American Journal of Energy Engineering
2023; 11(2): 61-66
http://www.sciencepublishinggroup.com/j/ajee
doi: 10.11648/j.ajee.20231102.14
ISSN: 2329-1648 (Print); ISSN: 2329-163X (Online)

Predicting Building Energy Consumption Using Stacking-Based Ensemble Model

Chang Lu[1], Yining Jing[1], Xinxue Lin[2], Kun Li[1, *]

[1]School of Urban Design, Wuhan University, Wuhan, China
[2]School of Resource and Environmental Sciences, Wuhan University, Wuhan, China

Email address:
luchang@whu.edu.cn (Chang Lu), jingyining@whu.edu.cn (Yining Jing), xinxuelin@whu.edu.cn (Xinxue Lin), kunli@whu.edu.cn (Kun Li)
[*]Corresponding author

To cite this article:
Chang Lu, Yining Jing, Xinxue Lin, Kun Li. Predicting Building Energy Consumption Using Stacking-Based Ensemble Model. *American Journal of Energy Engineering*. Vol. 11, No. 2, 2023, pp. 61-66. doi: 10.11648/j.ajee.20231102.14

Received: February 18, 2023; **Accepted**: May 27, 2023; **Published**: May 28, 2023

Abstract: With the development of urbanization, the smart city is gradually receiving widespread attention. The construction industry accounts for nearly half of China's total carbon emissions, which is one of the main contributors to the greenhouse effect. Accurate prediction of building energy consumption is of vital significance for energy conservation and smart city development. However, the traditional statistical methods for building energy consumption prediction have some problems such as low fitting accuracy and inaccurate prediction results. Machine learning algorithms are developing rapidly, which have significantly improved the predictive accuracy of building energy consumption. In this study, a stacking-based ensemble learning model based on architectural features is proposed. The building energy consumption mainly consists of the heating load and cooling load, which are used as target variables to predict the building energy consumption. Firstly, the normalized preprocessing is performed on the raw data. Subsequently, the ensemble model is obtained by integrating the optimal base predictors using stacking-based ensemble learning method. In the experiments, the dataset from the building energy area is tested with three metrics to evaluate the performance of the proposed model in building energy consumption prediction. The experimental results show that the proposed ensemble model outperforms the base predictors in solving the building energy consumption prediction problem.

Keywords: Machine Learning, Prediction, Stacking, Ensemble Model, Building Energy Consumption

1. Introduction

The construction industry accounts for nearly half of China's total carbon emissions, which is one of the main contributors to the greenhouse effect. Accurate prediction of building energy consumption can effectively reduce carbon emissions in the construction industry and contribute to the construction of smart cities and the promotion of green buildings. The building energy consumption mainly consists of the heating load and cooling load. Therefore, it is important and urgent to study the prediction of building energy consumption.

As urbanisation continues to advance, the construction of smart cities is gradually gaining widespread attention [1]. Smart energy system, the focus of smart city, aims to the analysis and integration of data resources and information to avoid waste of resources. The core computing technologies for smart city construction include Internet of Things, big data, cloud computing, intelligent algorithms and so on [2, 3]. Many scholars have studied the key issues of building energy consumption in smart building energy systems, and have successfully solved the problems of energy consumption characteristic analysis, energy consumption influence factor analysis and energy coordination and optimization [4, 5]. At present, building energy consumption is further increased and data formats are more diversified. Therefore, the development of smart energy systems is facing a bottleneck and there is an urgent need for a more general and robust prediction method to monitor and support the system [6, 7].

However, the traditional statistical methods for building energy consumption prediction including simulation methods encounter some problems such as low fitting accuracy and inaccurate prediction results. Meanwhile, machine learning techniques have been increasingly mature and significant in discovering hidden information behind various complex data, which fortunately enables architects and governments to better predict building energy consumption. Machine learning methods are evolving at a rapid pace, which are becoming a powerful tool for predicting building energy consumption. Nowadays, some researchers have achieved acceptable results of building energy consumption prediction through machine learning methods, but prediction accuracy still needs to be improved.

