Review

A Review of Data-Driven Approaches for Measurement and Verification Analysis of Building Energy Retrofits

Abdurahman Alrobaie and Moncef Krarti *

Building Systems Program, University of Colorado Boulder, Boulder, CO 80309, USA;
abdurahman.alrobaie@colorado.edu
* Correspondence: moncef.krarti@colorado.edu; Tel.: +1-303-492-3389

Abstract: Although the energy and cost benefits for retrofitting existing buildings are promising, several challenges remain for accurate measurement and verification (M&V) analysis to estimate these benefits. Due to the rapid development in advanced metering infrastructure (AMI), data-driven approaches are becoming more effective than deterministic methods in developing baseline energy models for existing buildings using historical energy consumption data. The literature review presented in this paper provides an extensive summary of data-driven approaches suitable for building energy consumption prediction needed for M&V applications. The presented literature review describes commonly used data-driven modeling approaches including linear regressions, decision trees, ensemble methods, support vector machine, deep learning, and kernel regressions. The advantages and limitations of each data-driven modeling approach and its variants are discussed, including their cited applications. Additionally, feature engineering methods used in building energy data-driven modeling are outlined and described based on reported case studies to outline commonly used building features as well as selection and processing techniques of the most relevant features. This review highlights the gap between the listed existing frameworks and recently reported case studies using data-driven models. As a conclusion, this review demonstrates the need for a flexible M&V analysis framework to identify the best data-driven methods and their associated features depending on the building type and retrofit measures.

Keywords: baseline models; data-driven modeling; energy conservation measures; measurement and verification; retrofitted buildings

1. Introduction

Buildings are the largest energy-consuming sector, with a global share of 35% of energy consumption, exceeding industry and transportation [1]. It is estimated that 85% of the building energy consumption is attributed to heating, ventilation, and air conditioning (HVAC), lighting, and plug loads. Moreover, residential buildings account for approximately 63% [1] of the total energy used by the building sector. In terms of electricity consumption, buildings take 50% of the world's electricity consumption [1]. According to the U.S. Energy Information Administration (EIA), projections show that the residential and commercial buildings will increase by 1.3% per year from 2018 to 2050 for countries in the Organization for Economic Cooperation and Development (OECD), while non-OECD countries will experience an average of 2% growth annually [2]. Several studies have analyzed the historical and current status of energy consumed by buildings and have projected future increases in building-related energy use globally [3] or in specific regions such as China [4], the European Union [5], and Gulf Cooperation Council countries [6]. The high energy consumption by the built environment has significant detrimental effects on the environment and the climate. Several governmental agencies and global organizations are adopting initiatives and programs that target the reduction of energy consumption in the building sector. For instance, the U.S. Department of Energy has set a 2030 goal of tripling

2020 levels of commercial and residential buildings' energy efficiency [7]. Similarly, the UK has developed a net-zero energy strategy for buildings so that by 2050, buildings will be completely decarbonized [8]. The goal included a plan that is driven by decisions to fund several research projects, support owners to shift the buildings' efficiency, and subsidize clean and efficient projects [8]. In addition, China, the largest carbon dioxide emitter, has pledged to reach neutral carbon emissions before 2060 [9].

Such initiatives and pledges can be achieved by a combination of several approaches including enhancing renewable energy sources, setting more stringent energy efficiency regulations, and funding research to develop effective and transformative technologies in the building energy sector. However, improvements in the energy efficiency levels of existing buildings are required to attain the desired goals. Indeed, the average annual rate of replacing existing buildings is low reaching only 1% in the UK [10]. It is argued that the environmental and economic benefits of retrofitting existing buildings outweigh those achieved by replacing them with more efficient new buildings. Hasik et al. [11] performed a life cycle assessment (LCA) of both retrofitted and newly constructed buildings and found that retrofitting results in a reduction ranging between 53% and 75% for over six different environmental impact factors compared to new construction. Economic benefits are highest when retrofitting the least energy-efficient buildings when considering aspects such as the creation of employment and reduction of carbon emissions compared to constructing new buildings [12].

Retrofitting existing buildings includes renovations of mechanical, structural, and electrical systems with a range of options such as refurbishment, replacement, or addition of new equipment. In the case of energy efficiency retrofits, the replacement and addition of new equipment is usually referred to as an energy conservation measure (ECM). Several ECMs can be considered for existing buildings such as changing HVAC equipment, lighting systems, and envelope features such as glazing types and wall assemblies. The deployment of ECM aims primarily at reducing the energy use and cost of the buildings. The required investments for ECMs are justified based on economic and environmental benefits. However, the implementation of ECMs can face several challenges, especially during assessment and identification, as well as installation and verification. In the first period, any missing information and documentation can hinder good assessment of the existing building energy performance and thus effective identification. Similarly, uncertainty and lack of data can affect the installation and validation process. For the validation analysis, an energy model of the building is typically needed to predict the energy use before the deployment of any ECM with minimum prediction uncertainty. Additionally, this process includes an essential step in justifying the effectiveness of the installed employed EMCs, that is, measurement and verification (M&V) analysis.

Several literature reviews have discussed data-driven models for tasks that are related to the energy performance of existing buildings. Wei et al. [13] categorized data-driven approaches into two applications including prediction and classification. However, this literature review considers only prediction with a subset of prediction that is related to times series instead of cross-sectional data. Cross-sectional data represent observations that are not collected at unique timestamps or identified by a chronological order. Typically, data are represented in a tabular form of columns and rows. If each row represents a value assigned for a specific timestamp (e.g., the energy consumption of a building at 13:00), then the data is time series. In M&V baseline modeling, the data characteristics may require different modeling methods. On the other hand, prediction of energy consumption with cross-sectional data is performed with each realization being a single building and its response variable being a single value representing total energy consumption over a specific period. Deb et al. [14] did a review of forecasting in building's energy consumption using nine different techniques with a hybrid approach that represents a combination of more than one technique. Unlike M&V baseline modeling, forecasting in building's energy consumption focuses on predicting future values using correlation from closely past values which is not possible for the created baseline when performing M&V after retrofitting.

Grillone et al. [15] conducted a literature review of deterministic and data-driven methods that can be used to estimate energy savings from retrofits. Their analysis is focused on two areas: M&V and prediction and recommendation while this literature review aims to focus on M&V and prediction extensively. Deb and Schlueter [16] reviewed data-driven approaches in retrofitting applications including benchmarking, energy signature, and feature extraction. However, the review did not specifically discuss the baseline modeling approaches for M&V Applications.

In terms of applications, numerous reported studies have considered data-driven models to predict the energy consumption of existing buildings. Many of the reported data-driven models are based on historical data that is used for training the models and testing their prediction accuracy. However, few of such models have been applied to create a baseline for M&V analysis to determine energy savings achieved by installed ECMs. Additionally, reported data-driven models for M&V applications have been developed and tested only for specific building types and ECMs. Furthermore, models have not been evaluated for multiple buildings and ECMs. With the increasing interest in applying data-driven models for building energy retrofit analysis, there are limited guidelines on the suitability of these models for M&V applications. Therefore, this literature review examines various methods and algorithms that have been applied to develop data-driven models for M&V analysis of building energy retrofits. The contribution of this paper is to build a summary of every step in building a data-driven model for M&V analysis. The focus is specifically on only prediction with applicability to M&V analysis along with the necessary steps in processing and creating features before training the baseline model. The literature review summarizes the previous endeavors of studies and frameworks and extracts the most listed requirements and modeling approaches.

2. Overview of Measurement and Verification Analysis

M&V analysis is a process of quantifying the energy use savings due to the deployment of ECMs when retrofitting existing buildings. A baseline energy model allows the prediction of the energy use of an existing building due to variations in environmental and behavioral factors such as different climatic conditions or changes in occupancy levels before any ECM implementation. The baseline energy model is often used as a benchmark to estimate energy savings due to installing one or several EMCs. Figure 1 shows the difference between metered and modeled energy use of an existing building over three periods. The first period corresponds to the pre-retrofit operation of the building with the energy use data being metered using historical data collected from utility bills or a building management system (BMS). The baseline energy model is typically developed and tested during this pre-retrofit period. During the retrofit period, ECMs are installed in the building resulting in a gradual reduction in the energy consumption compared to the predictions of the baseline energy model as noted in Figure 1. After completing the ECM installation phase, the building typically consumes less energy than during the pre-retrofit period, as demonstrated in Figure 1 during the post-retrofit period. Indeed, the baseline energy model predicts higher energy consumption than the metered data during the post-retrofit period. The difference between the baseline and the metered energy consumption during the post-retrofit period represents energy savings incurred by the installed ECMs during the post-retrofit period. A building energy model baseline is often established for M&V analysis of retrofitting existing buildings, especially for those with energy use historical data that meet certain requirements such as date range, missing values, or reporting frequency. These requirements, however, are not well-defined and vary from one case to another depending on a wide range of factors including the nature of occupancy. The process of constructing a data-driven baseline model requires that a building has been in operation during a sufficiently long period to gather enough data to establish correlation between its energy performance and other independent factors such as weather and occupancy parameters.
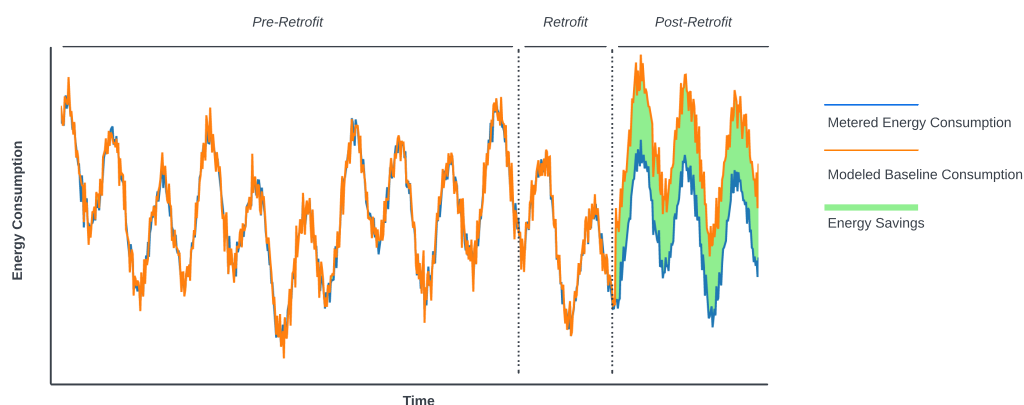
Figure 1. Actual energy consumption and modeled building baseline vs. time.

2.1. Measurement and Verification Protocols

Several M&V protocols have been developed to improve consistency and reduce uncertainty in estimating the energy savings attributed to retrofitting existing buildings, such as the International Performance Measurement and Verification Protocol (IPMVP) [17] and ASHRAE Guideline 14 [18]. The analysis approaches outlined in these protocols differ depending on the geographical regulatory requirements, the types of ECMs, and building typologies. Additionally, specific frameworks and methodologies have been proposed to achieve the desired objectives of retrofitting projects. For instance, Ma et al. [19] developed a systematic methodology for carrying out retrofitting projects and successfully completing the various phases and analyses including the M&V analysis.

2.1.1. International Performance Measurement and Verification Protocol (IPMVP)

The International Performance Measurement and Verification Protocol or IPMVP is one of the most common frameworks in performing M&V analysis of retrofitting existing buildings, with four evaluation options as outlined in Table 1. The selection of the most appropriate analysis option depends on the boundary of the deployed ECMs. Options A and B are applied when the retrofit is restricted to only one specific and isolated building energy system. These two options differ depending on the analysis method and the availability of metered data. In particular, option A can be used for a M&V analysis for a lighting system retrofit using only key parameters including power ratings and operation schedules to calculate energy savings. On the other hand, option B is applied for systems whose energy performance can be monitored such as chillers and boilers. In addition, Options C and D can be applied when the retrofit affects the energy performance of the entire building. When metered building energy data can be collected before and after the retrofit periods, Option C is suitable for conducting the M&V analysis. When the historical data of metered energy consumption are not available or are unreliable before or after the retrofit, option D is considered using calibrated energy models [17].

Table 1. IPMVP M&V options.

