Title
Energy Demand Data Analysis and Prediction Using Machine Learning for a Campus Dining Facility: Segundo Dining Commons

Author
Khadka, Pratik Raj

Publication Date
2023

Peer reviewed|Thesis/dissertation

Energy Demand Data Analysis and Prediction Using Machine Learning for a Campus Dining Facility: Segundo Dining Commons

By

PRATIK RAJ KHADKA

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Energy Systems

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Kurt Kornbluth, Chair

Alan Meier

Alan Jenn

Committee in Charge

2023

ACKNOWLEDGEMENTS

I would like to extend my heartfelt gratitude to my major advisor, Professor Kurt Kornbluth, and my thesis committee advisors Professor Alan Meier and Professor Alan Jenn whose insightful guidance, valuable feedback, and unwavering support were crucial in making this thesis possible. This thesis work would not have been possible without their expertise in the domain of Energy Efficiency and Energy Data Analysis.

I am also deeply grateful to the Facilities Management: Energy Conservation Office and their engineers for providing me the opportunity to work in the campus buildings looking at the opportunities for energy usage optimization. I express my special thanks to Daniel Colvin for guiding me in developing the machine learning models and helping me tackle every data science issue I encountered during this thesis research. I am very thankful to Hiroko Masuda, Nicolas Fauchier-Magnan, Joshua Morejohn for shaping my understanding of building energy systems and helping me formulate the research questions for this thesis.

I would also like to acknowledge the Student Housing and Dining Services, who supported my graduate studies and research and provided me the opportunity to look at their facilities for energy saving opportunities. I am thankful to all the staff within this team that worked with me and contributed to my academic and professional growth.

Finally, I would like to thank my parents for their unconditional love, support, and encouragement. Their constant belief in me and their unwavering support has been the driving force behind my academic and personal accomplishments, and I am forever grateful for their presence in my life.

ABSTRACT

With 30% of the world's final energy consumption and 40% of carbon dioxide emissions being attributed to the building sector, accurately predicting building energy consumption has become essential for various energy management applications such as identification of energy efficiency measures. With the increasing availability of smart utility meters and data related to building energy consumption, multiple techniques in Artificial Intelligence (AI) such as machine learning are being successfully applied to identify building energy usage patterns and develop focused energy efficiency plans. The UC Davis campus buildings are highly information-intensive and effective interpretation of this building data can help identify energy demand patterns, predict future energy demands, and plan varied types of energy saving strategies. This thesis performs energy demand prediction using multiple machine learning models and analyzes the prediction outcome to identify major areas to focus energy efficiency efforts at a student dining facility on campus: Segundo Dining Commons. Six machine learning models for each of the three energy commodities: Steam, Chilled Water, and Electricity were created and the best-performing model for each commodity was selected to generate a prediction of future demand. The model developed was also fed with four different meteorological scenarios: 0.5°C, 1°C, 2°C rise in outside air temperatures and a typical year with max recorded temperatures for each day of each hour in the past 5 years. The major observations made were simultaneous heating and cooling demands in summer, high energy demands even during school breaks, constant electricity demands and minor changes in demand for different temperature scenarios. Broad energy saving areas were then identified from the observations that can help develop focused energy efficiency plans.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

Chapter 1: INTRODUCTION

1.1 Motivation

In 2021 the operation of buildings accounted for 30% of global final energy consumption and 27% of total energy sector emissions (IEA, Buildings, 2022) The built environment generates 40% of annual global CO_2 emissions of which building operations are responsible for 27% (9.9 GT) annually while building and infrastructure material and construction are responsible for an additional 13% (4.7 GT) of emissions annually. (IEA, Buildings, 2022) These emissions from the built environment are expected to increase with the rise in the building stock if strict energy efficiency measures are not implemented. In 2021, 56.61% of the world's population lived in cities and this trend is expected to continue and reach 68% by 2050, at which point nearly 7 of 10 people will live in cities (United Nations Department of Economic and Social Affairs, 2022) With this large wave of urban growth, global building floor area is expected to double by 2060, it is expected that 2.4 trillion sq. ft. of new floor area will be added to the existing building stock, the equivalent of adding an entire New York City to the world, every month for 37 years. (United Nations Environment Programme, 2022) With the expected rise in the global urban population, building energy efficiency will be an important challenge to address as we plan to meet climate targets. In 2040, approximately 2/3 of the global building stock will be buildings that exist today, so without widespread existing building decarbonization across the globe, these buildings will still be emitting significant amounts of CO_2 in 2040 if we do not increase building energy efficiency and increase the use renewable energy. (IEA, Energy Technology Perspectives, 2020)

The construction of energy efficient buildings and improvement of current buildings' energy usage is important to contribute to the reduction of CO_2 emissions and reduce global warming. In this regard, data-driven models are useful tools for optimization of Energy Management System (EMS) and Heating, Ventilation and Air Conditioning system (HVAC). (S. Seyedzadeh, 2018) UC Davis buildings are rich in data, with variety of sensors recording instantaneous data for years.

Therefore, there is a great opportunity for analysis of building energy consumption patterns and prediction of future energy demand using the data-driven models to facilitate energy efficiency measures and energy planning.

1.2 Problem Statement

UC Davis Energy and Engineering Office under the Facilities Management have been able to implement a variety of energy conservation measures in campus buildings and have been successful in optimizing energy usage in them, saving thousands of dollars in utility bills. (Energy and Engineering, 2023) However, the focus had never been on the auxiliary units in campus. Campus auxiliary units are the facilities that pay their own utility bills and are not funded through the campus funds. The campus auxiliary units include student housing and dining buildings, student community centers, recreational facilities, etc. Among these auxiliaries, student dining facilities are one of the top energy consumers that are not yet explored for energy conservation. Segundo Dining Commons is the biggest in terms of energy usage among the dining facility with utility bills of around $300,000 per year. The facility has kitchens with large dining spaces and office spaces with equipment serving the area 24 hours every day.

Despite such high energy usage and utility costs, the facility is not explored for energy efficiency measures yet. All the HVAC equipment in the facility is kept on all the time and no energy efficiency measures have been implemented. From the experience of the Energy and Engineering team's energy conservation project with other similar facilities on campus, it is safe to say that there should exist multiple measures that can conserve energy in the facility.

Segundo Dining Commons has most of the equipment under direct digital control and uses the SIEMENS building management system for managing the HVAC system. The data from BMS which includes temperature, humidity, CO_2 data is stored in PI system (data storage and analysis tools developed by software manufacturer OSIsoft) and predates many years. While all the resources exist for Segundo Dining Commons for identification of energy saving opportunities, it had never been explored before. Identifying energy usage patterns in this building can help identify strategies to increase energy efficiency in the facility. The campus has long-term plans of electrifying its buildings, and understanding the time of use demands in the building could help demand response and other electrification initiatives. Therefore, there exists an ideal opportunity of energy usage analysis in this building that has multiple benefits for the facility.

1.3 Research Questions

Based on the motivation for this research, the following questions were generated as the major research questions.

- What is the typical energy usage pattern of the dining commons? What can we learn broadly from the energy demand data about the building characteristics?
- What is the predicted future energy usage pattern of the building based on future climatic data? What broad energy saving strategies can be identified, by the predicted energy demand data?
- How does the demand change with changes in outside meteorological conditions based on the model?

The objective of this research is to inform the dining commons for better operation planning and identify strategies to optimize their energy efficiency.

Chapter 2: LITERATURE REVIEW

2.1 Importance of Analyzing Building Energy Usage Data

Analyzing building energy usage data is an essential step toward achieving energy efficiency and sustainability goals. Energy data analytics provides valuable insights into the consumption patterns of a building and can help identify areas of inefficiency. Building Management System(BMS) is capable of generating a large amount of data through the use of sensors that can allow smart energy control (Shaikh, Bin Mohd Nor, Nallagownden, Elamvazuthi, & Ibrahim, 2014), fault detection in the system (Magoulès, Zhao, & Elizondo, 2013) and potential options of energy efficiency and calculation of achieved energy savings. Analyzing energy usage data can also aid in the development of energy management strategies that are tailored to the specific needs of a building. It can help optimize HVAC systems to save energy consumption and reduce operational costs through accurate estimation of heating and cooling loads. This is even more significant for air-conditioned buildings that use thermal energy storage, where such predictions are necessary to optimize the system. However, calculating loads, especially for non-domestic buildings, can be expensive and time-consuming for consulting firms, as noted by Kalogirou (Kalogirou, et al., 2001). As a result, analyzing energy usage data instead of calculating loads of the building to identify energy efficiency opportunities can help operate HVAC systems effectively, providing comfortable temperature and humidity conditions, as suggested by Kumar (Kumar, Aggarwal, & Sharma, 2013). Additionally, advanced forecasting of electricity loads through building's historical energy data can help identify excessive usage periods, reduce peak demand, and ease the load of the electrical HVAC system. (Seyedzadeh, Rahimian, Glesk, & Roper, 2018) These studies highlight the importance of analyzing building energy usage data and demonstrate how data analytics can be leveraged to improve energy efficiency.

In addition to the benefits of energy savings and environmental impact reduction, analyzing building energy usage data can also help building owners and managers make informed decisions regarding building operations and maintenance. In their study, John (2007) used energy usage data analytics to assess the performance of HVAC systems in commercial buildings. The analysis revealed several issues that were affecting the efficiency of the HVAC system, including ineffective controls and equipment failure (Seem, 2007). The analysis of energy usage data can also aid in the identification of equipment failures and maintenance needs. In a study by Magoules et al. (2013), energy usage data was used to identify malfunctioning equipment, allowing for prompt repair and reduced downtime. These examples highlight the benefits of energy usage data analytics in building operations and maintenance and underscore the importance of data analysis in ensuring the efficient and sustainable operation of buildings.

Two main modeling approaches are prevalent in the forecasting of building energy consumption:

1. Engineering-based models (white-box models): These models estimate building energy consumption by mainly using physical characteristics of buildings (building construction details, floor layout, building operational details, etc.) and environmental parameters (Outdoor climate data). EnergyPlus, TRANSYS are some examples of software used for this calculation.
2. Data-driven models (black-box models): These models use historical data available from the buildings to train Machine Learning (ML) algorithms in order to make forecasts and create valuable information from the building's usage data. They use historical data of energy usage and meteorological information to make predictions. They are called black—box models because they do not use data from inside the building and treat it as a black-box. They typically use three main inputs: historical energy consumption data in time-series format, outdoor meteorological data and time index variables that differentiate business days or time. In the last few years, they are being widely used in the field of sustainability and tackling climate change. (Khalil, McGough, Pourmirza, Pazhoohesh, & Walker, 2022)

This research uses data-driven models that use data science tools to predict energy consumption.

2.2 Data Science and Machine Learning in the Built Environment

Data science techniques are gaining popularity in the analysis of building energy usage. Data Science builds systems and algorithms to discover knowledge, detect patterns and generate insights and predictions from large-scale data that would not be possible otherwise. It involves the whole data analysis process that starts with data extraction and cleaning and extends to data analysis using different models and prediction of new values and their summarization. (Molina-Solana, Ros, Ruiz, Gómez-Romero, & Bautista, April 2017) The application of data science also involves managing and interpreting data in order to obtain valuable information. The process starts with raw data, after that it's important to clean the data and select specific segments that have valuable information. Different filters can be applied in the process to obtain data that can eliminate irrelevant information. After the data preparation, exploring the data is useful to decide the methods and algorithms that can be used to obtain the desired knowledge. The application of such algorithms can produce results that can guide decision making.

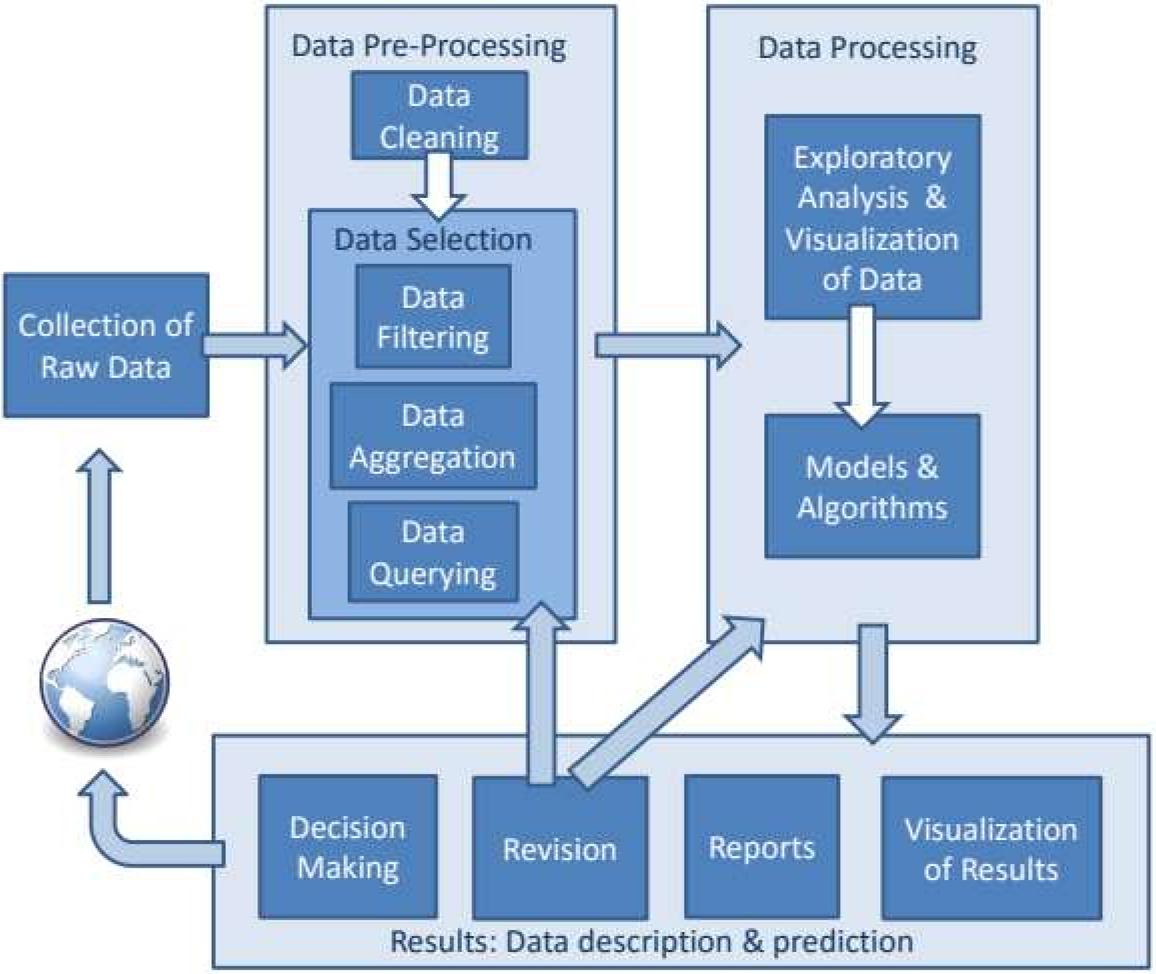


Figure 1 Data Science Process (Molina-Solana, Ros, Ruiz, Gómez-Romero, & Bautista, April 2017)

One of the popular data science techniques for building energy is machine learning, which can learn patterns from past data and apply those patterns to future predictions. Some of these most popular machine learning techniques are classification, clustering, regression, and neural networks. The widely used machine learning techniques in the built environment are as shown in the figure below.

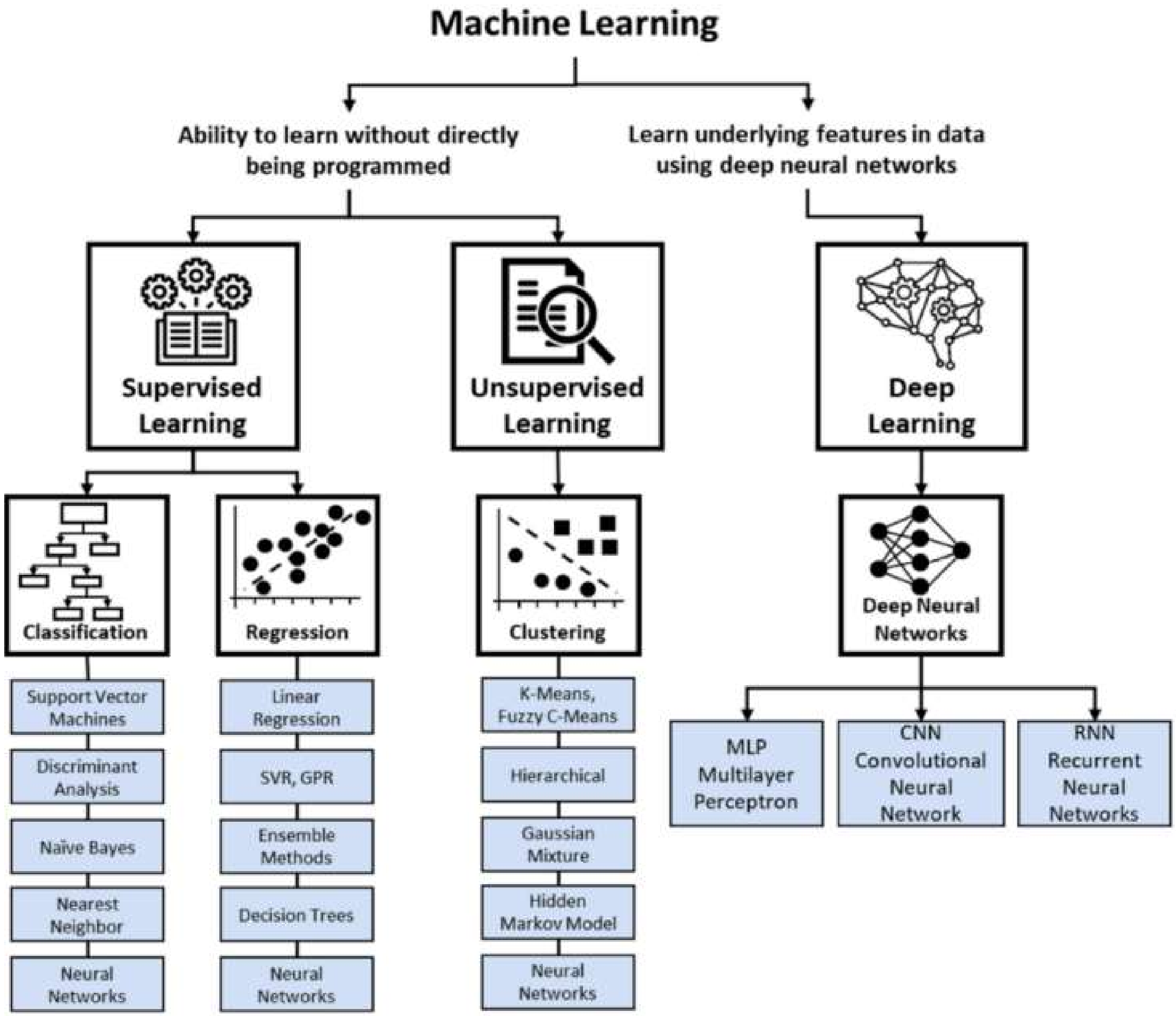


Figure 2 Machine Learning techniques in Building Energy (Tien & Shuangyu Wei, 2022)

Supervised machine learning algorithm like regression and classification are designed to be trained using datasets where the desired outcome is known. Clustering can be used to analyze energy consumption patterns, identify potential energy-saving opportunities, and classify buildings based on their energy performance while Regression can be used to predict the future energy usages learning from the correlations and patterns in the past energy usage data. The supervised machine learning algorithm tries to identify the patterns in the data and presents the solution in terms of the required dependent variable, whereas unsupervised machine learning attempts to analyze the data when the desired outcome is not known. It extracts the patterns or features in data and presents the hidden information in the overall data. Another type of machine learning is the deep learning which do not require any form of human intervention and can learn from its own mistakes.

