7-6-2022

A Customer Agnostic Machine Learning Based Peak Electric Load Days Forecasting Methodology for Consumers With and Without Renewable Electricity Generation

Omar Aponte
oxa4944@rit.edu

Kate Gleason College of Engineering
Department of Industrial and Systems Engineering

A Customer Agnostic Machine Learning Based
Peak Electric Load Days Forecasting Methodology
for Consumers With and Without Renewable Electricity Generation

by

Omar Aponte

(née Omar Aponte Contreras)

A dissertation submitted in partial fulfillment
of the requirements for the Degree of
Doctor of Philosophy in Mechanical and Industrial Engineering

Department of Industrial and Systems Engineering
Kate Gleason College of Engineering

Rochester Institute of Technology
Rochester, New York
July 6th, 2022

Committee Approval:

We, the undersigned committee members, certify that we have advised and/or supervised the candidate on the work described in this dissertation. We further certify that we have reviewed the dissertation manuscript and approve it in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Mechanical and Industrial Engineering.

Katie T. McConky *Aug. 18th, 2022*

Prof. Katie T. McConky, Ph.D. Date
Advisor, Department of Industrial and Systems Engineering, KGCOE

Brian K. Thorn *Aug. 17th, 2022*

Prof. Brian K. Thorn, Ph.D. Date
Committee Member, Department of Industrial and Systems Engineering, KGCOE

Bing Yan *Aug. 18th, 2022*

Prof. Bing Yan, Ph.D. Date
Committee Member, Department of Electrical and Microelectronic Engineering, KGCOE

Qi Yu *Aug. 18th, 2022*

Prof. Qi Yu, Ph.D. Date
Committee Member, School of Information, GCCIS

Katie T. McConky *Aug. 18th, 2022*

Prof. Katie T. McConky, Ph.D. Date
Director, Ph.D. in Mechanical and Industrial Engineering Program, KGCOE

Acknowledgements

As a believer in the existence of a higher power (which I have personally chosen to refer to as God) that controls all of the variables that we as humans can't control and don't understand, I would like to first express my gratitude to God for allowing everything to line up and work out so that I can reach this moment in my life. Second, I feel like every adjective that I can think of, falls short to describe how grateful I am to my amazing advisor Dr. Katie T. McConky. Dr. K was nice enough to grant me a meeting after showing up unannounced at her office and without knowing me at all. After that meeting she gave me the most important opportunity that I had gotten up to that point in my life, she supported me both personally and academically throughout this journey, and without her, this dissertation would not be a reality and I would not be the researcher and professional that I am today. Thank you Dr. Katie! You are an amazing role model!

I would like to thank the members of my committee, who have been instrumental in improving my research and have provided me with insights and guidance that I will hold on to for the rest of my career as a researcher and a professional. Dr. Brian K. Thorn, Dr. Bing Yan, and Dr. Qi Yu, I thank you for your time, your valuable insights, and all of your advice! The list of mentors, RIT departments, staff, faculty members, and peers that have been instrumental in the completion of this journey is extensive. You are all in my heart and I thank you immensely for your support!

This research would not have been possible without the financial support provided by the Kate Gleason College of Engineering (KGCOE) at RIT, the Industrial and Systems Engineering department at KGCOE, the New York State Energy Research and Development Authority (NYSERDA) through the agreement number 137653, and the International Institute of Forecasters through the 2019-2020 SAS Grant to Promote Research on Forecasting. Thank you!

I would also like to thank the industrial consumer that gracefully provided me with and allowed me to use their electricity consumption data to perform this research. This consumer asked to remain anonymous. A huge thank you! Goes to the Bedon family for providing me and allowing me to use their residential electricity consumption data to perform this research, thank you!

Dedication

This accomplishment is dedicated to:

My parents, José and Angela, for bringing me into this world and giving me the love and support that have allowed me to reach my goals.
All of my successes are also your successes!

My siblings, Joel and Lidia, for being my partners in crime in many occasions, a source of sound and honest advice about anything and everything, and supporters of my wildest plans.
I will always love you both and your families!

To my only living grandmother, Gloria, who called me at least once every week and sent me care packages every two weeks as if I was deployed in a war zone.
She has always gone above and beyond for those that she loves and I love her for loving me like that!

To my extended family, uncles, aunts, cousins, every family member that has kept an eye on me and sent me their love and support.
You have all being instrumental in this accomplishment!

To those that I call my friends, without your company and support during the good times and especially during the not so good times when I needed a break or a scape with good company, I would not be here at this point in my life right now.
This accomplishment would not be possible without you!

Abstract

The adoption of electricity generation from renewable sources, as well as the push for a speedy electrification of sectors such as transportation and buildings, makes peak electric load management an essential aspect to ensure the electric grid's reliability and safety. Utilities have established peak load charges that can amount to up to 70% of electricity costs to transfer the financial burden of managing these loads to the consumers. These pricing schemes have created a need for efficient peak electric load management strategies that consumers can implement in order to reduce the financial impact of this type of load. Research has shown that the impact of peak load charges can be reduced by acting on the intelligence provided by peak electric load days (PELDs) forecasts. Unfortunately, published PELDs forecasting methodologies have not addressed the increasing number of facilities adopting behind the meter renewable electricity generation. The presence of this type of intermittent generation adds substantial complexity and other challenges to the PELDs forecasting process.

The work reported in this dissertation is organized in terms of its three main contributions to the body of knowledge and to society. First, the development and testing of a first of its kind PELDs forecasting methodology able to accurately predict upcoming PELDs for a consumer regardless of the presence or absence of renewable electricity generation. Experimental results showed that 93% and 90% of potential savings (approximately US$ 142,129.01 and US$ 123,100.74) could be achieved by a consumer with and a consumer without behind the meter solar generation respectively. The second contribution is the development and testing of a novel methodology that allows virtually any type of consumer to determine an efficient electricity demand threshold value before the start of a billing period. This threshold value allows consumers to proactively trigger demand response actions and reduce peak demand charges without receiving any type of signal or information from the utility. Experimental results showed 65% to 82% of total potential demand charge reductions achieved during a year for three different consumers: residential, industrial, and educational with solar generation. These results translate to US$ 149.09, US$ 23,290.00, and US$ 107,610.00 in demand charges savings a year respectively. As a third contribution, we present experimental results that show how the implementation of machine learning based ensemble classification techniques improves the PELDs forecasting methodology's performance beyond previously published ensemble techniques for three different consumers.

Table of Contents

List of Figures

Page

List of Tables

List of Abbreviations and Acronyms
(In alphabetical order)

AdaBoost:	Adaptive boosting
AID:	Automatic interaction detection
ANN:	Artificial Neural Network(s)
ARIMA:	Autoregressive integrated moving average
ARIMAX:	Autoregressive integrated moving average with exogenous variables
AUC:	Area under the curve
BTMREG:	Behind the meter renewable electricity generation
CART:	Classification and regression trees
Class:	Classification
HVAC:	Heating, ventilation, and air conditioning
kW:	Kilowatt \ kilowatts
MAPD:	Mean absolute percentage deviation
MAPE:	Mean absolute percentage errors
MARS:	Multivariate adaptive regression splines
MW:	Megawatt \ megawatts
NOAA:	National Oceanic and Atmospheric Administration
PEL:	Peak electric load
PELD(s):	Peak electric load day(s)
REG:	Renewable electricity generation
RF:	Random forest
ROC:	Receiver operating characteristic
RSS:	Residual sum of squares
SMOTE:	Synthetic minority oversampling technique
SVR:	Support vector regression
SVRM:	Support vector regression machine
THAID:	Theta automatic interaction detection
USA:	United States of America

Chapter 1: Introduction

This first chapter is divided into eight sections that will serve to introduce the reader into the research motivation, the problem statement, and the societal context surrounding this dissertation. A brief state of the art summary will be provided covering topics such as the worldwide impact of electricity consumption, peak electric loads, peak electric load days, and behind the meter renewable electricity generation adoption. The chapter will conclude with a description of the research questions and dissertation objectives to be addressed, and a summary of the methodology.

1.1 Research Motivation

The constant evolution of the electric grid with the integration of generation from renewable sources and "smart" components makes peak electric load management an essential aspect to ensure the grid's reliability and safety. In order to pass the financial burden of managing these loads on to the consumers, utilities around the world have established peak load charges that can amount to up to 70% of electricity costs in the case of the United States of America (Xu et al., 2019; Zhang and Augenbroe, 2018; McLaren et al., 2017). These pricing schemes have created a need for efficient electric load management strategies that consumers can implement in order to reduce the financial impact of peak electric loads.

1.2 Worldwide Impact of Electricity Consumption

Commercial and residential facilities require a significant amount of energy and contribute a considerable amount of greenhouse gas emissions worldwide. International and domestic agencies that focus on energy related statistics include a distinct category to report the energy consumed by commercial and residential buildings. This practice is a testament to the significant impact of these consumers' energy usage. The International Energy Agency reported that buildings were responsible for 28% of global energy-related CO2 emissions in 2018 (IEA, 2019). The United States' Energy Information Administration reported that the residential and commercial sectors represented 39% of the total energy consumption in the United States of America (USA) during 2019 including losses (USEIA, 2020). During 2018, sustained ongoing efforts to decarbonize energy generation worldwide increased the share of renewable energy in global power capacity to 33% (REN21, 2019). However, the increase in building electricity consumption was five-times faster than the improvements in the carbon intensity of the power sector during the 2000-2018 period (IEA, 2019). Given buildings' significant energy requirements and contributions to greenhouse gas emissions, it is imperative that research efforts continue to focus on ways to increase buildings' energy efficiency in order to reduce their energy related costs.

1.3 Peak Electric Loads

Many commercial and residential buildings are billed under dynamic pricing schemes that can include peak demand charges (Dutta and Mitra, 2017; McLaren et al., 2017; Hledik, 2014). Peak load charges are typically based on the highest electric load (measured in kilowatts (kWs)) observed during a billing period, typically a month, and are charged in $/kW (McLaren et al., 2017;

Hledik, 2014). These peak load charges can amount to up to 70% of an electric bill in the USA (Xu et al., 2019; Zhang and Augenbroe, 2018; McLaren et al., 2017). However, the number of days in a month contributing to peak load charges is typically very small. Figure 1 illustrates real electric load data for an educational consumer in the USA during the month of April 2019. Figure 1 clearly shows that April 8th, 12th and 18th have significantly higher load levels than the other days of the month. If these peak load days are forecasted ahead of time, demand response actions could be executed to mitigate the electric load during these specific days and reduce the peak load charges described earlier. The lead-time provided by these peak electric load days (PELDs) forecasts is very important because some demand response actions require several hours either to be executed, to show results, or both.

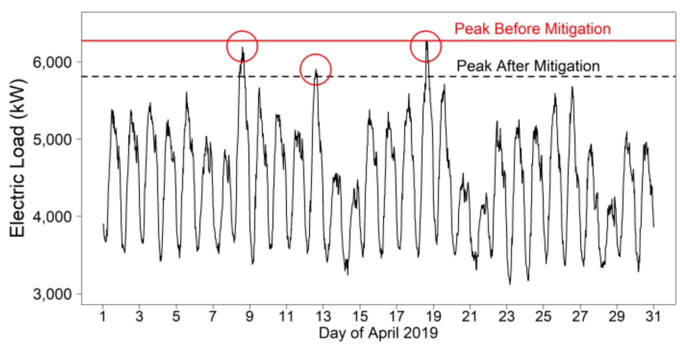

Fig. 1. Electric load for a circuit during April 2019.

1.4 Peak Electric Load Days

Most of the published research on electric load forecasting focuses on generating accurate electric load forecasts for both utilities and consumers, but there is limited research on the application of these forecasts to avoid the peak load charges described earlier. Saxena et al. (2019) noted that studies focusing on forecasting a billing period's peak electric load days (PELDs) in order to trigger demand response actions to reduce peak load charges are scarce. These authors reviewed how machine learning based models have been used to develop forecasting methodologies for next-day building electric load and peak load. However, being able to predict the next day's electric load does not provide actionable intelligence to determine if the next day will contribute to peak load charges for the billing period. Saxena et al. developed an ensemble machine-learning model focused specifically on predicting if the next day will be a peak electric load day for a billing period. Saxena et al. tested their model using data from an educational consumer; the ensemble model predicted 70% of actual peak electric load days and revealed potential savings in the neighborhood of US $80,000 after a yearlong testing period. The testing phase was limited to one out of four total distribution circuits at the campus.

As Saxena et al.'s methodology was being prepared for implementation by the educational consumer, the campus' electricity infrastructure was combined into just two main distribution circuits. Each of these main circuits now included a solar field designed to provide up to 2 megawatts (MW) of behind the meter renewable electricity generation (BTMREG). Even though the Saxena et al. methodology had been validated for an electrical circuit without BTMREG at an educational consumer's campus, the methodology had not been tested for circuits with BTMREG.

1.5 Impact of Behind the Meter Renewable Electricity Generation Adoption

Researchers had already noted that renewable electricity generation (REG) output is as variable as weather itself (Staffell and Pfenninger, 2018; Chaiamarit and Nuchprayoon, 2014). Aponte and McConky (2019) documented how behind the meter renewable electricity generation (BTMREG) output could represent a challenge for the accuracy of current peak electric load days (PELDs) forecasting methodologies. REG output can fluctuate for periods ranging from minutes to hours to multiple days (Staffell and Pfenninger, 2018). This characteristic of REG challenges the accuracy of both electric load forecasts (Tushar et al., 2018) and peak electric load days forecasts (Aponte and McConky, 2019).

1.6 Problem Statement and Societal Context

Mismanaged peak electric loads are not only a threat to the electric grid’s reliability and safety; they can also cause unplanned increases in electricity costs for consumers, and negatively affect the environment. Even though efforts to decarbonize the electric sector have been increasing significantly worldwide, electricity consumption has increased at a more accelerated pace. The amount of commercial and residential electricity consumers adopting some type of REG worldwide has been increasing year after year. Utilities worldwide pass the financial burden of managing peak electric loads on to the consumers through dynamic pricing schemes. Some of these schemes include peak load charges that can amount to up to 70% of electricity costs in the USA. Given this market reality, efficient electric load management strategies become even more important for consumers to reduce the significant financial impact of peak electric loads. Additionally, these strategies need to be applicable not only to consumers without behind the meter renewable electricity generation (BTMREG) but also to the increasing number of consumers adopting BTMREG worldwide.

1.7 Research Questions and Dissertation Objectives

The main objective of this dissertation is inspired by the thorough study of the state of the art that will be detailed in Chapter 2 of this manuscript, as well as by the problem and societal context previously described in Section 1.6. The main objective is to provide electricity consumers with and without behind the meter renewable electricity generation (BTMREG), with accurate and accessible machine learning based peak electric load days (PELDs) forecasting techniques able to help them react in order to reduce the significant financial impact of peak electric loads. Three research questions related to the use of PELDs forecasting techniques are addressed throughout this dissertation in order to achieve the main objective:

1. Can threshold based and/or classification based forecasting methodologies accurately forecast more than 70% of a year’s peak electric load days for consumers with behind the meter renewable electricity generation using autoregressive integrated moving average (ARIMA), classification and regression trees (CART), random classification and regression forest, and artificial neural network (ANN) based machine learning techniques?

    Even though there is an abundance of published work related to future load forecasting methodologies, published research detailing PELDs forecasting methodologies based on accessible ARIMA, CART, and ANN based techniques applicable to the increasing number of facilities adopting BTMREG is very limited.

Furthermore, published studies comparing the performance of PELDs forecasting methodologies with and without BTMREG while providing insights into the impacts of BTMREG adoption on forecasting strategies are not available.

2. Can regression tree and random regression forest based machine learning models outperform widely used expert based and arithmetic based methodologies at forecasting an efficient electricity demand threshold value for triggering cost saving peak demand shaving actions, before the start of a billing period, and without receiving any signal or information from the utility?

Threshold-based peak electric load classification is one of the approaches for electric load classification currently available in the published literature (Saxena et al., 2019). The vast majority of the published methodologies on peak demand shaving and other demand response actions has focused on the consumer reacting to signals or information coming from the utility. Electric load forecasting techniques that rely on the use of machine learning based models have been widely used in the published literature to forecast future electric load values. However, this work will be the first to use these techniques to empower consumers under demand charges to proactively determine an appropriate electricity demand threshold value in order to trigger peak demand shaving and other demand response actions even without receiving any signal or information from the utility.

3. Can classification tree, random classification forest, adaptive boosting (AdaBoost), and artificial neural network (ANN) based ensemble modeling techniques outperform majority based and single-vote based ensemble modeling techniques at forecasting peak electric load days (PELDs)?

At present, a published methodology for forecasting PELDs for consumers without BTMREG produced the best results by using a majority-classifier based ensemble approach (Saxena et al., 2019). However, this ensemble approach has not been contrasted with machine learning based alternatives in order to ensure that the consumers are indeed achieving the best possible results.

Chapter 2: State of the Art

This second chapter is divided into ten sections that will allow the reader to take a deep dive into the state of the art of electric load forecasting, peak electric load forecasting, electricity demand threshold forecasting, ensemble peak electric load days forecasting, and the most relevant models to be discussed within this dissertation. The initial sections (2.1 to 2.4) will be dedicated to presenting the most important contributions found during an extensive review of the published literature relevant to the different forecasting tasks to be discussed within this dissertation. Five sections (2.5 to 2.9) will be dedicated to tracing the origins, mathematical foundations, and characteristics of each of the four key models that will be used to develop the methodology described within this dissertation. The chapter will conclude with a brief and all encompassing summary of the state of the art relevant to the topics discussed within this dissertation.

2.1 Electric Load Forecasting

Future electric load forecasts have been a core activity for utilities since the electricity industry began in the late 1800's (Hong, 2014). Utilities rely on electric load forecasts to plan their supply and generating capacities (Dutta and Mitra, 2017; Hong and Fan, 2016; Alfares and Mohammad, 2002), to inform revenue projections, rate design, energy trading, and more (Hong and Fan, 2016). The capacity of electric utilities to ensure a reliable service to their clients depends heavily on these demand forecasts. Electric load forecasting methods have been extensively researched over the past few decades. The literature provides an ample range of studies featuring various methodologies and models for this purpose (Yildiz et al., 2017; Hong and Fan, 2016; Garulli et al., 2015; Alfares and Mohammad, 2002).

Alfares and Mohammad (2002) conducted a review of more than 100 works published between the comprehensive review by Moghram and Rahman (1989) and February 2000. Alfares and Mohammad classified the published methodologies into the first nine categories shown in Table 1. The researchers also provided a brief description along with the advantages and disadvantages identified for each category. The authors observed what they described as a clear trend towards new, stochastic, and dynamic forecasting techniques. Fuzzy logic, expert systems and artificial neural network (ANN) were specific techniques highlighted by the authors. They also highlighted a trend towards hybrid methods that combine two or more techniques.

Table 1

Categorization of techniques found in electric load forecasting literature reviews.

Category	**Alfares and Mohammad (2002)**	**Hong and Fan (2016)**	**Yildiz et al. (2017)**
1) Regression models *(Including Semi-Parametric Additive Models)*	X	X	X
2) Exponential smoothing	X	X	
3) Iterative reweighted least-squares	X		
4) Adaptive load forecasting	X		
5) Stochastic time series	X		
6) Autoregressive moving average with external variables models based on genetic algorithms *(Including other autoregressive models)*	X	X	X
7) Fuzzy logic	X	X	
8) Artificial neural network (ANN)	X	X	X
9) Knowledge-based expert systems	X		
10) Support vector regression and machine (SVRM)		X	X
11) Gradient boosting machine		X	
12) Thermal models			X
13) Classification and regression trees (CART)			X

Hong and Fan (2016) published a tutorial review based on more than 25 representative load forecasting papers (13 of which were literature review papers) published between the work by Abu-El-Magd and Sinha (1982) and November 2015. The techniques evaluated by the authors are included within the categories specified in Table 1. One of the authors' conclusions was that a universally best load forecasting technique does not exist. Hong and Fan concluded that the data and jurisdictions are the factors that determine the appropriate technique and not the other way around.

Yildiz et al. (2017) reviewed more than 50 commercial building load forecasting works published between 1984 and March 2016 and identified the techniques within the categories specified in Table 1. The authors concluded that the machine learning models reviewed (ANN, support vector regression machine (SVRM), and classification and regression tree (CART) based) had a superior forecasting performance than the regression models included in the review. Forecasting daily peak electric load proved to be a more difficult task than forecasting day ahead hourly electric load for Yildiz et al. Kim and Kim (2019) performed a study comparing the performance of more than 10 models (including ARIMA, Holt-Winters exponential smoothing, and ANN) at forecasting peak electricity demand for buildings at the Chung-Ang University campus in Seoul, South Korea. The researchers concluded that all models performed similarly when forecasting 1 hour ahead, but the nonlinear autoregressive network with exogenous inputs (NARX) model outperformed the rest of the models at complete day ahead forecasts. The wide range of electric load forecasting methodologies and techniques currently found in the literature can be simplified into a general approach. This general approach entails the estimation of a load model from past data, and then using this model to predict future loads (Garulli et al., 2015).

Deep learning methods using Long Short Term Memory (LSTM) sequential models and Convolutional Neural Networks (CNN) have been used successfully to address building level electricity demand forecasts. Gao et al. (2019) presented a deep learning methodology using LSTM with an attention layer that achieved a better MAPE than just using LSTM at predicting electricity demand for an office building in Qingdao, China. Khan et al. (2020) developed a hybrid model based on CNNs and LSTM models. The researchers showed how the proposed hybrid model outperformed other state-of-the-art models at forecasting electricity consumption for both residential and commercial scenarios. One drawback of deep learning methodologies is the expertise required to implement them. A consistent conclusion among the researchers is that there is no one best performing model for all types of facilities.

2.2 Peak Electric Load Days Forecasting

Most of the published research on electric load forecasting focuses on generating accurate electric load forecasts for both utilities and consumers, but there is limited research on the application of these forecasts to avoid the peak load charges described earlier in Section 1.3. Saxena et al. (2019) noted that studies focusing on forecasting a billing period's peak electric load days (PELDs) in order to trigger demand response actions to reduce peak load charges are scarce. These authors reviewed how autoregressive integrated moving average (ARIMA), support vector regression machine (SVRM), classification and regression trees (CART), artificial neural network (ANN), and multivariate adaptive regression splines (MARS) based models, among others, have been used to develop forecasting models for next-day building electric load and peak load. However, being able to predict the next day's electric load does not provide actionable intelligence to determine if the next day will contribute to peak load charges for the billing period.

Saxena et al. (2019) developed an ensemble machine-learning model focused specifically on predicting if the next day will be a PELD for a billing period. Saxena et al. tested their model using data from an educational consumer; the ensemble model predicted 70% of actual PELDs and revealed potential savings in the neighborhood of USD$80,000 after a yearlong testing period. This work provided evidence of how consumers could potentially reduce peak load charges by executing demand response actions based on the results of PELDs forecasting efforts.

As Saxena et al.'s (2019) methodology was being prepared for implementation at the educational consumer's campus; the campus' electricity infrastructure underwent a reconfiguration. The campus' electricity infrastructure was divided into two main circuits. Each of these main circuits now included a solar field designed to provide up to 2 megawatts (MW) of behind the meter renewable electricity generation (BTMREG). Even though Saxena et al.'s methodology had been validated for an electrical circuit without BTMREG at an educational consumer's campus, the methodology had not been tested for circuits with BTMREG able to make up for as much at 25% of the electric load.

Researchers had already noted that renewable electricity generation (REG) output is as variable as weather itself (Staffell and Pfenninger, 2018; Chaiamarit and Nuchprayoon, 2014). Aponte and McConky (2019) documented how REG output could represent a challenge for the accuracy of current PELDs forecasting methodologies. These findings will be described later in this section. REG output can fluctuate for periods ranging from minutes to hours to multiple days (Staffell and Pfenninger, 2018). Figure 2 illustrates the intermittency of solar-based REG at the educational consumer's campus for three non-consecutive days, each day with different predominant weather from 6:00 AM to 6:00 PM.

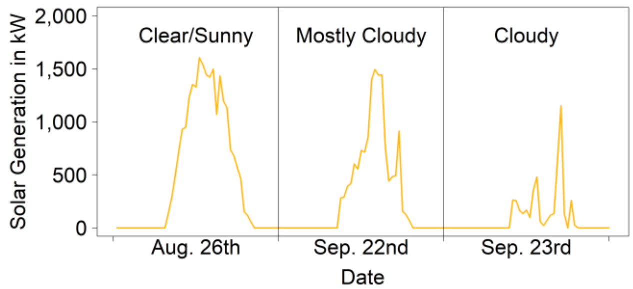

Fig. 2. Solar Generation during three days with different weather from 6:00 AM to 6:00 PM in 2019.

This intermittency of REG challenges the accuracy of both electric load forecasts (Tushar et al., 2018) and PELDs forecasts (Aponte and McConky, 2019). Chaiamarit and Nuchprayoon (2014) demonstrated that REG affects electric load characteristics and net demand. Net demand is defined as the result of subtracting the electricity generated behind the meter (on the consumer's side) from the total load required by the consumer. From this point on, whenever the term net demand is used, it will be referring to a scenario with BTMREG; and whenever the term demand is used, it will be referring to a scenario without BTMREG.

Aponte and McConky (2019) performed a data-driven analysis of a yearlong electric load and solar generation data for an educational consumer that highlighted five main findings. First, as expected the load values for the net demand scenario were lower than the load values for the demand scenario when the BTMREG was active. Second, the peak loads observed when BTMREG was present, happened during the hours when the BTMREG was either low or inactive and normal operations were still ongoing at the facilities. Third, as a direct consequence of the previous finding, demand response strategies need to be reevaluated to ensure that demand response actions can be performed during the new times with high concentration of peak loads.

Fourth, the number of PELDs during a month changed with the adoption of BTMREG. Consequently, the number of days during which demand response actions needed to be executed also changed with the adoption of BTMREG. Fifth, the adoption of BTMREG also changed the potential savings after executing demand response actions. The study concluded that new demand response strategies have to be developed as soon as facilities adopt BTMREG in order to ensure maximum reduction of peak demand charges.

A preliminary analysis of the electric load and solar generation data from an educational consumer's campus revealed that the presence of BTMREG increases the hour-to-hour net demand variability substantially. This net demand profile is the most important component of a consumer's electricity cost. Figure 3 illustrates the difference between the electric load to be forecasted when BTMREG is present (net demand) and the electric load to be forecasted in its absence (demand) for May 9th, 10th, and 11th, 2019.

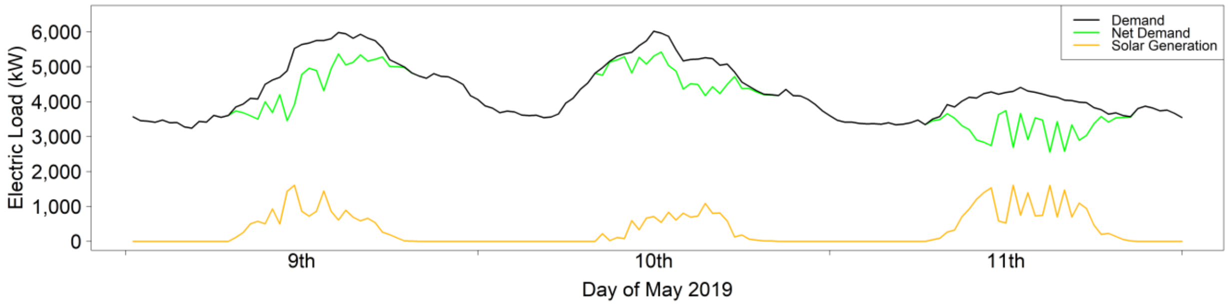

Fig. 3. Demand, net demand, and solar generation during May 9th-11th, 2019.

Figure 4(a) provides a closer view of May 9th 2019 as a sample case. This figure includes a smoothed load curve using a 2 points (One-Hour) moving average along with the corresponding Mean Absolute Percentage Deviation (MAPD) calculated according to Equation 1.

$$MAPD = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{|Actual_i - Smoothed\ Value_i|}{|Actual_i|}\right) \times 100 \qquad (1)$$

The higher MAPD value (2.2519 vs 1.3518) along with the noticeably worse fit illustrates how net demand exhibits a higher variability than demand. Figure 4(b) further supports this claim by illustrating how the hourly standard deviation tends to be higher for net demand during most of the hours with active BTMREG (6-19). In the absence of BTMREG, the load profile can be predicted using the consumer's past electric load data, weather and operations data, along with some minor influence from other factors. With the introduction of BTMREG, the influence of highly variable and difficult to predict weather conditions in the load profile is anticipated to make the forecasting process significantly more challenging. These initial findings, along with those documented by Aponte and McConky (2019), motivated a search for published research detailing accurate PELDs forecasting methodologies for facilities with BTMREG that revealed a lack of published literature on the topic.

