Article

Predicting Energy Consumption in Residential Buildings Using Advanced Machine Learning Algorithms

Fateme Dinmohammadi [1,2,*], Yuxuan Han [2] and Mahmood Shafiee [3]

1   School of Computing and Engineering, University of West London, London W5 5RF, UK
2   The Bartlett Center of Advanced Spatial Analysis (CASA), University College London (UCL), Gower Street, London WC1E 6BTL, UK
3   School of Mechanical Engineering Sciences, University of Surrey, Guildford GU2 7XH, UK
*   Correspondence: fateme.dinmohammadi@uwl.ac.uk

Abstract: The share of residential building energy consumption in global energy consumption has rapidly increased after the COVID-19 crisis. The accurate prediction of energy consumption under different indoor and outdoor conditions is an essential step towards improving energy efficiency and reducing carbon footprints in the residential building sector. In this paper, a PSO-optimized random forest classification algorithm is proposed to identify the most important factors contributing to residential heating energy consumption. A self-organizing map (SOM) approach is applied for feature dimensionality reduction, and an ensemble classification model based on the stacking method is trained on the dimensionality-reduced data. The results show that the stacking model outperforms the other models with an accuracy of 95.4% in energy consumption prediction. Finally, a causal inference method is introduced in addition to Shapley Additive Explanation (SHAP) to explore and analyze the factors influencing energy consumption. A clear causal relationship between water pipe temperature changes, air temperature, and building energy consumption is found, compensating for the neglect of temperature in the SHAP analysis. The findings of this research can help residential building owners/managers make more informed decisions around the selection of efficient heating management systems to save on energy bills.

Keywords: Net-Zero; energy consumption; residential building; machine learning; prediction

1. Introduction

With the excessive consumption of fossil fuels such as petroleum, natural gas, and coal, environmental pollution and the energy crisis have become two major problems in many developed and developing countries around the world. In response to these problems, many national governments and international agencies have made significant efforts to identify high-energy consumption areas, improve energy efficiency, and replace high-carbon fuels with low- or zero-carbon energy sources such as renewables. According to the U.S. Energy Information Administration (EIA) [1], the residential energy sector accounts for 23% of global energy consumption, making it the third-highest energy-consuming sector after industry (37%), and transportation (28%). With working from home becoming more common and widely accepted by employers after the COVID-19 crisis, it is expected that energy consumption in residential buildings will increase substantially in the future [2]. Therefore, in order to improve energy efficiency and reduce the carbon footprint in the residential building sector, it is important to be able to accurately predict the rising energy consumption demand and determine the different indoor and outdoor factors that contribute the most to residential heating energy consumption.

A brief review of the literature shows that a lot of studies have sought to predict the energy consumption in non-residential buildings, including commercial and educational buildings [3]. However, residential buildings have received limited attention in past studies.

Energy consumption prediction plays an important role in enabling buildings' energy management and control, energy strategy development, and quantification of energy saving potential [4]. In general, there are two types of models used for predicting the building energy consumption: *physical* models and *data-driven* models. The former models are known as 'white box' models, while the latter are known as 'black box' models [3]. The physical models require detailed information about the buildings' physical properties and environmental parameters, such as the thermophysical properties of the material, technical construction constraints, efficiency of the heating, ventilation, and air conditioning (HVAC) system, and outdoor climate, to be able to make predictions about energy consumption. These models are useful during the design phase of building projects, as they can support designers in evaluating the energy performance of various design alternatives. However, as the physical models often simulate activities at a very detailed level and make many evaluations of the model at each simulation iteration, they face computational efficiency problems. Data-driven models have emerged as a computationally efficient alternative to physical-based models. They are simple to operate and rely on historical data to uncover relationships between features and the target variable.

In the field of building energy consumption prediction, the data-driven methods based on machine learning (ML) have advanced considerably in recent years [5]. Many scientists and scholars have started developing and examining ML algorithms that are able to automatically learn from patterns and features in the data and make predictions [6]. Current research and applications of ML-based methods in building energy analysis are mainly in the areas of energy consumption prediction, detecting abnormal energy consumption, and energy efficiency optimization considering uncertainty [7,8]. Several researchers have applied ML-based methods such as multiple linear regression, decision tree, random forest, K-nearest neighbors (KNN), support vector machine (SVM), and artificial neural network (ANN) to predict energy consumption in residential buildings, and they have achieved reasonably good results (e.g., see, [9–14]). ML-based methods, as a data-driven approach, do not require detailed physical information about the buildings, their energy infrastructure, or their surroundings. Meanwhile, a new class of ML-based methods called deep learning (DL) has gained widespread attention and application in the field of building energy consumption prediction [15,16].

Despite all the advantages that traditional ML-based models offer, they are unable to perform the feature extraction procedure successfully, especially when there is a high degree of uncertainty in the data. With the development of interpretable ML-based techniques, the challenge is gradually being overcome [17]. However, the interpretable ML techniques rely more on correlations (rather than forming associations) between features and outcome variables. This may lead to a failure to understand the impact of a particular intervention change on the model output. Causal relationships, which are usually stable over multiple scenarios, are subject to less interference than correlations [18]. Therefore, decisions or judgements based on cause-and-effect are more stable, which is the type of relationship that ML-based methods are expected to learn.

The aims of this paper are two-fold. The first aim is to identify the main factors influencing energy consumption in residential buildings from a cause-and-effect perspective, beyond the limits of correlation. The second aim is to develop an ML model based on the stacking method for accurately predicting the energy consumption in residential buildings. To achieve these aims, we will pursue the following main actions. First, the energy consumption levels of residential buildings are classified into different classes by a data mining algorithm. Next, a technique based on particle swarm optimization (PSO) and a random forest classifier is employed to select the building attributes to be used in this study. Then, a framework is designed to examine the level of building energy consumption, with sensor data, weather data, and building attributes. After that, the stacking model is used to classify and predict residential heating energy consumption. Finally, causal inferences are introduced for effective causal analysis of the data, providing an effective theoretical basis

for the daily operation and maintenance of residual buildings as well as supporting the decisions for the selection of efficient heating management systems.

The rest of this study is structured as follows. Section 2 reviews the literature relevant to the prediction of energy consumption in residential buildings. Section 3 describes the methodology of the research. Section 4 presents the experimental results and discusses the findings. Section 5 provides concluding remarks and suggestions for further research.

2. Literature Review

2.1. The Methods Applied to Prediction of Building Energy Consumption

The research on the prediction of building energy consumption began in the 1970s, when an energy crisis forced countries to start thinking about ways to cut their energy consumption and carbon emissions. The early-developed models of building energy consumption prediction relied on the use of simplified calculation methods that were empirical models based on extensive engineering practices, allowing the estimates to be performed at the early stages of building design to guide the relevant design work. However, it was recognized that simplified calculation methods were not able to adequately capture the dynamicity and complexity of the environment. To tackle this problem, scholars in the mid-1980s started to adopt statistical methods for predicting building energy consumption [19]. Since then, significant progress has been made in the field of building energy consumption prediction. Nowadays, the three most popular methods for predicting energy consumption in buildings include engineering simplification, physical modeling, and ML-based methods [5]. The physical modeling methods enable accurate prediction of energy consumption by adopting a thermodynamic model of the building that helps to simulate the energy consumption process in actual operation. The thermodynamic model is built based on the physical properties of the building itself as well as the principles of heat transfer using some popular software tools such as EnergyPlus, DesignBuilder, and IES [20]. These tools often require many inputs, including building parameters and environmental factors, that might not always be available. Furthermore, they are difficult to replicate as the building parameters and environmental factors may vary from one case to another, and hence, each building project will need to be modeled separately with its own data type, constraints, etc.

With the flourishing development of computer science and information technology in the 1990s, many attempts have been made to employ ML techniques to predict building energy consumption [21]. ML is a class of data-driven statistical models that involve mining historical datasets to identify patterns and trends and then using this information to make informed decisions [22]. In recent years, ML algorithms have played a significant role in conducting predictive analysis, and there are numerous practical examples reporting the successful implementation of ML in a wide range of industry sectors. However, the recording of building parameters, heating methods, wall materials, etc. was not undertaken in a detailed way in the past, resulting in inadequate information to support model training. As energy efficiency in buildings has become more important and energy monitoring technologies have advanced greatly over the past few years, more detailed data about the buildings has become accessible. As a result, ML has been gradually recognized as a competent modeling tool for predicting building energy consumption. Nevertheless, reproducing the predictions of ML algorithms is an important challenge in the current research.

Good training data is the backbone of ML-based building energy consumption prediction models. This training data can be provided from two main sources: the historical building energy consumption datasets and the actual building energy consumption data gathered by smart meter technology [23]. With the rising adoption of energy smart meters and industrial internet of things (IIoT) sensors across the globe, an enormous amount of building energy consumption data has become available [24]. At the same time, however, there are some inevitable problems with the energy consumption data collection practices, such as sensor failures, loss of memory, data transmission network fluctuations, etc., resulting in numerous outliers and missing values [25]. Such outliers and missing values can have a negative impact on the accuracy of predictions if the data is not pre-processed.

Therefore, it is necessary to employ appropriate tools to remediate the outliers and fill in the missing values to enhance the quality of the data.

A review of the relevant literature revealed that most studies have focused on algorithm development rather than data preprocessing techniques [3]. The existing outlier detection methods (such as the box-line graph method) can identify the most prominent data in the original dataset by employing a mathematical approach. Such methods are accessible, but their efficiency and accuracy significantly vary depending on the size of the data. Several studies have proposed data pre-processing approaches for building energy consumption, and they have achieved good results. Mena et al. [26] developed a neural network model for short-term forecasting of electricity usage in an ecological building in the southeast of Spain and achieved an average accuracy of 88.52%. Naji et al. [27] used an adaptive neuro-fuzzy inference system to estimate the building's energy consumption based on several parameters, such as material thickness and insulation K values. In another work, Naji et al. [28] proposed a method based on Extreme Learning Machines (ELM) to predict building energy consumption and then compared their results with those of a genetic algorithm and an ANN technique. Zhao et al. [29] performed an analysis based on ANN, SVM, and autoregressive moving average techniques to predict the energy consumption of variable cooling systems in office buildings. The results showed that the ANN outperformed the other two methods. Li et al. [30] proposed an approach based on stacked autoencoders and ELM to predict building energy consumption. Robinson et al. [31] employed national survey data from the Commercial Buildings Energy Consumption Survey (CBECS) to test the applicability of their building energy consumption model on the New York City Local Law 84 dataset. Ma et al. [32] proposed an SVM method to predict building energy consumption by incorporating weather data (such as mean outdoor temperature, relative humidity, and global solar radiation) as well as economic factors. Li et al. [33] proposed an ANN model based on the methods of characterization decomposition and spatial homogenization decomposition to predict the building's energy consumption at an early stage of design. Fu et al. [34] presented a deep neural network model integrated with a transfer reinforcement learning algorithm to predict building energy consumption based on the data provided by American Power Balti Gas and Electric Power Company. Seyedzadeh et al. [35] proposed a method for optimizing the ML algorithms used to predict the building energy loads. They utilized simulated data generated by EnergyPlus to validate their model and then compared the results with a grid search tuning approach. Jiang et al. [36] proposed a method based on deep neural networks to predict the energy consumption of buildings in Manhattan city. Ngo et al. [37] proposed a hybrid model based on a seasonal autoregressive moving average, the firefly optimization algorithm, and support vector regression (SVR) to predict building energy consumption. They used a dataset of 30 min energy consumption, temporal data, and weather data from six buildings in Vietnam to train and test their model. In a recent study, Olu-Ajayi et al. [38] compared the performance of different ML techniques such as ANN, random forest, KNN, SVM, and decision trees in predicting building energy consumption.

