Computer
& Robotics

A Model for Predicting Building Energy Consumption Based on the Stacking of Machine Learning Regression Models

Mohammad Hosein Khodadadi[a], Ladan Riazi[a,*], Samaneh Yazdani[b]

[a]*Department of Information Technology Management, Science and Research Branch, Islamic Azad University, Tehran, Iran*
[b]*Department of Computer Engineering,, Tehran North Branch, Islamic Azad University, Tehran, Iran*

Received 18 April 2023; Accepted 15 June 2023

Abstract

In different societies, buildings are considered one of the main energy consumers in the world, and accordingly, they are responsible for a significant percentage of greenhouse gas emissions. Due to the upward growth of the population, the demand for energy consumption is increasing day by day. In such a situation, the prediction of energy consumption has become a vital issue to control the efficiency of energy consumption. To obtain an effective solution to solve this problem, a number of machine learning methods were examined and Xgboost and MLP methods were selected as the best available methods. In order to obtain more suitable results in this research, a system based on stacking was proposed. In the proposed method based on stacking, XGBoost and MLP methods were used in the first level so that the advantages of both methods can be used. The predictions made by each of these methods, in the second level, were used as input to another XGBoost algorithm, which was used as a meta-learner. To obtain better results, the hyperparameters of the basic techniques were optimized using the successive halving search. For a better comparison, machine learning regression techniques were implemented to solve the problem of energy consumption intensity prediction, and the results obtained from them were analyzed on WiDS Datathon. The results showed that the proposed system has improved the MAE, MAPE, and R2 criteria by 0.6, 0.03, and 0.07, respectively, compared to the best existing method.

Keywords: Energy Consumption, Stacking, Regression, XGBoost, MLP, AI ,Deep Learning

1.Introduction

Currently, buildings are the cause of a significant part of the total global energy consumption (Sun et al., 2020). In addition, One-third of the world's greenhouse gas emissions are due to energy consumption in buildings. And in this way, this issue can be one of the main reasons for global warming and extensive climate changes (Wei et al, 2018).

On the one hand, many international efforts are being made to deal with climate change, which can lead to irreversible changes on the planet. On the other hand, due to the increase in population, the need for energy resources is increasing day by day. These fundamental issues have increased the importance of energy consumption.

Due to the high consumption share of buildings, many measures have been taken for energy saving and efficiency in different parts of the world, the main goal of which is to minimize energy consumption and emission of greenhouse gases in the field of buildings.

One of the modeling approaches in the literature on building energy consumption prediction is data-driven models in which historical and existing data are used to train machine learning methods. This information typically includes data related to the energy of buildings, occupant information, and external variables (for example, meteorological information). The main advantage of data-driven methods is not require detailed physical information about the building (Chen et al, 2022). Data-driven approaches usually use three types of data to predict future energy consumption. These data include the

*Corresponding Author. Email: l.riazi@gmail.com

energy consumption of the building in the form of a time series, meteorological data inside or outside the building (such as temperature, relative humidity, wind speed, and sunlight), and time index variables (such as the type of day and working hours).

In recent years, the growth of the Internet of Things networks and information technology provide many advantages and opportunities to collect large data sets in the field of building energy consumption. In this way, according to the advantages that data-driven systems have in the field of solving the problem of predicting building energy consumption, effective solutions can be found in this field. In this regard, this research presents an approach based on stacking machine learning techniques to solve the problem of predicting building energy consumption, which uses the available data in this field. The proposed approach of this research combines the advantages of Xgboost and MLP methods.

In the continuation of this research, in section 2, the recent works done in the field of research will be reviewed and a summary of them will be given at the end. In section 3, the used data set is described and the pre-processing applied to it is discussed. In section 4, the important features of the data set and how to achieve them are described. The proposed stacking-based system is described in section 5. The section 6, discusses how to evaluate regression methods and their evaluation criteria. In addition, the results of the implementation were discussed. For a better evaluation of the proposed method in this section, several regression methods were applied to the data set and the results are shown. In addition, the results obtained from the proposed system are shown and will be compared with the results obtained from other regression methods. Finally, the conclusion of the research is given in section 7.

2. Related Works

The increase in global energy demand and its environmental effects increase the need to design reliable energy demand forecasting models. According to the research conducted, one of the high consumption areas in this field is construction. In this regard, many models have been presented in recent years to predict energy consumption in buildings, and some of them are reviewed in this section.

In the research of Lee et al (Lee et al, 2019), it was stated that the amount of energy consumption in the building depends on the behavioral characteristics of its residents in addition to the building characteristics. In this regard, in this research, an approach based on artificial neural network was proposed, which predicts the energy consumption of the building based on its user. In this research, the real daily information of 5240 single-person households was used. The energy consumed in the building of people was clustered and analyzed according to gender, age, occupation, income, education level, and period of occupation of the building. The simulation results showed that there are interesting behavioral similarities in the field of energy consumption for people with the same characteristics.

