**Advances in Energy Research**, *Vol. 6, No. 2 (2019) 91-101*
DOI: http://doi.org/10.12989/eri.2019.6.2.091


# A gradient boosting regression based approach for energy consumption prediction in buildings

Ali S. Al Bataineh*

*Department of Electrical Engineering and Computer Science, University of Toledo, 2801 Bancroft Street, Toledo, OH 43606, U.S.A.*

*(Received January 20, 2019, Revised June 20, 2019, Accepted June 21, 2019)*

**Abstract.** This paper proposes an efficient data-driven approach to build models for predicting energy consumption in buildings. Data used in this research is collected by installing humidity and temperature sensors at different locations in a building. In addition to this, weather data from nearby weather station is also included in the dataset to study the impact of weather conditions on energy consumption. One of the main emphasize of this research is to make feature selection independent of domain knowledge. Therefore, to extract useful features from data, two different approaches are tested: one is feature selection through principal component analysis and second is relative importance-based feature selection in original domain. The regression model used in this research is gradient boosting regression and its optimal parameters are chosen through a two staged coarse-fine search approach. In order to evaluate the performance of model, different performance evaluation metrics like r2-score and root mean squared error are used. Results have shown that best performance is achieved, when relative importance-based feature selection is used with gradient boosting regressor. Results of proposed technique has also outperformed the results of support vector machines and neural network-based approaches tested on the same dataset.

**Keywords:** energy consumption; ensemble methods; gradient boosting regression

## 1. Introduction

Energy plays vital role in all aspect of daily life. It is an ultimate important resource, and directly linked with social and economic growth of any country. Considering the fact in recent years, there has been an increasing interest in optimization of energy which is achieved by energy consumption prediction (Pérez-Lombard *et al.* 2008). Recent researches attempt to understand and predict appliances energy use in buildings. Energy prediction refers to forecasting the demand. Research shows, a major portion of electrical energy at household is consumed by electrical appliances (Khosravani *et al.* 2016). Proper monitoring and consumption prediction of these appliances, can help in making a building smart and energy efficient. Thus, prediction can help in improving energy management system, building performance, building automation, load forecasting and recognition of patterns in electrical energy consumption (Candanedo *et al.* 2017). It has previously been observed that, the type, number and use of appliances in a building affects

*Corresponding author, Ph.D., E-mail: ali.albataineh@utoledo.edu


http://www.techno-press.org/?journal=eri&subpage=7 ISSN: 2287-6316 (Print), 2287-6324 (Online)

the electrical energy consumption. These factors are correlated and of immense significance (Candanedo *et al.* 2017).

In order to predict energy consumption, a mathematical model is required to model the energy consumption behaviour. Prediction models are classified into main categories: The first principal models and data driven models. First principal methods use physical properties to develop thermal dynamic equations of the system. First principle models to predict energy consumption require very detailed information of parameters like thermal and structural properties of buildings, environmental conditions, occupants living there and their activities and parameters of HVAC system installed in that building ((Panwar *et al.* 2011, Scarlat *et al.* 2015, Bajçinovci 2017, Tzikopoulos *et al.* 2005, hua *et al.* 2016, Foucquier *et al.* 2013). This detailed information is very difficult to obtain in most of the cases. In case this information is available, high level domain expertise is still needed to develop such complex models which are accurate enough and can be used for prediction.

On the other hand, data driven models do not require any detailed information regarding system's physical parameters and domain knowledge to develop the prediction models. These techniques use historical data of energy consumption and other correlated variables to develop a mathematical model. Data-driven techniques utilize the historical data representing the behaviour of the process and try to learn mapping between inputs and outputs. As energy prediction is a regression problem, therefore researchers have used different regression approaches to solve the problem of energy prediction.