2.2. The Methods Used to Improve the Performance of ML Prediction Models

As mentioned above, the ML algorithms have received a lot of attention for predicting building energy consumption thanks to their remarkable capabilities and satisfactory results. However, in many studies, ML is viewed as a black box that takes relevant data as input (independent values) and produces output (dependent values) without discussing the relationship between dependent and independent variables. To overcome this drawback, researchers have suggested efficient ways to improve the transparency and reproducibility of ML algorithms [39]. A brief description of some of these approaches is given below:

Feature selection: With the advent of the big data era as well as the continuous development of ML models, the pre-processing of features has become the foremost priority in many fields such as pattern recognition, remote sensing, and image retrieval. Feature selection is an efficient pre-processing technique for high-dimensional data analysis, that

aims to remove redundant and irrelevant data and obtain related features. The advantage of feature selection is that it compresses the size of the feature space, reducing the dimensionality of the data and model complexity too. A proper selection of features can enhance the classification performance and thus increase the learning speed of the model. With the development of mathematical feature-based evaluation measures, search technology, statistics, and other multidisciplinary areas, many practical and effective feature selection algorithms have been proposed. Some intelligent evolutionary optimization algorithms adopted for feature selection include ant colony optimization (ACO), particle swarm optimization (PSO), bacterial foraging optimization (BFO), the shuffled frog-leaping algorithm, and the artificial bee colony algorithm (ABC) [40]. The common feature of these algorithms is that they are inspired by natural phenomena in the universe, such as the collective behavior of ant colonies or bees. PSO is one of the most popular intelligence optimization algorithms for feature selection in high-dimensional datasets. This algorithm is inspired by the motion of bird flocks and fish swarms. It provides a basis for the movement of the whole group in the problem-solving space by using the individuals in the population to share their current information so as to produce the evolution process from disorder to order and then obtain an optimal set of solutions. When the PSO algorithm is used, random particles are generated and iterated continuously until the optimal solution is found.

Feature dimensionality reduction: The recent dramatic growth in the size of data logs and attributes has led to the emergence of big data processing platforms and parallel data analysis algorithms. At the same time, it has driven the application of data dimensionality reduction techniques. These techniques are often of two types: linear methods and non-linear methods. Principal component analysis (PCA) and linear discriminant analysis (LDA) are two of the most popular linear methods [41]. PCA finds lines, planes, and hyperplanes in the multi-dimensional space that approximate the data in the least square sense. LDA is a supervised classification technique which takes into the account the labeled data while carrying out dimensionality reduction. This method finds the boundaries around clusters of classes and projects the data on a line so that clusters are separated, with each class having a relative distance to a centroid. Even though both PCA and LDA are great tools to reduce the dimensionality of the features, they are not suitable for cases where the data contain strong non-linear dependencies. To tackle this problem, Kohonen [42] proposed a new statistical method called self-organizing mapping (SOM) to interpret multidimensional, non-linear, and noisy data. SOM is an unsupervised learning method that transforms a complex, high-dimensional input space into a simpler, low-dimensional (typically two-dimensional) discrete output space. It is a brain-inspired model that is used for classification and clustering purposes. It usually generates a dimensional grid or map by learning the data in the input space, and hence, in some ways, it is an inherently dimensionality reduction algorithm.

Causal relationship analysis: Causal relationship analysis techniques are useful to study the intervention-to-outcome effects. The correlations are viewed as undirected relationships (where the features and outcomes affect each other, adjusting one side and the other will follow) and causation as directed relationships (where features determine outcomes, and the changes in features make outcomes change, and there is no inverse relationship). The two most widely used causal relationship models are the Potential Outcome Model (POM), proposed by Rubin [43], and the Structural Causal Model (SCM), proposed by Pearl [44]. The POM is based on the idea that causality is linked to the treatment (or action, manipulation, or intervention) applied to a unit. Treatment effects are obtained by comparing the potential treatment outcomes of units. SCM describes the causal mechanisms of a system in which a set of variables and the causal relationships between them are modeled by a set of simultaneous structural equations. When modeling causal inferences, prior knowledge is used to make hypotheses about the relationship between variables as input to the model. For this purpose, the DoWhy framework can be used to encode prior knowledge into a causal graph. DoWhy is a Python-based library combining causal

graphical models and potential outcomes for causal inference and analysis (https://github.com/py-why/dowhy, accessed on 19 March 2023).

In the next section, our proposed methodology for predicting building energy consumption is presented.

3. The Methodology of Research

The methodology of this study includes several stages, including dataset selection, data exploration, data preprocessing, feature engineering, model construction, model evaluation, and feature interpretation. These stages are illustrated in Figure 1 and described in detail below.
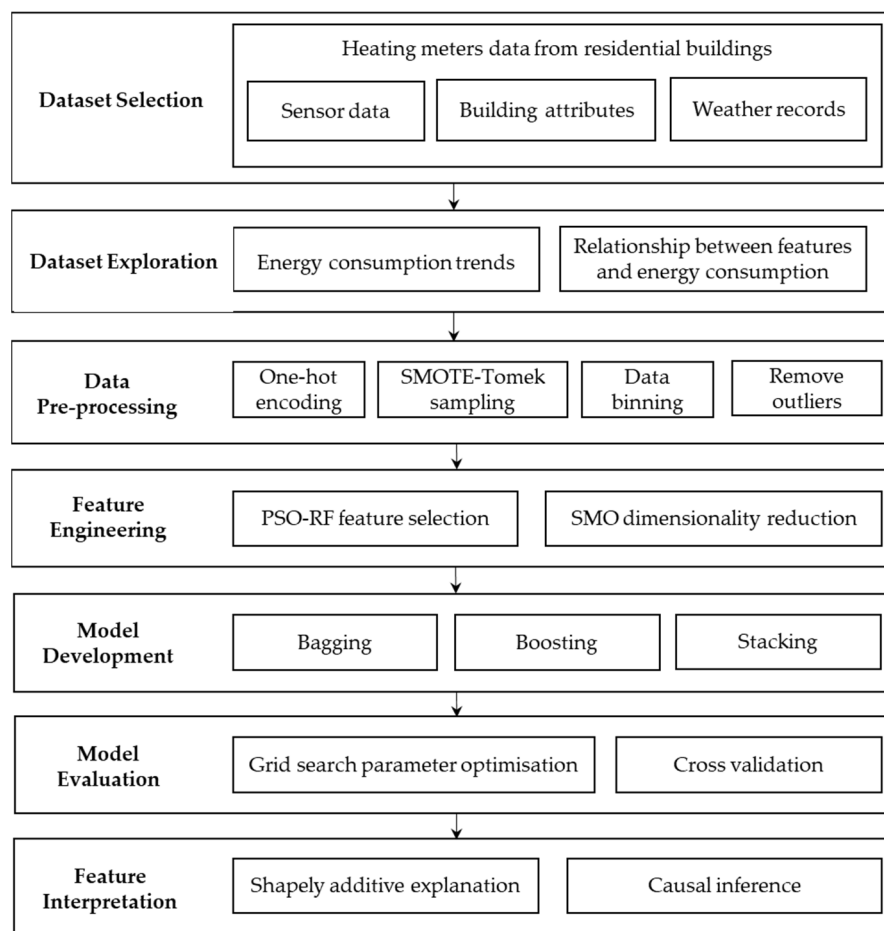
Figure 1. The Methodology of Research.

3.1. Database Selection

There are several public databases and internet resources with energy consumption information from different buildings in different locations with different characteristics that can be used to describe and test the model. In this study, a dataset from smart meters in residential buildings in Tomsk city in Russia was collected. This dataset was published for the first time in 2020 by IEEE DataPort [45], and, since then, it has been used in some research studies, e.g., [46]. The dataset contained sensor data, building attribute features, and weather records during five heating seasons. The data was also supplemented by floor number, wall material, and year of construction of buildings, as well as data on average daily outdoor temperatures. The thermal energy consumed by buildings was measured at an indoor room temperature of 24 °C. The description of the features in the dataset is presented in Table 1.

Table 1. Description of the features in heating meters dataset.

Dimension	Name	Description	Type
Sensor	M1	Mass of the input water (heat carrier) per day	Ratio
	M2	Mass of the output water	Ratio
	delta M	Difference between input and output in water volumes	Ratio
	T1	Average temperature of the heating carrier in the input	Interval
	T2	Average temperature of the heating carrier in the output	Interval
	delta T	Temperature difference	Interval
	Q	Amount of the consumed heating	Ratio
Attribute	USPD	ID of the heating meter	Nominal
	Registered	What is registered, heating or heating plus hot water	Nominal
	YYYYMM	Date in the format year-month YYYYMM	Date
	Scheme	Type of the heating system (opened or closed)	Nominal
	Type	Code system-load (4 digits)	Nominal
	Area	Area of building that heating meter is served	Ratio
	Floors	Amount of building floors	Ordinal
	Walls	Walls material	Nominal
	Year of construction	Year of building construction	Date
	Area of building	Total area of building	Ratio
Weather	Temp	Outdoor temperature	Interval