Deep learning techniques are being increasingly applied to building energy data for improved energy management and cost reduction. One popular application of deep learning is in the forecasting of building energy consumption using historical data. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are a type of deep learning models used to analyze time series data, such as energy consumption and weather data, to predict future energy usage. They can also be used for anomaly detection, identifying areas of energy waste and inefficiency in buildings. This approach allows for targeted interventions to reduce energy consumption, improving overall energy efficiency. These techniques offer powerful tools for analyzing building energy usage and can help building operators and energy engineers make data-driven decisions to optimize energy efficiency, reduce costs, while improving occupant comfort. Tien et al. (2022) summarized the different applications of machine learning in the built environment sector as shown below.

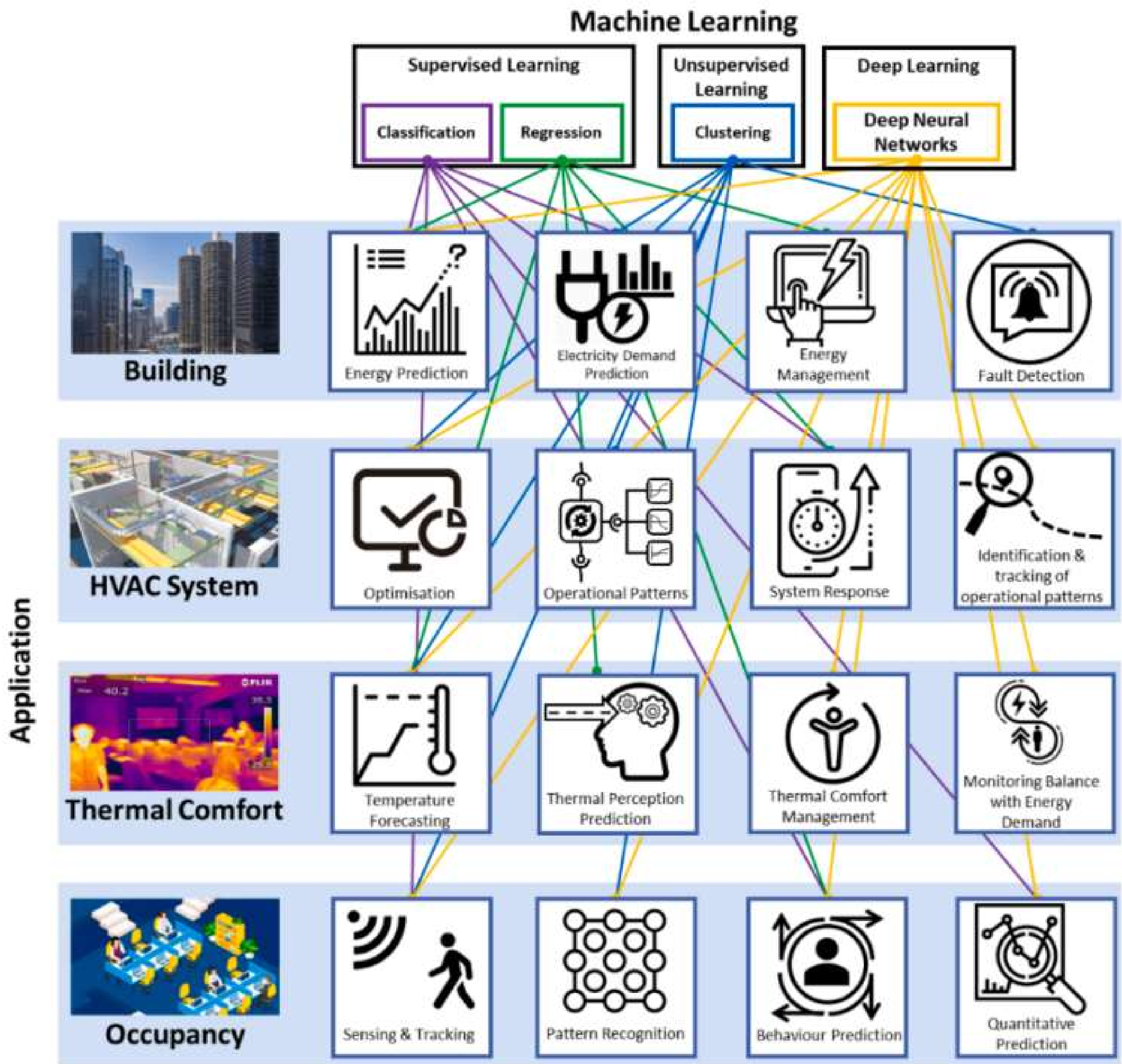


Figure 3 Applications of ML based methods in the built environment sector (Tien & Shuangyu Wei, 2022)

2.3 Popular ML Models in building energy use prediction

Some of the most widely used machine learning algorithms for energy use prediction are Random Forest (RF), Support Vector Machines (SVM) and Artificial Neural Networks (ANN). To see which model works best for the purposes of this research, we implemented a few models in the order of increasing complexity. The models used are Linear Regression, Decision Trees Random Forest, Gaussian Based Regressor, ANN and SVM.

Linear Regression

Linear regression is a statistical approach used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the data. Linear regression works by finding the best-fitting line through the data that minimizes the sum of the squared differences between the predicted and actual values. Linear regression is a simple yet powerful tool for analyzing and predicting continuous outcomes. Linear Regression is used in predicting the commodities demand for the purposes of this research.

Decision Trees

Decision tree regression builds a tree-like model of decisions and their possible consequences. Each node in the tree represents a decision based on a particular feature, and each branch represents the outcome of that decision. The leaves of the tree represent the predicted outcome of the model. Decision tree regression is a simple yet powerful tool for analyzing and predicting continuous outcomes. It can handle both numerical and categorical data, and it can capture non-linear relationships between variables, the decision tree model used in this research uses both numerical and categorical data to make prediction. Decision trees are easy to interpret and can be useful for identifying the most important features in a dataset. However, they can also be prone to overfitting and may not generalize well to new data.

Random Forest Model

Random forest is a supervised machine learning algorithm and is an ensemble learning method involving combining multiple decision tree models to improve the accuracy of predictions. It's prediction is usually better than any individual tree could make alone. Each decision tree works by recursively splitting the data based on the values of the input features, until a stopping criterion is met. They can be visualized as a combination of if-else statements that branches the data set until the final stopping criterion is met.

At a high level, the random forest algorithm works by creating multiple decision trees on random subsets of the training data and then combining the predictions of these trees to give the final prediction. The randomness in the algorithm comes from two sources: bootstrapped sampling and random feature selection.

Bootstrapped sampling involves randomly selecting a subset of the training data with replacement. This means that some data points will be selected multiple times, while others may not be selected at all. This helps to create more diverse trees, which in turn leads to a more robust model. Random feature selection involves selecting a subset of the features to be considered for splitting at each node in the decision tree. This helps to reduce overfitting and improve the performance of the model.

Once the decision trees are created, the random forest algorithm combines their predictions to make the final prediction. For classification problems, the most common class predicted by the decision trees is chosen as the final prediction. For regression problems, the average of the predictions made by the decision trees is taken as the final prediction.

Gaussian Based Regression model

Gaussian-based regression models the relationship between input variables and a continuous output variable as a weighted sum of Gaussian basis functions. In this approach, each basis function is centered at a particular point in the input space and has a width that controls how much influence it has on the output. The weights for each basis function are learned from the training data using maximum likelihood estimation or a Bayesian approach. Gaussian-based regression is a flexible and powerful method that can capture complex non-linear relationships between variables.

Support Vector Machine (SVM) model

Support Vector Machines (SVMs) are highly robust models for solving non-linear problems and used for both classification and regression purposes. SVMs work by finding the best line or boundary that separates two classes of data points. The boundary that SVMs find is the one that has the largest margin or distance between it and the closest data points from each class. The margin helps the SVMs generalize better to new, unseen data.

SVMs are based on the Structural Risk Minimization (SRM) principle that tries to find a hyperplane that is as far away as possible from the data points, while also trying to fit the data as closely as possible. In contrast to SVM classification, where the goal is to find a hyperplane that separates the two classes with the largest margin, the goal of SVM regression is to find a hyperplane that passes through as many data points as possible within a given margin.

In building sector, SVM is used for forecasting of cooling and heating loads, electricity consumption and classification of energy usage of buildings. (Seyedzadeh, Rahimian, Glesk, & Roper, 2018)

Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) are machine learning algorithm that are modeled after the structure and function of the human brain. ANNs consist of many interconnected nodes called neurons, organized into layers.

The first layer is called the input layer, which receives the input data. The last layer is called the output layer, which produces the output of the network. In between the input and output layers, there can be one or more hidden layers, which perform computations on the input data before passing it on to the next layer.

Each neuron in the network receives input from other neurons and performs a computation on that input. The output of the neuron is then passed on to other neurons in the network. The computation performed by a neuron can be a simple mathematical function like a weighted sum, followed by an activation function like the sigmoid function or ReLU function.

The weights on the connections between neurons in the network are adjusted during training, in order to minimize a loss function that measures the difference between the predicted output of the network and the true output. This is done using an optimization algorithm like stochastic gradient descent.

There has recently been an increased interest in utilizing ANN models to forecast building energy consumption due to the ability of ANNs to solve complex non-linear tasks and process big datasets. (Khalil, McGough, Pourmirza, Pazhoohesh, & Walker, 2022)

Detailed description of each of these models can be found easily, therefore they are not dealt in detail here.

Chapter 3: METHODOLOGY AND APPROACH TO DATA ANALYSIS

3.1 Approach and Methodology

This thesis aims to understand the energy usage pattern of the dining commons, analysis, and prediction of future consumption by using historical data and applying machine learning methods and suggests appropriate energy saving strategies based on wholesome analysis of data from the regression results, in-person site visits, and interviews with the building manager.

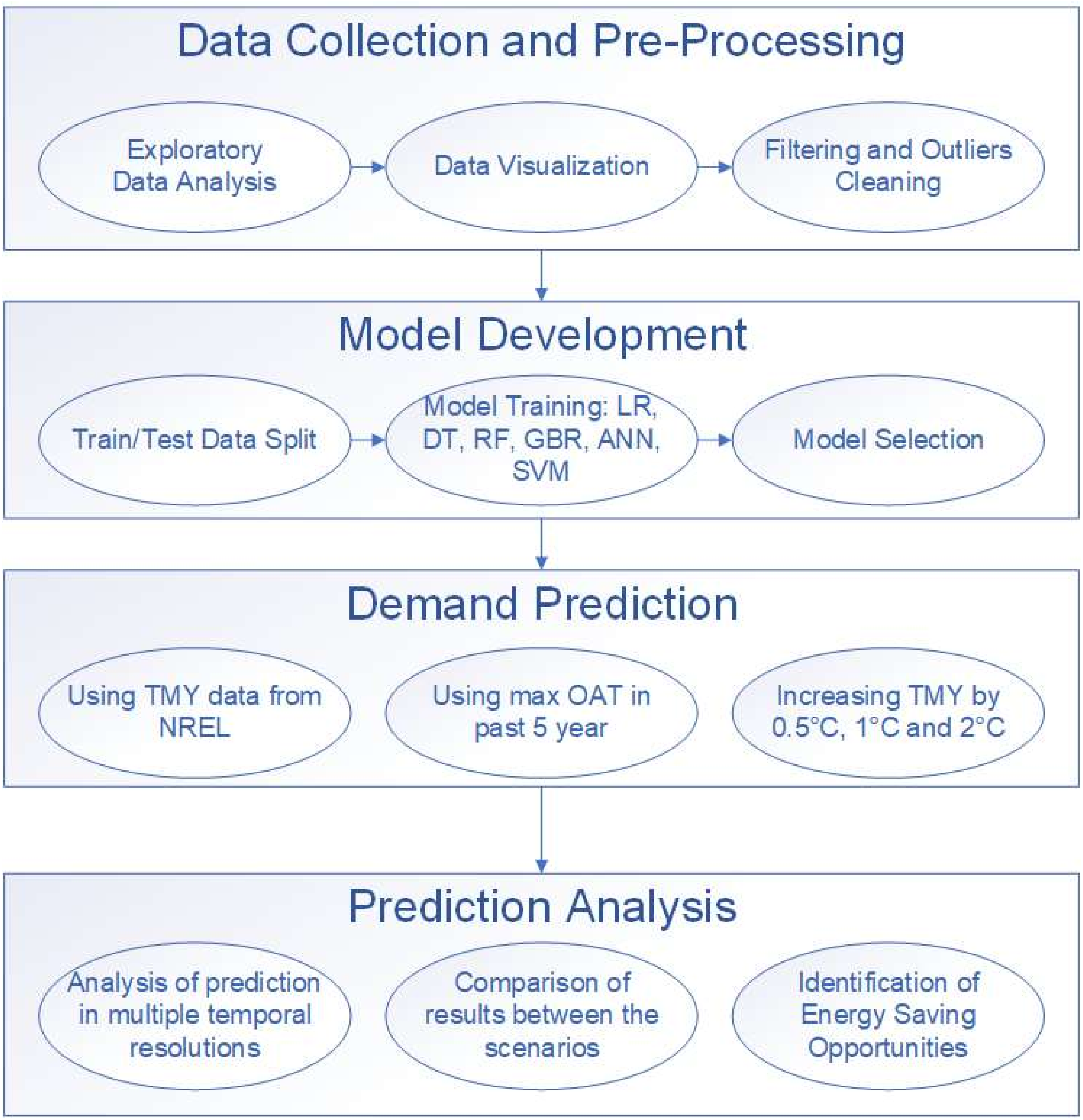


Figure 4 Methodology of the study

For the energy use prediction, a thorough literature review of different techniques used in machine learning for buildings were performed. The energy data of these buildings were collected from the UC Davis PI database which serves as a rich repository for all energy data of all campus buildings. This data was then pre-processed as shown in the figure above. First the data was explored to understand its nature, different graphs were looked at and appropriate date ranges were filtered, and outliers were cleaned. Then different machine learning models were implemented, and multiple metrics were compared to analyze the differences in their performance; then a model was selected based on their performance to predict future demands using NREL published meteorological data. Prediction results for four other scenarios were also looked at: Increasing TMY by 0.5°C, 1°C, 2°C and maximum OAT in the past 5 years. Comparison of model results for each of these scenarios were performed.

For recommendations of energy saving opportunities, the results of the regression were analyzed on multiple temporal resolutions, multiple graphs were made for each of the scenarios including the base case and suggestions were made based on the energy usage pattern and knowledge gained from site-visits.

3.2 Reasons for selecting Segundo Dining Commons

There are four main dining facilities in campus serving the students: Segundo Dining Commons, Tercero Dining Commons, Cuarto Dining Commons and Latitude. Out of these four main restaurants, Segundo DC is the biggest one in terms of Energy Usage Intensity (EUI) and the number of servings. Three of these dining commons: Segundo DC, Tercero DC and Latitude are in the campus steam and chilled water network whereas Cuarto DC uses PGE utilities. The energy consumption of these facilities and utility bills for the year 2022 was looked at and the following table was generated. It summarizes the comparison between these dining commons.

Facilities	**Energy Use Intensity (EUI, kBtu/sq. ft.)**	**Floor Area (sq. ft.)**	**Utility Bills (per year for electricity, steam and chilled water)**
Segundo Dining Commons	347	46,712	$ 365,000
Tercero Dining Commons	208	59,253	$ 307,000
Cuarto Dining Commons	280	18,776	$ 192,000
Latitude	153	33,566	$141,000

Table 1 Comparison of Segundo DC with other DCs in campus

As shown in the table, Segundo DC has the largest Energy Usage Intensity and pays the highest in utility bills. The facility is also one of the largest in size and serves the catering needs for events in campus. Therefore, Segundo DC was chosen for the purposes of energy data analysis for this project.

3.3 Segundo Dining Commons Overview

Segundo DC was built in 2005 and situated in the heart of campus near the student housing facilities. It has a culinary support center which is a production kitchen for some specific meals served in other dining commons and support culinary needs in various campus events. It prepares around 18,000 meals per day and is open throughout the week for students. (Mabberley, 2022).

Figure 5 Segundo Dining Commons Entrance

The facility is one storied facility that is divided into food production/ kitchen area and dining area. It has three Air Handling Units (AHUs) and 12 Make-up Air Units (MAUs) that condition the air in the space. The dining areas are served by two AHUs and multiple Make-up Air Units that provide makeup air for the exhausted air from the cooking island areas. The kitchen/ food production area is served by multiple MAUs and one AHU. The mechanical drawings of the facility were studied, and the following simplified figure was made to represent the mechanical systems of the building.