M&V Options	Boundary	Parameters	Process
Option A	System	Key system parameters with estimation	Simple calculations with the estimated parameters
Option B	System	All system parameters with no estimation	More rigorous calculations with all related parameters
Option C	Whole Building	Whole building energy consumption historical data	whole building baseline modeling with building energy consumption and related parameters
Option D	Whole Building and/or System	Whole building and/or system parameters with energy bills	Calibrated simulation of the boundary using data and modeling tools

2.1.2. ASHRAE Guideline 14

ASHRAE has developed Guideline 14 for Measurement of Energy, Demand, and Water Savings to standardize the M&V calculations used to estimate achieved energy demand and water savings from retrofit projects. ASHRAE Guideline 14 utilizes three M&V analysis options that are similar to those specified by the IPMVP including retrofit isolation, whole facility, and whole building calibrated simulation. Instead of having two analysis options for isolated systems, ASHRAE Guideline 14 allows only one method with the flexibility of the parameters that can be used in the calculations. The whole facility option is similar to Option C of the IPMVP using the whole facility metered energy consumption along with independent variables to establish the building's baseline energy model. The third approach is similar to the IPMVP's option D using a calibrated baseline model to quantify savings from the retrofit. While ASHRAE Guideline 14 shares some features with IPMVP, it does not cover specific details such as energy performance contracting and metering provisions as IPMVP does [18].

2.1.3. Advanced Measurement and Verification

Advanced M&V, usually referred to as M&V 2.0, encompasses detailed analysis approaches using high frequency metered data (i.e., sub-hourly), and end-use loads using advanced metering infrastructures (AMI) [20]. In fact, M&V 2.0 enables metered data to be more effective for building real-time performance assessment, occupant engagement, and resource management using various analysis tools and algorithms. The improvements of both hardware and software over the last decade have resulted in better accuracy in performing various M&V tasks such as developing baseline models, detecting non-routine events, and benchmarking energy consumption. Furthermore, retrieval of metered data at higher frequencies and shorter time intervals facilitates performing data analytics and automating savings quantification for retrofit projects, which reduces the time lag between implementation and evaluation phases [21].

2.2. Baseline Modeling

Three approaches are commonly used to establish building baseline models: deterministic (also referred to as direct or white-box), data-driven (also referred to as indirect or back-box), and hybrid (also referred to as gray-box) methods. All approaches reach the same objective in M&V (i.e., constructing a baseline) with different inputs and processes. The comparison between such means for a single case is time-consuming and rarely performed, as each approach has sub-approaches, which alone will take more time and effort. Therefore, this subsection aims to provide a concise comparison between them.

Deterministic modeling relies on physics-based tools to predict the energy consumption of buildings due to their thermal interactions with the outdoor environment. Such interactions are often represented using heat and mass balance equations that are solved using a set of algorithms that are the basis for a deterministic building energy modeling tool. There is a wide range of commercially available and open-source deterministic modeling tools that can be utilized for developing building energy models including EnergyPlus [22], TRNSYS [23], DOE-2 [24], DesignBuilder [25], Matlab/Simulink [26], and Modelica/Dymola buildings library [27]. Most of these deterministic modeling tools require comprehensive input data about the building features such as envelope thermal properties, mechanical equipment efficiency, and operation schedules. Ke et al. [28] developed a deterministic (also referred to as white-box) baseline energy model using eQUEST software (based on DOE-2 simulation engine) for an existing office building with a mean bias error (MBE) of 0.37%. The building energy model includes over 50 input variables indicating the types and operation characteristics of chillers, indoor air-conditioning units, and cooling towers performance in addition to several variables describing other building systems such as the envelope elements and lighting fixtures. The study has demonstrated high levels of interpretability in understanding the specific interactions between energy end-uses of various building systems and occupancy behaviors that deterministic building

energy modeling can offer. However, the interpretability as well as the high prediction accuracy of the deterministic models come with significant computing times and input data collection efforts.

Data-driven models represent relationships between energy performance indicators and environmental parameters identified using historical data. These relationships are then applied to predict the building response when all or some environmental variables would change. Thus, data-driven models are based on developing correlations between the desired input and output parameters using various statistical and machine learning approaches. In particular, the development as well as the accuracy level of data-driven models rely heavily on historical data for both input and output variables. Types and applications of data-driven modeling are discussed in detail in Section 4. Typically, the accuracy and interpretability levels of data-drive models are lower than those achieved by white-box models, as data is usually noisy and the occupancy behavior is not consistent.

Hybrid, also referred to as gray-box, models utilize a data-driven analysis approach to tune and improve physics-based (also referred to as deterministic or white-box) models through value estimations of input parameters values using historical data. A common deterministic model using in the hybrid analysis approach is based on resistance and capacitance (RC) modeling to account for building thermal mass. Piccinini et al. [29] developed a framework for building a hybrid modeling approach using historical monthly electricity and natural bills of a primary school building to calibrate a building energy model using the Dymola Environment. The study achieved a normalized mean bias error (NMBE) of 1.8% while using far less parameters compared to a white-box model developed using detailed simulation tools such as EnergyPlus or TRNSYS. Similarly, Giretti et al. [30] compared the performance of reduced-order modeling using Modelica with Buildings Library against calibrated detailed models belonging to three cases: a hospital, library, and an educational building. The calibrated reduced-order models obtained a coefficient of variation of the root mean squared error (CV(RMSE)) between 5% and 8% compared to the detailed models while using only 25 parameters that are categorized into building envelope, heating/cooling system, occupancy, and weather components.

Chen et al. [31] compared three energy modeling approaches including black-, white-, and gray-box models. The comparative analysis considered several performance metrics including development efforts, computational times, and analysis limitations. Their study found a trade-off between each metric category, with black-box models requiring the least effort and time, while white-box models had far more input parameters. Gray-box models are in the middle in terms of development effort and required input parameters as it still requires significant data correlating energy consumption and weather variables. In terms of interpretability, white-box models allow for better understanding of the impact contributed by each input on building energy performance followed by gray-box then black-box modeling approaches. This capability is due to the fact that relationships between energy consumption and input parameters are well-established for deterministic models based basic physical principles rather than inferred from historical data as required by the data-driven (black-box) models.

3. Data-Driven Trend in Building Energy Modeling

Data-driven approaches have gained an increasing popularity especially in the last 10 years as more methods have been proposed for baseline modeling and energy consumption forecasting due to the higher availability of reliable building energy consumption data and advances in machine learning techniques. Indeed, significant valuable and granular data can be collected from smart metering building technologies. For instance, measured building data currently have short frequencies of 15-min or 1-h intervals instead of monthly frequency data like utility bills. Additionally, the energy end-use granularity allows better evaluation of various building energy systems (i.e., lighting, HVAC, and appliances). Moreover, higher data storage capabilities through databases and clouds permit access to a significant amount of building performance data for various analyses and applications. Due

to the aforementioned factors, advances and interests in using data-driven approaches for building energy analyses have been significantly increased in the last decade as described in the following subsections.

3.1. Interest in Data-Driven Approaches

To gauge the level of interest in using data-driven approaches for building energy assessments, a bibliometric analysis is performed by using Web of Science database [32]. The analysis is specific to the literature published between 2010 and 2021, focusing on technical papers that develop and apply data-driven approaches to model or predict building energy consumption. Web of Science is used to identify the number of published papers that are relevant to data-driven approaches during a specific time period. Specifically, the following keywords are considered for the Web of Science search:

- Data-driven Building Energy Modeling.
- Building Energy Prediction.
- Building Electricity Prediction.
- Machine Learning Building Energy Modeling.

Figure 2 shows the results of the bibliometric analysis, clearly depicting the rapid and consistent increase over the last decade in published papers with a focus on data-driven approaches applied to building energy analysis.
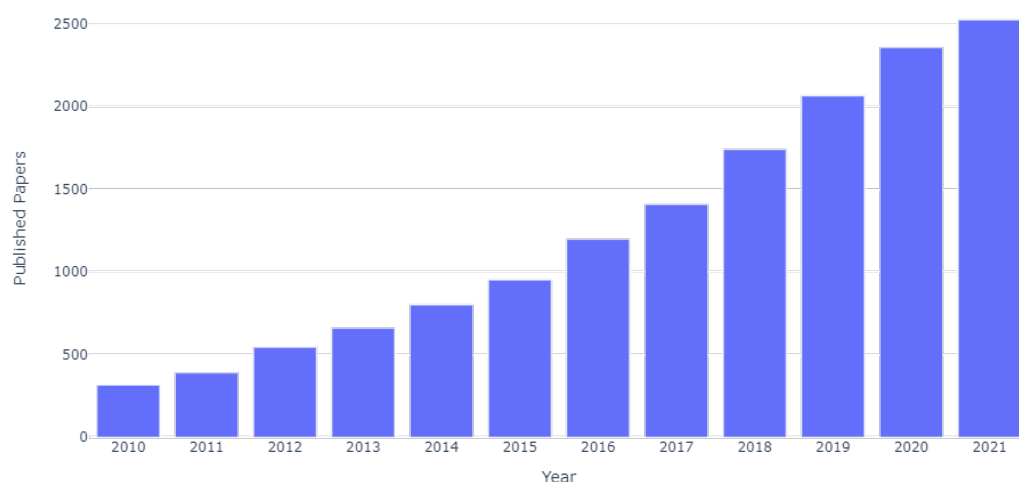

Figure 2. Number of research papers using data-driven approaches for building energy modeling.

3.2. Data-Driven Approaches

The bibliometric analysis is also performed to understand the most common methods used in the literature for data-driven building energy modeling. The search query of peer-reviewed papers published from 2010 to 2021 to identify the used modeling approaches was categorized into machine learning methods categories by Sklearn library [33] in Python. The models are further organized to represent relevant models since Sklearn library includes many models while the energy modeling research papers were more focused on a set of these models. However, this approach was not successful, since the search query cannot determine the objective of the research paper, but rather filter results based on matching keywords and similarity. This issue is amplified when the research paper addresses certain methods but does not actually use that method. Hence, an alternative approach is followed by using the resulted papers in the queries mentioned in Section 3.1 and randomly selecting papers until reaching 75 research papers. The randomness is introduced by exporting the list of papers and using Python to shuffle titles. If a certain paper does not utilize data-driven methods to model the energy consumption, the next paper on the shuffled list will be analyzed until reaching 75 papers. Figure 3 shows the result of this reading. The main level categories are:

- Linear Regression (LR): it is a category that involves linearly regressed models as described in Section 4.1.
- Ensemble methods and Decision Tree (DT): The two methods are grouped in one category based on their similarity as explained in Section 4.2.
- Support Vector Machine (SVM): it is a modeling method of using supporting vectors to fit a hyperplane for regression and classification as demonstrated in Section 4.3.
- Artificial Neural Network (ANN): it is a category that utilizes deep learning and human brain-inspired function of neurons and layers as discussed in Section 4.4.
- Kernel regression: it is a family of non-parametric techniques to fit changing coefficients on data points as outlined in Section 4.5.

From the main categories, the most commonly used models branch based on the frequency of occurrence in peer-reviewed papers. The plot does not branch into the specific models used in each study, but rather into a general yet specifically sufficient level that can encompass the modeling approach.
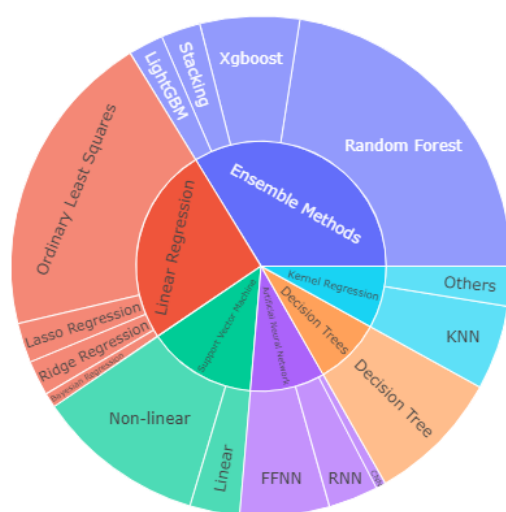

Figure 3. Sunburst chart of bibliometric analysis methods used in data-drive building energy modeling.