There has been growing interest in using support vector machines (SVM) for building different regression models from last few years due to their ease of implementation and strong generalization capabilities. Researchers have used SVMs as a quick method for building generalized models for energy prediction (Dong *et al.* 2006, Jung *et al.* 2015). In order to capture nonlinear relationships through SVMs, nonlinear kernels like radial basis functions (RBF) are used. However, accuracy of models built using SVM is highly dependent on set of hyperparameters used (e.g. kernel type, kernel parameters etc.) to train the model.

Besides SVM, research interest in artificial neural networks has increased tremendously in last two decades. Researchers have used neural networks to solve problems in domain of both classification and regression. One of the strengths of neural networks is that they are very good in modeling complex nonlinear relationships. As energy prediction is a very nonlinear problem, therefore researches have used different variants of neural networks to build models for energy prediction (Neto and Fiorelli 2008, Ferreira *et al.* 2009, Li *et al.* 2011, Karatasou *et al.* 2006). It has generally been observed that performance of neural network is highly dependent on amount of data available for training and they are best suitable for cases where a lot of training data is available. Hence using neural network in cases where ample amount of training data is not available is not a good choice.

In addition to SVM and neural networks, there has been growing interest in using ensemble methods for developing complex machine learning models from last two decades. Researchers have used ensembles methods to solve problems in different domains (Karatasou *et al.* 2006, Elith *et al.* 2008, Zhu *et al.* 2009). These ensemble techniques use combination of different models to generate a single strong model. These ensemble techniques are generally robust and perform well even in cases where training data is not enough for techniques like neural networks (Keneni *et al.* 2019, Bataineh and Kaur 2018).

This research has used an ensemble method which is gradient boosting regression for building energy prediction models. In section 2, general ensemble and gradient boosting techniques are explained in detail. In section 3, dataset used for analysis is explained in detail. In section 4, methodology used for implementation and results are explained and section 5 contains the conclusion of the research.

## 2. Methodology

As explained earlier, an ensembled technique is used in this research to build an energy prediction model. Detailed explanation of adopted methodology is explained in this section.

### 2.1 Ensemble methods

In contrast to ordinary modelling approaches, ensemble methods use collection of estimators to train a model rather than using a single estimator. The main strength of ensemble methods is that they combine different weak estimators to generate a strong estimator. An ensemble made up off estimators of same type is known as homogeneous ensemble, whereas ensemble made up off estimators of different types is known as heterogeneous ensemble (Zhou 2012). Ensemble methods are further classified into Bagging and Boosting methods.

Bagging or bootstrap aggregation is an ensemble technique in which multiple parallel estimators are trained on different random sub samples of data taken from training set. These independent estimators are then combined using some averaging techniques (e.g., normal average, weighted average etc.). Boosting on the other hand is an ensemble technique in which different estimators are trained sequentially. The main intuition behind training multiple estimators sequentially is that each estimator learns from mistakes of subsequent estimator and hence multiple weak estimators are combined together to form a strong estimator. Difference between bagging and boosting methods is also depicted in Fig. 1. The technique used in this paper to predict energy usage of different appliances is Gradient Boosting Regression, which belongs to boosting class of ensemble methods.

