Applied Energy 357 (2024) 122413

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

A machine learning framework to estimate residential electricity demand based on smart meter electricity, climate, building characteristics, and socioeconomic datasets

McKenna Peplinski[a,*], Bistra Dilkina[b], Mo Chen[c], Sam J. Silva[d], George A. Ban-Weiss[a], Kelly T. Sanders[a,*]

[a] Department of Civil and Environmental Engineering, University of Southern California, United States of America
[b] Department of Computer Science, University of Southern California, United States of America
[c] California Air Resources Board, United States of America
[d] Department of Earth Sciences and Department of Civil and Environmental Engineering, University of Southern California, United States of America

HIGHLIGHTS

- A generalized ML framework is developed.
- Smart meter data from ~58,000 homes used to predict residential electricity demand.
- Models also trained with weather, building, and socioeconomic datasets.
- Annual, monthly and daily usage estimated at household and census tract resolutions.
- Feature selection and importance used to improve models and their interpretability.

ARTICLE INFO

Keywords:
Smart meter
Residential electricity
Machine learning
Climate change
Building energy
Energy forecasting

ABSTRACT

Due to the substantial portion of total electricity use attributed to the residential sector and projected rises in demand, anticipating future energy needs in the context of a warming climate will be essential to maintain grid reliability and plan for future infrastructure investments. Machine learning has become a popular tool for forecasting residential electricity demand, but previous studies have been limited by lack of access to high spatiotemporal resolution at a regional scale, which reduces a model's ability to capture the relationship between electricity and its driving factors. In this study, we develop and execute a machine learning framework to predict residential electricity demand at varying temporal and spatial resolutions using hourly smart meter electricity records from roughly 58,000 homes provided by Southern California Edison as well as local weather data, building characteristics, and socioeconomic indicators. The best performing model at the household level, multilayer perceptron (MLP), was able to predict electricity demand most accurately at a monthly resolution, achieving an r^2 of 0.45, while the most accurate annual and daily models (also MLP) had r^2 values of 0.34 and 0.38, respectively. The results also show that models trained with data aggregated to the census tract level were more accurate (e.g., $r^2 = 0.82$ for the monthly MLP model) than at the household level across all three temporal resolutions analyzed. Total square footage and various climate indicators had the highest feature importance values. Square footage was ranked first in feature importance for the annual and daily models, while the month of the year, which is strongly tied to temperature, was most important to the monthly model. Through this analysis we gain insight into factors that drive electricity demand and the usefulness of machine learning for predicting residential electricity use.

* Corresponding authors.
E-mail addresses: peplinsk@usc.edu (M. Peplinski), ktsanders@usc.edu (K.T. Sanders).

https://doi.org/10.1016/j.apenergy.2023.122413
Received 28 September 2022; Received in revised form 30 October 2023; Accepted 21 November 2023
Available online 27 December 2023

1. Introduction

The residential sector is a significant consumer of electricity, accounting for 39% of US total end-use electricity consumption in 2020 [1]. Although per capita electricity consumption flattened in recent years [2], there is an expectation that a warming climate coupled with electrification trends will drive up electricity demand in the future [3–5]. Given the residential sector's significance in overall electricity demand, anticipating future household electricity consumption will be essential to maintaining grid reliability, managing peak demand, and planning for new power capacity investments.

In previous studies analyzing factors driving residential electricity consumption, temperature has been found to play one of the most significant roles [6–8]. Additionally, physical building characteristics (e.g., square footage, insulation, number of stories, number of appliances) [6,9,10], socio-economics, (i.e., occupation, income, education, class) [11,12] and occupant behavior and preferences [6,13–15] are significant factors in influencing electricity demand. While these studies provide some insight into the factors that shape electricity use, the accuracy of residential electricity demand models remain limited by the diverse and complex nature of the residential sector and the data available to capture that diversity, as housing stock can vary significantly both across and within regions according to home size, building materials, appliances, demographics, occupancy patterns, etc.

Residential energy modeling studies can be categorized into two distinct approaches: top-down [16–21] and bottom-up [22–29]. Top-down models rely on aggregate data to establish relationships between variables and energy use and predict energy demand [30]. In top-down studies, historical energy consumption is typically estimated at a city, state, or regional level and regressed against macroeconomic indicators, such GDP or unemployment [16,17], energy prices [18,19], housing stock trends [19,20], or weather variables [19,21]. The focus of many of these analyses is to capture how socioeconomic characteristics impact the electricity sector [31]. For example, one study implemented two statistical methods, ordinary least squares (OLS) and random coefficient (RC), to analyze the relationship between electricity consumption and socioeconomic variables, including per capita GNP, GDP growth, structure of the economy, urbanization, and level of literacy, using data from 93 countries and found that electricity consumption increases with socioeconomic development [32]. Salari and Javid estimated electricity and natural gas demand in 48 U.S. states while considering socioeconomic and demographic variables, building stock characteristics, energy prices, and weather data. The results from three different linear regression techniques, OLS, random effect (RE), and fixed effects (FE), show that the socioeconomic and demographic variables of per capita income, household size, and percentage of residents with a high school degree have a statistically significant impact on the residential energy demand [33]. These top-down approaches are advantageous because of model simplicity and the wide availability of data, but their lack of detail makes it difficult to identify local demand patterns and areas for improvement.

In contrast, bottom-up models use microdata, i.e., highly detailed building and appliance information, from an individual home or subset of homes to estimate energy demand and extrapolate to the region, using either a physics-based [22–25] or statistical approach [26–28]. Physics-based models simulate a region's electricity demand by utilizing a set of building archetypes, which are described based on an extensive selection of possible user-defined input variables, to broadly represent the region's building stock [34,35]. A representative building stock model for Los Angeles County was used to estimate the region's residential electricity and natural gas demand in 2020–2060 under climate change scenarios and energy efficiency trends [36]. The study found that under population growth and temperature increases, the total residential electricity demand for the region could increase by 41–87% between 2020 and 2060. However, the total increase in electricity demand could fall to 28% with aggressive energy efficiency policies. Physics based models are valuable because they describe current and prospective technologies with high detail, including a breakdown of end-use consumption, without requiring private residential electricity records and building-specific info that are often not publicly available. Because simulations depend on physical characteristics and thermodynamic principles, the impact of potential technological combinations and energy efficiency measures can be quantified, and policies that more effectively target consumption can be developed. The drawbacks of physics based models are that many assumptions have to be made regarding behavioral factors and their influence on energy [34], since the models do not rely on historical data, and the building stock of a region must be coarsened to a few types of buildings with estimations made for the number of buildings for each type.

Statistical models, a second type of bottom-up model, use historical data, such as energy bills or smart meter data, from a subset of homes to relate physical building characteristics, climate, and occupancy behavior to energy demand (see [30] for a survey). The benefit of using actual energy data is that the effect of a homeowner's individual behaviors and demographics can be considered, unlike physics-based models which require many assumptions to estimate behavior or top-down methods that apply broad socioeconomic indicators to their model. For example, Min et al. performed linear regression analysis of four different residential end use categories (space heating, water heating, cooling and appliance) to develop a mathematical relationship between energy use and predictor variables, including energy price, household characteristics, housing unit characteristics, regional fixed effects, and heating/cooling degree-days [37]. The regression models were used to estimate residential energy by end use and fuel type for every US zip code and provide an in depth look into how energy use varies across regions. In general, bottom-up models are advantageous because they reveal information about end usage and finer-scale resolution energy patterns and predictions. However, both bottom-up methods have higher complexity and computation time than top-down methods and require detailed input data that are typically not readily available [38].

Machine learning has emerged more recently as a method to forecast energy usage that can address the complexity, dynamics, and nonlinearity of building energy systems without requiring detailed information on the building properties and energy system configurations [39–41]. This approach has been proven effective in fast and accurate forecasting for building energy prediction studies due to its relative simplicity, particularly in comparison to physics based models [42]. Models are trained with historical data to determine the relationship between input parameters (e.g., weather, building characteristics, and socioeconomic data) and building energy consumption [40]. Like linear regression models, machine learning models are data-driven but can be better equipped to model nonlinear and complex patterns [43,44]. Machine learning models are also advantageous because they require less detailed building characteristics than physics based methods, which can be expensive and time consuming to gather and therefore difficult to extrapolate to a larger building stock [45]. Further, studies have shown that machine learning models can forecast energy demand with higher accuracy than linear regression and physics based models [46,47]. While there are advantages of using machine learning models for energy forecasting, several gaps exist in the literature, mainly due to constraints of the available data.

Machine learning models have been used to predict electricity demand for both commercial [48–50] and residential buildings [51–53] as well as for a mixed building stock [54–57] but substantially fewer studies have been conducted for residential buildings than other building types. (See Table 1 for a summary of studies that use machine learning to forecast residential electricity load). The lack of research in the residential sector is most likely due to two limitations: there are less data available from private residences versus commercial or industrial buildings, and residential consumption is highly variable and greatly driven by occupancy patterns that are difficult to model [58–60]. As the

Table 1
A summary of studies that use machine learning to forecast residential electricity load.

Model type	Temporal resolution	Spatial resolution	Number of Buildings	Training features	Region	Citation
SVR	10-min, Hourly, and Daily	Apartment Unit, Floor, Building	1	Weather Data	New York City	Jain et al. 2014 [60]
SVM	Hourly	Building	1	Weather Data, Building Characteristics, Occupant Behavior	France	Paudel et al. 2017 [69]
ANN, SVR, LS-SVM, GPR, GMM	Hourly	Building	4	Weather Data, Building Characteristics	San Antonio, Texas	Dong et al. 2016 [89]
ANN, SVR, GPR, BN	Hourly	Building	4	Weather Data	San Antonio, Texas	Rahman, Srikumar, and Smith 2017 [78]
SVR, MLP, LR	Hourly	Building	782	Weather Data	Ireland	Humeau et al. 2013 [90]
ANN	Hourly, Daily	Building	93	Building Characteristics, Occupant Behavior	Lisbon, Portugal	Rodrigues, Cardeira, and Calado 2014 [91]
ANN, SVM, Classification and Regression Tree, LR, ARIMA, Voting, Bagging, SARIMA-PSO-LSSVR, SARIMA-MetaFA-LSSVR	Daily	Building	1	Weather Data	New Taipei City, Taiwan	Chou and Tran 2018 [62]
SVR	Daily	Building	1, 20, 50	Weather Data, Building Characteristics, Occupant Behavior	France	Zhao and Magoules 2012 [92]
SVM, BPNN, RBFNN, GRNN	Annual	Building	59	Building Characteristics	Guandong, China	Li, Ren, Meng 2010 [93]
ANN, GB, DNN, RF, Stacking, KNN, SVM, DT, LR	Annual	Building	5000	Weather Data, Building Characteristics	UK	Olu-Ajayi et al. 2022 [94]
ElasticNet, Lasso, Ridge, LR, Bagging, RF, GB, Adaboost, Extra Trees	Annual	Zip Code	2246	Building Characteristics	Atlanta	Zhang et al. 2018 [73]
MLR, RF, MNN. GB	Annual	District		Building Characteristics, Socioeconomic Data	London	Gassar, Yun, and Kim 2019 [95]
ElasticNet, Lasso, Ridge, LR, Bagging, RF, GB, Adaboost, Extra Trees, MLP, KNN	Daily, Monthly, Annual	Building, Census Tract	58,537	Weather Data, Building Characteristics, Socioeconomics	Southern California	Our study

number of smart meter installations has increased in recent years, electricity data for residential homes have become more widely accessible and used in a growing number of machine learning studies. For example, one study used hourly consumption data from 6309 individual customers during the 2020 COVID mandates to predict how power consumption patterns could change under a new remote work era using a machine learning framework. The results showed that power consumption increased by 13% in the afternoon due to COVID mandates [61]. However, most existing machine learning studies that use high temporal resolution residential electricity data (i.e. 15-min or hourly intervals) only use data for one or a handful of buildings [62–65] as few studies have had access to high volumes of individual customer smart meter data [66,67].

Very few machine learning electric load forecasting studies have incorporated weather data, physical building characteristics, and socioeconomics together, and those studies that do often use detailed occupant information for a select number of homes that are not publicly available [43,68,69]. Instead of joining multiple datasets to build a diverse feature set, many studies include only the historic electricity data of an individual building to forecast its short-term electricity load [70–72]. Studies that do incorporate a combination of characteristics are often constrained by coarse resolution spatial or temporal data, or vice versa. For example, a study by Zhang et al. used household level information from the Residential Energy Consumption Survey (RECS), Public Use Microdata Survey (PUMS) and American Community Survey (ACS) datasets for ~2000 residential homes, but the study was limited by the course temporal granularity of annual consumption and dataset length of one year [73]. Another study trained various machine learning models with population, building, and weather data from Dubai to investigate the impact of different features on electricity demand, but predictions were made at a monthly, community-wide scale [66].

