Civil Engineering and Architecture 11(6): 3538-3551, 2023 http://www.hrpub.org
DOI: 10.13189/cea.2023.110623

Predicting and Optimizing the Energy Efficiency of Sustainable Residential Buildings Using Machine Learning Method

Dareen Qashmar*, Dema Khraisat, Hadeel Tariq Obeidat

Department of Architecture, Faculty of Engineering, Al-Balqa Applied University, Jordan

Received November 24, 2022; Revised July 10, 2023; Accepted August 15, 2023

Cite This Paper in the Following Citation Styles

(a): *[1] Dareen Qashmar, Dema Khraisat, Hadeel Tariq Obeidat , "Predicting and Optimizing the Energy Efficiency of Sustainable Residential Buildings Using Machine Learning Method," Civil Engineering and Architecture, Vol. 11, No. 6, pp. 3538 - 3551, 2023. DOI: 10.13189/cea.2023.110623.*

(b): *Dareen Qashmar, Dema Khraisat, Hadeel Tariq Obeidat (2023). Predicting and Optimizing the Energy Efficiency of Sustainable Residential Buildings Using Machine Learning Method. Civil Engineering and Architecture, 11(6), 3538 - 3551. DOI: 10.13189/cea.2023.110623.*

Abstract The increase in overall building energy consumption can be attributed to both socioeconomic advancement and increased urbanization. Forecasting energy consumption in buildings is critical for improving energy efficiency and sustainable development, ultimately resulting in lower energy prices and a negative environmental impact. Sustainable solutions in residential buildings aim to improve thermal comfort while lowering energy consumption. The difficulties and issues associated with residential structures may be resolved by using consumer behavior models and incorporating their inference into residential problem solutions. This article employs machine learning models that have been developed, tested, and trained to simulate energy usage in the building. Several appliances' energy data are used to evaluate the proposed predictive optimization technique. The proposed technique's results are compared to modules for prediction and optimization. To evaluate performance, regression performance measures are used. Furthermore, the results of the calculations demonstrated that the trained Random Forest Regressor model proposed in this work can accurately forecast the building's energy usage. Finally, the proposed model can be used to predict and optimize energy use in buildings that are similar to one another.

Keywords Machine-learning (ML), Optimization, Energy Efficiency, Residential Buildings, Prediction

1. Introduction

Energy conservation is both a global issue and one of the most prominent research topics in the current decade [1]. Population growth, deterioration of the interior environment, and the impact of climate change have all contributed to a significant increase in demand for environmentally friendly technologies and methods of reducing energy use in residential buildings [2-5].

Increased energy demands are a natural byproduct of the world's accelerating economic growth [6]. Because of the rapid growth of both urban residential and commercial areas, more buildings are consuming more energy. Only accurate energy consumption forecasts can now provide an accurate estimate of energy consumption [7]. Buildings that use less energy may benefit both the environment and the bottom line. One of the many advantages of commercial energy efficiency upgrades is lower utility bills [8]. In recent years, several countries' efforts to enact building-specific energy rules and regulations have accelerated. Furthermore, in recent years, a plethora of widely used software applications dedicated to the design of green buildings has emerged. As a result, developing an accurate optimization predictive model is critical for lowering building energy expenditures and increasing their positive environmental impact [9].

The number of countries enforcing energy efficiency regulations on all types of buildings is rapidly increasing. These laws require new constructions to meet minimum energy efficiency standards to reduce costs and greenhouse gas emissions [10].

According to research, improved energy efficiency at the point of use could significantly reduce global energy demand [7,11]. Even minor changes in peak demand may result in significant cost savings for customers and utilities [12]. As a result, increasing building energy efficiency is required to mitigate negative environmental effects while also keeping energy costs under control.

Using the most effective operational procedures for a building's energy management systems can result in the most efficient use of its energy resources. A manager's top priority is to keep an eye on the energy time series and the elements that affect the building's energy performance. Energy consumption must be constantly monitored and managed, and prediction is an important part of that. Precision energy usage forecasting can help you build a more energy-efficient building [13].

Predicting a building's energy consumption is difficult due to the numerous variables that influence it, including weather, location, building construction, occupancy, and so on. Several experts have worked on the problem of predicting future energy consumption over the last two decades. Physical methods [14, 15] and statistical methods (e.g., Auto-Regressive Integrated Moving Average; ARIMA) are the two most common approaches to estimating a building's future energy use [16]. Numerous studies use ML and AI methods, such as ANNs [14, 17, 18], SVMs [19,2,20], and Decision Trees [21]. The artificial neural network (ANN) and its advancements were the most widely used approach for predicting building energy consumption, employing a variety of methods such as input variable selection, network hyperparameter tuning, and training algorithm refinement. We use the ANN technique based on variable selection to assess and select all potentially significant input variables, as in [22].

The purpose of this research is to create a predictive machine learning model by combining the Decision Tree Regressor, the Random Forest Regressor, the Multi-Layer Perceptron Regressor, and the Gradient Boosting Regressor. The goal of this research is to create, train, and apply a set of machine learning models that can estimate the amount of cooling and heating required for an existing building, and then use these models to demonstrate a forecasting and diagnosis method for residential building energy consumption that takes into account the impact of these and other factors. We compared the projected energy consumption value with the observed energy consumption value to further investigate irrational usage occurrences in building operations. This study solved the problem of precise energy prediction at the level of residential buildings by using machine learning models to maximize the goal function. By conducting a longitudinal analysis of the factors that influence energy efficiency in buildings subject to the same climate, this study fills a significant knowledge gap which regarding the study question (How can machine learning algorithms be leveraged to predict and optimize the energy efficiency of sustainable residential buildings, specifically for heating and cooling purposes?).

2. Methods and Materials

This section provides an overview of the statistical concepts and machine learning algorithms that will be used to analyze the data.

2.1. Data-set Description

In this study, there are a total of ten components: eight variables and two findings. This information is generated for a total of 12 different types of buildings, each of which is comprised of 18 different components in its own right. The simulation is focused on a cluster of three apartment buildings in the city of Athens, Greece, which is assumed to have a population of seven. The combined volume of all of the structures is 771.75 m^3, yet their heights and footprints are quite different from one another. It makes no difference what kind of building you're putting together as long as all 18 parts are made out of the same materials. These structures each make use of one of three distinct glazing areas, with each one expressed as a proportion of total floor space: 10%, 25%, or 40%. In addition to this, five different distribution scenarios were modeled for each glazing area. These are some of the options: Five different glazing percentages are possible: I: uniform, with 25% glazing on all sides; ii: north, with 55% glazing on the north side and 15% on the other sides; iii: east, with 55% on the east side and 15% on the other sides; iv: south, with 55% on the south side and 15% on the other sides; and v: west, with 55% on the west side and 15% on the other sides [5]. I: uniform, with 25% glazing on all sides; II: north, with 55% glazing on the north side. In addition to that, several of the samples do not have any glazing on them. Last but not least, we checked to see that each figure was oriented in one of the cardinal directions. It has also been determined how much energy is necessary to heat and cool each of the 768 buildings [5]. Either heat must be added (referred to as the "heating load," or HL) or heat must be removed (referred to as the "cooling load," or CL) from a room in order to maintain it within a temperature range that is considered to be comfortable. These two quantities are referred to as the heating load and the cooling load, respectively.

2.2. Machine Learning Algorithms

2.2.1. Linear Regression

Techniques of predictive modeling, such as linear regression, are utilized rather frequently. This is represented by the equation $Y = a + bX + e$, in which a and b stand for the intercept and slope, respectively, and e stands for the error, respectively. This equation can be used to make predictions about the value of the dependent variable, given information about the predictor variables [23,24].