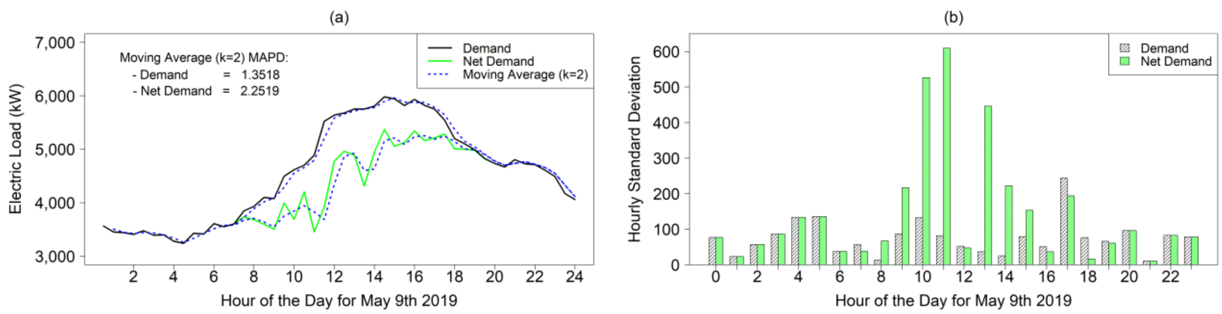

Fig. 4. (a) Demand, net demand, moving average and

(b) hourly standard deviation during May 9th, 2019.

2.3 Electricity Demand Threshold Value Forecasting

Some utilities provide signals to consumers in order to influence their consumption behavior by letting the consumer know when the electricity prices might be high because of peak demand (Dutta and Mitra, 2017). The expectation is that the consumer will respond to these signals by avoiding the use of non-essential loads thus contributing to a system-wide demand level reduction. However, in many pricing schemes that include peak load charges such as the one we described in Figure 1, these charges are determined at the end of the billing period. The consumer does not receive any signal from the utility during the billing period, and there is nothing that the consumer can do after the end of the billing period to avoid these charges. Therefore, it is up to these consumers to set their own signal, or electricity demand threshold, at the beginning of each month so that they can proactively trigger demand response actions and significantly reduce their peak load charges.

Figure 5 shows a consumer's demand profile for a billing period. Three days, 8, 12, and 18, can be identified as the days with the highest, or peak, electricity demand. At the end of the billing period, this consumer will be charged for the peak demand level represented by the red arrow. However, if the consumer proactively triggers demand response actions such as peak demand shaving during all of the days with peak consumption, and these actions effectively reduce the load to a more "normal" level such as the one represented by the blue arrow, then the consumer can avoid a significant amount of demand charges.

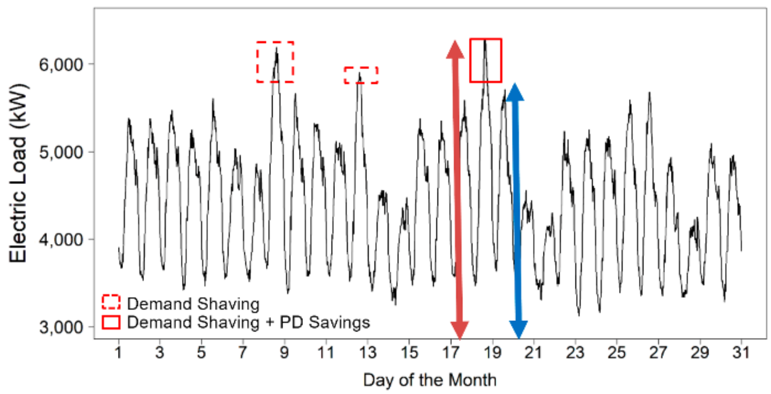

Fig. 5. Example of peak demand shaving.

The process of establishing an efficient electricity demand threshold right before the start of a billing period to proactively trigger cost saving demand response actions is challenging. For starters, the typical consumer does not have any information in regards to how his load profile will look during the upcoming billing period in order to determine what a "normal" demand level can look like. If the threshold is set higher than the optimal level, then the consumer will not be able to achieve all of the potential savings as illustrated by the red box with the solid outline in Figure 6(b). Alternatively, if the threshold is set below the optimal level, then the consumer could find himself in a situation where the number of demand response events necessary during the billing period will increase as well as user inconvenience. This scenario is illustrated in Figure 7 where the greater number of red boxes with dashed outlines in Figure 7(b) shows an increased number of demand response events needed. There is definitely the potential to achieve more savings under this scenario as illustrated by the red box with the solid outline in Figure 7(b). However, the additional savings might not be enough to offset the increased number of demand response events required and the user inconvenience that these events often produce.

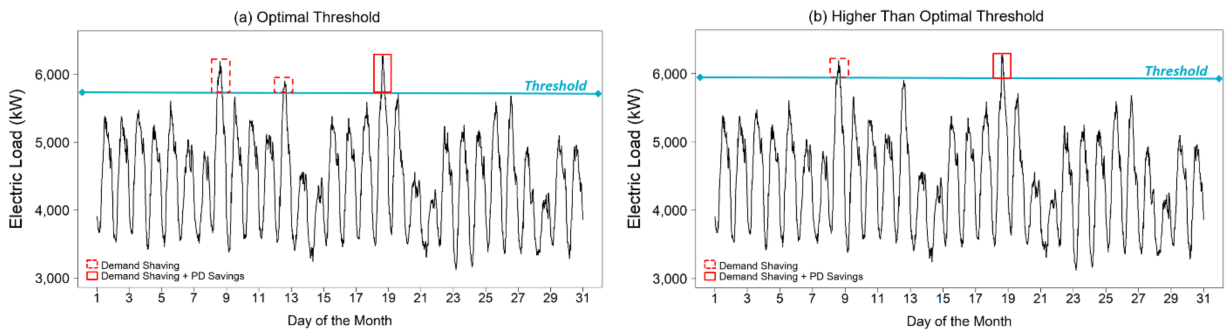

Fig. 6. Comparison between (a) an optimal threshold and (b) a higher than optimal threshold.

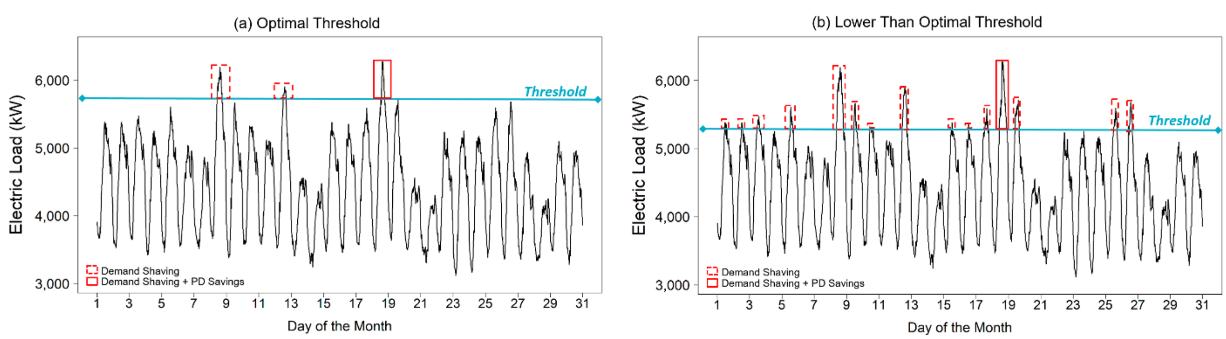

Fig. 7. Comparison between (a) an optimal threshold and (b) a lower than optimal threshold.

Researchers have placed significant emphasis on methodologies that focus on consumers reacting to time-based rate differentiation and price signals coming from the utility (Ganesan et al., 2022; Silva et al., 2020; Almahmoud et al., 2019; Li et al., 2018; Asadinejad & Tomsovic, 2017; Park et al., 2017; Siano, 2014). However, little attention has been paid to methodologies that would allow consumers under demand charges to proactively determine an appropriate electricity demand threshold value in order to trigger peak demand shaving and other demand response actions. As we mentioned earlier, peak demand shaving and other demand response actions are designed to minimize demand charges. However, these same actions often generate undesired inconvenience to users, such as sub optimal thermostat settings (Pi et al., 2021). The work of Saxena et al. (2019) detailed how both the demand charges and the need to perform demand response actions were minimized by performing these actions only when demand reached a set threshold.

Most of the published work on peak demand shaving and demand response actions on the consumer side focuses on scheduling algorithms that minimize the electric bill based on signals or information previously provided by the utility. Park et al. (2018) proposed a residential demand response methodology that achieved electricity cost reductions in the order of 10% while considering the user inconvenience. Asadinejad et al. (2017) developed a novel optimization model to determine an adequate threshold value to trigger incentive-based demand response actions. The model achieved monthly electricity costs savings of up to 10% for consumers and revenue increases of up to 50% for utilities when evaluated using utility level data collected from California, USA. Later on, Xu et al. (2019) developed an adaptive optimal monthly peak demand limiting strategy combining probabilistic demand profiles, weather forecasts, electricity charge tariffs, and measured electricity demand. The strategy successfully reduced the monthly peak demand and achieved considerable monthly net cost savings for an educational building in Hong Kong, China, by establishing adaptive optimal monthly thresholds. In addition, Pi et al. (2021) critically reviewed 15 published methodologies on peak demand shaving and 19 published methodologies on demand response. User satisfaction was identified in this work as a key element of successful demand response programs that has only been evaluated by a handful of studies.

2.4 Ensemble Peak Electric Load Days Forecasting

Alfares and Mohammad (2002) identified a trend in load forecasting research towards ensemble methods that combine two or more techniques. These authors reviewed more than 100 works about load forecasting published between a comprehensive review by Moghram and Rahman (1989) and February 2000. By 2014, research had shown that ensemble machine learning models often outperform the individual models that make them up (Fan et al., 2014). These models tend to deliver better generalization performance by integrating a number of base models to generate a final output. This integration can compensate for the individual imperfections of the base models. Fan et al. (2014) applied ensemble modeling to developed next-day energy consumption and peak power demand forecasts that achieved mean absolute percentage errors (MAPE) of 2.32% and 2.85% respectively for a building in Hong Kong, China. These results were better than the results of any of the eight individual models used to build the ensembles. Jingrui Xie's winning solution to the probabilistic load forecasting track of the Global Energy Forecasting Competition 2014 incorporated the use of simple arithmetic averaging to combine individual forecasts in order to enhance the accuracy of the final forecast (Xie and Hong, 2016).

Liu et al. (2017) applied quantile regression to produce probabilistic load forecasts based on an ensemble of point load forecasts. The results of the ensemble model showed higher levels of accuracy than those of the individual models based on several probabilistic scores. Li et al. (2018) developed an ensemble approach based on support vector regression (SVR) models to forecast short-term electric load. Li et al's approach was tested with utility level data from the Jiangxi Province in China and data from New South Wales Electric Utility in the United Kingdom. In both scenarios, the approach outperformed all of the non-SVR based models while achieving either better or similar results to those achieved by pure SVR based models. Lee et al. (2019) proposed a day-ahead electric load forecasting approach for a residential building using a stacking ensemble learning methodology. Lee et al.'s approach outperformed single models based on multiple linear regression, artificial neural network, and support vector regression, as well as other ensemble models such as gradient boosting machine, adaptive boosting, and extreme gradient boosting.

Within the realm of peak electric load forecasting, Saxena et al. (2019) presented a majority-classifier ensemble model that achieved the highest accuracy (86%), the lowest inaccuracy (14%) and the lowest percentage of false positives (9%) when compared to all of its individual base models at classifying upcoming days as either PELDs or Non-PELDs. Due to the novelty and relevance of Saxena et al.'s work, there is still an opportunity for published works evaluating the performance of additional ensemble methodologies for PELDs forecasting. A drawback of ensemble models based on a majority-vote as well as models based on simple arithmetic averaging is that the effect of all of the base models carries the same weight in determining the final answer. As will be described later in Sections 2.6 through 2.9, machine learning strategies such as classification and regression tree (CART), adaptive boosting (AdaBoost), and artificial neural network (ANN) models automatically assign weights to each input based on the inputs effect on the output's accuracy. The exploration of these ensembling options to potentially develop a better performing ensembling approach will not only improve Saxena et al.'s approach to classification-based PELDs forecasting, it will also ensure that the consumers are indeed achieving the best possible results in their forecasting efforts.

2.5 ARIMA Based Forecasting

Autoregressive integrated moving average (ARIMA) models are an adaptation of the Wiener filter, which was presented by the American mathematician Norbert Wiener first as part of a classified report during the Second World War in 1942, and later published as part of his book "Extrapolation, Interpolation, and Smoothing of Stationary Time Series" in 1949 (Wiener, 1949). The Wiener filter was used by the allies during the Second World War to predict the position of German bombers from radar reflections. British statisticians George Box and Gwilym Jenkins published systematic methods for applying the Wiener filter to business and economics data in their 1976 book titled "Time Series Analysis: Forecasting and Control" (Box and Jenkins, 1976). Since then, ARIMA models have also being referred to as Box-Jenkins models and have been widely used to model and forecast stationary and non-stationary data.

ARIMA forecasting models operate under the assumption that the future values to be forecasted are related to a finite combination of exponentially and non-exponentially weighted past disturbances. These models are applicable when evaluating time series with correlated observations and seasonal characteristics (Montgomery et al., 2015). ARIMA models do not take into account the effect of independent variables on the response, only the effects already embedded within the values in the time-series. A variation of ARIMA called autoregressive integrated moving average with exogenous variables (ARIMAX) has been developed in order to forecast future values taking into account the effect of independent variables. ARIMA based models have been good performers at forecasting future electric load values (Saxena et al., 2019; Yildiz et al., 2017; Hong and Fan, 2016; Alfares and Mohammad, 2002). However, Saxena et al. (2019) found that ARIMAX models significantly increased the computational complexity of forecasting future electric load values while failing to provide significantly better forecasts for a university campus. The work described in this manuscript will rely on arguably less computational intensive and more parsimonious models than ARIMAX such as classification and regression trees (CART), random classification and regression forest, and artificial neural network (ANN) in order to incorporate the effect of independent variables on the response.

Equation 2 illustrates Montgomery et al.'s (2015) mathematical representation of a generalized seasonal ARIMA model of orders (p,d,q) x (P,D,Q) with period *s*, also represented as ARIMA (p,d,q) x (P,D,Q)s.

$$\Phi*(B^s)\Phi(B)(1-B)^d(1-B^s)^D y_t = \delta + \Theta*(B^s)\Theta(B)\varepsilon_t \tag{2}$$

Where B is a backward shift operator such that $B(y_t) = y_{t-1}$,

s is the number of periods in a season,

$\Phi*(B^s) = (1-\Phi_1 B^s - \ldots - \Phi_p B^{sP})$ is a seasonal autoregressive operator of order P,

$\Phi(B) = (1-\Phi_1 B^s - \ldots - \Phi_p B^p)$ is a regular autoregressive operator of order p,

$(1-B)^d$ is the regular difference operator d of the time series y_t,

$(1 - B^s)^D$ is the seasonal difference operator D of the time series y_t,

δ is a constant term,

$\Theta{*}(B^s) = 1 - \Theta_1{*}(B^s) - \ldots - \Theta_Q{*}(B^{sQ})$ is the seasonal moving average operator of order Q,

$\Theta{*}(B) = 1 - \Theta_1{*}(B^s) - \ldots - \Theta_q{*}(B^q)$ is the regular moving average operator of order q, and

ε_t is a white noise process.

2.6 Classification and Regression Trees Based Forecasting

The use of classification and regression trees (CART) models to forecast future values taking into account the effect of independent variables was increased during the rise of the computer age after the publication of the book "Classification and Regression Trees" by Breiman et al. (1984) (Loh, 2014). These authors strengthened and extended the theta automatic interaction detection (THAID) methodology for classification trees presented by Messenger & Mandell (1972), and the automatic interaction detection (AID) methodology for regression trees presented by Morgan & Sonquist (1963) (Breiman et al., 1984). The manuscript by Morgan & Sonquist is considered the first regression tree algorithm published in the literature (Loh, 2014). Electric load forecasting techniques based on CART based models such as random forest have been used in the published literature to forecast future electric load values (Saxena et al., 2019; Yildiz et al., 2017). One of the most important reasons to include this type of models into the work described in this manuscript is that they are designed to complete a variable selection process. This process will be key to keep the techniques proposed by this manuscript on the more parsimonious side, by incorporating the least amount of independent variables, while also keeping the same techniques on the lower computational intensity side.

2.6.1 Classification Trees Based Forecasting

Classification trees are used to forecast a future discrete class based on independent variables. The process of developing a classification tree is formulated by James et al. (2013) as two main steps:

1) The predictor space (the set of possible values for the response y) gets divided into J distinct and non-overlapping high-dimensional regions based on the P amount of training observations $X_1, X_2, \ldots X_p$, such that the regions $R_1, R_2, \ldots, R_J$ minimize either:

    a. The classification error rate E given by Equation 3. If increased prediction accuracy is the main goal of the model.

    $$E = 1 - \max_{k}(\hat{p}_{mk}) \tag{3}$$

    Where $\hat{p}_{mk}$ is the proportion of training obervations in the j^{th} region that are from the k^{th} class.

    b. The Gini index G, is a measure of total variance across the K classes given by Equation 4. It is considered a measure of node purity. A small value of G indicates that most observations within a node belong to a single class k.

    $$G = \sum_{k=1}^{K} \hat{p}_{mk}\,(1 - \hat{p}_{mk}) \tag{4}$$

c. The cross-entropy index D as an alternative to the Gini index to measure the total variance across the K classes given by Equation 5. Just as with the Gini index, a small value of D indicates that most observations within a node belong to a single class k. G and D are actually very similar numerically.

$$D = -\sum_{k=1}^{K} \hat{p}_{mk} \log \hat{p}_{mk} \quad (5)$$

2) For any given test observation that falls into the region R_j, generate the same forecast value, which is the most commonly occurring class k among the training observations in the region R_j.

2.6.2 Regression Trees Based Forecasting

Regression trees are used to forecast a future quantitative value based on independent variables. Similar to classification trees, the process of developing a regression tree is also formulated by James et al. (2013) as two main steps:

1) The predictor space (the set of possible values for the response y) gets divided into J distinct and non-overlapping high-dimensional regions based on the P amount of training observations $X_1, X_2, \ldots X_p$, such that the regions $R_1, R_2, \ldots, R_J$ minimize the residual sum of squares (RSS) given by Equation 6.

$$\sum_{j=1}^{J} \sum_{i \in R_j} \left(y_i - \hat{y}_{R_j} \right)^2 \quad (6)$$

Where $\hat{y}_{R_j}$ is the mean response for the training observations within the j^{th} region.

2) For any given test observation that falls into the region R_j, generate the same forecast value, which is the mean of the response values for the training observations in the region R_j.

2.7 Random Classification and Regression Forest Based Forecasting

Classification and regression trees (CART) models are simple and useful for interpretation. However, they suffer from high variance and are likely to overfit the training data because of their selection of the strongest independent variable to produce each split (James et al., 2013). Random classification and regression forest models overcome this challenge by building a number of trees on randomly selected training samples. Additionally, each one of these trees is built by using a random selection of the available independent variables at each node to split. The final predicted value is the average of the predicted values of all trees built (James et al., 2013; Biau, 2012). In other words, random forest models force each split to consider only a subset of the independent variables. Therefore, some of the splits will not even consider the strongest predictor, and other predictors will have a chance to be selected. The average of the resulting trees that make up the random forest is less variable and hence more reliable because of the decorrelation of the trees that happens while developing the random forest (James et al., 2013; Biau, 2012).

Random forest models were first proposed by American statistician Leo Breiman in the 2000's (Biau, 2012). Breiman went on to develop these models after co-authoring the 1984 book "Classification and Regression Trees" (Breiman et al., 1984). The process of developing a random forest is formulated by James et al. (2013) as:

1) Generate B different training datasets by randomly splitting the original training dataset into smaller datasets.

2) Train the CART model on the b^{th} training dataset using a random sample of *m* independent variables from the full set of *p* independent variables as split candidates at each split considered while building the tree. The split is only allowed to use one of those *m* independent variables as split candidates and a new random selection of independent variables *m* is chosen at each split. Repeat $\forall b$ where $b=1, 2, \ldots, B$.

3) Obtain the prediction $\hat{y}^{*b}(x)$ corresponding to the CART model trained using the b^{th} training dataset. Repeat $\forall b$ where $b=1, 2, \ldots, B$.

4) Using Equation 7, calculate the random forest's final predicted value $\hat{y}(x)$ by determining the most occurring class (for classification cases) or averaging (for regression cases) all of the predictions $\hat{y}^{*b}(x)$ obtained across all of the *B* different training datasets.

$$\hat{y}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{y}^{*b}(x) \tag{7}$$

2.8 Artificial Neural Networks Based Forecasting

Artificial neural networks are multivariate statistical models structured to approach problems by mimicking the way a human brain would approach such problems (Montgomery et al., 2015). The origins of this technique can be traced back to the early 1940's when American neurophysiologist Warren McCulloch along with American logician Walter Pitts created a computational model for a neural network (Kleene, 1956). American psychologist Frank Rosenblatt developed the first algorithm for supervised learning of binary classifiers using a single layer of links, also known as the perceptron, in 1957 (Rosenblatt, 1957). More effective and accurate models integrating multiple layers of neurons were later developed by Ivakhnenko and Grigorʹevich Lapa in the late 1960’s (Ivakhnenko and Grigorʹevich Lapa, 1968).

The architecture of an ANN model consists of interconnected units known as neurons that are arranged in multiple layers in order to exchange information between each neuron. Each neuron in the ANN receives a weighted input that is processed through a transfer or activation function in order to generate a neuron output. The most basic architectures include three types of layers of neurons: the inputs layer, containing the independent variables or original predictors; the hidden layers, containing transformed variables; and the output layer, containing the responses (Montgomery et al., 2015; Yadav et al., 2015; Agatonovic-Kustrin and Beresford, 2000).

Montgomery et al. (2015) formulates a general ANN model as follows:

1. Let x_1, x_2, ..., x_p be the p independent variables or predictors to be incorporated in the model.

2. Let y be a response variable or future value to be predicted.

3. Let the value of each node a_u within each of the k hidden layers be a linear combination of its inputs according to Equation 8.

$$a_u = \sum_{j=1}^{p} w_{1ju} x_j + \theta_u \tag{8}$$

Where w_{1ju} are unknown parameters referred to as weights that must be estimated, and

θ_u is a parameter that plays the role of an intercept in linear regression and is sometimes referred to as the bias node.

4. Generate a transformed value z_u for each node a_u using a transfer or activation function $g(x)$ according to Equation 9.

$$z_u = g(a_u) \tag{9}$$

5. Use Equation 10 to form a linear combination b of the z_u inputs.

$$b = \sum_{u=1}^{k} w_{uev} z_u \tag{10}$$

6. Use Equation 11 to obtain the output response, also referred to as the response variable or future value to be forecasted y, which will be a transformation of b.

$$y = \tilde{g}(b) \tag{11}$$

Where $\tilde{g}(b)$ is the transfer or activation function for the output response, also referred to as the response variable or future value to be forecasted.

Figure 8 illustrates the architecture of an ANN model with three input variables (x_1, x_2, and x_3), one hidden layer containing four nodes, and a single output y as an example.

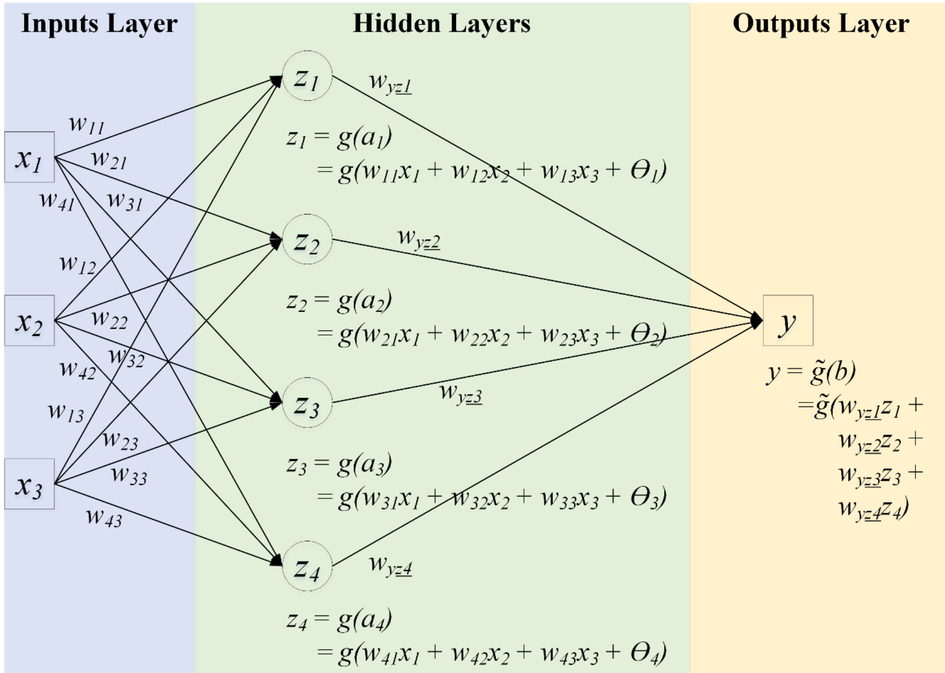

Fig. 8. Example of an ANN model architecture.

All ANN models evaluated in this study were generated using the function “nnet” from the R package “nnet” v7.3-14 (Ripley and Venables, 2022) with manually selected values for the required hyper parameters *size*, *decay*, and *maxit*. *Size* refers to the number of nodes in the single hidden layer. *Decay* specifies the value of the regularization parameter that determines the weight decay in each iteration during training in order to avoid over-fitting. *Maxit* establishes the maximum amount of iterations during training.

2.9 Adaptive Boosting Based Ensemble Forecasting

The adaptive boosting (AdaBoost) algorithm is an ensemble method used to improve the accuracy of a machine learning algorithm by overcoming overfitting and converting weak learners into strong learners (Freund and Schapire, 1999). A learner is considered "weak" or "strong" based on how correleated it is with the actual target variable. The methodology was developed in 1995 by Freund and Schapire (1997) as the first practical boosting algorithm and remains as one of the most widely used boosting options in several fields. Boosting models such as AdaBoost, use training samples to train one unit of a decision tree and pick over-weighted data as a replacement. Each decision tree generated learns from its predecessors and updates the residuals error. A weighted average of the estimates generated by these learners (trained decision trees) is used to produce a final prediction. Freund and Schapire (1999) developed the pseudocode in Figure 9 to describe their adaptive boosting algorithm AdaBoost.

Given: $(x_1, y_1), \ldots, (x_m, y_m)$ where $x_i \in X$, $y_i \in Y = \{-1, +1\}$
Initialize $D_1(i) = 1/m$.
For $t = 1, \ldots, T$:

- Train weak learner using distribution D_t.
- Get weak hypothesis $h_t : X \to \{-1, +1\}$ with error

$$\epsilon_t = \Pr_{i \sim D_t}\left[h_t(x_i) \neq y_i\right].$$

- Choose $\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t}\right)$.
- Update:

$$\begin{aligned} D_{t+1}(i) &= \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{if } h_t(x_i) \neq y_i \end{cases} \\ &= \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t} \end{aligned}$$

where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution).

Output the final hypothesis:

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right).$$

Fig. 9. Pseudocode for AdaBoost by Freund and Schapire (1999).

2.10 Summary of the State of the Art

Predicting future electric load forecasts have been a core activity for utilities since the electricity industry began in the late 1800's (Hong, 2014). Utilities rely on electric load forecasts to plan their supply and generating capacities (Dutta and Mitra, 2017; Hong and Fan, 2016; Alfares and Mohammad, 2002), to inform revenue projections, rate design, energy trading, and more (Hong and Fan, 2016). The capacity of electric utilities to ensure a reliable service to their clients depends heavily on these demand forecasts. Electric load forecasting methods have been extensively researched over the past few decades. The literature provides an ample range of studies featuring various methodologies and models for this purpose (Yildiz et al., 2017; Hong and Fan, 2016; Garulli et al., 2015; Alfares and Mohammad, 2002). Autoregressive integrated moving average (ARIMA), classification and regression trees (CART), random classification and regression forest, and artificial neural network (ANN) are among the most widely used models found within the electric load forecasting literature.

Most of the published research on electric load forecasting focuses on generating accurate electric load forecasts for both utilities and consumers, but there is limited research on the application of these forecasts to avoid the peak load charges described earlier in Section 1.3. Saxena et al. (2019) noted that studies focusing on forecasting a billing period's peak electric load days (PELDs) in order to trigger demand response actions to reduce peak load charges are scarce. Saxena et al. developed an ensemble machine-learning model focused specifically on predicting if the next day will be a PELD for a billing period. Saxena et al. tested their model using data from an educational consumer; the ensemble model predicted 70% of actual PELDs and revealed potential savings in the neighborhood of USD$80,000 after a yearlong testing period. This work provided evidence of how consumers could potentially reduce peak load charges by executing demand response actions based on the results of PELDs forecasting efforts.