Past machine learning energy forecasting studies have predicted large scale (e.g. regional or national) energy demand at short [74,75], medium [76–78], and long-term time horizons [68,79,80]. Short-term load forecasts aid daily grid operations such as energy transfers and load dispatch [81], while medium to long-term forecasts are necessary for infrastructure investments and future capacity installments [82]. However, most studies focus on the short-term, only forecasting load up to one day ahead. While there are studies that focus on long-term prediction (e.g. months, years) they most often use data with coarse spatial resolution, such as at a city-wide or countrywide scale [83–85]. Thus, building a more thorough understanding of how input parameters might affect long term demand, especially under changing conditions (e.g., rising temperatures, higher AC adoption rates, growing incomes), is prudent for grid planning over the longer term for aspects such as future grid capacity and storage investments.

The current body of electric load forecasting literature utilizing machine learning has been constrained by limited access to 1) high resolution data that can capture both spatial and temporal variations in energy consumption, 2) statistically representative data at a regional scale, and 3) combinations of weather, physical building, occupancy, and sociodemographic data. To our knowledge, no study has investigated how machine learning models perform under different spatial and temporal resolutions for residential electricity demand projections across entire regions, and because few machine learning studies in this field have used high resolution, regionally representative data with a diverse feature set, there is little insight into how to best optimize these models. To address these research gaps, we ask the following research questions:

1. To what extent can machine learning models accurately predict residential electricity demand with publicly available climate, building, and socioeconomic data?

2. How does the spatiotemporal resolution of historical electricity consumption data impact the ability of machine learning models to make precise predictions of electricity demand?
3. Which features are most useful for predicting the target variable of electricity consumption?

Here we develop a generalized, repeatable framework to predict household-level electricity consumption for the residential sector. We train machine learning models using smart meter electricity records for 58,537 households in the Greater Los Angeles region, as well as feature sets derived from publicly available local site weather, building characteristics, and socioeconomic data. The main contribution of our study is to use household-level smart meter data to capture differences in electricity usage in households across different regions, as well as differences across individual households within regions, to better understand the factors that drive trends in residential electricity consumption. Our study improves upon previous methods of load forecasting by leveraging a diversity of high spatiotemporal resolution datasets at a regional scale, previously unavailable to researchers, to test model efficacy across a selection of ML models, spatiotemporal aggregations, and feature sets.

The framework proposed here can serve as a guide for researchers in the energy domain utilizing ML to estimate residential electricity consumption for a variety of applications. Although our case study is performed in southern California, our framework utilizes standardized smart meter data and publicly available climate, building, and socioeconomic datasets so that it can be repeated in other regions that utilize smart meters. Southern California serves as a valuable case study as it consists of widely varying microclimates with socioeconomically diverse populations and building stocks, making it an ideal location to develop a methodology that can be repeated in cities around the world. The heterogeneity of dataset contributes to this study's novelty, as residential smart meter datasets used for electricity consumption analyses typically represent a more uniform climate, set of buildings, or demographics [6,11,73,86–88]. In an era where electricity reliability will be challenged by a changing climate, trends towards increasing electrification, and massive decarbonization investments, anticipating future demand at more granular resolutions will be important for informing decisions related to infrastructure investment, designing equitable demand response programs, and offsetting the need for additional power plant capacity.

2. Methods

The main objectives of this study are to 1) develop a predictive machine learning model that can be applied to new and changing scenarios (e.g., different regions, climates, and building stocks) to predict residential electricity demand, 2) identify which variables are most useful in predicting residential energy through feature selection and feature importance, and 3) optimize model performance by training models with various combinations of spatial and temporal data resolution. An overview of the methodology is depicted in Fig. 1.

2.1. Datasets

Southern California Edison (SCE), an Investor-Owned Utility (IOU), provided household electricity records for roughly 200,000 customers across Greater Los Angeles. These homes were selected to be statistically representative of the 4.5 million homes that are in the region at a 99% confidence level as described in [96] (note: following the data preparation steps in this analysis, the dataset was no longer statistically representative of the region). Households within the SCE dataset that were located in Orange County, roughly 50,000, were not included in the study as there were no publicly available building property data to match to the records. After the additional data processing steps (described in Section 2.2), the final dataset utilized for our study consisted of 58,537 unique single-family homes. The smart meter data were collected from each household at 15-min intervals over the course of two years from 2015 to 2016 and aggregated to the daily, monthly, and annual level for model training. To conduct this study at high geospatial

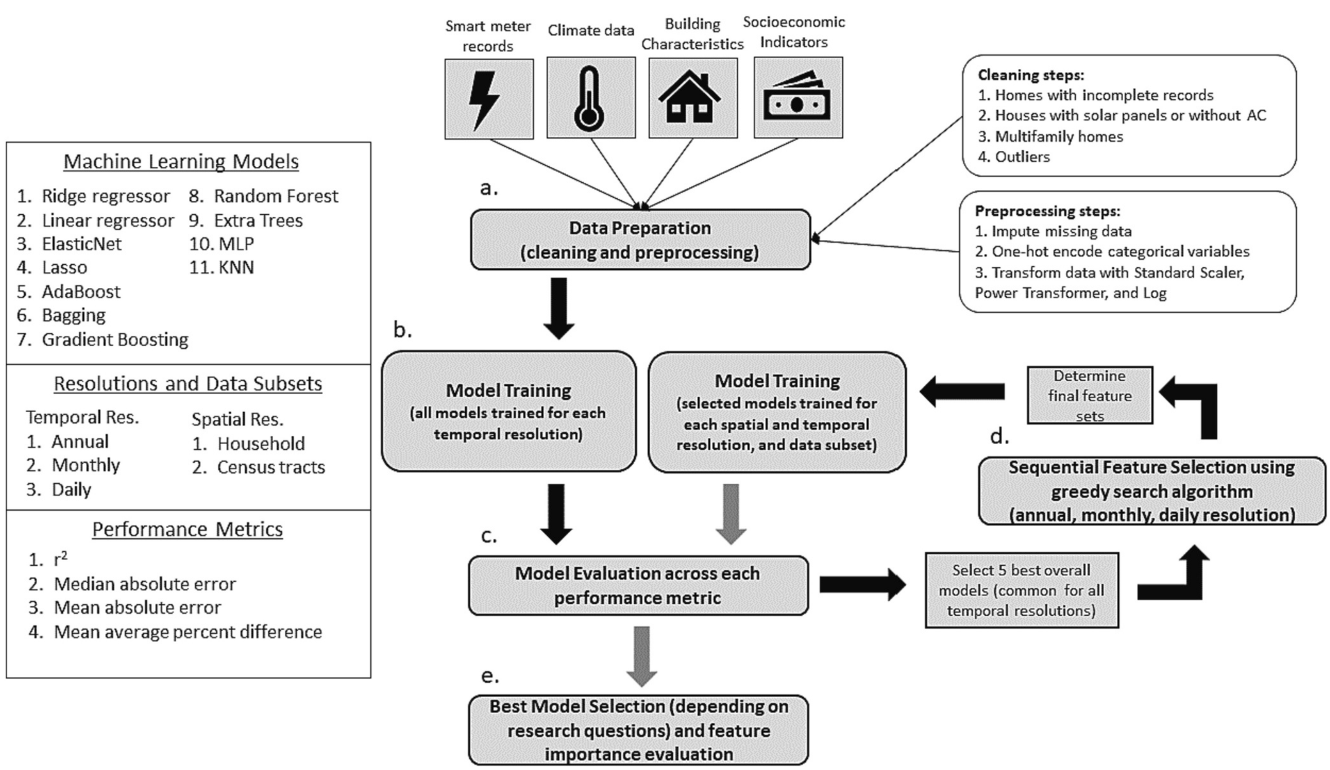

Fig. 1. Machine learning model development framework.

resolution, the street addresses of each home were provided by the utility. Due to the privacy concerns and security requirements of the IOU, the data were stored on the University of Southern California Center for High Performance Computing (HPC) cluster with a highly secure High Security Data Account.

To gain insight into the factors that influence electricity demand, site weather, building characteristics, and socioeconomic data were also obtained. Weather datasets with similar spatiotemporal resolution to the electricity data were necessary to accurately capture energy-climate interactions. Historical weather records were retrieved from two automated weather networks: the California Irrigation Management Information System (CIMIS) and the National Oceanic and Atmospheric Administration's National Centers for Environmental Information (NCEI) [97,98]. Both networks consist of hundreds of automated, land-based stations across California that record hourly observations of climatic indicators such as temperature, precipitation, dew point, and windspeed. For this study, we use only the ambient near-surface air temperature from 36 CIMIS stations and 43 NCEI stations. The stations were selected based on their proximity to the households, with each household being matched to the nearest weather station. The ambient temperature observations were used to calculate cooling degree days (CDD) and heating degree days (HDD). Degree days are a measure of how cold or warm a location is. CDD (HDD) is defined as the daily cumulative number of degrees above (below) a given temperature threshold. This threshold is defined on an application-specific basis. Here, we used 18 degrees Celsius as the threshold (approximately the temperature at which air conditioning (AC) is expected to be needed) to calculate the daily, monthly, and annual CDD and HDD [99]. We also computed a customized metric that we call "extreme cooling degree days" (ECDD) with a threshold of 35 degrees Celsius as an indicator of extreme heat to further differentiate climates.

Various building characteristics for individual households were retrieved from the Property Information Systems database, established by the Office of the Assessor, for Los Angeles, San Bernardino, and Riverside Counties, which were the three counties containing the households analyzed in this study [100–102]. The county databases contain public records for all the properties in each of the three counties including square footage, number of bedrooms and bathrooms, year of construction, address and more, shown in Table 2. To merge datasets, we matched electricity records provided by SCE with each building's physical characteristics using the given street addresses.

Demographic information was collected to explore the role of population characteristics on electricity use. Socioeconomic data were retrieved from CalEnviroScreen 3.0 [103], a mapping tool developed by the Office of Environmental Health Hazard Assessment, on behalf of California Environmental Protection Agency, that identifies which California communities are subjected to higher pollution levels and are often most vulnerable to the effects. CalEnviroScreen includes environmental, health, and socioeconomic information from state and federal government sources for the approximately 8000 census tracts in California. In this study, each individual home within a census tract is matched with the corresponding census indicators. The indicators used in this study are listed in Table 2.

2.2. Data preparation

Data preparation is an important step in machine learning that transforms the raw, collected data into a quality dataset that is more suitable for model training [104]. A few of the standard tasks that are commonly practiced include data cleaning, data transforms, and feature engineering [105–109]. The methods and algorithms used in an ML study depend on the specific dataset and modeling objectives, but broadly, the goal is to better uncover the underlying nature of the data by removing erroneous data and produce a dataset that the desired analysis can be carried out with. Data preparation measures applied to the datasets utilized in this study are outlined in Fig. 1 step a.

Data cleaning is a practice that filters flawed points from a dataset. In some cases, model performance improves by identifying and correcting for outliers and missing values in the data [110]. For this application, we first screened out customers with less than a year of electricity records and homes deemed uninhabited, defined as annual consumption less than 20 kWh, the average daily demand of a home in California [111]. Our analysis targets single family detached homes so electricity customers with an apartment indicator in the address line (e.g., unit

Table 2
Full feature set.

Category	Feature	Type	Mean	Units	Number of Categories
Physical Building Property	Square footage	Continuous	1808	Square feet	
	Bedrooms	Continuous	3.3	Bedrooms	
	Bathrooms	Continuous	3.2	Bathrooms	
	Presence of pool	Binary			
	Building vintage	Continuous	1971		
	Building vintage category	Categorical			3
Climate	Climate zone	Categorical			7
	Annual cooling degree days	Continuous	1293	Degree days	
	Annual heating degree days	Continuous	863	Degree days	
	Annual extreme cooling degree days	Continuous	141	Degree days	
	Monthly cooling degree days	Continuous	101	Degree days	
	Monthly heating degree days	Continuous	66.0	Degree days	
	Monthly extreme cooling degree days	Continuous	9.81	Degree days	
	Monthly average temperature	Continuous	19.1	degrees Celsius	
	Monthly temperature delta	Continuous	11.7	degrees Celsius	
	Daily cooling degree days	Continuous	3.58	Degree days	
	Daily heating degree days	Continuous	2.32	Degree days	
	Daily extreme cooling degree days	Continuous	0.39	Degree days	
	Daily average temperature	Continuous	19.2	degrees Celsius	
	Daily max temperature	Continuous	25.6	degrees Celsius	
	Daily min temperature	Continuous	13.1	degrees Celsius	
	Daily temperature delta	Continuous	12.5	degrees Celsius	
Socioeconomic	Education	Continuous	17.8	Percent	
	Linguistic Isolation	Continuous	8.1	Percent	
	Poverty	Continuous	32.5	Percent	
	Housing Burden	Continuous	17.3	Percent	
	Unemployment	Continuous	10.7	Percent	
Temporal	Month	Categorical			12
	Day of Week	Categorical			7

number) or that were designated as an apartment in the County Assessor databases were removed from the dataset. We adopted the method developed in our previous publication to identify homes with onsite electricity generation (e.g. solar panels) or homes without AC because this information was not provided by the utility [96]. With this method, any home with at least one hour of zero electricity consumption between 10:00 and 16:00 and one or more hours of positive electricity consumption between 17:00 and 23:00 on at least 5% of the days within the two-year time period (i.e., 36 days), was identified as a household with onsite electricity generation [96]. The electricity-temperature sensitivity of a home was also characterized to determine whether a household utilized AC during the period of study based on our framework detailed in [96]. Homes with solar panels and/or without AC were filtered from the data as not to distort the electricity-temperature relationship of the single-family homes remaining in dataset.