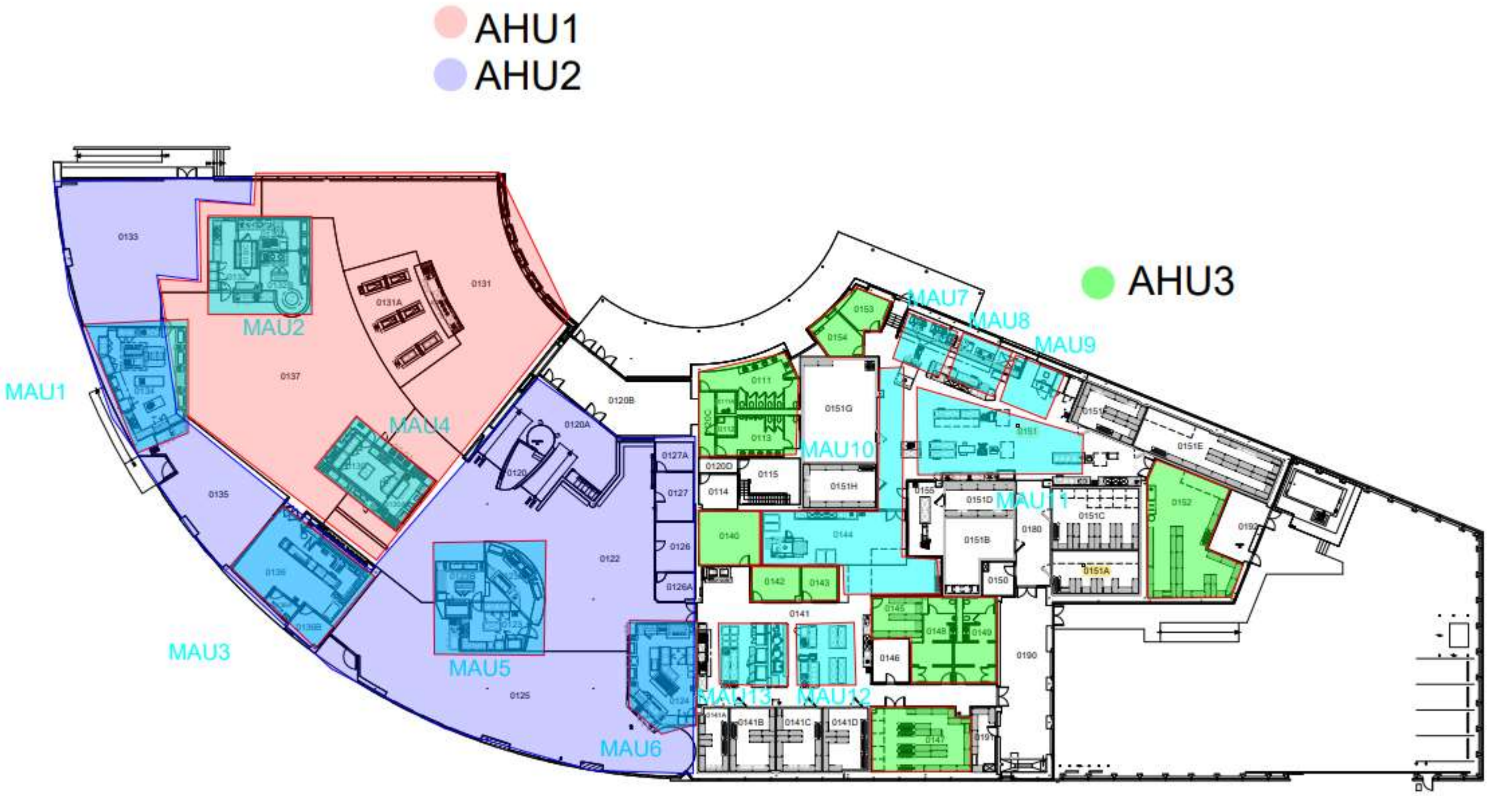


Figure 6 Mechanical System Schematics of Segundo DC

The facility is connected to the UC Davis campus steam and chilled water loop. The central heating and cooling plant supplies steam and chilled water to the building that is used for air conditioning of the space. It was also found that the facility uses steam for some meal preparation as well. The facility's electricity demand, condensate demand and chilled water demand for the year 2022 was analyzed and represented in the following pie chart. It should be noted that condensate is measured as a proxy for steam consumption and is used for billing purposes too.

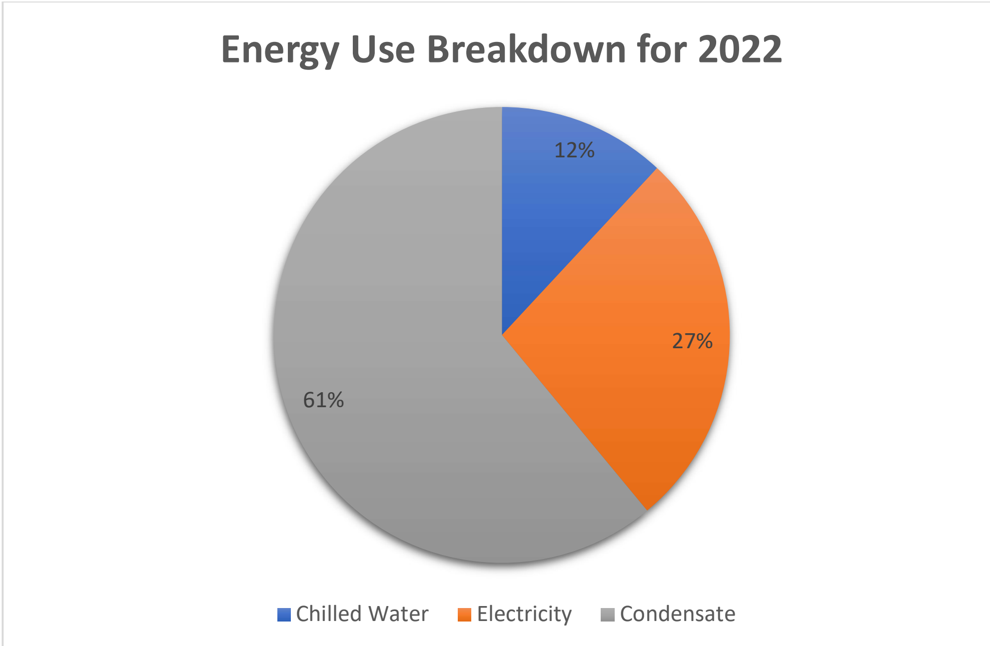


Figure 7 Segundo DC Energy Use Breakdown for 2022

The high steam usage as depicted by the condensate demand can be attributed to the space heating needs for the large square footage of the dining commons and its use of steam in food preparation. The electricity demand is influenced by refrigerators in the facility that helps keep food safe and the demand from HVAC equipment and lighting system.

3.4 Building Systems Control and Data Sources

Segundo DC is equipped with Building Management System (BMS), specifically Siemens Insight and Desigo to monitor and control the building's HVAC system. The BMS system feeds countless data points from the building control system including but not limited to supply air temperature, supply static pressure, return air temperature, outside air temperature, etc. into the building equipment control panels. It also allows facilities staff to easily monitor the system and change any parameters as needed.

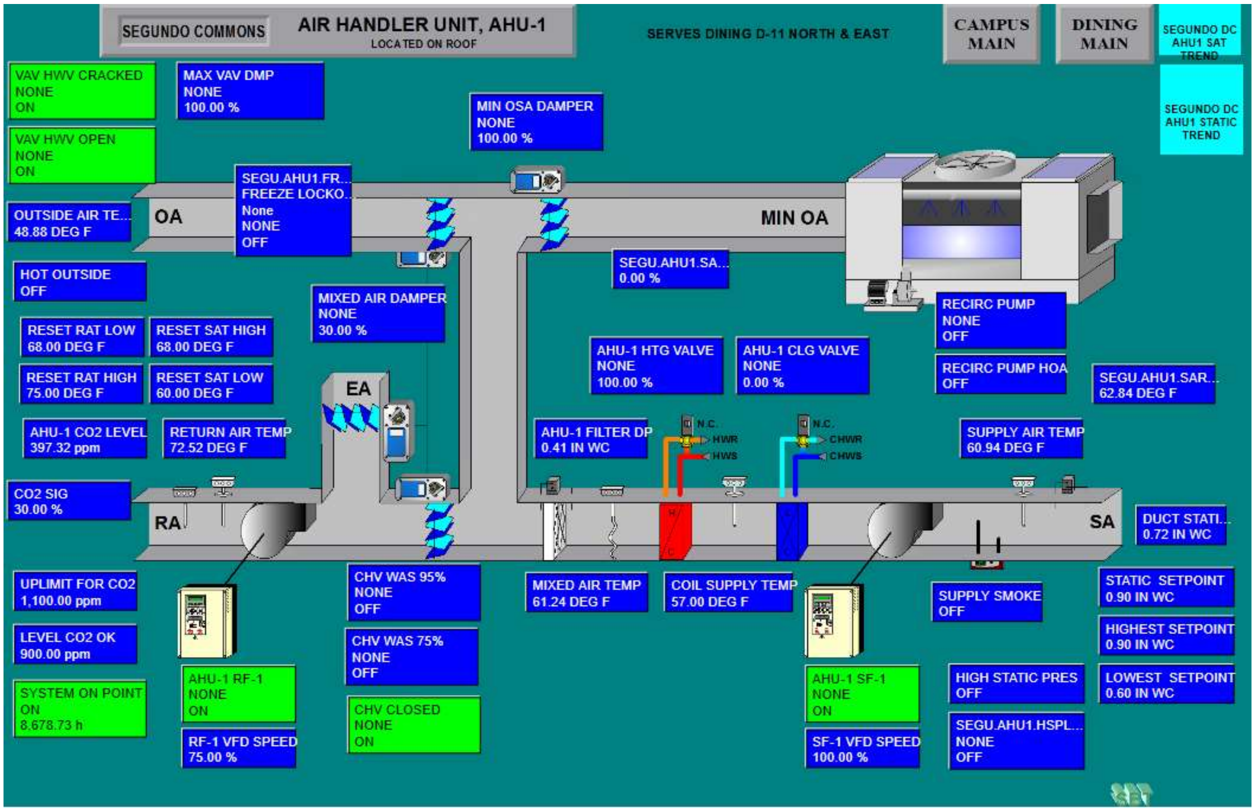


Figure 8 SIEMENS Desigo Snapshot

The numerous sensor data from BMS is stored in it only for a few months and important data points like data from building meters are transferred to the PI system which is the main repository for the campus energy data. This repository contains numerous years of data from the building control system of all buildings on campus. The data stored in this PI repository is pulled for analysis for the purposes of this thesis.

Since this research utilizes black box model of buildings' energy data modelling, it uses just the meteorological data and commodities demands data from the PI system. The facility is treated as a black box that responds to the outside meteorological inputs. The data imported and used for the modelling purpose in this thesis are as follows. All of them are time series data with the interval of one hour.

i. Condensate Demand
ii. Chilled Water Demand
iii. Electricity Demand
iv. Outside Air Temperature (OAT)

The OAT data is further divided into heating degree hours and cooling degree hours to feed into the model. The inflection point temperature of 65F is used for defining the heating and cooling degree hours. Two continuous variables, Heating Degree Hours (HDH) and Cooling Degree Hours (CDH) are created which are a modulus of the difference between 65F and OAT for that hour. Three categorical variables for months, days (weekdays and weekends) and hour are also created from the timeseries data to be fed into the model. A total of 6 variable is fed into the model, they are as follows.

i. Outside Air Temperature (OAT)
ii. Heating Degree Hours (HDH)
iii. Cooling Degree Hours (CDH)
iv. Month (Categorical variable; 1 through 12 for each month)
v. Day (Categorical variable: 1 for weekday, 0 for weekend)
vi. Hour (Categorical variable: 0 through 23 for each hour in a day)

These inputs are fed into the model for training along with the demand data for condensate, chilled water and electricity. Then the trained models are fed with the same input as above with TMY data for the future.

The meteorological data fed into the model is called Typical Meteorological Year (TMY) data made available by National Renewable Energy Laboratory (NREL) through National Solar Radiation Database (NSRDB). The data is called “typical” because the entirety of original meteorological data is condensed into one year’s worth of the most usual conditions. It is calculated using multiyear data for a particular location and 12 months of that time frame that best represents the median conditions is used. This data is intended to be used in computer simulations for building systems and widely used in the industry for these purposes. (Typical Meteorological Year (TMY) NREL, 2023)

Chapter 4: MODEL TRAINING, SELECTION AND DEMAND PREDICTION

4.1 Overview of the data

To understand the data before feeding it into the models, exploratory data analysis of the past demand data for condensate, chilled water and electricity was performed. These demand data for the period 2015 January 1st through 2022 December 31st was analyzed to determine the best period to feed into the model. The following results were obtained.