3.3. Building Typologies

Data-driven models have been applied for several building types including commercial and residential buildings as well as individual and group of buildings. The typology can significantly impact the building energy performance due to occupant behavior and the operation of various systems. According to a study performed by Liang et al. [34] in Phoenix, Arizona using data for 636 commercial and 201 residential buildings collected from Energize Phoenix program, retrofits can save about 8% and 12% in annual energy consumption for residential and commercial buildings, respectively. Wang et al. [35] performed a comparison between different building typologies including offices, shopping malls, and educational buildings. The study concluded that shopping malls have the highest potential for energy savings, followed by multifunctional buildings and hotels. Figure 4 shows a sunburst chart of building typology reported in various energy modeling studies categorized using three levels including data source, building the main function, and sub-category of building type as the following:

- Actual or metered data: these consist of energy consumption recorded using standalone measurement devices or Building Management Systems (BMS).
  - Non-residential: encompassing mostly commercial buildings and office spaces. Educational buildings represent cases where the building's purpose is mostly for classrooms and teaching such as schools and universities. Other buildings with a

commercial nature such as restaurants and retail buildings are grouped into one category.

– Residential: including buildings that are used mostly for housing and living spaces. Residential buildings are divided into detached houses, apartment buildings, and other types of residential buildings.

- Simulated or synthesized data: are typically generated using simulation analysis tools such as EnergyPlus and DOE-2.
- Public datasets: are obtained from public databases such as Open Energy Data Initiative (OEDI) [36].
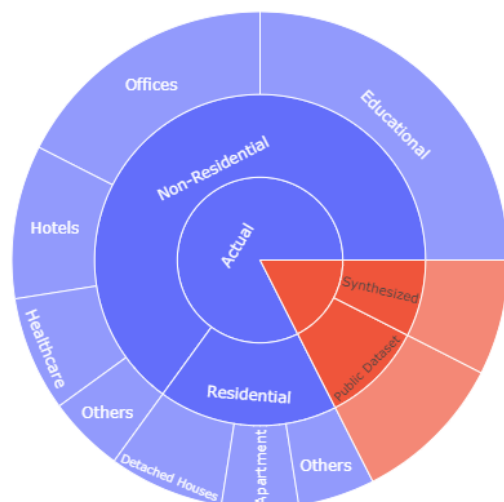

Figure 4. Sunburst chart of bibliometric analysis building types used in data-drive building energy modeling.

4. Data-Driven Approaches

This section outlines a brief description of each method used for data-driven modeling and the main reported applications for these methods. Each subsection discusses one of the main categories that are mentioned in Section 3.2 with an explanation of the general algorithm, sub-models within the category, and a list of publications that utilized one or more of the category's models. The tables in this section show summaries of such publications with a description of the applied case, data type, features, utilized category's models, and data granularity or frequency. The papers listed in this section include data-driven modeling suitable not only for M&V analysis, but also for baseline building energy development. Among the reported literature, there are very limited papers that perform full M&V analysis using data-driven models, as most of the reviewed applications evaluate the prediction performance of data-driven approaches. Features, predictors, and dependent variables are terms that are used interchangeably to list input parameters that are used to train the model to perform predictions about the response, target, or independent variable that represents the model's output. In each of the following sections' tables, the general category of the feature will be mentioned instead of the specific features for conciseness. In Section 5, the features will be explained further in terms of filtering and processing. Based on the conclusion reached by each study, the tables in this section will show, if the results are clearly indicating one best model, the best model within the categories mentioned in this section in bold font. Data granularity represents the interval of prediction, which can be 15-min, hourly, daily, weekly, or monthly.

4.1. Linear Regression

4.1.1. Definition

Linear regression (LR) is a term that encompasses a family of different techniques that aims to establish a linear relationship between the target y (i.e., output) and a set of predictors x_i (i.e., input parameters). Equation (1) shows the general form of a linear regression [37].

$$\hat{y}_i(\beta, X) = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n \tag{1}$$

where

- β_i: Linear regression coefficients.
- X_i: Linear regression features or predictors.
- $\hat{y}_i(\beta, X)$: Linear regression prediction of the output variable.

The LR modeling includes several methods, with the most basic approach being the Ordinary Least Square (OLS). Other methods can be more complex involving other equation forms and algorithms for estimating the regression coefficients.

4.1.2. Applications

The LR approach with its various forms is used extensively in building energy modeling including establishing baselines and benchmarks. Mathieu et al. [38] used an OLS method that is called Time of Week and Temperature (TOWT) to develop a building energy baseline model. The model considers two input parameters: time of the week and temperature. The time of week segments the week into 15-min intervals, while the temperature is featured into ranges that are a function of the maximum and minimum temperature from historical data. The ranges are fitted using piecewise linear regression analysis. Existing frameworks modified the method of TOWT by using a Weighted Least Squares (WLS) regression instead of OLS and allowed for recent data to be weighted more than old data. Granderson et al. [39] compared the prediction accuracy of 10 data-driven models including those based on linear regression methods using data from 537 buildings to gauge the accuracy of M&V modeling approaches. The study included two metrics where linear regression with appropriate feature engineering showed similar accuracy to complex models. Kim et al. [40] modeled the energy use of an educational facility based on a set of metered data using linear regression methods along with more complex techniques over both working and non-working periods. In the study, Kim et al. [40] found that the linear regression method predicted building energy use less accurately than the complex model during non-working days when occupancy stochastic behavior is difficult to capture. Further applications are shown in Table 2.

Table 2. Linear regression cases that are applicable to M&V analysis and utilize real historical data.

Building Type and Number	Features	Data Granularity	Model Type	References
Bakery, office, and furniture store	Date and temperature	Hourly	OLS	[38]
537 commercial buildings	Varies from model to model with temperature and date as main features	Hourly	**OLS** and **MARS**	[39]
Educational building	Date, Weather, Occupancy	Hourly	OLS	[40]
Health Center	Temperature	Monthly	OLS	[41]
Office Building	Temperature and Occupancy	Monthly	OLS	[42]
Genome Project 2 open dataset of 1578 non-residential buildings	Date, and Meteorological data	3-h	Baysian LR	[43]
Two office buildings, two shopping malls, one hotel, and one multi-function building	Date, Meteorological data, and Occupancy	15-min	GPR	[44]
2 Educational buildings	Date and Meteorological data	Daily	OLS	[45]

Reported studies showed that LR approaches can vary in complexity and accuracy. The LR approach is often used as a benchmark for more complex models or even as a method with similar accuracy compared to more complex approaches for modeling the building energy consumption. Although LR cannot fit complex non-linear relations, the accurate selection of features, analysis before modeling, and checking LR assumptions can improve its accuracy greatly. Raw features with LR usually do not fit relationships easily while processing features with other models or simple methods allows LR to capture relationships better. This highlights the importance of the LR approach to act at least as a benchmarking model.

4.2. Decision Tree and Ensemble Methods

4.2.1. Definition

DT is a basic non-parametric supervised learning method used for classification and regression analyses. The DT method can predict the value of a target variable using simple decision rules inferred from the data features. The training process for DT follows a piecewise constant approximation approach with different prediction models for various data groups [46]. In the context of M&V applications, decision trees act as regressors rather than classifiers using different metrics to measure their splitting homogeneity or commonly known as impurity. In regression, the case of M&V, the impurity of a leaf is measured by the residual sum of squares. The tree splits data points based on features until fitting the data or reaching specified stopping criteria. The splitting relies on different metrics to decide the goodness of the criterion set at a node, which acts as a decision point that splits the data to minimize a specified cost function (i.e., residual sum of squares for M&V applications). Typically, DTs use the splitting criterion as described in Equation (2) [46].

$$R_1 j, s = X | X_j \leq s \text{ and } R_2 j, s = X | X_j > s \tag{2}$$

where

- s: A decision dividing a node into two leaves.
- R_i: Resulted leaf.
- X_i: Feature from the dataset.
- X: Realizations from the dataset.

Figure 5 shows a simple DT for regression where X represent that data points and X_i to X_r represent features from the dataset. At each decision node, the tree divides the data based on criteria, s_i to s_r, where the resulted leaves can have additional decision nodes. The tree keeps branching until minimizing the considered cost function (i.e., Residual Sum of Squares (RSS)) as shown in Equation (3) or reaching the set stopping criteria. The end leaves represent the predicted value for the data points that fall into the leaf based on a series of decision nodes, y_j to y_r.

$$RSS = \sum_{i:x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1}) + \sum_{i:x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2}) \tag{3}$$

However, decision trees can form the basis for more complex models using ensemble methods. Random Forest (RF) is an ensemble method that fits several regressions or classification decision trees for various sub-samples of the dataset and aggregates them by averaging to improve prediction accuracy levels and control the overfitting problem. This ensemble approach is called "bagging" with sampling features and aggregating via averaging. Moreover, other ensemble approaches can be utilized instead of simple weighted averaging methods such as RF. "Stacking" is another ensemble process of generating several base models using training data such that meta-models use predictions from base models as features for out-of-sample predictions. "Blending" is a variation of stacking using testing data set to gauge the prediction accuracy of base models while a final test is applied for the meta-model [47]. The state-of-art ensemble methods include AdaBoost [48], Gradient Boosting Machine (GBM) [49], Extreme Gradient Boosting Machine (XGB) [50],

and Light Gradient Boosting Machine (LGBM) [51]. All these methods use the principle of multiple learners, except the boosting algorithm, which introduces weighting penalization before each successive learner rather than aggregating the final prediction from multiple learners directly. However, DT and DT-based models like RF and XGB are not effective at extrapolating beyond the range of the predictor's values [52]. Therefore, when a building's energy consumption data include values beyond the trained data for the predictors, other algorithms must be incorporated to overcome this issue.
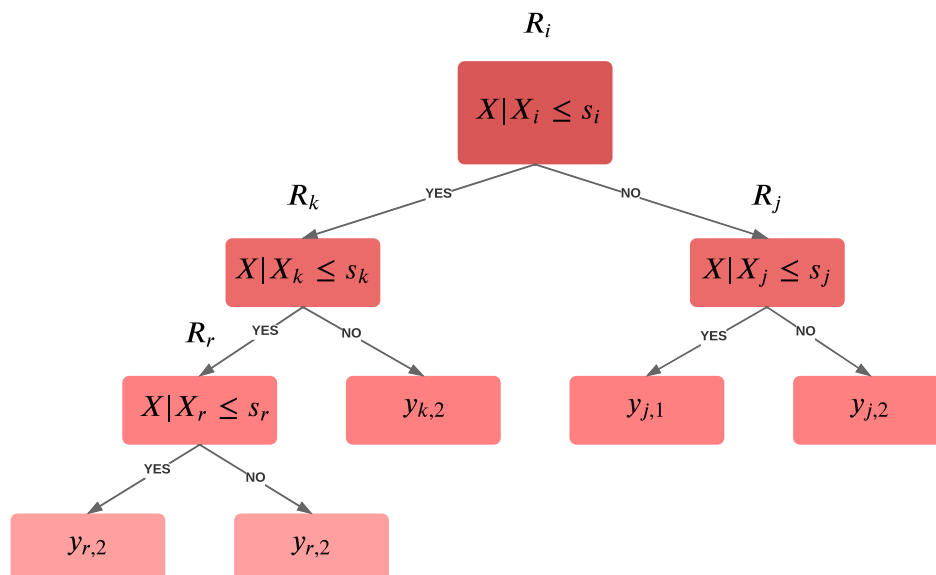

Figure 5. Simple structure of decision tree for regression.

4.2.2. Applications

DT is a machine learning method that is used in both classification and regression applications. Touzani et al. [53] used XGB to determine the improvements of boosting against TOWT by using date and temperature of buildings where the accuracy metric boxplots showed an improvement over the TOWT method. Afroz et al. [54] compared six data-driven models by predicting the energy consumption of 11 office buildings located in Ottawa, Canada. RF method is found to provide superior prediction accuracy levels than those of the DT method and even better than those achieved by some models except Nonlinear Autoregressive with Exogenous inputs (NARX). Agenis-Nevers et al. [55] applied 10 methods to model the energy performance of 11 UAE buildings including 10 commercial complexes and one housing unit. RF approach has achieved a global score that is above the average for the 11 buildings. Liu et al. [56] used simulated data generated using DesignBuilder model for an educational building in the Northern China region to compare the energy use predictions from three models. The study found that RF provides the highest prediction accuracy. Publications that utilized DT and ensemble methods are shown in Table 3.

Ensemble methods can be used to develop a set of new models different from base models. With several sampling and aggregating techniques, the choice of the best category approach can pose some challenges. Indeed, the best suitable model depends on several factors and is often not possible to generalize for different building types and retrofit measures. However, reported comparative studies have indicated the appropriateness of certain ensemble methods over others. For example, in several analyses, RF approach outperforms DT in regression modeling as the former prevents overfitting by introducing randomness while the latter tends to branch out until overfitting the training data. On the other hand, approaches such as stacking rely heavily on their base learners with different applications providing completely different results. While stacking can be effective,

it is still a computationally expensive approach with vague transparency and unclear interpretability.