Outliers were removed based on the total square footage of the home and electricity demand. The average daily electricity demand per square footage was calculated, and customers with an electricity demand three times greater than the standard deviation for 10% of the time period were filtered out to exclude possible multi-family units or very high consuming households that might skew models. Using the same reasoning, homes with square footage above 20,000 square feet were identified as outliers and removed. The outliers were located throughout the region and not biased towards certain areas. The features in each of the datasets were also processed, and individual variables with more than 10% of records missing were excluded. The number of stories in each building was the only omitted feature across all the originally included features due to the frequency at which it was missing. Table 2 summarizes the features used in the study.

Data preprocessing steps are performed to prepare the raw data for subsequent processing steps. A few basic preprocessing steps include handling missing data, converting formats, and data transformations. For the weather data, numerical imputation was used when hourly weather data were missing to compute degree days and daily average values. Additionally, date formats were converted to match weather data and electricity data. As stated, variables from the county assessor databases were discarded if missing more than 10% of the time. Categorical encoding is a key data preprocessing technique that converts categorical variables to numerical representation so that they are machine readable [112]. In this feature set, the climate zone, presence of a pool, building vintage category, month, and day of week are all unordered, categorical variables that are categorically encoded prior to model training using OneHotEncoder from the Python Scikit-learn library [113].

Data transformations are used to convert a dataset into a format that is more suitable for a given machine learning model. The transformations might be mandatory, meaning that they are necessary for data compatibility, or optional quality transformations, which help the model perform better [114]. Transformations are commonly used to scale and standardize features to the same range so that variables have equal influence [115]. Because there is a large difference in scale across the input variables for this study, the StandardScaler transform from the Python Scikit-learn library was selected to standardize the numerical features by subtracting the mean and scaling to unit variance. This ensures that one feature with high variance does not dominate the rest during training. A PowerTransformer, from the Scikit-learn library, and log transform were also implemented prior to model training, but both reduced model performance and were thus omitted in the final analysis.

2.3. Model training and evaluation

One of the main objectives of this study is to develop an optimized machine learning framework that can predict the electricity demand of individual households using the variables described in Section 2.1. Machine learning models take a set of features, X, as input variables and a target variable, Y. The models build mathematical functions that define Y in terms of X based on the relationships in the training set. Using these functions, target variable predictions are made on a test set based on the corresponding input variables. Machine learning models from varying machine learning model classes, including linear, nonlinear, ensemble, and tree models, were selected to see which models and model types are best suited for this application. In step b of Fig. 1, we trained the following 11 machine learning models from the scikit-learn Python library in our study: ridge regressor, linear regressor, elasticnet regressor, lasso regressor, adaboost regressor, bagging regressor, gradient boosting regressor (XGBoost), random forest regressor (RF regressor), extra trees regressor (ET regressor), multi-layer perceptron regressor (MLP regressor), and k-nearest neighbor regressor (KNN regressor). It is important to note that because the goal of this study is to build a repeatable framework that can be applied to other regions (as opposed to creating an optimal model for our particular dataset and region), we focus on machine learning models that are generalizable and easy to implement. However, a more complex set of models, or combinations of models, might lead to improved model performance. These models are optimized by finding the ideal coefficients, θ, that minimize the sum of losses between each data point and the predicted value calculated by a cost function, L, which varies by model. For each of the selected models, the minimized cost function was mean squared error with some models having added regularization penalties that are built in.

In machine learning, hyperparameters are parameters explicitly defined by the user that control a given model's learning process. The values and configurations for the hyperparameters can be adjusted prior to training in an effort to achieve optimal performance. However, determining the best values is often completed through rule of thumb or trial and error, which are both time intensive. For the scope of this study, the hyperparameters of the models were only slightly adapted from the default settings of the scikit-learn Python library version 0.24.2 in instances where the default settings might cause long run times. For the XGBoost regressor, the max depth was adjusted to 4 and the number of estimators was reduced to 20. Changes made to the RF regressor include setting the max depth to 3 and the number of estimators to 60. Lastly, the max depth for ET regressor was set to 3. For these three models, the rest of the hyperparameters remained as the default. For all other models, all hyperparameters were set to the defaults. Additional hyperparameter tuning could improve model performance, and we do not suggest that these are the optimal settings.

Resampling methods are commonly used in machine learning studies to reduce bias in the training set by repeatedly sampling from the original data [116–119]. The technique is used to avoid overfitting, which happens when a model has learned the training data too well instead of a generalizable relationship. Overfitting results in poor model performance when predicting on new data [120]. In the model training steps shown in Fig. 2, a bootstrap method was implemented in which n number of equally sized subsets are extracted from the dataset with replacement [121]. During training, data leakage can occur when information is shared between the training and testing, leading to unrealistically high levels of measured model performance [122]. To avoid data leakage in this study, the entirety of each of the household's data was included in either the training or test set for each split (i.e., the training and test sets have the entirety of a households two years of data).The model was trained on each of the sampled subsets (training set size of ~90,000, ~1,000,000, ~29,000,000 records, respectively, for the annual, monthly, and daily models) and was evaluated on the test set of remaining data, as illustrated in Fig. 2. To evaluate the models, we set n equal to 10 and recorded the bootstrapped mean score for the four error metrics described below.

For this study, several accuracy metrics were explored including mean absolute error, median absolute error, and r^2 score. The mean absolute error (MAE) and median absolute error (MdAE) are both scale dependent error metrics, meaning the error metrics are expressed in units. MAE measures the average magnitude of the absolute value of

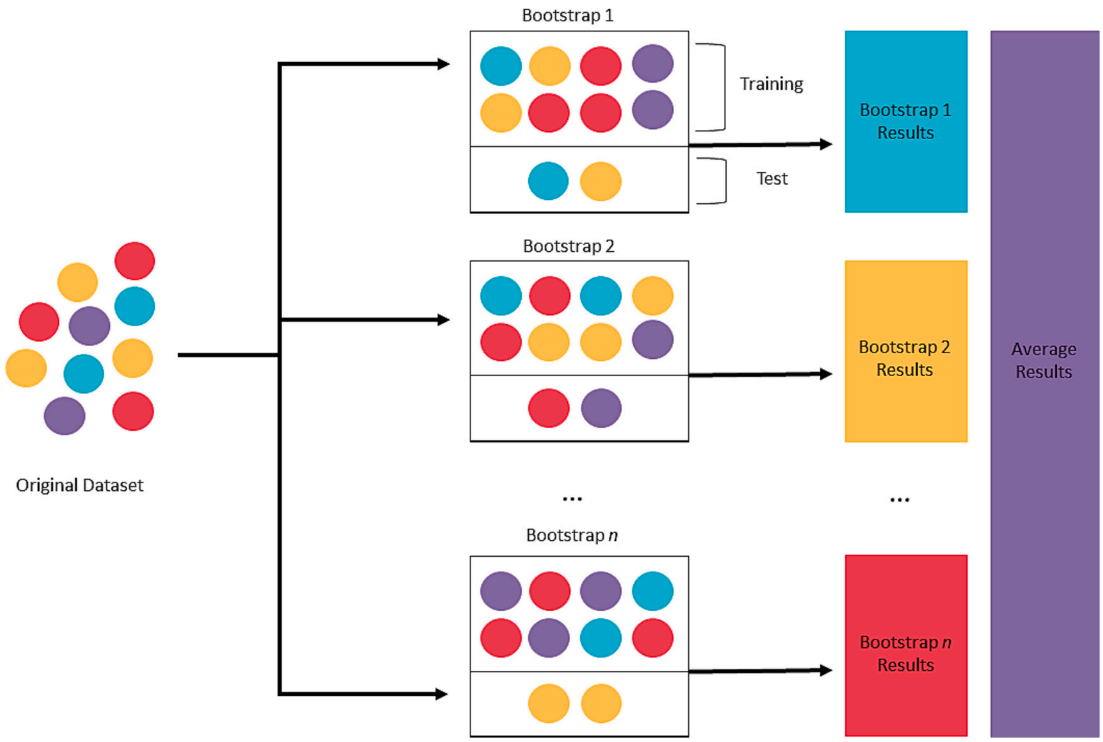

Fig. 2. Graphical representation of bootstrap resampling methods used during model training, where n is equal to 10.

errors in a set of predictions, while MdAE is the median value of all the absolute values of the residuals. Mean absolute percent error (MAPE) is the average difference between the forecasted value and the actual value given as a percentage [123]. The r^2 score, or the coefficient of determination, measures the amount of variance between the samples in the dataset and predictions in the model. The drawback of scale-based metrics for this application is that they cannot be directly compared across temporal resolutions with differing magnitudes of electricity demand. Conversely, percentage-based metrics are flawed because MAPE might be higher for values that tend towards zero (e.g., some daily energy values) and could have different values for two predictions with the same absolute error. Because it does not have these same interpretability limitations, we selected the r^2 metric to assess best model performance in step c of Fig. 1 [124]. The remaining metrics are still reported for completeness.

2.4. Feature selection

Following the data preparation steps, an initial round of model training is performed to determine the overall best models. Feature selection is conducted after the first round of model training on the top five performing models to attain the final feature set (See step d of Fig. 1). Feature selection is the process of identifying and removing redundant or irrelevant variables that are less useful in predicting the target variable. By removing extraneous or redundant features, model performance and computational time for training can both be improved [125–127]. Most commonly used feature selection algorithms can be broadly classified as filter or wrapper methods. Filter methods rank each feature by evaluating the relationship between the input and target variables and then select only the highly ranked features. Wrapper methods select the feature subset that leads to best model performance based on a specified performance indicator [128].

In this study we apply the wrapper method, using the sequential forward selection (SFS) algorithm from the mlxtend library to reduce the d-dimensional dataset into a k-dimensional dataset, where $k < d$ [129]. SFS is a greedy search algorithm in which features are added one at a time until the best feature subset of k features is determined based on the cross-validated r^2 score. In the first iteration, each feature is individually tested and the single feature, x, that leads to best model performance is selected. In the subsequent iteration, every combination of feature x plus an additional feature is tested to determine the two features that in combination achieve the highest performance. Iterations are repeated until a combination of features of size k is found. The value of k can be a specified number or range of numbers. For our model, we set the range as 0 to k, with k equal to the total number of features, to attain the feature set size with the overall best model performance. While feature selection is a valuable algorithm in the machine learning process because of its ability to reduce computational time and improve model accuracy, it does not aid in increasing the interpretability of models. Feature selection can inform hypotheses between features and the target variable, but it does not provide causal understanding for why specific features were selected or discarded from the final feature set.

2.5. Spatiotemporal resolution

To explore the impact of spatiotemporal data resolution on model performance, models were trained with daily, monthly, and annual electricity demand (step b of Fig. 1), and model performance was evaluated for each resolution (step c of Fig. 1). Features with a temporal dimension were averaged (e.g., daily average temperature) or aggregated (e.g., annual CDD) depending on the variable. The ability of our model to predict on larger spatial scales was evaluated using two different methods, illustrated in Fig. 3, to gain insight into how the spatial resolution of the electricity consumption dataset impacts the ability of the model to make accurate predictions. In the first method, referred to as pre-aggregation, the models were trained with household data, and the predicted electricity consumption for all the homes within a census tract was averaged and compared to the true mean of the test set observations. Conversely, in the second method, post-aggregation models were trained and tested with census tract averages of electricity consumption and input variables.