In the research (Zekić-Sušac et al, 2021), artificial neural networks, CART, and random forest regression trees were used to predict energy consumption and its cost in public buildings. In this work, real data from the Croatian public sector were used. In this research, three different methods for feature selection were tested: all variables, filter-based selection method, and packing-based method that integrates Boruta and random forest. Among the models examined in this research, the model with the integration of RF-Boruta and machine learning methods showed the best accuracy. Also, heating and occupational data were selected as the most important and influential features.

In the article (Banik et al, 2021), a model was proposed to forecast electricity consumption in India, which can predict the next day's energy consumption with appropriate accuracy, in addition, it can also make this forecast for the next week or month. In this paper, a cumulative approach based on Random Forest and XGBoost was proposed. The results showed that this combination improves accuracy by 15-29%.

In the research (Wang et al, 2020), a stack model was presented in which various features of basic algorithms were used. Algorithms considered as basic algorithms in this research include RF, GBDT, XGBoost, SVR, and kNN algorithms. These algorithms form the first layer. The output of these models was considered a meta-feature. These outputs were considered as the input of the second layer. The second layer works based on the GBDT model. This method was evaluated with the data from two real campus buildings. The performance of this stack model is significantly improved compared to the base models.

In the research (Lu et al, 2020), a hybrid model named CEEMDAN-XGBoost was proposed. In this model, CEEMDAN was used to remove non-conventional data with large fluctuations. The model based on XGBoost achieved good results. In this study, data from the CBU input tower in the United States was used to evaluate the system. According to the results, the results of the CEEMDAN-XGBoost model were recorded as better than the XGBoost model, and the performance results of this model were better than CEEMDAN-RF.

The research (Liu et al, 2020) implemented deep reinforcement learning techniques to predict energy consumption. In this research, A3C, DDPG, and RDPG methods were investigated. The performance of these algorithms was compared with three supervised models including MLR, BPNN, and RF. The results showed that the DDPG and RDPG algorithms performed well for the power prediction system, but required a relatively long calculation time. Studies on the A3C algorithm have shown that this algorithm is not suitable for solving this problem.

In the research (Ilbeigi et al, 2020), a neural network method based on multilayer perceptron was proposed. In this research, the optimization of important parameters was done using a genetic algorithm. The implementation results showed that with this method, energy consumption can be reduced by 35%. For the prediction of energy consumption, the obtained results were acceptable for the evaluation criteria studied.

Research (Shao et al, 2020) used the support vector machine model to predict energy consumption in hotels. In this model, the weather characteristics and air conditioning characteristics of the hotel were considered as system inputs. Performing data pre-processing, cross-validation, and parameter optimization improved system performance. The results of the implementation of the proposed system were satisfactory. The RBF kernel function was selected as the support vector machine kernel function, and the prediction accuracy of the model was improved by optimizing the kernel parameters. The predicted MSE value of the final model was 2.22%.

In the research (Luo et al, 2020), a model was implemented to predict building energy consumption based on deep neural networks. This model improved its performance by using a feature extraction technique and genetic algorithm. In this model, the K-means clustering method was used to extract features. After implementing the proposed model, it was evaluated on real office building data in England. The comparison results show the significant superiority of this model (FE-GA-DNN) over basic models such as GA-DNN. The value obtained for the R2 criterion in training data is 0.993 and for test data is 0.960.

In their research, Seydzadeh et al. (2020) proposed a multi-objective method based on MOO for forecasting cooling and heating energy. In this study, a synthetic data set was used to cover different cases, which includes 460,000 different records and contains different weather data of metropolises around the world. The model was reliable with an appropriate selection of values for various parameters. In addition, while reducing the time complexity of the system, it also achieved good accuracy.

The results of the research showed that the high potential of using Internet of Things devices and increasing the required information about buildings has made it possible to predict energy consumption in this field. In many types of research, machine learning and artificial intelligence techniques have been used to solve this problem, which is summarized in the table 1. In many of these methods, a data set that allows proper comparison and is available to everyone has not been used. So much of the information related to the data in the reviewed research is unclear. In this regard, in this research, a new data set is used that is accessible to the public and its full specifications are stated. In addition, different evaluation criteria have been used in the previous work, which has made the comparison in this field more difficult. For this reason, several common evaluation criteria for regression were used in this article to better examine the results.