3.2. Database Exploration
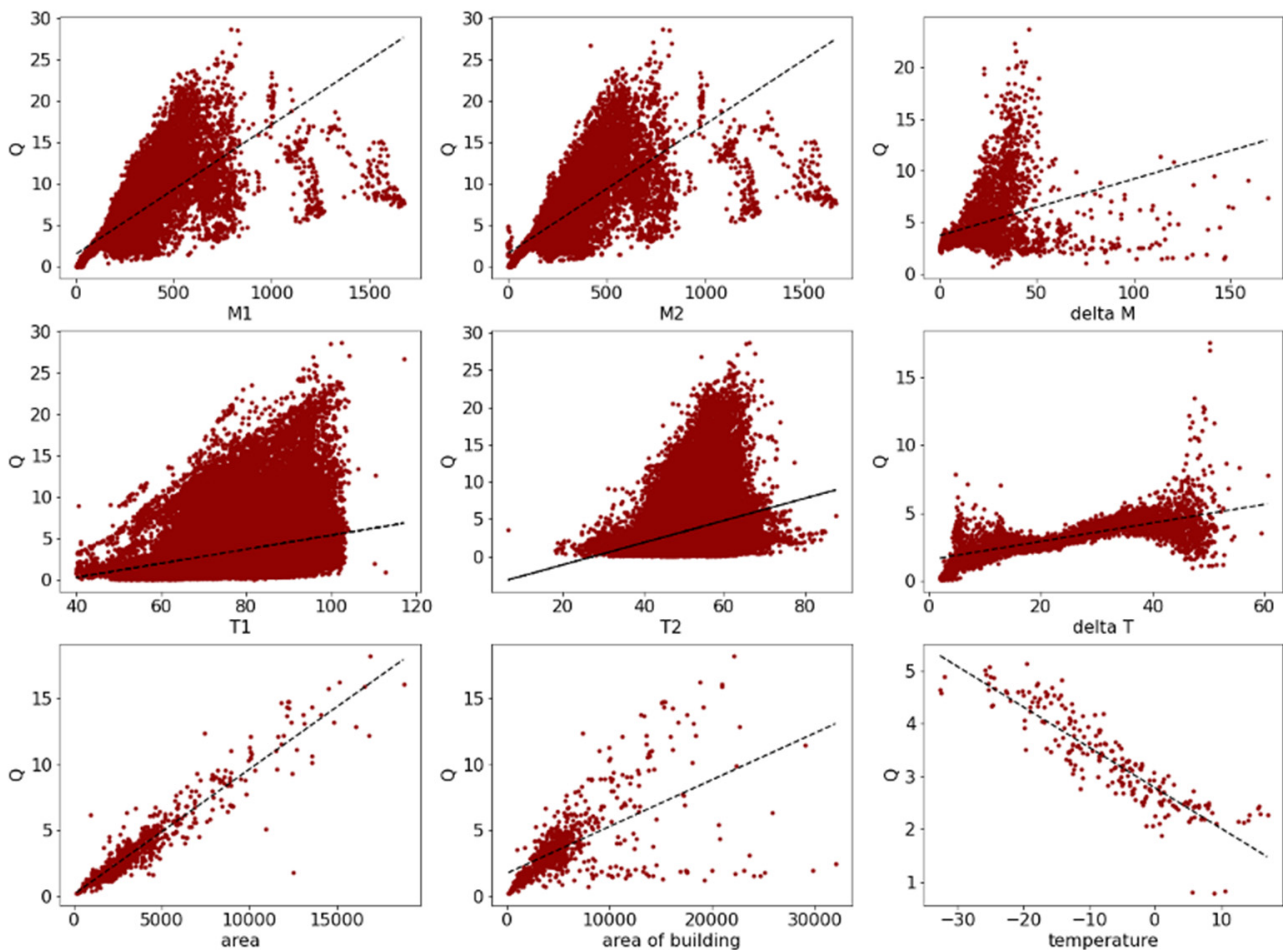
3.2.1. Energy Consumption Trends

As the dataset contains a large amount of daily recorded data, an exploration was first carried out by taking an average of the building energy consumption over one year for each day of the week. As shown in Figure 2a, except in July and August, when heating was not required, no significant trend in the daily energy consumption was observed. Therefore, a graph of the monthly energy consumption was plotted, and it is shown in Figure 2b. In the monthly energy consumption graph, there is a clear trend between different months. The thermal energy consumption in buildings gradually increases from September and reaches a peak level of around 3.8 kilogram-calories (kGcal) in November. It fluctuates from November to December and then hits a peak level of 4.5 kGcal in January. After January, the energy consumption level decreases steadily to 1.6 kGcal in May.

(a)   (b)
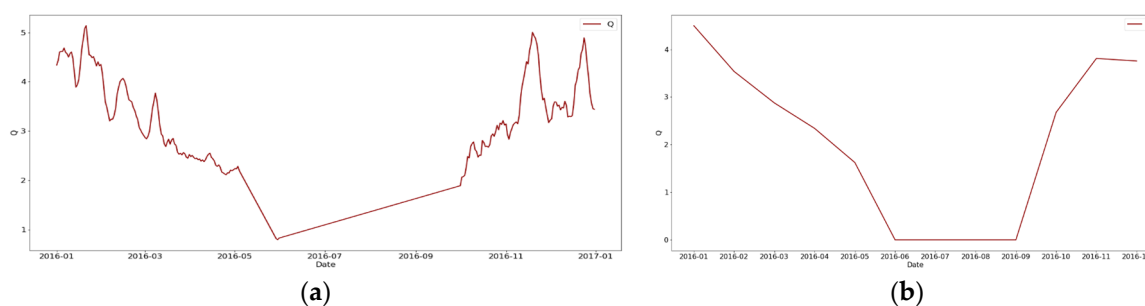
Figure 2. Trends of building energy consumption on a (**a**) daily basis, and (**b**) monthly basis.

3.2.2. Exploring the Relationship between Features and Energy Consumption Levels

As presented in Table 1, the variables are categorized into three types, namely, nominal, interval, and ratio. The nominal variables include USPD (ID of the heating meter), registered (what is registered, heating or heating plus hot water), scheme (type of the heating system, opened or closed), type (code system-load), and wall materials. The interval variables include T1, T2, delta T, and temperature. M1, M2, and delta M are ratio variables, as the volume of water for heating has an absolute zero point. Subsequently, the relationship between each of the interval and ratio variables and energy consumption Q will be explored. The energy consumption data is first aggregated and averaged over each

feature to be represented by a scatter plot, and then the Polyfit function is applied by fitting the points to a trend line using the least squares method. The trends of building energy consumption with respect to each feature variable are depicted in Figure 3. As shown, all variables are positively correlated with energy consumption, except weather temperature, which is negatively correlated with energy consumption. The three variables of delta T, area, and temperature are more evenly distributed on both sides of the trend line than other variables. Also, the actual area served by the heating system is more clearly correlated with energy consumption than the area of the building.

Figure 3. The distribution of building energy consumption with respect to nine feature variables.

To quantify the correlations between variables, Pearson's correlation coefficient was adopted. The Pearson's correlation coefficient is the most common way of measuring the degree of linear correlation between two variables. This coefficient is between −1 and 1, with larger absolute values indicating a stronger correlation. Figure 4 represents a heat map of the Pearson's correlation coefficient for features and the level of energy consumption. As can be seen, the service area of the heating system with a Pearson's coefficient of 0.84 has the highest correlation with the level of energy consumption, followed by the mass of the input water and output water per day with Pearson's coefficients of, respectively, 0.8 and 0.79. The temperature variables T1, T2, and outdoor temperature were weakly correlated to the building energy consumption with Pearson's coefficients of 0.37, 0.35, and −0.31, respectively.
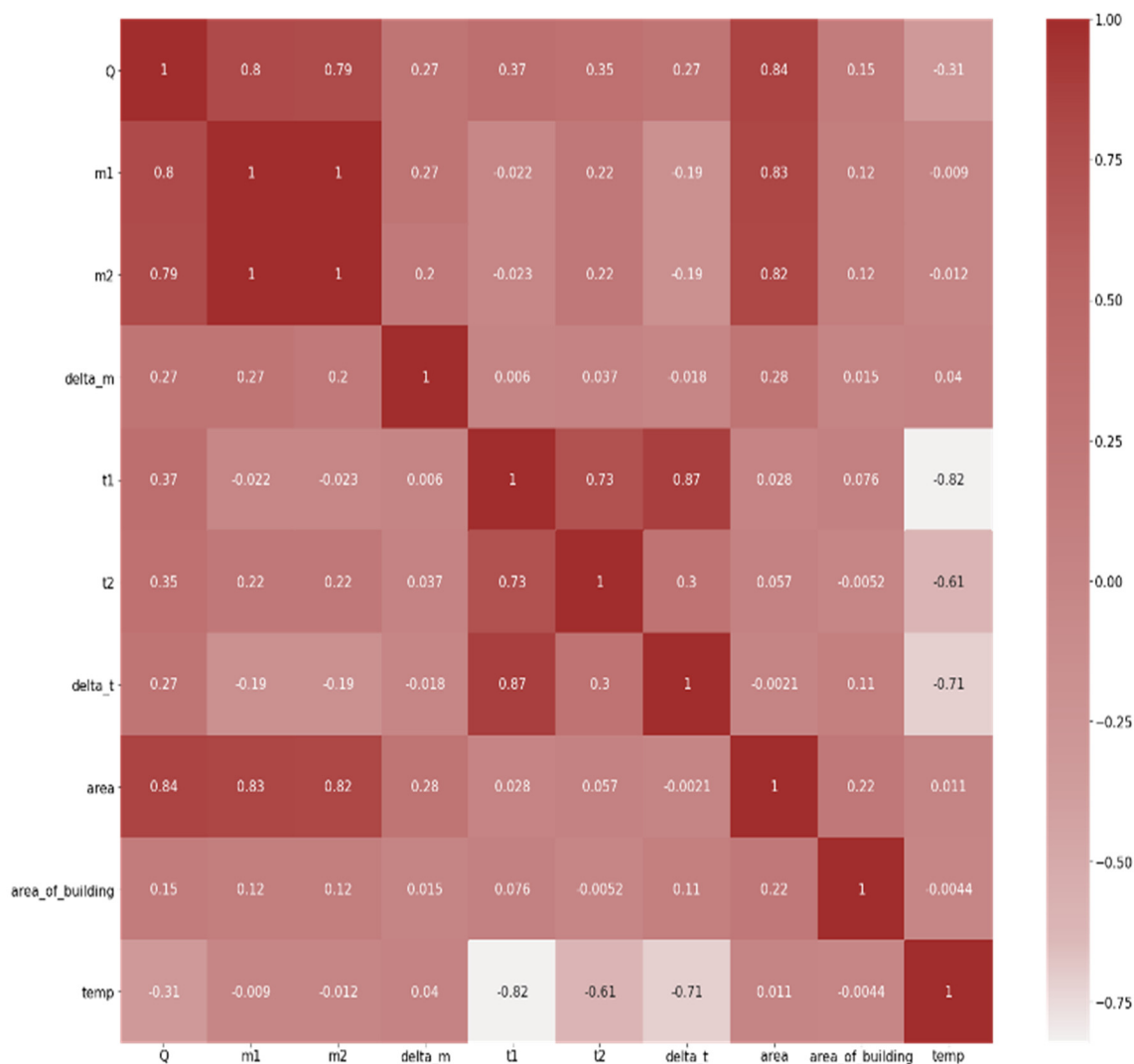
Figure 4. A heat map of Pearson's correlation coefficients for building energy consumption features.