During 2018, sustained ongoing efforts to decarbonize energy generation worldwide increased the share of renewable energy in global power capacity to 33% (REN21, 2019). Researchers have already noted that renewable electricity generation (REG) output is as variable as weather itself (Staffell and Pfenninger, 2018; Chaiamarit and Nuchprayoon, 2014). Aponte and McConky (2019) documented how REG output could represent a challenge for the accuracy of current PELDs forecasting methodologies. A preliminary analysis of the electric load and solar generation data from an educational consumer revealed that the presence of BTMREG increases the hour-to-hour net demand variability substantially. A search for published research detailing accurate PELDs forecasting methodologies for facilities with BTMREG revealed a lack of published literature on the topic.

Threshold-based peak electric load classification is one of the approaches for electric load classification currently available in the published literature (Saxena et al., 2019). Some utilities provide signals to consumers in order to influence their consumption behavior by letting the consumer know when the electricity prices might be high because of peak demand (Dutta and Mitra, 2017). The expectation is that the consumer will respond to these signals by avoiding the use of non-essential loads thus contributing to a system-wide demand level reduction. However, in many pricing schemes demand charges are determined at the end of the billing period based on each consumer's specific demand profile. Consumers under this type of pricing schemes do not receive any signal from the utility during the billing period, and there is nothing that the consumer can do after the end of the billing period to avoid these charges. The process of establishing an efficient electricity demand threshold right before the start of a billing period to proactively trigger cost saving demand response actions is challenging. Our survey of the published literature revealed that researchers have placed significant emphasis on methodologies that focus on consumers reacting to time-based rate differentiation and price signals coming from the utility (Ganesan et al., 2022; Silva et al., 2020; Almahmoud et al., 2019; Li et al., 2018; Asadinejad & Tomsovic, 2017; Park et al., 2017; Siano, 2014). However, little attention has been paid to methodologies that would allow consumers under demand charges to proactively determine an appropriate electricity demand threshold value in order to trigger peak demand shaving and other demand response actions even without receiving signals or information from the utility.

Alfares and Mohammad (2002) identified a trend in load forecasting research towards ensemble methods that combine two or more techniques. Research has shown that ensemble machine learning models often outperform the individual models that make them up (Lee et al., 2019; Saxena et al., 2019; Li et al., 2018; Ahmad et al., 2017; Liu et al., 2017; Xie and Hong, 2016; Fan et al., 2014). Ensemble models tend to deliver better generalization performance by integrating a number of base models to generate a final output. Within the realm of peak electric load forecasting, Saxena et al. (2019) presented a majority-classifier ensemble model that outperformed all of its individual base models at classifying upcoming days as either PELDs or Non-PELDs. However, this ensemble approach has not been contrasted with machine learning based alternatives in order to ensure that the consumers are indeed achieving the best possible results. The exploration of these ensembling options to potentially develop a better performing ensembling approach will not only improve Saxena et al.'s approach to classification-based PELDs forecasting, it will also ensure that the consumers are indeed achieving the best possible results in their forecasting efforts.

Chapter 3: Methodology

This chapter will present the three main methodologies developed as part of the research covered within this dissertation. First, the peak electric load days forecasting methodology developed for consumers regardless of the presence of behind the meter renewable generation (BTMREG). Second, the proposed methodology to forecast efficient electricity demand threshold values to trigger demand response actions. Third, the proposed methodology to use machine learning approaches to conduct ensemble based PELD forecasting.

3.1 Peak Electric Load Days Forecasting Methodology

This section will provide an overview of the methodology developed for the current study. The methodology developed to determine the performance of both load forecasting and PELDs forecasting methodologies was predominantly based on the previous work by Saxena et al. (2019). Saxena et al. established two general approaches for PELDs prediction. We will refer to the first approach as the threshold-based approach. This approach can be separated into two phases. During the first phase, regression-based load forecasting models are used to generate day ahead load forecasts. During the second phase, these forecasts are compared to a pre-determined monthly threshold (Dlim) in order to classify each day as either a PELD or a Non-PELD. The second approach will be referred to as the classification-based approach. For this approach, classification models are used to classify an upcoming day as either a PELD or a Non-PELD. Saxena et al. combined the results of some of the evaluated individual models into one ensemble approach and tested the methodology using electric load data from an electrical circuit without BTMREG at an educational consumer's campus. As an addition to the previous work by Saxena et al., the methodology for the current study includes an additional ensemble approach, it also includes the results of all individual models in both ensembles, and considers the presence of BTMREG by including electricity generation data (when applicable) and additional weather related features expected to affect REG. The proposed improved methodology is outlined in Figure 10.

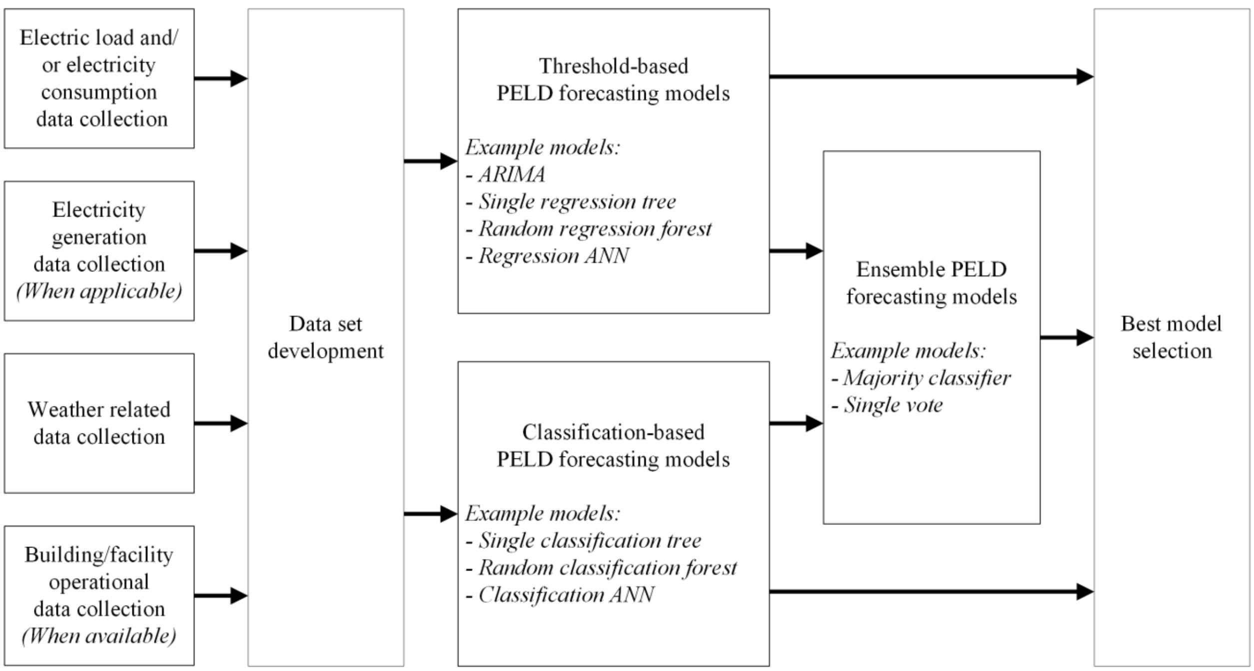

Fig. 10. PELDs forecasting methodology overview.

3.1.1 Methodology Overview

The methodology for the current study can be outlined in five phases. Data collection, dataset development, base machine learning models implementation within each of the two general approaches (threshold and classification based), ensemble models implementation using all of the base models as its components, and best PELDs forecasting model selection. This methodology was applied for a scenario with BTMREG and then repeated for a scenario without BTMREG. The methodologies tested were based on autoregressive integrated moving average (ARIMA), classification and regression trees (CART), random classification and regression forest, and artificial neural network (ANN) techniques. Details about the experimental implementation of this methodology will be provided in Sections 3.1.2 through 3.1.6.

3.1.2 Data Collection and Dataset Development

A dataset containing 29,952 records of electric load, electricity generation, weather, and operational data at 30 minute intervals was developed. The records in the dataset cover the period between June 16th, 2018 at 00:00 hours and February 29th, 2020 at 23:30 hours. All times are represented at the official local time of the educational consumer's campus. Table 2 provides a list of the 30 variables contained in the dataset along with each of the variable's description and type. Electric load and generation related data (measured in kW), such as that represented by variables 1, 4, and 8 in Table 2, was collected at 30 minute intervals from a smart metered circuit at an educational consumer's campus that included a solar field designed to provide up to 2 MW of BTMREG. Weather data was collected using hourly values from the publicly available local climatological data summaries corresponding to the airport weather station in closest proximity to the campus and provided by the National Oceanic and Atmospheric Administration (NOAA) of the USA (NOAA, 2020). This weather data was later imputed using linear interpolation for continuous variables and last value carried forward for categorical variables in order to generate a 30 minute intervals dataset. Building\facility operational data and calendar data, such as that represented by variables 9, 10, and 19 to 30 in Table 2, were collected from the campus' heating, ventilation, and air conditioning (HVAC) management system and the campus' academic calendar. The following data pre-processing steps were completed for the complete dataset as described by Saxena et al. (2019) in order to ensure the quality of the dataset: 1) uniformly-spaced time indices generation; 2) outlier detection and removal; and 3) missing value interpolation using linear interpolation for continuous variables and last value carried forward for categorical variables.

Table 2

Dataset variables.

Variable name	Description	Type
1) Demand	Load without BTMREG present (Demand) at the time of observation registered in kW	Continuous
2) DemDlim	Calculated monthly threshold (Dlim) for Demand as described by Saxena et al. (2019) in kW	Continuous
3) DemDmaxTm1	Maximum Demand registered during the previous day in kW	Continuous
4) NetDemand	Load with BTMREG present (Net Demand) at the time of observation registered in kW	Continuous
5) NetDemDlim	Calculated monthly threshold (Dlim) for Net Demand as described by Saxena et al. (2019) in kW	Continuous
6) NetDemDmaxTm1	Maximum Net Demand registered during the previous day in kW	Continuous
7) LastDemTM1	Last Demand registered during the previous day in kW (Demand and Net Demand are the same at this point because REG is not active during this time)	Continuous
8) SolarREG	Solar REG at the time of observation registered in kW	Continuous

9) OP_CoolReq	If at the time of observation, HVAC system cooling set point < indoor air temperature; Then, OP_CoolReq = Positive difference between HVAC system cooling set point and indoor air temperature in degrees Fahrenheit (°F); Else, OP_CoolReq = 0	Continuous
10) OP_HeatReq	If at the time of observation, HVAC system heating set point > indoor air temperature; Then, OP_HeatReq = Positive difference between HVAC system heating set point and indoor air temperature in degrees Fahrenheit (°F); Else, OP_HeatReq = 0	Continuous
11) NW_DBTemp	Outdoor dry bulb temperature at the time of observation in degrees Fahrenheit (°F) recorded by NOAA	Continuous
12) NW_RelHum	Outdoor relative humidity at the time of observation to the nearest whole percentage recorded by NOAA	Continuous
13) NW_WindSpe	Outdoor wind speed at the time of observation in miles per hour (mph) recorded by NOAA	Continuous
14) NW_WeatherClassShort	Outside weather classification at the time of observation recorded by NOAA and grouped into 5 categories. *Categories: 1 = Clear/Sunny , 2 = Cloudy, 3 = Rain, 4 = Snow, 5 = Windy*	Categorical
15) DemActPEL	If at the time of observation, Demand > DemDlim; Then, DemActPEL = 1; Else, DemActPEL = 0	Categorical
16) DemActPELD	Identification of the day as 1 for actual PELD or 0 for actual Non-PELD for the Demand data as described by Saxena et al. (2019)	Categorical
17) NetDemActPEL	If at the time of observation, NetDemand > NetDemDlim; Then, NetDemActPEL = 1; Else, NetDemActPEL = 0	Categorical
18) NetDemActPELD	Identification of the day as 1 for actual PELD or 0 for actual Non-PELD for the NetDemand data as described by Saxena et al. (2019)	Categorical
19) Time	Date (MM/DD/YYYY) and time (HH:MM) in 24 hours format at the time of observation	Categorical

20) Month	Month component of Time at the time of observation. *Categories: 1, 2, 3, …, 12*	Categorical
21) HoD	Hour component of Time at the time of observation. *Categories: 0, 1, 2, …, 23*	Categorical
22) DoW	Day of the week at the time of observation *Categories: 1 = Mon, 2 = Tue, …, 7 = Sun*	Categorical
23) OP_Semester	Academic semester at the time of observation *Categories: 1 = Fall, 2 = Spring, 3 = Summer*	Categorical
24) OP_Classes	If the day is an official class day; Then, OP_Classes = 1; Else, OP_Classes = 0	Categorical
25) OP_ResHallsOpen	If the on-campus residence halls are officially open during the day; Then, OP_ResHallsOpen = 1; Else, OP_ResHallsOpen = 0	Categorical
26) OP_CampusOpen	If the campus is officially open for administrative operations during the day; Then, OP_CampusOpen = 1; Else, OP_CampusOpen = 0	Categorical
27) OP_SpringBreak	If the day is part of spring break; Then, OP_SpringBreak = 1; Else, OP_SpringBreak = 0	Categorical
28) OP_FirstDayAfterBreak	If the day is the first after a break period; Then, OP_FirstDayAfterBreak = 1; Else, OP_FirstDayAfterBreak = 0	Categorical
29) OP_Increase	If there is an event during the day that can potentially cause an increase in electric load (festival, fair, convention, concert, etc); Then, OP_Increase = 1; Else, OP_Increase = 0	Categorical
30) OP_Decrease	If there is an event during the day that can potentially cause a decrease in electric load (holiday, half-day, exams week, etc); Then, OP_Decrease = 1; Else, OP_Decrease = 0	Categorical

The calculated monthly threshold (Dlim) values for the demand scenario (DemDlim) and the net demand scenario (NetDemDlim) included in the dataset, were determined using Equation 12 from the previous work by Saxena et al. (2019).

$$D_{lim,i} = \mu_i + 2\sigma_i \tag{12}$$

Where μ_i = the mean of every electric load observation at time interval t for the given month i, and

σ_i = the standard deviation of every electric load observation of the given month i.

3.1.3 Model Training, Validation, and Testing Process

Thirteen PELDs forecasting models were developed and tested for this study. The testing period selected for this study included 12 months from March 1st, 2019 at 00:00 hours to February 29th, 2020 at 23:30 hours. For each month *m* in the test period, a training data used was created using a random selection of 80% of all the available data in the dataset covering the period between June 16th, 2018 at 00:00 hours and the final day of the previous month, month *m-1*, at 23:30 hours. The remaining 20% of the data leading up to month *m* was used as validation dataset in order to optimize any model parameters. All final ANN models were selected based on their performance on the validation set. The parameters for each ANN model were optimized by testing the values specified in Table 3. After the parameter optimization process, the model for each month *m* was retrained using all of the training and validation data available prior to the start of month *m* before forecasting month *m* for testing purposes. This procedure was followed for all models, with the exception of the seasonal ARIMA model. Because of the continuity requirement of ARIMA based models, the Seasonal ARIMA model was retrained daily at the end of each day in the testing period at 23:59 hours using all of the available data from June 16th, 2018 at 00:00 hours up to the most recent record available before the retraining time.

Table 3

Values tested for ANN parameters.

Parameter	Values tested
# Hidden Nodes	2 to 30 by increments of 1
Decay Rate	0.0001, 0.001, 0.01, 0.05, and 0.1
# of Iterations	200 to 5000 by increments of 200

In order to test the models, all of the models were used at the end of each day at 23:59 hours to generate 48 predictions (each at 30 minute intervals) corresponding to the next day. Regression-based models generated a load prediction for each of the 48 time intervals. Classification-based models on the other hand, generated either a peak electric load (PEL) for the month or Non-PEL for the month class label for each of the 48 time intervals. A PEL for the month is defined as any load that is above the monthly threshold (Dlim) for the month (See variables 15 and 17 in Table 2). For final testing purposes during each month of the testing period, the load predictions generated by the regression-based models were compared to a monthly threshold (Dlim), any load found above this threshold was considered a PEL, and any day during which a PEL occurred was forecasted as a PELD. Similarly, the classification PEL and Non-PEL labels generated by the classification-based models were used to classify any day during which a PEL was forecasted as a PELD.

The process to generate the 48 predictions for October 5th, 2019 using any of the evaluated models except the seasonal ARIMA will be explained next as an example. On September 30th, 2019 at 23:59 hours, a new model to be used for the month of October 2019 is developed. The October 2019 model is initially trained using 80% of all the available data in the dataset covering the period between June 16th, 2018 at 00:00 hours through September 30th, 2019 at 23:30 hours, and validated using the remaining 20% of the data. Once optimal parameters are identified, the October 2019 model is retrained using all of the training and validation data available prior to September 30th, 2019. This October 2019 model is used to generate 48 predictions, one for each 30-minute interval starting with October 5th, 2019 at 00:00 hours and ending at 23:30 hours of the same day. In the case of the seasonal ARIMA model, the model available by October 4th, 2019 at 23:59 hours would have been trained using all of the available data between June 16th, 2018 at 00:00 hours up to the most recent record available before generating the first prediction for October 5th, 2019. The hyper parameters used to generate all seasonal ARIMA models for this study were determined using the "auto.arima" function from the R package "forecast" v8.12 (Hyndman and Khandakar, 2008) with a value s=336. All other parameters remained at their default value. The 48 predictions are then converted to a single PELD binary classification. If a PEL is found, then the day is forecasted as a PELD and this prediction is compared to the actual classification of the day in order to determine the model's performance. All the models evaluated during this research were implemented using the R language and environment for statistical computing (R Core Team, 2013). Information about specific R libraries and packages used to implement each model can be found on Tables 4, 7, 12, 13, and 17.

3.1.4 Threshold-based PELDs Forecasting Models

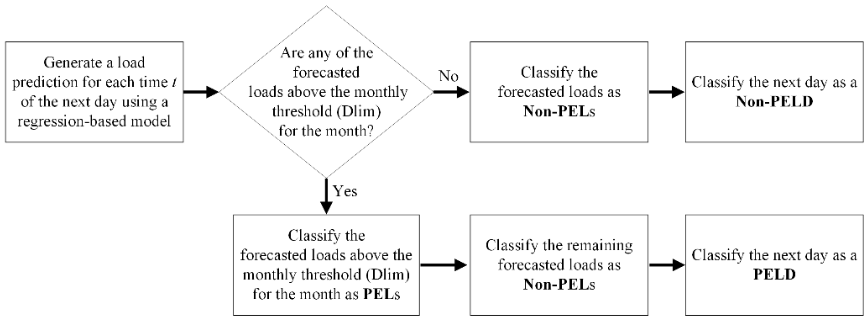

Fig. 11. Threshold-based PELDs forecasting process.

Five threshold-based PELDs forecasting models were developed for this study. These models were used to generate a day ahead load forecast that would later be compared to a pre-calculated monthly threshold (Dlim) in order to classify each day in the testing period as either a PELD or a Non-PELD (see Figure 11). Table 4 shows the characteristics of each of these models that were developed using the open source software R (R Core Team, 2013). Table 4 also includes details about the specific R libraries and packages used to develop each model. Table 5 shows the values used as the monthly threshold (Dlim) for each of the months in the testing period. These values were determined using Equation 12. The focus of the current study does not include evaluating the accuracy of the monthly threshold (Dlim) prediction method suggested by Saxena et al. (2019). For this reason, the current study used ground truth monthly thresholds (Dlim) for PELD determination.

Table 4

Threshold-based PELDs forecasting models characteristics.

Name	**Description**	**Response**	**Inputs used from Table 2**
M01_RegSARIMA	Seasonal ARIMA generated using the "auto.arima" function from the R package "forecast" v8.12 with a value s=336. All other parameters remained at their default value.	Electric load at the time of observation. (Variable 4 for electric load with BTMREG present (net demand) or Variable 1 for electric load without BTMREG present (demand) from Table 2)	Variable 4, for electric load with BTMREG present (net demand); Variable 1, for electric load without BTMREG present (demand).
M02_RegST	Regression single decision tree generated using the function "tree" from the R package "tree" v1.0-40 with default parameters.		Variables 2, 3, 5, 6, 7, 9:14, and 20:30
M03_RegRF	Regression random decision forest generated using the function "randomForest" from the R package "randomForest" v4.6-14 with values ntree=1000 and importance=TRUE. All other parameters remained at their default value.		
M04_RegANN	Regression feed-forward artificial neural network with a single hidden layer generated using the function "nnet" from the R package "nnet" v7.3-14 with manually selected values for size, decay, maxit, and MaxNWts, and linout = TRUE. All other parameters remained at their default value.		
M05_RegANNST	M04 but only using the variables selected by the regression single decision tree in M02 as inputs.		Variables selected by the regression single decision tree in M02.

Table 5

Monthly threshold (Dlim) values for the months in the testing period for the net demandand the demand scenarios.

Month and year	**Monthly threshold for net demand** (NetDemDlim)	**Monthly threshold for demand** (DemDlim)
Mar. 2019	5,277.62	5,584.79
Apr. 2019	5,442.61	5,759.57
May 2019	4,732.92	5,242.71
Jun. 2019	5,192.39	5,962.64
Jul. 2019	6,730.34	7,359.20
Aug. 2019	6,125.85	6,809.69
Sep. 2019	6,490.20	7,063.56
Oct. 2019	5,657.77	5,957.91
Nov. 2019	5,660.77	5,764.58
Dec. 2019	5,373.77	5,410.37
Jan. 2020	5,492.00	5,583.62
Feb. 2020	5,638.32	5,792.96

3.1.5 Classification-based PELDs Forecasting Models

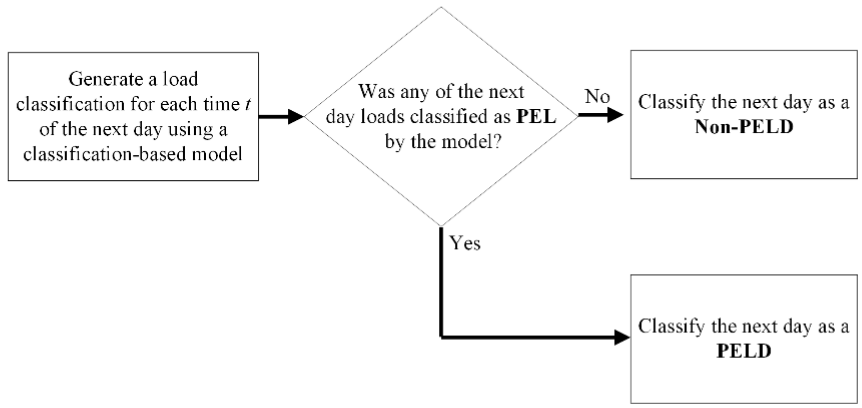

Fig. 12. Classification-based PELDs forecasting process.

Six classification-based PELDs forecasting models were developed to classify the electric load at time *t* as either a PEL for the month or not. Any day with a PEL present was automatically tagged as a PELD; otherwise, the day was classified as a Non-PELD (see Figure 12). Saxena et al. (2019) found a class imbalance while developing similar classification-based PELDs forecasting models for a circuit without BTMREG. The current study found similar class imbalances while evaluating circuits with and without BTMREG. Table 6 shows comparisons between the amount of PELs and Non-PELs, and PELDs and Non-PELDs to illustrate the class imbalance present in the complete dataset. After observing the class imbalance while developing the first two classification-based PELDs forecasting models (M06_ClassST and M07_ClassRF), the full training set (before splitting into the training and validation datasets) was balanced using the synthetic minority oversampling technique (SMOTE) developed by Chawla et al. (2002) and also applied by Saxena et al.. The SMOTE technique was applied using the function “SMOTE” from the R package “DMwR” v0.4.1 with default parameters. The remaining four classification-based PELDs forecasting models were developed using the balanced full training dataset. Table 7 shows the characteristics of each of the six classification-based PELDs forecasting models developed using the open source software R (R Core Team, 2013).

Table 6

Amount of PELs and Non-PELs, and PELDs and Non-PELDs for the net demand and the demand scenarios.

	Net demand	**Demand**
PELs	608	559
Non-PELs	29,344	29,393
Total observations	29,952	
PELs to non-PELs ratio	**1:48.26 (0.021)**	**1:52.58 (0.019)**
PELDs	85	65
Non-PELDs	539	559
Total observations	624	
PELDs to non-PELDs ratio	**1:6.34 (0.158)**	**1:8.6 (0.116)**

Table 7

Classification-based PELDs forecasting models characteristics.

Name	Description	Response	Inputs used from Table 2
M06_ClassST	Classification single decision tree generated using the function “tree” from the R package “tree” v1.0-40 with default parameters.	Is the electric load at time t a Peak Electric Load (PEL) for the month? 1 (Yes) | 0 (No)	Variables 2, 3, 5, 6, 7, 9:14, and 20:30
M07_ClassRF	Classification random decision forest generated using the function “randomForest” from the R package “randomForest” v4.6-14 with values ntree=1000 and importance=TRUE. All other parameters remained at their default value.		
M08_ClassSTwSMOTE	M06 trained and validated using the dataset balanced with the SMOTE technique.		
M09_ClassRFwSMOTE	M07 trained and validated using the dataset balanced with the SMOTE technique.		
M10_ClassANNwSMOTE	Classification feed-forward artificial neural network with a single hidden layer generated using the function “nnet” from the R package “nnet” v7.3-14 with manually selected values for size, decay, maxit, and MaxNWts, linout = FALSE, and softmax = TRUE. All other parameters remained at their default value. Trained and validated using the dataset balanced with the SMOTE technique.		
M11_ClassANNSTwSMOTE	M10 but only using the variables selected by the classification single decision tree in M08 as inputs.		Variables selected by the classification single decision tree in M08.

3.1.6 Ensemble PELDs Forecasting Models and Best Model Selection Process

Two ensemble PELDs forecasting models were developed by combining the results generated by all of the eleven base models evaluated in Sections 3.1.4 and 3.1.5 to classify each day in the testing period as either a PELD or a Non-PELD. These ensemble models were developed based on the ensemble model proposed by Saxena et al. (2019) to classify an upcoming day as either a PELD or a Non-PELD using demand data. The first ensemble model, E01_Majority, was a majority class classifier. This model follows the same ensemble approach proposed by Saxena et al. The majority class identifier can be represented mathematically using Equation 13.

$$C_j = \begin{cases} 1 \; if \; \sum_{i \in \mathrm{M}} X_{i,j} > \frac{|M|}{2} & \forall \, \mathrm{j} \in \mathrm{D} \\ 0 \; otherwise & \end{cases} \tag{13}$$

Where M: Set of base models used for day classification,

D: Set of days in the billing period,

|M|: Represents the cardinality of set M,

X*i*, *j*: Binary variable, takes a value of 1 when model *i* classifies day *j* in set D as a PELD, otherwise it takes a value of 0, and

C*j*: Returns the proposed ensemble model's forecasted classification for day *j* as a binary result of 1 for PELD or 0 for Non-PELD.

The second ensemble model, E02_SingleVote, was a single vote classifier. This model differs from the first ensemble model in that it only needs one the component models to classify a day as a PELD in order to classify the observed day as a PELD. This methodology was included in this study to account for the possibility of having PELDs that were only detected by a minority of the models because of certain special characteristics not noticeable by the majority of the base models in the ensemble.

The model results for all thirteen models (M01 through M11, E01, and E02) for each month of a 12 months testing period were compared to the actual values. The best model was selected based on the Total Score obtained by evaluating the ranked scorecard presented in Table 8. A higher Total Score identifies a better model.

Table 8

Ranked scorecard for selecting the best PELDs forecasting model.

Performance Measure (PM)	**Rank**	**Score** *(PM x Rank)*
False negatives during testing period *(FN / Number of days in testing period) * (-1)*	5	
False positives during testing period *(FP / Number of days in testing period) * (-1)*	4	
Model complexity *(Model complexity level * (-1))*	3	
Average monthly sensitivity during testing period	2	
Average monthly balanced accuracy during testing period	1	
TOTAL SCORE		

The elements of the ranked scorecard for selecting the best PELDs forecasting model in Table 8 are defined as follows:

FN = number of the false negatives i.e. amount of PELDs incorrectly predicted as non-PELDs,

FP = the false positives i.e. amount of non-PELDs incorrectly predicted as PELDs

The values for sensitivity were calculated according to Equation 14. This performance measure refers to the probability of a positive test, conditioned on the actual state of the instance being positive.

$$Sensitivity = \frac{TP}{TP+FN} \tag{14}$$

Where *TP* = the true positives i.e. amount of correctly predicted instances of PELDs; and *FN* has already being defined in this section.

The balanced accuracy values were calculated according to Equation 15. This performance measure refers to the probability of accurate positive and negative tests when one class appears much more than the other.

$$Balanced\ Accuracy = \frac{\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right)}{2} \tag{15}$$

Where *TN* = the True Negatives i.e. amount of correctly predicted instances of Non-PELDs; and *TP*, *FP, and FN* have already being defined in this section.