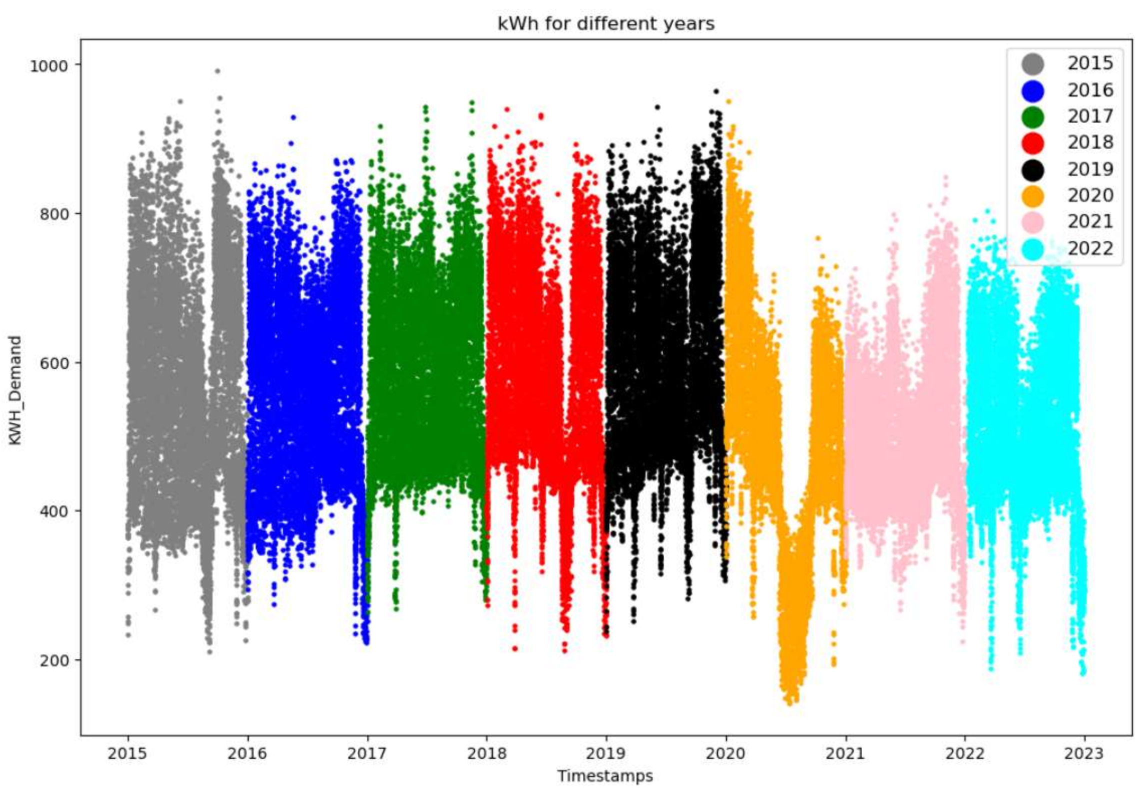


Figure 9 Past Electricity Demand for Segundo DC

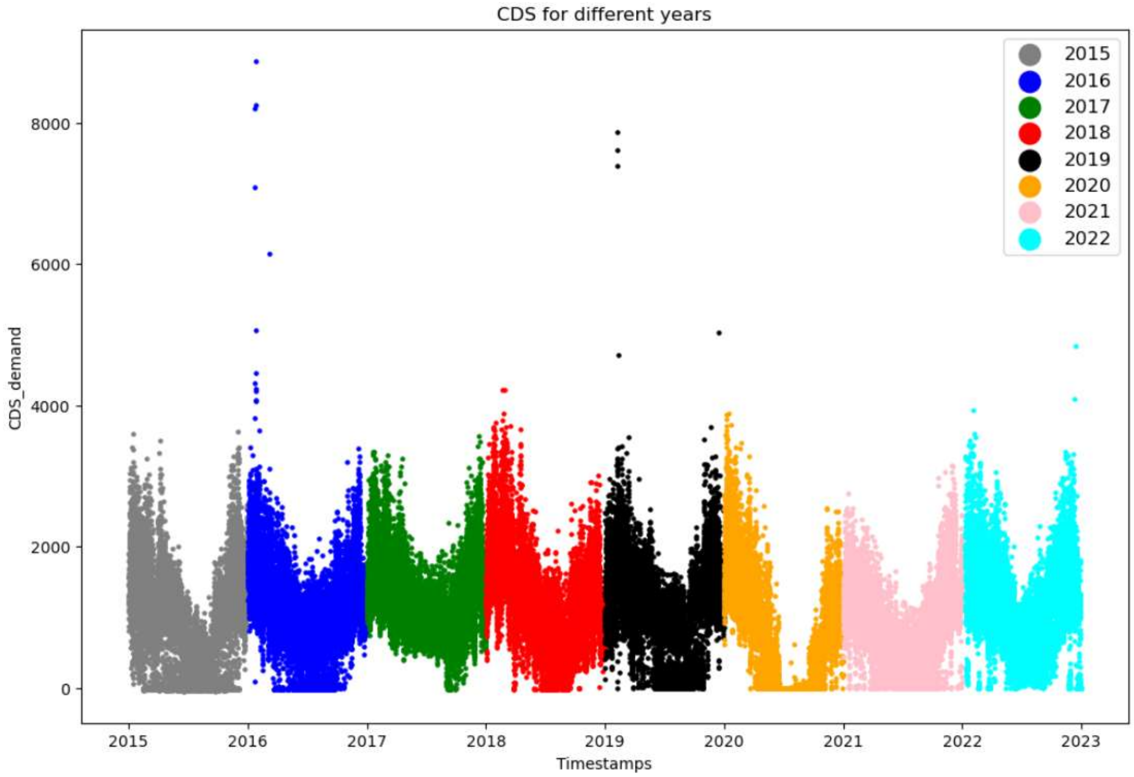


Figure 10 Past Condensate Demand for Segundo DC

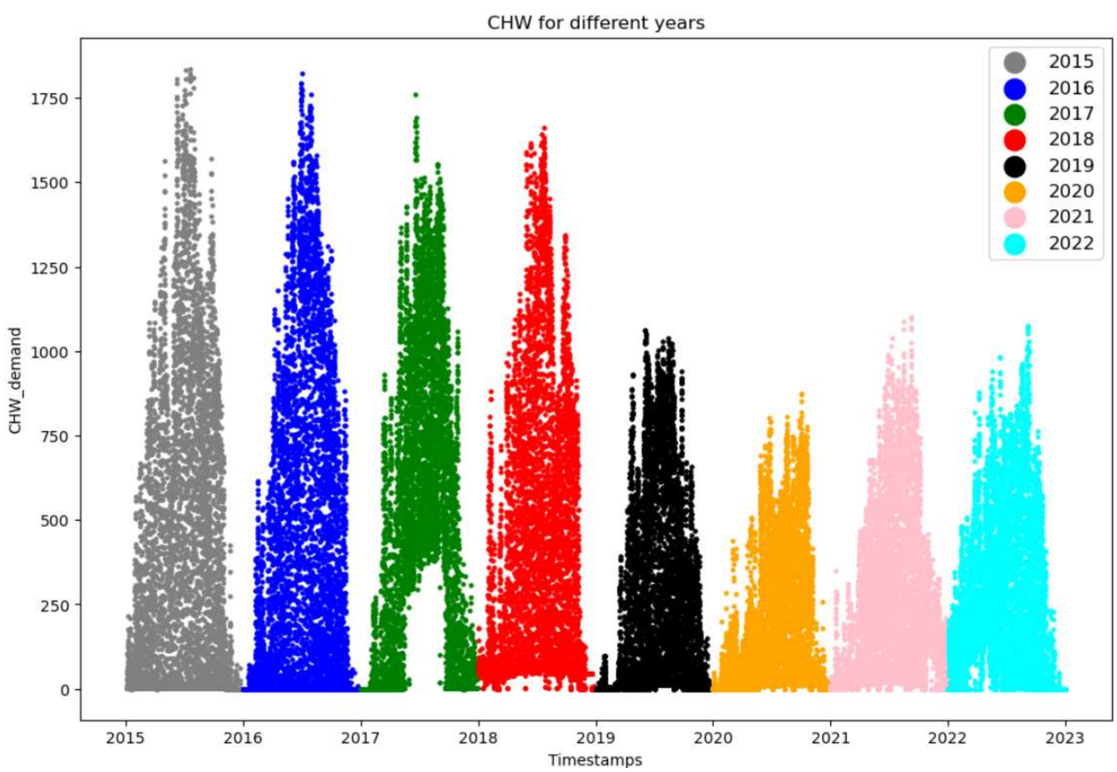


Figure 11 Past Chilled Water Demand for Segundo DC

As observed from the above data, the demands for Electricity and Chilled Water are significantly different for the years 2021 and 2022 compared to the previous years. The significant dip in demands for the year 2020 can be attributed to the COVID campus lockdown when the facility was closed. The demands for years 2021 and 2022 are similar and represent the time periods of full operation, therefore the time-period between 2021 January 1st through 2022 December 31st was used for model training.

The OAT data was also analyzed for this time-period and represented as shown in the following graph.

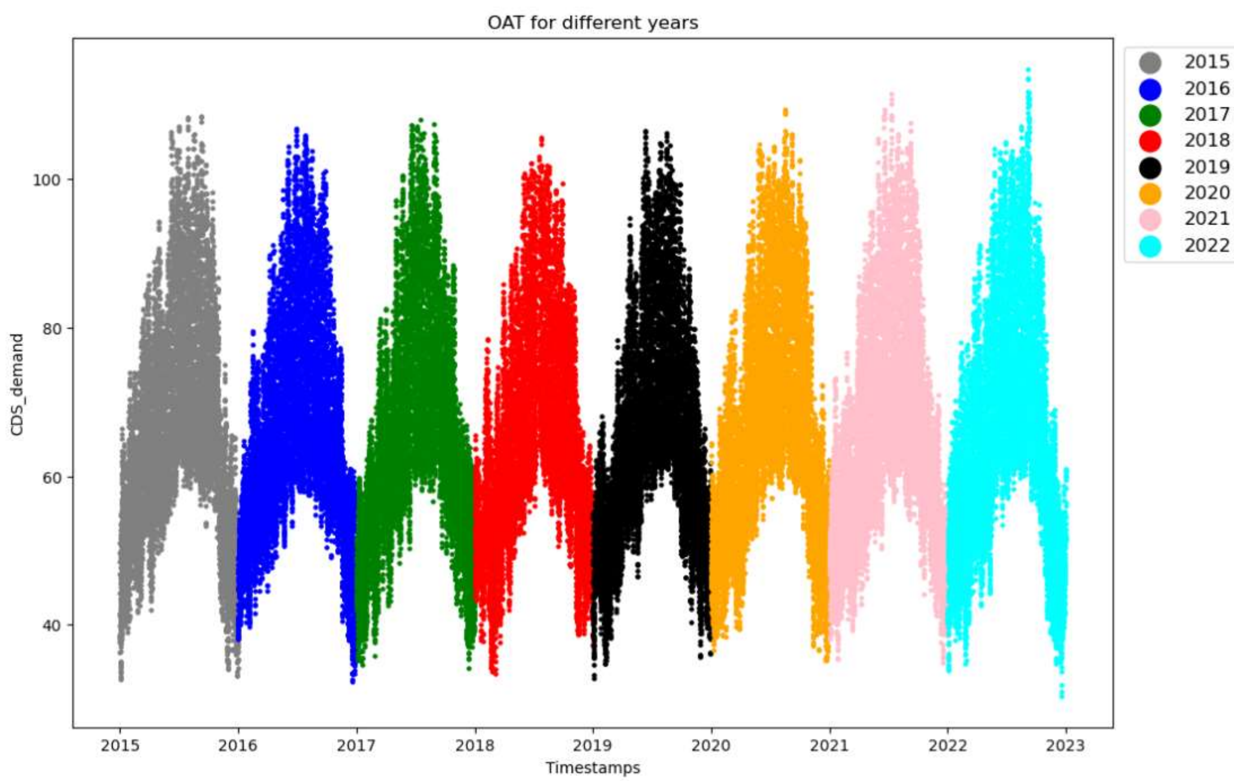


Figure 12 Outdoor Air Temperatures

The OAT data for the entire period is very similar every year, so the time period for OAT that is good for other demand data was taken to use for the data analysis.

4.2 Exploratory Data Analysis and Data Cleanup

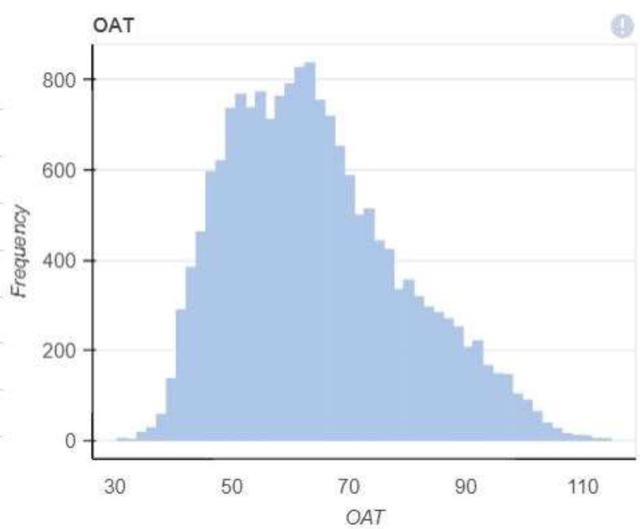
After selection of the year for the data analysis, the timestamps were divided into categorical variables that represented the month (1 to 12), day (representing 1 for weekday and 0 for weekend) and time of the day (0 to 23). All the columns in the data were then explored to understand the data in more detail. Using the python “dataprep” package, following summary details for each column of the data was obtained.

Dataset Statistics

Number of Variables	43
Number of Rows	17494
Missing Cells	557
Missing Cells (%)	0.1%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	1.7 MB
Average Row Size in Memory	100.0 B
Variable Types	Numerical: 4 DateTime: 1 Categorical: 38

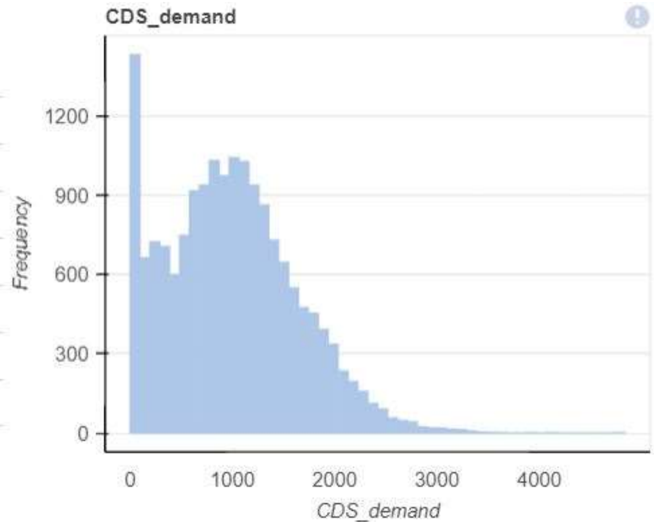
Approximate Distinct Count	17466
Approximate Unique (%)	100.0%
Missing	28
Missing (%)	0.2%
Infinite	0
Infinite (%)	0.0%
Memory Size	272.9 KB

Mean	64.4618
Minimum	30.4137
Maximum	114.854
Zeros	0
Zeros (%)	0.0%
Negatives	0
Negatives (%)	0.0%



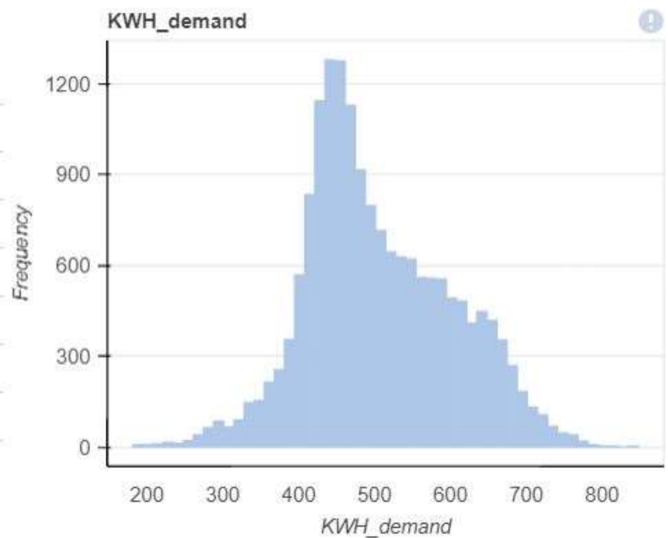
Approximate Distinct Count	16726
Approximate Unique (%)	97.1%
Missing	264
Missing (%)	1.5%
Infinite	0
Infinite (%)	0.0%
Memory Size	269.2 KB

Mean	988.8055
Minimum	0
Maximum	4842.0316
Zeros	476
Zeros (%)	2.7%
Negatives	0
Negatives (%)	0.0%



Approximate Distinct Count	17224
Approximate Unique (%)	100.0%
Missing	264
Missing (%)	1.5%
Infinite	0
Infinite (%)	0.0%
Memory Size	269.2 KB

Mean	505.0612
Minimum	181.2982
Maximum	848.326
Zeros	0
Zeros (%)	0.0%
Negatives	0
Negatives (%)	0.0%



Approximate Distinct Count	12642
Approximate Unique (%)	72.3%
Missing	1
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Memory Size	273.3 KB

Mean	205.4797
Minimum	0
Maximum	1100.0313
Zeros	2651
Zeros (%)	15.2%
Negatives	0
Negatives (%)	0.0%

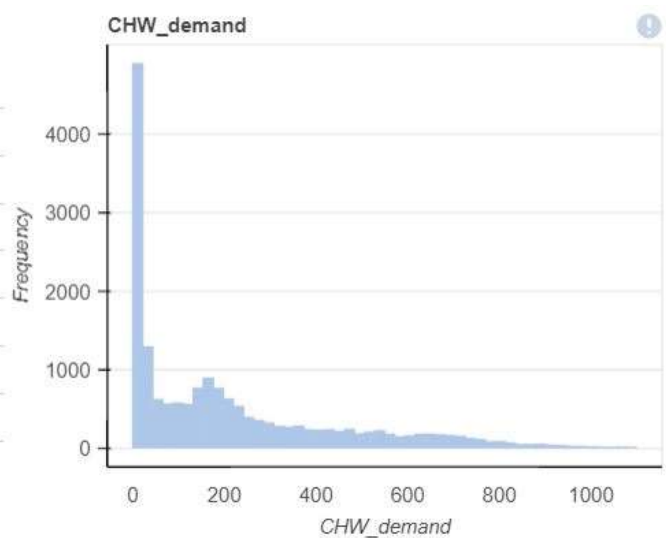


Figure 13 Exploratory Data Analysis for all variables

As shown in the summary above, some of the columns were missing data, but the number of rows with missing data was negligible compared to the total number of rows in the data, maximum missing percentage being 1.5%, therefore, the rows with missing data were deleted to prevent inconsistency in model learning.

All the demand data were also checked for outliers. Two data points were found as outliers in the condensate demand data as shown in the figure below. The condensate demand value above 4000 kBtu was considered as outlier as it was significantly higher than condensate demand at the same OAT.

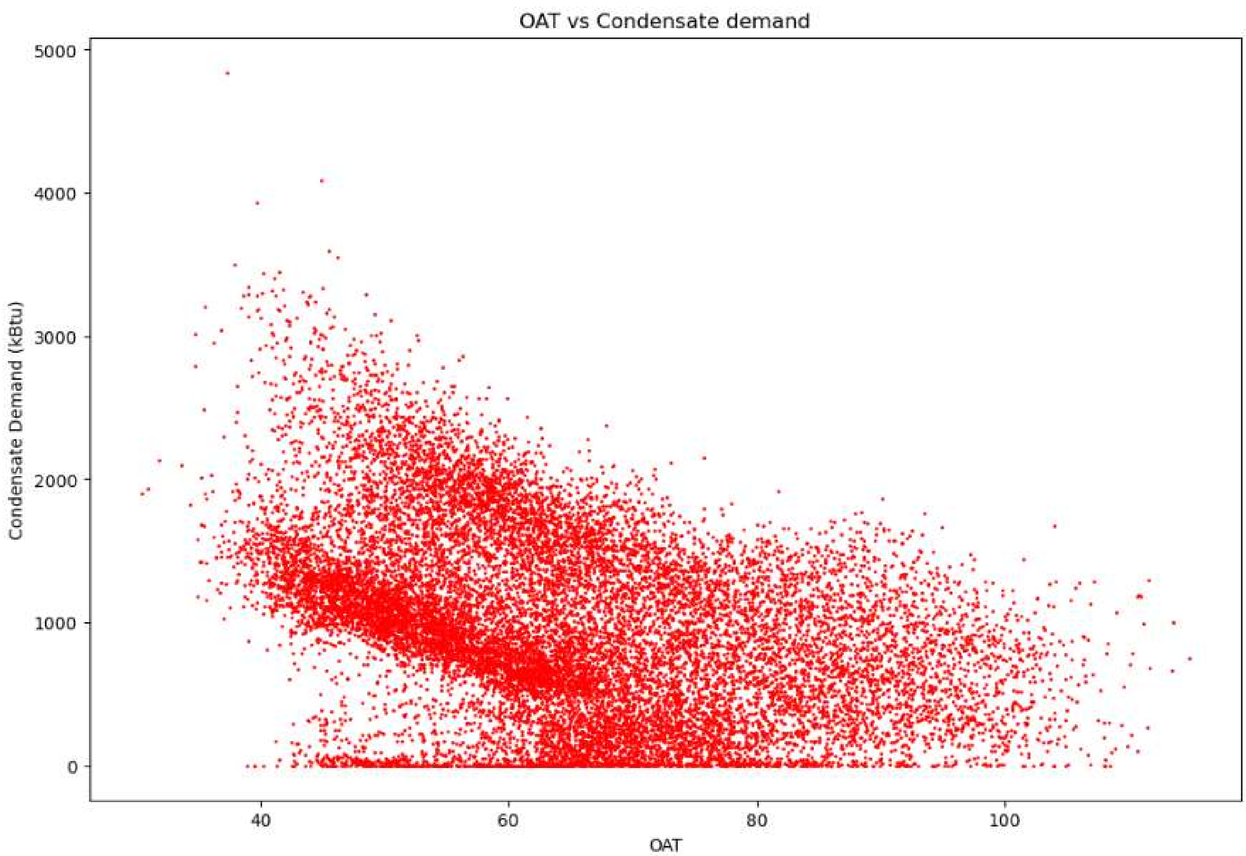


Figure 14 OAT vs Condensate Demand

After cleaning the outliers, Heating degree hours (HDH) and Cooling degree hours (CDH) variables were created with the inflection point of 65F distinguishing the heating and cooling degree hours. These variables were compared against Condensate demand and Chilled water demand to better visualize their relationship.

