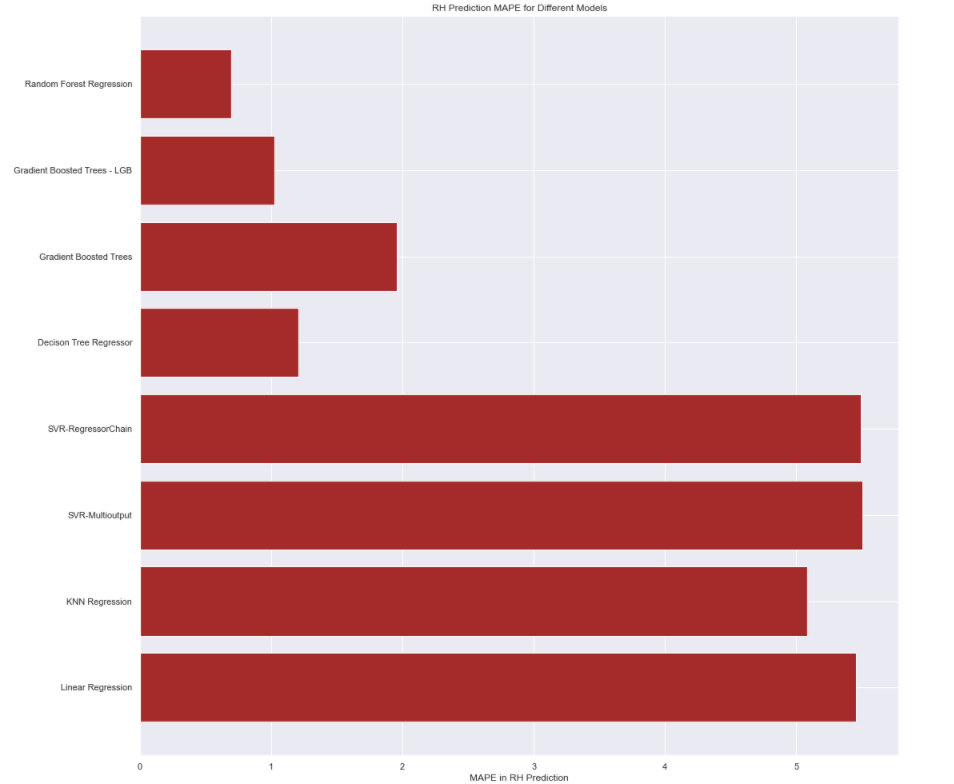
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model Name** | **MSE Temp** | **MSE RH** | **RMSE Temp** | **RMSE RH** | **R2**  **Temp** | **R2 RH** | **MAE Temp** | **MAE RH** | **MAPE  Temp** | **MAPE RH** |
| Linear Regression | 0.21 | 23.61 | 0.45 | 4.86 | 0.99 | 0.93 | 0.33 | 3.35 | 3.30 | 5.46 |
| KNN Regression | 0.88 | 21.92 | 0.94 | 4.86 | 0.98 | 0.93 | 0.69 | 3.33 | 5.81 | 5.08 |
| SVR-Multioutput | 0.22 | 24.60 | 0.47 | 4.96 | 0.99 | 0.92 | 0.33 | 3.26 | 3.61 | 5.50 |
| SVR-RegressorChain | 0.22 | 24.64 | 0.47 | 4.96 | 0.99 | 0.92 | 0.33 | 3.27 | 3.58 | 5.49 |
| Decision Tree | 0.33 | 1.59 | 0.57 | 1.26 | 0.99 | 1.00 | 0.34 | 0.66 | 3.14 | 1.21 |
| GDBT | 0.17 | 2.93 | 0.41 | 1.71 | 1.00 | 0.99 | 0.30 | 1.11 | 3.00 | 1.96 |
| GDBT - LGBM | 0.08 | 0.81 | 0.28 | 0.90 | 1.00 | 1.00 | 0.19 | 0.60 | 2.73 | 1.03 |
| RF | 0.11 | 0.48 | 0.33 | 0.69 | 1.00 | 1.00 | 0.20 | 0.38 | 2.03 | 0.69 |

**Table 4. Metrics for Model comparison**

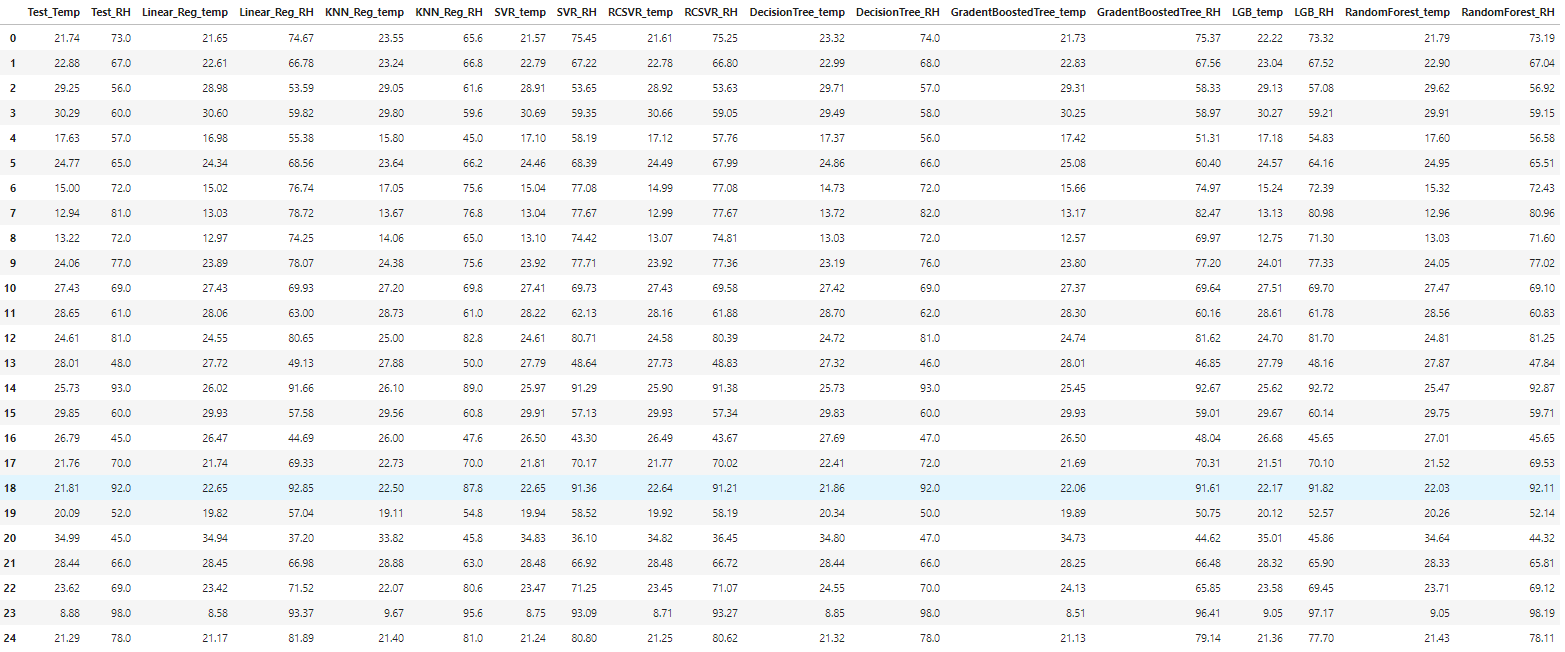
The MAE and MAPE are considered as the most suitable error metrics for the analysis of the best performing model as these metrics directly correspond to the absolute difference and percentage of absolute difference from actual values to the predicted values. Especially MAPE provides a better insight into the predicting accuracy of the models compared to other metrics as illustrated in fig 22 and fig 23. Table 4 shows the sample predictions using different models.