2.2.2. Robust Regression

In the same way that the least squares method is applicable everywhere, the use of robust regression is not

limited to any particular context. In the least squares regression, specific data points may stand out as outliers or have high leverage. We determined that there was no human error in the data entry and that these samples do not come from a different demographic than the rest of our data [25]. This was accomplished by comparing these samples' demographics to the rest of our data. As a result, there is no compelling argument for not taking them into account. Robust regression may be a useful method because it strikes a balance between including all data points and treating them equally in OLS regression, on the one hand, and excluding these points entirely from the study, on the other. Data must be given more or less weight depending on how well-behaved it is to perform robust regression. This weight could range between 0 and 100. One way to think of it is as an alternative to the least-squares regression method, but with weights and reweights [26].

2.2.3. Ridge Regression

If multicollinearity analysis is performed on the data from multiple regressions, it may be well suited to ridge regression. When multicollinearity exists, estimates calculated using the least number of squares are free of bias; however, these estimates frequently have large standard deviations and are thus inaccurate [27]. To reduce the standard errors produced by ridge regression, a small amount of purposeful bias is added to the regression estimates. It is expected that the final product will result in more accurate estimates [28].

2.2.4. LASSO Regression

The Least Absolute Shrinkage and Selection Operator is the full name of the concept represented by the acronym LASSO. Other regularization methods, such as lasso regression, are available [29]. It is preferred over regression algorithms because it produces more accurate forecasts. As part of its construction, this model uses the shrinking technique. Normalized data values tend to decrease as they approach a single number known as the mean. The Lasso regression technique employs the L1 regularization approach to carrying out its analysis. Because of the large number of features that we own, we can take advantage of its automatic feature selection [30].

2.2.5. Elastic Net

Elastic net regression was created in response to the observation that traditional OLS-based regression attempts to reduce bias but still exhibits significant variance. The elastic net, like the L1 (Ridge) and L2 (LASSO) techniques, considers the trade-off between bias and variance by selecting the appropriate penalties that minimize an information criterion or prediction error [31]. This is done similarly to the L1 (Ridge) and L2 (LASSO) techniques. The interaction of these two variables causes not only some coefficients to become smaller, but also some coefficients to become zero, which aids in the feature selection process. Since the regularization parameter λ can be set to a wide range of values, it is common to use methods like k-fold cross-validation to try out different λ settings and choose the one that best meets a certain goal, like reducing the amount of prediction error [32].

2.2.6. Polynomial Regression

The regression equation is referred to as a polynomial regression equation if the independent variable has a power that is greater than 1. The following equation displays an example of a polynomial equation. $Y = a + bX + cX2$. In this particular kind of regression analysis, the line that provides the best fit is not straight. It is a curve that fits the data points rather well and smoothly [33,34].

2.2.7. Stochastic Gradient Descent

Stochastic gradient descent is a method for optimizing machine learning programs based on stochastic differential equations. The strategy looks for a function that has a derivative that is very similar to the function being targeted over a number of iterations or epochs, which is known as a loss function [22].

Stochastic gradient descent is a machine-learning approach that makes minute adjustments to model parameters to achieve optimal levels of predicted accuracy. The term "stochastic" refers to the fact that during each iteration of the stochastic gradient descent algorithm, only a random subset of all training samples is used to calculate the error and update the model coefficients; this is what gives the algorithm its name. This method reduces the total amount of time spent training by reducing the amount of work required at each step [35].

2.2.8. Artificial Neural Networks

Artificial neural networks are a type of predictive strategy created by simplifying mathematical models of the human brain. These models enable complex nonlinear interactions between the response and the predictors of the response. A neural network can be thought of as a hierarchical structure made up of "neurons" that are all connected. The inputs (or predictors) are at the very bottom of the pyramid, and the predictions (or outputs) are at the very top. There is also the possibility of "hidden neurons" in the intermediary layers [36].

The most basic types of networks are similar to linear regressions and do not have any hidden layers. A neural network is used to implement the linear regression model with four variables. Weights are the coefficients that are connected to these predictors. The forecasts are created by linearly combining the inputs. To determine which weights should be used, the neural network framework employs a learning algorithm. As part of the learning process, this "cost function," which is frequently the MSE, is minimized. Because this is a simple situation, we can train the model using linear regression, which takes far less time and effort than other methods [37].

2.3. Data Preprocessing

With the help of one-hot encoding, a large number of possible alternate values for an attribute can be condensed down to a single binary value. The One-Hot-Encoding standard is the only requirement that this transformer must

meet to be considered compliant. The input data must be written as an integer matrix with discrete (categorical) attributes [38]. As a result, the sparse matrix will have several columns, each of which will represent a different outcome. It is assumed that the input properties will have values ranging from 0 to n. To convert each class into a numeric value, the Label Encoder must be used to change the properties of each class. Each type of parameter contained in the data collection has its dedicated data file. As a result of this change, the names of all of their parameters have been changed [39].

2.4. Feature Selection

During the pre-processing phase of data mining, feature selection, which is also referred to by its acronym FSS, is typically utilized. Getting rid of any unnecessary particulars will result in a reduction in dimensionality, which will, in turn, result in an improvement in the measurement's precision [40]. "Feature selection" refers to the process of determining which data points are most useful for generating an accurate prediction for a specific class. The phrase "feature selection" can be applied to a variety of situations, including the one just described. When it comes to selecting characteristics, three distinct approaches can be taken. A filter is applied first, followed by a wrapper, and finally an embedded component [41].

2.5. Performance Measurement

2.5.1. R-Squared

The coefficient of determination, also known as R-squared (R2) or the coefficient of determination, shows how much of the total variation in the dependent variable can be assigned to the independent variables in a regression model. R squared is used to explain how much the variance of one variable accounts for the variance of another, whereas correlation is used to show how closely two variables are related to one another. So, if a model's R2 is 0.50, it means that the factors that went into creating the model account for roughly half of the differences observed [42].

The coefficient of determination, also known as R squared, provides a rough estimate of the extent to which changes in one variable's value can be rationalized in terms of changes in another variable's value. It doesn't tell you if the data and forecasts are skewed, and it doesn't tell you anything about the model's quality. R-square is neither the best nor the worst metric to use when assessing the quality of a model because it does not indicate whether or not the correct regression was chosen. R-squared values greater than one indicate that the model does not fit the data well, while [43].

$$R^2 = 1 - \frac{\text{Unexplained Variation}}{\text{Total Variation}} \quad (1)$$

Several steps must be taken to calculate R squared. The best line of fit between the data points (observations) representing the dependent variable and the independent variable is an important step in this procedure. The observed data are then subtracted from the forecasted figures, and the difference is squared. This yields a table of squared errors, which when added up yields the same value as the residual variance.

2.5.2. Mean Squared Error (MSE)

A statistical measure that quantifies the average magnitude of errors is the mean squared error. Following that, it computes the square root of the difference between the estimated and actual values, computes the average of the squared values produced, and then generates a new value. MSE can only ever take on positive values due to the way errors are squared. Because almost all processes have some level of noise and unpredictability, MSE is not always zero.

$$MSE = \frac{1}{n}\sum_{i=\ 1}^{n}(Y_i - \hat{Y}_i)^2 \quad (2)$$

The MSE, like the variance, is expressed as a percentage with the same units as the square of the quantity being estimated. Furthermore, the MSE is named after the variation. The presence of outlying data points affects the square root of the mean error in the same way that it affects the variance. If one of the samples contains a "y" and an associated error that is significantly larger than the errors associated with the other samples, the square of the error will be significantly larger. As a result of the fact that MSE computes an average of errors, it is especially susceptible to values that are on the end of the scale.