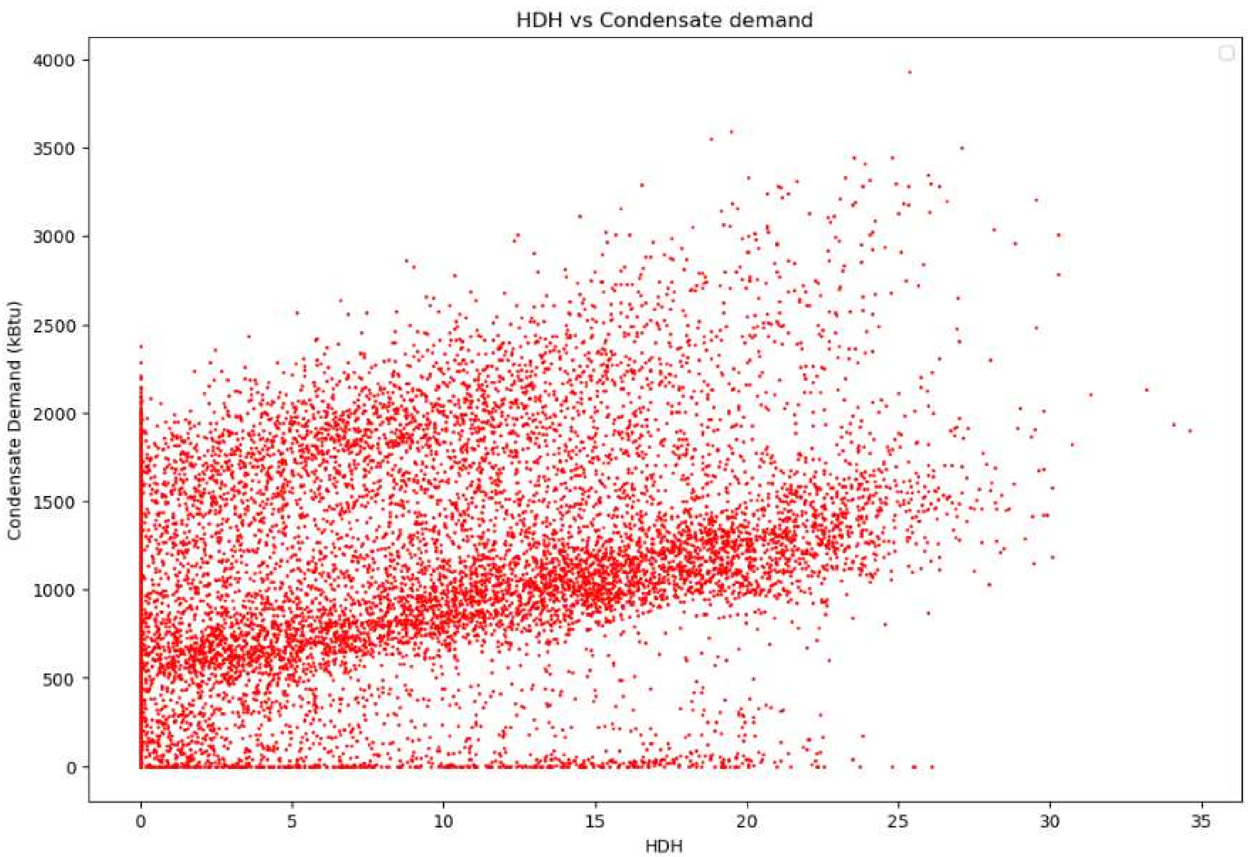


Figure 15 HDH vs Condensate Demand

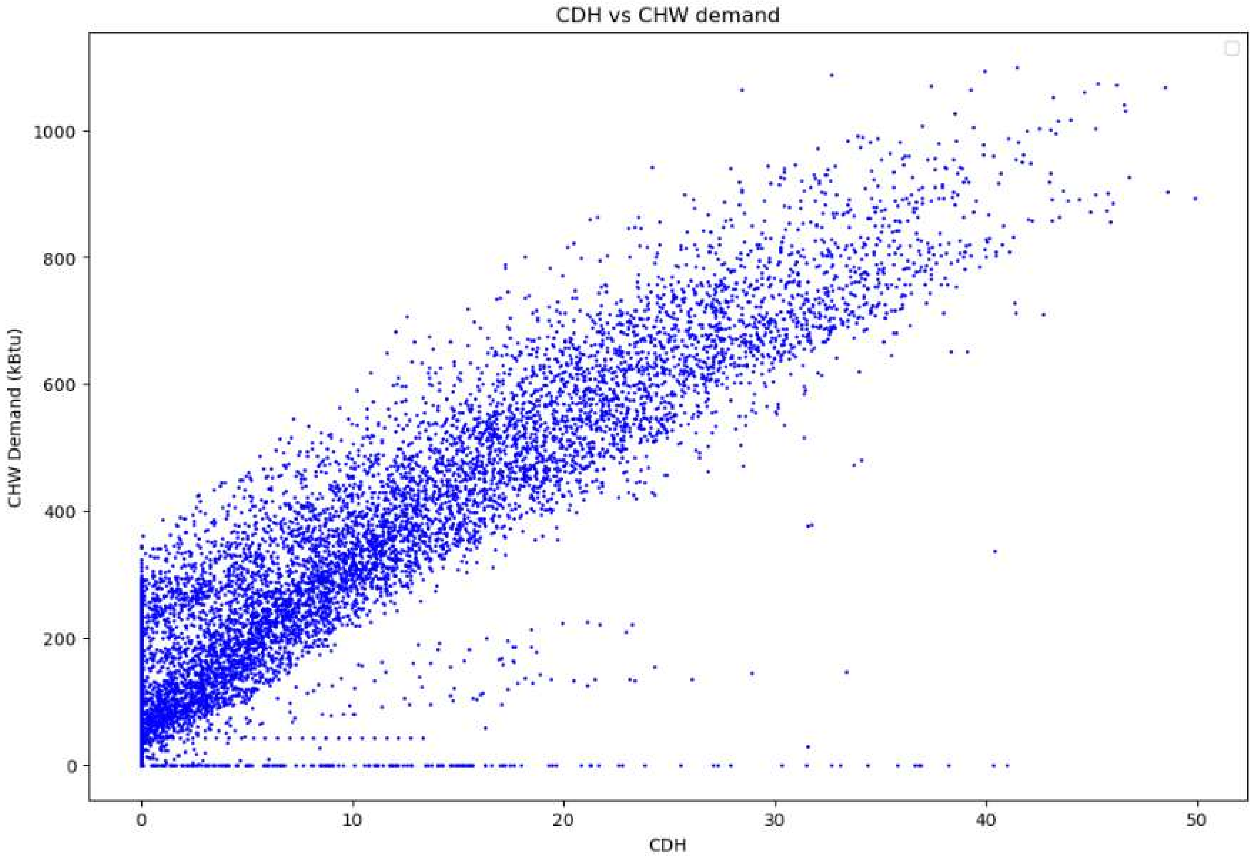


Figure 16 CDH vs Chilled Water Demand

The degree hours against the condensate or chilled water demand graphs are as expected showing a nearly linear relationship against each other. They were not plotted against electricity demand because electricity demand is also influenced by lighting and plug loads that do not have any relationship with the temperature. After completing the data exploration and cleanup part, the data was ready to be fed into the model.

4.3 Model Comparison and Selection

The cleaned data was divided into train and test split of 80% and 20% respectively including demand data for condensate, chilled water, and electricity. The same set of data was fed into all the models considered and their prediction on the test data was compared using some metrics. Following models were fed with the data prepared to predict all three commodities demand.

i. Linear Regression
ii. Decision Trees
iii. Random Forest
iv. Gradient Boosting Regressor
v. Artificial Neural Network (ANN)
vi. Support Vector Mechanism (SVM)

The comparison metrics for all three commodities using each of the models mentioned above are as follows.

For Condensate prediction models

Models	MSE	RMSE	MAE	R2	Adjusted R2
Linear Regression	179028	423.117	328.867	0.570726	0.565674
Decision Tree	303051	550.5	396.494	0.273344	0.264792
Random Forest	183876	428.807	318.088	0.559102	0.553913
Gradient Boosting Regressor	177024	420.742	324.494	0.575531	0.570536
ANN	152830	390.934	297.844	0.633545	0.629232
SVM	361281	601.067	479.48	0.133718	0.123524

For KWH prediction models

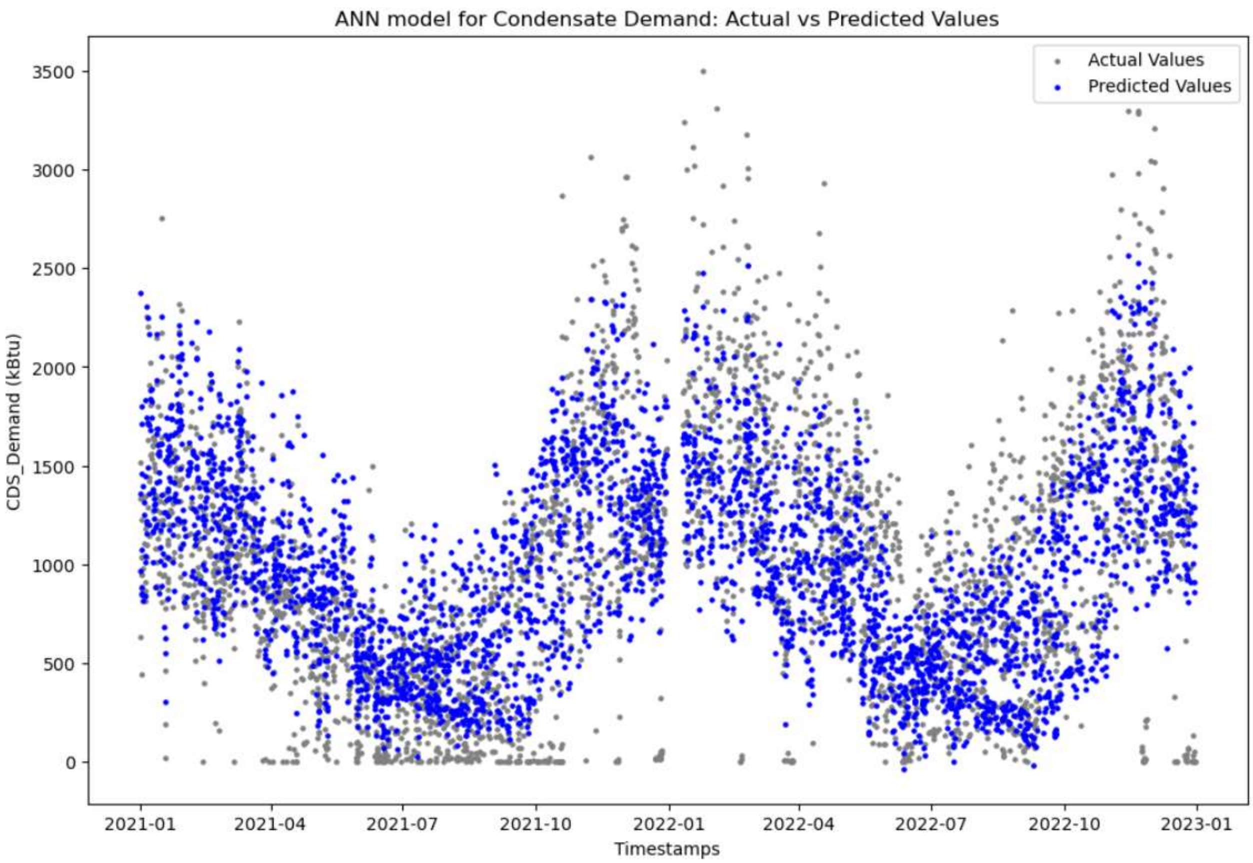
Models	MSE	RMSE	MAE	R2	Adjusted R2
Linear Regression	4967.62	70.4814	53.4517	0.494246	0.488294
Decision Tree	7679.04	87.6301	61.2944	0.218196	0.208995
Random Forest	4998.36	70.6991	49.4772	0.491116	0.485128
Gradient Boosting Regressor	4654.03	68.2205	50.9662	0.526173	0.520596
ANN	4106.08	64.0787	45.0474	0.581959	0.57704
SVM	6235.12	78.9628	59.0558	0.365202	0.357731

For CHW prediction models

Models	MSE	RMSE	MAE	R2	Adjusted R2
Linear Regression	7328.46	85.6064	62.9058	0.851823	0.850079
Decision Tree	13063.4	114.295	75.2869	0.735864	0.732756
Random Forest	7677.64	87.6222	61.1844	0.844762	0.842935
Gradient Boosting Regressor	6827.2	82.6269	61.5063	0.861958	0.860333
ANN	6547.22	80.9149	58.5003	0.867619	0.866061
SVM	20617.5	143.588	111.403	0.583126	0.578221

Table 2 Model Metrics Comparison

From the above table, it can be observed that ANN has the highest adjusted R2 value for all three commodities demand model (CHW, KWH and Condensate models), however, each of them have marginal differences with Gradient Boosting Regressor method and highest difference with SVM. The difference simply shows how effectively these methods work with our type of data. The method working best for this type of data can be worst for a different type of data (say building's internal temperature prediction model based on solar irradiance data) and vice-versa. The methods with marginal differences from ANN may not produce very different results from its prediction, but for this purpose ANN was chosen for all three commodities based on the highest score obtained for adjusted R2. The model's prediction and their residual values are as follows.



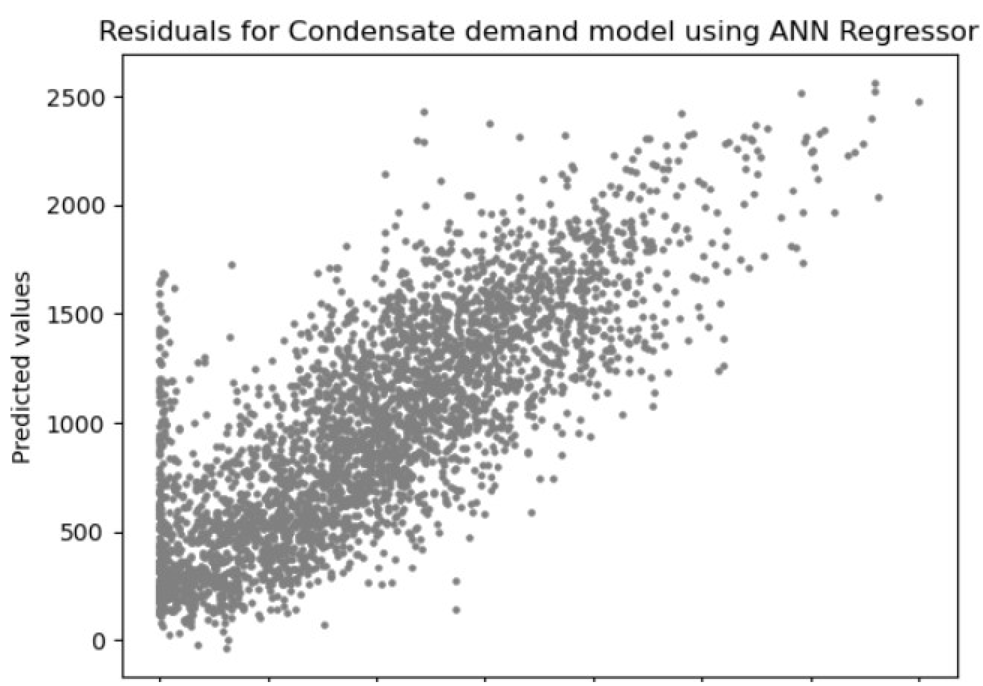


Figure 17 Prediction Visualization and Residuals for Condensate Demand using ANN

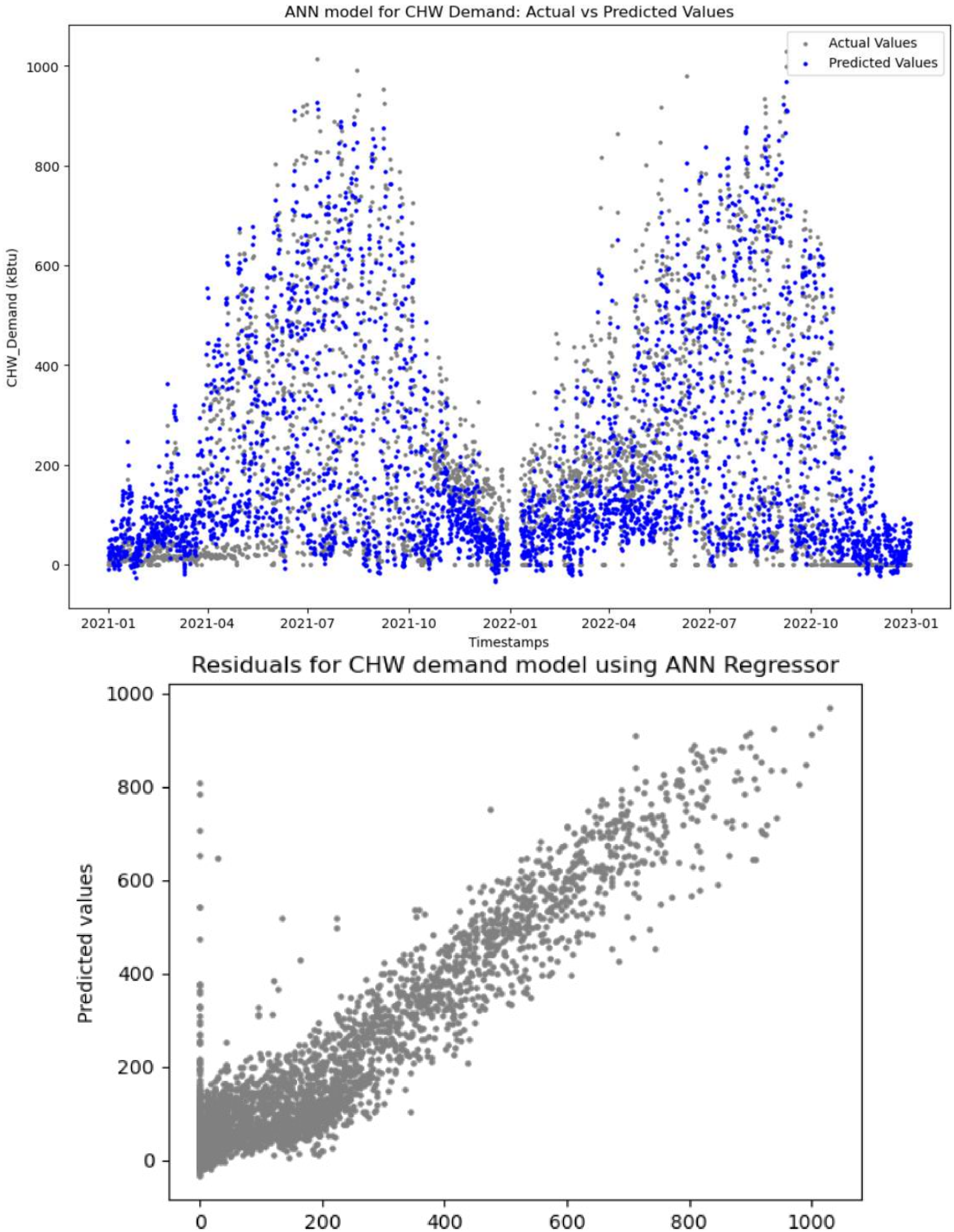


Figure 18 Prediction Visualization and Residuals for Chilled Water Demand using ANN