**Figure 22. Temperature Prediction MAPE**

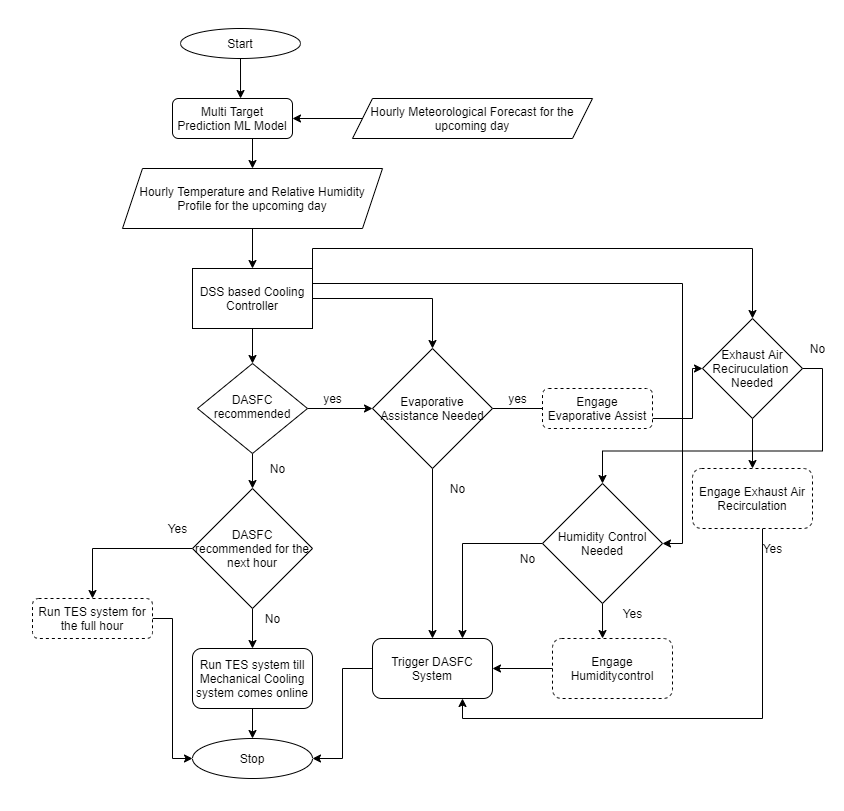


**Figure 23. RH Prediction MAPE**

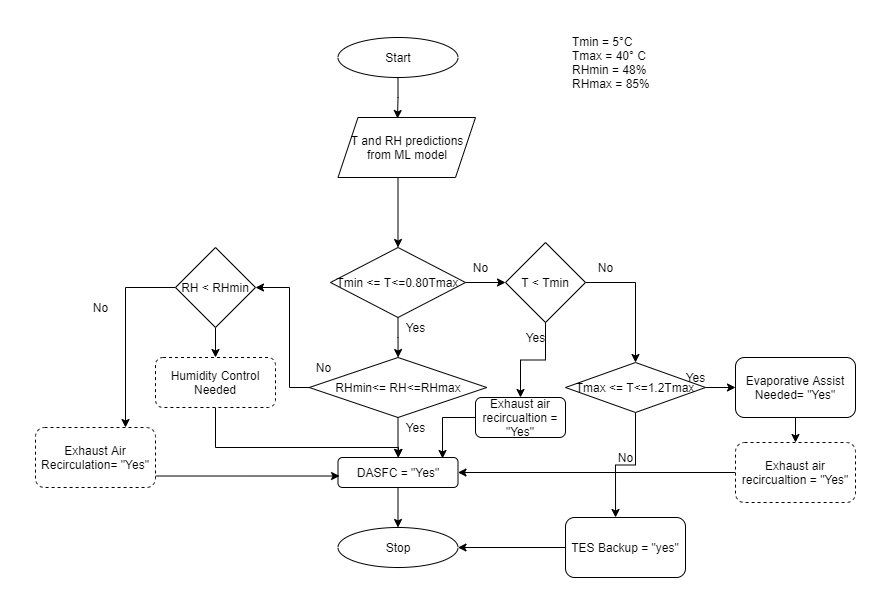
**Table 5. Sample Predictions Using different models**

After careful consideration, it is observed that the Random Forest model performs better when compared to all the other models in terms of different error and performance metrics discussed above. Also, the prediction accuracy of the model for both temperature and RH fares much better than that of other models that have considerable differences in the prediction accuracies for the multiple targets.

## **3.3 DSS system for cooling control:**

****

**Figure 24. Flowchart for Cooling Control Scheme**

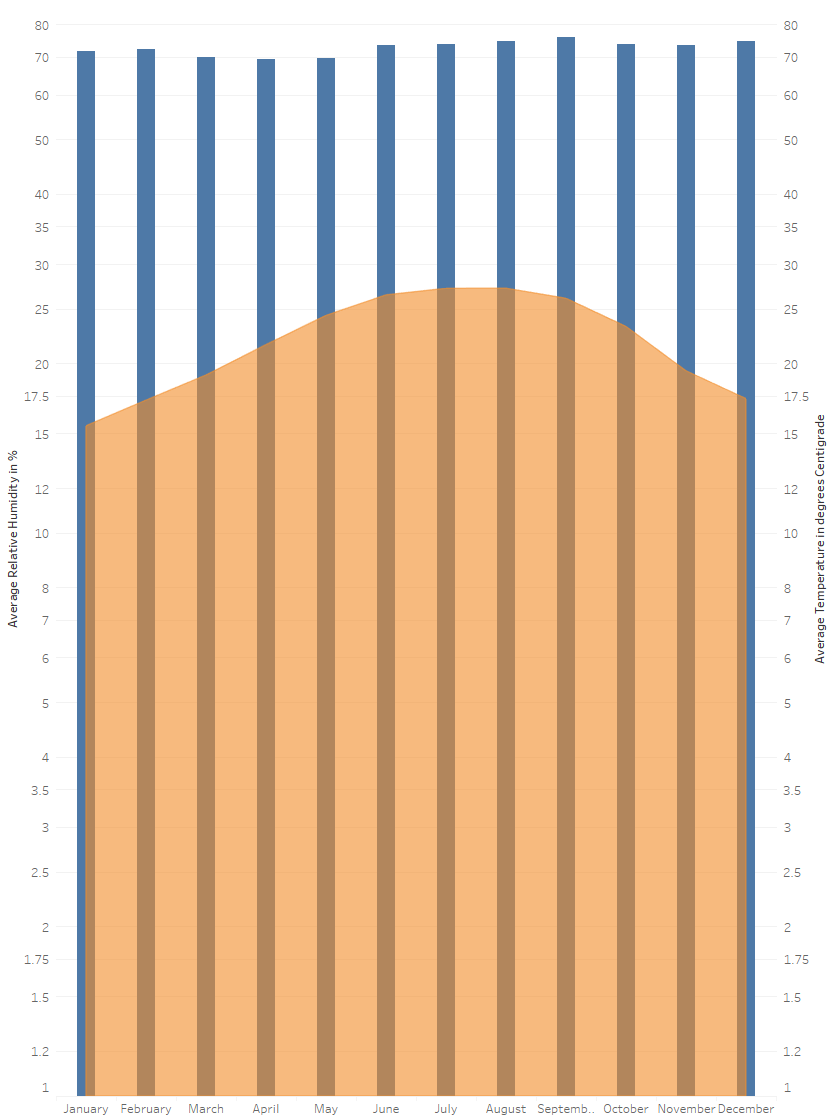
****

**Figure 25. Flowchart for DSS Logic**

The overall cooling control scheme is illustrated in figure 24 and the associated DSS logic involved is depicted in figure 25. The hourly meteorological forecast is provided as input to the trained ML model which generates a temperature and RH profile for the upcoming day. This profile is then passed onto the DSS system that determines the operating parameters of the cooling control for each hour and forms a switching scheme. This switching scheme provides the data needed for operating different components of the mixed mode cooling system. If the operating temperature and RH ranges are within the limits, DASFC system will be triggered directly.

If the temperature ranges are over the specified limits but within 1.2 times of the max temperatures, evaporative assist with exhaust air recirculation(Park and Seo, 2018) or humidity control is employed to bring down the temperature by at least 10°C while maintaining humidity levels in the DASFC system. If the temperature is below the specified lower limit of 5°C or if the Relative Humidity levels are higher than 80% then Exhaust Air Recirculation is engaged as shown in Fig 24. If the predicted temperature is greater than 1.2 times Tmax then TES backup(Capozzoli and Primiceri, 2015) or Mechanical cooling options are engaged in accordance with the succeeding hour cooling recommendations from the switching profile. If the subsequent hours have DASFC recommended, then TES backup will be able to cover for the initial hour and DASFC will be triggered for the next hour. Otherwise, mechanical cooling will be primed for triggering while TES is running.