On the other hand, XGB and RF can indicate the contribution of each variable and increase the model interpretability. Given the results of the bibliometric analysis and reported applications of this modeling category, RF and XGB are the two most commonly suitable techniques in ensemble approach with limited drawbacks.

Table 3. Decision tree and ensemble cases that are applicable to M&V analysis and utilize real historical data.

Building Type and Number	Features	Data Granularity	Model Type	References
410 Commercial building	Date and Temperature	15-min	XGB	[53]
12 Office building	Date and Meteorological data	Hourly	**RF** and DT	[54]
10 Commercial and 1 residential buildings	Meteorological data	Monthly and Daily	**RF** and DT	[55]
2 Educational Buildings	Date, Meteorological data, and Occupancy	Hourly	**RF** and DT	[57]
Residential Quarter	Date and Meteorological data	Hourly	DT, GBM, **XGB**	[58]
507 Non-residential Buildings from Genome Database	Date and Meteorological data	Hourly	Stacking	[59]
Educational Building	Date, Meteorological data, and Occupancy	Hourly	Bagging Trees	[60]
Healthcare	Date, Meteorological data, and Occupancy	Daily and Weekly	XGB and **RF**	[61]
Hotel	Date, Meteorological data, and Occupancy	Hourly	RF	[62]
1325 air conditioners	Date, Meteorological data, and indoor environmental parameters	Daily	**XGB**, RF, GBDT, AdaBoost	[63]
Heat pump in a residential building	Date, Meteorological data, and HVAC system operating parameters	30-min	XGB, **Stacking**	[64]
House	Meteorological data, and indoor environmental parameters	10-min	XGB	[65]

4.3. Support Vector Machine

4.3.1. Definition

Support vector machine (SVM) is a common machine learning tool used for classification and regression analyses. A SVM model is developed by fitting a hyperplane that aims to determine the underlying relationship between predictors (i.e., input parameters) and target (i.e., output). The hyperplane is supported by two vectors as shown in Figure 6 such that the error measured with respect to these two vectors and the hyperplane is minimized by including the maximum number of points within the boundary lines and close to the hyperplane. The two parallel lines represent the supporting vectors while the middle line is the hyperplane. Equation (4) shows the hyperplane equation where the data is mapped to a higher dimension by a dot product between points and weights. Then, SVM aims to minimize the cost function, which is shown in Equation (5), where ϵ represents the distance of the supporting vectors from the hyperplane and ζ represents the distance from the supporting vectors to the points outside the supporting vectors. The more points that lie within the boundary, the less the cost function [66].

$$F(x_i) = (w, \phi(x_i)) + b \tag{4}$$

where

- b: Model bias.
- $\phi(x_i)$: Kernel function that maps data to higher dimension.
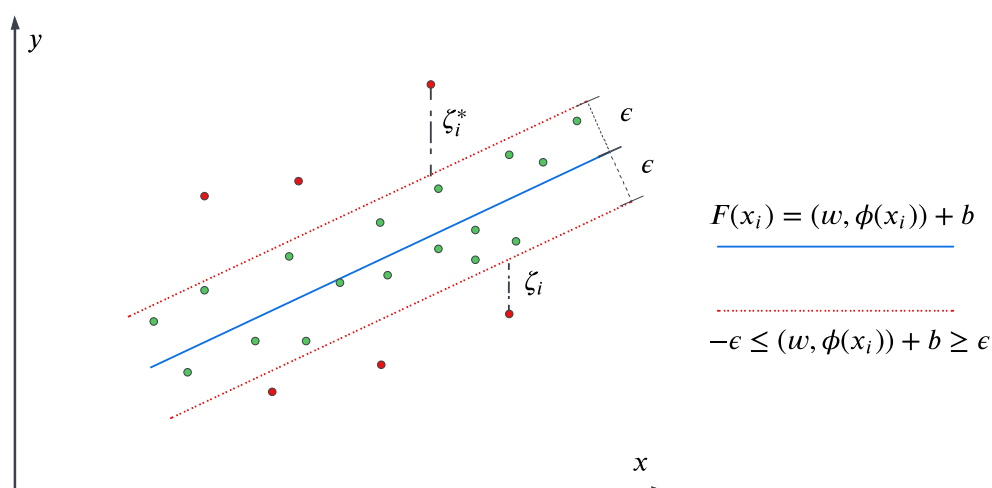- w: Model weights.

$$F(x_i) = (w, \phi(x_i)) + b$$

$$-\epsilon \leq (w, \phi(x_i)) + b \geq \epsilon$$

Figure 6. One-dimensional support vector machine for regression.

$$Min(L(w,C)) = Min\left(\frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\zeta_i + \zeta_i^*)\right) \tag{5}$$

- $L(w,C)$: Loss or cost function.
- C: Direction regularization coefficient.
- ζ_i: The distance from data observation to any of the supporting vectors which is minimized by the cost function.

4.3.2. Applications

Edwards et al. [67] compared using two variations of SVM against other modeling techniques including LR and ANN. The SVM is demonstrated to have a better performance compared to complex models when applied to residential buildings and to provide similar prediction accuracy levels compared to complex models for commercial buildings. Amber et al. [68] utilized parameters denoting working and non-working days to predict energy demand for an office building which resulted in SVM models that are trained on a subset of data denoting a specific type of day (i.e., work or non-work day) outperforming SVM models that were trained on all the data in prediction accuracy. This result highlights the importance of consistency in occupancy and how the model prediction accuracy can be degraded with more stochasticity in occupant behavior. Although SVM can be computationally expensive, several fitting algorithms can be utilized to minimize the computational time such as parallelizing the training work [69]. Zhao and Magoulès [70] utilized a parallel implementation approach for predicting a building's energy consumption that reduces the training time by parallelizing kernel evaluations and gradients compared to a sequential approach and provides similar prediction accuracy. Table 4 provides some reported studies applying SVM for building energy predictions.

The Support Vector Machine is a powerful yet computationally expensive algorithm. The mapping of observation to a higher dimension makes SVM superior in fitting complex relationships and minimizing the model prediction errors. Parallelization can mitigate the slow-fitting performance of the SVM approach, especially when dealing with large datasets and when accuracy distribution over the entire dataset is required. The proper choice of kernel when using SVM is not straightforward, as the resulted mapped data points can change the prediction accuracy of the model and the process of fitting a hyperplane with no direct relation to model accuracy. Additionally, the choice of a kernel can be determined usually through k-fold cross validation. However, the number of studies using non-linear models as found by the bibliometric analysis suggests that using kernels such as Gaussian or RBF are more common. Furthermore, linear kernels can fit linear hyperplanes

for non-complex applications as well as non-linear models that are more computationally expensive.

Table 4. Support Vector Machine cases that are applicable to M&V analysis.

Building Type and Number	Features	Data Granularity	Model Type	References
3 Residential Buildings	Date, and Meteorological data	Hourly	SVM, **LS-SVM**	[67]
Simulated Office building	Date, and Meteorological data	Hourly	PI-SVM	[70]
4 Commercial Buildings	Temperature	Monthly	SVM	[71]
Hotel	Date, Meteorological data, and Occupancy	Hourly	SVM with RBF	[72]
Commercial Building	Date, Meteorological data	15-min	SVM with RBF	[73]
Hotel	Date, Meteorological data, and HVAC operation parameters	Hourly	SVM with RBF	[74]

4.4. Artificial Neural Network

Deep learning or artificial neural network (ANN) is a subfield of machine learning where algorithms mimic the human brain functioning process. The ANN involves a set of neurons forming layers that are inter-connected starting from an input layer to an output layer. The connections between neurons are determined using weight coefficients that are determined based on a training process using input–output data sets. As discussed in Section 3, the majority of ANNs used in data-driven building energy modeling are Feed Forward Neural Network (FFNN) [75], as detailed in the following sections.

Feed Forward Neural Network

FFNN is the most commonly used ANN-based approach in building energy modeling. Each layer's neurons, comprised of various features' signals, are multiplied by weights $w_{i,j,k}$ that connect them to the other layer's neurons. Following this, a bias term $b_{j,k}$ is added to the summation of each weight and signal multiplication. The result is then inserted into an activation function, which can be either a Rectified Linear Unit (ReLU) or a linear activation function. Without activation functions, the FFNN would be just a linear regression model. Equation (6) shows the process of multiplying weights with signals and adding bias [75]. Figure 7 illustrates the basic FFNN architecture. Figure 7 shows the same variables in Equation (6) with different indices, where i denotes the layer number, j the neuron in certain layer, and k the connection. For example, $w_{i,j}$ represents the weight $w_{i,j}$ of the connection between node i and j.

$$\hat{y}_{X,W,b,g(h)} = g\left(W^T X + b\right) \tag{6}$$

where

- W: Weights associated with the connection between neurons.
- X: Inputs form the input layer or the output of an activation layer.
- b: Bias term for each neuron.
- $g(h)$: Activation function.

FFNN can have multiple hidden layers (i.e., Multi-Layer Perceptrons, MLP) or a single hidden layer (i.e., Single Layer Perceptrons, SLP). Other forms can have different processes with the same network architecture such as Radial Basis Function Neural Network (RBFNN) [76] or Extreme Learning Machine (ELM) [56]. Both forms have instead of multi-hidden layers, a single hidden layer. RBFNN has radial basis functions that map data to a higher dimension instead of simply activating h. ELM is also a single hidden layer network where initial weights bias terms are initialized using a different method than MLP or SLP and fixed during the tuning phase. Table 5 shows some reported stuides that apply FFNN to predict building energy consumption.
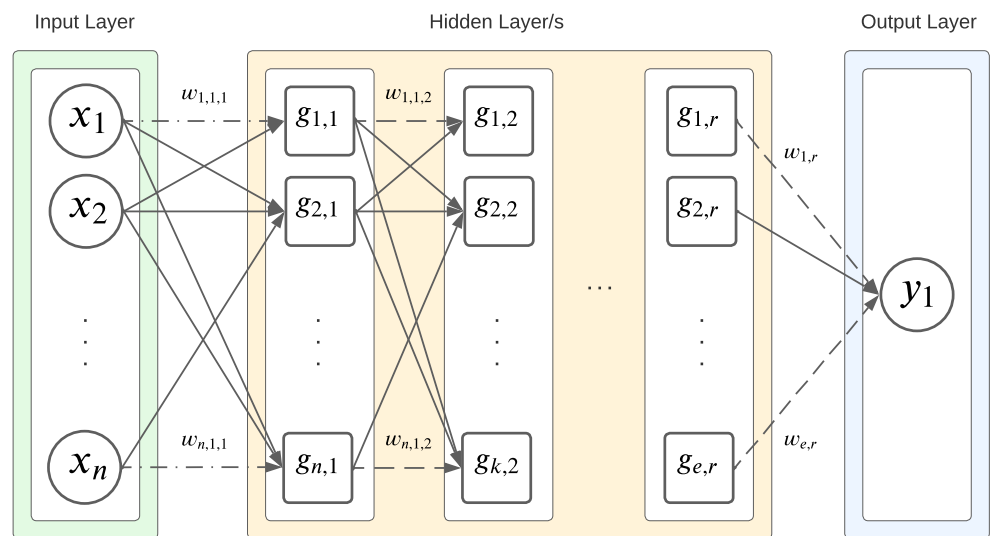
Figure 7. Feed forward neural network architecture.

Table 5. Neural network cases that are applicable to M&V analysis.