The values for *Rank* are consumer defined, based on how the consumer would like to prioritize the performance measures. The higher the value, the higher the importance of that specific performance measure. There should be no duplicate values for *Rank*. For this study it is suggested to assign the highest priority to the number of false negatives produced during the testing period by each model. A false negative means that the consumer will be billed demand charges that could have been avoided. The next priority suggested is the number of false positives produced during the testing period by each model. Even though these events do not produce demand charges, they generate unnecessary user inconvenience that can reduce productivity and negatively affect the work environment. Model complexity according to the values in Table 9 is recommended as the third most important performance measure in order to ensure the selection of the simplest yet also accurate model. The model complexity values in Table 9 are suggested by this study based on factors related to model implementation such as hyper parameter tuning, training time, and amount of model inputs. A *Rank* value of 2 is recommended for the average monthly sensitivity given that it is desirable to maximize the probability of a positive test, conditioned on the actual state of the instance being positive, over the remaining performance measure.

Using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve as an additional performance metric was considered but dismissed for the current study. The ROC curve is a probability curve that plots True Positive Rate against False Positive Rate at different threshold values. This is a very useful performance metric for situations where classification is more accurate if performed considering different classification thresholds. The AUC delivers a combined performance evaluation across all possible classification thresholds included in the ROC. At first glance, this might seem like a good performance metric to evaluate PELDs classification models given the fact that we have a threshold that varies from month to month. However, the proposed methodology only forecasts PELDs within the same month, which means that the threshold is constant for all classifications. Furthermore, given the significant financial impact of false negatives, the proposed methodology favors PELDs forecasting models that minimize this type of classification error over any other errors. One of the characteristics of the AUC-ROC curve optimization method is that it is classification-threshold-invariant. This means that it is designed to not favor any type of classification error over another. Nonetheless the use of some variation of the AUC-ROC curve to develop PELDs forecasting models that might eliminate the need to forecast a monthly threshold can be a very interesting avenue for future research.

Table 9

Suggested values for model complexity level.

Type of Model	**Model Complexity Level**
Base Model : Arithmetic	0.1
Base Model : Single Decision Tree	0.2
Base Model : Random Decision Forest	0.3
Base Model : ARIMA | Seasonal ARIMA | ARIMAX	0.4
Base Model : Artificial Neural Network (ANN)	0.5
Ensemble Model : Arithmetic	0.6
Ensemble Model : Single Decision Tree	0.7
Ensemble Model : Random Decision Forest	0.8
Ensemble Model : Adaptive Boosting (AdaBoost)	0.9
Ensemble Model : Artificial Neural Network (ANN)	1.0

3.2 Electricity Demand Threshold Value Forecasting Methodology

This section will provide an overview of the methodology developed for forecasting electricity demand threshold values. After collecting monthly data and developing a dataset, six base forecasting models were developed to forecast future demand threshold values. Using the results generated by the base models, seven ensemble forecasting models were developed. The results generated by all 13 models will be evaluated in order to determine the best model. The proposed methodology is outlined in Figure 13.

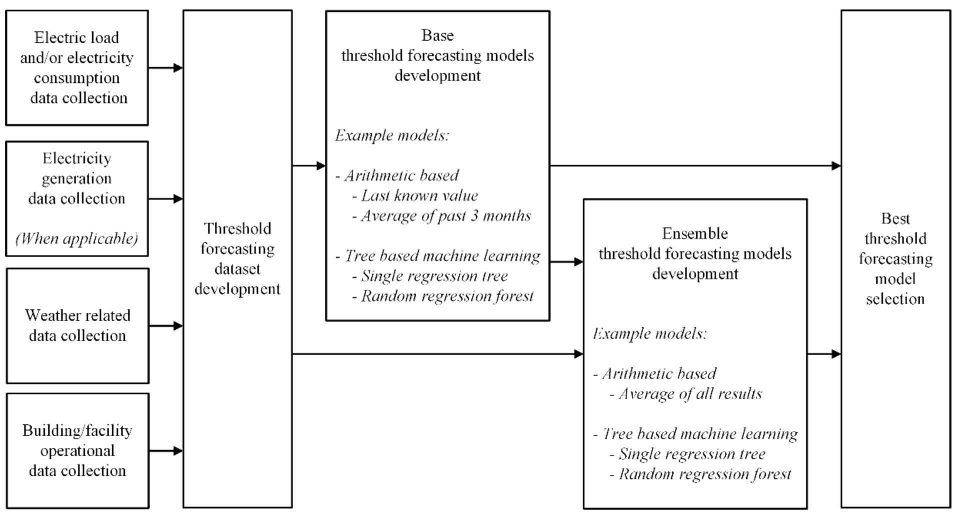

Fig. 13. Electricity demand threshold value forecasting methodology overview.

<u>*3.2.1 Methodology Overview*</u>

The methodology developed for forecasting electricity demand threshold values can be outlined in five phases (See Figure 13): data collection, threshold forecasting dataset development, base threshold forecasting models development, ensemble threshold forecasting models development, and best threshold forecasting model selection. During the first phase, data collection, hourly (or higher resolution) electric load and/or electricity consumption data as well as hourly (or higher resolution) electricity generation data (when applicable) and monthly operational data is collected from the building or facility of interest. Monthly weather data is obtained from a local weather data source such as the National Oceanic and Atmospheric Administration (NOAA) for buildings or facilities in the USA. The period covered by the data collected will depend on the availability of data about the consumer of interest. A period of at least 24 months is recommended. Section 3 provides details about the period of data collected and the data resolution for each of the consumers evaluated during this study. It is important to highlight that the period covered by all of the data collected should be the same for each type of data.

At the beginning of the second phase, threshold forecasting dataset development, the hourly (or higher resolution) electric load and/or electricity consumption data is used to calculate an actual electricity demand threshold value (Dlim) for each month. These monthly values will serve as ground truth for future model training and testing purposes. Equation 16, a slightly modified version of an original proposal by Saxena et al. (2019) is used to calculate the actual electricity demand threshold value (Dlim) for each month.

$$Dlim_i = \mu_i + \phi\sigma_i \tag{16}$$

Where

μ_i = the mean of every electric load observation for the given month *i*,

ϕ = consumer defined factor (2 for industrial and educational consumer | 3 for residential consumer), and

σ_i = the standard deviation of every electric load observation of the given month i.

The resulting threshold forecasting dataset will contain values for the actual electricity demand threshold value (Dlim), electricity consumption and generation related data, weather related data, and operational data all at a monthly resolution.

The third phase of the proposed methodology involves the development of base threshold forecasting models. These models can be arithmetic based such as a last known value model. This model will use the threshold value observed during the same month on the previous year or any other last known value for that month as the forecasted value for the month of interest. Another arithmetic based model can entail calculating the average of the threshold values observed during the previous three months and using this average as the forecasted value for the month of interest. Machine learning based models such as regression single decision tree and regression random decision forest can also be developed as base forecasting models to determine the threshold value using data collected from the threshold forecasting dataset as inputs.

Considering that many researchers agree that ensemble models often outperform the individual models that make them up, the fourth phase of the proposed methodology involves the development of ensemble threshold forecasting models. These models can also be arithmetic based such as an average of the results obtained from all models. This model will calculate the average of the values forecasted by all of the base models and use this average as the forecasted value for the month of interest. Machine learning based models such as regression single decision and regression random decision forest can also be developed as ensemble forecasting models to determine the threshold value using the values forecasted by all of the base models as inputs. These models could also use a combination of the values forecasted by all of the base models and data collected from the threshold forecasting dataset as inputs. The overall best performing threshold forecasting model will be selected during the fifth phase based on performance metrics such as the mean absolute percentage error (MAPE), the percentage of savings achievable by using the model, and the amount of user inconvenience. User inconvenience was measured in terms of the number of days during the testing period when peak demand shaving actions are unnecessarily triggered.

3.2.2 Data Collection and Dataset Development

Four threshold forecasting datasets, each containing 24 records of monthly actual electricity demand threshold values (Dlim), electricity consumption and generation (when applicable) related data, weather related data, and operational data were developed. Real electricity demand data from three different consumers, an industrial, an educational with behind the meter solar generation, and a residential were collected to develop the datasets. The amount of data available from the industrial consumer was enough to develop two different threshold forecasting datasets using the data collected from this consumer. The researchers decided to take advantage of this opportunity and develop a dataset to forecast the threshold values during the 12 months prior to the beginning of the COVID-19 pandemic in the USA (approximately March 2020). A second threshold forecasting dataset for the same consumer is used to forecast the values during the full first year of the pandemic. As part of the response to this pandemic near-global shutdowns occurred that expanded from weeks to moths. These shutdowns came accompanied with stay at home orders, curfews, and business disruptions that completely altered electricity consumption worldwide. Table 10 shows details about the location, data availability, presence of behind the meter renewable electricity generation (BTMREG), and availability of operational data for each of the threshold forecasting datasets. Appendix 1 provides a detailed list of the 62 variables considered for this study, each variable's description and type, and a checkbox identifying each dataset in which the variable is present.

Table 10

Details about each threshold forecasting dataset.

Threshold Forecasting Dataset	**Consumer Location**	**Period of Monthly Data Available**	**BTMREG Present**	**Operational Data Available**
Industrial_PRE (Pre-Pandemic)	Rochester, NY, USA Zip Code 14607	Mar. 2018 to Feb. 2020 24 Months	No	Calendar
Industrial_YR1 (Year 1 of Pandemic)	Rochester, NY, USA Zip Code 14607	Mar. 2019 to Feb. 2021 24 Months	No	Calendar
Educational	Rochester, NY, USA Zip Code 14623	Sep. 2019 to Aug. 2021 24 Months	Yes (Solar)	Calendar Open/Closed Days Special Events
Residential	Wethersfield, CT, USA Zip Code 06109	Dec. 2019 to Nov. 2021 24 Months	No	Calendar

Electric load and generation related data was collected at hourly intervals from each consumer's smart meter. Weather data was collected using monthly values from the publicly available local climatological data summaries corresponding to the weather station in closest proximity to each consumer and provided by the National Oceanic and Atmospheric Administration (NOAA) of the USA. Operational related data, such as calendar data, open/closed days (when available), and days with special events (when available), was collected from each consumer. The educational consumer's threshold forecasting dataset contained the most detailed operational data, which included days when classes, residence halls, and other specific areas of the campus were open/closed, also days with special events such as graduation, and calendar related data.

3.2.3 Model Training, Validation, and Testing Process

Thirteen threshold forecasting models were developed and tested for each of the four datasets generated for this study. Ten of these models were of the tree-based machine learning type, specifically either regression single decision tree based or regression random decision forest based. These two families of models were purposely selected to take advantage of their ability to clearly perform variable selection and provide insight into the effect of each variable selected, their auto-validation capabilities, and the reduced number of parameters required to setup as compared to other machine learning based models. These characteristics will be key to keep the methodology proposed on the more parsimonious side by incorporating the least number of independent variables in the models while also keeping the techniques on the low computational intensity side. In addition to the implementation complexity, computational costs, and unclear variable importance hierarchy of other more complex techniques such as artificial neural network, the researchers were also concerned about the risk of overfitting given the reduced number (24 records) of data points in each dataset. All of the training and validation data available prior to the start of month *m* was used for training and validation by the all of the tree-based machine learning models before forecasting a threshold value for month *m* for testing purposes. Table 11 provides details in regards to the training and validation period as well as the testing period for each of the threshold forecasting datasets.

Table 11

Training, validation, and testing periods for each threshold forecasting dataset.

Threshold Forecasting Dataset	**Training & Validation Period** *(12+ Months)*	**Testing Period** *(12 Months)*
Industrial_PRE *(Pre-pandemic)*	Mar. 2018 to Feb. 2019[1]	Mar. 2019 to Feb. 2020
Industrial_YR1 *(Year 1 of pandemic)*	Mar. 2019 to Feb. 2020[1]	Mar. 2020 to Feb. 2021
Educational	Sep. 2019 to Aug. 2020[1]	Sep. 2020 to Aug. 2021
Residential	Dec. 2019 to Nov. 2020[1]	Dec. 2020 to Nov. 2021

[1] The training and validation set increases by one month at the beginning of each new month in the testing period. For example, in order to test the month of Mar. 2019, the training and validation set contains all of the data points from Mar. 2018 to Feb. 2019 (12 months). However, in order to test the following month of Apr. 2019, the training and validation set increases by one month and contains all of the data points from Mar. 2018 to Mar. 2019 (13 months).

3.2.4 Base Threshold Value Forecasting Models

Six base threshold forecasting models were developed for this study. Table 12 shows implementation parameters, response, and inputs used for each model. The six models were used to generate a forecasted electricity demand threshold value. Models BM05 and BM06 are reduced versions of models BM03 and BM04 respectively. These reduced versions were obtained by first performing a manual feature selection applying a variation of the elbow method (first discussed by Thorndike (1953)) to the variable importance results generated by model BM04, specifically the percentage increase in mean standard error (%IncMSE), and selecting only the input variables before the elbow to include in each reduced model BM05 and BM06. The elbow method is the oldest and still state-of-the-art method to determine the potential optimal amount of clusters when performing cluster analysis (Shi et al., 2021; Fritz et al., 2020). The method has been adapted for other selection tasks by researchers. Similarly, the method has been adapted and used to perform manual variable selection for this research. The MAPE was calculated for each of these first reduced versions (BM05 and BM06) of the BM04 model. Using the variable importance results generated by the first BM06 model, the input variable with the lowest %IncMSE was removed, new versions of the reduced models (BM05 and BM06) were generated using the new list of input variables, and the MAPE was calculated for each of these new reduced versions. This process of removing one input variable at a time and calculating the MAPE was repeated until there were no more input variables to remove. Out of all the reduced versions for model BM05, the version with the lowest MAPE value was selected as the best reduced version for model BM05. The same process was completed for model BM06.

Table 12

Details about the development of each base threshold value forecasting model.

Name	**Type**	**Description**	**Response**	**Inputs used from Appendix 1**
BM01_LKV	Arithmetic based	Use the monthly threshold value observed during the same month on the previous year. If this value is not available, use the most recent known value for the month of interest. $Dlim_{i} = Dlim_{i-12periods}$ where i = period of interest expressed as Month.Year, and each period is one month. For example, the value for the forecasted threshold in Mar.2020 ($Dlim_{Mar2020}$) would be equal to the threshold value observed during Mar.2019 ($Dlim_{Mar2019}$). $Dlim_{Mar2020} = Dlim_{Mar2019}$	Forecasted electricity demand threshold value. (Variable 4 from Appendix 1)	Variable 5
BM02_AvPast3		Use the average of the past *n* observed threshold values. The value of *n* selected for this study was 3.		Previous 3 values of Variable 4
BM03_ST	Tree based machine learning	Regression single decision tree [1]		Varies by threshold forecasting dataset.[3]
BM04_RF1000		Regression random decision forest [2]		

BM05_ST_BestRed		Regression single decision tree [1]		Varies by threshold forecasting dataset.[3]
BM06_RF1000_ BestRed		Regression random decision forest [2]		

[1] Regression single decision tree generated using the function “tree” from the R software (R Core Team, 2013) package “tree” (Ripley, 2019) v1.0-40 with default parameters.

[2] Regression random decision forest generated using the function “randomForest” from the R software (R Core Team, 2013) package “randomForest” (Liaw and Wienner, 2002) v4.6-14 with values ntree=1000 and importance=TRUE. All other parameters remained at their default value.

[3] Please see Appendix 2 within the Supporting Information document accompanying this paper for details.

3.2.5 Ensemble Threshold Value Forecasting Models

Seven ensemble threshold forecasting models were developed for this study. Table 13 shows implementation parameters, response, and inputs used for each model. The seven models were used to generate a forecasted electricity demand threshold value. Models EM04 and EM05 are reduced versions of models EM02 and EM03 respectively. These reduced versions were obtained by following the same process previously described in Section 3.2.4 to obtain models BM05 and BM06. Models EM06 and EM07 only include a variable identifying the month and the results of all of the six base models as inputs.

Table 13

Details about the development of each ensemble threshold value forecasting model.

Name	Type	Description	Response	Inputs
EM01_Ave	Arithmetic based	Use the average of the values forecasted by all of the base models.	Forecasted electricity demand threshold value. (Variable 4 from Appendix 1)	Electricity demand threshold values forecasted by each of the 6 base models.
EM02_ST	Tree based machine learning	Regression single decision tree [1]		Varies by threshold forecasting dataset.[3]
EM03_RF1000		Regression random decision forest [2]		
EM04_ST_BestRed		Regression single decision tree [1]		Varies by threshold forecasting dataset.[3]
EM05_RF1000_BestRed		Regression random decision forest [2]		
EM06_ST_BM		Regression single decision tree [1]		Variable 3 from Appendix 1 (Month) and electricity demand threshold values forecasted by each of the 6 base models.[4]
EM07_RF1000_BM		Regression random decision forest [2]		

[1] Regression single decision tree generated using the function “tree” from the R software (R Core Team, 2013) package “tree” (Ripley, 2019) v1.0-40 with default parameters.

[2] Regression random decision forest generated using the function “randomForest” from the R software (R Core Team, 2013) package “randomForest” (Liaw and Wienner, 2002) v4.6-14 with values ntree=1000 and importance=TRUE. All other parameters remained at their default value.

[3] Please see Appendix 3 within the Supporting Information document accompanying this paper for details.

[4] Electricity demand threshold value forecasted by base model BM01 was not included for the residential consumer because of insufficient data.

3.2.6 Best Model Selection Process

The overall best performing threshold forecasting model was selected based on the Total Score obtained by evaluating the ranked scorecard presented in Table 14. A higher Total Score identifies a better model. The ranked scorecard evaluates 4 performance metrics. The first performance metric was the percentage of model savings achievable by using the threshold value forecasted by each model each month of the full testing period. This metric was calculated using the method proposed by Aponte et al. (2020) to calculate potential savings. Potential savings in kW after executing demand response actions for each month were determined according to Equation 17.

$$Potential\ savings\ in\ kw = HPEL - Dlim \tag{17}$$

Where

HPEL = highest peak electric load of the month in kW, and

Dlim = monthly threshold established for the month.

This methodology assumes that all peak loads predicted in the month are reduced to the level of the monthly threshold (Dlim) established for the month. The potential savings using the actual threshold were calculated first and then the same process was repeated using the threshold value forecasted by each model each month. For each month in the testing period, out of the potential savings calculated using the actual threshold value, those achieved by using the threshold value forecasted by each model were calculated. Totals for all savings during the testing period calculated using actual threshold values and for all model savings during the same period were calculated in order to determine the percentage of model savings. Higher values were considered as indicators of better performance.

Table 14

Ranked scorecard for selecting the best threshold forecasting model.

Performance Measure (PM)	**Rank**	**Score** *(PM x Rank)*
Percentage of model savings during testing period *Expressed in decimal form*	4	
User inconvenience during testing period *(FP / Number of days in testing period) * (-1)*	3	
Model complexity *(Model complexity level * (-1))*	2	
MAPE *(MAPE during testing period) * (-1)*	1	
TOTAL SCORE		

The second performance metric calculated was the amount of unnecessary user inconvenience produced by each model in terms of the number of days during the testing period when peak demand shaving actions were unnecessarily triggered. Similarly to what was previously illustrated in Figures 6 and 7, the electricity demand profile for each month was used along with the actual threshold value to identify true peaks. The results were then compared to those obtained by repeating the same procedure but using the threshold values forecasted by each model. All of the days identified by using the forecasted threshold values as having peaks that were not also identified using the actual threshold values as having peaks were considered days when peak demand shaving actions were unnecessarily triggered. This is the same procedure described previously in Section 3.1.6 for the false positives performance metric.

The third performance metric, model complexity, was calculated using the same procedures described previously in Section 3.1.6 to calculate the same performance metric. The final performance metric was the mean absolute percentage error (MAPE) achieved by each model during the full testing period. The values for MAPE were calculated according to Equation 18 and the lowest values were considered as indicators of better performance.

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{|Actual_i - Forecast_i|}{|Actual_i|}\right) \times 100 \tag{18}$$

3.3 Ensemble Peak Electric Load Days Forecasting Methodology

This section will provide an overview of the methodology developed to test machine learning based approaches to for ensemble peak electric load days forecasting. Six ensemble PELDs forecasting models were developed by combining the results generated by models M01 through M05 (previously described in Section 3.1.4) and those generated by models M08 through M11 (previously described in Section 3.1.5). The results generated by models M06 and M07 (previously described in Section 3.1.5) were not included in any of the ensemble models because of the poor preliminary results for unbalanced datasets to be described in Section 4.1.2. The ensemble models were developed based on the ensemble model proposed by Saxena et al. (2019) to classify an upcoming day as either a PELD or a Non-PELD. The results generated by all base and ensemble models will be evaluated in order to determine the best model. The proposed methodology is outlined in Figure 14.

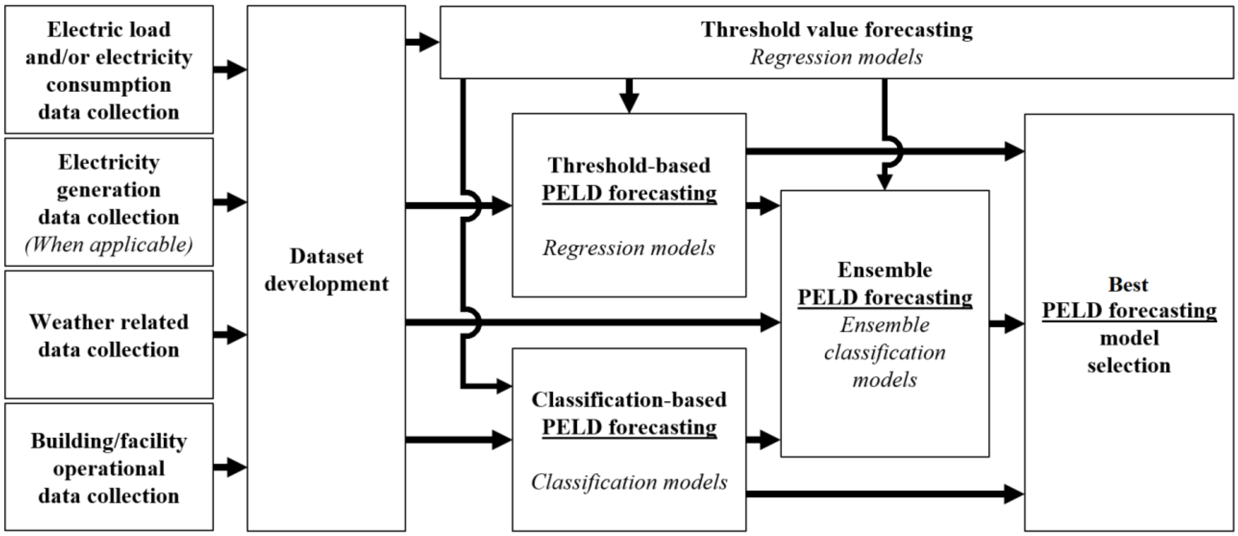

Fig. 14. Ensemble PELDs forecasting methodology overview.

3.3.1 Methodology Overview

The methodology developed for ensemble peak electric load days forecasting can be outlined in five phases. Data collection, dataset development, base models implementation, ensemble models implementation using the results generated by all of the base models as its components, and best overall forecasting model selection. The best performing model was selected based on the scorecard combining Sensitivity, Balanced Accuracy, number of False Negatives, and number of False Positives values as previously described in Section 3.1.6. Real electricity demand data from an industrial, an educational with behind the meter solar generation, and a residential consumer was collected to test the methodology. Details about the experimental implementation of this methodology will be provided in Section 4.3.

3.3.2 Data Collection and Dataset Development

Three ensemble PELD forecasting datasets, two containing 2 years of data and one containing 1 year and 9 months of data on daily actual electricity demand threshold values (Dlim), electricity consumption and generation (when applicable) related data, weather related data, and operational data were developed. Real electricity demand data from an industrial, an educational with behind the meter solar generation, and a residential consumer was collected to develop the datasets. Each dataset contains the results generated by models M01 through M05 (previously described in Section 3.1.4) as well as those generated by models M08 through M11 (previously described in Section 3.1.5). In addition, weather and building/facility operational data (when available) was also included. Table 15 shows details about the location, data availability, and availability of operational data for each dataset. Appendix 4 provides a detailed list of 62 variables, each variable's description and type, and a checkbox identifying each dataset in which the variable is present.

Weather data was collected from the publicly available local climatological data summaries corresponding to the airport weather station in closest proximity to the consumer and provided by the National Oceanic and Atmospheric Administration (NOAA) of the USA. When necessary, the weather data was imputed using linear interpolation for continuous variables and last value carried forward for categorical variables. Calendar data was constructed from the date and time of each reading in the dataset and building/facility operational data was collected from the consumer when available. The following data pre-processing steps were completed for the complete datasets as described by Saxena et al. (2019) in order to ensure the quality of the dataset: 1) uniformly-spaced time indices generation; 2) outlier detection and removal; and 3) missing value interpolation using linear interpolation for continuous variables and last value carried forward for categorical variables.

Table 15

Details about each ensemble PELD forecasting dataset.

Ensemble PELD Forecasting Dataset	**Consumer Location**	**Period of Daily Data Available**	**BTMREG Present**	**Operational Data Available**
Industrial	Rochester, NY, USA Zip Code 14607	May. 2020 to Apr. 2022 2 years	No	Calendar
Educational	Rochester, NY, USA Zip Code 14607	Jun. 2018 to Feb. 2020 1 year + 9 months	Yes	Calendar Events
Residential	Wethersfield, CT, USA Zip Code 06109	May. 2020 to Apr. 2022 2 years	No	Calendar

3.3.3 Model Training, Validation, and Testing Process

Six ensemble PELDs forecasting models were developed and tested for this study. The testing period selected for this study included 12 months. The first two models (E01 and E02) followed the same methodology previously described in Section 3.1.6 without including models M06 and

M07 as previously explained at the beginning of Section 3.1. The next four models were classification single tree, classification random forest, AdaBoost, and single layer classification artificial neural network respectively. All of these six models combined the results generated by models M01 through M05 (previously described in Section 3.1.4) and those generated by models M08 through M11 (previously described in Section 3.1.5) in order to generate their own classification.

Because of the imbalanced nature of the dataset previously explained in Section 3.1.5, all of the training and validation datasets were balanced using the SMOTE technique previously described in Section 3.1.5. All of the training and validation data available prior to the start of month *m* was used for training and validation by the all of the models before the start of a new month *m* for testing purposes. Table 16 provides details in regards to the training and validation period as well as the testing period for each of the threshold forecasting datasets.

Table 16

Training, validation, and testing periods for each threshold forecasting dataset.

Threshold Forecasting Dataset	**Training & Validation Period** *(9+ Months)*	**Testing Period** *(12 Months)*
Industrial	May 2020 to Apr. 2021[1]	May 2021 to Apr. 2022
Residential		
Educational	Jun 2018 to Feb. 2019[1]	Mar 2019 to Feb. 2020

[1] The training and validation set increases by one month at the beginning of each new month in the testing period. For example, in order to test the month of Mar. 2019, the training and validation set contains all of the data points from Mar. 2018 to Feb. 2019 (12 months). However, in order to test the following month of Apr. 2019, the training and validation set increases by one month and contains all of the data points from Mar. 2018 to Mar. 2019 (13 months).

3.3.4 Model Implementation and Best Model Selection Process

Six ensemble PELD forecasting models were developed for this study. Table 17 shows implementation parameters, response, and inputs used for each model. The six models were used to generate a classification of an upcoming day as either a PELD or a Non-PELD. The model results for each month of a 12 months testing period were compared to the actual values and the best performing model will be selected based on the same ranked scorecard previously described in Section 3.1.6.

Table 17

Details about the development of each ensemble threshold value forecasting model.

Name	**Type**	**Description**	**Response**	**Inputs**
E01_Majority	Arithmetic based	Majority classifier previously described in Section 3.1.6.	0 – The day will not be a PELD 1 – The day will be a PELD (Variable 1 from Appendix 4)	Please see Appendix 4 for details.
E02_SingleVote		Single vote classifier previously described in Section 3.1.6.		
E03_ST	Machine learning based	Regression single decision tree [1]		
E4_RF1000		Regression random decision forest [2]		
E5_ANN		Single layer artificial neural network [3]		
E6_AdaBoost		Adaptive boosting model [4]		

[1] Regression single decision tree generated using the function “tree” from the R software (R Core Team, 2013) package “tree” (Ripley, 2019) v1.0-40 with default parameters.

[2] Regression random decision forest generated using the function “randomForest” from the R software (R Core Team, 2013) package “randomForest” (Liaw and Wienner, 2002) v4.6-14 with values ntree=1000 and importance=TRUE. All other parameters remained at their default value.

[3] Single layer artificial neural network generated using the function "nnet" from the R software (R Core Team, 2013) package "nnet" (Ripley and Venables, 2022) v7.3-17 with parameters optimized as previously described in Section 3.1.3. All other parameters remained at their default value.

[4] Adaptive boosting model generated using the function "adaboost" from the R software (R Core Team, 2013) package "fastAdaboost" (Chatterjee, 2016) v1.0-0 with value nIter=100. All other parameters remained at their default value.