Table 1
Comparison of related works

Evaluation	Data collection	Method	Research
MSE/ 1.0203 × 104	Household information in South Korea	ANN	Lee et al, 2019
NRMSE/ 0.1371	Real data from the Croatian public sector	Random forest-Boruta	Zekić-Sušac et al, 2021
RMSE/ 0.955	Tripura data in India	RF-XGBoost	Banik et al, 2021
RMSE/ 18.40	Real data of two campus buildings	GBDT- RF, GBDT, XGBoost, SVR, kNN	Wang et al, 2020
MAPE/ 5.31%	The city of Bloomington daily data	CEEMDAN-XGBoost	Lu et al, 2020
RMSE/ 28.787	Data of office building located in Henan Province, China	RDPG	Liu et al, 2020
MSE/ 0.39 e^{-2}	Data from the Sustainable Research Center in North Tehran	ANN-GA	Ilbeigi et al, 2020
MSE/ 22.2 %	Hotel data in Shanghai	SVM	Shao et al, 2020
R2/ 0.960	Real office building in England	FE-GA-DNN	Luo et al, 2020
Cooling RMSE/ 4.48±1.64 Heating RMSE/ 6.19±1.55	Data of metropolitan cities around the world	MOO	Seyedzadeh et al, 2020

3.Dataset and Data Preprocessing

To implement the proposed method in this research, the WiDS Datathon dataset will be used. This dataset was created in collaboration with the Climate Change Artificial Intelligence Institute and Lawrence Berkeley National Laboratory and is continuously being developed by various research and academic teams. WiDS Datathon seeks to find a suitable way to reduce the effects of climate change with a focus on energy efficiency. WiDS Datathon 2022 includes approximately 100,000 observations of building energy consumption records. In this data set, each record shows the energy consumption information of a building during a given year.

This information was collected over 6 years from several different states of the United States. The data set includes building characteristics (eg area and type of facility), geographic climate data (eg average annual temperature, total annual rainfall at the construction site), and energy consumption for the building and its year. By using the information related to the characteristics of the building and the climate data of the region that is in this data set, it is possible to predict the efficiency and energy consumption of each building (Kaggle, n.d).

In figure 1, the Pie chart shows the data for different years. In this chart, circular segments are used to show the relative size of the data recorded in different years. As can be seen in this figure, about half of the data was collected in the last two years.The features used in the WiDS Datathon and a brief explanation about each of them are given in table 2. The first 30 features of this table are known as input features. The last feature, which is the intensity of energy onsumption, is the target feature. Figure 2 shows the frequency distribution histogram for energy consumption intensity. In this figure, the horizontal axis represents the variable to be measured, i.e. the energy consumption intensity, while the vertical axis shows the number of frequencies (number of observations) of the variable in certain ranges of values. As can be seen in this figure, the highest frequency for the energy consumption intensity is in the range of 0 to 160 units.

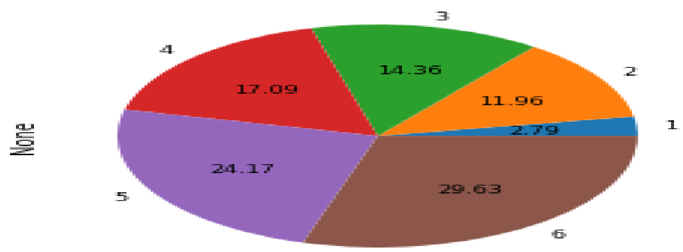

Fig 1: pie chart of data for different years

Table 2
WiDS Datathon Features

ID	Feature	Explanation
1	Id	Building id
2	Year_Factor	Anonymized year in which the weather and energy usage factors were observed
3	building_class	Building classification
4	State_Factor	Anonymized state in which the building is located
5	facility_type	Building usage type
6	floor_area	Floor area (in square feet) of the building
7	year_built	Year in which the building was constructed
8	energy_star_rating	The energy star rating of the building
9	ELEVATION	Elevation of the building location
10	january_min_temp	The minimum temperature in January (in Fahrenheit) at the location of the building
11	january_avg_temp	The average temperature in January (in Fahrenheit) at the location of the building
12	january_max_temp	Maximum temperature in January (in Fahrenheit) at the location of the building
13	cooling_degree_days	Cooling degree day for a given day is the number of degrees where the daily average temperature exceeds 65 degrees Fahrenheit. Each month is summed to produce an annual total at the location of the building.
14	heating_degree_days	The heating degree day for a given day is the number of degrees where the daily average temperature falls under 65 degrees Fahrenheit. Each month is summed to produce an annual total at the location of the building.
15	precipitation_inches	Annual precipitation in inches at the location of the building
16	snowfall_inches	Annual snowfall in inches at the location of the building
17	snowdepth_inches	Annual snow depth in inches at the location of the building
18	days_below_30F	Total number of days below 30 degrees Fahrenheit at the location of the building
19	avg_temp:	The average temperature over a year at the location of the building
20	days_below_20F	Total number of days below 20 degrees Fahrenheit at the location of the building
21	days_below_10F	Total number of days below 10 degrees Fahrenheit at the location of the building
22	days_below_0F	Total number of days below 0 degrees Fahrenheit at the location of the building
23	days_above_80F	Total number of days above 80 degrees Fahrenheit at the location of the building
24	days_above_90F	Total number of days above 90 degrees Fahrenheit at the location of the building
25	days_above_100F	Total number of days above 100 degrees Fahrenheit at the location of the building
26	days_above_110F	Total number of days above 110 degrees Fahrenheit at the location of the building
27	direction_max_wind_speed	Wind direction for maximum wind speed at the location of the building. Given in 360-degree compass point directions (e.g. 360 = north, 180 = south, etc.).
28	direction_peak_wind_speed	Wind direction for peak wind gust speed at the location of the building. Given in 360-degree compass point directions (e.g. 360 = north, 180 = south, etc.).
29	max_wind_speed	Maximum wind speed at the location of the building
30	days_with_fog	Number of days with fog at the location of the building
31	site_eui	Site Energy Usage Intensity is the amount of heat and electricity consumed by a building as reflected in utility bills