Building Type and Number	Features	Data Granularity	Model Type	References
Biomedical manufacturing's chilled water system	HVAC system operating variables	15-min to weekly	SLP	[77]
Office building HVAC hot water system	Outside dry-bulb temperature	Hourly and Daily	MLP	[78]
5 Office buildings	Date, Meteorological data, and HVAC loads	Hourly	SLP	[79]
1 Educational building, 1 real and 2 simulated office buildings	Date and Temperature	Hourly	SLP	[80]
47 buildings in an educational campus	Date and Meteorological data	Hourly	SLP	[81]
7 Dormitory buildings	Date, Meteorological data, and Occupancy	Hourly	SLP	[82]
Library and ASHRAE Energy Prediction Competition I dataset	Date and Meteorological data	Hourly	SLP	[83]
Educational building	Date, Meteorological data, and Occupancy	15-min	SLP	[84]

ANNs are gaining more popularity in building energy modeling due to the availability of better computing machines to perform cumbersome and time-consuming approaches. Furthermore, the development in ANN architecture and algorithms that enable the capture and identification of complex relationships. Nevertheless, the superiority of such methods remains the subject of debate since only slight improvements in prediction accuracy can be achieved at the expense of significant computational efforts. FFNN-based models can take several forms with the choice between them being difficult to generalize to all building energy modeling applications. Typically, the development of FFNN-based models relies on a trial and error process using cross-validation to obtain the best model's parameters with no clear choice in the reported literature on the best general approach that leads to an accurate model's prediction. Some papers recommended certain methods to find the first iteration's parameters' values such as the number of hidden layers and neurons. Ahmad et al. [62] chose only a single hidden layer and performed a stepwise searching method to select the optimum number of neurons. On the other hand, Amber et al. [85] and Ye and Kim [86] relied on a formula that is a function of both the output and input layer sizes to determine the number of neurons.

4.5. Kernel Regression

4.5.1. Definition

Another category of data-driven approaches used for building energy modeling is kernel regression. This category of regression analysis approaches is also called time-varying coefficients, where response values are predicted using different coefficients for different intervals. Kernel, in this context, is a function that assigns weights to data points based on a specific metric [87]. An example of kernel regression is K-Nearest Neighbor (KNN) [88] regression, where Euclidean distance is used as a metric for weighing nearby points where a subset of all data points is selected and each is given an equal weight. Equation (7) defines the K-nearest neighbor regression. However, this method can have boundary issues as the regression becomes inaccurate at the endpoints. Additionally, the method generates a curve with several discontinuities as each point has an equal weight. Another approach is Nadaraya–Watson kernel-weighted average [89], which minimizes the weight points based on distance. Equation (8) shows the calculation of the model predictions.

$$\hat{y} = \frac{1}{h} \sum_{x_i \in N_{h(x)}} y_i \tag{7}$$

where

- N: Neighborhood of points similar based on Euclidean distance.
- h: Count of points in neighborhood N.

$$y = \frac{\sum_{i=1}^{N} K_\lambda(x_0, x_i) y_i}{K_\lambda(x_0, x_i)} \tag{8}$$

where

- K_λ: Kernel equation that weighs points in neighborhood N.

The kernel equation K_λ can be Epanechnikov quadratic, Tri-cubic, or Gaussian [87]. In each kernel function, a hyperparameter λ, named smoothing parameter, determines the local neighborhood's widths where lower and higher values can change the variance and bias of the model.

4.5.2. Applications

Ho and Yu [90] applied a kernel regression using KNN using measured data for an educational building with a special focus on the energy performance of a chilled water plant. The model included typical features for a building and chiller operating variables such as water flow rate, water supply, and return temperatures, as well as outdoor air dry-bulb temperature, and relative humidity. The model achieved reasonable prediction accuracy levels by selecting the optimal number of clusters based on the lowest mean square error. These results highlight the ability of kernel regression to consider several factors and weight them based on Euclidean distance. Gallagher et al. [91] modeled energy use of a biomedical facility using over 18 features (i.e., input parameters) including dry-bulb temperature data and equipment manufacturing variables such as production machinery electricity consumption, facility operation schedule, and chilled water system electricity consumption. The study showed that KNN achieved the best accuracy metrics when using weekly data compared to SVM, ANN, LR, and DT. Wang et al. [92] compared energy use predictions for several data-driven models, stacking, RF, GBM, SVM, XGB, and KNN. The reported results indicate that the KNN-based model has mixed performance as it achieved better accuracy levels than RF and XGB in one case, but provided the worst prediction accuracy in another case. Table 6 shows a summary of the reported studies using kernel regression for building and retrofit baseline energy modeling.

Table 6. Kernel regression cases that are applicable to M&V analysis and utilize real historical data.

Building Type and Number	Features	Data Granularity	Model Type	References
2 Educational buildings	Date and Meteorological data	30-min and Hourly	KNN	[92]
Biomedical manufacturing facility	Date, Temperature, and manufacturing factors	Intervals from 15-min to Monthly	KNN	[91]
Educational building	Date, Meteorological data, and chiller operating variables	15-min	KNN	[90]
68 Commercial and 54 residential buildings	Date	Hourly	KNPTS and KNFTS	[93]
Chiller in a public building	Meteorological data and Chiller operating parameters	15-min	KNN	[94]

Kernel regression approach provides a powerful tool when modeling relations that are observed frequently over the dataset. By developing neighborhoods of similar points, the kernel-based models can make predictions that are based on weighted values. The similarity provides a mean for the kernel-based model to link the mapping between inputs and outputs and easily fit non-linear relations. However, several hyperparameters are encountered when selecting a kernel-based modeling approach. From the reviewed applications, there appears to be no specific selection guidelines for these parameters other than experimentation and trial and error mechanisms. Although complex kernels can produce smooth curves to fit the building energy consumption, there is no clear procedure to develop the set of complex kernels. The common recommendation from reported analyses is that kernel-based models need to be tested over a set of different data and compared against each other to determine the best modeling approach.

5. Feature Engineering

Feature engineering is an important step in developing data-driven models as it can significantly affect the models' accuracy. This process involves manipulating the available dataset to transform it into a set of features that data-driven models use to make predictions specific to a response variable. Feature engineering is similar to data processing where raw data is cleaned and any missing values are replaced or dropped. However, feature engineering utilizes processed data instead of raw data to identify a set of features that enables the model to capture the relationship between predictors and the desired response variable. Machine learning models can be applied to cleaned data without performing any feature engineering, but multiple issues can arise reducing their predictive and explanatory capabilities. Typically, the prediction accuracy levels using training data can increase when a higher number of predictors are included. However, the use of several predictors can lead to over-fitting with data-driven models fitting the noise variations instead of the actual relationship and thus limiting the model's predictive and explanatory capabilities. The aforementioned detrimental effects are important for M&V applications as only historical data sets are available to tune and test the models. Thus, feature engineering can have a significant role in developing sound data-driven models that quantify energy savings accurately when conducting M&V analyses.

For M&V applications, feature engineering can be structured into two main goals: identifying the features to be used in the models and selecting the methods to perform feature engineering. The former emphasizes the mostly used features to predict the energy consumption regardless of the used models while the latter focuses on the approaches to adopt for selecting and/or creating relevant features.

5.1. Features

As discussed in Section 4, the selected features for reported data-driven models vary significantly. For many studies, the model's features are considered based on their availability as limitations and challenges often arise in obtaining and collecting relevant data that are effective in predicting building energy performance. Table 7 lists categories and features used in predicting building energy consumption in the published literature. The most used categories of features include outdoor dry-bulb temperature and time-related parameters. Indeed, time-related features are often considered to predict the time dependency of the energy used by buildings. For instance, Mathieu et al. [38] used the time of week feature, where the week is segmented into 15-min intervals to capture patterns and correlations that occur on a weekly basis. The process of converting a numerical time value into a grouping factor is usually called hot-encoding [95]. Time-related features can include parameters such as month, day type, and holidays. The larger the interval of a time-related feature, the more historical data are needed to assess the contribution of that feature in the development of a relationship between the predictors and the response variable. However, certain relationships to predict a building's energy consumption may target large time periods as demonstrated by Wang et al. [57], who evaluated the effect of occupancy levels during three academic semesters to determine the energy use of an institutional building at the University of Florida.

Another set of important features is meteorological data that include measurements of site weather parameters during the period when the building's energy consumption data are collected. The structure of such data can vary in terms of time granularity and the number of variables depending on the weather station's capabilities. In most weather stations, at least six variables denoting outdoor temperature, humidity, pressure, wind, and precipitation are recorded on a sub-hourly basis [96]. Table 7 lists common meteorological features used to develop the building's energy models and some reported studies. Some of the analyses have indicated the diminishing return of including all meteorological features as the model's prediction accuracy tends to increase slightly, but its overfitting issues increase [38,55]. When including all meteorological features, it is important to consider the multicollinearity between predictors. Indeed, highly correlated predictors can lead to inaccurate estimates of predictors' contributions and ultimately reduce the model's prediction accuracy levels.

The occupancy level is an important feature that can significantly reduce the unexplained variance in predicting a building's energy consumption. Anand et al. [97] used recorded occupancy presence based on Wi-Fi traffic monitoring devices to predict energy use for an institutional building. Although the study found that a large portion of the building's energy is due to office equipment and plug loads, occupancy level is determined to be a contributing feature. Time-related features can be used as occupancy indicators by using schedules. Zeng et al. [73] incorporated six buildings' schedules in modeling the energy consumption and found that the use of occupancy schedules has a significant impact on the model's prediction accuracy, especially for buildings with stable occupants' rate.

Based on Table 2 through Table 6, the minimum required features for developing a baseline building energy model for M&V analysis include date and outdoor dry-bulb temperature. The date feature must be in a timestamp format to indicate the frequency of recording observations, starting, and ending date of historical values. Outdoor dry-bulb temperature is the main feature commonly reported for establishing the relationship between weather and building energy consumption. Meteorological features, other than outdoor temperature, have been utilized by certain reported studies with no or little improvement in prediction accuracy [38,55]. In a number of applications, occupancy rate can significantly improve the prediction accuracy of data-driven models if this data is readily and accurately available. As alternatives to occupancy rates, time-related features using operation schedules and building use patterns have been successfully considered in some applications. It is important to note, however, that as the modeling time step becomes

shorter (e.g., hourly to daily), accurate estimations of occupancy patterns become more challenging.

Table 7. Features used in building energy consumption prediction.

Feature Categories	Feature	References
Date-Related	15-min of an hour	[53,65,76,84]
	Hour	[40,61,63,74,82,84,98,99]
	Day	[40,45,55,57,74,84,84,99]
	Week	[45,57]
	Holiday	[53,55,100]
	Month or/and Biannually	[57,84,99]
Meteorological	Outside Dry-Bulb and/or Wet-bulb Temperature	[38,40,45,53–55,57,65,70,72,74,76,84]
	Relative Humidity and/or Humidity Ratio	[40,45,57,74,84]
	Solar Irradiance	[40,55,57,84]
	Enthalpy	[45,55]
	Wind Direction and/or Speed	[40,55,57,84]
Occupancy-Related	Infrared Sensors and/or recorded Equipment use	[40,74,84]
	Schedules and Records	[57,73]
Operation-Related	Indoor Dry-Bulb Temperature	[63,74]
	Building Systems' Operating Variables	[45,65,74,99]

5.2. Feature Processing and Extraction

In processing analysis for developing baseline building energy models for M&V analysis, time and outdoor air temperature (i.e., dry-bulb and/or wet-bulb air temperatures) are the most commonly used features. Often, outdoor air temperature, especially when it fluctuates significantly, can reduce the ability of data-driven models in providing accurate predictions of building energy consumption. Therefore, this feature is usually processed to have a predictor with less variation frequency such that the building's thermal dynamics can be easily explained [101]. Table 8 lists some predictors that are calculated based on outdoor air temperatures including Cooling Degree Days (CDD), change-point temperatures, and piece-wise fitting models. Such predictors provide surrogate variables to determine the impact of the outdoor air temperature variations on the building's thermal performance. In terms of time-related features, hot-encoding and factoring provide a set of features with categorical instead of numerical variables. Examples of these categorical features are listed in Table 7, where timestamps of 15-min intervals can be converted into a set of categorical variables such as hour, day type, and holidays. The last two processing methods indicated in Table 8 do not rely on domain knowledge, but are based on algorithms such as PCA and Deep Learning Extraction to develop a set of new features to enhance the predictive capabilities of data-driven models.

5.3. Feature Selection

The selection analysis in feature engineering involves the conversion of the original dataset into a smaller dataset with fewer features. As noted earlier, increasing the number of features has not only a limited effect on improving the model's prediction accuracy, but may result in an overfitted model with reduced explanatory capabilities. At the initial development phase of a data-driven model, EDA-based methods can be used to extract insights into the relationship between the predictors and the response variable. The EDA-based methods include pairwise correlation and plotting against the response variable [56].

However, data double-dipping must be avoided as identified relationships between the response variable and the predictors using the EDA approach based on the same training dataset can lead to enforcing relationships that do not necessarily exist outside the analyzed dataset. Gallagher et al. [77] used developed a M&V modeling methodology where a feature selection pipeline of two metrics reduced a dataset of 504 to 15 feature. The pipeline sorts several features by iteratively using Spearman correlation coefficient between features to select the optimum set of features that maximize the coefficient of determination R^2. Following this, the variance inflation factor (VIF) is utilized to remove any possible multicollinearity where features with a VIF greater than 5 are dropped.