Chapter 4: Experimental Results and Discussion

This chapter will present the results of the experimentation detailed in the previous Chapter 3 in order to investigate the three main research questions motivating this dissertation. These research questions were previously defined in Section 1.7. Section 4.1 will present the results related to peak electric load days forecasting for consumers with and without BTMREG following the methodology previously described in Section 3.1. Section 4.2 will present the results related to electricity demand threshold value forecasting following the methodology previously described in Section 3.2. The final Section 4.3 will present the results related to ensemble peak electric load days forecasting following the methodology previously described in Section 3.3. Each section will include a discussion of the results presented.

4.1 Peak Electric Load Days Forecasting Results and Discussion

4.1.1 Threshold-based PELDs Forecasting Results

Figure 15 shows the monthly mean absolute percentage error (MAPE) achieved by the five threshold-based PELDs forecasting models previously described in Section 3.1.4 during their regression-based load forecasting stage. The MAPE values are presented for both the net demand (see Figure 15.a) and the demand (see Figure 15.b) scenarios. The values in Figure 15 show how most of the evaluated models achieved better electric load forecasting performance (lower MAPE values) when applied to a scenario without BTMREG instead of a scenario with BTMREG. A paired T-test for a mean difference equal to 0 (vs $\neq 0$) was performed for each of the five models using the monthly MAPE values for both scenarios (net demand and demand). The null hypothesis was rejected for all models with a 95% of confidence demonstrating that the results between the scenarios show a statistical significant difference. Detailed results of these paired T-tests are available in Appendix 5. The results of these paired T-tests demonstrate that it is more challenging for the regression-based electric load forecasting models evaluated to achieve high performance levels when BTMREG is present. Table 18 illustrates how the M03_RegRF and M05_RegANNST models outperformed the remaining two models at achieving the lowest average monthly MAPE values for both the net demand and the demand scenarios.

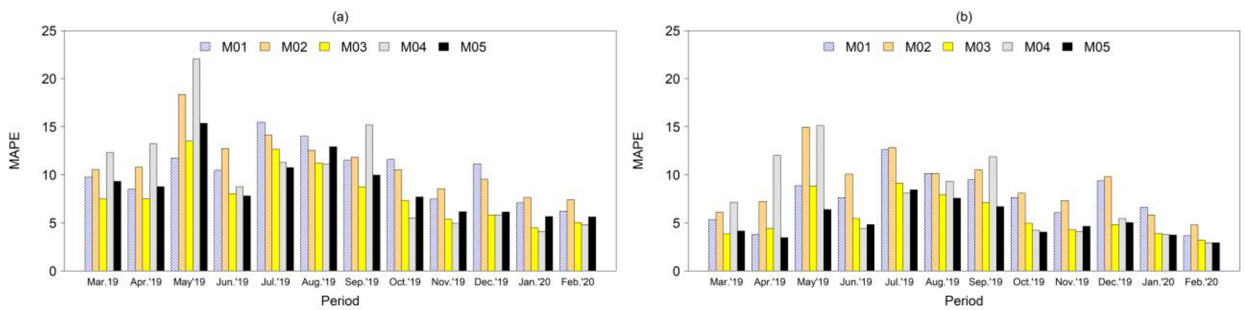

Fig. 15. MAPE achieved by the M01 to M05 models for the (a) net demand and the (b) demand scenarios.

Table 18

Average monthly MAPE values achieved by the M01 to M05 models for the net demand and the demand scenarios.

Model	**Average monthly MAPE**	
	Net demand	**Demand**
M01_RegSARIMA	10.4121	7.6004
M02_RegST	11.2159	8.9645
M03_RegRF	8.2921	5.6494
M04_RegANN	9.9221	7.3715
M05_RegANNST	8.8404	5.1544

Figure 16 shows the monthly values for sensitivity achieved by the five threshold-based PELDs forecasting models during their threshold-based PELDs classification stage. The sensitivity values are presented for both the net demand (see Figure 16.a) and the demand (see Figure 16.b) scenarios. The November 2019, January 2020, and February 2020 periods are not shown in Figure 16 because there were no PELD occurrences during these periods. Paired T-tests for a mean difference equal to 0 (vs ≠ 0) were performed for both the sensitivity and the balanced accuracy monthly results of each of the five models for both scenarios (net demand and demand). Detailed results of these paired T-tests are available in Appendix 5. The null hypothesis failed to be rejected for all tests except the test comparing the balanced accuracy obtained by model M05_RegANNST with a 95% of confidence. These results demonstrate that the sensitivity values for all five models and the balanced accuracy values for four of the models do not show a statistical significant difference caused by the presence of BTMREG. These results suggest that the performance of the five threshold-based PELD forecasting models at the classification stage is statistically the same regardless of the presence of BTMREG, which means that the PELD forecasting methodology developed is equally effective for both scenarios.

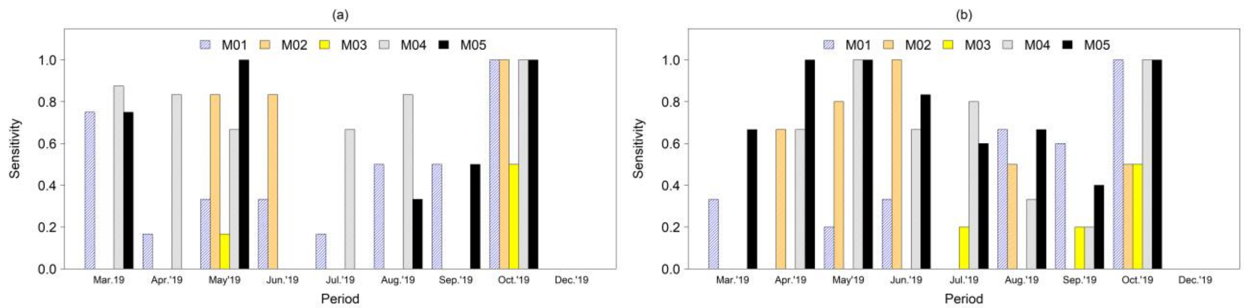

Fig. 16. Sensitivity achieved by the M01 to M05 models for the (a) net demand and the (b) demand scenarios.

Table 19 shows the average monthly sensitivity and balanced accuracy values achieved by the threshold-based PELDs forecasting models during their threshold-based PELDs classification stage. This table illustrates how the ANN-based models M04_RegANN and M05_RegANNST outperformed the remaining models at achieving the highest values for average monthly sensitivity and balanced accuracy for both the net demand and the demand scenarios.

Table 19

Average monthly sensitivity and balanced accuracy values achieved by the M01 to M05 models for the net demand and the demand scenarios.

Model	**Average monthly sensitivity**		**Average monthly balanced accuracy**	
	Net demand	**Demand**	**Net demand**	**Demand**
M01_RegSARIMA	0.4167	0.3481	0.6294	0.6102
M02_RegST	0.2963	0.3852	0.6258	0.6508
M03_RegRF	0.0741	0.1000	0.5370	0.5500
M04_RegANN	0.5417	0.5185	0.6953	0.6963
M05_RegANNST	0.3981	0.6852	0.6652	0.8236

4.1.2 Classification-based PELDs Forecasting Results

Figure 17 shows the monthly values for sensitivity achieved by the six classification-based PELDs forecasting models previously described in Section 3.1.5. The sensitivity values are presented for both the net demand (see Figure 17.a) and the demand (see Figure 17.b) scenarios. The November 2019, January 2020, and February 2020 periods are not shown in Figure 17 because there were no PELD occurrences during these periods. These results demonstrate how the class imbalance issue described in Section 3.1.5 needs to be addressed in order to achieve the best sensitivity values when implementing the classification-based PELDs forecasting approach regardless of the presence or absence of BTMREG. Figure 17 shows how the models using a balanced training and validation dataset overwhelmingly outperformed those obtained when using the original datasets during eight or more of the months in the testing period for the net demand (see Figure 17.a) and the demand (see Figure 17.b) scenarios.

Paired T-tests for a mean difference equal to 0 (vs ≠ 0) were performed for both the sensitivity and the balanced accuracy monthly results of each of the six models for both scenarios (net demand and demand). Detailed results of these paired T-tests are available in Appendix 5. The null hypothesis failed to be rejected for all tests with a 95% of confidence. These results demonstrate that the sensitivity and balanced accuracy values for all six models do not show a statistical significant difference caused by the presence of BTMREG. These results suggest that the performance of the six classification-based PELD forecasting models developed is statistically the same regardless of the presence of BTMREG, which means that the PELD forecasting methodology developed is equally effective for both scenarios.

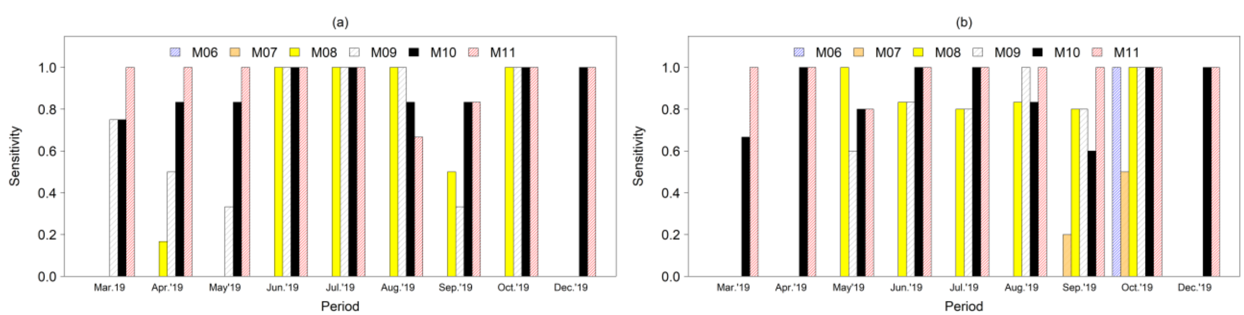

Fig. 17. Sensitivity achieved by the M06 to M11 models for the (a) net demand and the (b) demand scenarios.

Table 20 shows the average monthly sensitivity and balanced accuracy values achieved by the classification-based PELDs forecasting models. This table illustrates how the ANN-based models M10_ClassANNwSMOTE and M11_ClassANNSTwSMOTE outperformed the remaining models at achieving the highest values for average monthly sensitivity and balanced accuracy for both the net demand and the demand scenarios. The values in this table do not provide any clear evidence of a reduction in the performance level of the classification-based models caused by the presence of BTMREG.

Table 20

Average monthly sensitivity and balanced accuracy values achieved by the M06 to M11 models for the net demand and the demand scenarios.

Model	**Average monthly sensitivity**		**Average monthly balanced accuracy**	
	Net demand	**Demand**	**Net demand**	**Demand**
M06_ClassST	0.0000	0.1111	0.5000	0.5517
M07_ClassRF	0.0000	0.0778	0.5000	0.5389
M08_ClassSTwSMOTE	0.5185	0.5852	0.6315	0.6770
M09_ClassRFwSMOTE	0.6574	0.5593	0.7211	0.7055
M10_ClassANNwSMOTE	0.8981	0.8778	0.8329	0.8430
M11_ClassANNSTwSMOTE	0.9444	0.9778	0.8913	0.8906

4.1.3 Ensemble PELDs Forecasting Results and Best Model Selection

Figure 18 shows the monthly values for sensitivity achieved by the two ensemble PELDs forecasting models previously described in Section 3.1.6. This figure also includes the monthly values for sensitivity achieved by the M11_ClassANNSTwSMOTE model. This model achieved the best average monthly sensitivity and balanced accuracy values out of the eleven base models evaluated for both the net demand and the demand scenarios (see Tables 19 and 20). Figure 18 shows the results for both the net demand (see Figure 18.a) and the demand (see Figure 18.b) scenarios. The November 2019, January 2020, and February 2020 periods are not shown in Figure 18 because there were no PELD occurrences during these periods. These results show how both the proposed E02_SingleVote model and the M11_ClassANNSTwSMOTE model outperformed the E01_Majority model previously proposed by Saxena et al. (2019) for both the net demand and the demand scenario. The E02_SingleVote model outperformed the M11_ClassANNSTwSMOTE model on two out of nine months for the net demand scenario and on one month for the demand scenario.

Paired T-tests for a mean difference equal to 0 (vs $\neq 0$) were performed for both the sensitivity and the balanced accuracy monthly results of each of the ensemble models for both scenarios (net demand and demand). Detailed results of these paired T-tests are available in Appendix 5. The null hypothesis failed to be rejected for all tests with a 95% of confidence. These results demonstrate that the sensitivity and balanced accuracy values for the two ensemble models do not show a statistical significant difference caused by the presence of BTMREG. These results suggest that the performance of the ensemble PELD forecasting models developed is statistically the same regardless of the presence of BTMREG, which means that the PELD forecasting methodology developed is equally effective for both scenarios.

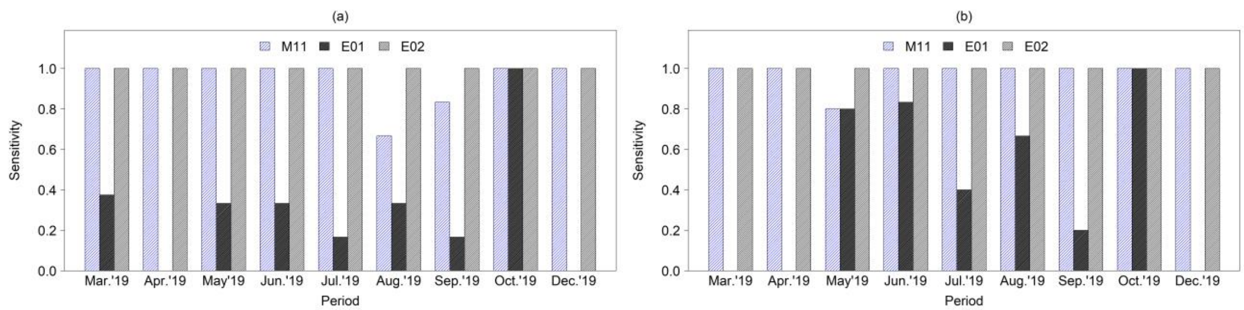

Fig. 18. Sensitivity achieved by the M11, E01 and E02 models for the (a) net demand and (b) the demand scenarios.

Table 21 shows the average monthly sensitivity and balanced accuracy values, as well as the total number of false positives and false negatives predictions produced by the two ensemble PELDs forecasting models evaluated and the M11_ClassANNSTwSMOTE model. This table illustrates how the M11_ClassANNSTwSMOTE model outperformed the remaining models at achieving the highest values for average monthly balanced accuracy. In terms of average monthly sensitivity and total number of false negatives, the M11_ClassANNSTwSMOTE model was only slightly outperformed by the E02_SingleVote model. However, the total number of false positives produced by the E02_SingleVote model is significantly greater than that produced by the other two models. Based on these results and the intent to select the most parsimonious of the models, the M11_ClassANNSTwSMOTE was selected as the best model to use for PELDs prediction with BTMREG for this facility because of the model's performance and lower complexity. Table 22 shows the ranked scorecard results that corroborate this model selection for both scenarios, net demand and demand.

Table 21

Average monthly sensitivity and balanced accuracy values, number of false positives and false negatives produced by the M11, E01, and E02 models for the net demand and the demand scenarios.

		M11 ClassANNST wSMOTE	**E01** Majority	**E02** SingleVote
Average monthly sensitivity	**Net demand**	0.9444	0.3009	1
	Demand	0.9778	0.4333	1
Average monthly balanced accuracy	**Net demand**	0.8913	0.6505	0.7046
	Demand	0.8906	0.7124	0.7333
Total number of false negatives	**Net demand**	3	34	0
	Demand	1	18	0
Total number of false positives	**Net demand**	37	0	136
	Demand	48	2	133

Table 22

Ranked scorecard results for selecting best overall performing PELDs forecasting model.

	False Negatives (FN)			False Positives (FP)			Model Complexity			Sensitivity			Balanced Accuracy			
	Number of FN	Rank	Score	Number of FP	Rank	Score	Value	Rank	Score	Value	Rank	Score	Value	Rank	Score	**Total Score**
Net Demand																
M11	***3***	***5***	***-0.04***	***37***	***4***	***-0.41***	***0.5***	***3***	***-1.50***	***0.944***	***2***	***1.89***	***0.891***	***1***	***0.89***	***0.83***
E01	*34*	*5*	*-0.47*	*0*	*4*	*0.00*	*0.6*	*3*	*-1.80*	*0.301*	*2*	*0.60*	*0.651*	*1*	*0.65*	*-1.01*
E02	*0*	*5*	*0.00*	*136*	*4*	*-1.49*	*0.6*	*3*	*-1.80*	*1.000*	*2*	*2.00*	*0.705*	*1*	*0.70*	*-0.59*
Demand																
M11	***1***	***5***	***-0.01***	***48***	***4***	***-0.53***	***0.5***	***3***	***-1.50***	***0.978***	***2***	***1.96***	***0.891***	***1***	***0.89***	***0.81***
E01	*18*	*5*	*-0.25*	*2*	*4*	*-0.02*	*0.6*	*3*	*-1.80*	*0.433*	*2*	*0.87*	*0.712*	*1*	*0.71*	*-0.49*
E02	*0*	*5*	*0.00*	*133*	*4*	*-1.46*	*0.6*	*3*	*-1.80*	*1.000*	*2*	*2.00*	*0.733*	*1*	*0.73*	*-0.52*

4.1.4 Potential and model savings calculation

Table 23 shows the potential and model savings expected upon implementation of the selected M11_ClassANNSTwSMOTE model for both the net demand and the demand scenarios. Potential savings were calculated using the methodology previously described in Section 3.2.6. Model savings were only applicable for months during which the day with the highest peak load of the month was predicted by the model as a PELD. These savings were determined according to Equation 19.

$$Model\ savings\ in\ kw = HPEL - \max\{HFN, Dlim\} \tag{19}$$

Where *HFN* = highest non-detected (ergo not reduced) peak load (or false negative PELD prediction) of the month in kW, and HPEL and Dlim are the same as in Equation 19.

Table 23

Potential and model savings, and model achievement percentage during the testing period.

	Net demand			Demand		
Period	**Potential savings** *(in kW)*	**Model savings** *(in kW)*	**% Model achievement**	**Potential savings** *(in kW)*	**Model savings** *(in kW)*	**% Model achievement**
Mar. 2019	247.38	247.38	100%	38.20	38.20	100%
Apr. 2019	749.38	749.38	100%	513.42	513.42	100%
May 2019	766.07	766.07	100%	775.28	0.00	0%
Jun. 2019	884.60	884.60	100%	1,045.36	1,045.36	100%
Jul. 2019	1,657.66	1,657.66	100%	1,281.79	1,281.79	100%
Aug. 2019	1,247.15	1,008.00	81%	835.31	835.31	100%
Sep. 2019	1,169.79	753.00	64%	731.44	731.44	100%
Oct. 2019	2,163.22	2,163.22	100%	2,701.08	2,701.08	100%
Dec. 2019	131.22	131.22	100%	94.62	94.62	100%
Aggregate	**9,016.47**	**8,360.53**	**93%**	**8,016.50**	**7,241.22**	**90%**

Potential and model savings in US$ were calculated by applying a US$17.00 per kW peak load rate to the previously calculated potential and model savings in kW (see Table 23). This peak load rate was obtained from the utility that serves the educational consumer evaluated. However, this rate is not necessarily the actual rate that the consumer pays for demand charges. The actual rates are negotiated confidentially. This was still the approximate active peak load rate at the time of this manuscript submission. The results presented in Tables 23 and 24 demonstrate how the selected M11_ClassANNSTwSMOTE model would have achieved 93% of the potential savings in kW and US$ 142,129.01 savings in electricity costs for the educational consumer within the net demand scenario. The results also show how there are more potential and model savings to be achieved after adopting BTMREG. At first glance, this is a very counterintuitive finding because by definition a customer's load profile is reduced when BTMREG is present as we have seen in Figures 3 and 4. Figure 19 illustrates how this finding can be explained by the fact that the demand reduction targets set for demand response actions (based on the monthly threshold (Dlim)) when BTMREG is present, are typically lower than the targets set when BTMREG is not present. The values for monthly threshold (Dlim) for the complete testing period can be compared by looking back at Table 5. In addition, there is always the possibility of peak loads within the net demand scenario to be as high as those within the demand scenario if there is a considerable drop in BTMREG levels.

Table 24

Potential, model, and missed savings in US$ during the testing period.

	Net demand			Demand		
Period	**Potential savings**	**Model savings**	**Missed savings**	**Potential savings**	**Model savings**	**Missed savings**
Mar. 2019	4,205.46	4,205.46	0.00	649.40	649.40	0.00
Apr. 2019	12,739.46	12,739.46	0.00	8,728.14	8,728.14	0.00
May 2019	13,023.19	13,023.19	0.00	13,179.76	0.00	13,179.76
Jun. 2019	15,038.20	15,038.20	0.00	17,771.12	17,771.12	0.00
Jul. 2019	28,180.22	28,180.22	0.00	21,790.43	21,790.43	0.00
Aug. 2019	21,201.55	17,136.00	4,065.55	14,200.27	14,200.27	0.00
Sep. 2019	19,886.43	12,801.00	7,085.43	12,434.48	12,434.48	0.00
Oct. 2019	36,774.74	36,774.74	0.00	45,918.36	45,918.36	0.00
Dec. 2019	2,230.74	2,230.74	0.00	1,608.54	1,608.54	0.00
Aggregate	**153,279.99**	**142,129.01**	**11,150.98**	**136,280.50**	**123,100.74**	**13,179.76**

Fig. 19. Model savings calculations for the (a) net demand and (b) the demand scenarios during July 2019.

Figure 20 provides more insight into this finding by illustrating the demand, net demand, solar generation, monthly thresholds (Dlim), and model savings during the day with the highest peak electric load for the month. Which was the 19th of July 2019. The figure shows how the peak electric load in the scenario without BTMREG is higher than the peak electric load in the scenario with BTMREG. This figure also shows how the peak electric load in the scenario with BTMREG was caused by a drop in solar generation. However, more model savings (1,911 kW vs 1,282 kW) are achieved because the presumptive demand reduction target set for demand response actions (based on the monthly threshold (Dlim)) is lower when BTMREG is present (net demand). The results shown in Tables 23 and 24 also indicate that the highest savings for the educational consumer are achieved during the summer months (June to August) and the first two months of fall, September and October. This can be explained by the fact that these are typically the months with the highest outside temperatures and consequently the highest energy usage for cooling purposes at the educational consumer's campus.

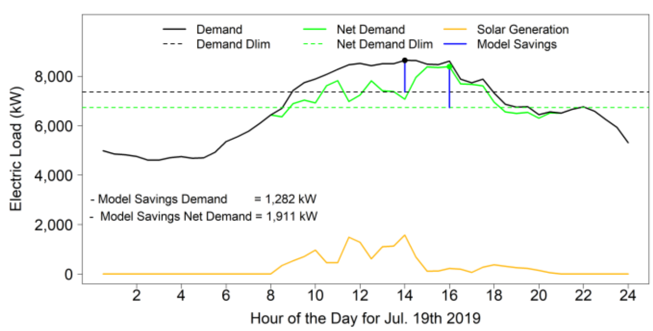

Fig. 20. Demand, net demand, solar generation, monthly thresholds (Dlim), and model savings during July 19th, 2019.

4.2 Electricity Demand Threshold Value Forecasting Results and Discussion

4.2.1 Industrial Consumer Threshold Forecasting Results

Figure 21 presents a side by side comparison of the top 5 lowest MAPE values achieved by the models when applied to the (a) Industrial_PRE (Pre-pandemic) and (b) Industrial_YR1 (Year 1 of pandemic) datasets developed with data collected from an industrial consumer. Appendix 6 provides the MAPE values achieved by all of the models. The regression random decision forest based ensemble model EM05 was the best performing model on this performance metric for this consumer before and after the COVID-19 pandemic. Figure 21(b) illustrates how the performance of the models on this performance metric was negatively affected with the pandemic overall. It is also interesting to notice that none of the arithmetic based forecasting models made it to the list of the top 5 performers. These results show that tree-based machine learning models achieve better MAPE values than arithmetic based models for this consumer.

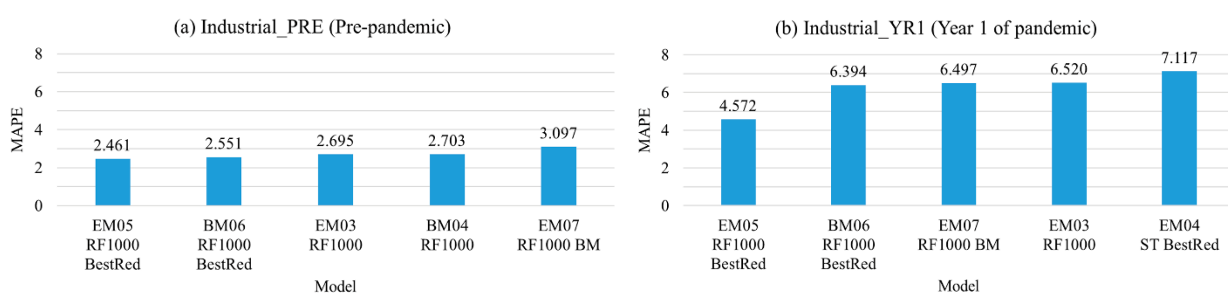

Fig. 21. Top 5 lowest MAPE value results for the (a) Industrial_PRE and (b) Industrial_YR1 datasets.

Figure 22 presents a side by side comparison of the top 5 high percentage model savings results achieved by the models when applied to the (a) Industrial_PRE and (b) Industrial_YR1 datasets. Appendix 6 provides the percentage model savings results achieved by all of the models. Once again, the regression random decision forest based ensemble model EM05 was the best performing model before and after the pandemic for this consumer. The EM05 model performed even better during the pandemic than before the pandemic in this performance metric for this consumer. It is interesting to notice how the arithmetic based BM02 model makes it to the list of the top 5 performers during the pandemic. Based on the results presented in Figure 22, tree-based machine learning models achieved a higher percentage of model savings before the pandemic, but the same cannot be said during the pandemic for this consumer.

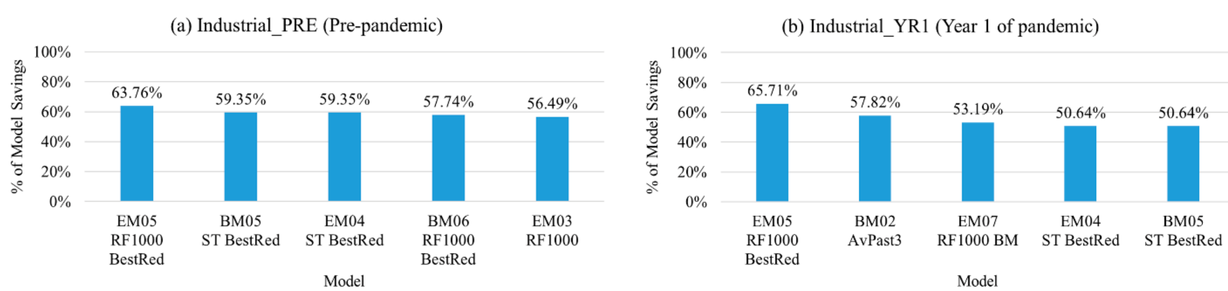

Fig. 22. Top 5 high percentage model savings results for the (a) Industrial_PRE and (b) Industrial_YR1 datasets.

Figure 23 presents a side by side comparison of the top 5 least number of false positive days achieved by the models when applied to the (a) Industrial_PRE and (b) Industrial_YR1 datasets. Appendix 6 provides the number of false positive days achieved by all of the models. The regression random decision forest based ensemble model EM05 that clearly outperformed all other models in the previous two performance metrics, comes third in this performance metric before and after the pandemic for this consumer.

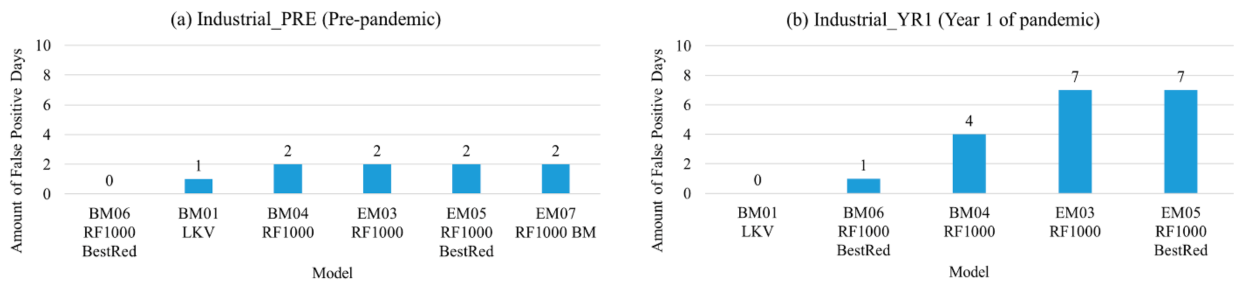

Fig. 23. Top 5 least number of false positive days for the (a) Industrial_PRE and (b) Industrial_YR1 datasets.