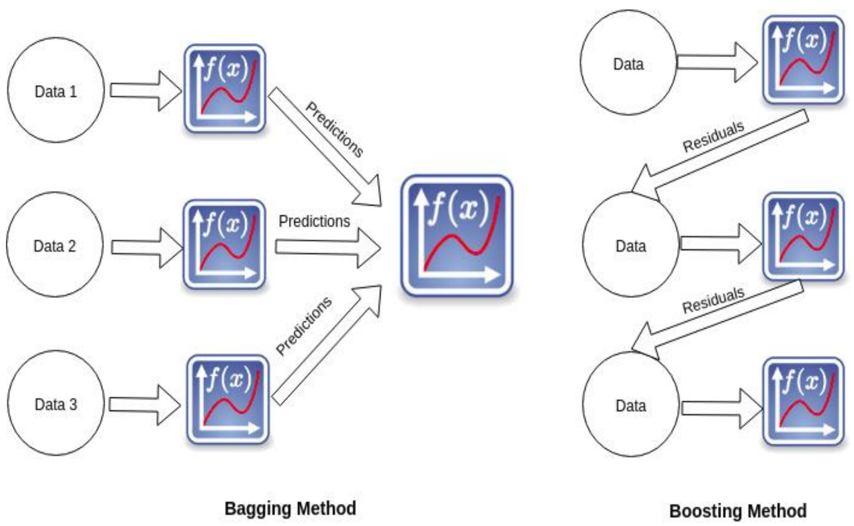


Fig. 1 Difference between bagging and boosting methods

### 2.2 Gradient boosting regression

Objective of any regression technique is to learn a mapping $f(x)$ through data points ($\boldsymbol{x\_i}$, $y\_i$) that best describes the given data. Where $\boldsymbol{x\_i}$ represents the input feature vector (can be scalar as well) and $y\_i$ represents the target value. Gradient boosting regression methods use combination of multiple weak models $f(x)$ in a stage wise manner to generate a strong model. Let's assume that $y_{hat}$ represents prediction made by a Gradient boosting regression model, then this can be represented by following mathematical expression

$$y_{hat} = \sum_{i=1}^{m} f_i(\boldsymbol{x}) \tag{1}$$

Here '$m$' represents the total number of models. Multiple estimators $f(x)$ used in gradient boosting regression technique can be of same as well as of different types, whereas we used same type of estimators to build our gradient boosting regression model. Base estimators used in our implementation are decision trees of same sizes. Reason behind using decision trees as our base estimator is that they have ability to model complex functions and also, they are good in handling data of mixed types (Keneni *et al.* 2019).

In training phase, objective of gradient boosting learning algorithm is to define a loss function and minimize it. Loss function which we used is 'mean squared error'. Mathematical expression of loss function is

$$L(y, y_{hat}) = \sum (y - y_{hat})^2 \tag{2}$$

$$L(y, y_{hat}) = \sum \left( y - \sum f_m(\boldsymbol{x}) \right)^2 \tag{3}$$

The task of gradient boosting learning algorithm is to choose that specific set of parameters (of the model) for which loss function or residual is close to zero. This is done by adding multiple weak estimators in stage wise fashion to improve overall performance of the model. Gradient boosting regression is a greedy kind of approach in a sense that choosing $f_m$ does not alters the previous estimators. Total number of estimators to be added in gradient boosting regression model is a tunable parameter which is decided based on criteria that, we stop adding weak estimators when adding these estimators does not further improve performance of the overall model. As in boosting methods, generally one estimator is added at a time, therefore we can express gradient boosting regression by equivalent recursive mathematical formulation as follows:

$$F_m = F_{m-1} + fm \tag{4}$$

Here $F_m$ represents the overall composite model. In addition to number of estimators, maximum depth of each fixed size decision tree is also a tunable parameter and determined through k-fold cross validation.

## 3. Case study

In this work we picked up a case study of prediction of appliances energy consumption in a house using data of other available variables (e.g., temperature, humidity, wind speed etc.). The data used in this research is collected by installing energy meters at different appliances in a home

Table 1 Data variables and their units

| Data Variables (Sensors) | Units |
|---|---|
| Energy consumption of appliances | Watt-hour |
| Energy consumption of lights | Watt-hour |
| Kitchen's Temperature | Centigrade |
| Living Room's Temperature | Centigrade |
| Office Room's Temperature | Centigrade |
| Bathroom's Temperature | Centigrade |
| Building's Outside Temperature | Centigrade |
| Ironing Room's Temperature | Centigrade |
| Teenagers Room's Temperature | Centigrade |
| Parents Room's Temperature | Centigrade |
| Outside's Temperature (from weather station) | Centigrade |
| Kitchen's Humidity | Percentage |
| Living Room's Humidity | Percentage |
| Office Room's Humidity | Percentage |
| Bathroom's Humidity | Percentage |
| Building's Outside Humidity | Percentage |
| Ironing Room's Humidity | Percentage |
| Teenagers Room's Humidity | Percentage |
| Parents Room's Humidity | Percentage |
| Outside's Humidity (from weather station) | Percentage |
| Windspeed (from weather station) | m/s |
| Tdewpoint (from weather station) | Centigrade |
| Visibility (from weather station) | Kilometer |
| Pressure (from weather station) | mm Hg |