This proactive determination and priming of the mechanical system in combination with TES backup will provide seamless switching between different cooling modes with maximum priority given to the economizer mode. For the purpose of this analysis, ASHRAE level 3 envelope recommendations are used for the DBT and RH values as shown in Table 2. Also, from the analysis of the meteorological data of the DC location as shown in fig 26, it is observed that the average temperature and RH ranges falls in the operating ranges of ASHRAE level 3. The ML model trained using Random Forest algorithm has been used in the creation of the Cooling Controller Schematic.

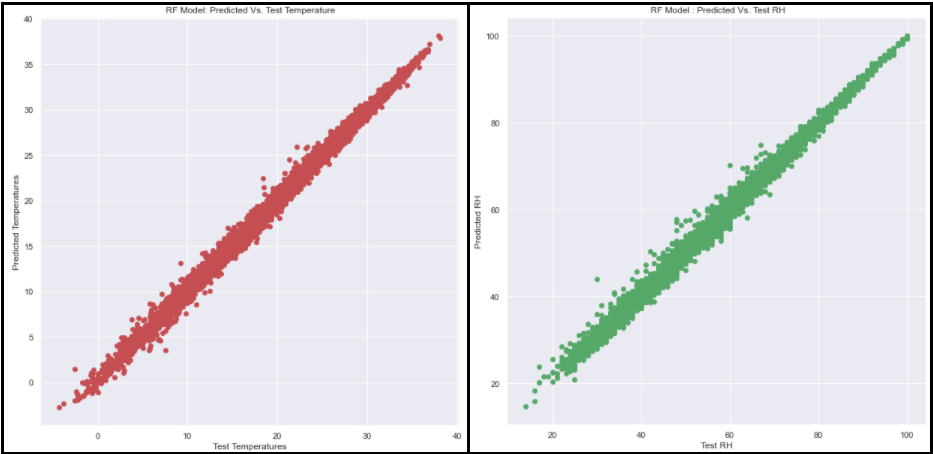


**Figure 26. Monthly Average Temperature and Relative Humidity Levels**

# **Chapter 4 – Results and Analysis**

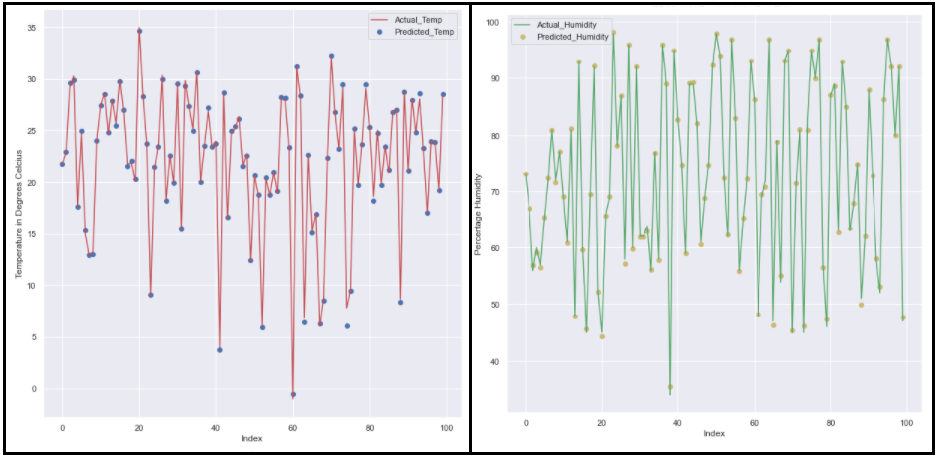
## **4.1 RF Model Performance Evaluation**

The Random Forest algorithm trained model recorded the least possible error in terms of MSE, MAE, RMSE and MAPE. It has also recorded the highest levels of explained variance approximating the value of 1. The prediction accuracies of both temperature and RH are comparatively high with corresponding MAPE values of 2.03% and 0.69%. A sample prediction accuracy of the model for both Temperature and RH is shown in fig 27. It can be seen that the Predicted Vs Test scatter plots for temperature and RH shows a linear trend corresponding to the lower errors in the multi target prediction using this model.



**Figure 27. Predicted Vs Test Scatter Plots for Temperature and RH (Random Forest Model)**

A similar line and dot plot for analysing the accuracy of model prediction is depicted in fig 28 where the random predictions for 100 datapoints is plotted against the associated test values. The plot shows little to no deviations between the predicted and actual values plotted. This is a very important consideration as very high level of reliability is expected in the DC operation and the lowest possible predictive errors helps in achieving the same.



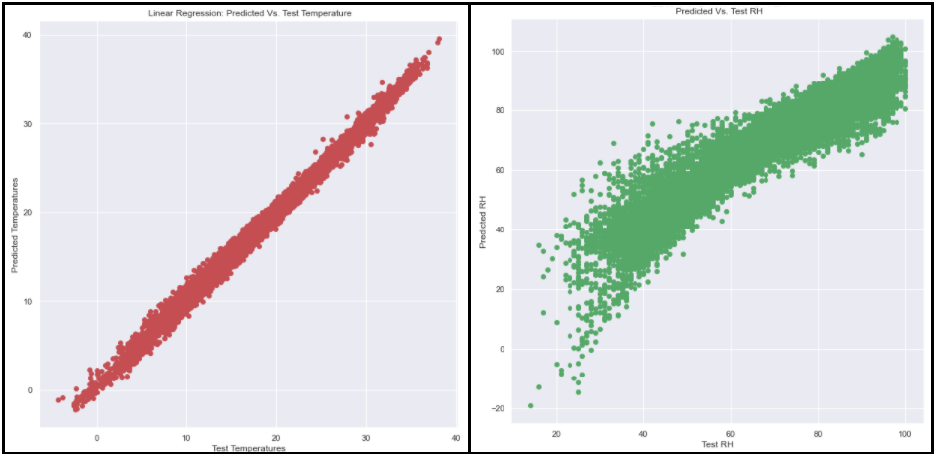
**Figure 28. Prediction Performance of the model for Temperature and RH (RF Model)**

## **4.2 Performance evaluation of other algorithms considered**

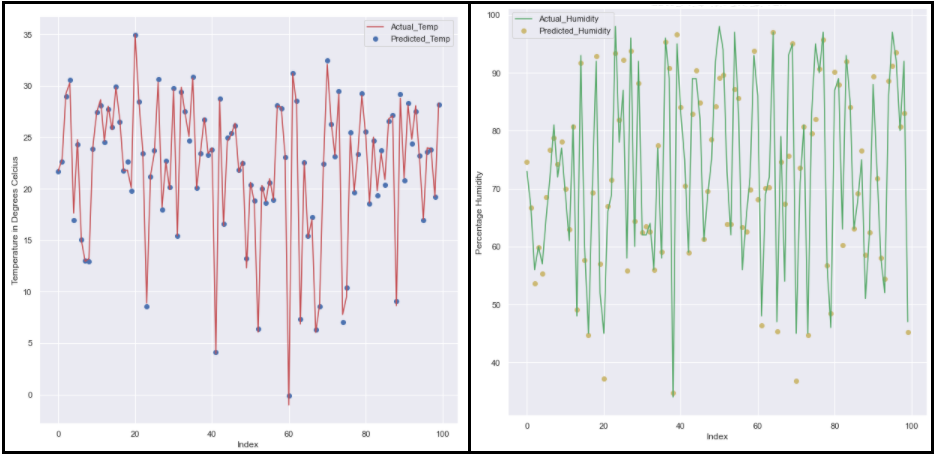
Although other algorithms fared well in terms of Temperature estimation, many of them failed in producing lower error results for RH prediction. A model wise depiction of these results is illustrated in the following sections.

### **4.2.1 Scikit Regression Models – Linear, KNN, SVR, DT and GBT**

The linear regression has a very high deviations in the RH predictions as compared to that of the test values according to fig 29. Also, the test vs predicted plot from fig 30 shows minor deviations in the temperature predictions while RH predictions falls short by a wide range for random data points from the validation set.