Table 25 shows the ranked scorecard results used to select the BM06 RF1000 BestRed as the best overall performing model for threshold forecasting for this consumer before the COVID-19 pandemic and the EM05 Ensamble RF1000 BestRed model after the pandemic. All top performing models were included in the scorecard evaluation. These models achieved model savings, in terms of electricity, of 922 kW and 1,370 kW during the testing period before and during the pandemic respectively. In financial terms, these savings translate into US$ 15,674 and US$ 23,290 for each test year period. These values were calculated based on a peak load rate of US$ 17.00 per kW charged by the local utility for the industrial and the educational consumer by the time that this research was completed.

Table 25

Ranked scorecard results for selecting best overall threshold forecasting model for the Industrial consumer.

	% Model Savings			User Inconvenience			Model Complexity			MAPE			
	Value	Rank	Score	Number of FP	Rank	Score	Value	Rank	Score	Value	Rank	Score	**Total Score**
Industrial_PRE													
BM01 - LKV	0.5167	4	2.07	1	3	-0.01	0.1	2	-0.20	4.440	1	-4.44	-2.58
BM04 - RF1000	0.5630	4	2.25	2	3	-0.02	0.3	2	-0.60	2.703	1	-2.70	-1.07
BM05 - ST BestRed	0.5935	4	2.37	16	3	-0.13	0.2	2	-0.40	4.363	1	-4.36	-2.52
BM06 - RF1000 BestRed	**0.5774**	**4**	**2.31**	**0**	**3**	**0.00**	**0.3**	**2**	**-0.60**	**2.551**	**1**	**-2.55**	**-0.84**
EM03 - Ens. RF1000	0.5649	4	2.26	2	3	-0.02	0.8	2	-1.60	2.695	1	-2.70	-2.05
EM04 - Ens. ST BestRed	0.5935	4	2.37	16	3	-0.13	0.7	2	-1.40	4.363	1	-4.36	-3.52
EM05 - Ens. RF1000 BestRed	0.6376	4	2.55	2	3	-0.02	0.8	2	-1.60	2.461	1	-2.46	-1.53
EM07 - RF1000 BM	0.5617	4	2.25	2	3	-0.02	0.8	2	-1.60	3.097	1	-3.10	-2.47
Industrial_YR1													
BM01 - LKV	0.1522	4	0.61	0	3	0.00	0.1	2	-0.20	11.469	1	-11.47	-11.06
BM02 - AvPast3	0.5782	4	2.31	46	3	-0.38	0.1	2	-0.20	12.175	1	-12.18	-10.44
BM04 - RF1000	0.3826	4	1.53	4	3	-0.03	0.3	2	-0.60	7.406	1	-7.41	-6.51
BM05 - ST BestRed	0.5064	4	2.03	8	3	-0.07	0.2	2	-0.40	7.554	1	-7.55	-5.99
BM06 - RF1000 BestRed	0.4293	4	1.72	1	3	-0.01	0.3	2	-0.60	6.394	1	-6.39	-5.29
EM03 - Ens. RF1000	0.4744	4	1.90	7	3	-0.06	0.8	2	-1.60	6.520	1	-6.52	-6.28
EM04 - Ens. ST BestRed	0.5064	4	2.03	8	3	-0.07	0.7	2	-1.40	7.117	1	-7.12	-6.56
EM05 - Ens. RF1000 BestRed	**0.6571**	**4**	**2.63**	**7**	**3**	**-0.06**	**0.8**	**2**	**-1.60**	**4.572**	**1**	**-4.57**	**-3.60**
EM07 - RF1000 BM	0.5319	4	2.13	8	3	-0.07	0.8	2	-1.60	6.497	1	-6.50	-6.04

Figure 24 presents a side by side comparison of the evolution of the top variable importance values throughout the testing set with model BM06 for the (a) Industrial_PRE and (b) Industrial_YR1 datasets. Even though the EM05 was selected as the best performing model for the Industrial_YR1 dataset by the scorecard, the only input for this model is the base BM06 model. It is interesting to notice how the last known value of the threshold (Demand_Dlim_LKV) becomes one of the most important input variables for this model to forecast the electricity demand threshold during the pandemic (Industrial_YR1). Also, more consumption related inputs such as “AvDemand_PM”, “AvLowDemand_PM”, and “AvHighDemand_PM” become important during the pandemic (Industrial_YR1). This shows how the model starts to rely more on historical consumption data than on weather data to forecast the demand threshold for this consumer during the pandemic. These results show that calendar related data is not very relevant for this consumer. Only the input “Month” is seen and this only happens in the pre-pandemic scenario.

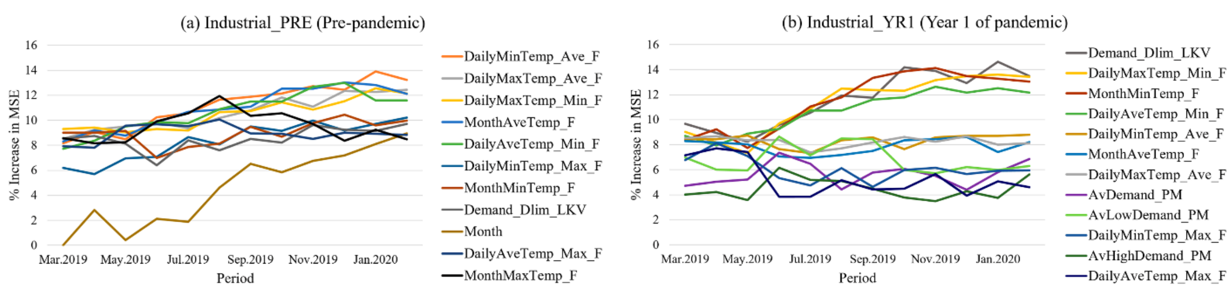

Fig. 24. Evolution of the variables with the highest importance values with model BM06 for the (a) Industrial_PRE and (b) Industrial_YR1 datasets.

4.2.2 Educational Consumer Threshold Forecasting Results

Figure 25 presents the top 5 lowest MAPE values achieved by the models when applied to the educational consumer. Appendix 7 provides MAPE values achieved by all of the models. The regression random decision forest based ensemble model EM05 was the best performing model on this performance metric for this consumer. None of the arithmetic based as well as none of the regression single decision tree based forecasting models made it to the list of the top 5 performers.

These results show that tree-based machine learning models, especially regression random decision forest based models in this case; achieve better MAPE values than arithmetic based models for this consumer.

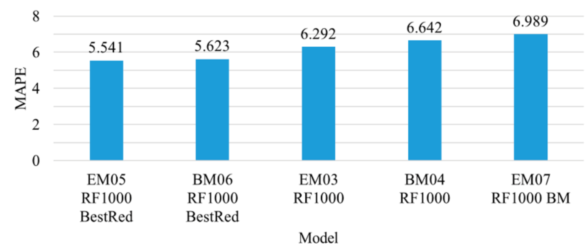

Fig. 25. Top 5 lowest MAPE value results for an educational consumer.

Figure 26 presents the top 5 highest percentage model savings results achieved by the models when applied to the educational consumer. Appendix 7 provides the percentage model savings results achieved by all of the models. The regression random decision forest based ensemble model EM03 was the best performing model on this performance metric for this consumer. None of the regression single decision tree based forecasting models made it to the list of the top 5 performers. These results show that both arithmetic and tree-based machine learning models, especially regression random decision forest based models in this case; can achieve high percentage model savings results for this consumer.

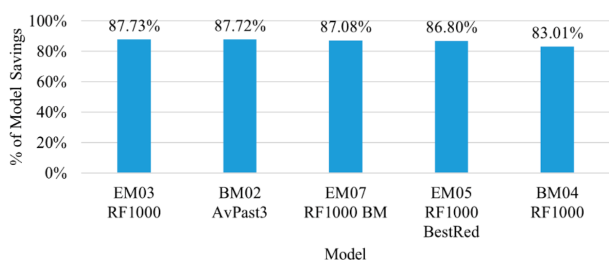

Fig. 26. Top 5 highest percentage model savings results for an educational consumer.

Figure 27 presents the top 5 lowest number of false positive days achieved by the models when applied to the educational consumer. Appendix 7 provides the number of false positive days achieved by all of the models. The regression random decision forest based base model BM06 was the best performing model on this performance metric for this consumer. None of the arithmetic based forecasting models made it to the list of the top 5 performers. These results show that tree-based machine learning models can generate a lower number of false positive days than the number generated by arithmetic based models for this consumer.

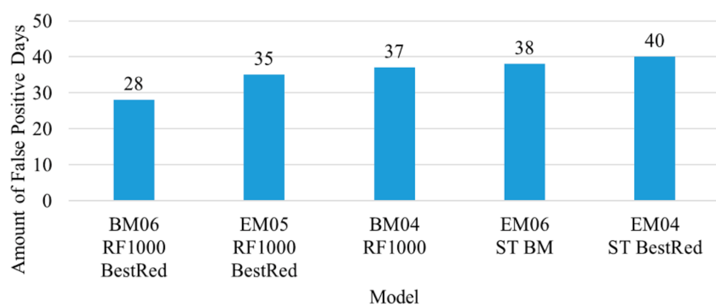

Fig. 27. Top 5 least number of false positive days for an educational consumer.

Table 26 shows the ranked scorecard results used to select the BM06 RF1000 BestRed as the best overall performing model for this consumer. All top performing models were included in the scorecard evaluation. This model achieved model savings of 6,330 kW during the testing period, which translates into US$ 107,610 for the full testing period using the same US$/kW rate previously described at the end of Section 4.2.1.

Table 26

Ranked scorecard results for selecting best overall threshold forecasting model for the Educational consumer.

	% Model Savings			User Inconvenience			Model Complexity			MAPE			
	Value	Rank	Score	Number of FP	Rank	Score	Value	Rank	Score	Value	Rank	Score	**Total Score**
BM02 - AvPast3	0.8772	4	3.51	70	3	-0.58	0.1	2	-0.20	10.158	1	-10.16	-7.42
BM04 - RF1000	0.8301	4	3.32	37	3	-0.30	0.3	2	-0.60	6.642	1	-6.64	-4.23
BM06 - RF1000 BestRed	**0.8185**	**4**	**3.27**	**28**	**3**	**-0.23**	**0.3**	**2**	**-0.60**	**5.623**	**1**	**-5.62**	**-3.18**
EM03 - Ens. RF1000	0.8773	4	3.51	41	3	-0.34	0.8	2	-1.60	6.292	1	-6.29	-4.72
EM04 - Ens. ST BestRed	0.7989	4	3.20	40	3	-0.33	0.7	2	-1.40	6.990	1	-6.99	-5.52
EM05 - Ens. RF1000 BestRed	0.8680	4	3.47	35	3	-0.29	0.8	2	-1.60	5.541	1	-5.54	-3.96
EM06 - Ens. ST BM	0.6793	4	2.72	38	3	-0.31	0.7	2	-1.40	8.015	1	-8.02	-7.01
EM07 - RF1000 BM	0.8708	4	3.48	56	3	-0.46	0.8	2	-1.60	6.989	1	-6.99	-5.57

Figure 28 presents the evolution of the highest variable importance values throughout the testing set with model BM06 for the educational consumer. Inputs related to operational data that were available for the educational consumer such as “Ev_Classes”, “Ev_CampusOpen”, and “Ev_ResHallsOpen” were determined to be of high importance value by the model. This means that an additional effort should be made to gather these type of input variables when implementing the proposed methodology for an educational consumer.

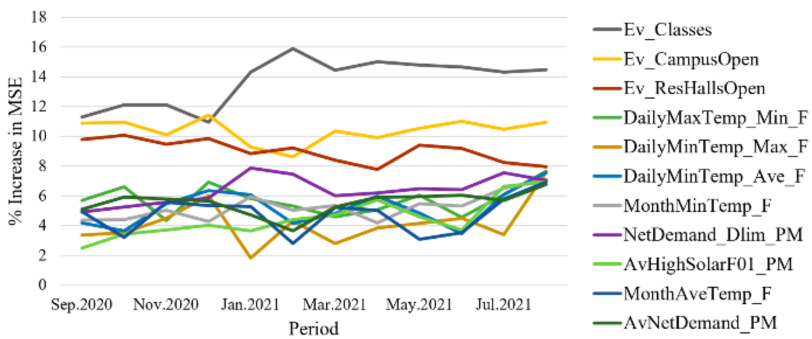

Fig. 28. Evolution of the variables with the highest importance values with model BM06 for an educational consumer.

The BM06 model also selected weather related data. However, similar to what was observed with the industrial consumer, calendar related data such as number of weekdays, number of weekend days, and the number of each day of the week are not of high importance for the educational consumer. Another interesting aspect is the presence of data related to previous months and historical electricity consumption. Input variables such as “NetDemand_Dlim_PM”, “AvHighSolarF01_PM”, and “AvNetDemand_PM” are considered of high importance by base model BM06.

4.2.3 Residential Consumer Threshold Forecasting Results

Figure 29 presents the top 5 lowest MAPE values achieved by the models when applied to the residential consumer. Appendix 7 provides the MAPE values achieved by all of the models. The regression random decision forest based base model BM06 was the best performing model on this performance metric for this consumer. None of the arithmetic based base models as well as none of the regression single decision tree based forecasting models made it to the list of the top 5 performers. These results show that tree-based machine learning models, especially regression random decision forest based models in this case; achieve better MAPE values although the arithmetic based ensemble model was also a top performer on this performance metric for this consumer.

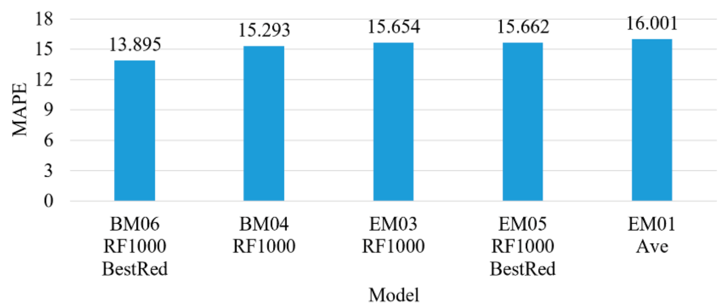

Fig. 29. Top 5 lowest MAPE value results for a residential consumer.

The MAPE values obtained for the residential consumer can look surprisingly high when compared with those obtained for the industrial and the educational consumer. Such a comparison might even raise questions in regards to the validity of the methodology for residential consumers. However, this difference can be explained by remembering that the MAPE is a percentage and taking a closer look at the actual and forecasted values for each consumer. Table 27 shows how the threshold values for the residential consumer are so low in comparison to those of the educational consumer, that a very small error value can represent a significant increase in the MAPE.

Table 27

Actual thresholds, forecast errors, and MAPE values for educational and residential consumer by month.

	Educational Consumer [BM06]			**Residential Consumer [BM06]**		
Period	**Actual Threshold**	**Forecast Error**	**MAPE**	**Actual Threshold**	**Forecast Error**	**MAPE**
Jan.	4,742.14	134	0.023	5.96	0.86	0.160
Feb.	5,038.34	501	0.097	6.15	-0.19	0.032
Mar.	5,166.36	314	0.061	10.91	-0.30	0.049
Apr.	5,041.99	485	0.109	9.59	-4.82	0.442
May.	5,199.10	190	0.040	9.39	-1.18	0.123
Jun.	6,863.66	104	0.021	8.02	-0.32	0.034
Jul.	6,326.17	-161	0.031	6.22	0.58	0.072
Aug.	6,873.45	61	0.012	5.63	1.35	0.217
Sep.	5,874.17	23	0.005	7.50	1.13	0.200
Oct.	5,176.91	-1080	0.157	7.01	-1.33	0.177
Nov.	5,125.73	-140	0.022	7.68	0.28	0.041
Dec.	4,456.31	-667	0.097	5.36	-0.93	0.121
			5.623			**13.895**

Figure 30 presents the top 5 high percentage model savings results achieved by the models when applied to the residential consumer. Appendix 7 provides the percentage model savings results achieved by all of the models. The arithmetic based base model *BM01 - Last known value* was the best performing model on this performance metric for this consumer. None of the regression single decision tree based forecasting models made it to the list of the top 5 performers. These results show that both arithmetic and tree-based machine learning models, especially regression random decision forest based models in this case; can achieve high percentage model savings results for this consumer.

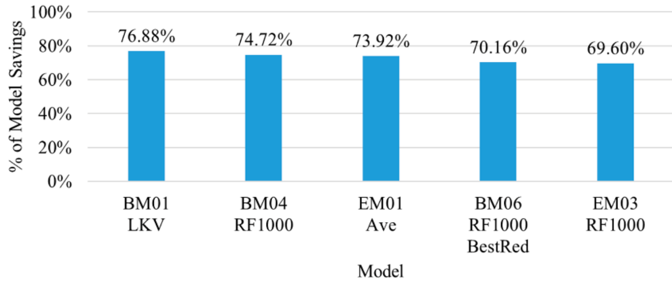

Fig. 30. Top 5 high percentage model savings results for a residential consumer.

Figure 31 presents the top 5 least number of false positive days achieved by the models when applied to the residential consumer. Appendix 7 provides the number of false positive days achieved by all of the models. The regression random decision forest based ensemble model EM05 was the best performing model on this performance metric for this consumer. None of the regression single decision tree based forecasting models made it to the list of the top 5 performers. These results show that both arithmetic and tree-based machine learning models, especially regression random decision forest based models in this case; can achieve the least number of false positive days for this consumer.

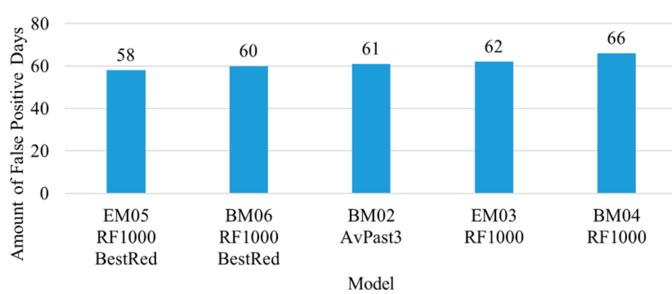

Fig. 31. Top 5 number of false positive days for a residential consumer.

Only the tree-based machine learning models BM04, BM06, and EM03 have consistently made the top 5 performers list for the three performance metrics being evaluated to select an overall best performing model for this consumer. Table 28 shows the ranked scorecard results used to select the BM06 RF1000 BestRed as the best overall performing model for this consumer. All top performing models were included in the scorecard evaluation. This model achieved model savings of 8.77 kW during the testing period, which translates into US$ 149 for the full testing period using the same US$/kW rate previously described at the end of Section 4.2.1.

Table 28

Ranked scorecard results for selecting best overall threshold forecasting model for the Residential consumer.

	% Model Savings			User Inconvenience			Model Complexity			MAPE			
	Value	Rank	Score	Number of FP	Rank	Score	Value	Rank	Score	Value	Rank	Score	**Total Score**
BM02 - AvPast3	0.8772	4	3.51	70	3	-0.58	0.1	2	-0.20	10.158	1	-10.16	-7.42
BM01 - LKV	0.7688	4	3.08	131	3	-1.08	0.1	2	-0.20	36.041	1	-36.04	-34.24
BM02 - AvPast3	0.5344	4	2.14	61	3	-0.50	0.1	2	-0.20	22.089	1	-22.09	-20.65
BM04 - RF1000	0.7472	4	2.99	66	3	-0.54	0.3	2	-0.60	15.293	1	-15.29	-13.45
BM06 - RF1000 BestRed	**0.7016**	**4**	**2.81**	**60**	**3**	**-0.49**	**0.3**	**2**	**-0.60**	**13.895**	**1**	**-13.90**	**-12.18**
EM01 - Ens. Ave	0.7392	4	2.96	75	3	-0.62	0.6	2	-1.20	16.001	1	-16.00	-14.86
EM03 - Ens. RF1000	0.6960	4	2.78	62	3	-0.51	0.8	2	-1.60	15.654	1	-15.65	-14.98
EM05 - Ens. RF1000 BestRed	0.6048	4	2.42	58	3	-0.48	0.8	2	-1.60	15.662	1	-15.66	-15.32

Figure 32 presents the evolution of the highest variable importance values throughout the testing set with model BM06 for the residential consumer. Contrary to what was observed with the industrial and the educational consumers, inputs related to historical electricity consumption, even those looking as far back as 3 months before, are the most important for this residential consumer. It is clear that weather and calendar related data did not play a high importance role in any of these models for this consumer. This observation can be explained by considering that the family that resides in this residence follows such a fixed routine month after month, that their consumption patterns are best captured by looking at their historical consumption without the need to incorporate additional weather and/or operational related data available.

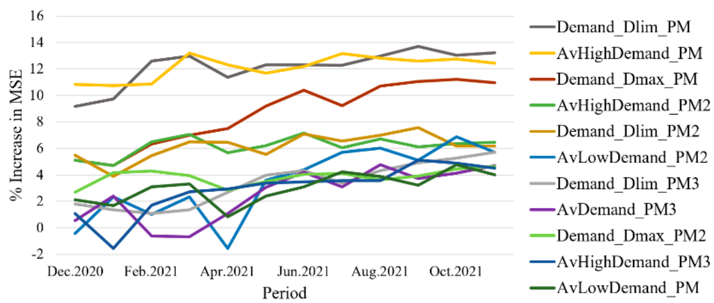

Fig. 32. Evolution of the variables with the highest importance values with model BM06 for a residential consumer.

Table 29 shows how there were no actual peak loads during 3 months (Mar., Apr., and May) of the test set for the residential consumer. When there are no actual peaks, most of the high loads of the month tend to stay at a similar level. Looking back at Table 27, the overall best performing model for the residential consumer (BM06) under-predicted the threshold value during these months. All of these events explain how, as shown in Table 29, 43 of the 60 false positive days for the whole year occurred during these 3 months.

Table 29

Number of actual peaks and of false positives for educational and residential consumer by month.

		Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	
Educational Consumer [BM06]	Number of actual peaks	6	11	8	5	7	4	6	3	9	4	4	3	70
	Number of false positives	0	3	13	1	0	4	0	5	0	0	0	0	28
Residential Consumer [BM06]	Number of actual peaks	11	7	0	0	0	2	8	6	7	10	4	11	66
	Number of false positives	3	1	23	15	5	0	0	0	5	0	9	0	60

4.2.4 Summary of Results and Model Savings Calculations

Table 30 provides a summary of the main results obtained through the case study developed. The overall best performing models for all three consumers evaluated were regression random decision forest based models. These models achieved the best results when manual feature reduction techniques were applied for the industrial and the educational consumer. The results also showed that the consumers evaluated could potentially achieve model savings within the range of 65% and 82% during a year. These results translate to US$ 149.09, US$ 23,290.00, and US$ 107,610.00 in savings for the residential, industrial, and educational consumer respectively. However, these results also showed that in order to achieve these savings, consumers would have to experience unnecessary user inconvenience during 0 to 60 days in a year. Consumers with larger differences between their peak loads and demand thresholds stand to gain the most benefit from implementing these models.

Table 30

Summary of main results obtained through the case study developed.

	Industrial_PRE *(Pre-pandemic)*	**Industrial_YR1** *(Year 1 of pandemic)*	**Educational**	**Residential**
Overall best performing model	BM06 Regression random decision forest based base model with manual feature reduction.	EM05 Regression random decision forest based ensemble model with manual feature reduction.	BM06 Regression random decision forest based base model with manual feature reduction.	BM06 Regression random decision forest based base model with manual feature reduction.
MAPE	2.55	4.57	5.62	13.90
Model savings achieved (Percentage)	57.74%	65.71%	81.85%	70.16%
Model savings achieved (Electricity)	922.00 kW	1,370.00 kW	6,330.00 kW	8.77 kW
Model savings achieved (Costs)	US$ 15,674.00	US$ 23,290.00	US$ 107,610.00	US$ 149.09
Number of false positive days per year (User inconvenience)	0	7	28	60
Average difference between monthly peak load and monthly threshold	177.24 kW	207.00 kW	617.33 kW	0.91 kW

4.3 Ensemble Peak Electric Load Days Forecasting Results and Discussion

4.3.1 Industrial Consumer Ensemble Forecasting Results

Table 31 shows the average monthly sensitivity and balanced accuracy values, as well as the total number of false positives and false negatives predictions produced by the top 6 performing models and the arithmetic based models across these indicators for the Industrial consumer. All of the top performing models achieved perfect predictions. These metrics were evaluated for all 15 models: 5 threshold-based regression models, 4 classification models, and 6 ensemble models. Appendix 8 provides a table with the results for all 15 models for this consumer. These results show how machine learning based ensemble models clearly outperform the majority classifier (E01) and the single-vote (E02) classifier for PELD forecasting for this consumer. Only two classification models were top performing models M07_ClassRF100 and M08_ClassANN. The machine learning based ensemble models clearly picked up on these models and therefore were able to adjust and achieve a perfect performance. The E03_Ens.ClassST model, a decision tree model, only picked the output of the M07_ClassRF1000 model as a selected feature during the 12 test months.

Figure 33 shows how the random forest model E04_Ens.ClassRF1000 consistently picked the output of the M07_ClassRF100 and M08_ClassANN models as two of the most important features during training. It is interesting to notice how the input of other base models that were not "top performing" is considered by this ensemble approach and adjusted through time. It is also interesting to notice how the ensemble models consider variables such as the Month, Day of the Week (DoW), Day of the Month (DoM), and the forecasted threshold (Dlim). Table 32 shows the ranked scorecard results used to select the E03 Ensemble Class ST as the best overall performing ensemble PELDs forecasting model for this consumer. All ensemble models were included in the scorecard evaluation.

Table 31

Average monthly sensitivity and balanced accuracy values, number of false positives and false negatives produced by the by the top 6 performing models across these indicators for the Industrial consumer.

	E01 Majority	**E02** Single-Vote	**M07** Class RF1000	**M08** Class ANN	**E03** Ens. Class ST	**E04** Ens. Class RF1000	**E05** Ens. Class ANN	**E06** Ens. Class AdaBoost
Average monthly sensitivity	0.344	1	1	1	1	1	1	1
Average monthly balanced accuracy	0.659	0.526	1	1	1	1	1	1
Total number of false negatives	23	0	0	0	0	0	0	0
Total number of false positives	12	313	0	0	0	0	0	0

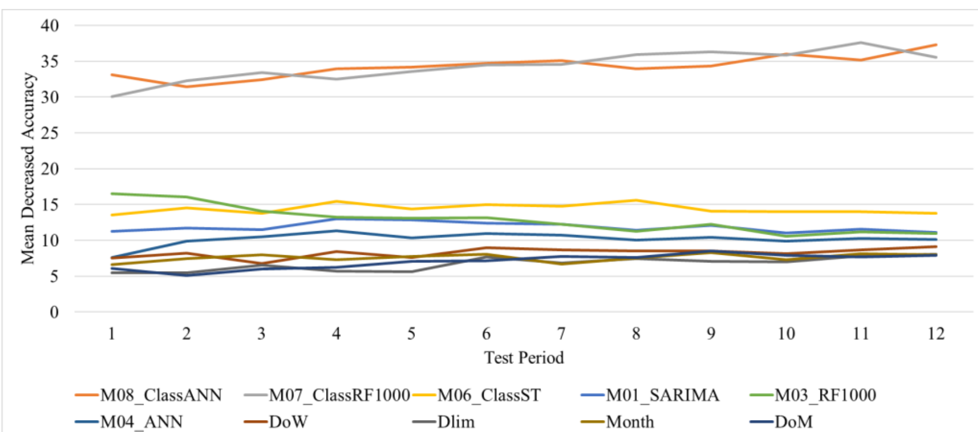

Fig. 33. Evolution of the variables with the highest importance values with model E04_Ens.ClassRF1000 for an Industrial consumer.

Table 32

Ranked scorecard results for selecting best overall performing ensemble PELDs forecasting model for the Industrial consumer.

	False Negatives (FN)			False Positives (FP)			Model Complexity			Sensitivity			Balanced Accuracy			
	Number of FN	Rank	Score	Number of FP	Rank	Score	Value	Rank	Score	Value	Rank	Score	Value	Rank	Score	**Total Score**
E01	23	5	-0.32	12	4	-0.13	0.6	3	-1.80	0.344	2	0.69	0.659	1	0.66	-0.90
E02	0	5	0.00	313	4	-3.43	0.6	3	-1.80	1.000	2	2.00	0.526	1	0.53	-2.70
E03	**0**	**5**	**0.00**	**0**	**4**	**0.00**	**0.7**	**3**	**-2.10**	**1.000**	**2**	**2.00**	**1.000**	**1**	**1.00**	**0.90**
E04	0	5	0.00	0	4	0.00	0.8	3	-2.40	1.000	2	2.00	1.000	1	1.00	0.60
E05	0	5	0.00	0	4	0.00	1.0	3	-3.00	1.000	2	2.00	1.000	1	1.00	0.00
E06	0	5	0.00	0	4	0.00	0.9	3	-2.70	1.000	2	2.00	1.000	1	1.00	0.30

4.3.2 Educational Consumer Ensemble Forecasting Results

Table 33 shows the average monthly sensitivity and balanced accuracy values, as well as the total number of false positives and false negatives predictions produced by the top 5 performing models and the arithmetic based models across these indicators for the Educational consumer. Appendix 9 provides a table with the results for all 15 models for this consumer. Contrary to what happened with the Industrial consumer, it is difficult to select a best performing models with this information. Table 34 shows the ranked scorecard results used to select the E03 Ensemble Class ST as the best overall performing ensemble PELDs forecasting model for this consumer. All ensemble models were included in the scorecard evaluation. Even though the best performing model is machine learning based, some of the machine learning based ensemble models were not able to outperform the arithmetic based models for this consumer. A combination of limited data compared to the other two consumers, more input variables, and the effect of renewable energy could be playing a role in making the forecasting task more challenging for these models.