Another method to filter the features is by a forward and backward elimination approach, with the model recursively trained on a different set of features and its performance is assessed using various evaluation metrics. These evaluation metrics, such as Akaike Information Criterion (AIC) [102], can be used to balance the number of features and the accuracy level. Feature importance methods [103], often used in DT-based models, such as RF and XGB, can be considered for the selection analysis with features splitting the leaves having a high contribution in improving the model's prediction accuracy are identified to have higher importance. In M&V applications, the process of selecting features relies on the availability of more features by having multiple meters, weather data that includes several parameters, or multiple occupancy sensors that are laid on several locations inside the building. Moreover, generated time-related features (e.g., day of a month, month of a year) might be dropped if the received information gain does not lead to an improvement in prediction accuracy. A list of selection processes in building energy modeling are summarized in Table 8.

Table 8. Feature engineering methods in building energy consumption prediction.

Feature Engineering Category	Feature	Method	References
Processing	Outdoor Temperature	CDD and HDD	[45,104]
		Change-Point	[54,104]
		Piece-wise Fitting	[38,53,105]
	Time	Hot-encoding	[38,53,55,100,105]
	All or Multiple Features	Clustering	[98,105]
	All or Multiple Features	PCA	[73,83,106]
	All or Multiple Features	Deep Feature Extraction	[107]
Selection	All or Multiple Features	Forward and Backward Selection	[54,67]
	All or Multiple Features	Feature Importance	[55,63–65]
	All or Multiple Features	EDA	[56,63,73,82,97]

6. Data Requirements

Data-driven models for M&V rely heavily on historical data to establish a relationship between input variables and building energy performance. The quality and quantity of historical data can significantly affect the accuracy of a data-driven model. In particular, the following three characteristics are often used to assess the quality and quantity of the data: time range, reporting frequency, and missing values. Data time range affects the re-occurrence of certain performance levels that can help models identify repeating patterns or ignore unusual activities. Grillone et al. [105] simulated 54 cases of three buildings with different parameters and trained two data-driven models using data specific to a period ranging from 9 to 12 months. The result showed that a significant decrease in the prediction accuracy distribution median and an increase in the distribution variance when using TOWT approach. OpenEEmeter [108] is an open-source framework used to calculate the energy use that could be avoided by retrofitting a building. The framework sets certain requirements on the data used for developing a building energy model including the data

time range. For data with hourly and daily frequency, an OpenEEmeter compliant baseline building energy model requires at least data for 365 days.

The time frequency provides the level and type of information that can be gained from data through data-driven modeling. Using hourly or sub hourly data for energy consumption, any patterns and correlations can be learned better, but more significant noise levels could be introduced as the building energy consumption becomes less consistent. On the other hand, aggregated consumption using daily or monthly frequencies exhibit less fluctuations at the expenses of extracting more information. Gallagher et al. [77] analyzed the effect of sub-hourly, hourly, daily, and weekly frequency on four data-driven models by using a recorded measurement of a chilled water system. They found that the frequency effect varied between models with daily frequency producing the lowest CV(RMSE), except for KNN, where the hourly-based model resulted in a lower CV(RMSE).

Missing values represent another important requirement for the quality of data needed for training. Missing values are identified by periods of disconnected metering, irregular values, or missing some features' values during a given timestamp. Although each case of missing data is usually unique and requires a certain imputation technique, several thresholds were established to prevent training models from using invalid data sets. Cal-TRACK [109] dictates missing data requirements for daily and hourly frequencies specific to data-driven models. Models based on daily data must not have more than 37 days (i.e., 10% for a full-year data) while hourly frequency data must have less than 10% missing hours of the total hours in every calendar month.

7. Existing Open-Source M&V Frameworks

Several papers developed specific data-driven models for building energy modeling and M&V analysis as shown in Table 2 through Table 6. The process of developing such models involves specific knowledge of the underlying statistics and building energy consumption patterns and is limited to the considered building type and location. To automate this process, only limited analysis frameworks have been proposed with varying degree of capabilities and applications. Although such frameworks still require some knowledge of the underlying process, they still facilitated the development of baseline energy modeling needed for performing an M&V analysis. Table 9 lists the limited open-source frameworks and their characteristics suitable for M&V analysis.

Table 9. Sample of existing M&V frameworks.

Framework	Models	Inputs	Frequency	Development Language	Reference
ECAM	LR	Date, Dry-bulb temperature, and Occupancy	Hourly, Daily, and Monthly	Excel add-in	[110]
EEMeter	LR, GBM	Date, Dry-bulb temperature, and Occupancy	Hourly, Daily, and Monthly	Python	[108]
NMECR	LR	Date, Dry-bulb temperature, Occupancy, and independent variables	Hourly, Daily, and Monthly	R	[111]
RMV2.0	LR	Date, Dry-bulb temperature, and Occupancy	Hourly	R	[112]

8. Models Evaluation

To accurately estimate energy savings associated with retrofitting buildings, baseline models need to be established to provide predictions with a satisfactory accuracy level using specific qualities. Several such qualities have already been discussed throughout this review including generality, predictive, and explanatory capabilities. The process of obtaining a model with the best aforementioned qualities requires a selection of both evaluation metrics and approaches.

8.1. Evaluation Approaches

To have good predictive performance, data-driven models must balance bias and variance effects to provide accurate predictions even using unforeseen conditions. To enhance their prediction accuracy, models are often trained using a fraction of the available historical dataset until acceptable accuracy levels are reached without compromising other model performance metrics such as overfitting. Then, the trained models are tested on a new data set not used in the training analysis to evaluate their performance using metrics such as prediction accuracy level. A basic approach of evaluation is to split the data into training and testing dataset. With cross-sectional data, the split is random, but with time series, the split must be carried to have two sets of data that are similar such that the model testing process is valid. An approach to split time series data is to set a date at the end of the data where observations after such date are left for testing. However, this approach is limited if the some of features' values does not occur in the training set, a case that can happen when splitting with a training data of less than a year time range.

Another approach is using k-fold cross validation where the whole data set is segmented into k consecutive blocks and the model is trained on all blocks except one which is left for testing. Following this, the process is iterated by changing training and testing blocks k times and averaging the resulted evaluation metric. There is no specific number for the cross validation folds (i.e., blocks) but in practice, a value of 5 or 10 is usually chosen [37].

8.2. Evaluation Metrics

Regardless of the followed evaluation approach, different evaluation metrics can be used to indicate a desired quality of the trained model. One metric commonly used for model performance evaluation is the coefficient of determination, R^2, the explained variance as expressed by Equation (9) where $\hat{y}$ represents predicted values.

$$R^2 = \sum_{i=1}^{n} \frac{(\hat{y} - \bar{y})^2}{(y - \bar{y})^2} \tag{9}$$

However, R^2 does provide an accurate performance metric when comparing multiple models having a different number of predictors. Therefore, the adjusted coefficient of determination R^2_{adj} is used such that a complex model with a high number of predictors is penalized as shown by Equation (10):

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - P - 1} \tag{10}$$

where p is the of predictors and n is the number of training data points. Similar metrics to R^2_{adj} are AIC and Bayesian Information Criterion (BIC), where the model is penalized as its complexity increases.

For M&V applications, the normalized mean bias error (NMBE) and coefficient of variation of the root mean squared error (CV(RMSE)) are the most commonly used performance metrics to assess the model's prediction accuracy levels. The NMBE measures the overall bias in the model's predictions as defined by Equation (11), with a positive value indicating that the model is on average over-predicting and a negative value being an indicator that the model is under-predicting.

$$NMBE = \frac{100}{n} \times \frac{\sum_{i=1}^{n} y - \hat{y}}{\bar{y}} \tag{11}$$

The NMBE is invariant of the time granularity as models with 15-min or hourly interval predictions result in the same NMBE value. The CV(RMSE) measures the difference

between actual and predicted values as expressed by Equation (12) where the differences are squared instead of being summed as the case for estimating the NMBE value.

$$CV(RMSE) = 100 \times \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2}}{\bar{y}} \tag{12}$$

The CV(RMSE) is sensitive to the model's prediction time granularity so its value for a model with 15-min prediction intervals is different from the value obtained from another model with hourly prediction intervals. Both metrics are important in evaluating and comparing the performance of several models.

9. Summary and Conclusions

To justify retrofitting buildings, M&V analysis is often needed to quantify the achieved energy savings and ultimately justify the cost-effectiveness of implemented energy efficiency measures. Data-driven modeling provides an effective approach to perform M&V analysis when compared to traditional deterministic modeling methods especially considering the growing availability of significant historical data specific to historical energy performance due to advancements in metering and monitoring of building energy systems.

In this review of the existing literature, several data-driven building energy models that are suitable for M&V applications have been described and evaluated. In particular, five categories of data-driven modeling approaches have been identified for M&V analysis of retrofitted building energy systems. The simplest data-driven modeling option consists of LR with the TOWT approach, which is found to be widely used for developing baselines of existing buildings. The TOWT method is incorporated in two of the mentioned existing frameworks (i.e., EEMeter and RMV2.0), and all the mentioned studies with an hourly frequency use such a method to build LR models. The ensemble modeling approach has two prominently applied methods for assessing building energy performance including RF and XGB. The two modeling approaches were mentioned in almost all the reported papers using ensemble approaches and, in every one of them, either XGB and RF were the models scoring best in prediction accuracy compared to the remaining ensemble methods. Several data-driven models have been developed using the SVM approach combined with a range of hyperparameters. However, there are no clear guidelines from the reported literature on determining the best combination of hyperparameters suitable for M&V analysis of building energy savings. In addition, a wide range of FFNN-based models has been considered to predict building energy performance with different architectures and features. Among the reported FFNN's architectures, SLP is mostly used in predicting building energy consumption. One study suggests that there was no improvement in prediction accuracy when changing SLP to MLP by one hidden layer [79]. Lastly, kernel regression methodology has been applied for building energy prediction, with KNN being mostly used, especially for M&V applications.

Two important features used in most of the data-driven models reported for building energy prediction and M&V analysis include date and outdoor dry-bulb temperature. Another effective feature considered for several data-driven modeling consists of occupancy pattern derived from indoor sensing and/or operating schedules. In terms of selecting features, EDA and feature importance by ensemble methods were demonstrated to be the widely used methods for selecting the optimum features. The popular processing techniques were applied to date and outdoor dry-bulb temperature with hot-encoding being popular for time-related features. For temperature, CDD and HDD transformation is popular for data with low frequency, while change-point and piece-wise fitting is used mainly for linear regression-based models.

Existing popular frameworks for M&V analysis were discussed along with features, modeling approaches, and used features. The usual data requirements for building a M&V baseline were derived from studies and frameworks' requirements. Important requirements were discussed including data range, frequency, and missing values. The

smallest data range for building a baseline was one year before retrofitting regardless of the data frequency. Results from reported studies demonstrated that the highest prediction accuracy usually comes with an hourly or daily frequency since sub-hourly data introduce more noise than information while lower frequencies such as weekly or monthly lack usage patterns. Few studies discussed the effect of missing data, but an emphasis was made to have less consecutive missing data as data imputation becomes difficult.

Finally, the paper discussed several evaluation performance metrics and approaches to assess the prediction accuracy of the baseline building energy model. In particular, evaluation metrics for both general building energy prediction and M&V analysis were discussed with CV(RMSE) and NMBE being the mostly used metrics to evaluate the building energy models. These two metrics complement each other and convey better information about the model's performance. Two other evaluation approaches were outlined with their drawbacks and benefits: split without shuffle and k-folds evaluation. With sufficient data covering more than a year of building energy consumption, split without shuffle approach provides an easy and efficient evaluation metric, while the k-folds approach tests the generality of the model better. However, the selection of the evaluation approach is still dependent on the building case.

It is clear from the presented review analysis that there is a need for a general framework and a set of guidelines to develop advanced data-driven models suitable for M&V analysis and capable to estimate accurately energy savings achieved by building retrofits. While the review has revealed some existing frameworks, all of them are based mostly on LR. Several papers indicated that modeling approaches deliver varying prediction accuracy and that there is no best modeling approach for every M&V analysis. Moreover, retrofit analyses with advanced data-driven modeling approaches are currently developed only for specific case studies and their application cannot be readily generalized to any building type and location. If established, the proposed framework will enhance the use of data-driven models for various applications of building energy analysis including M&V of energy saving from retrofit projects.