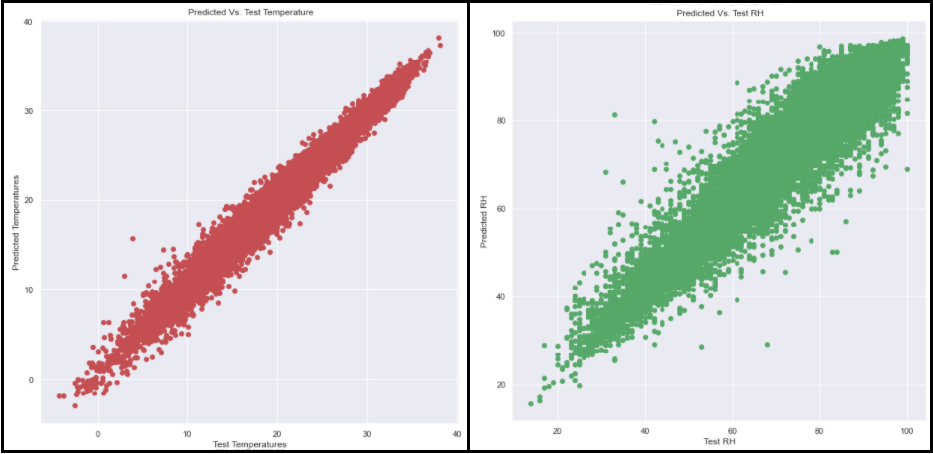


**Figure 29. Predicted Vs Test Scatter Plots for Temperature and RH (Linear Regression Model)**

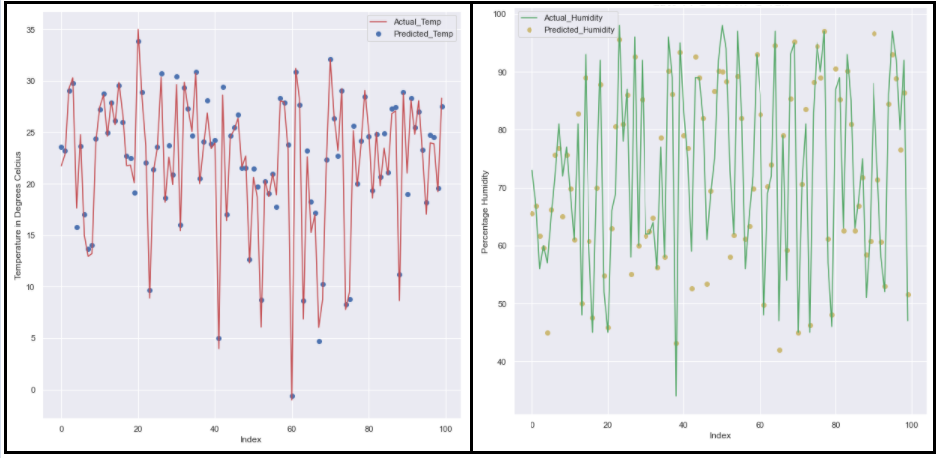


**Figure 30. Prediction Performance of the model for Temperature and RH (Linear Regression Model)**

KNN Regression model performs poorly on both temperature as well as RH fronts as shown in fig 31 and fig 32. The actual vs predicted plot shows considerable deviations away from the linear distribution for both temperature and RH.

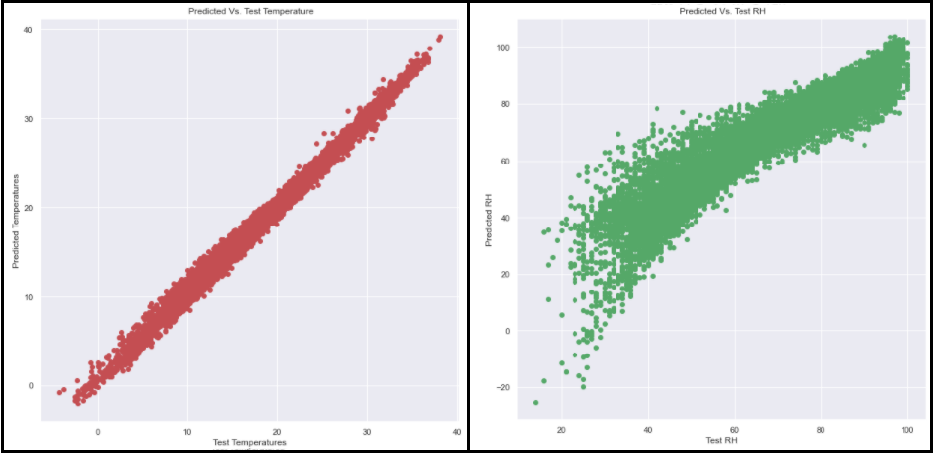


**Figure 31. Predicted Vs Test Scatter Plots for Temperature and RH (KNN Regression Model)**

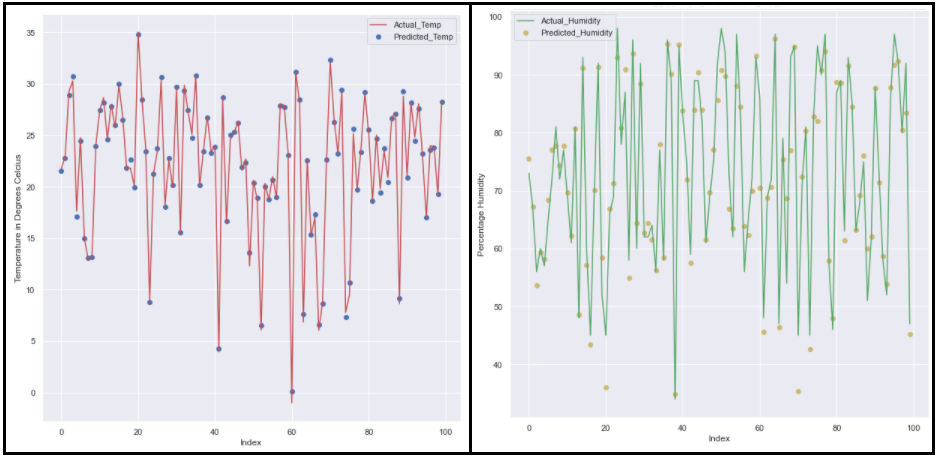


**Figure 32. Prediction Performance of the model for Temperature and RH (KNN Regression Model)**

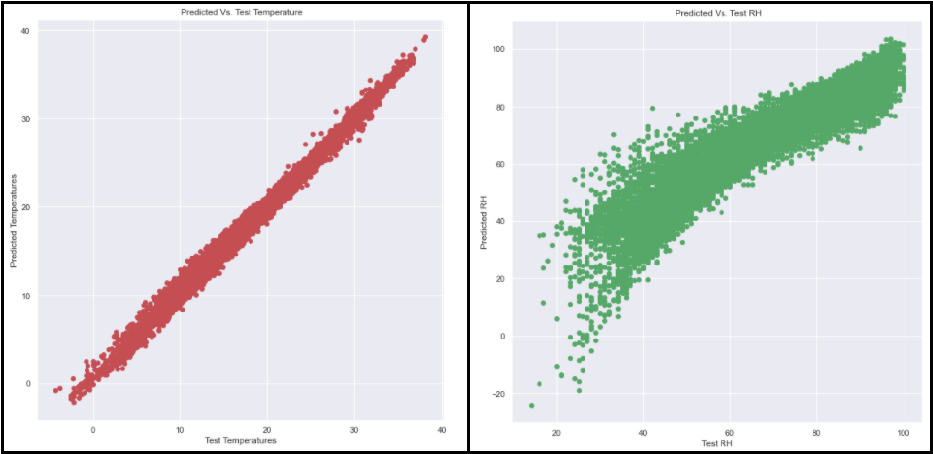
SVR has similar prediction error levels as that of Linear regression when it comes to RH prediction while the temperature prediction fares comparatively better. The RH prediction distribution for both Multi-Output and Chained multi-output regressor wrappers is similar. Fig 34 and fig 36 shows the deviations in the RH predictions. Fig 33 and fig 35 depicts the prediction performance of the SVR model for random datapoints.