In this study, heating load and cooling load are used as target variables to predict the building energy consumption. A stacking-based ensemble model based on architectural features is proposed to predict building energy consumption accurately. Firstly, the raw data is preprocessed with the normalization method, which aims to reduce the influence of the different orders of magnitude of input variables on the model performance. Subsequently, a stacking ensemble method is applied to integrate the optimal base predictors into an ensemble model for building energy consumption prediction. In the experiments, the dataset is tested with three metrics to evaluate the performance of the proposed model in building energy consumption prediction. The results of the experiment show that the proposed model outperforms other baseline models in the building energy consumption prediction.

The remainder of the study is organized as follows. Section 2 shows related work of data mining, machine learning techniques and building energy consumption prediction. Section 3 explores data preprocessing and the modeling method. Experimental results are analyzed in Section 4. In Section 5, conclusion of this study and future work are presented.

2. Related Work

2.1. Machine Learning

Machine learning techniques, the core of Artificial Intelligence, have been developed rapidly in recent years, which help people explore new information from massive data more efficiently. The machine learning methods have been applied to various fields, and have achieved better prediction results than traditional predicting methods.

Montazeri et al. [8] proposed a rule-based classification method for the prediction of survival from different types of Breast cancers, in which, Trees Random Forest (TRF) technique showed better results in comparison to benchmark machine learning techniques. Pai & Wang [9] employed four machine learning models-namely, least squares support vector regression (LSSVR), classification and regression tree (CART), general regression neural networks (GRNN), and backpropagation neural networks (BPNN), to forecast real estate prices. Phan & Dhar [10] investigated the advantages and limitations of various machine learning models in predicting burst pressure, and found that the selected machine learning models provide significant improvements in predicting burst pressure compared to the existing reference models. Kwon et al. [11] developed a machine learning-based model for predicting the production stage temperature of distillation process. Yan & Liu [12] proposed a stacking ensemble model to predict and analyze the student performance in academic competition, in which, a feature importance analysis was applied to identify important variables, and three machine learning algorithms, i.e., Random Forest (RF), Gradient Boosted Regression Tree (GBRT) and Multi-Layer Perceptron neural network (MLP) are integrated. Chen et al. [13] proposed a direction of stock selection strategy, by combining the AdaBoost lifting algorithm and Decision Tree to predict stock movements and select superior stocks. Park and Bae [14] used machine learning algorithms as a research methodology to develop a housing price prediction model.

2.2. Building Energy Consumption Prediction

The main research methods for building energy consumption prediction include building model simulations and data driven methods.

Common building energy simulation software includes TRNSYS, ECOTECT and Energy Plus [15-17]. However, simulation methods for predicting building energy consumption are inefficient. On the one hand, modelling a building by software is time consuming. What's more, it requires continuous improvement of the model based on operational results in order to make the results of prediction more accurate. On the other hand, it is difficult for the simulation models to provide accurate real-time control strategies in the model design phase. Therefore, traditional simulation software programs are not suitable for the building energy consumption prediction [18].

The data-driven methods include multivariate linear regression (MLR), support vector machines (SVM), auto-regression (AR) and so on. MLR models are highly non-linear and have relatively low prediction accuracy [7]. SVM models require massive data for model training, so the absence of variables has a significant influence on the effectiveness of the models. The input variables of the AR model are only historical data of building energy consumption, without considering other factors, making it difficult to meet the actual requirements for prediction accuracy. In recent years, with the rapid development of artificial intelligence, the Back Propagation (BP) neural network has been widely used in building energy consumption prediction with its outstanding characteristics of non-linear mapping, self-adaptation and fault tolerance [19]. However, BP neural networks have a strong dependence on slow convergence speed and training data, which hinders its practical application [7].

As seen in the above discussion, machine learning has a wide range of application in building energy consumption prediction, which has greatly improved the accuracy of predictive results. However, the few researches that focused on the use of machine learning methods to solve building energy consumption prediction used only a single machine learning model, whose prediction results are not robust enough. Therefore, this study aims to address the problem of predicting building energy consumption using a stacking-based ensemble learning model.

3. Methodology

3.1. Data Preprocessing

The dataset contains eight independent features and two target variables. The description to features and target variables are shown in Table 1, which may impact the building energy consumption. The eight features are used to predict the heating load and cooling load accurately, which represent the building energy consumption.

This dataset is available on Machine Learning Repository at the University of California–Irvine (UCI). The raw data should be preprocessed to improve the availability and standardization of dataset.