located in Stambruges, Belgium (Candanedo *et al.* 2017). Energy meters used are M-Bus energy counters, whose readings are recorded at sampling rate of 10 minutes. The appliances on which these energy meters are installed are located in different house zones which are laundry, dining, kitchen and garage etc. These reading of energy meters are recorded for a period of almost five months starting from January 1st, 2016 till May 22nd, 2016. Along with energy data, readings of humidity and temperature sensors are also recorded at same sampling rate. Including temperature and humidity data will help in identifying the role of these parameters in predicting energy consumption of different devices. In addition to these variables, weather data from nearby airport weather station is also included in the dataset to study the impact of weather conditions on energy consumption prediction. This weather data includes readings of temperature, pressure, humidity, windspeed, visibility and Tdewpoint recorded at Chievres weather station. The sampling rate of weather data collection is generally one hour which is then interpolated linearly to make its sampling rate uniform (10 minutes) with sampling rate of other sensors. List of all data variables is presented in Table 1.

Task of this research is to use appliances energy consumption as target variable and build data driven model by treating all other variables as input/independent features.

## 4. Our implementation

One of the most important tasks in building good data-driven models is to select important set of features which contribute significantly in predicting target variable. In most of the cases, knowledge of specific problem domain plays a very important role in feature selection. Domain knowledge-based feature selection contribute significantly in increasing performance of the trained model in most of the cases, but the drawback of this technique is that it makes performance of machine learning models dependent on human input which is not the ideal case. Researchers have explored different techniques to solve problem of efficient selection of feature set without using domain knowledge. There are some statistical tests, which can be used to compute relative importance of different features with respect to target variable. Other than these statistical tests, there are some techniques which involve transformation of features in some other domain e.g., principal component analysis (PCA). PCA projects original feature set on to orthogonal principal components computed based on variance of data.

In this research, we have explored relative importance-based feature selection in both PCA as well as in original (raw feature) domain. This workflow is shown in Fig. 2. Available data is splitted into training and test sets, where 75% of data is used for training and 25% of the data is used for testing purpose. In PCA based implementation, raw features are projected onto principal components. These principal components are arranged in such a way that first principal component captures the highest variance of the data and last principal component captures the least variance. In the other approach, relative importance of features is computed in original domain based on their respective contribution (or correlation) to target variable. There are different statistical techniques available in literature which can be used to compute feature importance. We have used information gain to compute relative feature importance and then these features are arranged in descending order based on their computed importance index.

After computation of relative feature importance, next task is to train gradient boosting regression model with optimal set of features and parameters. There are two independent models which are to be trained, one for each type of features (i.e., PCA based and original domain based). For each type of feature set, performance of model is evaluated on different combinations of features (based on their importance) and parameters and best combination is chosen for final