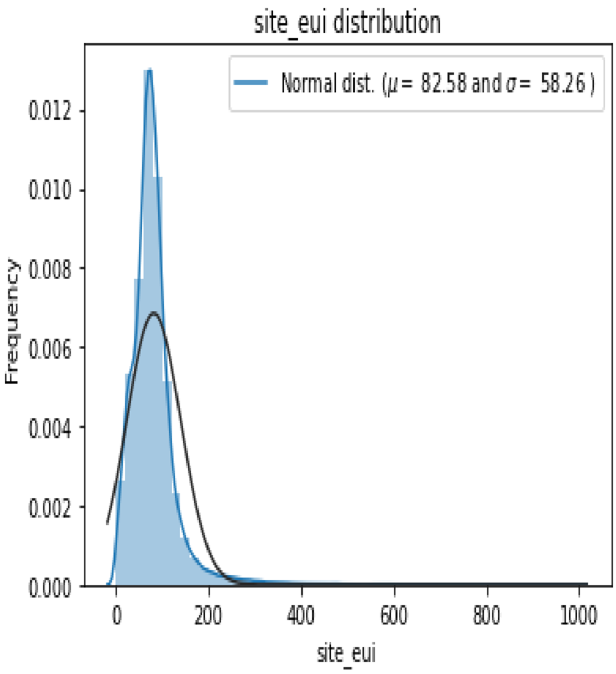

Fig 2: Frequency distribution histogram for energy consumption intensity

One of the main issues in the preprocessing of datasets is dealing with missing data. Table 3 shows the number of missing values and their percentage for several features. Considering that the values are a high proportion of the data, it is important to solve this problem. There are many simple methods to replace the missing values, but considering their use can introduce undue bias to the data set and have a bad effect on the final result. In this research to determine the appropriate alternative values for the missing values Imputer libraries are used. These libraries use learning-based methods to select the best alternative to missing values.

Table 3
The number of missing values in several attributes of WiDS Datathon

Feature	**Null**	**Null_percenatage**
year_built	1837	0.024249
energy_star_rating	26709	0.352561
direction_max_wind_speed	41082	0.542287
direction_peak_wind_speed	41811	0.551909
max_wind_speed	41082	0.542287
days_with_fog	45796	0.604512

4. Importance of Features

Depending on the model, the dataset may have too many features to handle. Many machine learning methods do not have a clear understanding of how they work. In many types of research, the final obtained accuracy is a significant amount, but at the end of the work, there is no significant understanding of the machine learning model and how it works on the data set. In this research, the importance of permutation is used to determine the importance of the features of the data set of the problem. This work is done by the Permutation Importance algorithm. First, the model is applied to the training dataset using xgboost and is trained. After that, the trained model is applied to the test data set and the rmse is calculated. This performance indicator calculates the average difference between the predicted values and the actual values.
After that, the column values of the data set whose it's feature importance should be calculated are randomly mixed (shuffled). Then the rmse is calculated for the test data set whose values of one column are shuffled. The new rmse is compared with the old rmse. An increase in the error indicates the importance of the investigated feature, and a decrease or no change indicates its lack of importance. This process is done for all the features (all columns) and the importance of each feature is obtained. Finally, the features are sorted according to their importance. In Table 8, 10 important features of the data set are shown in the order of their obtained weights.

Table 8
10 important features of the data set

Rank	**Feature**	**Weight**
1	energy_star_rating	0.2804 ± 0.0072
2	facility_type	0.1848 ± 0.0056
3	floor_area	0.0742 ± 0.0056
4	building_class	0.0418 ± 0.0060
5	State_Factor	0.0371 ± 0.0024
6	year_built	0.0127 ± 0.0015
7	snowdepth_inches	0.0044 ± 0.0005
8	january_avg_temp	0.0043 ± 0.0004
9	ELEVATION	0.0034 ± 0.0007
10	january_min_temp	0.0027 ± 0.0004

5.Proposed Stacking Model

Before presenting the proposed system, a set of machine learning methods including random forest, Xgboost, SVR and MLP were applied on a subset of features. The results of our investigations showed that XGBoost and MLP regression models had the best performance for solving the problem of predicting building energy consumption, respectively. Since both of these methods provided good performance on their subset of data, it was decided to use both of these methods to solve the problem. In the continuation of this research, the stacked method based on these methods is proposed to increase the efficiency of the system. The stacking technique is used to implement multiple regression models so that the final system performs better.