(1) Data normalization

In the raw building energy data, different features often have different magnitudes and units of magnitude, which may affect the predictive accuracy. Therefore, the raw data is preprocessed using the normalization method to reduce the impact of the different orders of magnitude of input variables on model performance.

(2) Feature analysis

Feature analysis can eliminate irrelevant or redundant features, thereby reducing the number of features, improving model accuracy and reducing runtime. Therefore, feature analysis is performed on the preprocessed data to reduce dataset dimensionality if necessary.

Feature analysis helps find the magnitude of correlation between among features and target variables of the dataset. For the independent features with very strong correlation, only one is selected as the main factor and the others are removed. For the features that have very weak correlation with the target variables, they are removed too.

Table 1. *Building feature & target variable.*

Features & target variables	Description
Relative compactness	the density of the building arrangement
Surface area	the sum of all the areas of the building, including walls, doors, Windows, roofs, etc
Wall area	the total area of all the walls of a building
Roof area	the total area of all the roofs of a building
Overall height	the height of a building including the height of the walls and the thickness of the roof
Orientation	the orientation of buildings, mainly related to the position of the sun
Glazing area	the sum of all the glass areas of the building
Glazing area distribution	the placement of glass in a building
Heating load	the heat supplied to the building by the heating system in order to maintain the required environment
Cooling load	the heat removed from the room by the air-conditioning system to maintain the required environment

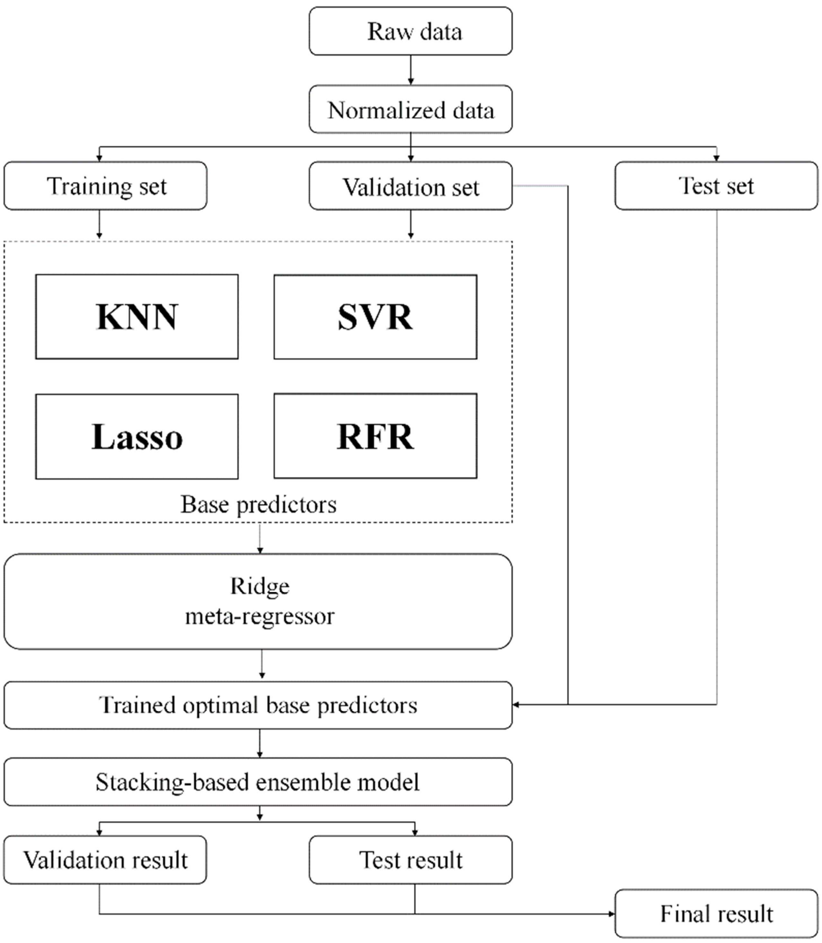

Figure 1. *Flow chart of the proposed stacking model.*

3.2. Modelling

(1) Splitting the dataset

The pre-processed data was divided into training, validation and test datasets. Training dataset is used to construct the machine learning model. Validation dataset is used to evaluate the model during the model construction process and thus adjust the model hyperparameters. Test dataset is used to evaluate the performance of the trained final model.