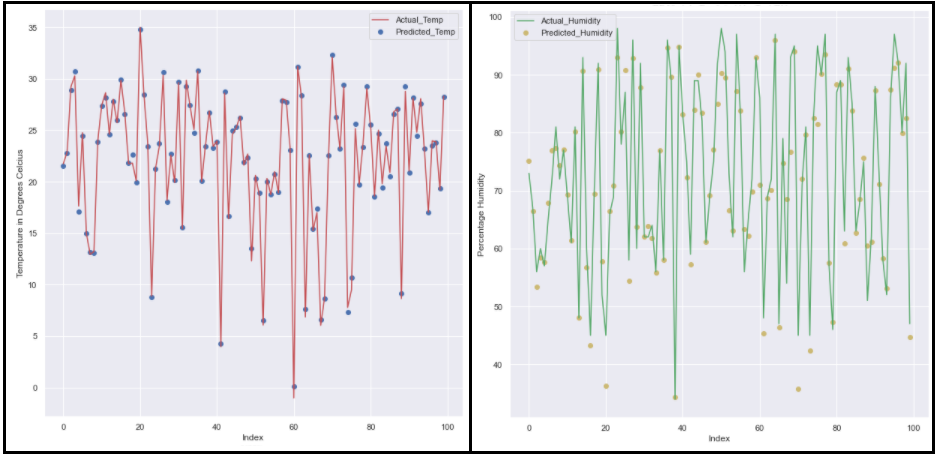
**Figure 33. Predicted Vs Test Scatter Plots for Temperature and RH (SVR- Multioutput Wrapper)**



**Figure 34. Prediction Performance of the model for Temperature and RH (SVR- Multioutput Wrapper)**

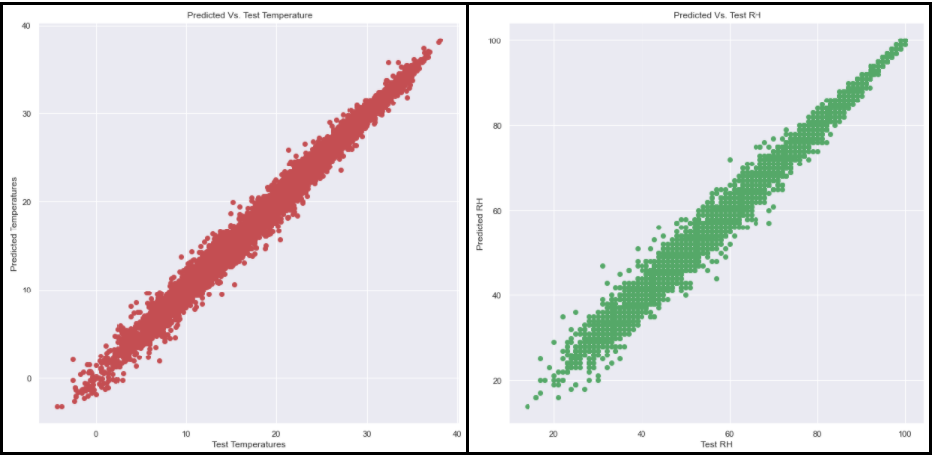


**Figure 35. Predicted Vs Test Scatter Plots for Temperature and RH (SVR- Regressor Chain)**

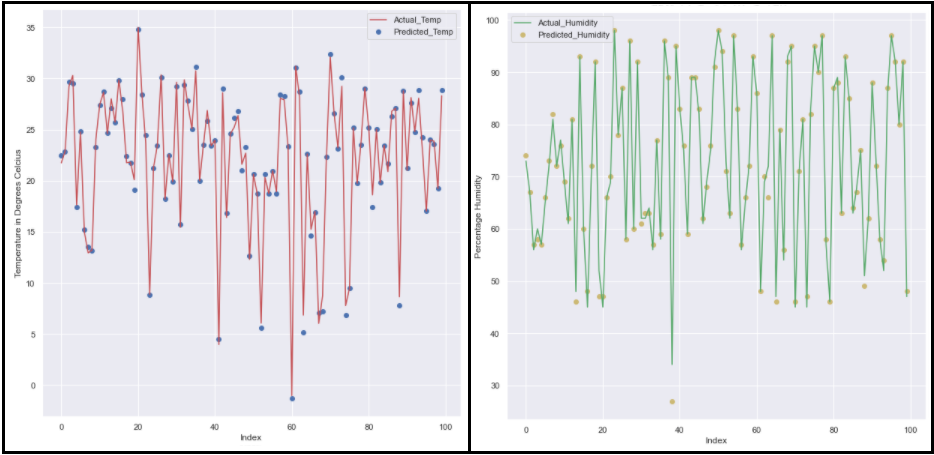


**Figure 36. Prediction Performance of the model for Temperature and RH (SVR – Regressor Chain)**

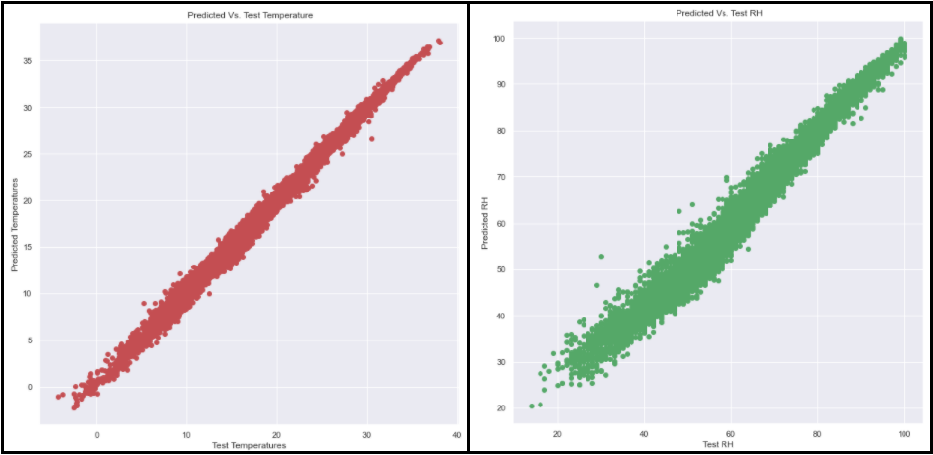
DT and GBT yielded similar results for temperature predictions but GBT ended up having lower prediction efficiency in terms of RH as shown in fig 37 and fig 38. Overall, DT model is the third best model in terms of both temperature and RH prediction in this analysis.



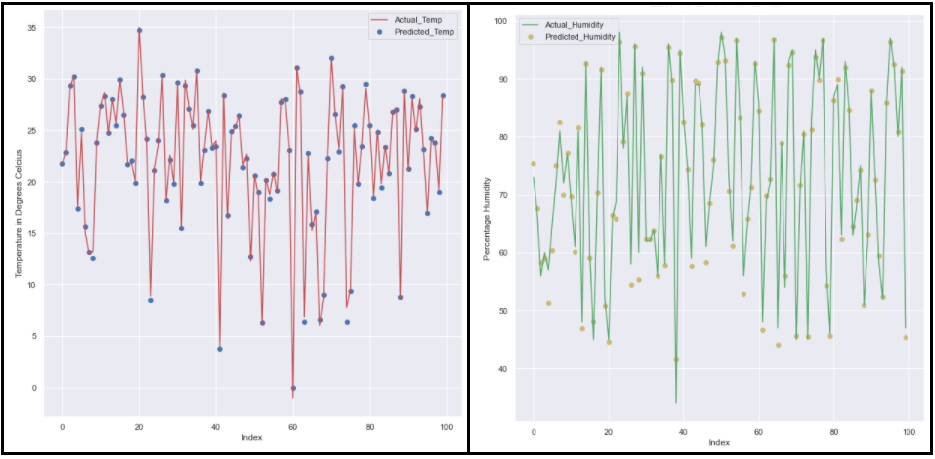
**Figure 37. Predicted Vs Test Scatter Plots for Temperature and RH (DT)**