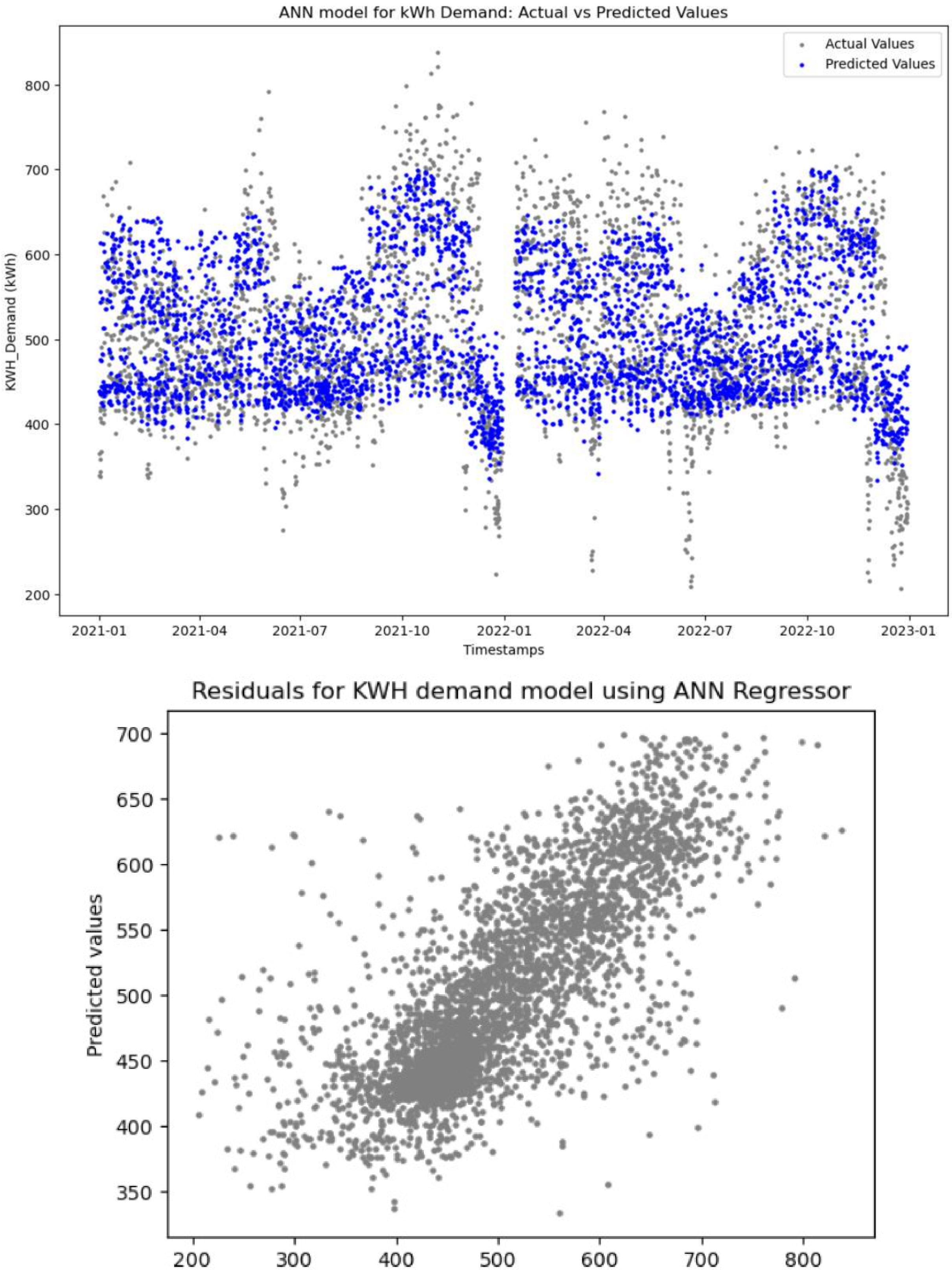


Figure 19 Prediction Visualization and Residuals for Electricity Demand using ANN

As observed from the visualization above, the prediction sometime overshoots and undershoots in prediction of all the commodities, however, this ANN model performs best compared to all the other models.

4.4 Preparing Feed Data and Final Model Prediction

The TMY data obtained from NREL contains OAT and timestamps for a typical year. This data was used for the prediction of demands for the year 2023 by feeding the trained ANN model with same variables as the variables used in training the model for consistency in prediction. They are as follows.

i. Outside Air Temperature (OAT)
ii. Heating Degree Hours (HDH)
iii. Cooling Degree Hours (CDH)
iv. Month (Categorical variable; 1 through 12 for each month)
v. Day (Categorical variable: 1 for weekday, 0 for weekend)
vi. Hour (Categorical variable: 0 through 23 for each hour in a day)

HDH and CDH were calculated as done for the training data with the inflection temperature of 65F. The OAT values in the feed data compared to the training data of 2021 and 2022 are shown as follows.

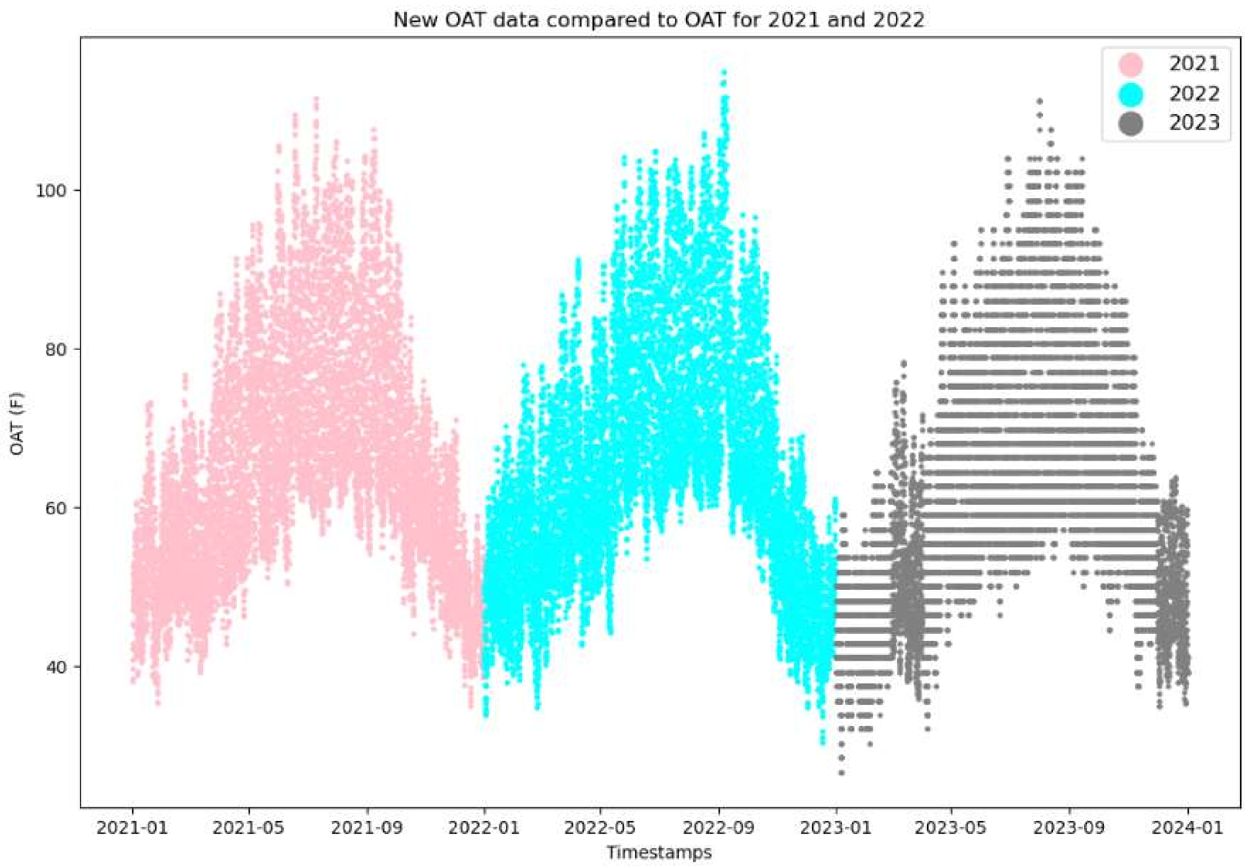
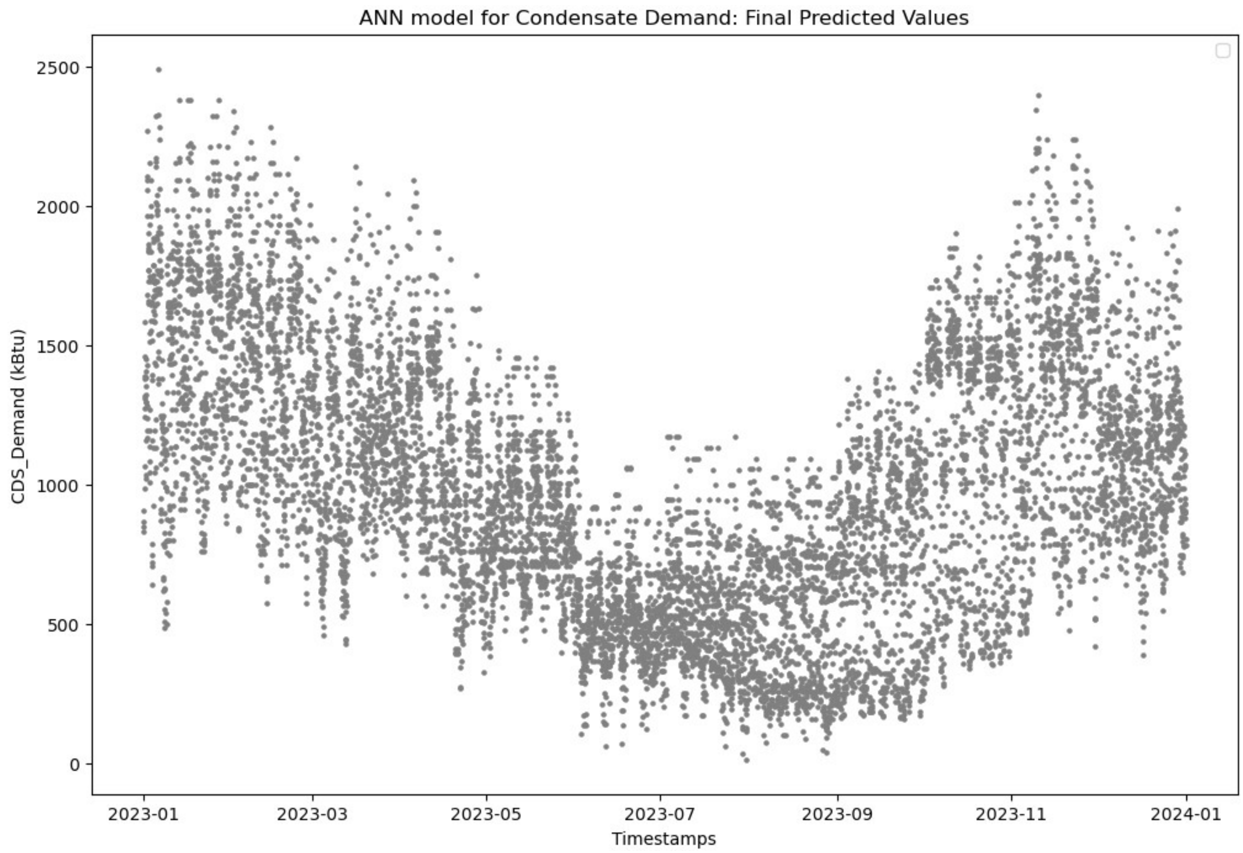
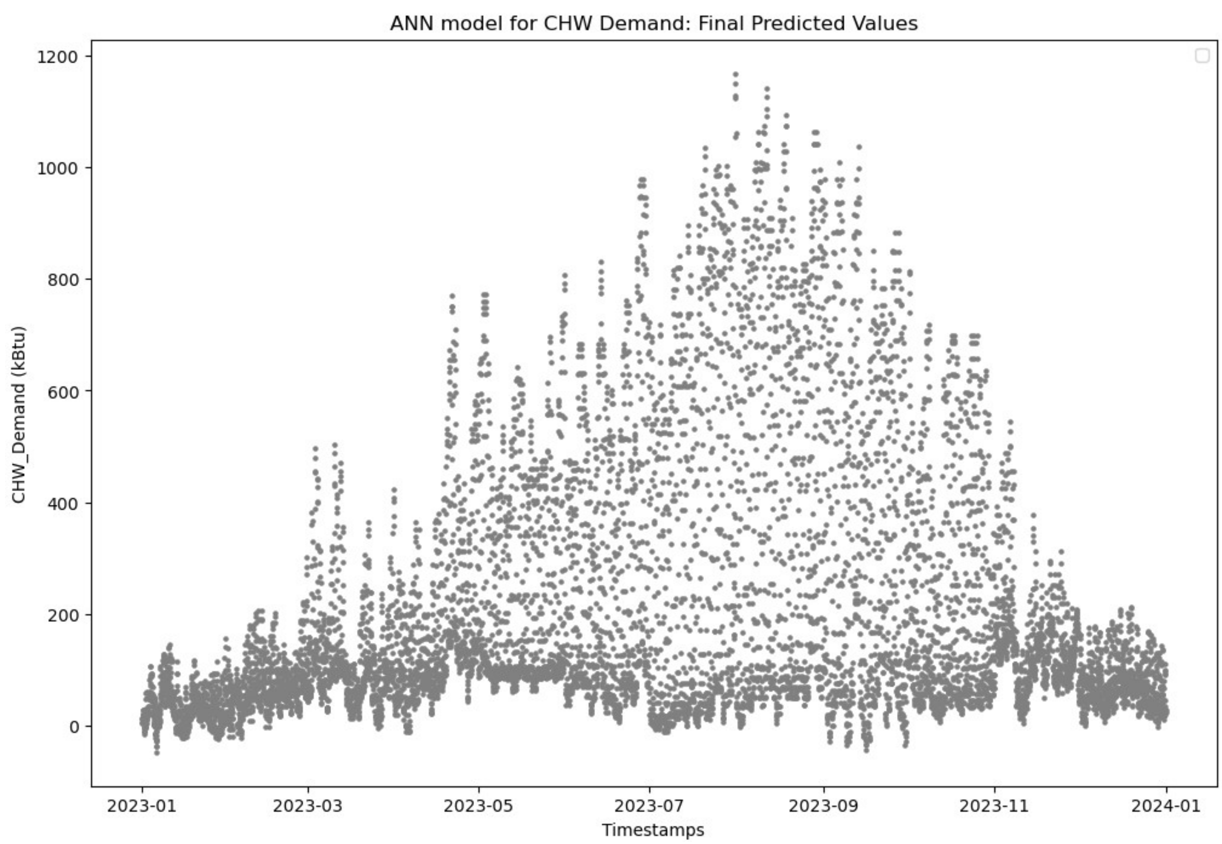


Figure 20 Feed OAT data to the trained ANN model

Since the model is trained with the same variables, it can be fed with new values that are similar in type but different values to get prediction for the future demands. The following demand prediction was obtained when above data was fed into the model.