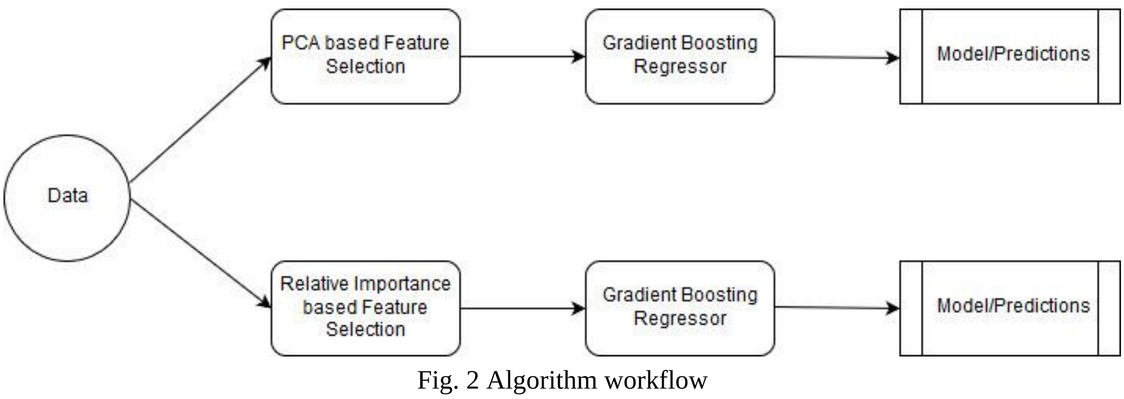


Fig. 2 Algorithm workflow

training. Feature selection is done in such a way that we keep on adding new features (arranged in descending order) in to the model up to the point where model score is maximum.

For each model, optimal set of features and parameters are computed using hyperparameter tuning. Technique adopted for hyperparameter tuning is k-fold cross validation. Model parameters which are to be tuned for gradient boosting regressor include depth of decision trees (base estimator), learning rate and total number of estimators. Performance measure used for cross validation is R2-score. For optimal hyperparameter search, a two-staged search approach is used. In first stage, hyperparameter search is conducted on a coarse grid and based on results of first stage, a fine grid is constructed in region where r2-score is high. In second stage, fine search is conducted on grid constructed based on results of first stage and results are used as final set of parameters and features for training the gradient boosting model.

### *4.1 Results*

Once model is trained, next task is to evaluate the performance of trained model on test set. As energy prediction is a regression problem, therefore performance evaluation metrics used are root mean squared error (RMSE), mean absolute error (MAE) and r2-score. Mathematical expressions these metrics are as follows

$$r2-score = 1-\frac{\sum(y_i - y_{pred_i})^2}{\sum(y_i - y_{mean})^2} \tag{5}$$

$$RMSE = \sqrt{\frac{\sum(y_i - y_{pred_i})^2}{n}} \tag{6}$$

$$MAE = \frac{\sum|y_i - y_{pred_i}|}{n} \tag{7}$$

where $y_i$ is actual value if $i_{th}$ instance of target variable in test set, $y_{predi}$ is predicted value if $i_{th}$ instance of target variable in test set, $y_{mean}$ is the mean value of target variable in test set

Test set is chosen through random sampling from whole data set to remove any bias. Results have shown that using respective importance-based feature selection through hyperparameter tuning with gradient boost regressor gives the best performance with r2-score of 0.62 on test set and 0.99 on training set. Ten most important features, computed by gradient boosting algorithm are shown in Fig. 3. Performance of PCA based feature selection with gradient boosting regressor

Table 2 Model performance

| | Training | | | Test | | |
|---|---|---|---|---|---|---|
| | r2-score | RMSE | MAE | r2-score | RMSE | MAE |
| PCA + GBR | 0.99 | 1.76 | 1.35 | 0.51 | 65.68 | 31.65 |
| Importance based feature computation + GBR | 0.99 | 2.46 | 1.91 | 0.62 | 59.99 | 27.47 |
| SVM | 0.83 | 42.64 | 11.8 | 0.483 | 67.72 | 31.04 |
| MLP | 0.14 | 88.37 | 47.4 | 0.154 | 83.77 | 46.50 |