In the proposed model, at the first level, XGBoost and MLP regression models are run in parallel and independently from each other so that the advantages of both methods can be used. The results obtained from each of the basic methods in the first level are added as new features to the existing data set. In this way, the results of the base models are combined to learn a meta-learner at the second level. In the second level, an XGBoost is combined with the first level models sequentially. By applying the knowledge obtained from the previous level, the meta-learner tries to improve the learning ability and facilitates effective learning. Finally, the final results in the second level are built by XGBoost based on the results obtained from the first level. More details about setting the parameters of xgboost and MLP regression methods are given in Table 4.

To obtain more suitable results from the basic methods used in the system based on stacking, the meta-parameters of the basic methods are optimized before using the Successive Halving Search.

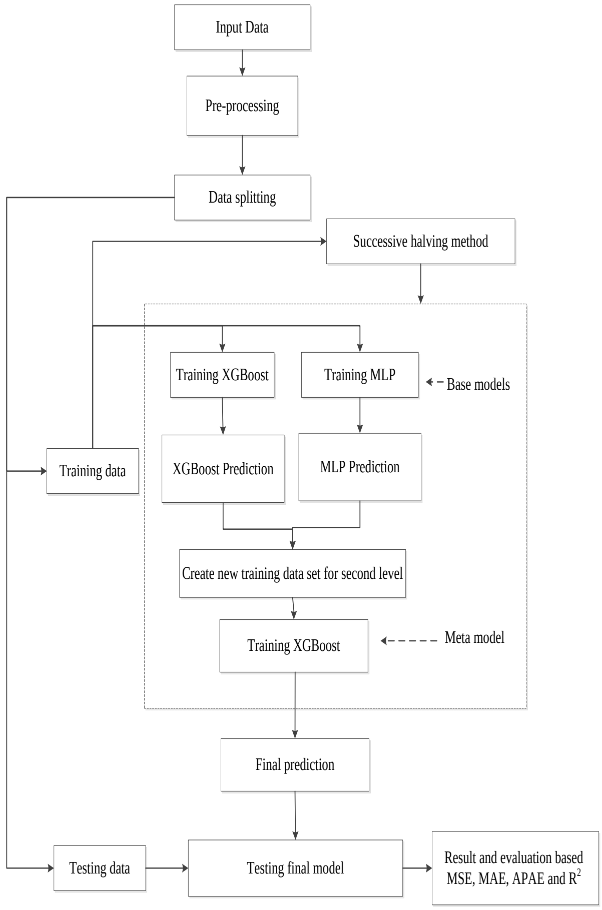

Fig 3:Flowchart of the Proposed Stacking System

Hyperparameters are parameters whose values are set before starting the learning process. Correct setting of hyperparameters is one of the most important tasks to implement any model and obtain better results. Considering that the setting of hyperparameters is different according to the type of problem and data set for the same techniques, this issue is a challenge at the front of the problem. In this research, to obtain the best results from the proposed model, the parameters of its basic techniques are performed using the successive halving method.

This method uses successive halving to search the parameter space (it starts with a large population of available options for hyperparameters). This method is a population-based algorithm. The search strategy in the first iteration starts to evaluate all candidates (combination of parameters) with a small number of resources (number of training samples). A limited number of the best candidates are selected for the next iteration and more resources are allocated to them (Kumar et al, 2018).

Table 4
Setting the parameters of xgboost and MLP regression methods

Method	Parameters	Value	Description
Xgboost.xgbregressor	n_estimators	100	number of boosting stages to perform
	learning_rate	0.02	shrinks the contribution of each tree by defined value
	gamma	0	Complexity control factor
	subsample	0.75	The fraction of samples to be used for fitting the individual base learners.
	colsample	0.4	Subsample ratio of columns during construction
	max_depth	5	Maximum depth of the individual regression estimators
Mlp regressor	random_state	1	Determines random number generation for weights and bias initialization
	max_iter	100	Maximum number of iterations.
	activation	relu	Name of the output activation function
	solver	adam	The solver for weight optimization, 'adam' refers to a stochastic gradient-based optimizer
	hidden_layer_size	100	Number of neurons in the hidden layer.

6. Experimental Results

Regression analysis is one of the fundamental parts of supervised machine learning. The difference between classification and regression problems is in the values that the target variable can have. There is no general agreement on a single standard criterion for regression analysis.
In the continuation of this section, first, the evaluation criteria used in this research will be explained. After that, the results of the implementation of the basic methods and the proposed method are described and at the end a comparison is made.