ANN model for Condensate Demand: Final Predicted Values
CDS_Demand (kBtu)
Timestamps
0
500
1000
1500
2000
2500
2023-01
2023-03
2023-05
2023-07
2023-09
2023-11
2024-01


ANN model for CHW Demand: Final Predicted Values
CHW_Demand (kBtu)
Timestamps
0
200
400
600
800
1000
1200
2023-01
2023-03
2023-05
2023-07
2023-09
2023-11
2024-01

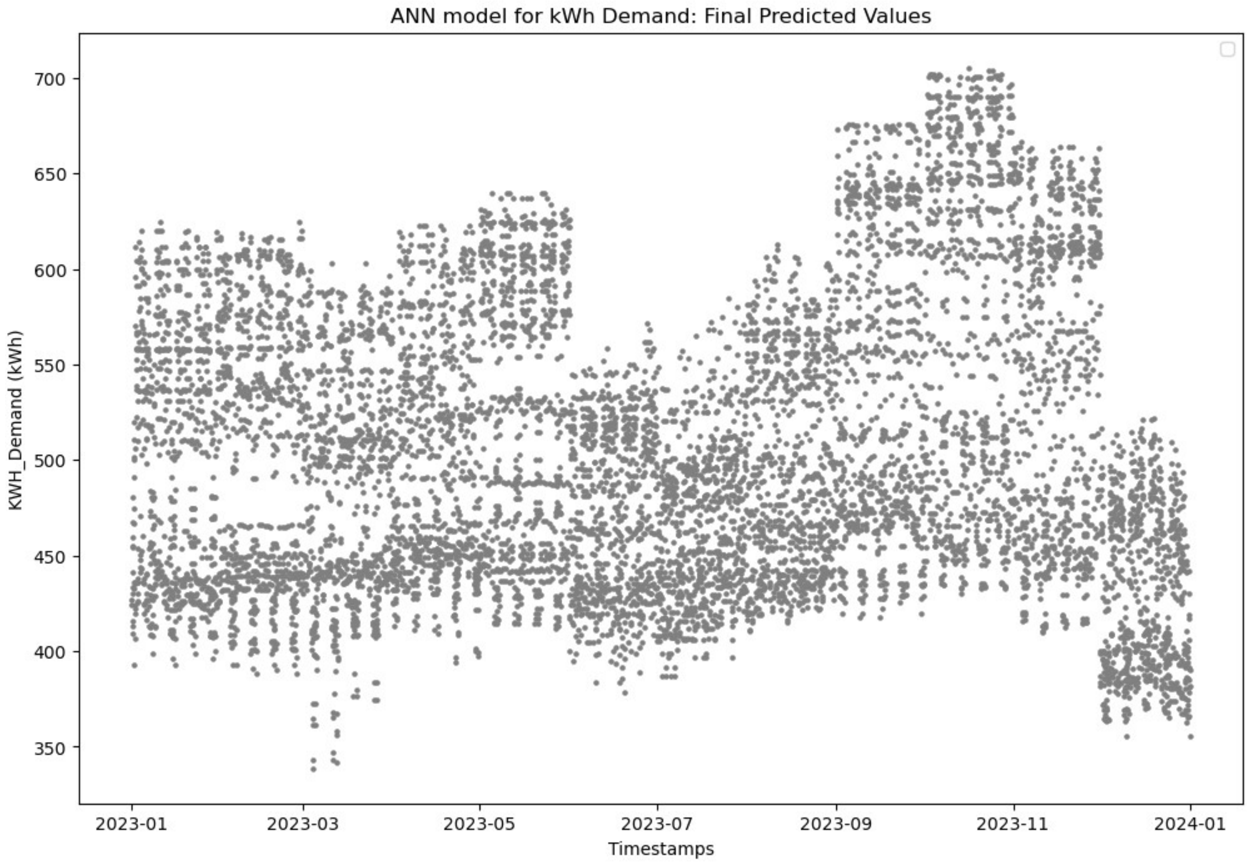


Figure 21 Condensate, Chilled Water and Electricity Demand Prediction with the trained ANN model

4.5 Analyzing Demands under different Meteorological Scenarios

The trained model was fed with the four different scenarios to analyze how the demand changes with the change in meteorological conditions.

For the first scenario, a typical year's meteorological data was created from the max outside air temperature for each hour of each day in the past five years of recorded temperatures in the UCD campus. This data was fed into the trained model as described above and the results were analyzed.

The remaining three scenarios were increasing the outside air temperature in the TMY data by 0.5°C, 1°C and 2°C to see how the commodities demand change with the change in climate due to global warming as projected by IPCC. (IPCC, 2018) The three temperatures were selected to capture all the high confidence and medium confidence levels of the temperature rise due to climate change. Though this is a crude approach to observe changes in demand prediction, it helps us see the effects of varied changes in temperatures and if it creates a significant change in future demands given the scenarios were to come true.

All four of these scenarios had a one-hour interval data for the entire year. The trained models as described previously were used for analysis and the following variables were created from the timestamps to feed into the model. These model inputs are same as the inputs used for its training and listed as follows.

i. Outside Air Temperature (OAT)

ii. Heating Degree Hours (HDH)

iii. Cooling Degree Hours (CDH)

iv. Month (Categorical variable; 1 through 12 for each month)

v. Day (Categorical variable: 1 for weekday, 0 for weekend)

vi. Hour (Categorical variable: 0 through 23 for each hour in a day)

Chapter 5: RESULTS AND DISCUSSION

The predicted values were analyzed in multiple ways to understand the nature of the demand in the building. For each of the model outputs: demand for the TMY data and demand for the four different meteorological scenarios, following charts were looked at to understand the characteristics of the demand and how they change with changing temperatures.

i. Average hourly demand for each month
ii. Average hourly demand on Weekdays for each month
iii. Average hourly demand on Weekends for each month
iv. Average hourly demand over the summer months (Jun-Sept) and winter months (Nov-Feb)
v. Average hourly demand for the day shift (7am-10pm) and the night shift(10pm-5am)
vi. Average hourly demand on weekdays and weekends throughout the year

The observations obtained for each of the model output was analyzed and the highlights of each of them are presented below.

5.1 Observations of the Model Output for TMY data

A total of 42 demand graphs were analyzed with the target to dissect the chilled water, condensate, and electricity demand on multiple temporal frames using the Typical Meteorological Year (TMY) data. Following observations were made with their graphs presented below.

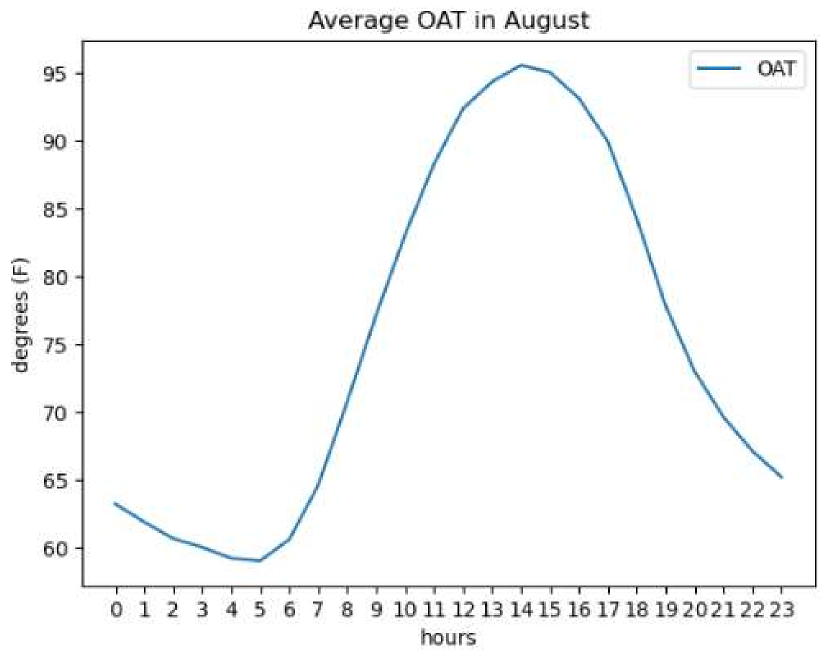


Figure 22 Average OAT for August

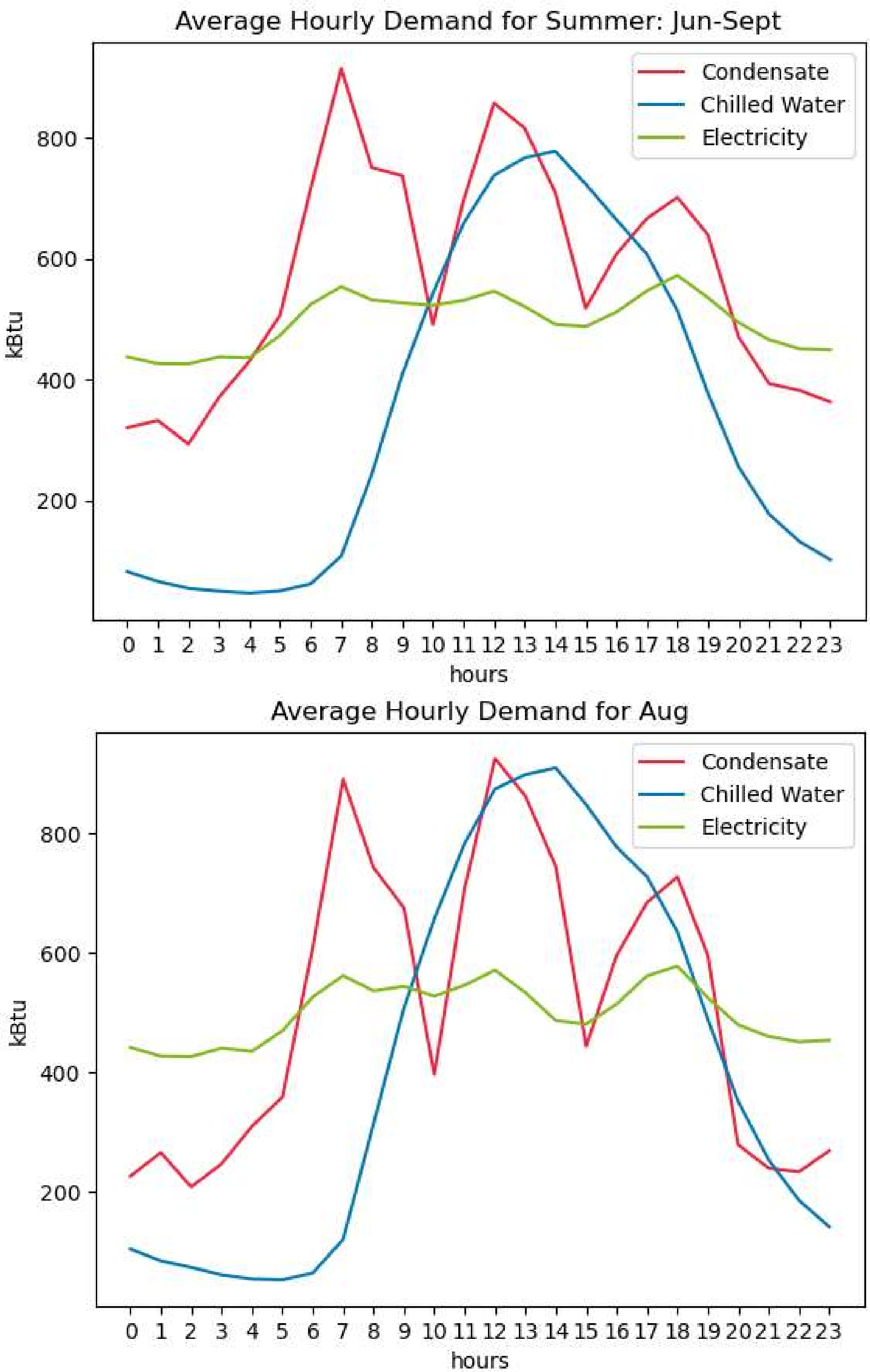


Figure 23 Average hourly demand for the summer and Max Summer Demand

Simultaneous Heating and Cooling Demand

As shown in the figures above, it can be observed that both the condensate and chilled water demands are high during the summer months and highest in the month of August. This was found to be very unusual and energy wasteful if this was only for the purposes of space heating or cooling.

After observation of these graphs, a walkthrough energy audit of the facility was conducted to understand the reason for this highly unusual activity. Even though a site inspection and analysis of internal factors affecting energy in a building is out of scope of a black-box demand data driven analysis, the site inspection was done to understand the reason for such an atypical demand characteristic. It was found from the site visit and interview of staffs in the facility that the reason for such a high condensate demand in summer was for cooking purposes. The dining commons was found to use steam to prepare some food items served in Segundo dining commons and other dining commons in campus. Though this might be the major contributor for the condensate demand spike, there are possibilities of leakage in heating hot water valves in the building's HVAC equipment that can raise condensate demand but in a lower extent.

By comparing OAT with the condensate demand graph, it can be observed that the reason for the decrease of condensate demand between hours 7am to 10am can be attributed to reduction in use of condensate for cooking purposes because the OAT rises in the same time-period, so space heating does not occur at this time. This is also validated by the rise in chilled water demand between 7am to 10am. The condensate demand can be seen rising again sharply from 10am to noon, which can again be mainly driven by cooking instead of space heating. The difference between this peak and trough was calculated to be 528 kBtu, which is the condensate demand for cooking in the dining commons.

Insignificant Difference in Demands on Weekdays and Weekends

The average hourly demand for weekdays and weekends throughout the year was compared to understand the difference in energy consumption between these periods. It was found as indicated in the following graph that the energy demand does not have a significant difference between these time periods.

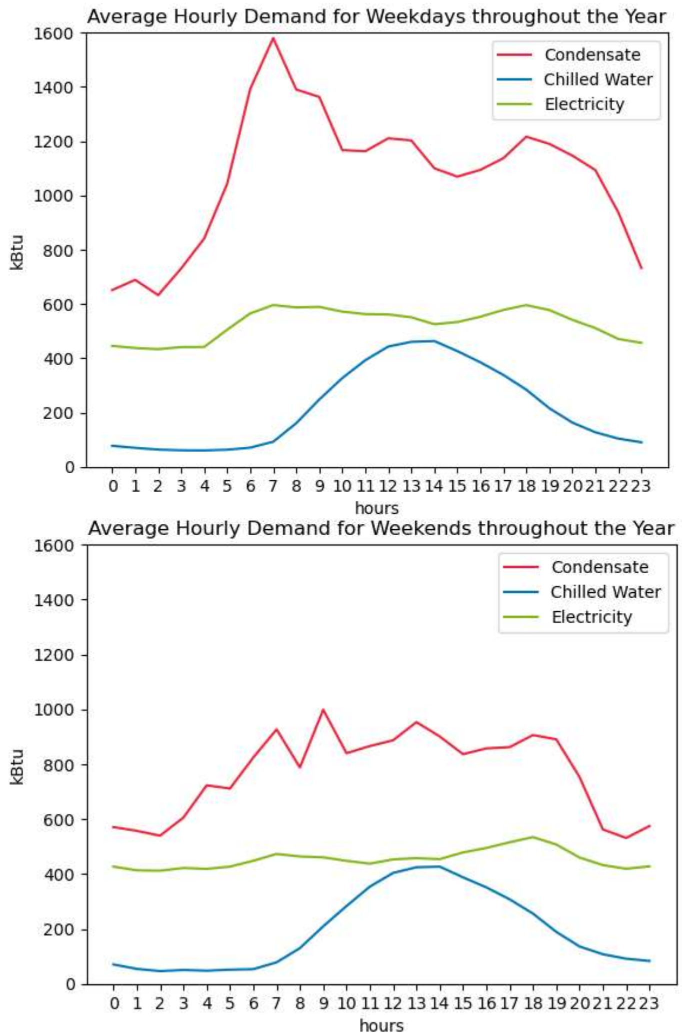


Figure 24 Average Hourly Demand for Weekdays and Weekends

It can be observed that the chilled water demand and electricity demand is almost the same in weekdays and weekends. Condensate demand during weekdays is slightly higher- around 600 kBtu on average, compared to weekends, but not as high as expected. It should be noted that the dining commons is open 15 hours on weekdays whereas its open only 8 hours on weekends, and with many students generally leaving campus on weekends, lower demands were expected during this time.

Stationary Electric Demand

The electricity demand for all temporal the resolutions- Weekdays, Weekends, Summer, Winter is same throughout the day which is around 500 kBtu.

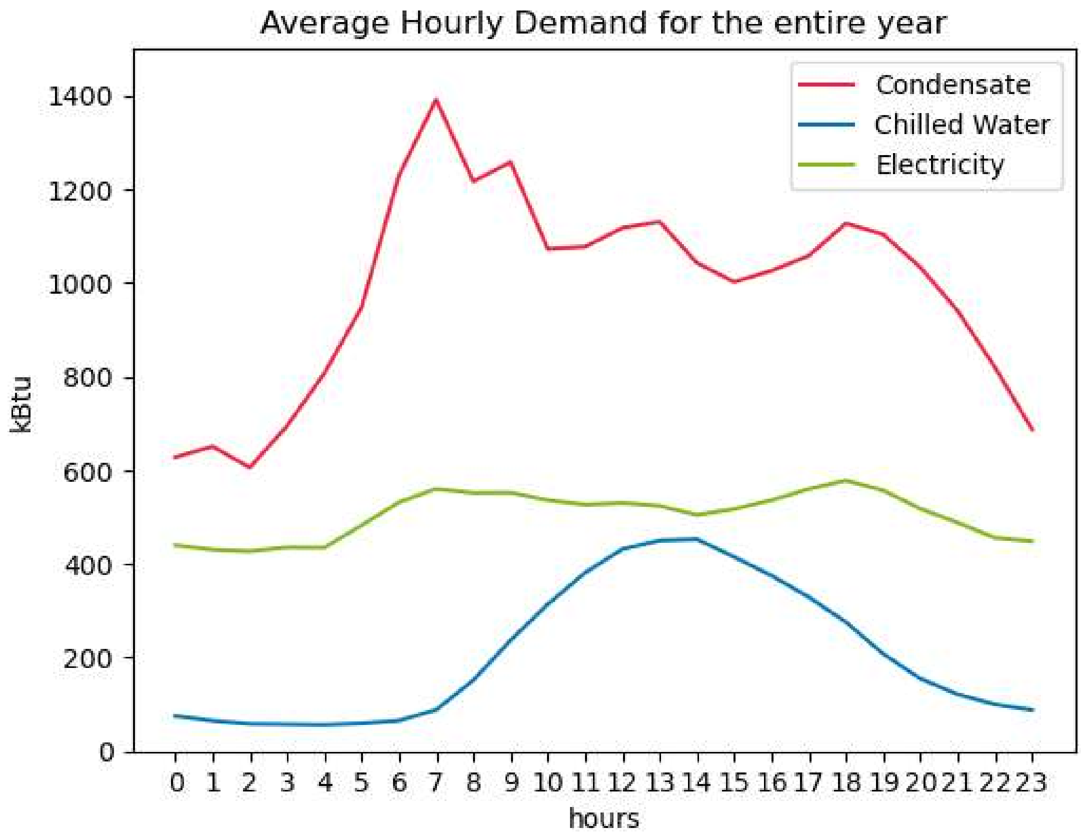


Figure 25 Average Hourly Demand for the entire year

This static demand of electricity throughout the year signifies that HVAC equipment electricity consumption in the facility is minimal compared to other electricity consuming equipment like Refrigerators, Lights, Plug loads, etc. in the facility. The electric demand is independent of time of day, OAT and work shifts in the facility. From the demand data analysis, it can be predicted that most of the electric consumption comes from equipment that continuously consumes electricity like freezers in the facility.

High Energy Consumption During School Breaks

Average demand for all three commodities has a high baseline even during long school breaks like summer and winter breaks. The average monthly demand for all three commodities is as represented below.