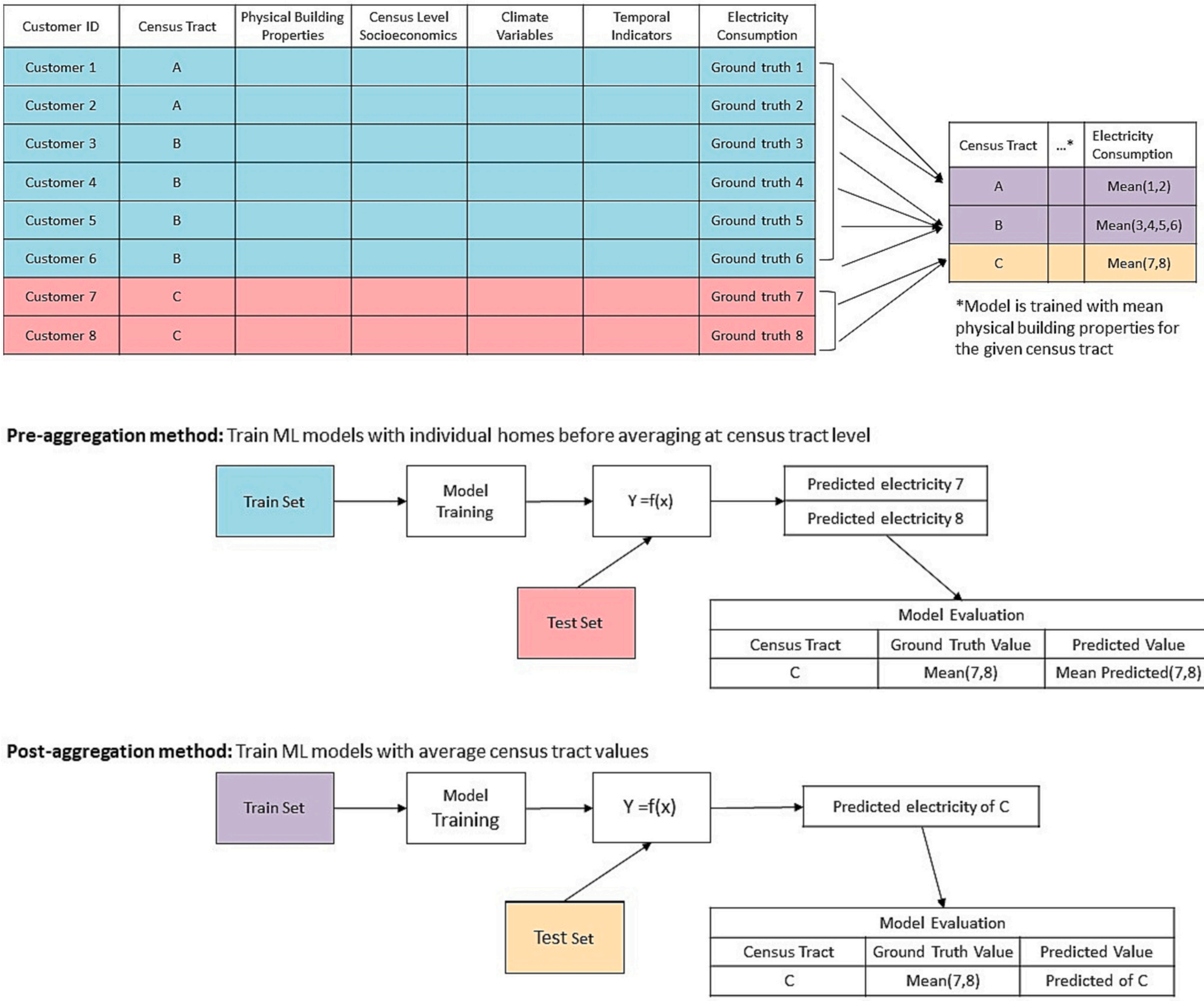

Fig. 3. Two data aggregation methods were executed to test how the spatial resolution of input data impacts a model's ability to predict electricity demand.

2.6. Feature importance

Feature importance techniques are beneficial because they improve the explainability of machine learning models that are often complex and difficult to unpack and reveal relationships between features and target variables [130–132]. There is often overlap between the techniques used for feature importance and feature selection; the key difference is that feature selection is a preprocessing technique that is applied before a model is trained to detect the most relevant features and discard the others. Feature importance algorithms are typically implemented following model training to determine which features are most useful to the model and explain the model behavior [133]. In step d of Fig. 1, we selected the permutation feature importance algorithm from the Python Scikit-learn library, because it has a fast calculation time, is easy to understand, and is applicable for all models in this study [134]. Permutation importance measures the deterioration of model performance after permuting each feature, which effectively breaks the relationships between the feature and the target variable. Because permutation importance is calculated after a model has been fitted, reordering the values of a feature does not impact the relationship learned by the model. The process is as follows: 1) train model, 2) individually shuffle the values of a single variable within the test set and compute the drop in performance score, and 3) return the dataset to the original order and repeat for each of the remaining variables. A feature that significantly impacts the target variable will greatly reduce model performance when shuffled, while one that is less important will have a smaller impact on the accuracy.

3. Results and discussion

The goal of this study was to develop, evaluate, and optimize ML models for residential electricity forecasting. Results from this study gauge the extent to which various machine learning models can accurately predict residential electricity demand with publicly available climate, building, and socioeconomic datasets and at differing spatio-temporal data resolutions. The feature selection and feature importance steps also provide better understanding of the models, giving insight into which features are most useful to make energy demand predictions at various scales.

3.1. Model performance

We wanted to understand and compare a model's ability to predict short-term (e.g., daily) versus longer-term electricity consumption (e.g., annual), as well as household level versus more aggregated scales (e.g., census tracts) of electricity consumption, as projections at each of these various spatiotemporal resolutions offer different insights and utility based on application. Daily household level data has the advantage of capturing day to day variations in energy use among households, which would be important in situations such as understanding the impacts of demand side management strategies or behind-the-meter generation or storage technologies across different populations. For instance, if a utility wanted to anticipate how time-of-use rate structures might affect wealthy versus marginalized communities within their service territories, understanding household level variability on finer timescales would be advantageous. There are other applications in which we might want to understand how electricity use is impacted at broader regional scales or across longer time scales. For example, regional scale forecasts would be most desirable for planning longer term investments in new utility-scale generation capacity.

Table 3 presents the model results for all models across the three temporal resolutions (complete results can be found in the SI). The best performing models and temporal resolutions are consistent across the four different performance metrics (e.g., MLP regressor is the best performing model for each temporal resolution when evaluated by each performance metric). In general, the MAE, MdAE, and MAPE are larger for the annual models than monthly and daily models and larger for monthly models than daily models. The results for MAE and MdAE are intuitive because the yearly electricity demand is greater than the monthly or daily demand and will then likely have larger absolute prediction errors as well. The MAPE results suggest that the models can more accurately predict monthly and daily electricity demand than annual demand.

For this study, the r^2 value was selected as the main indicator of model performance to allow for direct comparisons between the annual, monthly, and daily models. The results showed that prediction accuracy varied significantly across the different ML models and varying temporal resolutions. Fig. 4 summarizes the results of the top five best performing

Table 3
Annual, monthly, daily results before sequential feature selection for top 5 models.

Temporal resolution	Model	Mean absolute error	Median absolute error	r^2	Mean average percent difference
Annual	Ridge Regressor	2780 ± 16.4	2120 ± 13.45	0.31 ± 0.01	112 ± 2.33
	Linear Regressor	2780 ± 16.4	2120 ± 13.48	0.31 ± 0.01	112 ± 2.33
	GB Regressor	2780 ± 15.1	2140 ± 11.05	0.32 ± 0.01	113 ± 2.18
	RF Regressor	2830 ± 14.1	2180 ± 14.30	0.30 ± 0.01	113 ± 2.20
	MLPRegressor	**2740 ± 15.8**	**2090 ± 8.31**	**0.34 ± 0.01**	**111 ± 2.21**
Monthly	Ridge Regressor	250. ± 1.02	184 ± 0.69	0.38 ± 0.0	83.7 ± 9.65
	Linear Regressor	250. ± 1.02	184 ± 0.69	0.38 ± 0.01	83.7 ± 9.65
	GB Regressor	255 ± 0.98	195 ± 1.07	0.36 ± 0.01	89.8 ± 11.0
	RF Regressor	277 ± 2.61	212 ± 1.82	0.25 ± 0.01	94.6 ± 12.1
	MLPRegressor	**235 ± 1.47**	**171 ± 1.52**	**0.45 ± 0.01**	**81.5 ± 9.01**
Daily	Ridge Regressor	9.59 ± 0.0266	6.96 ± 0.0284	0.30 ± 0.007	74.9 ± 1.41
	Linear Regressor	9.04 ± 0.0255	6.47 ± 0.0251	0.37 ± 0.0074	69.9 ± 1.28
	GB Regressor	9.18 ± 0.028	6.81 ± 0.0308	0.35 ± 0.0067	73.9 ± 1.41
	RF Regressor	9.76 ± 0.0297	7.11 ± 0.0401	0.26 ± 0.0042	77.2 ± 1.47
	MLPRegressor	**8.72 ± 0.0508**	**6.13 ± 0.143**	**0.38 ± 0.026**	**67.4 ± 2.32**

models across all three temporal resolutions. Their r^2 values range from 0.25 to 0.45, suggesting that while these models can likely be useful in informing broader trends in residential electricity use, there is still a lot of behavioral variability across individual homes that cannot be captured by the feature set utilized in this study, limiting the models' performance above this r^2 range. It is important to note that there is temporal variation in the model performance (e.g., monthly model performs better in certain months), and a time-series plot of the performance variation is shown in the SI.

The results depicted in Fig. 4 also show how the temporal resolution of data impacts model performance; the r^2 values are slightly higher for all the ML models trained with monthly data, rather than annual or daily, except random forest regressor. The MLP Regressor model trained with monthly electricity data has the highest overall r^2 of 0.45, with the best r^2 for annual and daily data being 0.34 and 0.38 respectively, using MLP Regressor. Monthly models are likely more accurate in predicting the target variable of electricity demand because the monthly data average out some of the highly variable demand seen in the daily data but capture seasonal weather trends more accurately than annual models. While the monthly MLP model has the overall best model performance, in general, the linear models perform with similar accuracy to the non-linear models (e.g., MLP, RFR) when comparing r^2 values.

Direct comparisons of model performance between this study and studies in the literature are limited. This is because studies either 1) utilize household data to make short term predictions for a single household or set of households or 2) use aggregated data to make longer term estimates at larger spatial scales. Typically, these studies have higher model performance as the individual homeowner's behavior is either more easily learned when a single household is used or averaged out in aggregated demand loads. For example, a study that predicted household daily electricity for one home using neural networks had r^2 values ranging from 0.87 to 0.91 [63]. The results of our study are more consistent with the few studies that have access to large samples of household electricity data. Zhang et al. used ML models to predict annual residential electricity demand with r^2 scores ranging from 0.78 to 0.88. While these values are higher than those reported in our study, the household's annual electricity bill was used to predict demand and was shown to be most highly correlated [73]. A study by Williams and Gomez predicted monthly, household residential electricity demand in Texas with three methods: linear regression, regression trees, and multivariate adaptive regression splines and achieved r^2 values ranging from 0.41 to 0.48 [135].

Most prior electricity forecasting work has utilized coarser spatial scale data, which has limited analysis of home-to-home variability. While the smart meter data utilized in this study offer much better spatial resolution, there are applications where household level projections are less desirable than for more coarse regional spatial extents. While acknowledging this, this study is the first to analyze how different techniques for aggregating data can affect the accuracy of ML model performance across large spatial scales. In other words, no study has explored whether high-resolution regional ML model projections will be more accurate if 1) models are trained with household level data that preserve variations in demand, and then aggregate results, or 2) data are first aggregated to the spatial scale at which predictions are being made prior to running the ML model.

Table 4 summarizes the model results of the two different spatial aggregation methods. Again, the results are consistent across the four performance metrics within a specific spatial and temporal resolution combination (e.g., the MLP regressor is the most accurate monthly post-aggregation model regardless of the selected performance metric). For all three temporal resolutions and each evaluation metric, the predictions from the post-aggregation method were overall more accurate than the predictions from the pre-aggregation method.