2.5.3. Mean Absolute Error (MAE)

The mean absolute error, abbreviated as MAE, quantifies how far off the mark the actual value is compared to the value that was forecast. The researcher will outline an example to have a better understanding of how linear regression, which generates a line that is the best fit for the data input and output, might be applied.

One must now calculate the mean absolute error (MAE) that your model produced, which is a mistake made by the model. Determine the degree of deviation between the actual values and the predicted values, and then compute the mean absolute error across the entire dataset. The next step is to divide the total number of observations by the total number of errors. My name is Mae, and I'd like to introduce myself. Aside from this disheartening news, we intend to reduce the MAE as much as possible.

$$MAE = \frac{1}{n}\ \sum_{i=1}^{n}|x_i - x| \quad (3)$$

2.5.4. Root Mean Squared Error (RMSE)

The mean squared error (MSE) is calculated as the first step in calculating the root-mean-squared error (RMSE), which is determined by taking the square root of the MSE. The MSE is calculated by averaging the squared errors across all of the samples.

The root mean squared error (RMSE) is an error measure that employs the same measurement technique as the variable under consideration. The RMSE returns the error in dollars; however, the MSE returns the error in dollars squared, which is a much less obvious result. If our target

Y is the following year's sales in dollars, the RMSE will provide this information.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \hat{x}_i)^2}{N}} \quad (4)$$

2.5.5. Cross Validation

Cross-validation is one of the fundamental concepts involved in data modeling. "Before committing to a model, try evaluating it on a sample that wasn't used for training." This recommendation is short and easy to understand.

2.6. Study Model

The proposed algorithm has the following structure: The goal of these modeling strategies is to improve understanding and prediction of home energy efficiency. The following taxonomies were used to categorize the data:

-The dataset must be loaded before proceeding.

-The second phase involves any prerequisite processing that must be done.

-We next move into the next stage, which includes partitioning the dataset.

-In the fourth stage, the data set is split in half to create a training set and a test set.

-In the fifth stage, the feature subset selection measure is used to determine the most effective method of feature clustering. Uncertainty Symmetrical uncertainty (SU) cancels out any benefit that could come from having more information.

-The sixth stage, training for machine learning models, makes use of the data set.

-The next step is a group effort to establish an R2, MSE, RMSE, MAE, and cross-validation value.

Performance measurement scores were chosen as the primary evaluation statistic for this study because they are widely used in other machine learning metrics such as R2, MSE, RMSE, MAE, and cross-validation. The selected model's fit is evaluated to a horizontal line that serves as a standard. The Performance measurement value will be negative if the provided model results in a worse fit than a horizontal line. Although the "square" is used in the method for calculating Performance measurement scores, R2 can have a negative value. To put it another way, R2 will be negative if the model fits the data less well than a horizontal line and departs from the overall trend of the data.

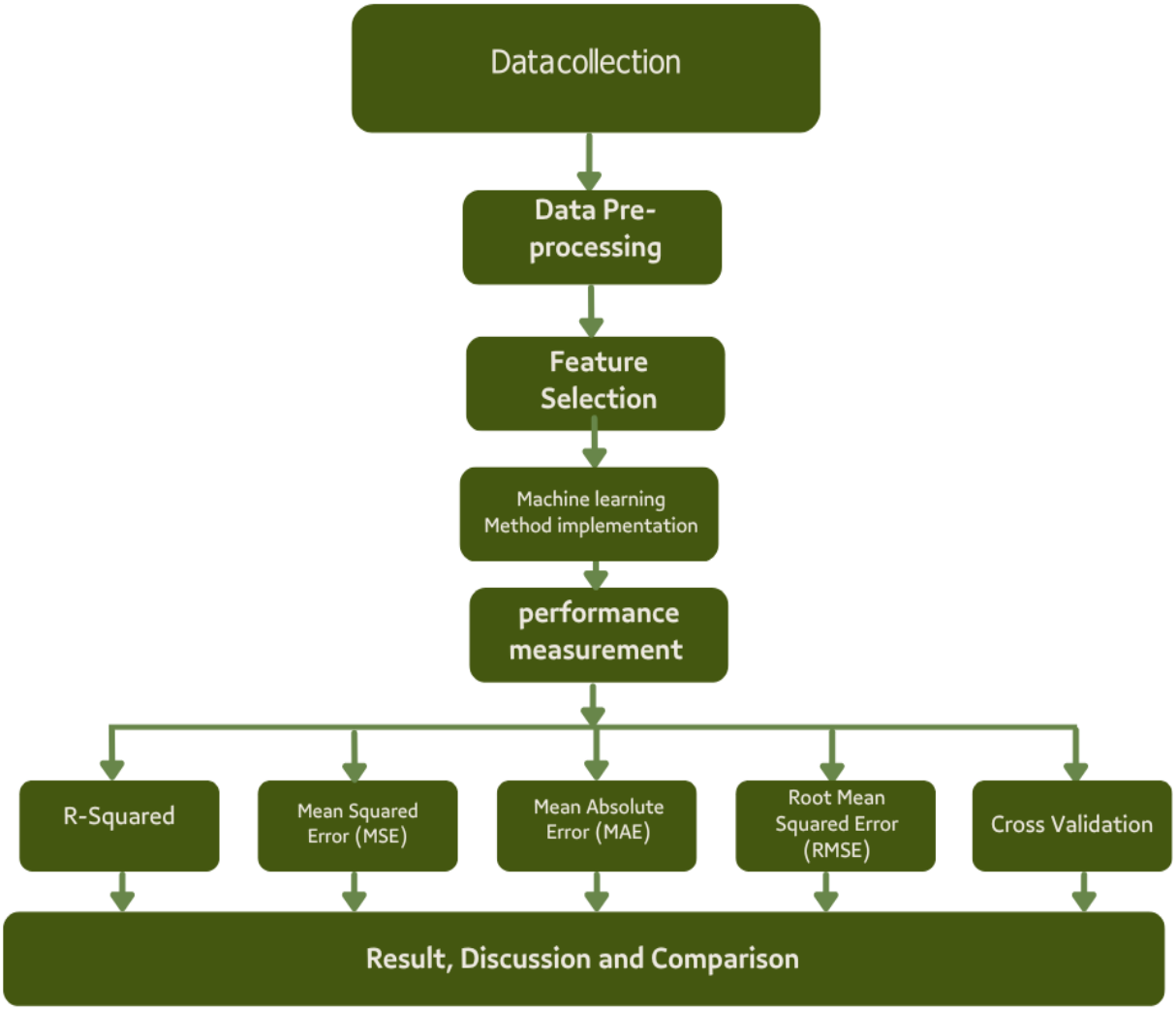

Figure 1. Study model

3. Experiment Results

The data collection on energy efficiency compiled by UCI was used as a benchmark. The dataset has 768 samples and 9 attributes, and its purpose is to make predictions about two real-valued responses. The dataset consists of eight different attributes and two response lists (Table 1). The objective is to make use of these eight characteristics to forecast not only the demand for heating (Y1) but also the demand for cooling (Y2). The overall height of the building (X5), the orientation of the building (X6), the surface area of the building (X7), the area of the building's walls (X3), the area of the building's windows (X2), and the distribution of the building's glass (X8) are the criteria that are employed (X6). Estimating the amount of cooling and heating needed is our primary objective, Table 2 illustrates a statistical analysis of the dataset parameters.