The main reason for the difference between the effects of two features, outdoor temperature and water temperature, on energy consumption levels can be derived from the Pearson's coefficient. The scatter plot is based on the average of energy consumption levels in the univariate characteristic dimension, which excludes the influence of other variables; however, the Pearson's coefficient is calculated directly from the original data after excluding outliers, which may be influenced by other variables. There are also some problems with exploring feature correlations in univariate dimensions. For example, Figure 5 shows the correlations between building age, heating service area, and energy consumption. As can be seen, the energy consumption level of the building decreases with the age of the building, but in fact, the energy consumption ratio of the building decreases over time with the development of new building technology and smart materials. The reason for such a phenomenon is that as the age of the building increases, the area served by the heating system decreases, whereas the heat map of the Pearson's correlation coefficient shows that the area served by the heating system is strongly correlated to the energy consumption level, and in this case, it is difficult to distinguish exactly which feature influences the energy consumption level of the building.
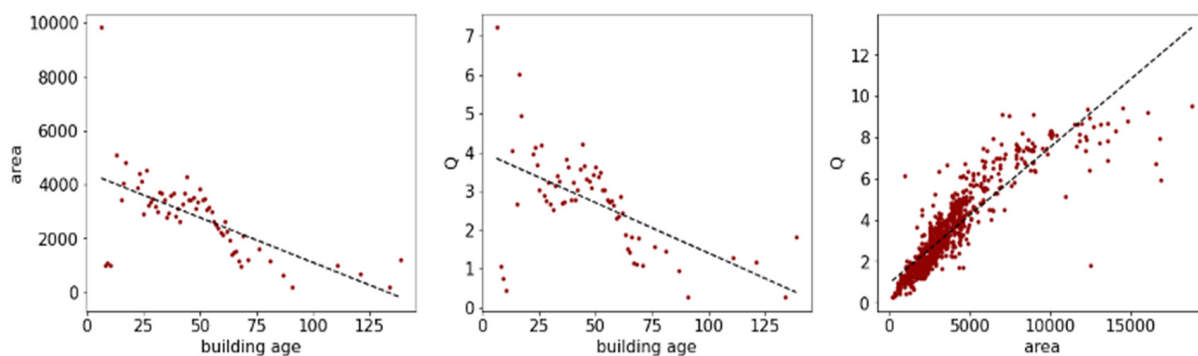
Figure 5. The correlations between building age, heating service area, and energy consumption.

3.3. Feature Engineering

3.3.1. PSO-RF Feature Selection Algorithm

A technique based on particle swarm optimization (PSO) and a random forest classifier is employed for feature selection, combining the advantages of both filter and wrapper methods. The PSO algorithm is applied to select the features, and the random forest classifier technique is used to evaluate the performance of feature subsets. The PSO is a population-based algorithm that consists of a group of particles. Each particle has two properties: velocity and position. The velocity parameter indicates the speed of movement, whereas the position parameter indicates the direction. The particles form a swarm, and the population moves progressively throughout the domain. During each move, PSO recalculates the objective function for each particle. For each initial subset of features, the performance is evaluated using a random forest classifier with a fitness function defined as follows: "Update each particle $x_{(t+1)}$, while restricting each dimension of the updated particle to be a binary vector between [0, 1]. The features with a value of 0 are selected and those with a value of 1 are not, thus obtaining an initial subset of features for each particle".

Random forest is a bagging technique and not a boosting technique, and, hence, the trees in random forest are run in parallel rather than sequentially. The final classification result is determined by the voting method, where each tree votes on the output with equal weight. The generalization error depends on the classification effectiveness of the single trees in the forest and the degree of correlation between the classification trees. We combine bagging and randomization to improve the performance of the combined classifier by reducing the correlation between the classification trees while ensuring the effectiveness of the single trees. The steps of our proposed algorithm are as follows:

1. Generate an initial particle population (with size *N*), convert it into a binary vector, and obtain an initial subset of features, including random position and velocity.
2. The fitness value of each particle is evaluated based on the fitness function. The initial position of each particle is regarded as the individual extreme value *pBest*, and the global extreme value *gBest* is the particle with the minimum fitness value.
3. Update the velocity and position of the particle. From the updated binary particle vector $x_{B(t+1)}$, features with a value of 0 are selected and features with a value of 1 are not selected to obtain a new subset of features, and the fitness value of all particles is calculated according to the fitness function.
4. Update the *pBest* and *gBest* values. If the updated binary particle vector $x_{B(t+1)}$ is all 1 and no features are selected, then $x_{B(t+1)}$ is updated to a set of randomly generated binary vectors.
5. Terminate the loop when the number of loops reaches *epoch*, producing the global optimal solution and the optimal subset of features; otherwise, repeat step 3.

3.3.2. Self-Organizing Feature Mapping

The SOM consists of two layers of neurons: an input layer that is connected to each vector of the dataset and a competition layer that forms a two-dimensional array of nodes.

In the SOM, not only the winning neuron has to adjust its weight vector, but the neurons around it also have to adjust their weight vectors to varying degrees under its influence. The winning neighborhood is the area within the radius of the neighborhood set at the center of the winning neuron. In the SOM, the weights of all neurons in the winning neighborhood are adjusted according to their distance from the winning neuron. Figure 6 shows a flowchart for implementing the SOM method.
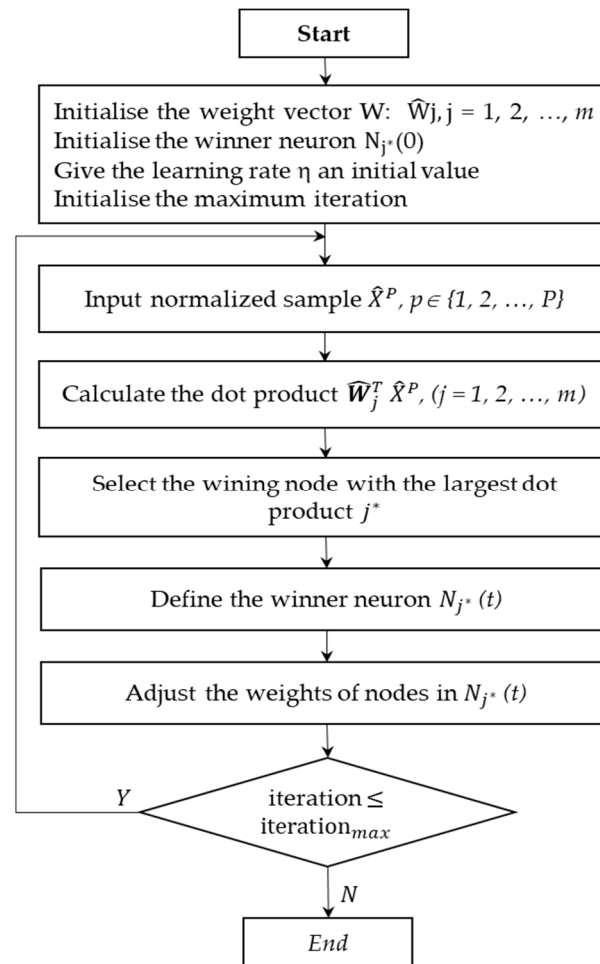
Figure 6. A flowchart for implementing the self-organizing feature mapping.

SOM can be applied to reduce the dimensionality of the data, allowing the projection of high-dimensional data onto a lower subspace while keeping the topology constant. A component plane is adopted to visualize the results of dimensionality reduction. After dimensionality reduction, the neural network produces a 7×7 matrix of weights, and the value of each square indicates what value the neuron at each position is most sensitive to (or is understood to be the best match to that feature value). The results of the SOM dimensionality reduction are represented in Figure 7. As can be seen, the correlation between features and energy consumption remains the same after the dimensionality reduction; the distribution of the weight blocks of M1, M2, heating service area, and energy consumption Q are approximately the same; the weight blocks of heating service area and building age are in a complementary color relationship. This indicates that while reducing the dimensionality of the data, the SOM has retained the distribution of the data as well as the correlation between features.
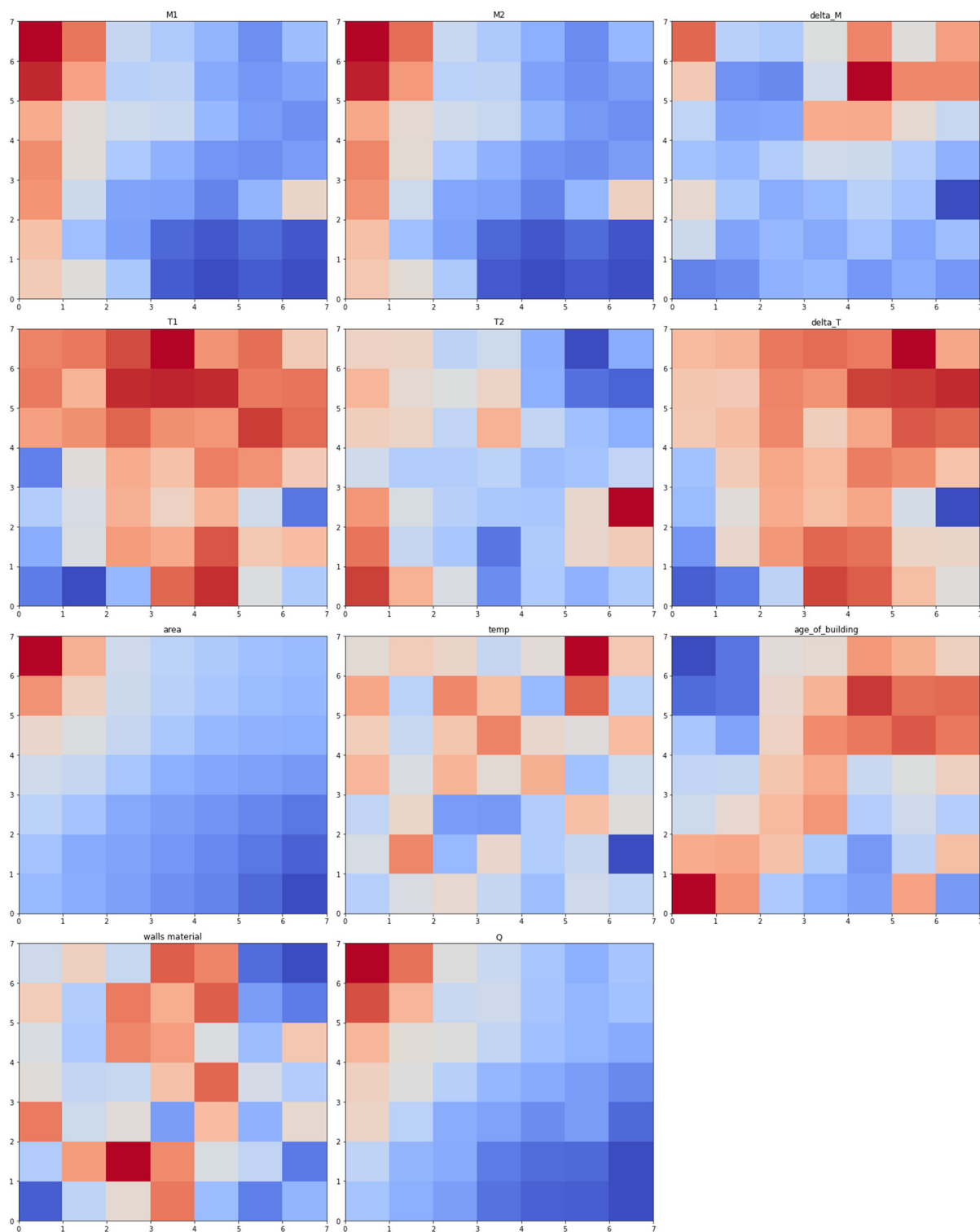
Figure 7. Visualization of SOM dimensionality reduction results.