(2) The proposed stacking-based ensemble learning model

The flow chart of the proposed ensemble model is shown in Figure 1, and its details are shown follows.

Ensemble learning is a machine learning algorithm in which multiple learners are trained to solve the problem together. Unlike individual machine learning methods which try to learn a hypothesis from the training data, ensemble methods try to construct a set of hypotheses and combine them. Ensemble methods include bagging, boosting, stacking and blending, and can make the predictive results more accurate.

In this study, the stacking ensemble method is employed to integrate the four base predictive models, including K-Nearest Neighbor (KNN), Support Vector Regression (SVR), Least Absolute Shrinkage and Selection Operator (Lasso) and Random Forest Regression (RFR). The predictions from the four base predictors are superimposed by the ensemble method and provided as input data to the meta-regressor (e.g., ridge regression algorithm). The entire dataset is fitted and the best predictions are obtained after training.

4. Experiment

This section introduces the statistical metrics for evaluating the performance of all models and analyzes their performance in predicting the building energy consumption. All models and methods were implemented with Python programming language.

4.1. Metrics of Model Performance

In order to evaluate the performance of the model, three credible statistical metrics are adopted, including mean square error (MSE), mean absolute error (MAE) and coefficient of determination (R^2). MAE and MSE can measure the absolute magnitude of the deviation of the true value from the predicted value. R^2 can measure the applicability of linear regression. The calculation process of these three metrics is defined as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{x}_i - x_i\right)^2 \quad (1)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{x}_i - x_i\right| \quad (2)$$

$$R^2 = \left[1 - \frac{\sum_{i=1}^{N}\left(\hat{x}_i - x_i\right)^2}{\sum_{i=1}^{N}\left(\overline{x}_i - x_i\right)^2}\right] \quad (3)$$

Where $x_i \left(1 \le i \le N\right)$ and $\hat{x}_i \left(1 \le i \le N\right)$ denote the test and predicted values in the house rent, respectively, N denotes the size of the test sample, $\overline{x}_i$ denotes the average building energy consumption of N test samples. As for MSE and MAE, the lower value indicates the better prediction performance. For R^2, the higher value indicates the better the prediction performance.

4.2. Experimental Results Analysis

Table 2 shows the evaluation results of baseline models and ensemble model for predicting both heating load and cooling load. Among the base predictive models, RFR performs best in all metrics. It was the found that the MSE and MAE values of the ensemble model are the lowest and the R^2 value of the ensemble model is the highest among all models. Therefore, it can be concluded that the stacking-based ensemble model has the best performance with the highest prediction accuracy.

Table 2. *Performance of baseline models and ensemble model.*

Model	Heating load			Cooling load		
	MSE	MAE	R^2	MSE	MAE	R^2
KNN	8.2804	1.1787	0.9192	10.1265	1.3532	0.8886
RFR	0.3006	0.2533	0.9970	3.6683	0.5476	0.9597
SVR	34.9288	3.2506	0.6591	31.7440	2.7884	0.6509
GBR	0.3585	0.3138	0.9775	3.0065	0.6705	0.9669
Ridge	8.6276	1.4368	0.9158	9.9932	1.5025	0.8901
Ensemble model	0.2956	0.2415	0.9971	3.6008	0.5474	0.9601

(1) The evaluation results of heating load

To present the evaluation results more intuitively, the evaluation results of four baseline models and ensemble model for predicting the heating load are depicted in column charts, as shown in Figure 2. The MSE of the best baseline model (RFR) is 0.3006, which is still higher than that of ensemble model (0.2956). The MAE of the best baseline model (RFR) is 0.2533, which is still higher than that of ensemble model (0.2415). The R^2 of the best baseline model (RFR) is 0.9970, which is still lower than that of ensemble model (0.9971). In conclusion, the stacking-based ensemble model has the best performance with the highest prediction accuracy for predicting the heating load.