Nomenclature

AIC	Akaike Information Criterion
AMI	Advanced Metering Infrastructure
ANN	Artificial Neural Network
ASHRAE	American Society of Heating Ventilation Refrigeration and Air-conditioning Engineers
BIC	Bayesian Information Criterion
BMS	Building Management System
CDD	Cooling Degree Days
DT	Decision Tree
ECM	Energy Conservation Measure
EIA	Energy Information Administration
ELM	Extreme Learning Machine
FFNN	Feed Forward Neural Network
GBM	Gradient Boosting Machine
HVAC	Heating Ventilation and Air Conditioning
IPMVP	International Performance Measurement and Verification Protocol

KNN	K-Nearest Neighbor
LCA	Life Cycle Assessment
LGBM	Light Gradient Boosting Machine
LR	Linear Regression
LS-SVM	Least Squares Support Vector Machine
M&V	Measurement and Verification
MBE	Mean Bias Error
MLP	Multilayer Percepton
NARX	Nonlinear Autoregressive with Exogenous inputs
NMBE	Normalized Mean Bias Error
OECD	Organization for Economic Cooperation and Development
OEDI	Open Energy Data Initiative
OLS	Ordinary Least Square
PI-SVM	Parallel Implemented Support Vector Machine
RBFNN	Radial Basis Function Neural Network
RC	Resistance and Capacitance
RF	Random Forest
ReLU	Rectified Linear Unit
RSS	Residual Sum of Squares
SLP	Single Layer Perceptron
SVM	Support Vector Machine
TOWT	Time of Week and Temperature
VIF	Variance Inflation Factor
WLS	Weighted Least Squares
XGB	Extreme Gradient Boosting Machine

References

1. Global Alliance for Buildings and Construction. *Global Status Report for Buildings and Construction*; Global Alliance for Buildings and Construction: Nairobi, Kenya, 2020.
2. EIA, US. *Energy Information Administration "International Energy Outlook"*; Technical Report; US Department of Energy: Washington, DC, USA, 2021.
3. Allouhi, A.; El Fouih, Y.; Kousksou, T.; Jamil, A.; Zeraouli, Y.; Mourad, Y. Energy consumption and efficiency in buildings: Current status and future trends. *J. Clean. Prod.* **2015**, *109*, 118–130. [CrossRef]
4. Guo, Y.Y. Revisiting the building energy consumption in China: Insights from a large-scale national survey. *Energy Sustain. Dev.* **2022**, *68*, 76–93. [CrossRef]
5. Thonipara, A.; Runst, P.; Ochsner, C.; Bizer, K. Energy efficiency of residential buildings in the European Union – An exploratory analysis of cross-country consumption patterns. *Energy Policy* **2019**, *129*, 1156–1167. [CrossRef]
6. Almasri, R.A.; Alshitawi, M. Electricity consumption indicators and energy efficiency in residential buildings in GCC countries: Extensive review. *Energy Build.* **2022**, *255*, 111664. [CrossRef]
7. Satchwell, A.; Piette, M.A.; Khandekar, A.; Granderson, J.; Frick, N.M.; Hledik, R.; Faruqui, A.; Lam, L.; Ross, S.; Cohen, J.; et al. *A National Roadmap for Grid-Interactive Efficient Buildings*; Technical Report; Lawrence Berkeley National Lab. (LBNL): Berkeley, CA, USA, 2021.
8. *Net Zero Strategy : Build Back Greener*; HM Government: London, UK, 2021.
9. Mallapaty, S. How China could be carbon neutral by mid-century. *Nature* **2020**, *586*, 482–484. [CrossRef]
10. Roberts, S. Altering existing buildings in the UK. *Energy Policy* **2008**, *36*, 4482–4486. [CrossRef]
11. Hasik, V.; Escott, E.; Bates, R.; Carlisle, S.; Faircloth, B.; Bilec, M.M. Comparative whole-building life cycle assessment of renovation and new construction. *Build. Environ.* **2019**, *161*, 106218. [CrossRef]
12. EIA, US. *Sustainable Recovery*; Technical Report; US Department of Energy: Washington, DC, USA, 2020.
13. Wei, Y.; Zhang, X.; Shi, Y.; Xia, L.; Pan, S.; Wu, J.; Han, M.; Zhao, X. A review of data-driven approaches for prediction and classification of building energy consumption. *Renew. Sustain. Energy Rev.* **2018**, *82*, 1027–1047. [CrossRef]
14. Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A review on time series forecasting techniques for building energy consumption. *Renew. Sustain. Energy Rev.* **2017**, *74*, 902–924. [CrossRef]
15. Grillone, B.; Danov, S.; Sumper, A.; Cipriano, J.; Mor, G. A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. *Renew. Sustain. Energy Rev.* **2020**, *131*, 110027. [CrossRef]
16. Deb, C.; Schlueter, A. Review of data-driven energy modelling techniques for building retrofit. *Renew. Sustain. Energy Rev.* **2021**, *144*, 110990. [CrossRef]
17. *International Performance Measurement & Verification Protocol*; Efficiency Valuation Organization: Washington, DC, USA 2016.

18. American Society of Heating; U.S. Green Building Council; Chartered Institution of Building Services Engineers. *Performance Measurement Protocols for Commercial Buildings*; American Society of Heating Refrigerating and Air-Conditioning Engineers: Atlanta, GA, USA, 2010.

19. Ma, Z.; Cooper, P.; Daly, D.; Ledo, L. Existing building retrofits: Methodology and state-of-the-art. *Energy Build.* **2012**, *55*, 889–902. [CrossRef]

20. Karnouskos, S.; Terzidis, O.; Karnouskos, P. An advanced metering infrastructure for future energy networks. In *New Technologies, Mobility and Security*; Springer: Dordrecht, The Netherlands, 2007; pp. 597–606.

21. Kupser, J.; Francois, S.; Rego, J.; Steele-Mosey, P.; Galvin, T.; McDonald, C. M&V 2.0: Hype vs. reality. In Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, USA, 21–26 August 2016.

22. Crawley, D.B.; Lawrie, L.K.; Winkelmann, F.C.; Buhl, W.; Huang, Y.; Pedersen, C.O.; Strand, R.K.; Liesen, R.J.; Fisher, D.E.; Witte, M.J.; et al. EnergyPlus: Creating a new-generation building energy simulation program. *Energy Build.* **2001**, *33*, 319–331. [CrossRef]

23. Solar Energy Laborataory, University of Wisconsin-Madison. *TRNSYS, a Transient Simulation Program*; Loose-leaf for updating; 31 March 1975; This manual, and the TRNSYS program it describes, were developed under grants from the RANN program of the National Science Foundation (Grant GI 34029), and from the Energy Research and Development Administration (Contract E(11-1)-2588); Solar Energy Laborataory, University of Wisconsin-Madison: Madison, WI, USA, 1975.

24. Winkelmann, F.C.; Birdsall, B.E.; Buhl, W.F.; Ellington, K.L.; Erdem, A.E.; Hirsch, J.J.; Gates, S. *DOE-2 Supplement: Version 2.1E*; Lawrence Berkeley Lab.: Berkeley, CA, USA; James J Hirsch & Associates: Camarillo, CA, USA, 1993. [CrossRef]

25. DesignBuilder Software. DesignBuilder. Available online: https://designbuilder.co.uk (accessed on 10 October 2022)

26. Siegele, D.; Leonardi, E.; Ochs, F. A new MATLAB Simulink Toolbox for Dynamic Building Simulation with BIM and Hardware in the Loop compatibility. In Proceedings of the Building Simulation, Rome, Italy, 2–4 September 2019.

27. Wetter, M.; Zuo, W.; Nouidui, T.S.; Pang, X. Modelica buildings library. *J. Build. Perform. Simul.* **2014**, *7*, 253–270. [CrossRef]

28. Ke, M.T.; Yeh, C.H.; Jian, J.T. Analysis of building energy consumption parameters and energy savings measurement and verification by applying eQUEST software. *Energy Build.* **2013**, *61*, 100–107. [CrossRef]

29. Piccinini, A.; Hajdukiewicz, M.; Keane, M.M. A novel reduced order model technology framework to support the estimation of the energy savings in building retrofits. *Energy Build.* **2021**, *244*, 110896. [CrossRef]

30. Giretti, A.; Vaccarini, M.; Casals, M.; Macarulla, M.; Fuertes, A.; Jones, R. Reduced-order modeling for energy performance contracting. *Energy Build.* **2018**, *167*, 216–230. [CrossRef]

31. Chen, Y.; Guo, M.; Chen, Z.; Chen, Z.; Ji, Y. Physical energy and data-driven models in building energy prediction: A review. *Energy Rep.* **2022**, *8*, 2656–2671. [CrossRef]

32. Web of Science. Clavirate Analytics London, UK. Available online: https://www.webofscience.com/wos/woscc/basic-search (accessed on 3 January 2022).

33. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

34. Liang, J.; Qiu, Y.; James, T.; Ruddell, B.L.; Dalrymple, M.; Earl, S.; Castelazo, A. Do energy retrofits work? Evidence from commercial and residential buildings in Phoenix. *J. Environ. Econ. Manag.* **2018**, *92*, 726–743. [CrossRef]

35. Wang, Z.; Liu, Q.; Zhang, B. What kinds of building energy-saving retrofit projects should be preferred? Efficiency evaluation with three-stage data envelopment analysis (DEA). *Renew. Sustain. Energy Rev.* **2022**, *161*, 112392. [CrossRef]

36. Brodt-Giles, D.; Rossol, M. *Open Energy Data Initiative: Advancing Analytics and Research Innovation through Improved Data Access*; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2019. Available online: https://www.osti.gov/biblio/1545983 (accessed on 12 May 2022).

37. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. Linear model selection and regularization. In *Springer Texts in Statistics*; Springer: New York, NY, USA, 2013; pp. 203–264.

38. Mathieu, J.L.; Price, P.N.; Kiliccote, S.; Piette, M.A. Quantifying Changes in Building Electricity Use, With Application to Demand Response. *IEEE Trans. Smart Grid* **2011**, *2*, 507–518. [CrossRef]

39. Granderson, J.; Touzani, S.; Custodio, C.; Sohn, M.D.; Jump, D.; Fernandes, S. Accuracy of automated measurement and verification (m andV) techniques for energy savings in commercial buildings. *Appl. Energy* **2016**, *173*, 296–308. [CrossRef]

40. Kim, M.K.; Kim, Y.S.; Srebric, J. Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: Artificial neural network vs. linear regression. *Sustain. Cities Soc.* **2020**, *62*, 102385. [CrossRef]

41. Wang, H.; Xue, Y.; Mu, Y. Assessment of energy savings by mechanical system retrofit of existing buildings. *Procedia Eng.* **2017**, *205*, 2370–2377. [CrossRef]

42. Aris, S.; Dahlan, N.; Mohd Nawi, M.N.; Nizam, T.; Tahir, M. Quantifying energy savings for retrofit centralized hvac systems at Selangor state secretary complex. *J. Teknol.* **2015**, *77*, 93–100. [CrossRef]

43. Grillone, B.; Mor, G.; Danov, S.; Cipriano, J.; Lazzari, F.; Sumper, A. Baseline Energy Use Modeling and Characterization in Tertiary Buildings Using an Interpretable Bayesian Linear Regression Methodology. *Energies* **2021**, *14*, 5556. [CrossRef]

44. Zeng, A.; Ho, H.; Yu, Y. Prediction of building electricity usage using Gaussian Process Regression. *J. Build. Eng.* **2020**, *28*, 101054. [CrossRef]

45. Shin, M.; Do, S.L. Prediction of cooling energy use in buildings using an enthalpy-based cooling degree days method in a hot and humid climate. *Energy Build.* **2016**, *110*, 57–70. [CrossRef]

46. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. Trees and rules. In *Data Mining*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 209–242.

47. Witten, I.H.; Frank, E.; Hall, M.A.; Pal, C.J. Ensemble learning. In *Data Mining*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 479–501.

48. Schapire, R.E. Explaining adaboost. In *Empirical Inference*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 37–52.

49. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

50. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International Conference On Knowledge Discovery And Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

51. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3149–3157.