3.4. Model Development

In this study, the stacking strategy is used to construct the model. According to this strategy, a two-layer structure is implemented to aggregate multiple models to obtain a combined model that outperforms every single model. In the first layer, learners are built to cross-validate the training data, with each learner separately predicting the divided training data and generating new features for the training of the next learner. The test set is substituted throughout the trained model to produce predictions, and new labels are

generated via voting to be used as target values for the next learner. The ensemble learning is applied to construct and combine a set of individual learners to complete a learning task. The ensemble learning is broadly divided into three categories: bagging, boosting, and stacking. These categories are briefly explained below:

3.4.1. Bagging

There is no strong dependency between the base learners in bagging, and therefore numerous weak learners could be generated in parallel at the same time. Each iteration randomly samples data from the original dataset to train the learner, and for each prediction result, the final result is selected by voting. The random forest algorithm is used for training, in which multiple decision trees are trained simultaneously. The predictions are made considering multiple outcomes, and the plurality of multiple nodes is considered as the basis for the final decision. This eliminates the disadvantage that decision trees are prone to overfitting and reduces the variance of predictions (i.e., the predictions will not change dramatically due to small changes in the training data). Randomness is reflected, and a subset of the original dataset is randomly brought back to the bootstrap and used as the training dataset for a particular decision tree in the forest. Each selection of features for the bifurcation is limited to finding a feature in a subset of the randomly selected features.

Random forests are a class of bagging algorithms where the base learner is a decision tree. The base learner of the bagging algorithm is assumed to be a decision tree, and that requires the construction of a decision tree on each dataset that has a put-back generation, while the base decision tree constructed by random forest will seek the optimal feature for node splitting from a random set of features. Therefore, the random forest algorithm not only satisfies the bagging algorithm's property of putting back perturbations on the sample set but also satisfies the randomness of the feature set when building the node feature splits of the tree. In addition, each tree in the forest takes the same splitting rule for tree construction, and the termination condition is satisfied when the training samples on the nodes belong to the same class or the maximum depth of the tree is reached. Therefore, the random forest is more efficient and generalizes better than the general bagging model.

3.4.2. Boosting

XGBoost (eXtreme Gradient Boosting) is used to train the model. It is a type of gradient boosting tree (GBDT) which consists of an additive model and a forward stepwise algorithm [47]. In XGBoost, there are strong dependencies between base learners, and the base classifiers must be generated serially. To prevent overfitting of the model, XGBoost includes regularization in the cost function. The regularization contains the fit error function and a penalty term for the complexity of the decision tree, including the number of leaf nodes and the squared term for the fraction of leaf nodes. The regularization limits the complexity of the tree and reduces the variance of the model, making the XGBoost model superior to GBDT. Also, XGBoost builds all subtrees from the top to the bottom and then prunes from the bottom to the top, allowing XGBoost to be less likely to fall into a local optimum than LightGBM.

LightGBM is based on the Gradient-Based One-Side Sampling (GOSS) algorithm and the Exclusive Feature Bundling (EFB) algorithm [48]. GOSS enhances the training speed of the model and saves time by adjusting the sampling rate of data with different gradients without changing the data distribution or losing learner accuracy. EFB detects each feature, assigns it to the bundle with the least conflicting value, and outputs the set of feature bundles. Finally, the histogram algorithm merges the mutually exclusive features. Although the histogram algorithm is coarse and loses some accuracy, the loss in accuracy of the base learner could be compensated by introducing more trees in the gradient boosting technique. In addition, LightGBM applies a leaf-wise growth strategy with depth limits to find the leaf with the maximum splitting gain from all the leaves, and increases the maximum depth limit to ensure high efficiency while preventing over-fitting. LighGBM

supports parallelized learning of large-scale data, achieving faster training and higher accuracy with lower memory usage.

3.4.3. Stacking

Stacking combines the predictions of several base learners as a new training set for a new learner. The base layer of the stacking model includes different learning algorithms; therefore, the stacking ensemble is heterogeneous. The stacking model employed in this study is shown in Figure 8.
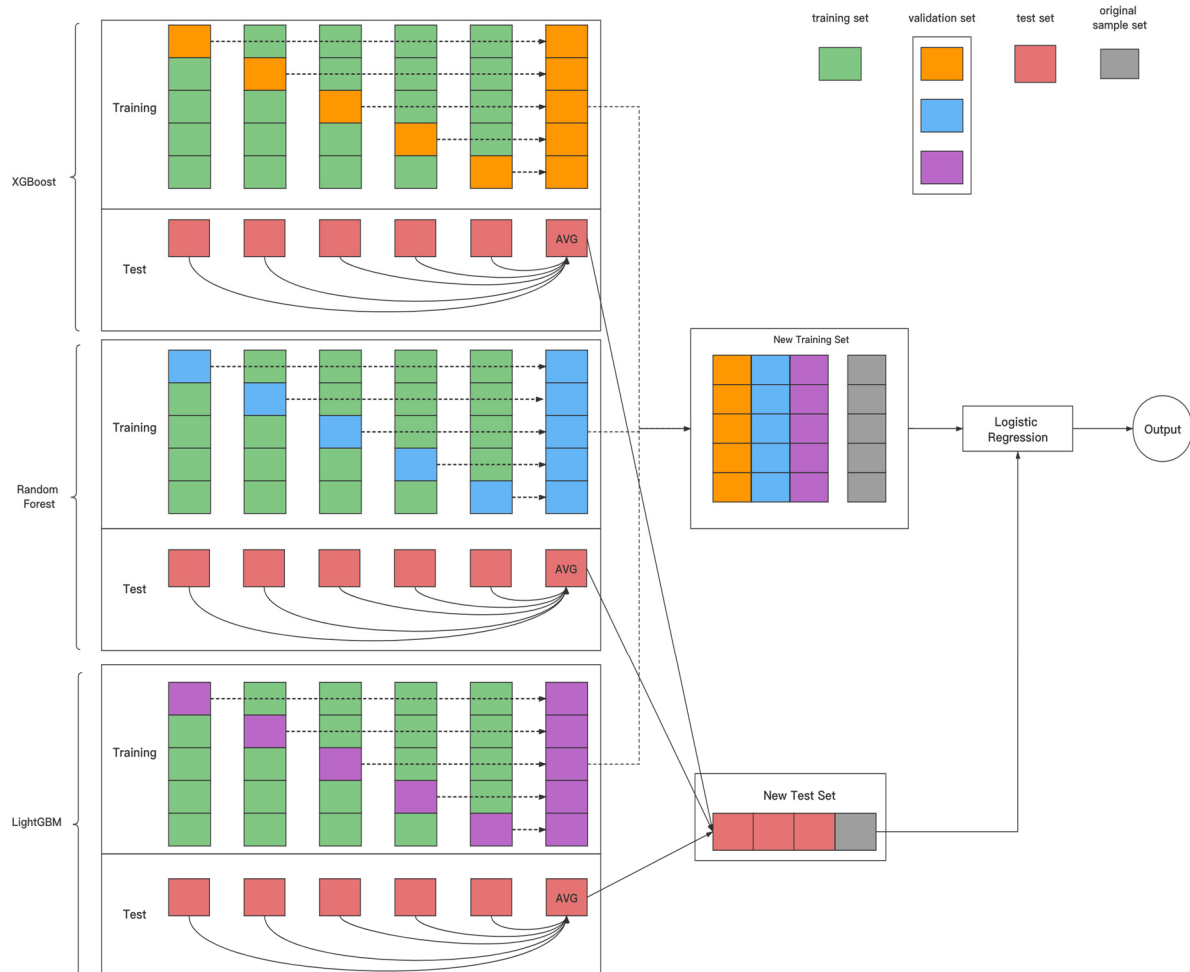
Figure 8. The stacking model employed in this study.

XGBoost, LightGBM, and random forest were adopted as the first layers of predictors. The training dataset was first trained so that the models could make predictions of labels for both the training and test sets. The predictions of multiple models for both the training set and the test set were then imported into the second-layer of models as new features along with the true labels of the training set for training. Finally, the second-layer model estimated the predictions from the previous multiple test sets to obtain the results.

3.5. Model Evaluation

The metrics considered for the evaluation of the proposed model include: (*i*) accuracy, which refers to the percentage of predictions that are correctly predicted; (*ii*) precision, which refers to the degree of reproducibility or repeatability of the results; and (*iii*) recall, which refers to the number of true positive predictions divided by the number of true positive predictions plus false negative predictions. In the confusion matrix, if an instance is a positive class and is predicted to be positive, it will be the true positive (TP) class; if an

instance is a positive class but is predicted to be negative, it will be the false negative (FN) class; if an instance is a negative class but is predicted to be positive, it will be the false positive (FP) class; and if an instance is a negative class and is predicted to be a negative class, it will be the true negative (TN) class. Figure 9 represents the calculation of three performance metrics (accuracy, precision, and recall) in the confusion matrix.
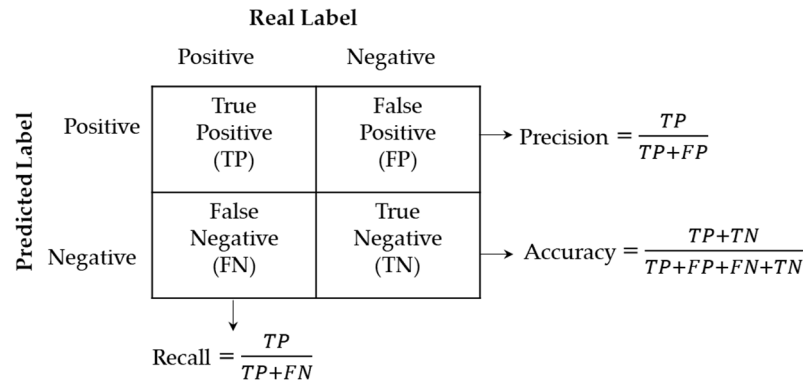

Figure 9. The performance evaluation metrics in the confusion matrix.

3.6. Feature Interpretation

Understanding the causal relationships between features in the energy consumption of heating systems is of great practical importance for both system design and day-to-day management, as causal inference can be useful to explain the importance of features. The causal effect is defined as the extent to which the outcome changes when a unit change in the intervention occurs. Therefore, a causal graph model was created to understand the effect of each feature on energy consumption using a priori knowledge. Figure 10 shows the casual graph developed for building energy consumption prediction.
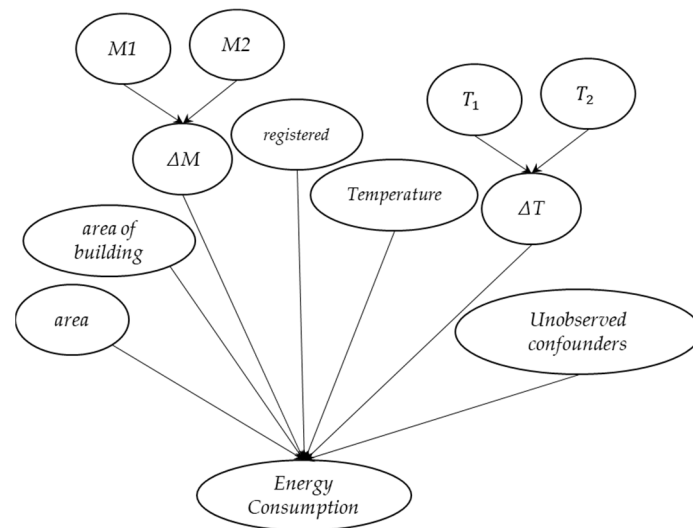

Figure 10. Casual relationships diagram for building energy consumption prediction.