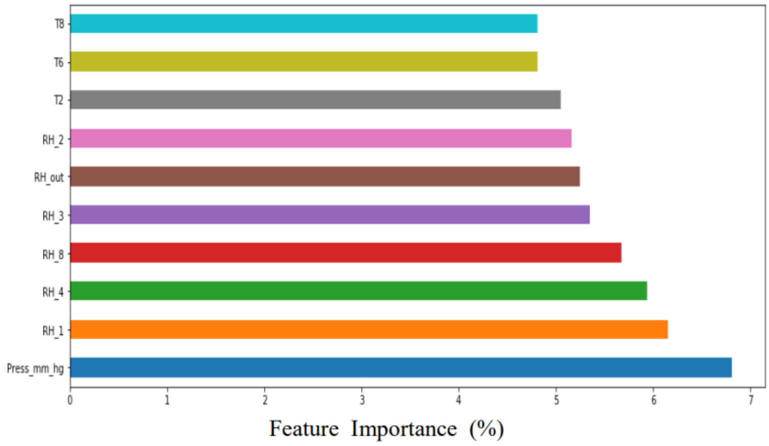


Fig. 3 Percentage importance of ten most important features

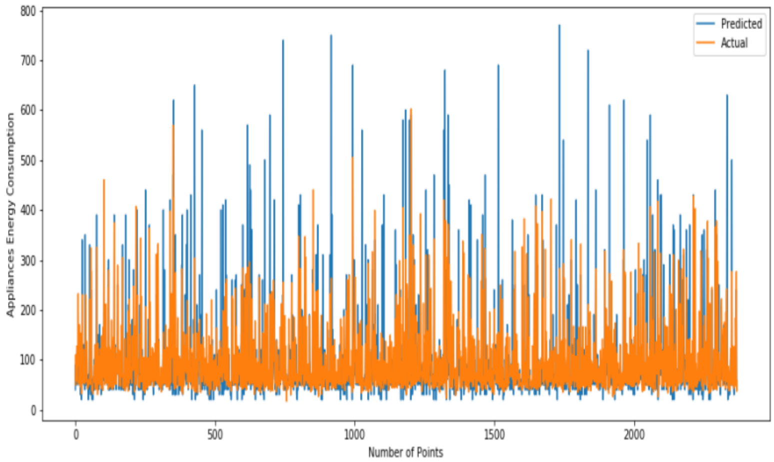


Fig. 4 Actual vs Predicted (PCA+GBR)

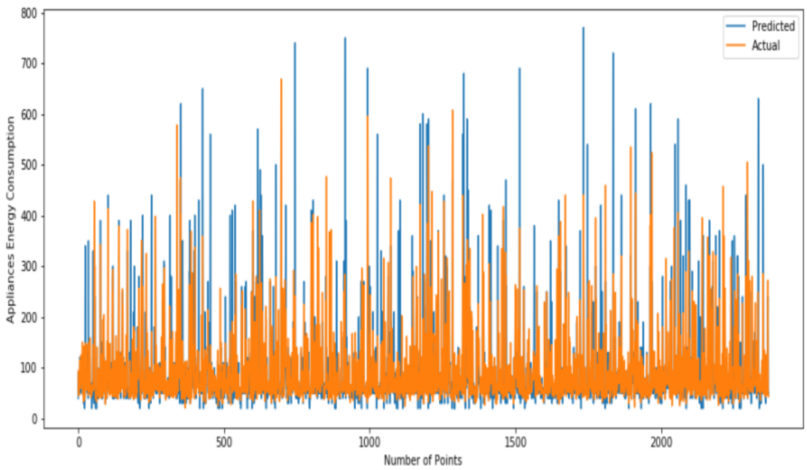


Fig. 5 Actual vs Predicted (Importance based Feature Computation + GBR)

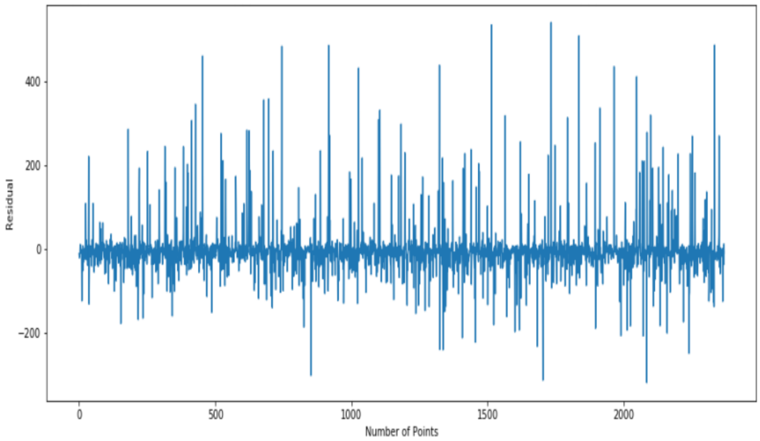


Fig. 6 Residual plot (PCA+GBR)