For direct comparison of annual, monthly, and daily results, the overall best r^2 values for each combination of temporal resolution and spatial aggregation method are highlighted in Fig. 5. The results show

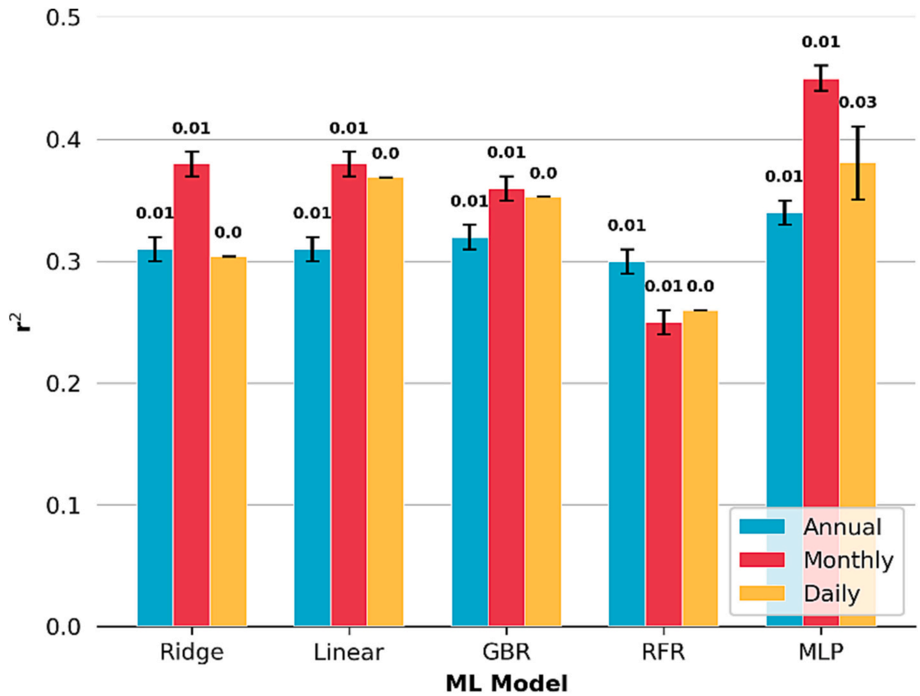

Fig. 4. Model performance measured by r^2 across the overall top 5 best performing models trained with annual, monthly, and daily data. The r^2 values, represented by the bars, range from 0.25 to 0.45, with the monthly MLP model achieving the highest score.

Table 4
Results of the pre-aggregation and post-aggregation training for the top 5 models and all 3 temporal resolutions.

Temporal/spatial resolution	Model	Mean absolute error	Median absolute error	r^2	Mean average percent difference
Annual Pre-aggregation	Ridge Regressor	1430 ± 28.1	1020 ± 24.2	0.47 ± 0.01	30.6 ± 2.60
	Linear Regressor	1430 ± 28.2	1020 ± 24.2	0.47 ± 0.01	30.6 ± 2.60
	GB Regressor	1410 ± 30.3	1030 ± 24.3	0.48 ± 0.01	31.1 ± 2.73
	RF Regressor	1470 ± 33.4	1090 ± 29.1	0.44 ± 0.01	32.0 ± 2.73
	MLP Regressor	**1370 ± 22.7**	**986 ± 20.7**	**0.51 ± 0.01**	**29.8 ± 2.63**
Annual Post-aggregation	Ridge Regressor	831 ± 30.2	618 ± 28.0	0.65 ± 0.04	13.3 ± 0.43
	Linear Regressor	**829 ± 30.5**	**618 ± 28.3**	**0.66 ± 0.04**	**13.3 ± 0.43**
	GB Regressor	824 ± 40.8	600 ± 17.4	0.65 ± 0.04	13.7 ± 0.77
	RF Regressor	888 ± 48.5	638 ± 20.5	0.60 ± 0.04	14.6 ± 0.73
	MLP Regressor	868 ± 45.5	628 ± 27.7	0.63 ± 0.03	13.8 ± 0.67
Monthly Pre-aggregation	Ridge Regressor	135 ± 1.25	97.9 ± 1.94	0.58 ± 0.01	25.9 ± 0.81
	Linear Regressor	135 ± 1.25	97.9 ± 1.94	0.58 ± 0.01	25.9 ± 0.81
	GB Regressor	147 ± 1.37	113 ± 1.37	0.53 ± 0.01	29.4 ± 1.00
	RF Regressor	175 ± 1.82	138 ± 1.88	0.36 ± 0.02	34.5 ± 1.07
	MLP Regressor	**112 ± 1.64**	**78.6 ± 1.44**	**0.70 ± 0.02**	**22.1 ± 1.10**
Monthly Post-aggregation	Ridge Regressor	104 ± 2.02	75.2 ± 1.29	0.68 ± 0.02	18.1 ± 0.44
	Linear Regressor	104 ± 2.02	75.2 ± 1.29	0.68 ± 0.02	18.1 ± 0.44
	GB Regressor	112 ± 3.57	84.7 ± 2.45	0.64 ± 0.02	20.0 ± 0.77
	RF Regressor	144 ± 3.16	106.9 ± 1.49	0.40 ± 0.03	25.4 ± 0.84
	MLP Regressor	**75.2 ± 2.83**	**51.6 ± 1.77**	**0.82 ± 0.02**	**13.5 ± 0.63**
Daily Pre-aggregation	Ridge Regressor	9.16 ± 0.025	6.58 ± 0.025	0.35 ± 0.006	70.7 ± 1.3
	Linear Regressor	9.16 ± 0.025	6.58 ± 0.025	0.35 ± 0.006	70.7 ± 1.3
	GB Regressor	9.18 ± 0.028	6.81 ± 0.031	0.35 ± 0.006	73.9 ± 1.40
	RF Regressor	9.76 ± 0.03	7.11 ± 0.04	0.26 ± 0.004	77.2 ± 1.47
	MLP Regressor	**8.71 ± 0.043**	**6.12 ± 0.125**	**0.38 ± 0.021**	**67.3 ± 2.68**
Daily Post-aggregation	**Ridge Regressor**	**3.60 ± 0.09**	**2.54 ± 0.06**	**0.69 ± 0.02**	**19.2 ± 0.82**
	Linear Regressor	3.60 ± 0.09	2.54 ± 0.06	0.69 ± 0.02	19.2 ± 0.82
	GB Regressor	3.70 ± 0.13	2.67 ± 0.07	0.67 ± 0.02	20.4 ± 0.79
	RF Regressor	4.49 ± 0.12	3.23 ± 0.06	0.53 ± 0.01	24.4 ± 0.76
	MLP Regressor	4.08 ± 0.11	2.98 ± 0.10	0.62 ± 0.03	22.6 ± 0.95

that for models being trained with all three temporal resolutions of data, predictions made at the census tract level with both methods are more accurate than at the household level. Models trained with monthly data again perform better than models trained with annual or daily data for both aggregation methods, with the highest r^2 value being 0.82 for the MLP Regressor trained with the post-aggregation method. The r^2 values for annual and daily models using the post-aggregation method are 0.66 and 0.69, respectively. The higher performance of the post-aggregation method suggests that, when assessing prediction accuracy for aggregated data, a model originally trained and optimized on data after aggregation will perform better than one that is trained prior to data aggregation. This means that household level data are not necessary for improving the accuracy of residential electricity projections for coarser spatial scales, such as the census tract level.

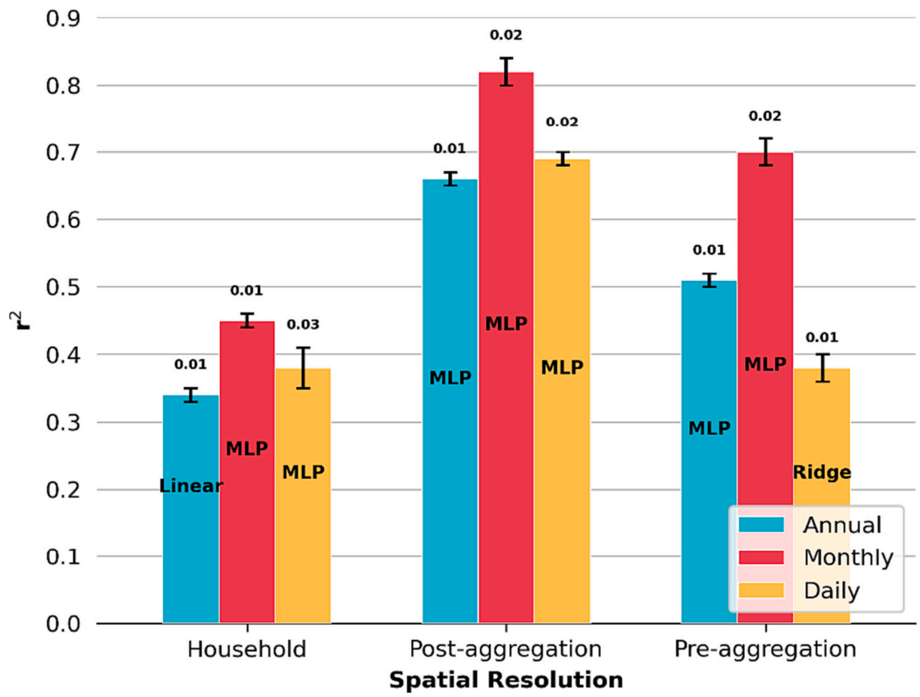

Fig. 5. Model performance of the best models for each combination of temporal resolution and spatial aggregation method measured by r^2. The r^2 values, represented by the bars, range from 0.34 to 0.81, with the monthly, post-aggregation MLP model being the most accurate.

3.2. Sequential feature selection

After the initial round of model training, feature selection was completed to find the most relevant subset of features. Sequential feature selection is commonly implemented in ML studies both to optimize model performance and to cut down on run times by reducing the feature set. The results from the sequential feature selection algorithm are shown in Fig. 6 for annual, monthly, and daily data resolutions. Certain variables were consistently selected across all models and temporal resolutions, such as the home's square footage and whether a home has a pool. Climate related features, such as the climate zone the house belongs to, which month it is, or differing temperature indicators, were also frequently selected. At the annual level, ECDD was selected for all the top five models except one, while CDD and HDD were selected for three and two of the models, respectively. For monthly models, which month it is (a proxy for weather) was selected for all models, and ECDD and HDD were selected for all but one. The CDD and monthly mean temperature were only kept in the final feature set for two of the top five monthly models. Lastly, the month variable and daily max temperature were selected by feature selection for all models at the daily level, followed by CDD, ECDD, daily mean temperature, and daily min temperature, which were selected for all but one model.

Socioeconomic indicators, including education, linguistic isolation, poverty, housing burden, and unemployment, were selected less frequently. These variables reflect census tract-level.

data so they are imprecise indicators in characterizing house to house variability. Across all three temporal resolutions, education and linguistic isolation were the two socioeconomic indicators that were most often kept in the final feature set. These features are often correlated to household financial insecurity, which impacts electricity usage. However, all the demographic variables are highly correlated, making it difficult to tease out their individual influences on energy behavior or determine why one socioeconomic indicator is more useful to model training than another.

3.3. Feature importance

After training the models, permutation feature importance was used to examine which of the features were most useful to predicting electricity consumption. The results of the algorithm showed a decrease in model score (here, r^2) when the records of a specific feature are randomly shuffled within the dataset, breaking the relationship between the feature and the target and revealing how much the model depends on that feature. Table 5 shows the permutation feature importance of all the features in the final feature set for household level data with the top performing ML model, which was MLP Regressor for annual, monthly, and daily trained models. Total square footage was consistently one of the most useful features to the model, ranking first for annual and daily, with feature importance values of 0.402 and 0.232 respectively, and second for monthly with a value of 0.272. The annual and daily feature importance values for square footage are an order of magnitude higher than any of the other annual or daily feature values which suggests that the rest of the variables do not matter much for these prediction cases.

These results show that feature importance varies significantly depending on the temporal resolution. In general, weather indicators were much more important to the monthly model than the daily and annual models. For example, the month of the year, which is strongly tied to temperature, was the most important feature for the monthly model and had a value of 0.336. Conversely, the month of the year is ranked second overall for the daily model, but with a much lower mean importance of 0.087. The low importance of weather features for the daily model could be because there are so many other uncaptured, highly variable daily occupancy factors that outweigh the impact of weather. Similarly, ECDD was the highest ranked climate indicator for the annual model with a low mean importance of 0.059. Because the model was only trained with two years of data, there might not have been enough variation in annual degree days for the model to learn, leaving building characteristics to be more useful in predicting annual electricity use.