52. Malistov, A.; Trushin, A. Gradient Boosted Trees with Extrapolation. In Proceedings of the 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, 16–19 December 2019; pp. 783–789. [CrossRef]

53. Touzani, S.; Granderson, J.; Fernandes, S. Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy Build.* **2018**, *158*, 1533–1543. [CrossRef]

54. Afroz, Z.; Gunay, H.B.; O'Brien, W.; Newsham, G.; Wilton, I. An inquiry into the capabilities of baseline building energy modelling approaches to estimate energy savings. *Energy Build.* **2021**, *244*, 111054. [CrossRef]

55. Agenis-Nevers, M.; Wang, Y.; Dugachard, M.; Salvazet, R.; Becker, G.; Chenu, D. Measurement and Verification for multiple buildings: An innovative baseline model selection framework applied to real energy performance contracts. *Energy Build.* **2021**, *249*, 111183. [CrossRef]

56. Liu, Y.; Chen, H.; Zhang, L.; Feng, Z. Enhancing building energy efficiency using a random forest model: A hybrid prediction approach. *Energy Rep.* **2021**, *7*, 5003–5012. [CrossRef]

57. Wang, Z.; Wang, Y.; Zeng, R.; Srinivasan, R.S.; Ahrentzen, S. Random Forest based hourly building energy prediction. *Energy Build.* **2018**, *171*, 11–25. [CrossRef]

58. Wang, R.; Lu, S.; Li, Q. Multi-criteria comprehensive study on predictive algorithm of hourly heating energy consumption for residential buildings. *Sustain. Cities Soc.* **2019**, *49*, 101623. [CrossRef]

59. Dong, Z.; Liu, J.; Liu, B.; Li, K.; Li, X. Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification. *Energy Build.* **2021**, *241*, 110929. [CrossRef]

60. Wang, Z.; Wang, Y.; Srinivasan, R.S. A novel ensemble learning approach to support building energy use prediction. *Energy Build.* **2018**, *159*, 109–122. [CrossRef]

61. Cao, L.; Li, Y.; Zhang, J.; Jiang, Y.; Han, Y.; Wei, J. Electrical load prediction of healthcare buildings through single and ensemble learning. *Energy Rep.* **2020**, *6*, 2751–2767. [CrossRef]

62. Ahmad, M.W.; Mourshed, M.; Rezgui, Y. Trees vs. Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy Build.* **2017**, *147*, 77–89. [CrossRef]

63. Yan, L.; Liu, M. A simplified prediction model for energy use of air conditioner in residential buildings based on monitoring data from the cloud platform. *Sustain. Cities Soc.* **2020**, *60*, 102194. [CrossRef]

64. Huang, Y.; Yuan, Y.; Chen, H.; Wang, J.; Guo, Y.; Ahmad, T. A novel energy demand prediction strategy for residential buildings based on ensemble learning. *Energy Procedia* **2019**, *158*, 3411–3416. Innovative Solutions for Energy Transitions, [CrossRef]

65. Bataineh, A.S.A. A gradient boosting regression based approach for energy consumption prediction in buildings. *Adv. Energy Res.* **2019**, *6*, 91–101.

66. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. Support Vector Machines. In *Springer Texts in Statistics*; Springer texts in statistics; Springer: New York, NY, USA, 2013; pp. 337–372.

67. Edwards, R.E.; New, J.; Parker, L.E. Predicting future hourly residential electrical consumption: A machine learning case study. *Energy Build.* **2012**, *49*, 591–603. [CrossRef]

68. Amber, K.; Ahmad, R.; Aslam, M.; Kousar, A.; Usman, M.; Khan, M. Intelligent techniques for forecasting electricity consumption of buildings. *Energy* **2018**, *157*, 886–893. [CrossRef]

69. Chang, E.Y. Psvm: Parallelizing support vector machines on distributed computers. In *Foundations of Large-Scale Multimedia Information Management and Retrieval*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 213–230.

70. Zhao, H.X.; Magoulès, F. Parallel Support Vector Machines Applied to the Prediction of Multiple Buildings Energy Consumption. *J. Algorithms Comput. Technol.* **2010**, *4*, 231–249. [CrossRef]

71. Dong, B.; Cao, C.; Lee, S.E. Applying support vector machines to predict building energy consumption in tropical region. *Energy Build.* **2005**, *37*, 545–553. [CrossRef]

72. Borowski, M.; Zwolińska, K. Prediction of Cooling Energy Consumption in Hotel Building Using Machine Learning Techniques. *Energies* **2020**, *13*, 6226. [CrossRef]

73. Zeng, A.; Liu, S.; Yu, Y. Comparative study of data driven methods in building electricity use prediction. *Energy Build.* **2019**, *194*, 289–300. [CrossRef]

74. Shao, M.; Wang, X.; Bu, Z.; Chen, X.; Wang, Y. Prediction of energy consumption in hotel buildings via support vector machines. *Sustain. Cities Soc.* **2020**, *57*, 102128. [CrossRef]

75. Aggarwal, C.C. Training Deep Neural Networks. In *Neural Networks and Deep Learning*; Springer International Publishing: Cham, Switzerland, 2018; pp. 105–167.

76. Li, C.; Ding, Z.; Zhao, D.; Yi, J.; Zhang, G. Building Energy Consumption Prediction: An Extreme Deep Learning Approach. *Energies* **2017**, *10*, 1525. [CrossRef]

77. Gallagher, C.V.; Leahy, K.; O'Donovan, P.; Bruton, K.; O'Sullivan, D.T. Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0. *Energy Build.* **2018**, *167*, 8–22. [CrossRef]

78. Zhang, Y.; O'Neill, Z.; Dong, B.; Augenbroe, G. Comparisons of inverse modeling approaches for predicting building energy performance. *Build. Environ.* **2015**, *86*, 177–190. [CrossRef]

79. Gunay, B.; Shen, W.; Newsham, G. Inverse blackbox modeling of the heating and cooling load in office buildings. *Energy Build.* **2017**, *142*, 200–210. [CrossRef]

80. Ridwana, I.; Nassif, N.; Choi, W. Modeling of building energy consumption by integrating regression analysis and artificial neural network with data classification. *Buildings* **2020**, *10*, 198. [CrossRef]

81. Walker, S.; Khan, W.; Katic, K.; Maassen, W.; Zeiler, W. Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. *Energy Build.* **2020**, *209*, 109705. [CrossRef]

82. Song, K.; Kwon, N.; Anderson, K.; Park, M.; Lee, H.S.; Lee, S. Predicting hourly energy consumption in buildings using occupancy-related characteristics of end-user groups. *Energy Build.* **2017**, *156*, 121–133. [CrossRef]

83. Li, K.; Hu, C.; Liu, G.; Xue, W. Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. *Energy Build.* **2015**, *108*, 106–113. [CrossRef]

84. Pombeiro, H.; Santos, R.; Carreira, P.; Silva, C.; Sousa, J.M. Comparative assessment of low-complexity models to predict electricity consumption in an institutional building: Linear regression vs. fuzzy modeling vs. neural networks. *Energy Build.* **2017**, *146*, 141–151. [CrossRef]

85. Amber, K.; Aslam, M.; Hussain, S. Electricity consumption forecasting models for administration buildings of the UK higher education sector. *Energy Build.* **2015**, *90*, 127–136. [CrossRef]

86. Ye, Z.; Kim, M.K. Predicting electricity consumption in a building using an optimized back-propagation and Levenberg–Marquardt back-propagation neural network: Case study of a shopping mall in China. *Sustain. Cities Soc.* **2018**, *42*, 176–183. [CrossRef]

87. Harrell, F.E., Jr. *Regression Modeling Strategies*; Springer Series in Statistics; Springer International Publishing: Cham, Switzerland, 2016.

88. Kramer, O. Unsupervised K-Nearest Neighbor Regression. *arXiv* **2011**, arXiv:1107.3600.

89. Mammen, E.; Marron, J.S. Mass recentred kernel smoothers. *Biometrika* **1997**, *84*, 765–777. [CrossRef]

90. Ho, W.; Yu, F. Chiller system optimization using k nearest neighbour regression. *J. Clean. Prod.* **2021**, *303*, 127050. [CrossRef]

91. Gallagher, C.V.; Bruton, K.; Leahy, K.; O'Sullivan, D.T. The suitability of machine learning to minimise uncertainty in the measurement and verification of energy savings. *Energy Build.* **2018**, *158*, 647–655. [CrossRef]

92. Wang, R.; Lu, S.; Feng, W. A novel improved model for building energy consumption prediction based on model integration. *Appl. Energy* **2020**, *262*, 114561. [CrossRef]

93. Gómez-Omella, M.; Esnaola-Gonzalez, I.; Ferreiro, S.; Sierra, B. k-Nearest patterns for electrical demand forecasting in residential and small commercial buildings. *Energy Build.* **2021**, *253*, 111396. [CrossRef]

94. Ho, W.; Yu, F. Measurement and verification of energy performance for chiller system retrofit with k nearest neighbour regression. *J. Build. Eng.* **2022**, *46*, 103845. [CrossRef]

95. Chandramitasari, W.; Kurniawan, B.; Fujimura, S. Building deep neural network model for short term electricity consumption forecasting. In Proceedings of the 2018 International Symposium on Advanced Intelligent Informatics (SAIN), Yogyakarta, Indonesia, 29–30 August 2018; pp. 43–48.

96. Henson, R. *Meteorology Today*, 12th ed.; CENGAGE Learning Custom Publishing: Mason, OH, USA, 2018.

97. Anand, P.; Deb, C.; Yan, K.; Yang, J.; Cheong, D.; Sekhar, C. Occupancy-based energy consumption modelling using machine learning algorithms for institutional buildings. *Energy Build.* **2021**, *252*, 111478. [CrossRef]

98. Li, K.; Zhang, J.; Chen, X.; Xue, W. Building's hourly electrical load prediction based on data clustering and ensemble learning strategy. *Energy Build.* **2022**, *261*, 111943. [CrossRef]

99. Lei, R.; Yin, J. Prediction method of energy consumption for high building based on LMBP neural network. *Energy Rep.* **2022**, *8*, 1236–1248. [CrossRef]

100. Gao, Y.; Ruan, Y. Interpretable deep learning model for building energy consumption prediction based on attention mechanism. *Energy Build.* **2021**, *252*, 111379. [CrossRef]

101. Bacher, P.; Madsen, H.; Nielsen, H.A.; Perers, B. Short-term heat load forecasting for single family houses. *Energy Build.* **2013**, *65*, 101–112. [CrossRef]

102. Faraway, J.J. *Linear Models with R*, 2nd ed.; Chapman & Hall/CRC Texts in Statistical Science; Chapman & Hall/CRC: Philadelphia, PA, USA, 2014.

103. Saeys, Y.; Abeel, T.; Peer, Y.V.d. Robust feature selection using ensemble feature selection techniques. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Antwerp, Belgium, 15–19 September 2008*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 313–325.
104. Lindelöf, D.; Alisafaee, M.; Borsò, P.; Grigis, C.; Viaene, J. Bayesian verification of an energy conservation measure. *Energy Build.* **2018**, *171*, 1–10. [CrossRef]
105. Grillone, B.; Mor, G.; Danov, S.; Cipriano, J.; Sumper, A. A data-driven methodology for enhanced measurement and verification of energy efficiency savings in commercial buildings. *Appl. Energy* **2021**, *301*, 117502. [CrossRef]
106. Zhang, C.; Cao, L.; Romagnoli, A. On the feature engineering of building energy data mining. *Sustain. Cities Soc.* **2018**, *39*, 508–518. [CrossRef]
107. Fan, C.; Sun, Y.; Zhao, Y.; Song, M.; Wang, J. Deep learning-based feature engineering methods for improved building energy prediction. *Appl. Energy* **2019**, *240*, 35–45. [CrossRef]
108. Phil, N. OpenEEmeter. 2021. Available online: https://github.com/openeemeter/eemeter (accessed on 12 May 2022).
109. CalTRACK. CalTRACK Methods. Available online: https://www.caltrack.org/ (accessed on 12 May 2022).
110. SBW. ECAM (ENERGY CHARTING & METRICS). 2022. Available online: https://sbwconsulting.com/ecam/ (accessed on 12 May 2022).
111. Engineering, K. NMECR (Normalized Metered Energy Consumption). 2022. Available online: https://github.com/kW-Labs/nmecr (accessed on 12 May 2022).
112. LBNL. RMV2.0—LBNL M&V2.0 Tool. 2020. Available online: https://github.com/LBNL-ETA/RMV2.0 (accessed on 12 May 2022).