Table 33

Average monthly sensitivity and balanced accuracy values, number of false positives and false negatives produced by the by the top 5 performing models across these indicators for the Educational consumer.

	E01 Majority	**E02** Single-Vote	**M09** Class ANN_ST	**M08** Class ANN	**M06** Class ST	**E03** Ens. Class ST	**E04** Ens. Class RF
Average monthly sensitivity	0.301	1	0.944	0.898	0.585	0.681	0.537
Average monthly balanced accuracy	0.651	0.705	0.891	0.833	0.677	0.796	0.741
Total number of false negatives	34	0	3	6	11	17	20
Total number of false positives	0	136	37	55	54	21	13

Table 34

Ranked scorecard results for selecting best overall performing ensemble PELDs forecasting model for the Educational consumer.

	False Negatives (FN)			False Positives (FP)			Model Complexity			Sensitivity			Balanced Accuracy			
	Number of FN	Rank	Score	Number of FP	Rank	Score	Value	Rank	Score	Value	Rank	Score	Value	Rank	Score	**Total Score**
E01	34	5	-0.47	0	4	0.00	0.6	3	-1.80	0.301	2	0.60	0.651	1	0.65	-1.01
E02	0	5	0.00	136	4	-1.49	0.6	3	-1.80	1.000	2	2.00	0.705	1	0.71	-0.59
E03	**17**	**5**	**-0.23**	**21**	**4**	**-0.23**	**0.7**	**3**	**-2.10**	**0.681**	**2**	**1.36**	**0.796**	**1**	**0.80**	**-0.41**
E04	20	5	-0.27	13	4	-0.14	0.8	3	-2.40	0.537	2	1.07	0.741	1	0.74	-1.00
E05	27	5	-0.37	33	4	-0.36	1.0	3	-3.00	0.315	2	0.63	0.615	1	0.62	-2.49
E06	29	5	-0.40	10	4	-0.11	0.9	3	-2.70	0.437	2	0.87	0.697	1	0.70	-1.64

Figure 34 shows how the random forest model E04_Ens.ClassRF1000 consistently picked the output of classification ANN based base models as two of the most important features during training. It is interesting to notice how calendar data such as Day of the Week (DoW) and Day of the Month (DoM) is selected as important by the model as well as the cooling requirements.

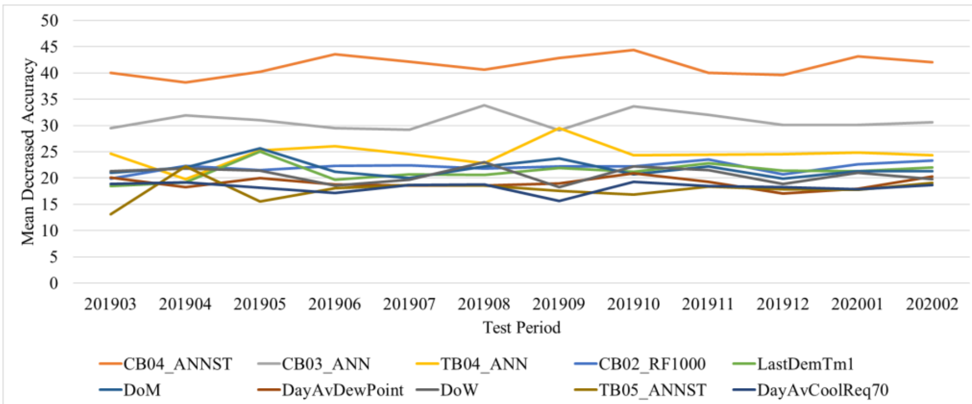

Fig. 34. Evolution of the variables with the highest importance values with model E04_Ens.ClassRF1000 for an Educational consumer.

4.3.3 Residential Consumer Ensemble Forecasting Results

Table 35 shows the average monthly sensitivity and balanced accuracy values, as well as the total number of false positives and false negatives predictions produced by the top 5 performing models and the arithmetic based models across these indicators for the Residential consumer. Four of these models achieved perfect predictions. These metrics were evaluated for all 15 models: 5 threshold-based regression models, 4 classification models, and 6 ensemble models. Appendix 10 provides a table with the results for all 15 models for this consumer. Similar to the results for the Industrial consumer, these results also show how machine learning based ensemble models clearly outperform the majority classifier (E01) and the single-vote (E02) classifier for PELD forecasting for this consumer. Only one classification base model was a top performing model (M07_ClassRF1000). The machine learning based ensemble models clearly picked up on this model and therefore were able to adjust and achieve a perfect performance, with an exception for the ANN model. The E03_Ens.ClassST model, a decision tree model, picked the output of the M07_ClassRF1000 model along with the output of the M06_ClassST model, and the maximum electric demand registered the day before as features during the seven test months.

Table 35

Average monthly sensitivity and balanced accuracy values, number of false positives and false negatives produced by the by the top 5 performing models across these indicators for the Residential consumer.

	E01 Majority	**E02** Single-Vote	**M07** Class RF1000	**E03** Ens. Class ST	**E04** Ens. Class RF1000	**E06** Ens. Class AdaBoost	**E05** Ens. Class ANN
Average monthly sensitivity	0.618	1	1	1	1	1	0.996
Average monthly balanced accuracy	0.777	0.500	1	1	1	1	0.991
Total number of false negatives	80	0	0	0	0	0	1
Total number of false positives	9	154	0	0	0	0	2

Table 36 shows the ranked scorecard results used to select the E03 Ensemble Class ST as the best overall performing ensemble PELDs forecasting model for this consumer. All ensemble models were included in the scorecard evaluation. Figure 35 shows how the random forest model E04_Ens.ClassRF1000 consistently picked the output of the M07_ClassRF1000 model as one of the most important features during training. It is interesting to notice how the input of other base models that were not "top performing" is considered by this ensemble approach and adjusted through time. It is also interesting to notice how the ensemble models also consider weather related variables in addition to the models as oppose to calendar related as it was observed for the Industrial and Educational consumers. This shows that the methodology truly adapts to each consumer.

Table 36

Ranked scorecard results for selecting best overall performing ensemble PELDs forecasting model for the Educational consumer.

	False Negatives (FN)			False Positives (FP)			Model Complexity			Sensitivity			Balanced Accuracy			
	Number of FN	Rank	Score	Number of FP	Rank	Score	Value	Rank	Score	Value	Rank	Score	Value	Rank	Score	**Total Score**
E01	80	5	-1.10	9	4	-0.10	0.6	3	-1.80	0.618	2	1.24	0.777	1	0.78	-0.98
E02	0	5	0.00	154	4	-1.69	0.6	3	-1.80	1.000	2	2.00	0.500	1	0.50	-0.99
E03	**0**	**5**	**0.00**	**0**	**4**	**0.00**	**0.7**	**3**	**-2.10**	**1.000**	**2**	**2.00**	**1.000**	**1**	**1.00**	**0.90**
E04	0	5	0.00	0	4	0.00	0.8	3	-2.40	1.000	2	2.00	1.000	1	1.00	0.60
E05	1	5	-0.01	2	4	-0.02	1.0	3	-3.00	0.996	2	1.99	0.991	1	0.99	-0.05
E06	0	5	0.00	0	4	0.00	0.9	3	-2.70	1.000	2	2.00	1.000	1	1.00	0.30

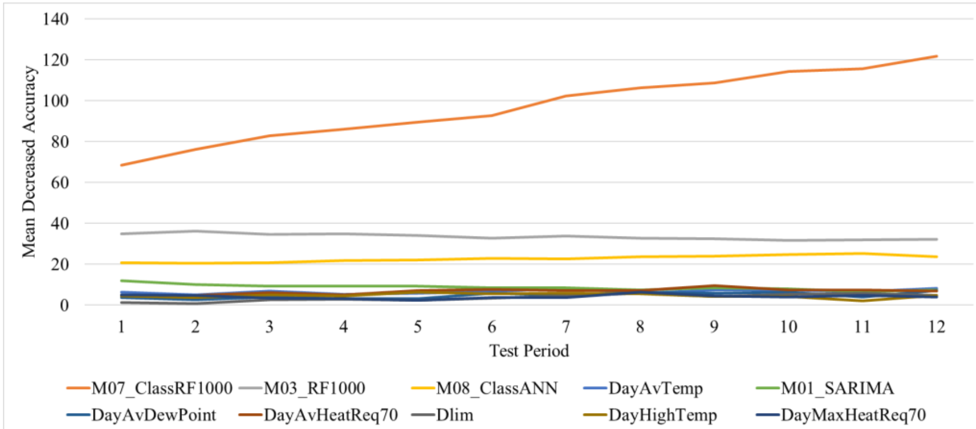

Fig. 35. Evolution of the variables with the highest importance values with model E04_Ens.ClassRF1000 for a Residential consumer.

Chapter 5: Conclusions

This chapter will present the author's conclusions drawn from the results of the completed stages of the research detailed within this dissertation. These results have been previously presented in Sections 4.1 through 4.3. Emphasis will be placed on the results' implications and contributions in the field of peak electric load days (PELDs) forecasting methodologies for consumers with and without behind the meter renewable electricity generation (BTMREG) when applicable.

5.1 Peak Electric Load Days Forecasting Conclusions

The results presented in this dissertation have provided evidence that shows that threshold based and/or classification based forecasting methodologies **can** accurately forecast more than 70% of a year's peak electric load days (PELDs) for consumers with behind the meter renewable electricity generation (BTMREG). This task was completed using autoregressive integrated moving average (ARIMA), classification and regression trees (CART), random classification and regression forest, and artificial neural network (ANN) based machine learning techniques. The research described in this manuscript has provided three main contributions in order to address the lack of published studies detailing accurate PELDs forecasting methodologies applicable to the increasing number of facilities adopting BTMREG, as well as the lack of published studies comparing the performance of these methodologies in both the presence and absence of BTMREG.

The most interesting insight provided by these contributions is that counterintuitively, there can be more potential and model savings to be achieved by facilities using PELDs forecasting methodologies after adopting BTMREG. The results show how implementing these methodologies after BTMREG adoption becomes even more important than before the adoption in order to achieve financial savings. At first, many researchers and practitioners might not consider this outcome because by definition, a customer's load profile is reduced when BTMREG is adopted which on the surface may appear to reduce the number of load reduction opportunities.

The first of the three main contributions is the development and testing of a PELDs forecasting methodology applicable to both consumers with and without BTMREG. This methodology was tested using ARIMA, CART, random regression and classification forest, ANN, and ensemble based models. However, the methodology is model agnostic and different models can be tested in future research efforts. The experimental results showed that an ANN based model using features selected by a CART based model (M11_ClassANNSTwSMOTE) and one of the ensemble models (E02_SingleVote) achieved the highest average monthly sensitivity values for both the net demand and the demand scenarios. Based on the average monthly sensitivity and balanced accuracy values, the number of false positives and negatives produced by the model, and the intent to select the most parsimonious of the models, the M11_ClassANNSTwSMOTE was selected as the preferred model to use for PELDs prediction for this facility with BTMREG present. This model showed superior performance and reduced complexity. Furthermore, this model demonstrated the capacity to have achieved 93% of the potential savings in kW and US$ 142,129.01 savings in electricity costs during a yearlong testing period for the scenario with BTMREG. Given these results, it was concluded that practitioners interested in achieving the best model performance using parsimonious models should start with the implementation of classification-based models. Based on the results obtained from these models, more elaborated approaches such as threshold-based PELDs classification and ensemble approaches might not be needed.

The second contribution is the documentation of the first of their kind side-by-side empirical comparisons between the performance of ARIMA, CART, random regression and classification forest, ANN, and ensemble based models at forecasting electric load and PELDs in both scenarios, with and without BTMREG. The results obtained while testing the proposed methodology in the scenario without BTMREG serve as additional validation of the work published by Saxena et al. (2019) about forecasting PELDs without BTMREG. The results obtained through the side-by-side empirical comparisons in the scenario with BTMREG provided four important insights in regards to past, present, and potential future research. First, both a random forest (M03_RegRF) and an ANN based regression model (M05_RegANNST) outperformed ARIMA and CART regression-based models at predicting future electric load levels for both the net demand and the demand scenarios. Second, comparing the results of the scenario with BTMREG and the scenario without BTMREG, empirical evidence suggesting that the presence of BTMREG affects the performance of the models was only observed for the regression-based models evaluated. The results obtained from the classification-based models as well as the ensemble models evaluated did not show evidence of an effect on the performance of these models due to the presence of BTMREG.

The third and fourth insights provide important details about the methodology to consider for future research based on past publications and the current results. The third insight was that class imbalance issues in the dataset need to be addressed in order to achieve the best performances when implementing the classification-based PELDs forecasting approach regardless of the presence or absence of BTMREG. The fourth insight was that the single vote ensemble approach outperformed the current majority vote approach proposed by Saxena et al. (2019) but produced a significantly greater number of false positive predictions when compared to the other models evaluated. The use of ensemble forecasting for PELDs forecasting can be further explored by evaluating other ensemble forecasting methodologies.

A first of its kind PELDs forecasting model savings comparison for scenarios with and without BTMREG was presented as the third and final contribution of this first phase of the research. We have already discussed the first insight provided by this contribution at the beginning of this section. This was also the most interesting insight, the discovery of the possibility for more potential and model savings to be achieved by facilities using PELDs forecasting methodologies when BTMREG is adopted. The second insight provided by this contribution was that the months with the highest outside temperatures and consequently the highest energy usage for cooling purposes were also the months with the greatest savings to be achieved at the educational consumer's campus.

5.2 Electricity Demand Threshold Value Forecasting Conclusions

The results disclosed within this dissertation also allow us to determine that regression tree and random regression forest based machine learning models **can** outperform widely used expert based and arithmetic based methodologies at forecasting an efficient electricity demand threshold value. Demand thresholds have proven useful to trigger cost saving peak demand shaving actions and can be determined before the start of a billing period without receiving any signal or information from the utility. This dissertation presented a novel methodology that empowers any electricity consumer paying for peak demand charges to proactively execute demand response actions even without receiving signals or information coming from the utility, and only when necessary to effectively reduce demand charges and user inconvenience. The results obtained using real data from three different types of consumers showed that the overall best performing models for all three consumers evaluated were regression random decision forest based models. These models achieved the best results when manual feature reduction techniques were applied for the industrial (both before and during the COVID-19 pandemic) and the educational consumer. The results also showed that the consumers evaluated could potentially achieve model savings within the range of 65% and 82% during a year. These results translate to US$ 149.09, US$ 23,290.00, and US$ 107,610.00 in savings for the residential, industrial, and educational consumer respectively.

The most influential input variables vary from consumer to consumer. In addition to the values forecasted by the base models, weather and historical electricity consumption related variables were the most important for the industrial consumer. In the case of the educational consumer, which had the most detailed operational data available, the input variables related to operation details were the most important in addition to the values forecasted by the base models. Calendar related data such as number of weekdays, number of weekend days, and the number of each day of the week did not achieve high importance values for any of the consumers, except “Month” for the industrial consumer and only before the pandemic. The residential consumer was the only one for whom variables related to historical electricity consumption were the most important input variables. These results show that some consumers can have such a fixed electricity consumption routine month after month, that their consumption patterns are best captured by looking at their historical consumption without the need to incorporate additional weather, calendar, and/or operational related data available.

5.3 Ensemble Peak Electric Load Days Forecasting Conclusions

In addition, the results provided by this dissertation show that classification tree, random classification forest, adaptive boosting (AdaBoost), and artificial neural network (ANN) based ensemble modeling techniques **can** outperform majority based and single-vote based ensemble modeling techniques at forecasting peak electric load days (PELDs). The machine learning based ensemble models demonstrated the ability to adjust to the strongest features in order to generate the best forecast. When comparing the features selected by the one of the best performing models for the Residential consumer and the Industrial consumer, it was evident that the model adapted to each consumer’s particular situation.

Choosing machine learning based models such as classification tree, random classification forest, adaptive boosting (AdaBoost), and artificial neural network (ANN), ensures that an ongoing PELD management system will adapt to the latest conditions for each consumer. Approaches that provide a clear understanding in regards to what features are being the most influential, such as classification trees and random classification forest can allow administrator to focus on those features and potentially simplify the PELD monitoring tasks. One drawback of using machine learning based ensemble models is that they will require more historical data. They might not be an option to start from zero, but definitely the option to build towards.

Chapter 6: Recommendations for Future Studies

There is a considerable amount of potential future work open for further exploration within the proposed methodologies and the PELD forecasting topic. Research into additional strategies to overcome the class imbalance problem for classification-based models as well as the exploration of additional ensemble forecasting methodologies have already been mentioned as potential future research opportunities. Before recommendations on policies can be made, the effects of this type of strategy being run on an entire set of consumers across a utility's service area needs to be completed. Future research should evaluate the input on utility demand curves as more consumers adopt such a strategy. It is unclear from the current research if this strategy would complement utility's attempt to curb peak demand or if it could potentially exacerbate the problem.

Taking advantage of the model agnostic characteristic of the proposed methodologies, additional models that might outperform the models already evaluated could be integrated in a future study. However, researchers exploring additional models, especially more complex models, are encouraged to pay close attention and take measures in order to minimize the risk of overfitting. The use of area under the curve (AUC) of the receiver operating characteristic (ROC) curve to develop PELDs forecasting models that might eliminate the need to forecast a monthly threshold can be a very interesting avenue for future research. As previously stated, this performance metric is especially useful in situations where classification is more accurate if performed considering different classification thresholds.

Future studies could also focus on developing model selection techniques based on optimization models designed specifically for the PELD forecasting application for every model evaluated. A potential total cost function could integrate the known cost of false negatives based on the peak demand charges applicable to the customer. The function could also include a cost of false positives based on the financial cost of user inconvenience given by evaluating elements such as reduced productivity and the labor cost of executing unnecessary demand response events. Researchers can determine these costs, as well as other applicable costs, establish applicable constraints, and develop an optimization model with the objective of minimization a total cost function in order to select the best performing model at every stage. The approach can also be modified to maximize a net profit, or in this case net savings function, that includes the savings generated by every true positive and the cost associated with executing the required demand response events to generate each of those true positives.

In regards to the task of forecasting an efficient monthly threshold, given a dataset bigger than 24 data points, models that are more complex might be considered. However, when considering these models, it should also be considered that a clear feature selection process might still be of interest to researchers and practitioners. Furthermore, another approach worth exploring and comparing to the proposed monthly threshold forecasting methodology is using forecasting techniques to produce an hourly forecast for the entire billing period of interest and calculating the threshold using Equation 12. Even though the accuracy of the threshold will be highly dependent on the accuracy of the hourly forecast, the researchers believe that this could be a natural next step to the research presented in this paper. From another angle, typical regression techniques such as the ones used during this study are optimized to produce results as close as possible to the actual value rather than favoring either under-prediction or over-prediction. Nevertheless, for this application, modifying already existing regression techniques or developing new ones that are optimized to favor slightly under-predicting models in order to reduce over-prediction could prove useful if the consumer is willing to allow a certain amount of user inconvenience in order to secure higher cost savings.

During the experimentation stage, it was noticed that the consumer with behind the meter renewable electricity generation (BTMREG) always had a positive net demand. This leads the researchers to also suggest the evaluation of how the proposed methodology would perform in cases where a consumer with BTMREG produces more electricity than it consumes. This type of consumer has a negative net demand and most likely either stores the excess energy or sells it back to the grid.

Other paths for future research that have been revealed by the insights discovered during the execution of this study as well as those provided by the results include:

- How does the resolution of the dataset (30 mins. vs 60 mins. vs x mins) affect the performance of the models?
- What is the optimal size of the training dataset for each model?
- What are the effects of training the models with just the hours when peak loads occur in order to reduce the class imbalance?
- How effective is the methodology for other types of REG sources such as wind and hydro?

Further research into these questions will help researchers and practitioners develop improved PELDs forecasting methodologies that support consumers with and without BTMREG on their task to reduce peak electric load related costs.

References

Abu-El-Magd, M. A., & Sinha, N. K. (1982). Short-term load demand modeling and forecasting: A review. IEEE Transactions on Systems, Man, and Cybernetics, 12(3), 370-382. https://doi.org/10.1109/TSMC.1982.4308827

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Elsevier B.V. https://doi.org/10.1016/S0731-7085(99)00272-1

Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy and Buildings, 147, 77-89. https://doi.org/10.1016/j.enbuild.2017.04.038

Alfares, H., & Mohammad, N. (2002). Electric load forecasting: Literature survey and classification of methods. International Journal of Systems Science - IJSySc. 33:1. 23-34. https://doi.org/10.1080/00207720110067421

Almahmoud, Z., Crandall, J., Elbassioni, K., Nguyen, T. T., & Roozbehani, M. (2019). Dynamic pricing in smart grids under thresholding policies. IEEE Transactions on Smart Grid, 10(3), 3415-3429. https://doi.org/10.1109/TSG.2018.2825997

Aponte, O., & McConky, K. T., (2019). Proceedings of the 2019 IISE annual conference: Renewable Electricity Generation Impact on Peak Demand and Response Strategies. Retrieved from https://584621ae7566faad44cf-fb5903404bc6e3be139ec7116b0f2cfc.ssl.cf1.rackcdn.com/H-967345-724160-1-001.pdf [Accessed 20 April 2021]

Asadinejad, A., & Tomsovic, K. (2017). Optimal use of incentive and price based demand response to reduce costs and price volatility. Electric Power Systems Research, 144(C), 215-223. https://doi.org/10.1016/j.epsr.2016.12.012

Biau, G. (2012). Analysis of a random forests model. Journal of Machine Learning Research, 13, 1063-1095.

Box, G. E. P., Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Prentice Hall.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth International Group.

Chaiamarit, K., & Nuchprayoon, S. (2014). Impact assessment of renewable generation on electricity demand characteristics. Renewable & Sustainable Energy Reviews, 39, 995-1004. https://doi.org/10.1016/j.rser.2014.07.102

Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002). SMOTE: Synthetic minority over-sampling technique. The Journal of Artificial Intelligence Research, 16, 321-357. https://doi.org/10.1613/jair.953

Dutta, G., & Mitra, K. (2017). A literature review on dynamic pricing of electricity. The Journal of the Operational Research Society, 68(10), 1131-1145. https://doi.org/10.1057/s41274-016-0149-4

Fan, C., Xiao, F., & Wang, S. (2014). Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Applied Energy, 127, 1-10. https://doi.org/10.1016/j.apenergy.2014.04.016

Freund, Y., & Schapire, R., (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1):119–139.

Freund, Y., & Schapire, R., (1999). A Short Introduction to Boosting. Journal of Japanese Society for Artificial Intelligence, 14(5):771-780, September, 1999. (In Japanese, translation by Naoki Abe.) https://cseweb.ucsd.edu/~yfreund/papers/IntroToBoosting.pdf

Fritz, M., Behringer, M., & Schwarz, H. (2020). LOG-means: Efficiently estimating the number of clusters in large datasets. Proceedings of the VLDB Endowment, 13(11), 2118-2131. https://doi.org/10.14778/3407790.3407813

Gao, Y., Fang, C., & Ruan, Y. (2019). A novel model for the prediction of long-term building energy demand: LSTM with attention layer. IOP Conference Series. Earth and Environmental Science, 294(1), 12033. https://doi.org/10.1088/1755-1315/294/1/012033

Garulli, A., Paoletti, S., & Vicino, A. (2015). Models and techniques for electric load forecasting in the presence of demand response. IEEE Transactions on Control Systems Technology, 23(3), 1087-1097. https://doi.org/10.1109/TCST.2014.2361807

Ganesan, K., Saraiva, J. T., & Bessa, R. J. (2022). Functional model of residential consumption elasticity under dynamic tariffs. Energy and Buildings, 255, 111663. https://doi.org/10.1016/j.enbuild.2021.111663

Hledik, R. (2014). Rediscovering residential demand charges. The Electricity Journal, 27(7), 82-96. https://doi.org/10.1016/j.tej.2014.07.003

Hong, T. (2014). Energy Forecasting: Past, Present, and Future. Foresight: The International Journal of Applied Forecasting, 32, 43-48.

Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. International Journal of Forecasting, 32(3), 914-938. https://doi.org/10.1016/j.ijforecast.2015.11.011

IEA (2019). The Critical Role of Buildings – Report April 2019. International Energy Agency, Paris, https://www.iea.org/reports/the-critical-role-of-buildings [Accessed 20 April 2021]

Ivakhnenko, A. G.; Grigor'evich Lapa, V. (1968). Cybernetics and forecasting techniques. American Elsevier Pub. Co.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: With applications in R. Springer. https://doi.org/10.1007/978-1-4614-7138-7

Khan, Z. A., Hussain, T., Ullah, A., Rho, S., Lee, M., & Baik, S. W. (2020). Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid CNN with a LSTM-AE based framework. Sensors (Basel, Switzerland), 20(5), 1399. https://doi.org/10.3390/s20051399

Kim, Y., Son, H., & Kim, S. (2019). Short term electricity load forecasting for institutional buildings. Energy Reports, 5, 1270-1280. https://doi.org/10.1016/j.egyr.2019.08.086

Kleene, S.C. (1956). Representation of Events in Nerve Nets and Finite Automata. Annals of Mathematics Studies (34). Princeton University Press. pp. 3–41.

Lee, J., Kim, J., & Ko, W. (2019). Day-ahead electric load forecasting for the residential building with a small-size dataset based on a self-organizing map and a stacking ensemble learning method. Applied Sciences, 9(6), 1231. https://doi.org/10.3390/app9061231

Li, Y., Che, J., & Yang, Y. (2018). Subsampled support vector regression ensemble for short term electric load forecasting. Energy (Oxford), 164, 160-170. https://doi.org/10.1016/j.energy.2018.08.169

Li, Y., Gao, W., Ruan, Y., & Ushifusa, Y. (2018). Demand response of customers in kitakyushu smart community project to critical peak pricing of electricity. Energy and Buildings, 168, 251-260. https://doi.org/10.1016/j.enbuild.2018.03.029

Liaw A., & Wiener M. (2002). Classification and Regression by randomForest. R News 2(3), 18-22.

Liu, B., Nowotarski, J., Hong, T., & Weron, R. (2017). Probabilistic load forecasting via quantile regression averaging on sister forecasts. IEEE Transactions on Smart Grid, 8(2), 730-737. https://doi.org/10.1109/TSG.2015.2437877

Loh, W. (2014). Fifty years of classification and regression trees. International Statistical Review, 82(3), 329-348. https://doi.org/10.1111/insr.12016

McLaren, J., Mullendore, S., Gagnon, P., & Laws, N. (2017;2015;). Identifying Potential Markets for Behind-the-Meter Battery Energy Storage: A Survey of U.S. Demand Charges. Golden CO: National Renewable Energy Laboratory, NREL/BR-6A20-68963, https://www.nrel.gov/docs/fy17osti/68963.pdf [Accessed 20 April 2021]

Messenger, R., & Mandell, L. (1972). A modal search technique for predictive nominal scale multivariate analysis. Journal of the American Statistical Association, 67(340), 768-772. https://doi.org/10.1080/01621459.1972.10481290

Moghram, I., & Rahman, S. (1989). Analysis and evaluation of five short-term load forecasting techniques. IEEE Transactions on Power Systems, 4(4), 1484-1491. https://doi.org/10.1109/59.41700

Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). Introduction to time series analysis and forecasting, second edition (2nd ed.). Wiley.

Morgan, J. N., & Sonquist, J. A. (1963). Problems in the analysis of survey data, and a proposal. Journal of the American Statistical Association, 58(302), 415-434. https://doi.org/10.1080/01621459.1963.10500855

NOAA (2020). NOAA National Centers for Environmental Information: Local Climatological Data (LCD). NOAA National Centers for Environmental Information. gov.noaa.ncdc:C00128 Subset used: 1 June 2018 – 29 February 2020, https://www.ncdc.noaa.gov/cdo-web/datatools/lcd [Retrieved 16 March 2020]

Park, L., Jang, Y., Bae, H., Lee, J., Park, C. Y., & Cho, S. (2017). Automated energy scheduling algorithms for residential demand response systems. Energies (Basel), 10(9), 1326. https://doi.org/10.3390/en10091326

Pi, Z. X., Li, X. H., Ding, Y. M., Zhao, M., & Liu, Z. X. (2021). Demand response scheduling algorithm of the economic energy consumption in buildings for considering comfortable working time and user target price. Energy and Buildings, 250, 111252. https://doi.org/10.1016/j.enbuild.2021.111252

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

REN21 (2019). Renewables 2019 Global Status Report. REN21 Secretariat, Paris, ISBN 978-3-9818911-7-1 https://www.ren21.net/wp-content/uploads/2019/05/gsr_2019_full_report_en.pdf [Accessed 20 April 2021]

Ripley B. (2019). Classification and regression trees. https://CRAN.R-project.org/package=tree

Rosenblatt, F. (1957). The Perceptron a perceiving and recognizing automaton. Report 85-460-1. Cornell Aeronautical Laboratory.