The results show that socioeconomic indicators are generally less useful to the model than climate and building characteristics, ranking low in importance across each of the temporal resolutions. Of the five

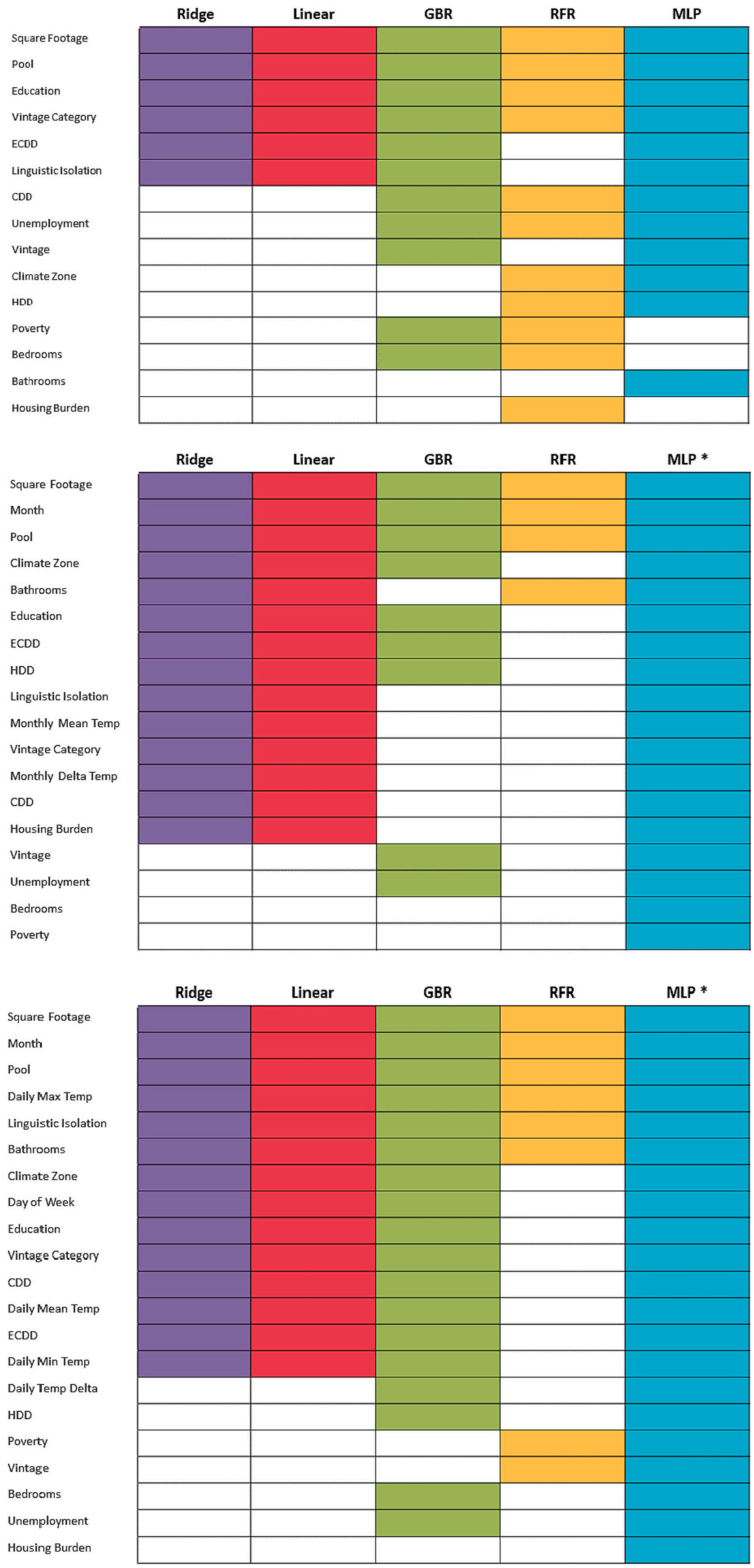

Fig. 6. Final feature set for top performing annual, monthly, and daily models after performing sequential feature selection. The features selected by the algorithm are filled in with color. *Feature selection algorithm was not completed for these combinations of model/temporal resolution because it was too computationally expensive.

Table 5
Feature importance with annual, monthly, and daily data by household.

ANNUAL MLP			MONTHLY MLP			DAILY MLP		
Feature	Mean importance	Std deviation	Feature	Mean importance	Std deviation	Feature	Mean importance	Std deviation
Total Sqft	0.402	0.004	Month	0.336	0.001	Total Square Feet	0.232	0.049
Pool	0.075	0.002	Total Sqft	0.272	0.001	Month	0.087	0.019
ECDD	0.059	0.001	Climate Zone	0.081	0.001	Pool	0.025	0.008
CDD	0.053	0.002	Pool	0.063	0.001	Daily Max Temp	0.022	0.017
Climate Zone	0.042	0.002	Linguistic Isolation	0.028	0.001	Linguistic Isolation	0.015	0.009
Vintage	0.023	0.001	Bathrooms	0.026	0.001	Bathrooms	0.013	0.010
Linguistic Isolation	0.018	0.001	Vintage	0.026	0.000	Climate Zone	0.010	0.008
Bathrooms	0.011	0.001	Monthly Mean Temp	0.021	0.000	Day of Week	0.004	0.003
Education	0.007	0.000	Education	0.020	0.001	Education	0.004	0.008
HDD	0.006	0.000	Bedrooms	0.016	0.000	Vintage Category	0.003	0.003
Vintage Category	0.004	0.000	Poverty	0.015	0.000	CDD	0.003	0.007
Unemployment	0.001	0.000	Vintage Category	0.011	0.001	Daily Mean Temp	0.000	0.000
			Monthly Delta Temp	0.010	0.000	ECDD	0.000	0.000
			ECDD	0.009	0.000	Daily Min Temp	0.000	0.000
			HDD	0.007	0.000	Daily Temp Delta	0.000	0.000
			CDD	0.006	0.000	HDD	0.000	0.000
			Unemployment	0.005	0.000	Poverty	0.000	0.000
			Housing Burden	0.003	0.000	Vintage	0.000	0.000
						Bedrooms	0.000	0.000
						Unemployment	0.000	0.000
						Housing Burden	0.000	0.000

socioeconomic indicators, linguistic isolation is shown to have the highest importance for annual, monthly, and daily models, with values of 0.018, 0.028, and 0.015 respectively. Since socioeconomic variables are only available at the census tract level, it follows that they would not be the best predictors for household electricity demand.

In comparing feature selection and feature importance, the features that were most frequently selected also had consistently higher values for feature importance. Accordingly, those that were not often included in the final feature set typically had lower values of importance in the instances that they were included. For example, the square footage and pool ownership variables were selected for every model and temporal resolution, and their mean importances also ranked in the top three for all three temporal resolutions.

4. Conclusion and future work

Machine learning models are capable of learning highly complex relationships between electricity demand and its driving factors, making them a promising tool for energy load forecasting. To date, studies utilizing ML models to predict residential electricity demand at a regional scale have only had access to coarse spatial (e.g., city, state, regional) and temporal (e.g., monthly or annual) electricity data.

The results show that ML models can predict household level electricity demand with a significant degree of accuracy in certain cases; the best performing model, MLP regressor trained with monthly data, achieves an r^2 value of 0.45. Monthly trained models may have superior performance to annual and daily models because some of the highly variable day to day differences in energy demand behavior are averaged out while still providing a greater distribution of training data than the annual model. Across all temporal resolutions, models predicted census tract level residential electricity demand with higher accuracy than for individual households. Using the post-aggregation training method for an MLP regressor model trained with monthly data, the mean electricity demand of census tracts was predicted with an accuracy of an r^2 of 0.82exi. These results are promising because they show that residential electricity demand can be predicted at relatively high-resolution spatial scales without needing private customer electricity data and can provide insight into patterns of energy demand which are necessary to understand for daily grid operation and future infrastructure investments.

The total square footage of a building as well as climate indicators were consistently selected to be in the final feature set across all the models. These features typically were found to be most important by the feature importance algorithm; total square footage, for example was ranked first for the annual and daily models and second for the monthly model. Socioeconomic indicators did not rank as high but because they were reported at the census tract level, it is harder to determine their influence on household demand. As this study serves as a framework for future grid modeling studies, feature selection and feature importance results can also give insight into where data retrieval efforts should be focused.

Certain limitations cap the extent to which predictions of residential electricity demand can be made. First, individual behavior of homeowners is highly variable and unpredictable and can vastly impact electricity demand. The socioeconomic data serves as a proxy to relate the occupants to their possible energy behavior, but it is not informative enough to account for many of their decisions pertaining to electricity use. As information about the demographics of homeowners is highly private and occupancy patterns would be almost impossible to extract, it would be difficult to surpass this limitation. Second, while the data is regionally representative, it is not necessarily representative of the conditions that are the focus of the study. For example, the annual models only have two years of data that may have, due to external factors, been easier or more difficult for models to predict on rather than other years.

The knowledge gained from this study can serve as a reference to optimize building energy prediction studies, which are crucial to anticipate future energy needs and develop climate adaption and mitigation plans. Researchers can build off the framework presented in this study and improve model performance through a number of techniques such as tuning models' hyperparameters, using a combination of models based on the results of a more granular performance assessment, and employing more complex ML models. Future work will incorporate highly resolved estimates of future temperature across the region of study into the optimized ML model to investigate how residential electricity demand might change due to urban warming. Under a warming climate, the distribution of temperatures, and any other weather data used in future studies, will be fundamentally different than the historical data that is available for model training. Building properties and socioeconomics will also shift, meaning the models will be trained with a feature distribution that no longer exists. The inconsistency between the feature set for training and real-world data will be a limitation for future studies as models will have to both interpolate and extrapolate well to make accurate predictions of electricity demand decades into the future.

Funding

The funding sources that supported this work include:

- CBET-CAREER 1845931
- CBET-CAREER 1752522
- Viterbi School of Engineering Graduate Student Fellowship

CRediT authorship contribution statement

McKenna Peplinski: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Bistra Dilkina:** Writing – review & editing, Methodology. **Mo Chen:** Methodology, Data curation, Conceptualization. **Sam J. Silva:** Writing – review & editing. **George A. Ban-Weiss:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Kelly T. Sanders:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

This work was funded in part by the National Science Foundation under grants CAREER Grant CBET-1752522 and CAREER Grant 1845931 and an internal fellowship from USC Viterbi School of Engineering. Dr. Bistra Dilkina was supported by funds allocated to her as the Dr. Allen and Charlotte Ginsburg Early Career Chair in Computer Science. Computation for the work described in this paper was supported by the University of Southern California's Center for High-Performance Computing (hpcc.usc.edu). We also thank Southern California Edison for access to the smart meter data. Finally, we would like to acknowledge the meaningful contributions to this work by the late Professor George Ban-Weiss, a great mentor and colleague who is deeply missed.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apenergy.2023.122413.