**Figure 38. Prediction Performance of the model for Temperature and RH (DT)**



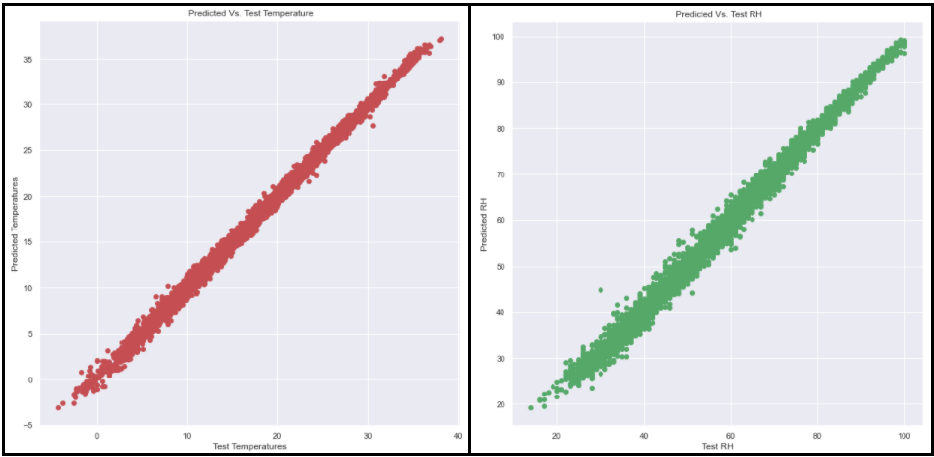
**Figure 39. Predicted Vs Test Scatter Plots for Temperature and RH (GBT)**



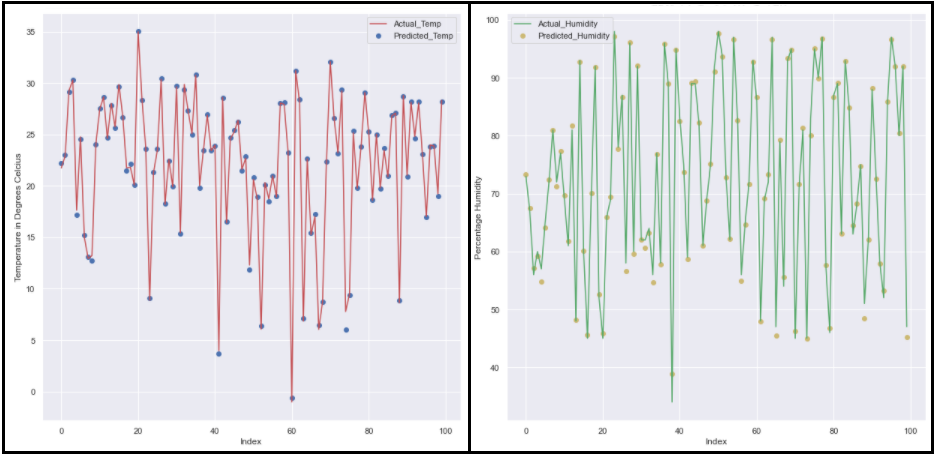
**Figure 40. Prediction Performance of the model for Temperature and RH (GBT)**

### **4.2.2 Gradient Boosted Trees – LightGBM**

The model trained using the Microsoft’s LightGBM based Gradient Boosted Trees with histogram binning is the second best model in terms of temperature and RH prediction metrics. It is also the fastest trained model and it only lags behind the RF model in terms of MAPE in RH prediction. The model performance is illustrated in fig 41 and fig 42.



**Figure 41. Predicted Vs Test Scatter Plots for Temperature and RH (GBT - LightGBM**)



**Figure 42. Prediction Performance of the model for Temperature and RH (GBT – LightGBM)**

## **4.3 Evaluation of the effectiveness of the control system**

For the purpose of this evaluation, a hypothetical DC of 1 MW rating with 50% IT load is considered. A load of 40% is considered for the traditional mechanical cooling system according to (Song, Zhang and Eriksson, 2015) which accounts for 400kW of the total load. In the remaining 100kW, 10% is considered for the operation of the auxiliary systems in the DC. The load ratings of the components involved in Air Side Economizer for a 1MW DC is assumed as shown in Table 6. The type and ratings of the components are obtained from a similar study for sub tropically located DC in Turkey (Gözcü and Erden, 2019).

The PUE in both the cases involving economizer and mechanical cooling options is calculated over a period of 365 days selected randomly from the validation set. A complete isolation of hot and cold aisles is assumed in this case. For the calculation of carbon footprint, the values of 0.038 m2/kW and 0.31 m2/kW are considered for DASFC and mechanical cooling options according to the DASFC economizer recommendations in (Niemann, Bean and Avelar, 2013) The various statistical readings with respect to the considered one year period spanning over 8760 hours is illustrated in table 8 and the number of hours each of the supporting systems runs is depicted in table 7.

|  |  |  |  |
| --- | --- | --- | --- |
| **Component** | **Type** | **Quantity** | **Rating** |
| Intake Fans | 744 | 1 | 39kW (87.9 m3/s air flow) |
| Exhaust Fans | 744 | 1 | 39kW |
| Evaporative Assistance and Humidity Control | 641 | 1 | 124kW |

**Table 6. Power Ratings of DASFC components** (Gözcü and Erden, 2019)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Criteria** | RH > 80% | RH < 48% | 32°C < T < 40°C | T > 48°C | T < 5°C |
| **Supporting System Involved** | Exhaust Air Recirculation (No Power Consumed) | Humidity Control | Vapour Assist |  | Exhaust Air Recirculation (No Power Consumed) |
| **No. of hours** | 3627 | 898 | 222 | 0 | 104 |

**Table 7. Hourly distribution of the supporting system involvement over 365 days**

|  |  |  |  |
| --- | --- | --- | --- |
| **Max Temp Predicted** | **Max RH Predicted** | **Min Temp Predicted** | **Min RH Predicted** |
| 36.62°C | 99.98% | -1.9588°C | 20.24% |

**Table 8. Min-Max values predicted for T and RH over 365 days**

From the PUE calculations (see APPENDIX), it is evident that there is a considerable improvement in the DC PUE levels when the DSS based system is employed for cooling control. A PUE improvement of 0.61 is observed between conventional cooling without DSS and multimode cooling with DSS controller with PUE values of 1.81 and 1.20, respectively. Power saving in the range of 2681840 kWhrs is observed for the considered scenario by employing the DSS controller based economizer operation. This is a considerable improvement which in turn reduces the amount of carbon footprint accordingly.

It can be observed that for the considered period of 365 days, the DC can be run in full DASFC economizer mode with the help of humidity regulators and exhaust air recirculation thereby eliminating the need to involve mechanical cooling component during this period. The DSS based control system is designed to give maximum priority to the free cooling option which enabled a considerable improvement in the PUE level as well as decrease in the carbon footprint. The exhaust air recirculation for maintaining the DC setpoints has also resulted in considerable power savings.

# **Chapter 5 – Conclusion and Future Work**

As part of this study, a novel DASFC based economizer implementation using a machine learning centered DSS cooling control scheme is recommended as one of the solutions for reducing the DCs cooling related impact on the environment. The impact of the rapidly growing DC infrastructure on the environment was also studied. The literature corresponding to the green cloud computing practices especially the implementation of the free cooling techniques in DCs to reduce the carbon footprint was explored. A sub tropical DC location is selected, and the historical meteorological data was collected for the region in order to create a DSS system that can be used in the economizer based multi mode cooling scheme for reliable operation of the DC cooling infrastructure.

Various data exploration and processing techniques were used for the analysis of the data. Different ML algorithms were considered for the creation of a Multi Target Prediction model that can be used for determining the temperature and RH levels needed for DSS based controller. The error metrics and prediction performance of these models was compared, and RF trained model was selected for the DSS logic. A novel operational scheme for the DSS based cooling controller was proposed which could seamlessly operate between different cooling modes ensuring maximum reliability.

A hypothetical DC was considered in which the results of the prediction for a period of 365 days and the DSS controller scheme were used to quantify its effectiveness. The total savings in terms of kWhrs and the improvement in PUE levels between the traditional cooling and DSS based multimode cooling were demonstrated. The reduction in the corresponding carbon footprint was also illustrated. It is also observed that using the DSS system for ASHRAE level 3 recommended operating thresholds for cooling the DC will eliminate the need for Mechanical cooling backup for most of the scenarios.