Based on the constructed causal graph, we used the following strategies for causal effect identification: back-door criterion, front-door criterion, instrumental variables, and mediation analysis (direct and indirect effect identification). The average treatment effect (ATE) was chosen as the causal effect estimate of the outcome. Finally, the correctness of the estimates was verified by the refutation method, random common cause, data subset refuter, and placebo treatment. The causal hypothesis is tested by adding random confounders, and if the hypothesis is correct, the causal effect should not change much with the addition of random confounders. The causal effect is tested by replacing the intervention with a random variable via a placebo intervention, and if the hypothesis is correct, the causal effect

should be close to 0. The causal effect is tested by estimating the causal effect on a subset of the data, and if the hypothesis is correct, the causal effect should not change much.

4. Results and Discussion

The experiments were implemented using the VScode compiler and in the Anaconda (python 3.8) environment on a macOS 11.4 system with an Intel Core i5 2 GHz processor and 16 GB RAM. The results of the experiments are presented and discussed below.

4.1. Results of Feature Engineering

The distribution of actual building energy consumption Q is depicted in Figure 11.
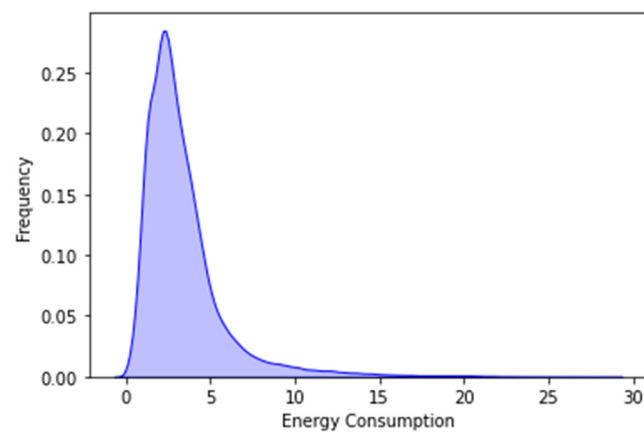
Figure 11. The distribution of actual building energy consumption.

Data binning, also called discrete binning or bucketing, is a data pre-processing technique used to reduce the effects of minor observation errors. This is a method of grouping multiple continuous values into a smaller number of 'bins'. The easy addition and reduction of discrete features to the boxed data allows for rapid iteration of the method. At the same time, the discrete features make the model more stable and reduce the risk of overfitting the model. In this study, the energy consumption Q was classified into six bins (classes), and the results are shown in Figure 12a.

| (a) | (b) |
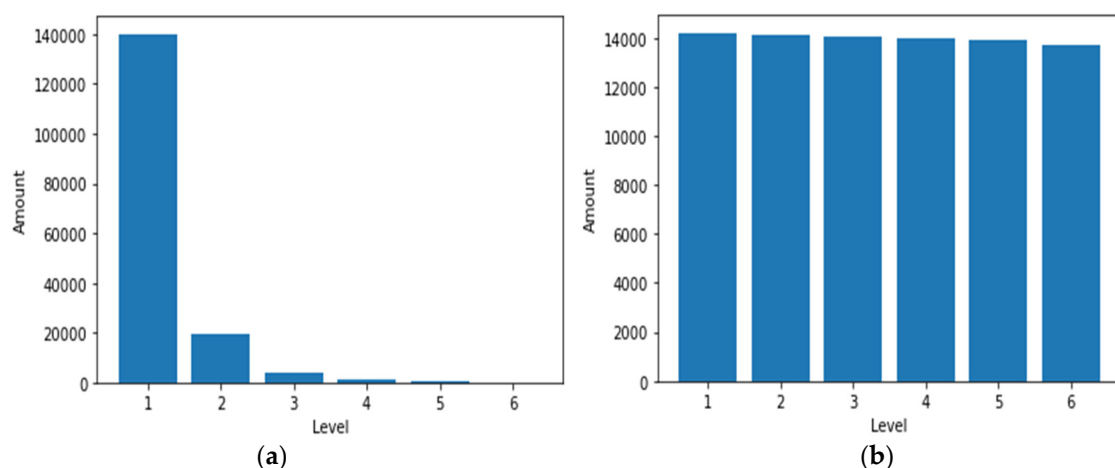
Figure 12. The building energy consumption (**a**) after data binning (**b**) using SMOTE-Tomek method.

As can be seen, the energy consumption of bin #1 is significantly higher than that of other bins. As other bins have a sparse number of samples, this created a very chaotic feature space, which ultimately diminishes the effectiveness of the model in learning the features. Therefore, we employed a Synthetic Minority Over-Sampling Technique (SMOTE) to oversample the minority class samples and eliminate the Tomek Link pairs in

the expanded dataset. Figure 12b shows the results after the SMOTE-Tomek algorithm is used. As the energy consumption data is highly dimensional, we used feature bagging for outlier detection. Feature bagging detectors can be fitted with multiple base detectors on individual subsamples of the dataset, using averaging or other combinations of methods to improve prediction accuracy. In this research, the Local Outlier Factor (LOF) was used as a feature bagging algorithm for outlier detection, and eventually 1.58% of the outliers in the sample were removed. The accuracy of random forest feature selection improved from 0.898 to 0.928 after 10 epochs of the PSO algorithm. The features included in the training process were M1, M2, delta M, T1, T2, delta T, area, temperature, age of the building, and wall material. After selecting the features, the SOM algorithm was employed to reduce the dimensionality of the filtered data from 10 to 7 dimensions and thereby reduce the training time for model fitting and comparison. The results of the training for PSO-RF feature selection are given in Table 2.

Table 2. The accuracy of random forest feature selection with different epochs of PSO.

Epoch	1	2	3	4	5	6	7	8	9	10
Accuracy	0.898	0.918	0.922	0.925	0.926	0.926	0.926	0.927	0.927	0.928

4.2. Results of ML Models

KNeighborsNearest, NeuralNet, LightGBM, Random Forest, CatBoost, XGBoost, and Stacking Model were trained and tested on the building energy consumption dataset. 80% of the data was used for training and the remaining 20% for testing. The experiments were repeated 10 times for every model, and the results were compared in terms of average accuracy and training time. The results of comparisons between the ML models are summarized in Table 3. As can be seen, the neural network model was less accurate than the other models, and it took longer to train the model. Bagging and boosting methods both performed better. However, due to the use of Grid Search CV for parameter tuning and cross-validation, the training rate of the model was significantly higher than that of the other models and was 0.059 higher than that of the random forest model, which was the most accurate of the single methods. The treatment of category imbalance by SMOTE-Tomek contributes significantly to the performance of the stacking model, and the training time has been significantly reduced by the SOM dimensionality reduction. However, the training time of the model was as long as three hours, which was significantly higher than the other models.

Table 3. The accuracy and run-time of different ML models.

	KNeighborsUnif	KNeighborsDist	NeuralNet FastAI	LightGBM	SMOTE-Tomek + SOM + Stacking
Accuracy	0.864	0.888	0.697	0.828	0.954
Run-time (s)	6.55	0.12	165.59	14.28	10,963
	Random Forest	**CatBoost**	**XGBoost**	**Stacking**	**SMOTE-Tomek + Stacking**
Accuracy	0.895	0.748	0.8568	0.904	0.942
Run-time (s)	20.99	7.69	62.56	12,486	13,076

4.3. Results of Feature Interpretation

The Shapley Additive Explanation (SHAP) algorithm is applied to interpret the output of ML models. A SHAP value is assigned to each feature in the predicted sample while expressing the positive and negative impacts. The average of the absolute SHAP values for each feature was taken as the importance of the feature and sorted in descending order to obtain a statistical plot of feature importance, as shown in Figure 13. The SHAP analysis indicates that the feature M1 is the most important factor contributing to residential heating energy consumption, followed by delta T, area, and T1.
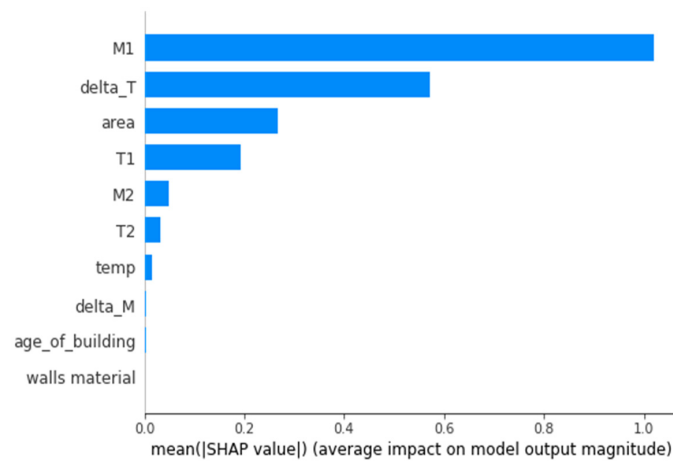
Figure 13. The results of SHAP analysis.