Saxena, H., Aponte, O., & McConky, K. T. (2019). A hybrid machine learning model for forecasting a billing period's peak electric load days. International Journal of Forecasting, 35(4), 1288-1303. https://doi.org/10.1016/j.ijforecast.2019.03.025

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. EURASIP Journal on Wireless Communications and Networking, 2021(1), 1-16. https://doi.org/10.1186/s13638-021-01910-w

Siano, P. (2014). Demand response and smart grids—A survey. Renewable & Sustainable Energy Reviews, 30, 461-478. https://doi.org/10.1016/j.rser.2013.10.022

Silva, B. N., Khan, M., & Han, K. (2020). Futuristic sustainable energy management in smart environments: A review of peak load shaving and demand response strategies, challenges, and opportunities. Sustainability (Basel, Switzerland), 12(14), 5561. https://doi.org/10.3390/su12145561

Staffell, I., & Pfenninger, S. (2018). The increasing impact of weather on electricity supply and demand. Energy (Oxford), 145, 65-78. https://doi.org/10.1016/j.energy.2017.12.051

Thorndike, R. L. (1953). Who belongs in the family?. Psychometrika, 18, 267–276. https://doi.org/10.1007/BF02289263

Tushar, M. H. K., Zeineddine, A. W., & Assi, C. (2018). Demand-side management by regulating charging and discharging of the EV, ESS, and utilizing renewable energy. IEEE Transactions on Industrial Informatics, 14(1), 117-126. https://doi.org/10.1109/TII.2017.2755465

USEIA (2020). April 2020 Monthly Energy Review (Table 2.1), United States Energy Information Administration, Washington DC, https://www.eia.gov/totalenergy/data/monthly/archive/00352004.pdf [Accessed 20 April 2021]

Wiener, N. (1949). Extrapolation, interpolation, and smoothing of stationary time series. New York: Wiley. ISBN 978-0-262-73005-1.

Xie, J., & Hong, T. (2016). GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation. International Journal of Forecasting, 32(3), 1012-1016. https://doi.org/10.1016/j.ijforecast.2015.11.005

Xu, L., Wang, S., & Xiao, F. (2019). An adaptive optimal monthly peak building demand limiting strategy considering load uncertainty. Applied Energy, 253, 113582. https://doi.org/10.1016/j.apenergy.2019.113582

Yadav, N., Yadav, A., Kumar, M., & SpringerLink (Online service). (2015). An introduction to neural network methods for differential equations. Springer Netherlands.

Yildiz, B., Bilbao, J. I., & Sproul, A. B. (2017). A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renewable & Sustainable Energy Reviews, 73, 1104-1122. https://doi.org/10.1016/j.rser.2017.02.023

Zhang, Y., & Augenbroe, G. (2018). Optimal demand charge reduction for commercial buildings through a combination of efficiency and flexibility measures. Applied Energy, 221, 180-194. https://doi.org/10.1016/j.apenergy.2018.03.150

Appendix 1

Variables per dataset.

	Variable name	**Description**	**Type**	**Industrial PRE**	**Industrial YR1**	**Educational**	**Residential**
1	Label	Year and month using the format YYYYMM at the time of observation.	Categorical	X	X	X	X
2	Year	Year component of Label at the time of observation. *Categories: 2018, 2019, 2020, and 2021*	Categorical	X	X	X	X
3	Month	Month component of Label at the time of observation. *Categories: 1, 2, 3, …, 12*	Categorical	X	X	X	X
4	Demand_Dlim	Calculated monthly threshold (Dlim) for Demand using Equation 12 in kW.	Continuous	X	X	X	X
5	Demand_Dlim_LKV	Calculated monthly threshold (Dlim) for the same month on the previous year or any other last known value for the month of interest.	Continuous	X	X	X	X
6	MonthMaxTemp_F	Maximum outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
7	DailyMaxTemp_Ave_F	Average daily maximum outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X

8	DailyMaxTemp_Min_F	Minimum daily maximum outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
9	DailyAveTemp_Max_F	Maximum daily average outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
10	MonthAveTemp_F	Average outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
11	DailyAveTemp_Min_F	Minimum daily average outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
12	DailyMinTemp_Max_F	Maximum daily minimum outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
13	DailyMinTemp_Ave_F	Average daily minimum outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
14	MonthMinTemp_F	Minimum outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
15	AmtDoW1	Total number of Mondays in the month of interest.	Continuous	X	X	X	X
16	AmtDoW2	Total number of Tuesdays in the month of interest.	Continuous	X	X	X	X

17	AmtDoW3	Total number of Wednesdays in the month of interest.	Continuous	X	X	X	X
18	AmtDoW4	Total number of Thursdays in the month of interest.	Continuous	X	X	X	X
19	AmtDoW5	Total number of Fridays in the month of interest.	Continuous	X	X	X	X
20	AmtDoW6	Total number of Saturdays in the month of interest.	Continuous	X	X	X	X
21	AmtDoW7	Total number of Sundays in the month of interest.	Continuous	X	X	X	X
22	AmtWeekdays	Total number of weekdays in the month of interest.	Continuous	X	X	X	X
23	Ev_Classes	Total number of days with active classes in the month of interest.	Continuous			X	
24	Ev_ResHallsOpen	Total number of days with residence halls open during the month of interest.	Continuous			X	
25	Ev_CampusOpen	Total number of days with the campus open during the month of interest.	Continuous			X	
26	Ev_HalfDayClass	Total number of days with only half day of classes during the month of interest.	Continuous			X	
27	Ev_HalfDayCampus	Total number of days with only half day of campus open during the month of interest.	Continuous			X	
28	Ev_ExamOrReadDay	Total number of exam or reading days during the month of interest.	Continuous			X	

29	Ev_SpringBreak	Total number of SpringBreak days during the month of interest.	Continuous			X	
30	Ev_FirstDayAfterBreak	Total number of days that are a first day after a break period during the month of interest.	Continuous			X	
31	Ev_CareerFair	Total number of days with career fair during the month of interest.	Continuous			X	
32	Ev_Graduation	Total number of days with graduation during the month of interest.	Continuous			X	
33	Ev_Orientation	Total number of days with orientation during the month of interest.	Continuous			X	
34	Ev_Festival	Total number of days with festival during the month of interest.	Continuous			X	
35	Ev_Homecoming	Total number of days with homecoming during the month of interest.	Continuous			X	
36	Ev_Hackathon	Total number of days with hackathon during the month of interest.	Continuous			X	
37	Ev_EventInc	Total number of days with expected electricity demand increase during the month of interest.	Continuous			X	
38	Ev_EventRed	Total number of days with expected electricity demand decrease during the month of interest.	Continuous			X	
39	Demand_Dlim_PM	Calculated monthly threshold (Dlim) for Demand during the previous month using Equation 12 in kW.	Continuous	X	X	X	X
40	Demand_Dmax_PM	Maximum Demand registered during the previous month in kW.	Continuous	X	X	X	X

41	AvDemand_PM	Average Demand registered during the previous month in kW.	Continuous	X	X	X	X
42	AvHighDemand_PM	Average high daily Demand registered during the previous month in kW.	Continuous	X	X	X	X
43	AvLowDemand_PM	Average low daily Demand registered during the previous month in kW.	Continuous	X	X	X	X
44	MaxSolar_PM	Maximum solar generation registered during the previous month in kW.	Continuous			X	
45	AvSolar_PM	Average solar generation registered during the previous month in kW.	Continuous			X	
46	AvHighSolar_PM	Average maximum daily solar generation registered during the previous month in kW.	Continuous			X	
47	Demand_Dlim_PM2	Calculated monthly threshold (Dlim) for Demand during the second previous month using Equation 12 in kW.	Continuous	X	X	X	X
48	Demand_Dmax_PM2	Maximum Demand registered during the second previous month in kW.	Continuous	X	X	X	X
49	AvDemand_PM2	Average Demand registered during the second previous month in kW.	Continuous	X	X	X	X
50	AvHighDemand_PM2	Average high daily Demand registered during the second previous month in kW.	Continuous	X	X	X	X

51	AvLowDemand_PM2	Average low daily Demand registered during the second previous month in kW.	Continuous	X	X	X	X
52	MaxSolar_PM2	Maximum solar generation registered during the second previous month in kW	Continuous			X	
53	AvSolar_PM2	Average solar generation registered during the second previous month in kW	Continuous			X	
54	AvHighSolar_PM2	Average maximum daily solar generation registered during the second previous month in kW	Continuous			X	
55	Demand_Dlim_PM3	Calculated monthly threshold (Dlim) for Demand during the third previous month as described by Saxena et al. (2019) in kW	Continuous	X	X	X	X
56	Demand_Dmax_PM3	Maximum Demand registered during the third previous month in kW	Continuous	X	X	X	X
57	AvDemand_PM3	Average Demand registered during the third previous month in kW	Continuous	X	X	X	X
58	AvHighDemand_PM3	Average high daily Demand registered during the third previous month in kW	Continuous	X	X	X	X
59	AvLowDemand_PM3	Average low daily Demand registered during the third previous month in kW	Continuous	X	X	X	X
60	MaxSolar_PM3	Maximum solar generation registered during the third previous month in kW	Continuous			X	

61	AvSolar_PM3	Average solar generation registered during the third previous month in kW	Continuous			X	
62	AvHighSolar_PM3	Average maximum daily solar generation registered during the third previous month in kW	Continuous			X	

Appendix 2

Inputs used from Appendix 1 for base threshold value forecasting models.

<table>
<tr><th>Model name</th><th>Industrial
PRE</th><th>Industrial
YR1</th><th>Educational</th><th>Residential</th></tr>
<tr><td>BM01_LKV</td><td colspan="4">Variable 5</td></tr>
<tr><td>BM02_AvPast3</td><td colspan="4">Previous 3 values of Variable 4</td></tr>
<tr><td>BM03_ST</td><td colspan="2" rowspan="2">Variables 2,3, 5-22, 39-43, 47-51, and 55-59</td><td rowspan="2">Variables 2, 3, and 5-62</td><td>(Same as Industrial PRE & YR1)</td></tr>
<tr><td>BM04_RF1000</td><td>(Same as Industrial PRE & YR1)</td></tr>
<tr><td>BM05_ST_BestRed</td><td>Variable 13</td><td>Variable 14</td><td>Variables 5, 7, and 23</td><td>Variables 39, 40, and 47</td></tr>
<tr><td>BM06_RF1000_BestRed</td><td>Variables 3, 5, and 6-14</td><td>Variables 5, 7-14, and 41-43</td><td>Variables 5, 7, 8, 10-14, 21, 23-25, 39-43, 45, 53, and 54</td><td>Variables 6, 9, 11, 12, 39-43, 47, 48, 50, 51, 55, 57, and 58</td></tr>
</table>

Appendix 3

Inputs used from Appendix 1 for ensemble threshold value forecasting models.

<table>
<tr><th>Model name</th><th>Industrial
PRE</th><th>Industrial
YR1</th><th>Educational</th><th>Residential</th></tr>
<tr><td>EM01_Ave</td><td colspan="4">Electricity demand threshold values forecasted by each of the 6 base models (BM01-BM06).</td></tr>
<tr><td>EM02_ST</td><td colspan="2" rowspan="2">Variables 2,3, 6-22, 39-43, 47-51, and 55-59 from Appendix 1, as well as electricity demand threshold values forecasted by each of the 6 base models (BM01-BM06).</td><td rowspan="2">Variables 2,3, and 6-62 from Table Appendix 1, as well as electricity demand threshold values forecasted by each of the 6 base models (BM01-BM06).</td><td rowspan="2">(Same as Industrial PRE & YR1)</td></tr>
<tr><td>EM03_RF1000</td></tr>
<tr><td>EM04_ST_BestRed</td><td>Electricity demand threshold values forecasted by base model BM06</td><td>Electricity demand threshold values forecasted by base model BM06</td><td>Variable 7 from Table Appendix 1, as well as electricity demand threshold values forecasted by base models BM03 and BM06.</td><td>Variables 40 and 55 from Table Appendix 1, as well as electricity demand threshold values forecasted by base model BM06.</td></tr>
<tr><td>EM05_RF1000_BestRed</td><td>Variables 10 and 13 from Table Appendix 1, as well as electricity demand threshold values forecasted by base models BM04 and BM06.</td><td>Electricity demand threshold values forecasted by base model BM06</td><td>Variables 7, 8, 10-14, 23-25, 39-42, and 46 from Table Appendix 1, as well as electricity demand threshold values forecasted by each of the 6 base models (BM01-BM06).</td><td>Variables 39, 40, 42, 51, 55, and 57 from Table Appendix 1, as well as electricity demand threshold values forecasted by base models BM03-BM06.</td></tr>
<tr><td>EM06_ST_BM</td><td colspan="3" rowspan="2">Variable 3 from Table Appendix 1 (Month), as well as electricity demand threshold values forecasted by each of the 6 base models (BM01-BM06).</td><td rowspan="2">Variable 3 from Table Appendix 1 (Month), as well as electricity demand threshold values forecasted by base models BM02-BM06. BM01 values not included because of insufficient data.</td></tr>
<tr><td>EM07_RF1000_BM</td></tr>
</table>

Appendix 4

Variables per dataset for threshold forecasting.

	Variable name	Description	Type	Industrial PRE	Industrial YR1	Educational	Residential
1	Label	Year and month using the format YYYYMM at the time of observation.	Categorical	X	X	X	X
2	Year	Year component of Label at the time of observation. *Categories: 2018, 2019, 2020, and 2021*	Categorical	X	X	X	X
3	Month	Month component of Label at the time of observation. *Categories: 1, 2, 3, …, 12*	Categorical	X	X	X	X
4	Demand_Dlim	Calculated monthly threshold (Dlim) for Demand using Equation 1 in kW.	Continuous	X	X	X	X
5	Demand_Dlim_LKV	Calculated monthly threshold (Dlim) for the same month on the previous year or any other last known value for the month of interest.	Continuous	X	X	X	X
6	MonthMaxTemp_F	Maximum outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
7	DailyMaxTemp_Ave_F	Average daily maximum outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X

8	DailyMaxTemp_Min_F	Minimum daily maximum outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
9	DailyAveTemp_Max_F	Maximum daily average outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
10	MonthAveTemp_F	Average outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
11	DailyAveTemp_Min_F	Minimum daily average outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
12	DailyMinTemp_Max_F	Maximum daily minimum outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
13	DailyMinTemp_Ave_F	Average daily minimum outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
14	MonthMinTemp_F	Minimum outdoor temperature for the month of interest in degrees Fahrenheit (°F) recorded by NOAA.	Continuous	X	X	X	X
15	AmtDoW1	Total number of Mondays in the month of interest.	Continuous	X	X	X	X
16	AmtDoW2	Total number of Tuesdays in the month of interest.	Continuous	X	X	X	X

17	AmtDoW3	Total number of Wednesdays in the month of interest.	Continuous	X	X	X	X
18	AmtDoW4	Total number of Thursdays in the month of interest.	Continuous	X	X	X	X
19	AmtDoW5	Total number of Fridays in the month of interest.	Continuous	X	X	X	X
20	AmtDoW6	Total number of Saturdays in the month of interest.	Continuous	X	X	X	X
21	AmtDoW7	Total number of Sundays in the month of interest.	Continuous	X	X	X	X
22	AmtWeekdays	Total number of weekdays in the month of interest.	Continuous	X	X	X	X
23	Ev_Classes	Total number of days with active classes in the month of interest.	Continuous			X	
24	Ev_ResHallsOpen	Total number of days with residence halls open during the month of interest.	Continuous			X	
25	Ev_CampusOpen	Total number of days with the campus open during the month of interest.	Continuous			X	
26	Ev_HalfDayClass	Total number of days with only half day of classes during the month of interest.	Continuous			X	
27	Ev_HalfDayCampus	Total number of days with only half day of campus open during the month of interest.	Continuous			X	
28	Ev_ExamOrReadDay	Total number of exam or reading days during the month of interest.	Continuous			X	

29	Ev_SpringBreak	Total number of SpringBreak days during the month of interest.	Continuous			X	
30	Ev_FirstDayAfterBreak	Total number of days that are a first day after a break period during the month of interest.	Continuous			X	
31	Ev_CareerFair	Total number of days with career fair during the month of interest.	Continuous			X	
32	Ev_Graduation	Total number of days with graduation during the month of interest.	Continuous			X	
33	Ev_Orientation	Total number of days with orientation during the month of interest.	Continuous			X	
34	Ev_Festival	Total number of days with festival during the month of interest.	Continuous			X	
35	Ev_Homecoming	Total number of days with homecoming during the month of interest.	Continuous			X	
36	Ev_Hackathon	Total number of days with hackathon during the month of interest.	Continuous			X	
37	Ev_EventInc	Total number of days with expected electricity demand increase during the month of interest.	Continuous			X	
38	Ev_EventRed	Total number of days with expected electricity demand decrease during the month of interest.	Continuous			X	
39	Demand_Dlim_PM	Calculated monthly threshold (Dlim) for Demand during the previous month using Equation 1 in kW.	Continuous	X	X	X	X
40	Demand_Dmax_PM	Maximum Demand registered during the previous month in kW.	Continuous	X	X	X	X

41	AvDemand_PM	Average Demand registered during the previous month in kW.	Continuous	X	X	X	X
42	AvHighDemand_PM	Average high daily Demand registered during the previous month in kW.	Continuous	X	X	X	X
43	AvLowDemand_PM	Average low daily Demand registered during the previous month in kW.	Continuous	X	X	X	X
44	MaxSolar_PM	Maximum solar generation registered during the previous month in kW.	Continuous			X	
45	AvSolar_PM	Average solar generation registered during the previous month in kW.	Continuous			X	
46	AvHighSolar_PM	Average maximum daily solar generation registered during the previous month in kW.	Continuous			X	
47	Demand_Dlim_PM2	Calculated monthly threshold (Dlim) for Demand during the second previous month using Equation 1 in kW.	Continuous	X	X	X	X
48	Demand_Dmax_PM2	Maximum Demand registered during the second previous month in kW.	Continuous	X	X	X	X
49	AvDemand_PM2	Average Demand registered during the second previous month in kW.	Continuous	X	X	X	X
50	AvHighDemand_PM2	Average high daily Demand registered during the second previous month in kW.	Continuous	X	X	X	X

51	AvLowDemand_PM2	Average low daily Demand registered during the second previous month in kW.	Continuous	X	X	X	X
52	MaxSolar_PM2	Maximum solar generation registered during the second previous month in kW	Continuous			X	
53	AvSolar_PM2	Average solar generation registered during the second previous month in kW	Continuous			X	
54	AvHighSolar_PM2	Average maximum daily solar generation registered during the second previous month in kW	Continuous			X	
55	Demand_Dlim_PM3	Calculated monthly threshold (Dlim) for Demand during the third previous month as described by Saxena et al. (2019) in kW	Continuous	X	X	X	X
56	Demand_Dmax_PM3	Maximum Demand registered during the third previous month in kW	Continuous	X	X	X	X
57	AvDemand_PM3	Average Demand registered during the third previous month in kW	Continuous	X	X	X	X
58	AvHighDemand_PM3	Average high daily Demand registered during the third previous month in kW	Continuous	X	X	X	X
59	AvLowDemand_PM3	Average low daily Demand registered during the third previous month in kW	Continuous	X	X	X	X
60	MaxSolar_PM3	Maximum solar generation registered during the third previous month in kW	Continuous			X	

61	AvSolar_PM3	Average solar generation registered during the third previous month in kW	Continuous			X	
62	AvHighSolar_PM3	Average maximum daily solar generation registered during the third previous month in kW	Continuous			X	

Appendix 5

Results of paired T-Test of mean difference=0 (vs ≠ 0) for Net Demand - Demand.

Model : Metric	T-Value	P-Value	95% CI for Mean Difference
M01 : MAPE*	7.620	0.000	(2.000, 3.624)
M02 : MAPE*	6.190	0.000	(1.451, 3.052)
M03 : MAPE*	6.780	0.000	(1.650, 3.236)
M04 : MAPE*	4.260	0.001	(1.233, 3.868)
M05 : MAPE*	5.780	0.000	(2.283, 5.089)
M01 : Sensitivity	1.180	0.272	(-0.065, 0.202)
M02 : Sensitivity	-0.790	0.450	(-0.347, 0.169)
M03 : Sensitivity	-0.690	0.510	(-0.113, 0.061)
M04 : Sensitivity	0.150	0.882	(-0.326, 0.372)
M05 : Sensitivity	-2.030	0.077	(-0.613, 0.039)
M01 : Balanced Accuracy	0.610	0.562	(-0.054, 0.092)
M02 : Balanced Accuracy	-0.450	0.665	(-0.153, 0.103)
M03 : Balanced Accuracy	-0.690	0.510	(-0.056, 0.030)
M04 : Balanced Accuracy	-0.020	0.988	(-0.150, 0.148)
M05 : Balanced Accuracy*	-2.510	0.036	(-0.304, -0.013)
M06 : Sensitivity	-1.000	0.347	(-0.367, 0.145)
M07 : Sensitivity	-1.360	0.211	(-0.210, 0.054)

M08 : Sensitivity	-0.520	0.616	(-0.361, 0.228)
M09 : Sensitivity	0.800	0.447	(-0.185, 0.381)
M10 : Sensitivity	0.590	0.573	(-0.060, 0.100)
M11 : Sensitivity	-0.690	0.511	(-0.145, 0.078)
M06 : Balanced Accuracy	-1.000	0.347	(-0.171, 0.068)
M07 : Balanced Accuracy	-1.360	0.211	(-0.105, 0.027)
M08 : Balanced Accuracy	-1.340	0.216	(-0.123, 0.033)
M09 : Balanced Accuracy	0.310	0.768	(-0.102, 0.134)
M10 : Balanced Accuracy	-0.420	0.689	(-0.066, 0.046)
M11 : Balanced Accuracy	0.040	0.968	(-0.041, 0.043)
E01 : Sensitivity	-1.430	0.191	(-0.346, 0.081)
E02 : Sensitivity	X	X	No Difference | Equal Values
E01 : Balanced Accuracy	-1.350	0.214	(-0.168, 0.044)
E02 : Balanced Accuracy	-1.380	0.206	(-0.077, 0.019)

** The Null Hypothesis, mean difference = 0, was rejected for the results obtained by these models.*

Appendix 6

Industrial consumer results.

	Industrial_PRE (Pre-Pandemic)			Industrial_YR1 (Year 1 of pandemic)		
Model	**MAPE**	**%Model Savings**	**Number of False Positives**	**MAPE**	**%Model Savings**	**Number of False Positives**
BM01_LKV	4.440	51.67%	1	11.469	15.22%	0
BM02_AvPast3	9.610	50.49%	53	12.175	57.82%	46
BM03_ST	6.521	35.61%	16	8.531	32.39%	11
BM04_RF1000	2.703	56.30%	2	7.406	38.26%	4
BM05_ST_BestRed	4.363	59.35%	16	7.554	50.64%	8
BM06_RF1000_BestRed	2.551	57.74%	0	6.394	42.93%	1
EM01_Ave	3.157	54.86%	3	7.828	40.21%	8
EM02_ST	6.521	35.61%	16	8.531	32.39%	11
EM03_RF1000	2.695	56.49%	2	6.520	47.44%	7
EM04_ST_BestRed	4.363	59.35%	16	7.117	50.64%	8
EM05_RF1000_BestRed	2.461	63.76%	2	4.572	65.71%	7
EM06_ST_BM	5.305	52.07%	6	9.334	27.49%	8
EM07_RF1000_BM	3.097	56.17%	2	6.497	53.19%	8

Appendix 7

Educational and Residential consumer results.

	Educational			Residential		
Model	**MAPE**	**%Model Savings**	**Number of False Positives**	**MAPE**	**%Model Savings**	**Number of False Positives**
BM01_LKV	20.302	70.56%	121	36.041	76.88%	131
BM02_AvPast3	10.158	87.72%	70	22.089	53.44%	61
BM03_ST	8.113	75.92%	49	19.637	63.52%	77
BM04_RF1000	6.642	83.01%	37	15.293	74.72%	66
BM05_ST_BestRed	7.134	81.72%	45	17.786	56.88%	68
BM06_RF1000_BestRed	5.623	81.85%	28	13.895	70.16%	60
EM01_Ave	8.104	81.60%	53	16.001	73.92%	75
EM02_ST	8.227	72.95%	45	22.759	48.40%	71
EM03_RF1000	6.292	87.73%	41	15.654	69.60%	62
EM04_ST_BestRed	6.990	79.89%	40	20.354	54.08%	69
EM05_RF1000_BestRed	5.541	86.80%	35	15.662	60.48%	58
EM06_ST_BM	8.015	67.93%	38	22.759	48.40%	71
EM07_RF1000_BM	6.989	87.08%	56	17.653	64.72%	71

Appendix 8

Average monthly accuracy, sensitivity and balanced accuracy values; as well as the total number of false positives and false negatives predictions produced by all 15 PELD forecasting models for the Industrial consumer.

Model_Type	**Av. Acc.**	**Av. Sens.**	**Av. BalAcc.**	**Total FNs**	**Total FPs**
Reg_SARIMA	0.867	0.302	0.628	35	14
Reg_ST	0.872	0.074	0.535	46	1
Reg_RF1000	0.867	0.121	0.556	46	3
Reg_ANN	0.850	0.372	0.642	28	27
Reg_ANNST	0.671	0.364	0.512	26	95
Class_ST	0.781	0.742	0.764	5	74
Class_RF1000	1.000	1.000	1.000	0	0
Class_ANN	1.000	1.000	1.000	0	0
Class_ANNST	0.225	0.864	0.499	11	272
Ens_Majority	0.904	0.344	0.659	23	12
Ens_SingleVote	0.142	1.000	0.526	0	313
Ens_ST	1.000	1.000	1.000	0	0
Ens_RF1000	1.000	1.000	1.000	0	0
Ens_ANN	1.000	1.000	1.000	0	0
Ens_AdaBoost	1.000	1.000	1.000	0	0

Appendix 9

Average monthly accuracy, sensitivity and balanced accuracy values; as well as the total number of false positives and false negatives predictions produced by all 15 PELD forecasting models for the Educational consumer.

Model_Type	Av. Acc.	Av. Sens.	Av. BalAcc.	Total FNs	Total FPs
Reg_SARIMA	0.829	0.417	0.629	27	36
Reg_ST	0.878	0.296	0.626	35	10
Reg_RF1000	0.877	0.074	0.537	45	0
Reg_ANN	0.856	0.542	0.695	20	33
Reg_ANNST	0.883	0.398	0.665	28	15
Class_ST	0.824	0.585	0.677	11	54
Class_RF1000	0.832	0.657	0.721	14	48
Class_ANN	0.835	0.898	0.833	6	55
Class_ANNST	0.891	0.944	0.891	3	37
Ens_Majority	0.907	0.301	0.651	34	0
Ens_SingleVote	0.632	1.000	0.705	0	136
Ens_ST	0.897	0.681	0.796	17	21
Ens_RF1000	0.910	0.537	0.741	20	13
Ens_ANN	0.835	0.315	0.615	27	33
Ens_AdaBoost	0.894	0.437	0.697	29	10

Appendix 10

Average monthly accuracy, sensitivity and balanced accuracy values; as well as the total number of false positives and false negatives predictions produced by all 15 PELD forecasting models for the Residential consumer.

Model_Type	Av. Acc.	Av. Sens.	Av. BalAcc.	Total FNs	Total FPs
Reg_SARIMA	0.460	0.175	0.502	172	25
Reg_ST	0.419	0.000	0.495	211	1
Reg_RF1000	0.421	0.000	0.500	211	0
Reg_ANN	0.467	0.175	0.511	173	22
Reg_ANNST	0.502	0.525	0.503	101	81
Class_ST	0.579	1.000	0.500	0	154
Class_RF1000	1.000	1.000	1.000	0	0
Class_ANN	0.717	0.970	0.665	6	98
Class_ANNST	0.568	0.914	0.507	19	139
Ens_Majority	0.758	0.618	0.777	80	9
Ens_SingleVote	0.579	1.000	0.500	0	154
Ens_ST	1.000	1.000	1.000	0	0
Ens_RF1000	1.000	1.000	1.000	0	0
Ens_ANN	0.992	0.996	0.991	1	2
Ens_AdaBoost	1.000	1.000	1.000	0	0