Table 1. Details about the energy dataset

Mathematical	Input or output variable
X1	Relative compactness
X2	Surface area
X3	Wall area
X4	Roof area
X5	Overall height
X6	Orientation
X7	Glazing area
X8	Glazing area distribution
Y1	Heating load
Y2	Cooling load

Table 2 displays a statistical examination of the parameters of the energy dataset. The parameter of Relative Compactness is comprised of 768 data points, exhibiting a mean value of roughly 0.763885 and a standard deviation of 0.105490. Each of the parameters, namely Surface Area, Wall Area, and Roof Area, comprises a total of 768 data points. The calculated average surface area is approximately 671.708333 units, with a corresponding standard deviation of 88.086116 units. Certain parameters are assigned both minimum and maximum values, with the former being explicitly stated. For instance, the minimum surface area is set at 514.500000, while the maximum value is not explicitly indicated. The parameter denoting the Overall Height exhibits a mean value of 5.25000, accompanied by a standard deviation of 1.75114. The tabular data incorporates Orientation, Glazing Area, Glazing Area Distribution, Heating Load, and Cooling Load variables, presenting descriptive statistics such as measures of central tendency, dispersion, and range, including quartile ranges. The statistical measures provide valuable insights into the central tendencies, variability, and range of the parameters in the energy dataset.

Figure 2 shows a heat map that depicts the previously mentioned correlation matrix. It provides a graphical representation of the relationships that exist between all of the variables, with areas of lower correlation represented by darker colors and areas of higher correlation represented by lighter colors. A more vibrant color, for example, denotes a link that is either more positive or more direct (in which a change in one quality can be seen to have an obvious and immediate effect on another), whereas a more subdued color denotes a link that is either more negative or more indirect (the variation of one characteristic inversely affects that of the other). When viewing the correlations displayed on the heatmap, it is clear that the following characteristics have links ranging from moderate to strong (between 0 and 1 for direct correlation and between 0 and -1.0 for inverse correlation).

The remaining twenty percent of the total patterns were used for out-of-sample prediction with the trained network. Eighty percent of the total patterns were used for training an artificial neural network (ANN). To accomplish this, we constructed a network of input-hidden-output modules, which led to the formation of a population vector that contained 79 nodes (weights). The range of possible weight values has been narrowed down to the range beginning at -10 and ending at 10. Randomization is used to generate the weights and biases for each individual, and the population size is initially set at fifty. A family of fifty members, spanning ten generations, was wiped out. The sigmoid function is utilized as the activation function of the neurons to evaluate the effectiveness of various machine-learning techniques. To calculate the Performance measurement score that distinguishes the training pattern set from the testing pattern set, we carry out 50 independent runs.

Table 3 provides a comprehensive summary of the performance measurement scores acquired during the testing phase for every machine learning model. The Linear Regression model exhibits a robust performance, producing a Mean Absolute Error (MAE) of 2.276661, Mean Squared Error (MSE) of 10.403479, Root Mean Squared Error (RMSE) of 3.225443, and a noteworthy R-squared value of 0.884583. The aforementioned statement implies that the model's prognostications exhibit an average deviation of roughly 2.28 units from the factual values, and it effectively accommodates the fluctuations in the dataset with a commendable alignment. The model of Robust Regression demonstrates a comparable level of performance, with an MAE of 2.285175, MSE of 11.168496, RMSE of 3.341930, and an R-squared value of 0.876096. The aforementioned scores suggest that the Robust Regression model is capable of generating dependable forecasts, exhibiting an average absolute error of roughly 2.29 units and a satisfactory adherence to the dataset.

Table 2. Statistical analysis about the energy dataset parameters.

	Relative Compactness	Surface Area	Wall Area	Roof Area	Overall Height	Orientation	Glazing Area	Glazing Area Distribution	Heating Load	Cooling Load
count	768.000000	768.000000	768.000000	768.000000	768.00000	768.000000	768.000000	768.00000	768.000000	768.000000
mean	0.763885	671.708333	318.500000	176.604167	5.25000	3.500000	0.234375	2.81250	22.307201	24.587760
std	0.105490	88.086116	43.626481	45.165950	1.75114	1.118763	0.133221	1.55096	10.090196	9.513306
min	0.620000	514.500000	245.000000	110.250000	3.50000	2.000000	0.000000	0.00000	6.010000	10.900000
25%	0.682500	606.375000	294.000000	140.875000	3.50000	2.750000	0.100000	1.75000	12.992500	15.620000
50%	0.750000	673.750000	318.500000	183.750000	5.25000	3.500000	0.250000	3.00000	18.950000	22.080000
75%	0.820000	741.125000	343.000000	220.500000	7.00000	4.250000	0.400000	4.00000	31.667500	33.132500

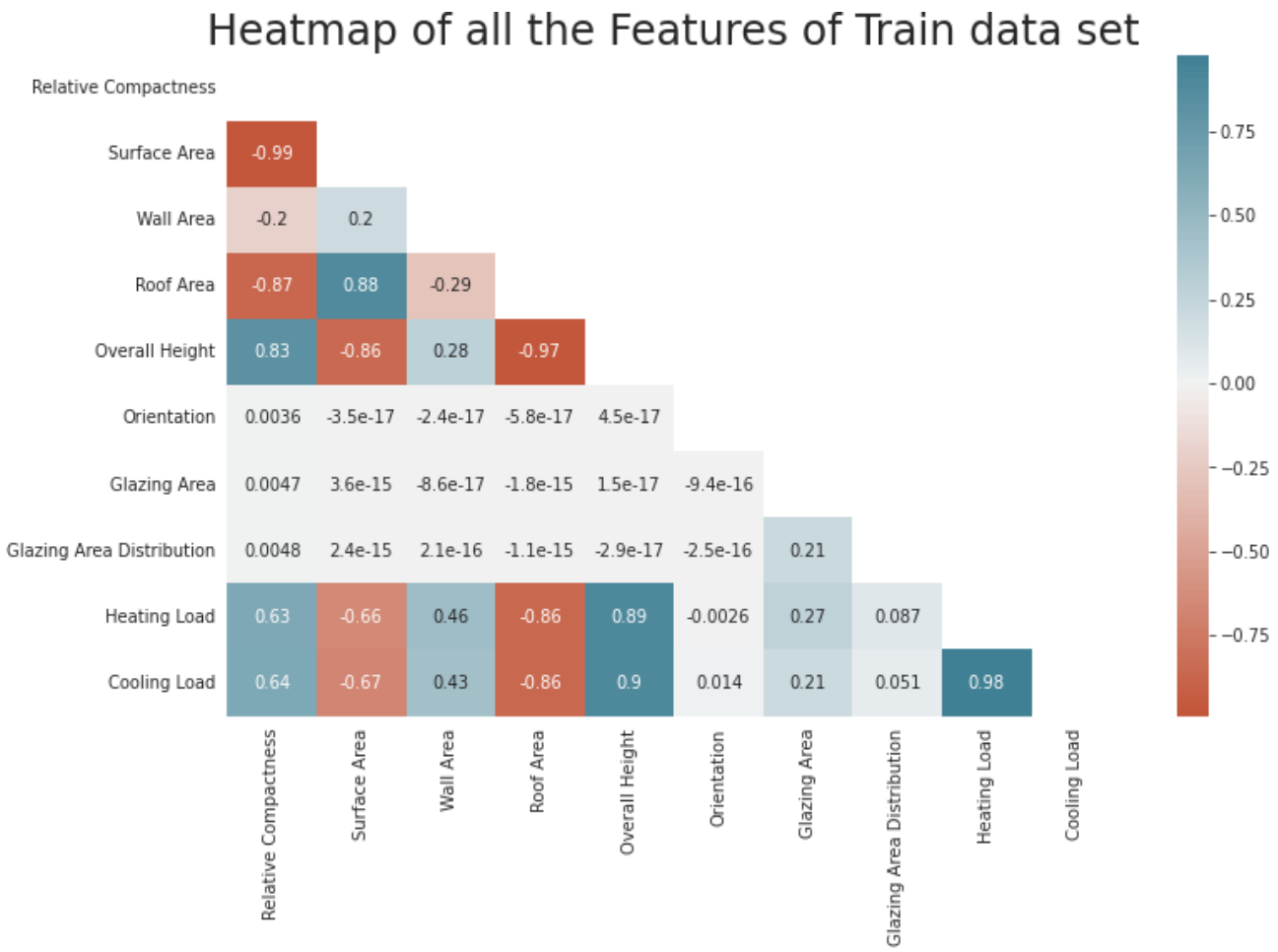

Figure 2. Correlation matrix.