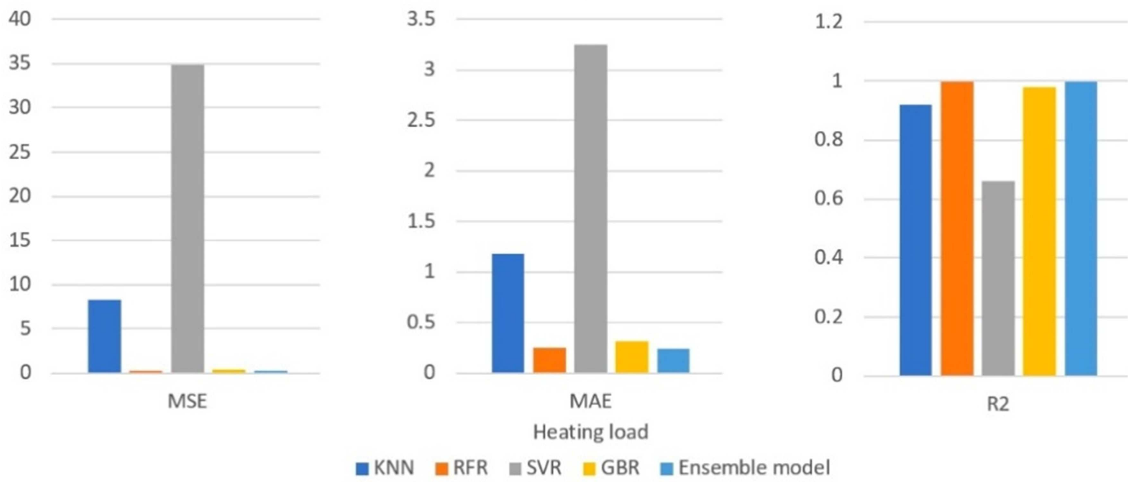

Figure 2. *The evaluation results of heating load.*

(2) The evaluation results of cooling load

Similarly, the evaluation results of four baseline models and ensemble model for predicting the cooling load are depicted in column charts, as shown in Figure 3. The MSE of the best baseline model (RFR) is 3.6683, which is still higher than that of ensemble model (3.6008). The MAE of the best baseline model (RFR) is 0.5476, which is still higher than that of ensemble model (0.5474). The R^2 of the best baseline model (RFR) is 0.9597, which is still lower than that of ensemble model (0.9601). In conclusion, the stacking-based ensemble model has the best performance with the highest prediction accuracy for predicting the cooling load.

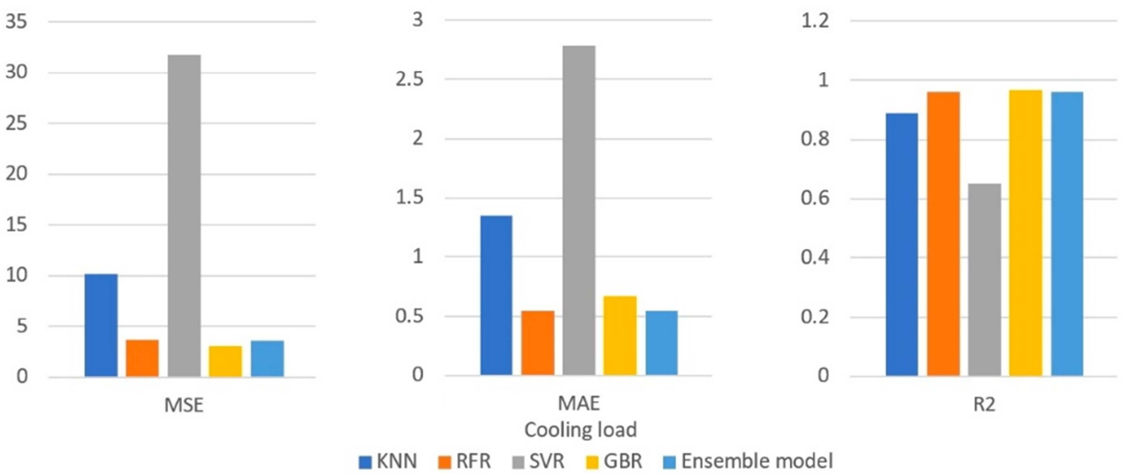

Figure 3. *The evaluation results of cooling load.*

Based on the experimental results above, it can be concluded that the stacking-based ensemble model is more promising than the baseline models.

5. Conclusion

The construction industry produces large amounts of carbon emissions, which is one of the main contributors to the greenhouse effect. Accurate prediction to building energy consumption can help reduce carbon emissions in the construction industry and contribute to the construction of smart cities and the promotion of green buildings. However, the traditional research methods for prediction of building energy consumption have some disadvantages such as low fitting accuracy and inaccurate prediction results.