As of now, the proposed DSS system concentrated on multi mode system cooling provision reliability with priority to increase the economizer operational hours. This can be extended to include additional parameters from the auxiliary systems and IT room to further expand the switching scheme based on IT load, VM states and other factors. Inclusion of renewable sources for humidity and evaporative assistance systems can also be considered to further improving the PUE levels and reduce associated environmental impact.

# **References**

Agarwal, S. and Nath, A. (2011) ‘“Green computing - A new horizon of energy efficiency and electronic waste minimization”: A global perspective’, *Proceedings - 2011 International Conference on Communication Systems and Network Technologies, CSNT 2011*, pp. 688–693. doi: 10.1109/CSNT.2011.148.

*Anaconda | Individual Edition* (2020). Available at: https://www.anaconda.com/products/individual (Accessed: 12 October 2020).

Anwar, M. (2013) ‘Green Computing and Energy Consumption Issues in the Modern Age’, *IOSR Journal of Computer Engineering*, 12(6), pp. 91–98. doi: 10.9790/0661-1269198.

Appasami, G., College, P. E. and K, S. J. (2011) ‘Optimization of Operating Systems towards Green Computing’, *International Journal of Combinatorial Optimization Problems and Informatics*, 2(3), pp. 39–51.

Ari, A. A. A. *et al.* (2017) ‘Efficient and Scalable ACO-Based Task Scheduling for Green Cloud Computing Environment’, *Proceedings - 2nd IEEE International Conference on Smart Cloud, SmartCloud 2017*, pp. 66–71. doi: 10.1109/SmartCloud.2017.17.

Berral, J. L. *et al.* (2010) ‘Towards energy-aware scheduling in data centers using machine learning’, *Proceedings of the e-Energy 2010 - 1st Int’l Conf. on Energy-Efficient Computing and Networking*, 2, pp. 215–224. doi: 10.1145/1791314.1791349.

Brownlee, J. (2020) *How to Develop Multi-Output Regression Models with Python*, *Machine Learning Mastery*. Available at: https://machinelearningmastery.com/multi-output-regression-models-with-python/ (Accessed: 20 October 2020).

Buyya, R., Beloglazov, A. and Abawajy, J. (2010) ‘Energy-Efficient Management of Data Center Resources for Cloud Computing : A Vision , Architectural Elements , and Open Challenges Clou d Computing and D istributed S ystems ( CLOUDS ) Laboratory Department of Computer Science and Software Engineering The’, *The University of Melbourne, Australia*, (Vm), pp. 1–12.

Capozzoli, A. and Primiceri, G. (2015) ‘Cooling systems in data centers: State of art and emerging technologies’, *Energy Procedia*. Elsevier B.V., 83, pp. 484–493. doi: 10.1016/j.egypro.2015.12.168.

Chi, C. *et al.* (2020) ‘Jointly Optimizing the IT and Cooling Systems for Data Center Energy Efficiency based on Multi-Agent Deep Reinforcement Learning’, pp. 489–495. doi: 10.1145/3396851.3402658.

Chung, K., Kalbarczyk, Z. T. and Iyer, R. K. (2018) ‘Poster: Indirect cyber attacks by perturbation of environment control: A data driven attack model’, *ACM International Conference Proceeding Series*, p. 2016. doi: 10.1145/3190619.3191681.

Daraghmeh, H. M. and Wang, C. C. (2017) ‘A review of current status of free cooling in datacenters’, *Applied Thermal Engineering*. Elsevier Ltd, 114, pp. 1224–1239. doi: 10.1016/j.applthermaleng.2016.10.093.

*Decision Trees — scikit-learn 0.24.0 documentation* (2020) *Scikit Learn Documentation*. Available at: https://scikit-learn.org/stable/modules/tree.html#tree (Accessed: 27 December 2020).

*Dry Bulb, Wet Bulb and Dew Point Temperatures* (2020). Available at: https://www.engineeringtoolbox.com/dry-wet-bulb-dew-point-air-d\_682.html (Accessed: 22 December 2020).

Elgelany, A. and Nada, N. (2013) ‘Energy Efficiency for Data Center and Cloud Computing : A Literature Review’, *International Journal of Engineering and Innovative Technology*, 3(4), pp. 16–20.

*Ensemble methods — scikit-learn 0.24.0 documentation* (2020) *Scikit Learn Documentation*. Available at: https://scikit-learn.org/stable/modules/ensemble.html#gradient-boosting (Accessed: 6 October 2020).

Farahnakian, F., Liljeberg, P. and Plosila, J. (2014) ‘Energy-efficient virtual machines consolidation in cloud data centers using reinforcement learning’, *Proceedings - 2014 22nd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2014*. IEEE, pp. 500–507. doi: 10.1109/PDP.2014.109.

Gao, X. *et al.* (2017) ‘Why “some” Like It Hot Too’, *Performance Evaluation Review*, 45(1), pp. 23–24. doi: 10.1145/3078505.3078545.

Glasser, P. M. *et al.* (2016) ‘Energy efficient VM migration revisited: SLA assurance and minimum service disruption with available hosts’, *2015 12th International Conference on High-Capacity Optical Networks and Enabling/Emerging Technologies, HONET-ICT 2015*, pp. 22–27. doi: 10.1109/HONET.2015.7395427.

Google Colab (2020) *Welcome To Colaboratory - Colaboratory*. Available at: https://colab.research.google.com/notebooks/intro.ipynb (Accessed: 10 September 2020).

Goyal, Y., Arya, M. S. and Nagpal, S. (2016) ‘Energy efficient hybrid policy in green cloud computing’, *Proceedings of the 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015*. IEEE, pp. 1065–1069. doi: 10.1109/ICGCIoT.2015.7380621.

Gözcü, O. and Erden, H. S. (2019) ‘Energy and economic assessment of major free cooling retrofits for data centers in Turkey’, *Turkish Journal of Electrical Engineering and Computer Sciences*, 27(3), pp. 2197–2212. doi: 10.3906/elk-1811-177.

Hambali, M. A. *et al.* (2019) ‘Adaboost Ensemble Algorithms for Breast Cancer Classification’, *Journal of Advances in Computer Research Quarterly*, 10(2), pp. 1–10. Available at: www.jacr.iausari.ac.ir.

Hossain, M. S. *et al.* (2017) ‘A belief rule based expert system for datacenter pue prediction under uncertainty’, *IEEE Transactions on Sustainable Computing*, 2(2), pp. 140–153. doi: 10.1109/TSUSC.2017.2697768.

Hu, Y. J. *et al.* (2012) ‘Decision tree-based learning to predict patient controlled analgesia consumption and readjustment.’, *BMC medical informatics and decision making*, 12, p. 131. doi: 10.1186/1472-6947-12-131.

*Humidity | North Carolina Climate Office* (2020) *North Carolina Climate Office*. Available at: https://climate.ncsu.edu/edu/Humidity (Accessed: 2 October 2020).

IBM Cloud Education (2020) *What is Machine Learning? | IBM*, *IBM Cloud Learn Hub*. Available at: https://www.ibm.com/cloud/learn/machine-learning (Accessed: 16 December 2020).

Jin, X. *et al.* (2016) ‘Green Data Centers: A Survey, Perspectives, and Future Directions’. Available at: http://arxiv.org/abs/1608.00687.

Karthik, C. *et al.* (2016) ‘Green intelligence for cloud data centers’, *2016 3rd International Conference on Recent Advances in Information Technology, RAIT 2016*. IEEE, pp. 591–597. doi: 10.1109/RAIT.2016.7507965.

Van Le, D. *et al.* (2019) ‘Control of air free-cooled data centers in tropics via deep reinforcement learning’, *BuildSys 2019 - Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 306–315. doi: 10.1145/3360322.3360845.

Lee, K. P. and Chen, H. L. (2013) ‘Analysis of energy saving potential of air-side free cooling for data centers in worldwide climate zones’, *Energy and Buildings*. Elsevier B.V., 64, pp. 103–112. doi: 10.1016/j.enbuild.2013.04.013.

Li, Y. *et al.* (2020) ‘Transforming Cooling Optimization for Green Data Center via Deep Reinforcement Learning’, *IEEE Transactions on Cybernetics*. IEEE, 50(5), pp. 2002–2013. doi: 10.1109/TCYB.2019.2927410.

Lin, P., Niemann, J. and Long, L. (2015) ‘Choosing Between Direct and Indirect Air Economization for Data Centers’, *Schneider Electric*. Available at: http://www.thegreengrid.org.