Table 3. Performance measurement score during testing for each machine learning model.

	Model	MAE	MSE	RMSE	R Square	Cross Validation
1	Linear Regression	2.276661	10.403479	3.225443	0.884583	0.863782
2	Robust Regression	2.285175	11.168496	3.341930	0.876096	0.854787
3	Ridge Regression	2.512286	13.776656	3.711692	0.847161	0.869961
4	Lasso Regression	2.377280	11.165299	3.341452	0.876132	0.751754
5	Elastic Net Regression	2.368227	11.179321	3.343549	0.875976	0.762353
6	Polynomial Regression	1.161049	2.893625	1.701066	0.967898	-
7	Stochastic Gradient Descent	2.597367	13.920465	3.731014	0.845566	-
8	Artificial Neural Network	3.245003	20.063715	4.479254	0.777412	-
9	Random Forest Regressor	1.046031	2.827819	1.681612	0.968628	-

The table presents the performance scores for the Ridge Regression, Lasso Regression, and Elastic Net Regression models. The models demonstrate diverse levels of performance, as evidenced by their corresponding MAE, MSE, RMSE, and R-squared metrics. The aforementioned scores provide valuable insights regarding the models' efficacy in minimizing errors and accurately capturing the variance present in the data. The Polynomial Regression model exhibits outstanding performance, which is noteworthy. The achieved results exhibit notably low levels of error, with a mean absolute error (MAE) of 1.161049, mean squared error (MSE) of 2.893625, and root means squared error (RMSE) of 1.701066. Moreover, the analysis showcases a noteworthy R-squared coefficient of 0.967898, which suggests a robust association between the projected and observed data points. The absence of a cross-validation score for the Polynomial Regression model in the table raises concerns regarding its ability to generalize to new

and unseen data.

Various models, including Stochastic Gradient Descent, Artificial Neural Network, and Random Forest Regressor, demonstrate their performance through metrics such as MAE, MSE, RMSE, and R-squared values. The aforementioned scores offer significant insights regarding the precision and prognostic capabilities of said models in approximating the target variable. In general, Table 3 presents a thorough assessment of the efficacy of diverse machine learning models, elucidating their respective merits and demerits with respect to predictive precision and congruity with the data.

The performance measurement scores for each machine learning model during testing are presented in Table 4. The results obtained from the Linear Regression model indicate a Mean Absolute Error (MAE) of 2.285851, Mean Squared Error (MSE) of 10.352179, Root Mean Squared Error (RMSE) of 3.235463 and an R-squared value of 0.895248. The results demonstrate that the Linear Regression model exhibits an average absolute error of roughly 2.29 units and accounts for approximately 89.52% of the variance in the target variable. The reported cross-validation score of the model is 0.863254, indicating a satisfactory level of generalization to unobserved data.

The model of Robust Regression exhibits comparable performance, as evidenced by an MAE of 2.278525, MSE of 11.952096, RMSE of 3.31253, and an R-squared value of 0.885623. The aforementioned scores suggest that the Robust Regression model offers dependable prognostications with a marginally lesser Mean Absolute Error in comparison to the Linear Regression model. The model explains approximately 88.56% of the variability observed in the dependent variable. The model demonstrates a cross-validation score of 0.875214, which suggests a satisfactory level of generalizability. The Ridge Regression, Lasso Regression, and Elastic Net Regression models yield performance metrics such as MAE, MSE, RMSE, R-squared values, and cross-validation scores. The Ridge Regression model exhibits a Mean Absolute Error (MAE) of 2.535846, Mean Squared Error (MSE) of 13.78826, Root Mean Squared Error (RMSE) of 3.77432, and an R-squared coefficient of determination value of 0.852314. The Lasso Regression model exhibits an MAE of 2.38658, MSE of 11.155219, RMSE of 3.336522, and an R-squared value of 0.863254. The Elastic Net Regression model demonstrates an MAE of 2.352547, MSE of 11.165201, RMSE of 3.3493179, and an R-squared value of 0.885236. The aforementioned scores serve as an indication of the model's performance in relation to error metrics and their aptitude for elucidating the variance in the target variable.

The Polynomial Regression model exhibits outstanding performance, as evidenced by its minimal errors. The model attains a Mean Absolute Error (MAE) of 1.171529, Mean Squared Error (MSE) of 2.885215, Root Mean Squared Error (RMSE) of 1.784266, and a noteworthy R-squared coefficient of determination of 0.978523. Nevertheless, the model lacks a cross-validation score. The Stochastic Gradient Descent, Artificial Neural Network, and Random Forest Regressor models exhibit their respective performance scores, which reflect their efficacy in terms of MAE, MSE, RMSE, R-squared values, and cross-validation scores.

To summarize, Table 4 presents a comprehensive examination of the performance measurement scores of every machine learning model, offering valuable insights into their predictive aptitude and capacity to generalize to novel data.

The primary criterion for comparison was the Performance of the model with which each approach to machine learning performed local searches. The results of the Performance measurement test for each machine learning algorithm used in the model are presented in Table 2 and Figure 3.

Table 4. Performance measurement score during testing for each machine learning model.

	Model	MAE	MSE	RMSE	R Square	Cross Validation
1	Linear Regression	2.285851	10.352179	3.235463	0.895248	0.863254
2	Robust Regression	2.278525	11.952096	3.31253	0.885623	0.875214
3	Ridge Regression	2.535846	13.78826	3.77432	0.852314	0.885236
4	Lasso Regression	2.38658	11.155219	3.336522	0.863254	0.785214
5	Elastic Net Regression	2.352547	11.165201	3.3493179	0.885236	0.79632
6	Polynomial Regression	1.171529	2.885215	1.784266	0.978523	-
7	Stochastic Gradient Descent	2.589637	13.929665	3.763584	0.896351	-
8	Artificial Neural Network	3.232013	20.002515	4.485264	0.796325	-
9	Random Forest Regressor	1.096331	2.898419	1.631412	0.985236	-