6.1.Regression Evaluation Criteria

The most widely used measures that are used in many studies are mean squared error, mean absolute error, mean absolute percentage error, and coefficient of determination.

Mean Squared Error (MSE): The MSE criterion is one of the most widely used criteria, which is used both in model training and model comparing.

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

The behavior of this criterion is proportional to the error of each data. The best possible value for it is 0 and the worst value is +∞. If there are outliers, MSE can be used to detect the existence of these data. The MSE measure enlarges the squared part of the error function if the model makes a very bad prediction.

Mean Absolute Error (MAE): The MAE criterion uses the absolute value function instead of the square and is calculated as follows:

$$MAE(Y,\hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i|$$

Unlike MSA, the MAE unit will be identical to the target feature. If there are outliers in the data set that we are sure of their correctness, the MAE criterion is appropriate. The MAE criterion does not overly penalize training outliers. The best value for this criterion is 0 and the worst value is ∞+.

Mean Absolute Percentage Error (MAPE): The MAPE measure is similar to the mean absolute error, but instead of the error, the relative error is used:

$$MAPE(Y,\hat{Y}) = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right|$$

In this way, a unitless measure will be obtained that is suitable for reporting results. This measure also has MSE problems. Another problem with MAPE is the possibility of values very close to 0 in the data. In this situation, the fraction of the expression will be very small and will have a strong impact on the criterion. This criterion makes it inappropriate for forecasting models where large errors are expected.

Coefficient of Determination (R^2): In this criterion, unlike the previous criteria, the high accuracy of the model is shown by increasing the value. It is calculated as follows:

$$R^2(Y,\hat{Y}) = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

This criterion is always a number smaller than 1. If a model always produces the average of the target feature in the output, the value of this criterion will be equal to 0. The coefficient of determination is usually expressed as a percentage. The coefficient of determination is only used to compare models and report results (Chicco et al, 2021).

6.2 Result of Base Method

For a better comparison and evaluation of the proposed system, several regression algorithms in the field of machine learning were used for prediction. These methods include linear regression, random forest, XGBoost, and MLP. Below is a brief definition of these methods and the results obtained from each of them.

Linear Reggression: Linear regression is one of the statistical methods used in the field of data science and machine learning. This algorithm determines the linear relationship between the independent variables and the dependent variable and is used for predictive modeling and analysis (Maulud & Abdulazeez, 2020). Due to its simplicity and appropriate training time and wide application, this method was used as the first method to solve the problem. In the rest of this section, the results of implementing linear regression on the WiDS Datathon are given.

Table 5
The results of applying linear regression on the dataset

Evaluation criteria	The obtained values
Mean Squared Error	2535.43
Mean Absolute Error	27.86
Mean Absolute Percentage Error	0.68
Coefficient of Determination	0.21

GBoost Reggressor: XGBoost is one of the most popular ensemble algorithms based on decision trees. This algorithm can be used to solve regression and classification problems (Bhattacharya et al, 2020). XGBoost is an optimized gradient tree-boosting system that was designed to increase the accuracy of regression and classification. XGBoost performs better compared to GBM, due to regular and parallel processing as well as providing additional meta-parameters. In table 6, the results obtained from applying the XGBOOST method to the data set are given. As can be seen in this table, the results obtained from XGBOOST are better than the linear regression model.

Table 6
The results of applying XGBoost on the dataset

Evaluation criteria	The obtained values
Mean Squared Error	2335.31
Mean Absolute Error	25.12
Mean Absolute Percentage Error	0.55
Coefficient of Determination	0.27

MLP Reggressor: MLP stands for Multilayer Perceptron, which represents a fully connected multilayer neural network. MLP consists of at least three layers including an input layer, a hidden layer, and an output layer (Savalia & Emamian, 2018). In table 7 the results obtained from applying the MLP technique to the research data set are shown. The result obtained in this section shows that the MLP technique performed better than the linear regression model, but was weaker compared to XGBOOST.

Table 7
The results of applying MLP on the dataset

Evaluation criteria	The obtained values
Mean Squared Error	2528.87
Mean Absolute Error	27.51
Mean Absolute Percentage Error	0.67
Coefficient of Determination	0.21

Random Forest Reggressor: Random forest is an ensemble learning method for classification and regression. Random forests perform relatively low training time by building a large number of decision trees that are relatively unrelated (Svetnik et al, 2003). The constructed parallel trees cover each other's errors and provide a good prediction. In table 7 the results obtained from applying the random forest regression technique on the research dataset are shown. The result obtained in this section shows that the random forest has obtained worse results than the other investigated techniques. The values obtained for the MSE, MAE, and APAE criteria were higher than the other investigated techniques and the R^2 value was lower.