Figure 14 illustrates the prediction results for four features of area: T1, M1, and delta T. The predictions of the model were derived from the base value and the model output. The features that push the predictions higher (positive impact) are shown with red arrows, and the features that push the predictions lower (negative impact) are shown with blue arrows. For the records with lower energy consumption (c,d), the main negative factors were essentially the same, all being M1, delta T, area, and T1, consistent with the performance of feature importance. However, for the records with high energy consumption (a,b), the main positive and negative factors differ, with (a) having a positive factor of area and M1 and a negative factor of delta T, while for (b), T1 and delta T are the main positive factors and M1 and area are the negative factors.
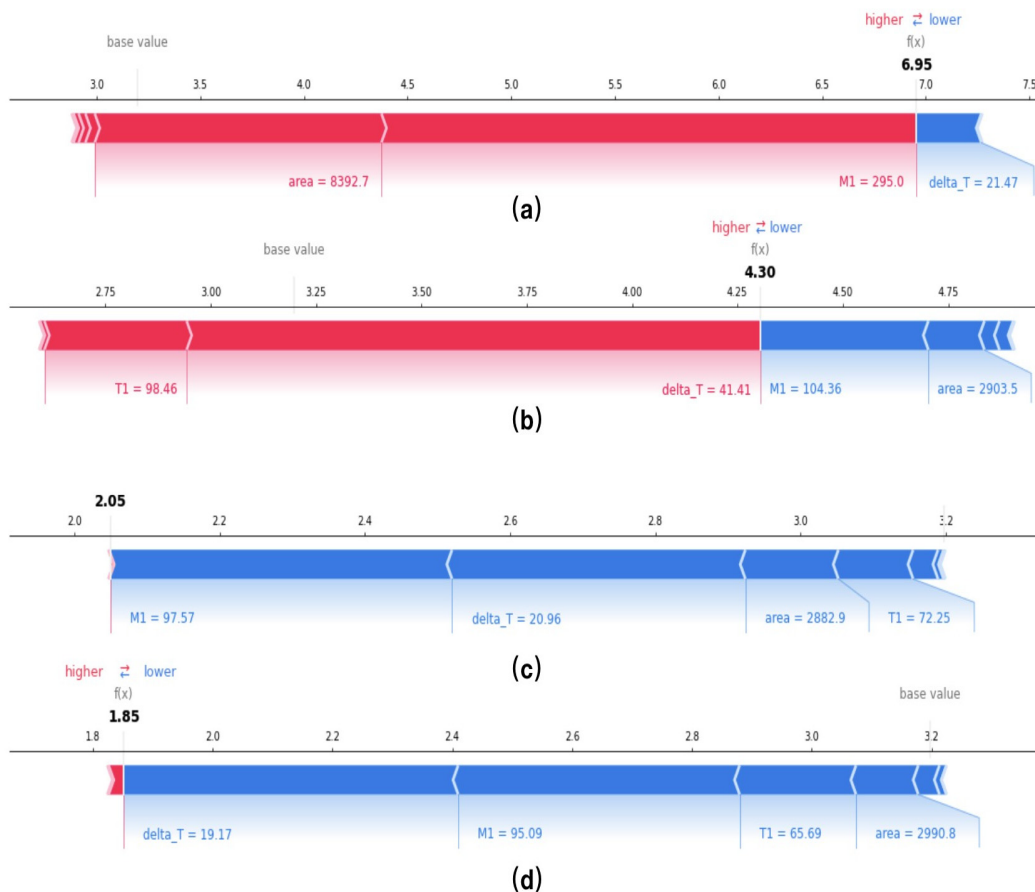

Figure 14. SHAP analysis results for four features of (**a**) area, (**b**) T1, (**c**) M1 and (**d**) delta T.

Figure 15 shows the SHAP values for all the features contributing to building energy consumption, which provides a more intuitive understanding of the analysis as a whole. The figure shows that for the features of M1, delta T, area, and T1, the SHAP values increase with increasing feature values and decrease with increasing T2 and temperature. At the same time, although M1 is a more important factor than delta T, delta T has a higher negative influence on energy consumption than M1.
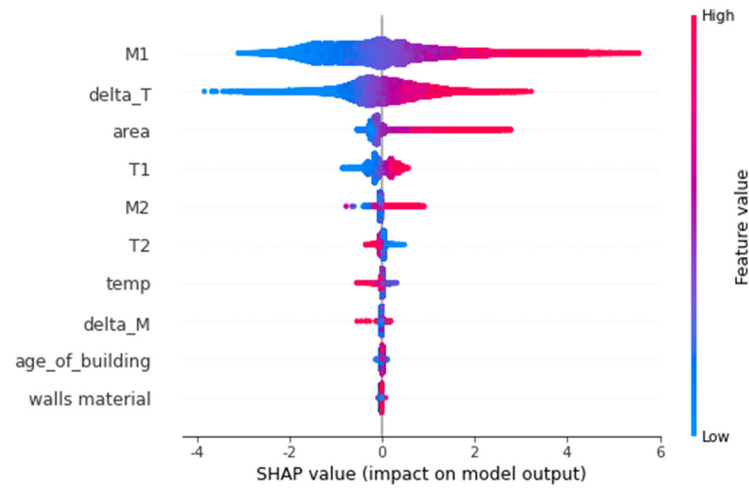

Figure 15. SHAP results for all the features.

The dependence plot is designed to demonstrate the way in which a single feature affects the model output. This is developed by comparing the SHAP value of each feature with all other features in the dataset. Figure 16a shows the dependence plot for delta T and the interaction with temperature. As can be seen, the SHAP values of the features showed a significant upward trend with increasing delta T. At the same time, the higher the temperature, the smaller the change in water heating temperature. Figure 16b demonstrates the dependence plot for area and the interaction with M1. It is found out that as the area served by water heating increases, the SHAP value appears to be on the rise, but the SHAP values vary from one building to another due to confounding factors. At the same time, a clear break between area and M1 exists at area 5000 sqm, with most of M1 in the samples below area 5000 sqm being below 150 m^3 and most of M1 above area 5000 sqm being above 150 m^3. Figure 16c reveals the variation of SHAP values with temperature and the interaction between temperature and M1. The SHAP values show an overall decreasing trend with increasing temperature, with positive SHAP values at −3 degrees Celsius and negative SHAP values above −3 degrees Celsius. Interestingly, when M1 is a small value, the SHAP value is 0, illustrating the absolute influence of M1 on the final results.
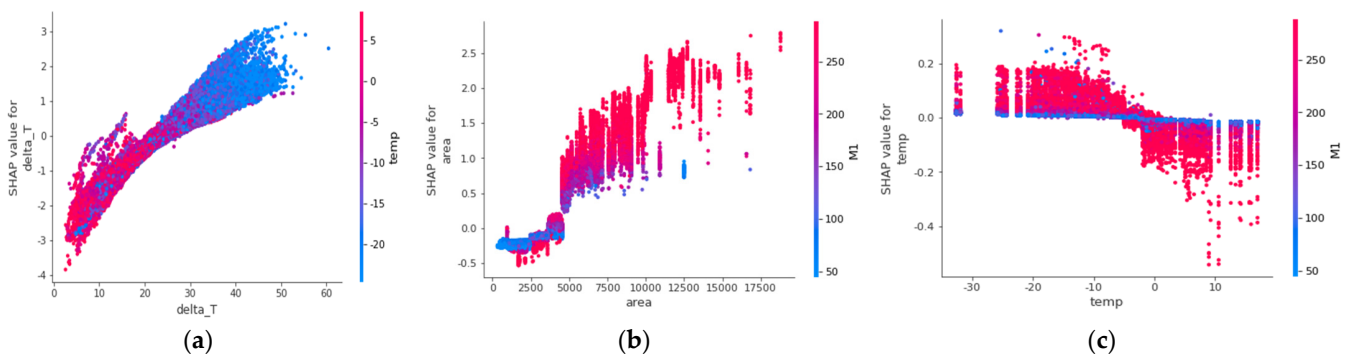

Figure 16. SHAP for (**a**) delta T and temperature, (**b**) area and M1, and (**c**) temperature and M1.

The results of the causal relationship analysis are presented in Table 4. The experimental results show that the estimates obtained from the random common cause refuter and the

data subset refuter are both similar to the mean value when the treatment is changed while other characteristics are kept constant; however, the estimates obtained from the placebo treatment refuter all tend to be close to zero. This effectively proves the correctness of the causal graph model, indicating that the experiments have obtained accurate estimates. From the results of the causal inference of the features, it is observed that the main factors influencing the final energy consumption of the building include the variation in the water temperature of the heating system and the weather temperature of the day, followed by the volume of water used for heating and the heating service area.

Table 4. Results of the causal analysis.

		T1	Delta M	M1	M2	Delta T	Temp	Area	Area of Building
Estimate	Mean Value	0.0157	0.0388	0.0813	0.0813	0.186	−0.177	0.0813	0.0921
Refute	Random Common Cause	0.0157	0.0386	0.0813	0.0813	0.228	−0.0837	0.0813	0.101
	Data Subset Refuter	0.0149	0.038	0.0812	0.0812	0.216	−0.18	0.0813	0.095
	Placebo treatment refuter	6.32×10^{-6}	1.38×10^{-5}	-2.49×10^{-5}	2.78×10^{-5}	9.08×10^{-5}	2.19×10^{-5}	-2.15×10^{-5}	3.83×10^{-5}

From the experimental results, it was evident that the PSO-optimized random forest feature selection filtered out most irrelevant features and greatly reduced the computational complexity of feature selection. The PSO algorithm continuously updates the particles, making them diverse and avoiding the problem of local optima. The stacking framework integrates three algorithms, LightGBM, random forest, and XGBoost, to make full use of different algorithms from different data space and data structure perspectives on different observations of the data. The results obtained by the stacking method are more robust and accurate since the Grid Search CV parameters were adjusted several times. However, the stacking framework cannot be parallelized, so it takes significantly longer than other models to be trained. In practice, the balance between model accuracy and efficiency has to be considered in most cases, as models with higher accuracy usually require a longer run time. It should be ensured that both the model accuracy and runtime are within an acceptable range, rather than just pursuing accuracy. An accuracy of 89.5% was achieved using only the random forest algorithm. Without feature engineering, the stacking method only improves accuracy by about 1% compared with the random forest method, but at the expense of a longer runtime.

The SHAP analysis of the features was very informative. For example, by analyzing a single sample, it was found that the importance of the impact of features is more stable in buildings with lower energy consumption, while the importance of the impact of features varies more between buildings with higher energy consumption. The deviations that emerged in the experimental results between causal inference analysis and feature interpretation also illustrate the limitations of current ML methods in general. ML models could predict what the energy consumption level of a building will be at a certain date and time in the future, but stakeholders and relevant authorities ultimately wish to understand what interventions or policies can be adopted in order to reduce the energy consumption level of a building for the purpose of saving energy. For example, given that changes in outdoor temperature and heating water temperature have a significant impact on building energy consumption, the materials of water/heating pipes can be upgraded to a more thermally efficient material in order to reduce the loss of heat energy. Therefore, it is essential to incorporate causal relationships between features in ML models to provide more accurate and realistic predictions for building energy consumption.

5. Conclusions

This study proposed advanced machine learning (ML) algorithms to predict heating energy consumption in residential buildings under different indoor and outdoor conditions. The main factors influencing the building's energy consumption were identified from a cause-and-effect perspective. Features were selected using a technique based on particle swarm optimization (PSO) and a random forest classifier, and the data outliers were

detected by a local outlier factor (LOF) estimator-based feature bagging method. Finally, the residential heating energy consumption levels were classified by employing the stacking algorithm and using three ML models, namely, LightGBM, Random Forest, and XGBoost. Comparing the performance of the proposed model with traditional methods, the results demonstrated the superiority of the stacking algorithm in terms of residential heating consumption classification. Causal inference was also adopted with the explanation of factors influencing residential building energy consumption. The causal inference analysis revealed the significant impact of changes in weather temperature and water heating temperature on building energy consumption. These findings can help residential building owners/managers make more informed decisions on the selection of efficient heating management systems at a macro level, especially during peak heating periods in the winter.

This study established a framework to study the level of residential heating energy consumption based on smart meter data, weather data, and building attributes such as heating methods, wall materials, etc. The smart building and smart city sectors are currently working to promote the use of sensor networks to optimize energy use. The computerization of buildings' heating systems will enable the construction of accurate district plumbing configurations and more balanced energy supply and demand for maximum efficiency in heat management. This can increase energy savings while ensuring the heating requirements of residents in buildings.