References

[1] Electric power annual: Table 1.2. summary statistics for the United States, 2010–2020 [Online]. Available, https://www.eia.gov/electricity/annual/html/epa_01_02.html; 2023.
[2] Per capita U.S. residential electricity use was flat in 2020, but varied by state. https://www.eia.gov/todayinenergy/detail.php?id=49036; 2021.
[3] Franco G, Sanstad AH. Climate change and electricity demand in California. Clim Change 2007;87(1 SUPPL):139–51. https://doi.org/10.1007/s10584-007-9364-y.
[4] Eskeland GS, Mideksa TK. Electricity demand in a changing climate. Mitig Adapt Strat Glob Chang 2010;15(8):877–97. https://doi.org/10.1007/s11027-010-9246-x.
[5] Sugiyama M. Climate change mitigation and electrification. Energy Policy May 2012;44:464–8. https://doi.org/10.1016/j.enpol.2012.01.028.
[6] Kavousian A, Rajagopal R, Fischer M. Determinants of residential electricity consumption: using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. Energy Jun. 2013;55:184–94. https://doi.org/10.1016/j.energy.2013.03.086.
[7] Psiloglou BE, Giannakopoulos C, Majithia S, Petrakis M. Factors affecting electricity demand in Athens, Greece and London, UK: a comparative assessment. Energy Nov. 2009;34(11):1855–63. https://doi.org/10.1016/j.energy.2009.07.033.
[8] Lam JC. Climatic and economic influences on residential electricity consumption. Energ Conver Manage 1998;39(7):623–9. https://doi.org/10.1016/S0196-8904(97)10008-5.
[9] Ahmad T, et al. Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment. Energy 2018;158:17–32. https://doi.org/10.1016/j.energy.2018.05.169.
[10] Bartusch C, Odlare M, Wallin F, Wester L. Exploring variance in residential electricity consumption: household features and building properties. Appl Energy Apr. 2012;92:637–43. https://doi.org/10.1016/j.apenergy.2011.04.034.
[11] McKenna E, et al. Explaining daily energy demand in British housing using linked smart meter and socio-technical data in a bottom-up statistical model. Energ Buildings 2022;258:111845. https://doi.org/10.1016/j.enbuild.2022.111845.
[12] Huebner G, Shipworth D, Hamilton I, Chalabi Z, Oreszczyn T. Understanding electricity consumption: a comparative contribution of building factors, socio-demographics, appliances, behaviours and attitudes. Appl Energy 2016;177. https://doi.org/10.1016/j.apenergy.2016.04.075.
[13] Dietz T, Gardner GT, Gilligan J, Stern PC, Vandenbergh MP. Household actions can provide a behavioral wedge to rapidly reduce US carbon emissions. Proc Natl Acad Sci U S A 2009;106(44):18452–6. https://doi.org/10.1073/pnas.0908738106.
[14] Schot J, Kanger L, Verbong G. The roles of users in shaping transitions to new energy systems. Nat Energy 2016;1(May). https://doi.org/10.1038/nenergy.2016.54.
[15] Yan D, et al. Occupant behavior modeling for building performance simulation: current state and future challenges. Energ Buildings 2015;107. https://doi.org/10.1016/j.enbuild.2015.08.032.
[16] Vojtovic S, Stundziene A, Kontautiene R. The impact of socio-economic indicators on sustainable consumption of domestic electricity in Lithuania. Sustain. 2018;10(2). https://doi.org/10.3390/su10020162.
[17] Ziramba E. The demand for residential electricity in South Africa. Energy Policy 2008;36(9):3460–6. https://doi.org/10.1016/j.enpol.2008.05.026.
[18] Alberini A, Filippini M. Response of residential electricity demand to price: the effect of measurement error. Energy Econ 2011;33(5):889–95. https://doi.org/10.1016/j.eneco.2011.03.009.
[19] Dergiades T, Tsoulfidis L. Estimating residential demand for electricity in the United States, 1965-2006. Energy Econ 2008;30(5):2722–30. https://doi.org/10.1016/j.eneco.2008.05.005.
[20] Wiesmann D, Lima Azevedo I, Ferrão P, Fernández JE. Residential electricity consumption in Portugal: findings from top-down and bottom-up models. Energy Policy 2011;39(5):2772–9. https://doi.org/10.1016/j.enpol.2011.02.047.
[21] Hor CL, Watson SJ, Majithia S. Analyzing the impact of weather variables on monthly electricity demand. IEEE Trans Power Syst 2005;20(4). https://doi.org/10.1109/TPWRS.2005.857397.
[22] Huang J, Akbari H, Rainer L. Lawrence Berkeley National Laboratory Recent Work Title 481 Prototypical Commercial Buildings for 20 Urban Market Areas Permalink [Online]. Available: https://escholarship.org/uc/item/1g90f5gj; 1991.
[23] Opitz MW, Norford LK, Matrosov YA, Butovsky IN. Energy consumption and conservation in the Russian apartment building stock. Energ Buildings 1997;25(1):75–92. https://doi.org/10.1016/s0378-7788(96)00995-4.
[24] Charlier D, Risch A. Evaluation of the impact of environmental public policy measures on energy consumption and greenhouse gas emissions in the French residential sector. Energy Policy 2012;46(2012):170–84. https://doi.org/10.1016/j.enpol.2012.03.048.
[25] Shimoda Y, Fujii T, Morikawa T, Mizuno M. Residential end-use energy simulation at city scale. Build Environ 2004;39(8 SPEC. ISS):959–67. https://doi.org/10.1016/j.buildenv.2004.01.020.
[26] Lopes MAR, Antunes CH, Martins N. Towards more effective behavioural energy policy: an integrative modelling approach to residential energy consumption in Europe. Energy Res Soc Sci 2015;7:84–98. https://doi.org/10.1016/j.erss.2015.03.004.
[27] Pukšec T, Mathiesen BV, Novosel T, Duić N. Assessing the impact of energy saving measures on the future energy demand and related GHG (greenhouse gas) emission reduction of Croatia. Energy 2014;76:198–209. https://doi.org/10.1016/j.energy.2014.06.045.
[28] Gouveia JP, Fortes P, Seixas J. Projections of energy services demand for residential buildings: insights from a bottom-up methodology. Energy 2012;47(1):430–42. https://doi.org/10.1016/j.energy.2012.09.042.
[29] Ferrando M, Causone F, Hong T, Chen Y. Urban building energy modeling (UBEM) tools: a state-of-the-art review of bottom-up physics-based approaches. Sustain Cities Soc 2020;62. https://doi.org/10.1016/j.scs.2020.102408.
[30] Swan LG, Ugursal VI. Modeling of end-use energy consumption in the residential sector: a review of modeling techniques. Renew Sustain Energy Rev Oct. 01, 2009;13(8. Pergamon):1819–35. https://doi.org/10.1016/j.rser.2008.09.033.
[31] Gul M, Qa SA. Incorporating g economic and demo ographic variablesfor fore casting electricity consumption in Pakistan. 2023.

[32] Burney NA. Socioeconomic development and electricity consumption A cross-country analysis using the random coefficient method. Energy Econ 1995;17(3): 185–95. https://doi.org/10.1016/0140-9883(95)00012-J.
[33] Salari M, Javid RJ. Residential energy demand in the United States: analysis using static and dynamic approaches. Energy Policy 2016;98:637–49. https://doi.org/10.1016/j.enpol.2016.09.041.
[34] Kavgic M, Mavrogianni A, Mumovic D, Summerfield A, Stevanovic Z, Djurovic-Petrovic M. A review of bottom-up building stock models for energy consumption in the residential sector. Build Environ 2010;45(7):1683–97. https://doi.org/10.1016/j.buildenv.2010.01.021.
[35] Uihlein A, Eder P. Policy options towards an energy efficient residential building stock in the EU-27. Energ Buildings 2010;42(6):791–8. https://doi.org/10.1016/j.enbuild.2009.11.016.
[36] Reyna JL, Chester MV. Energy efficiency to reduce residential electricity and natural gas use under climate change. Nat Commun 2017;8(May):1–12. https://doi.org/10.1038/ncomms14916.
[37] Min J, Hausfather Z, Lin QF. A high-resolution statistical model of residential energy end use characteristics for the United States. J Ind Ecol 2010;14(5): 791–807. https://doi.org/10.1111/j.1530-9290.2010.00279.x.
[38] Zhao HX, Magoulès F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16(6):3586–92. https://doi.org/10.1016/j.rser.2012.02.049.
[39] Zhang L, et al. A review of machine learning in building load prediction. Appl Energy 2021;285(July 2020):116452. https://doi.org/10.1016/j.apenergy.2021.116452.
[40] Amasyali K, El-gohary NM. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev March 2017;81:1192–205. https://doi.org/10.1016/j.rser.2017.04.095.
[41] Bourdeau M, Zhai X Qiang, Nefzaoui E, Guo X, Chatellier P. Modeling and forecasting building energy consumption: a review of data-driven techniques. Sustain Cities Soc 2019;48. https://doi.org/10.1016/j.scs.2019.101533.
[42] Seyedzadeh S, Rahimian FP, Glesk I, Roper M. Machine learning for estimation of building energy consumption and performance: a review. Vis Eng 2018;6(1). https://doi.org/10.1186/s40327-018-0064-7.
[43] Robinson C, et al. Machine learning approaches for estimating commercial building energy consumption. Appl Energy 2017;208(August):889–904. https://doi.org/10.1016/j.apenergy.2017.09.060.
[44] Yildiz B, Bilbao JI, Sproul AB. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renew Sustain Energy Rev 2017;73(February):1104–22. https://doi.org/10.1016/j.rser.2017.02.023.
[45] Foucquier A, Robert S, Suard F, Stéphan L, Jay A. State of the art in building modelling and energy performances prediction: a review. Renew Sustain Energy Rev 2013;23:272–88. https://doi.org/10.1016/j.rser.2013.03.004.
[46] Neto AH, Fiorelli FAS. Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energ Buildings 2008;40(12):2169–76. https://doi.org/10.1016/j.enbuild.2008.06.013.
[47] Turhan C, Kazanasmaz T, Uygun IE, Ekmen KE, Akkurt GG. Comparative study of a building energy performance software (KEP-IYTE-ESS) and ANN-based building heat load estimation. Energ Buildings 2014;85:115–25. https://doi.org/10.1016/j.enbuild.2014.09.026.
[48] Jing R, Wang M, Zhang R, Li N, Zhao Y. A study on energy performance of 30 commercial office buildings in Hong Kong. Energ Buildings 2017;144:117–28. https://doi.org/10.1016/j.enbuild.2017.03.042.
[49] Deng H, Fannon D, Eckelman MJ. Predictive modeling for US commercial building energy use : a comparison of existing statistical and machine learning algorithms using CBECS microdata. Energ Buildings 2018;163:34–43. https://doi.org/10.1016/j.enbuild.2017.12.031.
[50] Mohammadiziazi R, Bilec MM. Application of machine learning for predicting building energy use at Di ff erent temporal and spatial resolution under climate change in USA. 2020.
[51] Bassamzadeh N, Ghanem R. Multiscale stochastic prediction of electricity demand in smart grids using Bayesian networks. Appl Energy 2017;193:369–80. https://doi.org/10.1016/j.apenergy.2017.01.017.
[52] Ma J, Cheng JCP. Identifying the influential features on the regional energy use intensity of residential buildings based on random forests. Appl Energy 2016;183: 193–201. https://doi.org/10.1016/j.apenergy.2016.08.096.
[53] Ma J, Cheng JCP. Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology. Appl Energy 2016;183:182–92. https://doi.org/10.1016/j.apenergy.2016.08.079.
[54] Xu X, Wang W, Hong T, Chen J. Energy & Buildings Incorporating machine learning with building network analysis to predict multi-building energy use. Energ Buildings 2019;186:80–97. https://doi.org/10.1016/j.enbuild.2019.01.002.
[55] Mocanu E, Nguyen PH, Kling WL, Gibescu M. Unsupervised energy prediction in a smart grid context using reinforcement cross-building transfer learning. Energ Buildings 2016;116:646–55. https://doi.org/10.1016/j.enbuild.2016.01.030.
[56] Papadopoulos S, Bonczak B, Kontokosta CE. Pattern recognition in building energy performance over time using energy benchmarking data. Appl Energy 2018;221(March):576–86. https://doi.org/10.1016/j.apenergy.2018.03.079.
[57] Kolter JZ, Ferreira J. A large-scale study on predicting and contextualizing building energy usage. In: Proceedings of the National Conference on Artificial Intelligence. vol. 2; 2011.
[58] Wei Y, et al. A review of data-driven approaches for prediction and classification of building energy consumption. Renew Sustain Energy Rev 2018;82(August 2017):1027–47. https://doi.org/10.1016/j.rser.2017.09.108.
[59] Wang Z, Srinivasan RS. A review of arti fi cial intelligence based building energy use prediction: contrasting the capabilities of single and ensemble prediction models. Renew Sustain Energy Rev 2017;75(September 2015):796–808. https://doi.org/10.1016/j.rser.2016.10.079.
[60] Jain RK, Smith KM, Culligan PJ, Taylor JE. Forecasting energy consumption of multi-family residential buildings using support vector regression : investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Appl Energy 2014;123:168–78. https://doi.org/10.1016/j.apenergy.2014.02.057.
[61] Ku AL, Qiu Y Lucy, Lou J, Nock D, Xing B. Changes in hourly electricity consumption under COVID mandates: a glance to future hourly residential power consumption pattern with remote work in Arizona. Appl Energy 2022;310. https://doi.org/10.1016/j.apenergy.2022.118539.
[62] Chou JS, Tran DS. Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. Energy 2018;165:709–26. https://doi.org/10.1016/j.energy.2018.09.144.
[63] Biswas MAR, Robinson MD, Fumo N. Prediction of residential building energy consumption: a neural network approach. Energy 2016;117:84–92. https://doi.org/10.1016/j.energy.2016.10.066.
[64] Edwards RE, New J, Parker LE. Predicting future hourly residential electrical consumption: a machine learning case study. Energ Buildings 2012;49:591–603. https://doi.org/10.1016/j.enbuild.2012.03.010.
[65] Fan C, Wang J, Gang W, Li S. Assessment of deep recurrent neural network-based strategies for short-term building energy predictions. Appl Energy 2019;236(July 2018):700–10. https://doi.org/10.1016/j.apenergy.2018.12.004.
[66] Abdallah M, Abu Talib M, Hosny M, Abu Waraga O, Nasir Q, Arshad MA. Forecasting highly fluctuating electricity load using machine learning models based on multimillion observations. Adv Eng Inform Aug. 2022;53:101707. https://doi.org/10.1016/J.AEI.2022.101707.
[67] Burlig F, Bushnell J, Rapson D, Wolfram C. Low energy: estimating electric vehicle electricity use. AEA Pap Proc 2021;111. https://doi.org/10.1257/pandp.20211088.
[68] Ma Z, Ye C, Li H, Ma W. Applying support vector machines to predict building energy consumption in China. Energy Procedia 2018;152:780–6. https://doi.org/10.1016/j.egypro.2018.09.245.
[69] Paudel S, et al. A relevant data selection method for energy consumption prediction of low energy building based on support vector machine. Energ Buildings 2017;138:240–56. https://doi.org/10.1016/j.enbuild.2016.11.009.
[70] Taylor JW, McSharry PE. Short-term load forecasting methods: an evaluation based on European data. IEEE Trans Power Syst 2007;22(4):2213–9. https://doi.org/10.1109/TPWRS.2007.907583.
[71] Hernandez L, Baladrón C, Aguiar JM, Carro B, Sanchez-Esguevillas AJ, Lloret J. Short-term load forecasting for microgrids based on artificial neural networks. Energies 2013;6(3):1385–408. https://doi.org/10.3390/en6031385.
[72] Velasco LCP, Villezas CR, Palahang PNC, Dagaang JAA. Next day electric load forecasting using artificial neural networks. In: 8th Int. Conf. Humanoid, Nanotechnology, Inf. Technol. Commun. Control. Environ. Manag. HNICEM 2015. December; 2016. p. 1–6. https://doi.org/10.1109/HNICEM.2015.7393166.
[73] Zhang W, et al. Estimating residential energy consumption in metropolitan areas: a microsimulation approach. Energy 2018;155:162–73. https://doi.org/10.1016/j.energy.2018.04.161.
[74] Baklrtzis AG, Petrldis V, Klartzls SJ, Alexladls MC. A neural network short term load for the greek powers department of electrical and computer engineering. Neural Netw 1995:858–63.
[75] Ryu S, Noh J, Kim H. Deep neural network based demand side short term load forecasting. Energies 2017;10(1):1–20. https://doi.org/10.3390/en10010003.
[76] Masum S, Liu Y, Chiverton J. Multi-step time series forecasting of electric load using machine learning models. In: Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 10841. LNAI; 2018. p. 148–59. https://doi.org/10.1007/978-3-319-91253-0_15.
[77] Azadeh A, Ghaderi SF, Sohrabkhani S. A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran. Energy Policy 2008;36(7):2637–44. https://doi.org/10.1016/j.enpol.2008.02.035.
[78] Rahman A, Srikumar V, Smith AD. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Appl Energy 2018;212(December 2017):372–85. https://doi.org/10.1016/j.apenergy.2017.12.051.
[79] Wang J, Li L, Niu D, Tan Z. An annual load forecasting model based on support vector regression with differential evolution algorithm. Appl Energy 2012;94: 65–70. https://doi.org/10.1016/j.apenergy.2012.01.010.
[80] Azadeh A, Ghaderi SF, Tarverdian S, Saberi M. Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. Appl Math Comput 2007;186(2):1731–41. https://doi.org/10.1016/j.amc.2006.08.093.
[81] Rahman S, Senior RB, Member M. An expert system based algorithm for short term load forecast. 1988. https://doi.org/10.1109/59.192889.
[82] Kong F, Song GP. Middle-long power load forecasting based on dynamic grey prediction and support vector machine. Int J Adv Comput Technol 2012;4(5). https://doi.org/10.4156/ijact.vol4.issue5.18.
[83] Mostafavi ES, Mostafavi SI, Jaafari A, Hosseinpour F. A novel machine learning approach for estimation of electricity demand: an empirical evidence from