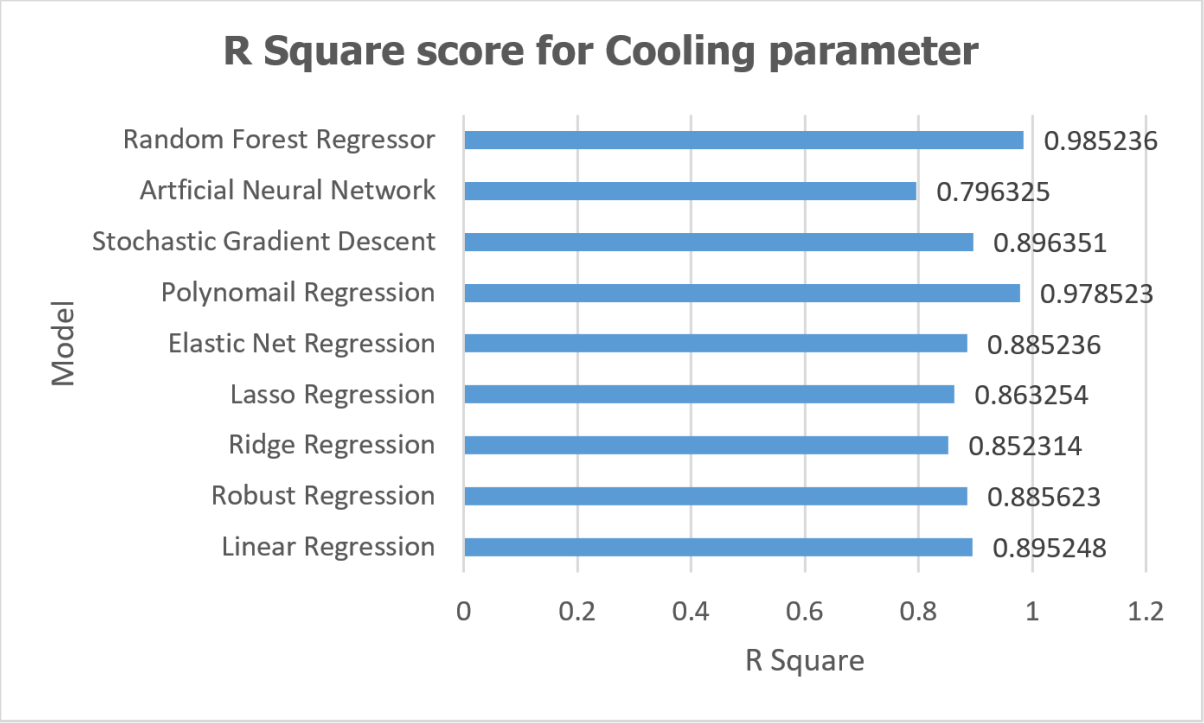

Figure 3. Performance measurement score comparison.

The R Square score for the Cooling parameter, as depicted in Figure 4, demonstrates the efficacy of the machine learning techniques in forecasting the outcome variable using the provided dataset. The Random Forest Regressor attained a superior R-squared value of 0.985236, signifying its exceptional ability to capture the interrelationships between the independent and dependent variables. The Random Forest Regressor is an ensemble learning technique that amalgamates numerous decision trees, rendering it efficacious in managing intricate associations and furnishing precise prognostications. The performance of the Polynomial Regression was strong, as evidenced by an R-squared score of 0.978523. The polynomial regression technique is utilized to fit a polynomial equation to the given dataset, thereby enabling it to effectively capture non-linear associations among the variables. The model exhibited a high degree of similarity to the Random Forest Regressor, indicating its efficacy in accurately representing the data. The R-squared score attained by the Artificial Neural Network was 0.796325. Neural networks possess significant modeling capabilities that enable them to acquire intricate patterns, albeit their efficacy is contingent upon the architecture and training methodology. Although the score exhibits a comparatively lower value in contrast to the preceding methodologies, it still signifies a reasonable degree of predictive capacity.

Stochastic Gradient Descent, Linear Regression, Elastic Net Regression, Robust Regression, Lasso Regression, and Ridge Regression were evaluated and found to have achieved R-squared scores within the range of 0.852314 to 0.896351. Various methodologies employ distinct techniques to approximate the association among variables and generate prognostications. Despite exhibiting a slightly lower performance in comparison to the highest-performing models, these models still demonstrate a respectable level of predictive capability. In light of the R-squared scores, it can be inferred that the Random Forest Regressor and Polynomial Regression techniques exhibited superior performance compared to the remaining methods in capturing the associations within the provided dataset. The optimal selection of a machine learning technique is contingent upon several factors, such as the particular demands of the task, the characteristics of the data, and the balance between interpretability and predictive precision.

Regarding Figure 4, the Random Forest Regressor captured the correlations between independent factors and the Heating parameter with a high R-squared score of 0.968628. The Random Forest Regressor handles complicated interactions and accurately predicts the Heating parameter using an ensemble of decision trees. Polynomial Regression scored 0.967898 R-squared. This approach uses a polynomial equation to capture non-linear variables. It models the Heating parameter well, matching the Random Forest Regressor.

R-squared was 0.777412. Neural networks can learn complicated patterns; hence this score is lower than earlier techniques. However, design and training substantially impact artificial neural network performance. This problem's score is lower, but it still shows some predictive ability. Stochastic Gradient Descent, Elastic Net Regression, Lasso Regression, Ridge Regression, Robust Regression, and Linear Regression have R-squared values

from 0.884583 to 0.845566. These techniques estimate and predict variables using various ways. Their Heating parameter prediction power is comparable to the best models, despite their lower scores. According to R-squared scores, the Random Forest Regressor and Polynomial Regression algorithms are best at predicting the Heating parameter. When picking a machine learning approach, interpretability, computational complexity, and problem requirements are all significant.

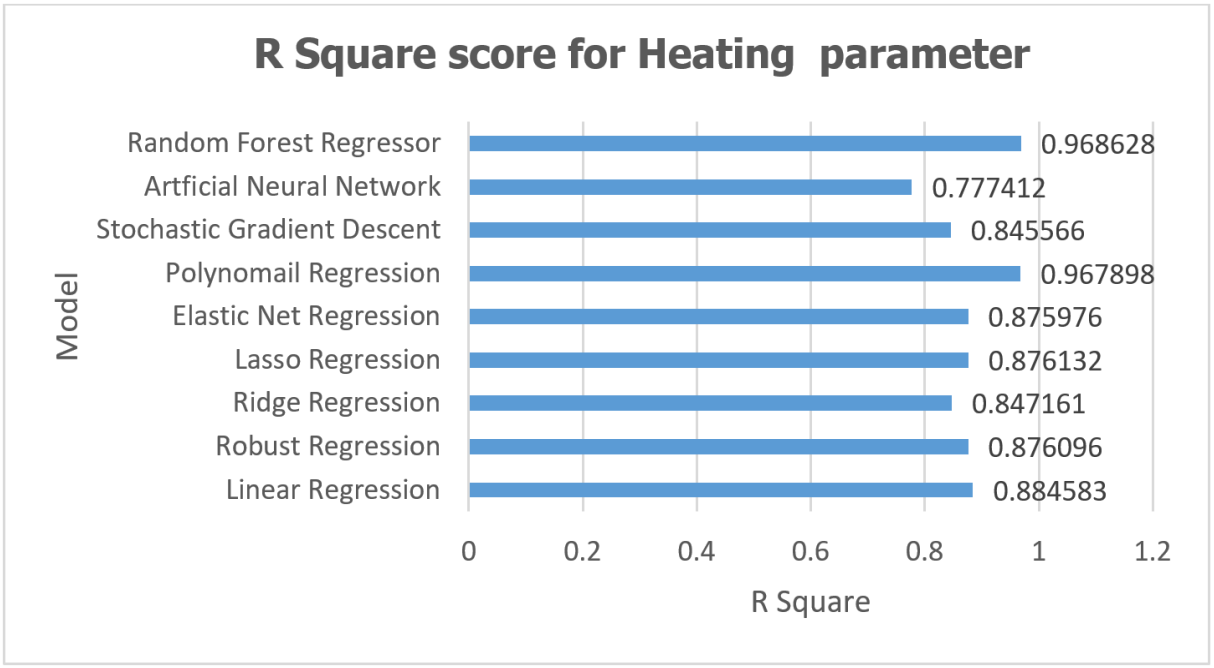

Figure 4. Performance measurement score comparison.