Table 7
The results of applying Random Forest on the dataset

Evaluation criteria	The obtained values
Mean Squared Error	2674.97
Mean Absolute Error	28.88
Mean Absolute Percentage Error	0.76
Coefficient of Determination	0.16

6.3 Results of the proposed system

In this section, the results obtained from the proposed stacking system are compared with LR, XGBoost, MLP, and RF methods. Table 8 shows the results obtained from these methods on the data set. The graphs presented in figure 4 compare the performance of the base and proposed stacking methods based on the MSE, MAE, APAE, and R^2 evaluation criteria. As can be seen from the results of these graphs, among the basic methods, XGBoost and MLP regression techniques provided the best performance. These techniques are used as basic methods in the proposed stacking model. The graphs in the figure show that the proposed stacking model has obtained the lowest value in the MSE, MAE, and APAE criteria and the highest value in the R2 criterion.

Table 8
Comparison of the results obtained from LR, XGBoost, MLP, RF, and the proposed stacking system

	LR	XGBoost	MLP	RF	Proposed stacking method
Mean Squared Error	2535.43	2335.31	2528.87	2674.97	**2100.55**
Mean Absolute Error	27.86	25.12	27.51	28.88	**24.52**
Mean Absolute Percentage Error	0.68	0.55	0.67	0.76	**0.52**
Coefficient of Determination	0.21	0.27	0.21	0.16	**0.34**

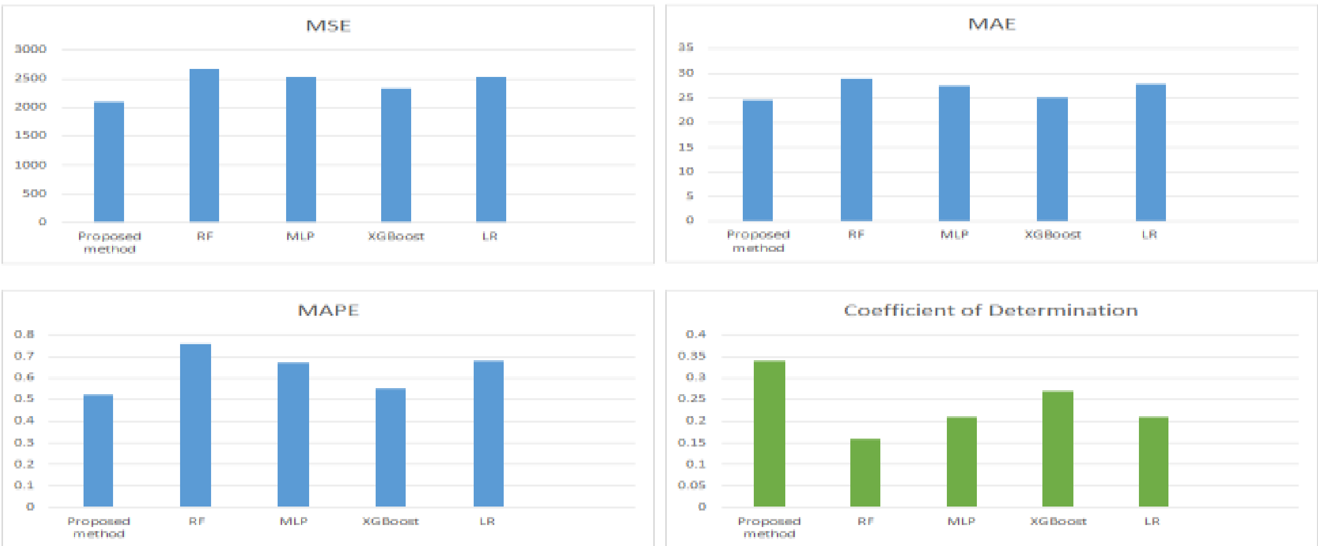

Fig 4: Performance comparison of LR, XGBoost, MLP, RF, and the proposed stacking system in different evaluation criteria

7. Conclusion

In this work, a number of machine learning regression models for predicting building energy consumption on WiDS Datathon were presented and evaluated. The effectiveness of the models for predicting the intensity of energy consumption was evaluated with criteria such as MAE, MSE, MAPE, and R2. Among the investigated models, the XGBoost method performed best with R^2 equal to 0.27. In this work, a number of machine learning regression models were presented and evaluated to predict building energy consumption on WiDS Datathon. The effectiveness of the models for predicting the intensity of energy consumption were evaluated with criteria such as MAE, MSE, MAPE and R2. Among the examined models, the XGBoost method achieved a value of 0.27 in the R2 criterion

and had the best performance. After the XGBoost method, the MLP method had the best performance.