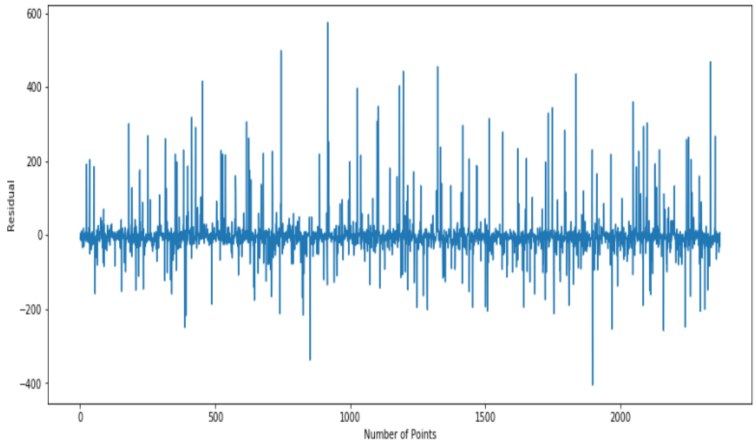


Fig. 7 Residual plot (Importance based Feature Computation + GBR)

was also quite impressive with r2_score of 0.59 on test set and 0.98 on training set. Results of different evaluation metrics for both above mentioned techniques and SVM, MLP are given in Table 2. Actual vs predicted plots for both feature selection techniques with gradient boost regressor are given in Figs. 4-5, whereas residual plots are given in Figs. 6-7.

## 5. Conclusions

Prediction of energy consumption in buildings is a very challenging task due to its dependence on large number of variables like environmental conditions, occupants living in that building and their activities etc. In this paper, we have proposed gradient boosting regression-based model for energy prediction. We have explored feature selection through PCA as well as by selecting features based on their relative importance in original domain. Results have shown that best performance is achieved, when relative importance-based feature selection is used with gradient boosting regressor. Another contribution of this research is that it has not used any sort of domain knowledge in feature selection and still outperformed the solutions which are highly dependent on domain knowledge for feature selection. To verify the results, different performance evaluation metrics are used like r2-score and root mean squared error etc. By using our proposed approach, we have achieved an r2-score of 0.62 and root mean squared error of 59.99 on test set, which clearly outperformed other techniques like SVMs and neural networks which are used to develop energy prediction models on the same data set.

## References

Bajçinovci, B. (2017), "Achieving thermal comfort and sustainable urban development in accordance with the principles of bioclimatic architecture: A case study of Ulcinj (Montenegro)", *Quaestiones Geographicae*, **40**(3), 131-140.

Bataineh, A.A. and Kaur, D. (2018), "A comparative study of different curve fitting algorithms in artificial neural network using housing dataset", *Proceedings of the NAECON2018-IEEE National Aerospace and Electronics Conference*, Ohio, U.S.A., July.

Candanedo, L.M., Feldheim, V. and Deramaix, D. (2017), "Data driven prediction models of energy use of appliances in alow-energy house", *Energy Build*., **140**, 81-97. https://doi.org/10.1016/j.enbuild.2017.01.083.

Dong, B., Cao, C. and Lee, S.E. (2006), "Applying support vector machines to predict building energy consumption in tropical region", *Energy Build*., **37**(5), 545-553. https://doi.org/10.1016/j.enbuild.2004.09.009.

Elith, J., Leathwick, J.R. and Hastie, T. (2008), "A working guide to boosted regression trees", *J. Animal Ecol*., **77**(4), 802-813. https://doi.org/10.1111/j.1365-2656.2008.01390.x.