Although the proposed methods possess various advantages, they also have certain limitations. Firstly, there is no comparative experimentation and analysis of feature selection and dimensionality reduction algorithms in building energy consumption prediction. Therefore, a subsequent work could provide a comparative analysis to justify the methods. Secondly, the dataset did not contain geographical information. New data can be collected to supplement the dataset, and further research can be completed to produce more stable results. Finally, corrections between some other features, such as the heating service area and the age of buildings, could be further explored by clustering analyses of different types of buildings in the future.

Author Contributions: Conceptualization, F.D., Y.H. and M.S.; methodology, F.D. and Y.H.; software, Y.H.; validation, F.D., Y.H. and M.S.; formal analysis, F.D., Y.H. and M.S.; investigation, F.D. and Y.H.; resources, F.D.; data curation, Y.H.; writing—original draft preparation, F.D., Y.H. and M.S.; writing—review and editing, F.D. and M.S.; visualization, F.D. and Y.H.; supervision, F.D.; project administration, Y.H.; funding acquisition, F.D. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The raw data are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Energy Information Administration. U.S. Energy Consumption by Source and Sector. 2021. Available online: https://www.eia.gov/energyexplained/us-energy-facts/images/consumption-by-source-and-sector.pdf (accessed on 15 February 2023).
2. Krarti, M.; Aldubyan, M. Review analysis of COVID-19 impact on electricity demand for residential buildings. *Renew. Sustain. Energy Rev.* **2021**, *143*, 110888. [CrossRef] [PubMed]
3. Amasyali, K.; El-Gohary, N.M. A review of data-driven building energy consumption prediction studies. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1192–1205. [CrossRef]
4. 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]
5. Zhao, H.-X.; Magoulès, F. A review on the prediction of building energy consumption. *Renew. Sustain. Energy Rev.* **2012**, *16*, 3586–3592. [CrossRef]
6. Alizamir, M.; Kim, S.; Kisi, O.; Zounemat-Kermani, M. A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: Case studies of the USA and Turkey regions. *Energy* **2020**, *197*, 117239. [CrossRef]
7. Fahim, M.; Fraz, K.; Sillitti, A. TSI: Time series to imaging based model for detecting anomalous energy consumption in smart buildings. *Inf. Sci.* **2020**, *523*, 1–13. [CrossRef]

8. Yu, J.; Chang, W.-S.; Dong, Y. Building energy prediction models and related uncertainties: A review. *Buildings* **2022**, *12*, 1284. [CrossRef]

9. Yu, Z.; Haghighat, F.; Fung, B.C.; Yoshino, H. A decision tree method for building energy demand modeling. *Energy Build.* **2010**, *42*, 1637–1646. [CrossRef]

10. Chae, Y.T.; Horesh, R.; Hwang, Y.; Lee, Y.M. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy Build.* **2016**, *111*, 184–194. [CrossRef]

11. Runge, J.; Zmeureanu, R. Forecasting energy use in buildings using artificial neural networks: A review. *Energies* **2019**, *12*, 3254. [CrossRef]

12. Seyedzadeh, S.; Pour Rahimian, F.; Glesk, I.; Roper, M. Machine learning for estimation of building energy consumption and performance: A review. *Vis. Eng.* **2018**, *6*, 1–20. [CrossRef]

13. Zhong, H.; Wang, J.; Jia, H.; Mu, Y.; Lv, S. Vector field-based support vector regression for building energy consumption prediction. *Appl. Energy* **2019**, *242*, 403–414. [CrossRef]

14. Chakraborty, D.; Elzarka, H. Advanced machine learning techniques for building performance simulation: A comparative analysis. *J. Build. Perform. Simul.* **2019**, *12*, 193–207. [CrossRef]

15. Ardabili, S.; Abdolalizadeh, L.; Mako, C.; Torok, B.; Mosavi, A. Systematic review of deep learning and machine learning for building energy. *Front. Energy Res.* **2022**, *10*, 786027. [CrossRef]

16. Wang, Y. Application of deep learning model in building energy consumption prediction. *Comput. Intell. Neurosci.* **2022**, *9*, 4835259. [CrossRef] [PubMed]

17. Wu, C.; Gales, M.J.; Ragni, A.; Karanasou, P.; Sim, K.C. Improving interpretability and regularization in deep learning. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *26*, 256–265. [CrossRef]

18. Alvarez-Melis, D.; Jaakkola, T.S. A causal framework for explaining the predictions of black-box sequence-to-sequence models. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 7–11 September 2017; pp. 412–421.

19. Sullivan, R.; Nozari, S.; Johnson, R.; Selkowitz, S. Commercial building energy performance analysis using multiple regression. *ASHRAE Trans.* **1985**, *91*, 337–353.

20. Li, Z.; Lin, B.; Zheng, S.; Liu, Y.; Wang, Z.; Dai, J. A review of operational energy consumption calculation method for urban buildings. *Build. Simul.* **2020**, *13*, 739–751. [CrossRef]

21. Kissock, J. Modeling commercial building energy use with artificial neural networks. In Proceedings of the Intersociety Energy Conversion Engineering Conference, Monterey, CA, USA, 7–12 August 1994.

22. Seyedzadeh, S.; Pour Rahimian, F.; Rastogi, P.; Glesk, I. Tuning machine learning models for prediction of building energy loads. *Sustain. Cities Soc.* **2019**, *47*, 101484. [CrossRef]

23. 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]

24. Nutkiewicz, A.; Yang, Z.; Jain, R.K. Data-driven Urban Energy Simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow. *Appl. Energy* **2018**, *225*, 1176–1189. [CrossRef]

25. Zhu, C.; Tian, W.; Yin, B.; Li, Z.; Shi, J. Uncertainty calibration of building energy models by combining approximate Bayesian computation and machine learning algorithms. *Appl. Energy* **2020**, *268*, 115025. [CrossRef]

26. Mena, R.; Rodríguez, F.F.; Castilla, M.; Arahal, M.R. A prediction model based on neural networks for the energy consumption of a bioclimatic building. *Energy Build.* **2014**, *82*, 142–155. [CrossRef]

27. Naji, S.; Shamshirband, S.; Basser, H.; Keivani, A.; Alengaram, U.J.; Jumaat, M.Z.; Petkovic, D. Application of adaptive neuro-fuzzy methodology for estimating building energy consumption. *Renew. Sustain. Energy Rev.* **2016**, *53*, 1520–1528. [CrossRef]

28. Naji, S.; Keivani, A.; Shamshirband, S.; Alengaram, U.J.; Jumaat, M.Z.; Mansor, Z.; Lee, M. Estimating building energy consumption using extreme learning machine method. *Energy* **2016**, *97*, 506–516. [CrossRef]

29. Zhao, D.; Zhong, M.; Zhang, X.; Su, X. Energy consumption predicting model of VRV (Variable refrigerant volume) system in office buildings based on data mining. *Energy* **2016**, *102*, 660–668. [CrossRef]

30. Li, C.; Ding, Z.; Zhao, D.; Yi, J.; Zhang, G. Building energy consumption prediction: An extreme deep learning approach. *Energies* **2017**, *10*, 1525. [CrossRef]

31. Robinson, C.; Dilkina, B.; Hubbs, J.; Zhang, W.; Guhathakurta, S.; Brown, M.A.; Pendyala, R.M. Machine learning approaches for estimating commercial building energy consumption. *Appl. Energy* **2017**, *208*, 889–904. [CrossRef]

32. Ma, Z.; Ye, C.; Li, H.; Ma, W. Applying support vector machines to predict building energy consumption in China. *Energy Procedia* **2018**, *152*, 780–786. [CrossRef]

33. Li, Z.; Dai, J.; Chen, H.; Lin, B. An ANN-based fast building energy consumption prediction method for complex architectural form at the early design stage. *Build. Simul.* **2019**, *12*, 665–681. [CrossRef]

34. Fu, Q.; Liu, Q.; Gao, Z.; Wu, H.; Fu, B.; Chen, J. A Building energy consumption prediction method based on integration of a deep neural network and transfer reinforcement learning. *Int. J. Pattern Recognit. Artif. Intell.* **2020**, *34*, 2052005. [CrossRef]

35. Seyedzadeh, S.; Pour Rahimian, F.; Oliver, S.; Glesk, I.; Kumar, B. Data driven model improved by multi-objective optimisation for prediction of building energy loads. *Autom. Constr.* **2020**, *116*, 103188. [CrossRef]

36. Jiang, F.; Ma, J.; Li, Z.; Ding, Y. Prediction of energy use intensity of urban buildings using the semi-supervised deep learning model. *Energy* **2022**, *249*, 123631. [CrossRef]

37. Ngo, N.-T.; Pham, A.-D.; Truong, T.T.H.; Truong, N.-S.; Huynh, N.-T. Developing a hybrid time-series artificial intelligence model to forecast energy use in buildings. *Sci. Rep.* **2022**, *12*, 15775. [CrossRef]

38. Olu-Ajayi, R.; Alaka, H.; Sulaimon, I.; Sunmola, F.; Ajayi, S. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *J. Build. Eng.* **2022**, *45*, 103406. [CrossRef]

39. Miller, C.; Picchetti, B.; Fu, C.; Pantelic, J. Limitations of machine learning for building energy prediction: ASHRAE Great Energy Predictor III Kaggle competition error analysis. *Sci. Technol. Built Environ.* **2022**, *28*, 610–627. [CrossRef]

40. Li, K.; Xue, W.; Tan, G.; Denzer, A.S. A state of the art review on the prediction of building energy consumption using data-driven technique and evolutionary algorithms. *Build. Serv. Eng. Res. Technol.* **2020**, *41*, 108–127. [CrossRef]

41. Abdi, H.; Williams, L.J. Principal component analysis. *WIREs Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]

42. Kohonen, T. *Self-Organizing Maps*, 3rd ed.; Springer Series in Information Sciences; Springer: Berlin/Heidelberg, Germany, 2001.

43. Rubin, D.B. Causal inference using potential outcomes: Design, modeling, decisions. *J. Am. Stat. Assoc.* **2005**, *100*, 322–331. [CrossRef]

44. Pearl, J. Causal diagrams for empirical research. *Biometrika* **1995**, *82*, 669–688. [CrossRef]

45. Zorin, P.; Stukach, O. Data of Heating Meters from Residential Buildings in Tomsk (Russia) for Statistical Modeling of Thermal Characteristics of Buildings. Published on 5 October 2020. Available online: https://ieee-dataport.org/documents/data-heating-meters-residential-buildings-tomsk-russia-statistical-modeling-thermal (accessed on 10 June 2021).

46. Kuznetsov, M.; Kaldybaeva, B.; Tsibulskiy, S.; Boldyryev, S. Optimisation of heat distribution in Tomsk City by inter-cluster connections model implemented in P-graph environment. In Proceedings of the 6th International Conference on Smart and Sustainable Technologies, Bol and Split, Croatia, 8–11 September 2021.

47. 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.

48. 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. In Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.