Manousakis, I. *et al.* (2015) ‘CoolProvision: Underprovisioning datacenter cooling’, *ACM SoCC 2015 - Proceedings of the 6th ACM Symposium on Cloud Computing*, pp. 356–367. doi: 10.1145/2806777.2806938.

Maurya, K. and Sinha, R. (2013) ‘Energy Conscious Dynamic Provisioning of Virtual Machines using Adaptive Migration Thresholds in Cloud Data Center’, *Ijcsmc*, 2(3), pp. 74–82. Available at: www.ijcsmc.com.

MeteoBlue-HistoryPlus (2020) *Weather history download Orlando - meteoblue*. Available at: https://www.meteoblue.com/en/weather/archive/export/orlando\_united-states-of-america\_4167147 (Accessed: 12 October 2020).

Mohiddin, S. K. (2016) ‘Green Computing an Eco Friendly It Environment for Upcoming Technologies Green Computing an Eco Friendly It Environment for Upcoming Technologies’, (April 2015).

Nagpal, S. (2018) ‘Comparative Analysis of the Power Consumption Techniques in the Cloud Computing Environment’, 8(8), pp. 32–49.

Niemann, J., Bean, J. and Avelar, V. (2013) ‘Economizer Modes of Data Center Cooling Systems’, *White Paper 132*, p. 19.

Niemann, J., Brown, K. and Avelar, V. (2020) ‘Impact of Hot and Cold Air containment.pdf’. Schneider Electric.

Nixon, J. S., Francis, A. and Devaraj, S. (2015) ‘Green Computing: Awareness, Current Issues and best practices’, *International Journal of Computer Systems, ISSN*, 520(12), pp. 2394–1065. Available at: http://www.ijcsonline.com/.

Oracle EU (2020) *What is Artificial Intelligence (AI)? | Oracle Ireland*, *Oracle EU Publications*. Available at: https://www.oracle.com/ie/artificial-intelligence/what-is-ai/ (Accessed: 10 October 2020).

Oró, E. *et al.* (2015) ‘Overview of direct air free cooling and thermal energy storage potential energy savings in data centres’, *Applied Thermal Engineering*, 85, pp. 100–110. doi: 10.1016/j.applthermaleng.2015.03.001.

Pang, B. *et al.* (2017) ‘Statistical Downscaling of Temperature with the Random Forest Model’, *Advances in Meteorology*, 2017. doi: 10.1155/2017/7265178.

Park, S. and Seo, J. (2018) ‘Analysis of air-side economizers in terms of cooling-energy performance in a data center considering exhaust air recirculation’, *Energies*, 11(2). doi: 10.3390/en11020444.

Pawlish, M. and Varde, A. S. (2010) ‘A decision support system for green data centers’, *International Conference on Information and Knowledge Management, Proceedings*, pp. 47–55. doi: 10.1145/1871902.1871912.

Pedregosa, F. *et al.* (2011) *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research*. Available at: https://scikit-learn.org/stable/about.html#citing-scikit-learn (Accessed: 25 December 2020).

*Predictive Analytics Software | RapidMiner Studio* (2020). Available at: https://rapidminer.com/products/studio/?campaignid=RapidMiner Branded - Worldwide--Brand - US&gclid=Cj0KCQiAifz-BRDjARIsAEElyGLDzEfm04Q\_AZIiJCpy\_TJCiBXt5zFxE0O8A18NOSJ9mDZV0phAh1kaAsHcEALw\_wcB (Accessed: 10 June 2020).

Rezai, H. and Speily, O. R. B. (2017) ‘Energy aware resource management of cloud data centers’, *International Journal of Engineering, Transactions B: Applications*, 30(11), pp. 1730–1739. doi: 10.5829/ije.2017.30.11b.14.

S, I. P., Niranjan, V. and N, S. R. (2014) ‘Improved Performance through Carbon Aware Green Cloud Policy’, pp. 10–15.

Saxena, R. (2017) *How Decision Tree Algorithm works*. Available at: https://dataaspirant.com/how-decision-tree-algorithm-works/ (Accessed: 27 December 2020).

Song, Z., Zhang, X. and Eriksson, C. (2015) ‘Data Center Energy and Cost Saving Evaluation’, *Energy Procedia*. Elsevier B.V., 75, pp. 1255–1260. doi: 10.1016/j.egypro.2015.07.178.

Suresh, G. (2015) ‘Power Efficient Management In Data Centers Using Server Consolidation For Green Computing’, *Ijirst.Org*, 1(8), pp. 212–215. Available at: http://www.ijirst.org/articles/IJIRSTV1I8068.pdf.

*Tableau Desktop* (2020). Available at: https://www.tableau.com/products/desktop (Accessed: 10 August 2020).

Tafani, D. *et al.* (2013) ‘Towards energy efficiency for cloud computing services’, *Communication Infrastructures for Cloud Computing*, (Vm), pp. 306–328. doi: 10.4018/978-1-4666-4522-6.ch014.

Thakur, D. S. (2016) ‘Energy Efficient Task Scheduling Algorithms In Cloud Data Center Energy Efficient Task Scheduling Algorithms In Cloud Data Center’, 6(December), pp. 12–18.

The Green Grid (2012) ‘Updated Air-Side Free Cooling Maps: The Impact of ASHRAE 2011 Allowable Ranges’, *The Green Grid*. Available at: http://www.thegreengrid.org/~/media/WhitePapers/WP46UpdatedAirsideFreeCoolingMapsTheImpactofASHRAE2011AllowableRanges.pdf.

Torell, W., Brown, K. and Avelar, V. (2016) ‘The Unexpected Impact of Raising Data Center Temperatures’, *Write paper 221, Revision 0*.

Uddin, M., Shah, A. and Memon, J. (2014) ‘Energy efficiency and environmental considerations for green data centres’, *International Journal of Green Economics*, 8(2), pp. 144–157. doi: 10.1504/IJGE.2014.065852.

Vikram, S. (2016) ‘Green computing’, *Proceedings of the 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015*, 7(2), pp. 767–772. doi: 10.1109/ICGCIoT.2015.7380566.

Wilcox, G. (2020) *Orlando - Environmental Air Quality*, *Griffinairquality*. Available at: https://griffinairquality.weebly.com/orlando.html (Accessed: 16 December 2020).

Woods, A. (2010) ‘Cooling the data center’, *Communications of the ACM*, 53(4), pp. 36–42. doi: 10.1145/1721654.1721671.

Zapater, M. *et al.* (2015) ‘Leakage-Aware Cooling Management for Improving Server Energy Efficiency’, *IEEE Transactions on Parallel and Distributed Systems*, 26(10), pp. 2764–2777. doi: 10.1109/TPDS.2014.2361519.

Zhang, H. *et al.* (2014) ‘Free cooling of data centers: A review’, *Renewable and Sustainable Energy Reviews*, 35, pp. 171–182. doi: 10.1016/j.rser.2014.04.017.

Zhou, J. and Turner, W. D. (2008) ‘Airside Economizer – Comparing Different Control Strategies and Common Misconceptions’, *Building*.

Zou, Z., Yu, X. and Ergan, S. (2020) ‘Towards optimal control of air handling units using deep reinforcement learning and recurrent neural network’, *Building and Environment*. Elsevier Ltd, 168(October 2019), p. 106535. doi: 10.1016/j.buildenv.2019.106535.

# **APPENDIX**

## **PUE Calculation for Conventional Cooling System**

Mechanical Cooling System Power consumption = 3504000 kWhrs

Auxiliary Equipment related consumption = 87600 kWhrs

IT Equipment Power Consumption = 4380000kWhrs

PUE = Total Power Consumption in DC/Power Consumed by the IT Equipment

= 7971600/4380000

= 1.81

## **PUE Calculation for Novel DSS based Controller**

Economizer Fans Power Consumption = 683280 kWhrs

Humidity Control and Vapour Assist related power consumption = 138880 kWhrs

Auxiliary Equipment related consumption = 87600 kWhrs

IT Equipment Power Consumption = 4380000kWhrs

PUE = Total Power Consumption in DC/Power Consumed by the IT Equipment

= 5289760/4380000

= 1.20