Ferreira, P.M., Ruano, A.E., Pestana, R. and Kóczy, L.T. (2009), "Evolving RBF predictive models to forecast the Portuguese electricity consumption", *IFAC Proc. Vol*., **42**(19), 414-419. https://doi.org/10.3182/20090921-3-TR-3005.00073.

Foucquier, A., Robert, S., Suard, F. and Stephan, L. (2013), "State of the art in building modelling and energy performances prediction: A review", *Renew. Sust. Energy Rev*., **23**, 272-288. https://doi.org/10.1016/j.rser.2013.03.004.

Hua, Y., Oliphant, M. and Hu, E.J. (2016), "Development of renewable energy in Australia and China: A comparison of policies and status", *Renew. Energy*, **85**, 1044-1051. https://doi.org/10.1016/j.renene.2015.07.060.

Jung, H. C., Kim, J.S. and Heo, H. (2015), "Prediction of building energy consumption using an improved real coded genetic algorithm based least squares support vector machine approach", *Energy Build*., **90**, 76-84. https://doi.org/10.1016/j.enbuild.2014.12.029.

Karatasou, S., Santamouris, M. and Geros, V. (2006), "Modeling and predicting building's energy use with artificial neural networks: Methods and results", *Energy Build*., **38**(8), 949-958. https://doi.org/10.1016/j.enbuild.2005.11.005.

Keneni, B.M., Kaur, D., Bataineh, A.A., Devabhaktuni, V.K., Javaid, A.Y., Zaientz, J.D. and Marinier, R.P. (2019), "Evolving rule-based explainable artificial intelligence for unmanned aerial vehicles", *IEEE Access*, **7**, 17001-17016. https://doi.org/10.1109/ACCESS.2019.2893141.

Khosravani, H.R., Castilla, M.D., Berenguel, M., Ruano, A.E. and Ferreira, P.M. (2016), "A comparison of energy consumption prediction models based on neural networks of a bioclimatic building", *Energies*, **9**(1), 57. https://doi.org/10.3390/en9010057.

Li, K., Su, H. and Chu, J. (2011), "Forecasting building energy consumption using neural networks and hybrid neuro-fuzzy system: A comparative study", *Energy Build*., **43**(10), 2893-2899. https://doi.org/10.1016/j.enbuild.2011.07.010.

Neto, A.H. and Fiorelli, F.A. (2008), "Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption", *Energy Build*., **40**(12), 2169-2176. https://doi.org/10.1016/j.enbuild.2008.06.013.

Panwar, N.L., Kaushik, S.C. and Kothari, S. (2011), "Role of renewable energy sources in environmental protection: A review", *Renew*. *Sust*. *Energy Rev*., **15**(3), 1513-1524. https://doi.org/10.1016/j.rser.2010.11.037.

Pérez-Lombard, L., Ortiz, J. and Pout, C. (2008), "A review on buildings energy consumption information", *Energy Build*., **40**(3), 394-398. https://doi.org/10.1016/j.enbuild.2007.03.007.

Scarlat, N., Dallemand, J.F., Monforti-Ferrario, F., Banja, M. and Motola, V. (2015), "Renewable energy policy framework and bioenergy contribution in the European Union–An overview from National Renewable Energy Action Plans and Progress Reports", *Renew*. *Sust*. *Energy Rev*., **51**, 969-985. https://doi.org/10.1016/j.rser.2015.06.062.

Tzikopoulos, A., Karatza, M.C. and Paravantis, J. (2005), "Modeling energy efficiency of bioclimatic buildings", *Energy Build*., **37**(5), 529-554. https://doi.org/10.1016/j.enbuild.2004.09.002.

Zhou, Z.H. (2012), *Ensemble Methods*.

Zhu, C., Chen, W., Zhu, Z.A., Wang, G., Wang, D. and Chen, Z. (2009), "A general magnitude-preserving boosting algorithm for search ranking", *Proceedings of the 18th ACM Conference on Information and Knowledge Management-CIKM 09*, Hong Kong, China, November.

*CC*