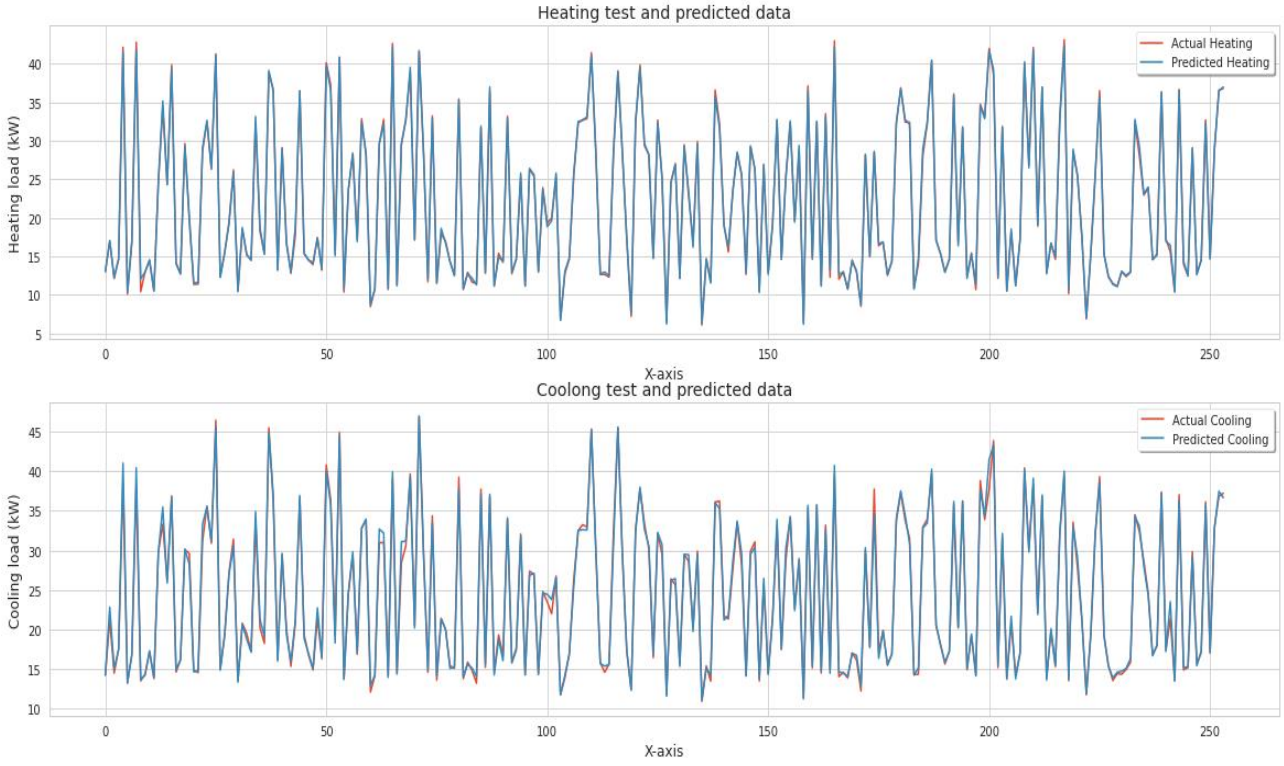

Figure 5. Heating and cooling test and predicted data.

Figure 5 visually presents the comparison between the heating and cooling loads calculated using predictions from an Artificial Neural Network (ANN) model and the corresponding experimental data. The figure serves as evidence that the machine learning (ML) predictor's predictions align well with the target values. This indicates that the ANN model accurately captures the patterns and relationships in the data, enabling it to make reliable predictions. Moreover, if given the opportunity, the Polynomial regressor model shows promise in providing even more precise forecasts. By further training and optimizing the model, its accuracy and predictive capability can potentially be enhanced. It is noteworthy that the method used for training the ML models demonstrated the production of more compact neural networks while reducing the number of testing errors. This was achieved by initiating the evolution process with an increased number of generations. The favorable performance of the models, as depicted in Figure 5, where the predicted data closely aligns with the actual data, reinforces the credibility and reliability of the gathered information.

4. Conclusions

The researcher used ten different stacking models to get an accurate estimate of the HL and CL parameters. These machine-learning models were created using datasets that were collected and analyzed. Input variable included relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, and glazing area distribution. The parameters considered for output were the HL and CL values. The first step in the process involved compiling data sets to determine which factors provided the most reliable forecasts for these two variables. Group 9 depicts the ideal state, which is the best that can be achieved by combining six different factors. The difficulty of the problem was significantly simplified, which made it much easier to deal with. After that, sensitivity analyses were carried out to locate the parameters of the model's foundation that should have the most desirable values. We find that the Random Forest Regressor has the highest R2 of all the models. This means that it is the most accurate model for making predictions.

Based on the evaluation of performance metrics, the Random Forest Regressor and Polynomial Regression models emerge as the top performers. The Random Forest Regressor consistently demonstrates accurate predictions with the lowest errors and high R-squared values. The Polynomial Regression model showcases low errors and a strong correlation with the target variable. However, the authors' decision not to choose the model with the lowest MSE and the inconsistency between the mention of the Artificial Neural Network and Polynomial Regression models require clarification. Overall, the Random Forest Regressor and Polynomial Regression models offer the best performance among the evaluated machine learning models.

In the future, more complex machine learning models, such as a combination of the data-handling fuzzy approach and the whale optimization method for group data, could improve the results. This will produce better results. We hope to achieve our research goal shortly by expanding and improving the existing database with new information. This will enable the development of methods that use soft computing to reduce the negative effects that buildings have on the environment. If we were trying to accelerate the process of achieving our goal, it would be useful to have a tool that could estimate how much it would cost to implement energy-saving measures in buildings.

REFERENCES

[1] X. Wu, Z. Feng, H. Chen, Y. Qin, S. Zheng, L. Wang, Y. Liu, and M. J. Skibniewski, "Intelligent optimization framework of Near Zero Energy Consumption Building Performance Based on a hybrid machine learning algorithm," Renewable and Sustainable Energy Reviews, vol. 167, p. 112703, 2022.

[2] J. S. Chou and D.-S. Tran, "Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders," Energy, vol. 165, pp. 709-726, 2018.

[3] A. Almalaq and J. J. Zhang, "Evolutionary deep learning-based energy consumption prediction for buildings," IEEE Access, vol. 7, pp. 1520-1531, 2019.

[4] Y. Liu, H. Chen, L. Zhang, X. Wu, and X.-jia Wang, "Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in China," Journal of Cleaner Production, vol. 272, p. 122542, 2020.

[5] I. Imran, N. Iqbal, and D. H. Kim, "IOT task management mechanism based on predictive optimization for efficient energy consumption in smart residential buildings," Energy and Buildings, vol. 257, p. 111762, 2022.

[6] A. Esmaeilzadeh, M. R. Zakerzadeh, and A. Y. Koma, "The comparison of some advanced control methods for energy optimization and comfort management in buildings," Sustainable Cities and Society, vol. 43, pp. 601-623, 2018.

[7] Y. M. Lee, L. An, F. Liu, R. Horesh, Y. T. Chae, and R. Zhang, "Applying science and mathematics to big data for Smarter Buildings," Annals of the New York Academy of Sciences, vol. 1295, no. 1, pp. 18-25, 2013.

[8] A. Habibollahzade, E. Gholamian, P. Ahmadi, and A. Behzadi, "Multi-criteria optimization of an integrated energy system with thermoelectric generator, parabolic trough solar collector and electrolysis for hydrogen production," International Journal of Hydrogen Energy, vol. 43, no. 31, pp. 14140-14157, 2018.

[9] M. Ilbeigi, M. Ghomeishi, and A. Dehghanbanadaki, "Prediction and optimization of energy consumption in an office building using artificial neural network and a genetic algorithm," Sustainable Cities and Society, vol. 61, p.