After the investigations carried out in this research, a method based on stacking the best available methods was proposed to solve the problem of energy consumption prediction. Stacking method is a group learning technique that uses the predictions of several methods to build a new model. In this proposed system, which is a two-level system of stacking technique, XGBoost and MLP methods were used in the first level. The predictions obtained in this section were used in the next level, where XGBoost was also used as a meta-model. By running on the output of the first-level models, the meta-model tries to minimize the weaknesses and maximize the strengths of each individual model. In this research, in order to obtain the best performance from the used models, the best values for the available parameters have been selected using the Successive Halving Search method. The results of the implementation of this research showed that the proposed system of this research was able to perform well compared to the basic methods, so that the R2 criterion was improved to 0.07. The proposed system has performed better than the investigated methods in other investigated criteria.

Reference

[1] Sun, Y., Haghighat, F., Fung, B.C.M., 2020. A review of the state-of-the-art data-driven approaches for building energy prediction. Energy Build. 221, 110022.

[2] Wei, Y., et al., 2018. A review of data-driven approaches for prediction and classification of building energy consumption. Renew. Sustain. Energy Rev. 82 (October 2017), 1027–1047.

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

[4] WiDS Datathon 2022: Data. Retrieved from Kaggle: https://www.kaggle.com/competitions/widsdatathon2022

[5] Lee, Seunghui, Sungwon Jung, and Jaewook Lee. "Prediction model based on an artificial neural network for user-based building energy consumption in South Korea." Energies 12, no. 4 (2019): 608.

[6] Zekić-Sušac, Marijana, Adela Has, and Marinela Knežević. "Predicting energy cost of public buildings by artificial neural networks, CART, and random forest." Neurocomputing 439 (2021): 223-233.

[7] Banik, Rita, Priyanath Das, Srimanta Ray, and Ankur Biswas. "Prediction of electrical energy consumption based on machine learning technique." Electrical Engineering 103, no. 2 (2021): 909-920.

[8] Wang, Ran, Shilei Lu, and Wei Feng. "A novel improved model for building energy consumption prediction based on model integration." Applied Energy 262 (2020): 114561.

[9] Lu, Hongfang, Feifei Cheng, Xin Ma, and Gang Hu. "Short-term prediction of building energy consumption employing an improved extreme gradient boosting model: a case study of an intake tower." Energy 203 (2020): 117756.

[10] Liu, Tao, Zehan Tan, Chengliang Xu, Huanxin Chen, and Zhengfei Li. "Study on deep reinforcement learning techniques for building energy consumption forecasting." Energy and Buildings 208 (2020): 109675.

[11] Ilbeigi, Marjan, Mohammad Ghomeishi, and Ali Dehghanbanadaki. "Prediction and optimization of energy consumption in an office building using artificial neural network and a genetic

algorithm." Sustainable Cities and Society 61 (2020): 102325.

[12] Shao, Minglei, Xin Wang, Zhen Bu, Xiaobo Chen, and Yuqing Wang. "Prediction of energy consumption in hotel buildings via support vector machines." Sustainable Cities and Society 57 (2020): 102128.

[13] Luo, X. J., Lukumon O. Oyedele, Anuoluwapo O. Ajayi, Olugbenga O. Akinade, Hakeem A. Owolabi, and Ashraf Ahmed. "Feature extraction and genetic algorithm enhanced adaptive deep neural network for energy consumption prediction in buildings." Renewable and Sustainable Energy Reviews 131 (2020): 109980.

[14] Seyedzadeh S, Rahimian FP, Oliver S, Glesk I, Kumar B. Data-driven model improved by multi-objective optimization for prediction of building energy loads. Automation in Construction. 2020 Aug 1;116:103188.

[15] Chicco, Davide, Matthijs J. Warrens, and Giuseppe Jurman. "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation." PeerJ Computer Science 7 (2021): e623.

[16] Maulud, Dastan, and Adnan M. Abdulazeez. "A review on linear regression comprehensive in machine learning." Journal of Applied Science and Technology Trends 1, no. 4 (2020): 140-147.

[17] Bhattacharya, Sweta, Praveen Kumar Reddy Maddikunta, Rajesh Kaluri, Saurabh Singh, Thippa Reddy Gadekallu, Mamoun Alazab, and Usman Tariq. "A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU." Electronics 9, no. 2 (2020): 219.

[18] Savalia, Shalin, and Vahid Emamian. "Cardiac arrhythmia classification by multi-layer perceptron and convolution neural networks." Bioengineering 5, no. 2 (2018): 35.

[19] Svetnik, Vladimir, Andy Liaw, Christopher Tong, J. Christopher Culberson, Robert P. Sheridan, and Bradley P. Feuston. "Random forest: a classification and regression tool for compound classification and QSAR modeling." Journal of chemical information and computer sciences 43, no. 6 (2003): 1947-1958.

[20] Kumar, Manoj, George E. Dahl, Vijay Vasudevan, and Mohammad Norouzi. "Parallel architecture and hyperparameter search via successive halving and classification." arXiv preprint arXiv:1805.10255 (2018).