Thailand. Energ Conver Manage 2013;74:548–55. https://doi.org/10.1016/j.enconman.2013.06.031.
[84] Aydinalp M, Ismet Ugursal V, Fung AS. Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. Appl Energy 2002;71(2):87–110. https://doi.org/10.1016/S0306-2619(01)00049-6.
[85] Hong WC. Electric load forecasting by support vector model. App Math Model 2009;33(5):2444–54. https://doi.org/10.1016/j.apm.2008.07.010.
[86] Fan H, MacGill IF, Sproul AB. Statistical analysis of driving factors of residential energy demand in the greater Sydney region, Australia. Energ Buildings 2015;105. https://doi.org/10.1016/j.enbuild.2015.07.030.
[87] Haben S, Singleton C, Grindrod P. Analysis and clustering of residential customers energy behavioral demand using smart meter data. IEEE Trans Smart Grid 2016;7(1):136–44. https://doi.org/10.1109/TSG.2015.2409786.
[88] Czétány L, et al. Development of electricity consumption profiles of residential buildings based on smart meter data clustering. Energ Buildings Dec. 2021;252:111376. https://doi.org/10.1016/J.ENBUILD.2021.111376.
[89] Dong B, Li Z, Rahman SMM, Vega R. A hybrid model approach for forecasting future residential electricity consumption. Energ Buildings 2016;117:341–51. https://doi.org/10.1016/j.enbuild.2015.09.033.
[90] Humeau S, Wijaya TK, Vasirani M, Aberer K. Electricity load forecasting for residential customers: exploiting aggregation and correlation between households. Sustain Internet ICT Sustain Sustain 2013;2013:2013. https://doi.org/10.1109/SustainIT.2013.6685208.
[91] Rodrigues F, Cardeira C, Calado JMF. The daily and hourly energy consumption and load forecasting using artificial neural network method: a case study using a set of 93 households in Portugal. Energy Procedia 2014;62:220–9. https://doi.org/10.1016/j.egypro.2014.12.383.
[92] Zhao HX, Magoulès F. Feature selection for predicting building energy consumption based on statistical learning method. J Algorithms Comput Technol 2012;6(1):59–77. https://doi.org/10.1260/1748-3018.6.1.59.
[93] Li Q, Ren P, Meng Q. Prediction model of annual energy consumption of residential buildings. In: 2010 Int. Conf Adv Energy Eng ICAEE. 2010; 2010. p. 223–6. https://doi.org/10.1109/ICAEE.2010.5557576.
[94] Olu-Ajayi R, Alaka H, Sulaimon I, Sunmola F, Ajayi S. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. J Build Eng 2022;45. https://doi.org/10.1016/j.jobe.2021.103406.
[95] Ahmed Gassar AA, Yun GY, Kim S. Data-driven approach to prediction of residential energy consumption at urban scales in London. Energy 2019;187. https://doi.org/10.1016/j.energy.2019.115973.
[96] Chen M, Sanders KT, Ban-Weiss GA. A new method utilizing smart meter data for identifying the existence of air conditioning in residential homes. Environ Res Lett 2019;14(9). https://doi.org/10.1088/1748-9326/ab35a8.
[97] California Irrigation Management Information Sytem (CIMIS). CIMIS station reports [Online]. Available, https://cimis.water.ca.gov/Stations.aspx; 2023.
[98] National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI). NOAA NCEI Local Climatological Data (LCD) [Online]. Available, https://www.ncei.noaa.gov/cdo-web/datatools/lcd; 2023.
[99] Schoenau GJ, Kehrig RA. Method for calculating degree-days to any base temperature. Energ Buildings 1990;14(4). https://doi.org/10.1016/0378-7788(90)90092-W.
[100] Assessor Parcel Data 2016. County of Los Angeles Open Data. 2023. https://data.lacounty.gov/browse?category=Property%2FPlanning&utf8=✓.
[101] San Bernardino County Assessor's property characteristics 2016. Office of San Bernardino County Assessor-Recorder-Clerk; 2016. https://sbcountyarc.org/services/property-information/.
[102] Riverside County Assessor's property characteristics 2016. County of Riverside Assessor-County Clerk-Recorder; 2016. https://www.rivcoacr.org/obtaining-record-copies.
[103] Office of Environmental Health Hazard Assessment, C.E.P.A. CalEnviroScreen 3.0. https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-30; 2018.
[104] Zhang S, Zhang C, Yang Q. Data preparation for data mining 2010;9514(2003). https://doi.org/10.1080/713827180.
[105] Kotsiantis SB, Kanellopoulos D. Data preprocessing for supervised leaning. Int Dent J 2006;1(2):1–7. https://doi.org/10.1080/02331931003692557.
[106] Guyon I. An introduction to variable and feature selection 1 introduction 2003;3:1157–82.
[107] Luengo J, García S, Herrera F. On the choice of the best imputation methods for missing values considering three groups of classification methods. 2012.
[108] Alexandropoulos SAN, Kotsiantis SB, Vrahatis MN. Data preprocessing in predictive data mining 2020;34(April):2019.
[109] Ahmad T, Aziz MN. Data preprocessing and feature selection for machine learning intrusion detection systems. ICIC Express Lett 2019;13(2):93–101. https://doi.org/10.24507/icicel.13.02.93.
[110] Corrales DC, Corrales JC, Ledezma A. How to address the data quality issues in regression models: a guided process for data cleaning. Symmetry (Basel) 2018;10(4):1–20. https://doi.org/10.3390/sym10040099.
[111] Household Energy Use in California [Online]. Available, https://www.eia.gov/consumption/residential/reports/2009/state_briefs/pdf/ca.pdf; 2009.
[112] Potdar K, T. S, C. D. A comparative study of categorical variable encoding techniques for neural network classifiers. Int J Comput Appl 2017;175(4):7–9. https://doi.org/10.5120/ijca2017915495.
[113] Pedregosa F, et al. Scikit-learn: machine learning in Python. 2011. p. 2825–30.
[114] Crone SF, Lessmann S, Stahlbock R. The impact of preprocessing on data mining: an evaluation of classifier sensitivity in direct marketing 2006;173:781–800. https://doi.org/10.1016/j.ejor.2005.07.023.
[115] Huang J, Li Y, Xie M. An empirical analysis of data preprocessing for machine learning-based software cost estimation. Inf Softw Technol 2015;67:108–27. https://doi.org/10.1016/j.infsof.2015.07.004.
[116] Dodangeh E, et al. Science of the Total Environment Integrated machine learning methods with resampling algorithms for fl ood susceptibility prediction 2020;705. https://doi.org/10.1016/j.scitotenv.2019.135983.
[117] Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods 2005;21(15):3301–7. https://doi.org/10.1093/bioinformatics/bti499.
[118] Anguita D, Ghio A, Ridella S, Sterpi D. K-fold cross validation for error rate estimate in support vector machines. In: K – fold cross validation for error rate estimate in support vector machines. June; 2014. p. 2009.
[119] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selectionno. June; 2013.
[120] Gonçalves I, Silva S, Melo JB. Random sampling technique for overfitting control in genetic programming. In: Genetic programming. Berlin, Heidelberg: Lecture No., Springer; 2012. p. 218–29.
[121] Horowitz JL. The bootstrap. 2001. https://doi.org/10.1016/S1573-4412(01)05005-X.
[122] Kaufman S, Rosset S, Perlich C. Leakage in data mining: Formulation, detection, and avoidance. 2011. https://doi.org/10.1145/2020408.2020496.
[123] Shcherbakov MV, Brebels A, Shcherbakova NL, Tyukov AP, Janovsky TA, Kamaev VA Evich. A survey of forecast error measures. World Appl Sci J 2013;24(24). https://doi.org/10.5829/idosi.wasj.2013.24.itmies.80032.
[124] Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 2021;7. https://doi.org/10.7717/PEERJ-CS.623.
[125] Cai J, Luo J, Wang S, Yang S. Neurocomputing feature selection in machine learning: a new perspective. Neurocomputing 2018;300:70–9. https://doi.org/10.1016/j.neucom.2017.11.077.
[126] Langley PAT, Flamingo L, Edu S. Selection of relevant features in machine learning. 1994. p. 127–31.
[127] Hall MA, Smith LA. Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper CFS: correlation-based feature. 1999.
[128] Chandrashekar G, Sahin F. A survey on feature selection methods q. Comput Electr Eng 2014;40(1):16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024.
[129] Raschka S. Sequential feature selector. http://rasbt.github.io/mlxtend/user_guide/feature_selection/SequentialFeatureSelector/; 2014.
[130] Hooker S, Erhan D, Kindermans PJ, Kim B. A benchmark for interpretability methods in deep neural networks. Adv Neural Inf Process Syst 2019;32(NeurIPS).
[131] Huang N, Lu G, Xu D. A permutation importance-based feature selection method for short-term electricity load forecasting using random forest. Energies 2016;9(10). https://doi.org/10.3390/en9100767.
[132] Razmjoo A, Xanthopoulos P, Zheng QP. Online feature importance ranking based on sensitivity analysis. Expert Syst Appl 2017;85:397–406. https://doi.org/10.1016/j.eswa.2017.05.016.
[133] Saarela M, Jauhiainen S. Comparison of feature importance measures as explanations for classification models. SN Appl Sci 2021;3(2). https://doi.org/10.1007/s42452-021-04148-9.
[134] Permutation Feature Importance. https://scikit-learn.org/stable/modules/permutation_importance.html; 2023.
[135] Williams KT, Gomez JD. Predicting future monthly residential energy consumption using building characteristics and climate data: a statistical learning approach. Energ Buildings 2016;128. https://doi.org/10.1016/j.enbuild.2016.06.076.