102325, 2020.

[10] H. xiang Zhao and F. Magoulès, "A review on the prediction of Building Energy Consumption," Renewable and Sustainable Energy Reviews, vol. 16, no. 6, pp. 3586-3592, 2012.

[11] N. T. Ngo, A. D. Pham, T. T. H. Truong, N. S. Truong, and N. T. Huynh, "Developing a hybrid time-series artificial intelligence model to forecast energy use in buildings," in *Scientific Reports*, vol. 12, no. 1, p. 15775, 2022.

[12] K. Spees and L. Lave, "Impacts of responsive load in PJM: Load shifting and real-time pricing," The Energy Journal, vol. 29, no. 2, 2008. https://doi.org/10.5547/issn0195-6574-ej-vol29-no2-6.

[13] Q. Lin, K. Liu, B. Hong, X. Xu, J. Chen, and W. Wang, "A data-driven framework for abnormally high building energy demand detection with weather and block morphology at community scale," *Journal of Cleaner Production*, vol. 354, p. 131602, 2022.

[14] K. Amarasinghe, D. Wijayasekara, and M. Manic, "Neural network based downscaling of Building Energy Management System Data," in 2014 IEEE 23rd International Symposiu on Industrial Electronics (ISIE), 2014. https://doi.org/10.1109/isie.2014.6865042.

[15] A. I. Dounis, "Artificial Intelligence for energy conservation in buildings," Advances in Building Energy Research, vol. 4, no. 1, pp. 267-299, 2010. https://doi.org/10.3763/aber.2009.0408.

[16] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "Arima models to predict next-day electricity prices," IEEE Transactions on Power Systems, vol. 18, no. 3, pp. 1014-1020, 2003. https://doi.org/10.1109/tpwrs.2002.804943.

[17] C. Roldán-Blay, G. Escrivá-Escrivá, C. Roldán-Porta, and C. Álvarez-Bel, "An optimization algorithm for distributed energy resources management in micro-scale Energy Hubs," Energy, vol. 132, pp. 126-135, 2017. https://doi.org/10.1016/j.energy.2017.05.038.

[18] J. G. Jetcheva, M. Majidpour, and W.-P. Chen, "Neural network model ensembles for building-level electricity load forecasts," Energy and Buildings, vol. 84, pp. 214-223, 2014. https://doi.org/10.1016/j.enbuild.2014.08.004.

[19] L. Ghelardoni, A. Ghio, and D. Anguita, "Energy load forecasting using empirical mode decomposition and support vector regression," IEEE Transactions on Smart Grid, vol. 4, no. 1, pp. 549-556, 2013. https://doi.org/10.1109/tsg.2012.2235089.

[20] R. Cheng, J. Yu, M. Zhang, C. Feng, and W. Zhang, "Short-term hybrid forecasting model of ice storage air-conditioning based on improved SVR," *Journal of Building Engineering*, vol. 50, p. 104194, 2022.

[21] M. A. Al-Gunaid, M. V. Shcherbakov, D. A. Skorobogatchenko, A. G. Kravets, and V. A. Kamaev, "Forecasting energy consumption with the data reliability estimation in the management of hybrid energy system using fuzzy decision trees," in 2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA), 2016. https://doi.org/10.1109/iisa.2016.7785413.

[22] Y. Li and Y. Liang, "Learning overparameterized neural networks via stochastic gradient descent on structured data," in Advances in Neural Information Processing Systems, 2018.

[23] P. L. Bartlett, P. M. Long, G. Lugosi, and A. Tsigler, "Benign overfitting in linear regression," *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30063-30070, 2020.

[24] D. Maulud and A. M. Abdulazeez, "A review on linear regression comprehensive in machine learning," Journal of Applied Science and Technology Trends, vol. 1, no. 4, pp. 140-147, 2020.

[25] A. Liu, G. Shi, S. J. Chung, A. Anandkumar, and Y. Yue, "Robust regression for safe exploration in control," in Learning for Dynamics and Control, 2020, pp. 608-619.

[26] A. Jambulapati, J. Li, T. Schramm, and K. Tian, "Robust regression revisited: Acceleration and improved estimation rates," in Advances in Neural Information Processing Systems, 2021, pp. 4475-4488.

[27] A. Tsigler and P. L. Bartlett, "Benign overfitting in ridge regression," arXiv preprint arXiv:2009.14286, 2020.

[28] P. Fan, R. Deng, J. Qiu, Z. Zhao, and S. Wu, "Well logging curve reconstruction based on kernel ridge regression," Arabian Journal of Geosciences, vol. 14, no. 16, pp. 1-10, 2021.

[29] J. Ranstam and J. Cook, "LASSO regression," Journal of British Surgery, vol. 105, no. 10, p. 1348, 2018.

[30] R. Alhamzawi and H. T. M. Ali, "The Bayesian adaptive lasso regression," Mathematical biosciences, vol. 303, pp. 75-82, 2018.

[31] R. Xiao, X. Cui, H. Qiao, X. Zheng, Y. Zhang, C. Zhang, and X. Liu, "Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net," Biomedical Signal Processing and Control, vol. 66, p. 102362, 2021.

[32] J. Liu, G. Liang, K. D. Siegmund, and J. P. Lewinger, "Data integration by multi-tuning parameter elastic net regression," BMC bioinformatics, vol. 19, no. 1, pp. 1-9, 2018.

[33] X. Cheng, B. Khomtchouk, N. Matloff, and P. Mohanty, "Polynomial regression as an alternative to neural nets," arXiv preprint arXiv: 1806.06850, 2018.

[34] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos, "Polylanenet: Lane estimation via deep polynomial regression," in 2020 25th International Conference on Pattern Recognition (ICPR), 2021.

[35] R. Bassily, V. Feldman, C. Guzmán, and K. Talwar, "Stability of stochastic gradient descent on nonsmooth convex losses," in Advances in Neural Information Processing Systems, 2020, pp. 4381-4391.

[36] G. R. Yang and X.-J. Wang, "Artificial neural networks for neuroscientists: A primer," Neuron, vol. 107, no. 6, pp. 1048-1070, 2020.

[37] Q. Zhang, H. Yu, M. Barbiero, B. Wang, and M. Gu, "Artificial neural networks enabled by nanophotonics," Light: Science & Applications, vol. 8, no. 1, pp. 1-14, 2019.

[38] O. O. Kartashov, A. V. Chernov, D. S. Polyanichenko, and M. A. Butakova, "XAS Data Preprocessing of Nanocatalysts for Machine Learning Applications," Materials, vol. 14, no. 24, p. 7884, 2021.

[39] D.-J. Kim, C.-G. Hwang, and C.-P. Yoon, "Learning data preprocessing technique for improving indoor positioning performance based on machine learning," Journal of the Korea Institute of Information and Communication Engineering, vol. 24, no. 11, pp. 1528-1533, 2020.

[40] B. Remeseiro and V. Bolon-Canedo, "A review of feature selection methods in medical applications," Computers in biology and medicine, vol. 112, p. 103375, 2019.

[41] P. Agrawal, H. F. Abutarboush, T. Ganesh, and A. W. Mohamed, "Metaheuristic algorithms on feature selection: A survey of one decade of research (2009-2019)," IEEE Access, vol. 9, pp. 26766-26791, 2021.

[42] N. Redell, "Shapley Decomposition of R-Squared in Machine Learning Models," arXiv preprint arXiv:1908.09 718, 2019.

[43] A. P. Varshini, K. A. Kumari, D. Janani, and S. Soundariya, "Comparative analysis of Machine learning and Deep learning algorithms for Software Effort Estimation," Journal of Physics: Conference Series, 2021.