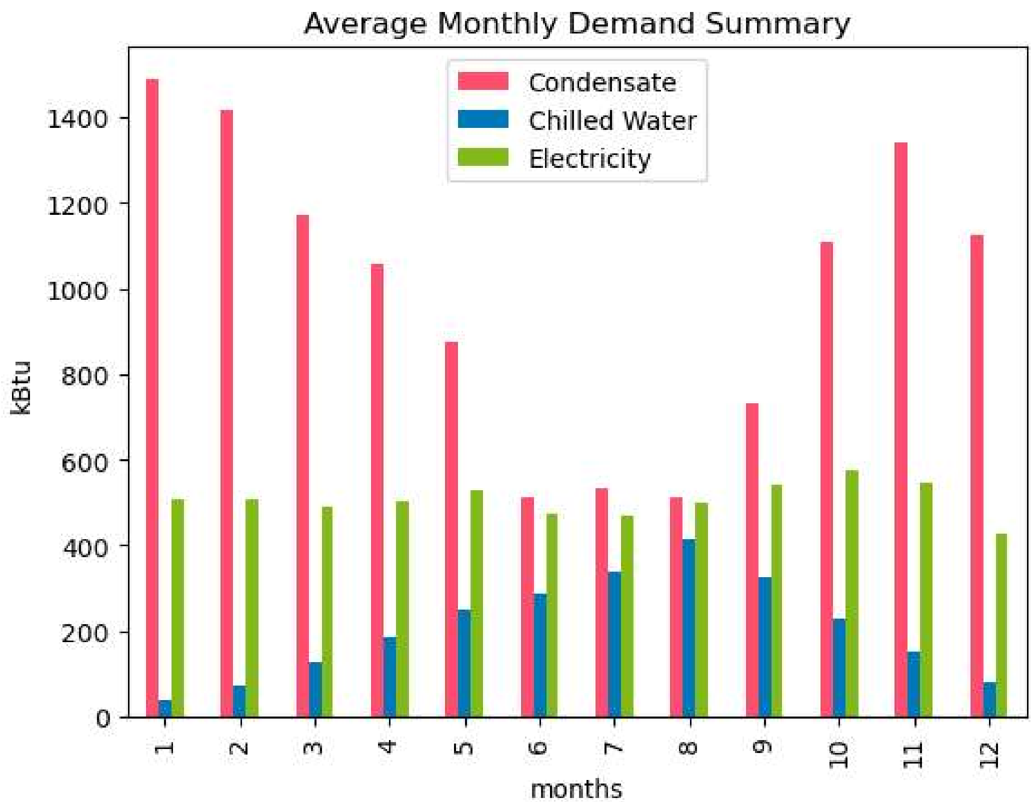


Figure 26 Average Monthly Demand for the year

The two major school year breaks are three months long summer break from the third week of June to third week of September and a month-long winter break from first week of December to first week of January. As can be seen from the graph, the baseline energy consumption during these school breaks is still quite high.

5.2 Observations of the Model Output for the Scenarios

Like the observations from TMY data, 42 graphs for different temporal frames were analyzed for each of the four scenarios. They were compared against the original prediction from TMY data and against each other to identify the differences in demand for these times.

The first three scenario of increase in temperature by 0.5°C, 1°C, 2°C did not produce very different prediction results from the TMY data. The graphs from the prediction showed similar characteristics as described in the section above besides minor differences. Therefore, it can be concluded that the commodities demand for Segundo dining commons does not change significantly with the specified rises in outdoor air temperature.

The final scenario of analyzing the model results using the max temperature produced slightly different results than the base scenario of using TMY data. Some of the main observations from the model result is as follows.

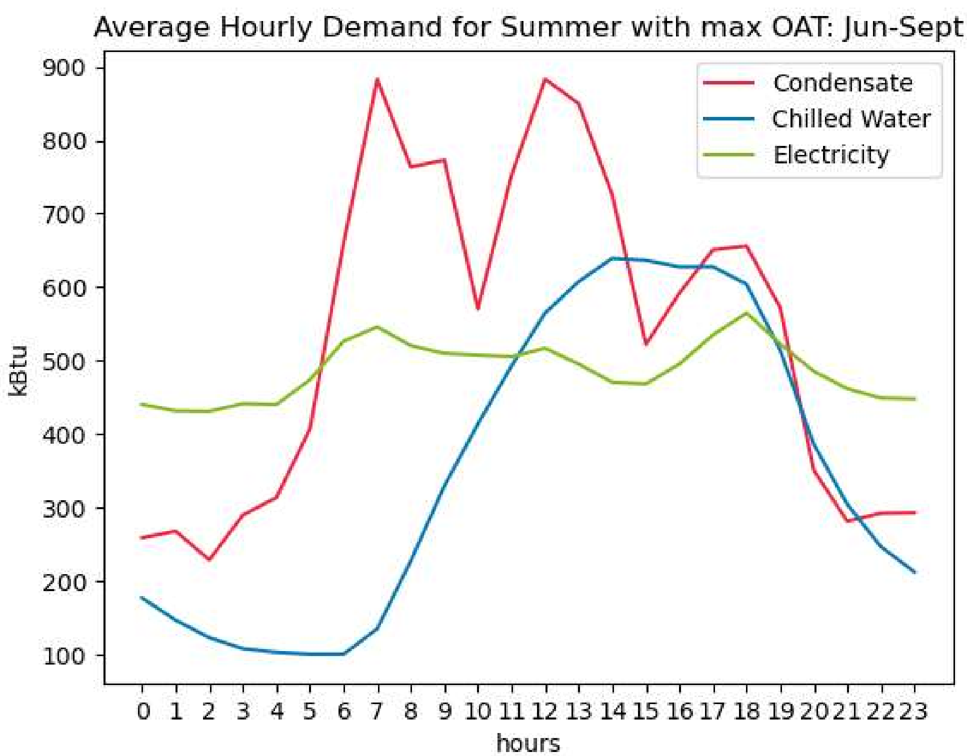


Figure 27 Average hourly demand for the summer using max OAT data

The hourly demand graph presented above is slightly different than the base case scenario in terms of the nature of the peaks for condensate and chilled water demand. The condensate demand in the evening is slightly lower than the base case condensate demand. The chilled water demand is also slightly shifted towards the evening compared to the base case scenario.

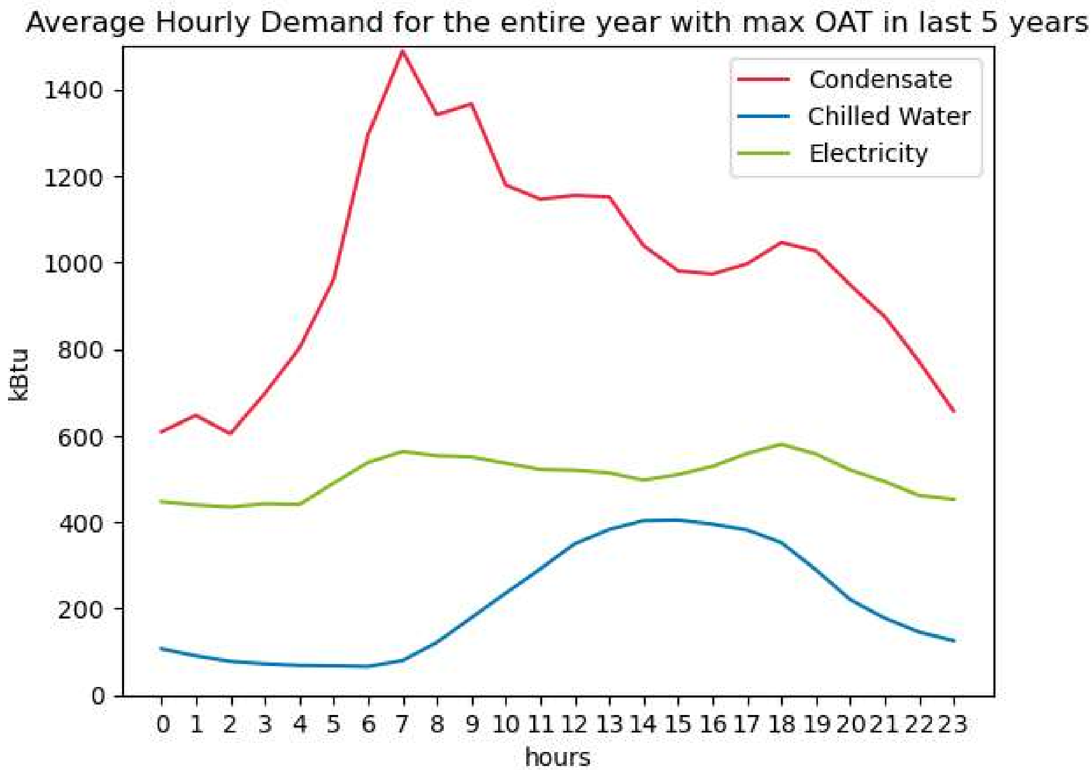


Figure 28 Average Hourly Demand for the entire year with max OAT

The other observation from this scenario is that the peak demands for condensate and chilled water are slightly higher compared to the base scenario. The difference between them is however not very significant. The demands of the three commodities were summed for the entire year and it was found that the Chilled water demand was 2.54% higher than the base case scenario. The other two commodities: Electricity and Condensate demand only increased by 0.5%. This shows that the condensate demand is highly influenced by cooking than for space heating. It can also be seen that the energy consumption in the dining commons is more influenced by internal characteristics of the building like equipment efficiency, setpoints, processes in the facility etc. than by outside air temperature.

5.3 Identified Energy Saving Opportunities

Several energy saving opportunities were identified at a high level from the observation of the model results. They are as listed below.

i. Heat from the cooking process using steam can be leaking to the ambient space and increasing the demand for cooling since the steam demand is considerably high. The simultaneous condensate and chilled water demand could be reduced to some extent by thoroughly auditing the cooking equipment and process with the intent to minimize heat escape into the surrounding spaces.

ii. All three commodities demand are almost similar for both weekdays and weekends while weekend serving hours are lower compared to weekdays. A detailed internal analysis of building equipment and processes can help identify exact strategies to understand the reason for the similar demand and optimize energy use.

iii. The electricity demand is almost constant throughout the year and times of the day. This shows that most of the equipment are kept on throughout the year, so there should be opportunities in optimizing electricity usage through a detailed calculation of power demand for all equipment and processes. Lighting and other equipment power demand can be calculated and optimized using occupancy sensors and meal demands.

iv. School breaks should be a period of very low demand unless the dining commons is used for other purposes, so the commodities demand should be minimal during this period. However, chilled water demand is highest, electricity demand is same as other periods and condensate demand is around half of the highest demand in winter. This shows that a lot of energy use during summer can be saved by understanding how the facility is used at that time and adjusting building HVAC schedules.

5.4 Discussion

Broader areas to focus on were identified for the energy saving strategies through the use of model prediction, however, for identification of specific activities to realize energy savings, a detailed analysis of energy system of the building, equipment used, processes and occupancy is necessary. This thesis could not explore into the facility's equipment level evaluation because of time constraints. Other researchers can have an idea from the identified energy savings opportunities in this thesis and investigate the details for each equipment in the facility to devise a specific action plan for energy savings.

The scope of this research was only limited to the use of supervised machine learning methods for the facility demands. Other unsupervised learning methods could also be applied to equipment level data from the facility to identify relationships within the building's energy system. This would result in specifically understanding the reason of patterns identified through this research.

The data sets used to train the machine learning models were relatively shorter datasets that represented only two years of demand data. More than two years of data could not be used because of significant differences in the energy demand compared to previous periods. If the demand had not changed and if more years of data could be used, the machine learning models may have been better trained and might have resulted in better model metrics. However, using machine learning models to analyze demand characteristics is better than observing a particular year of demand data as it might not be able to represent typical operation for the facility because of year specific differences in operations or equipment use or even outside air temperatures. Also with the development of the model, input data can be changed to see how it affects demand prediction.

Chapter 6: CONCLUSIONS

This thesis applied different types of machine learning methods for future demand prediction and analyzed the results of one of the highest energy consuming facilities on campus. It presented procedures and tools to facilitate detailed insights into Segundo Dining Common's energy consumption. The meter data for the past 8 years were collected, analyzed and 2 years of it was fed into 6 different types of machine learning models to train them for future demand prediction. These machine learning models were compared against test data and ANN model was chosen based on model metrics comparison. This model was then fed with a future meteorological data with four other scenarios: 0.5°C, 1°C, 2°C rise in outside air temperatures and max recorded temperatures for each day of each hour in the past 5 years. The prediction was then analyzed in multiple temporal scales to understand the nature of energy use in the facility. From the analysis, four specific areas to focus energy saving opportunities were presented that can help develop detailed energy saving strategies.

This analysis facilitated a visual understanding of the nature of energy consumption in the building and gave a direction to prioritize energy efficiency efforts in the facility. The result allows future energy efficiency engineers and researchers working on this facility to prioritize areas that have high energy saving potential. Analyzing building's internal equipment based on the identified opportunities presented will help develop detailed plans for energy savings. This energy demand analysis of the building was possible by training the model with two years of data and observing the results instead of just looking at a specific year's past data. If just a specific year's demand data was observed, it might have led to a different observation as that demand is likely to be influenced by changes in that particular year and might not represent typical demand. Therefore, building a model and analyzing its output presents a better way of looking at the nature of energy demand than just observing past data. This benefit of using the model instead of a year's demand data is further enhanced when multiple years of data that is not significantly different from each other is available for model training.

Finally, the result of this thesis is useful in developing detailed energy saving plans in the facility and understanding the change in demand when outside conditions are altered.

6.1 Future Work

There are multiple directions for future work based on the results of this research. One direction if for future engineers and researchers working in this facility to focus on the building's internal characteristics like the type and nature of HVAC equipment used, operational characteristics like times of cooking, number of staffs in each area of the facility and their occupancy timing, etc. This information can help devise specific plans for optimizing energy usage.

Another direction of future work can be to explore the effect of using steam cooking in chilled water demand through heat leakage from the cooking areas. If the high simultaneous chilled water and condensate demand as presented in the results is because of the heat leakage, strategies can be devised to exit the heat outside of the building instead of it leaking into the interior spaces and driving cooling demand.

The other direction for future work can be understanding the steam demand for cooking and exploring electric options for it. This would help reduce the steam demand in the central heating and cooling plant and help in the UC Davis' instead of using the steam as it would help the campus' ambitious carbon neutrality goal by 2025.

Optimizing the machine learning models used in this research can also another direction of future work. Large demand datasets with higher temporal resolution could be used to check if the other energy usage patterns can be identified. Appropriate data that can represent building occupancy could be identified like recording ID swipes, Wi-Fi connected devices, footfall door counters, etc to include in the model training and see the correlation between occupancy and energy demand.

Another machine learning based future work could be using unsupervised learning methods on the building's equipment data like supply and return air temperatures, heating and cooling coil opening percentages, airflows in each space, etc. to identify correlation between energy use parameters in the facility.

Chapter 7: REFERENCES

Energy and Engineering. (2023). *M&V savings from Winter Break Savings Strategies.*

IEA. (2020). *Energy Technology Perspectives.* IEA, Paris: https://www.iea.org/reports/energy-technology-perspectives-2020, License: CC BY 4.0.

IEA. (2022). *Buildings.* IEA, Paris: https://www.iea.org/reports/buildings, License: CC BY 4.0.

IPCC. (2018). *Global Warming of 1.5°C.An IPCC Special Report.* Retrieved from https://www.ipcc.ch/sr15/resources/headline-statements/

Kalogirou, S.A., Florides, G.A., Neocleous, & C. (2001). Estimation of Daily Heating and Cooling Loads Using Artificial Neural Networks. (http://ktisis.cut.ac.cy/bitstream/10488/883/3/C41-CLIMA2001.pdf).

Khalil, M., McGough, A., Pourmirza, Z., Pazhoohesh, M., & Walker, S. (2022). Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption — A systematic review. *Engineering Applications of Artificial Intelligence*, ttps://doi.org/10.1016/j.engappai.2022.105287.

Kumar, R., Aggarwal, R., & Sharma, J. (2013). Energy analysis of a building using artificial neural network: A review. *Energy and Buildings*(https://doi.org/10.1016/j.enbuild.2013.06.007).

Mabberley, C. L. (2022). Food Service Supervisor, Segundo DC.

Magoulès, F., Zhao, H.-x., & Elizondo, D. (2013). Development of an RDP neural network for building energy consumption fault detection and diagnosis. *Energy and Buildings*(https://doi.org/10.1016/j.enbuild.2013.02.050).

Molina-Solana, M., Ros, M., Ruiz, M., Gómez-Romero, J., & Bautista, M. (April 2017). Data science for building energy management: A review. *Renewable and Sustainable Energy Reviews, vol. 70*, 598-609.

S. Seyedzadeh, T. e. (2018). Machine Learning for estimation of building energy consumption and performance: a review. *Visualization in Engineering*.

Seem, J. E. (2007). Using intelligent data analysis to detect abnormal energy consumption in buildings. *Energy and Buildings 39 (2007) 52-58*.

Seyedzadeh, S., Rahimian, F., Glesk, I., & Roper, M. (2018). Machine learning for estimation of building energy consumption and performance: a review. *Visualization in Engineering*(https://doi.org/s40327-018-0064-7).

Shaikh, P., Bin Mohd Nor, N., Nallagownden, P., Elamvazuthi, I., & Ibrahim, T. (2014). A review on optimized control systems for building energy and comfort management of smart sustainable buildings. *Renewable and Sustainable Energy Reviews*(https://doi.org/10.1016/j.rser.2014.03.027).

Tien, P. W., & Shuangyu Wei. (2022). Machine Learning and Deep Learning Methods for Enhancing Building Energy Efficiency and Indoor Environmental Quality – A Review. *Energy and AI*, https://doi.org/10.1016/j.egyai.2022.100198.

Typical Meteorological Year (TMY) NREL. (2023, 03 10). Retrieved from https://nsrdb.nrel.gov/data-sets/tmy

United Nations Department of Economic and Social Affairs, P. D. (2022). *World Population Prospects: Summary of Results.* UN DESA/POP/2022/TR/NO. 3.

United Nations Environment Programme. (2022). *Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector.* Nairobi.