On the other hand, machine learning algorithms are developing rapidly, which have significantly improved the predictive accuracy of building energy consumption. Therefore, it is of vital importance to study building energy consumption prediction using machine learning techniques.

In this study, a stacking-based ensemble model based on architectural features is proposed. Firstly, the raw building energy data is preprocessed using the normalization method to reduce the impact of the different orders of magnitude of input variables on model performance. Subsequently, the stacking-based ensemble learning method is employed to integrate the four base predictive models, including KNN, SVR, Lasso and RFR. The prediction results from the four baseline models are superimposed by the ensemble method and provided as input data to the meta-regressor (e.g., ridge regression algorithm in this study). The entire dataset is fitted and the optimal predictions are obtained after training.

In the experiments, three metrics are used to evaluate the performance of the proposed model for predicting the building energy consumption. Based on the experimental results, it is found that the stacking-based ensemble model outperformed the base predictive models in solving the building energy consumption prediction problem. This study provides policy makers and managers with powerful tools, with which, they can make informed decision based on the prediction results.

Nevertheless, there is still room for improvement in this study. To improve the prediction accuracy, some other advanced and extended algorithms including Bootstrap Aggregating and Boosting can be evaluated and integrated into the ensemble model in the future. In addition, the more evaluation metrics can be employed to give a more accurate and comprehensive evaluation of the model performance.

References

[1] Silva, B. N., Khan, M., & Han, K. (2018). Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. *Sustainable cities and society*, 38, 697-713.

[2] Visvizi, A., & Lytras, M. D. (2018). Rescaling and refocusing smart cities research: From mega cities to smart villages. *Journal of Science and Technology Policy Management*, 9 (2), 134-145.

[3] Chui, K. T., Lytras, M. D., & Visvizi, A. (2018). Energy sustainability in smart cities: Artificial intelligence, smart monitoring, and optimization of energy consumption. *Energies*, 11 (11), 2869.

[4] Kim, T., Kang, B., Kim, H., Park, C., & Hong, W. H. (2019). The study on the Energy Consumption of middle school facilities in Daegu, Korea. *Energy Reports*, 5, 993-1000.

[5] Xu, Z., Hu, G., & Spanos, C. J. (2017). Coordinated optimization of multiple buildings with a fair price mechanism for energy exchange. *Energy and Buildings*, 151, 132-145.

[6] Enayet, A., Razzaque, M. A., Hassan, M. M., Alamri, A., & Fortino, G. (2018). A mobility-aware optimal resource allocation architecture for big data task execution on mobile cloud in smart cities. *IEEE Communications Magazine, 56* (2), 110-117.

[7] Jing W., Guan H., Luo W., Li C., & Liu X. (2022). Energy consumption prediction model of office building based on improved GWO-BP. *Building Energy Efficiency, 50* (8), 125-129.

[8] Montazeri, M., Montazeri, M., Montazeri, M., & Beigzadeh, A. (2016). Machine learning models in breast cancer survival prediction. *Technology and Health Care*, 24 (1), 31-42.

[9] Pai, P. F., & Wang, W. C. (2020). Using machine learning models and actual transaction data for predicting real estate prices. *Applied Sciences,* 10 (17), 5832.

[10] Phan, H. C., & Dhar, A. S. (2021). Predicting pipeline burst pressures with machine learning models. *International Journal of Pressure Vessels and Piping*, 191, 104384.

[11] Kwon, H., Oh, K. C., Choi, Y., Chung, Y. G., & Kim, J. (2021). Development and application of machine learning-based prediction model for distillation column. *International Journal of Intelligent Systems*, 36 (5), 1970-1997.

[12] Yan, L., & Liu, Y. (2020). An ensemble prediction model for potential student recommendation using machine learning. *Symmetry*, 12 (5), 728.

[13] Chen, Y., Li, X., & Sun, W. (2020). Research on stock selection strategy based on AdaBoost algorithm. In *Proceedings of the 4th International Conference on Computer Science and Application Engineering*, Sanya, China, October 20-22, 2020, pp. 1-5.

[14] Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert systems with applications*, 42 (6), 2928-2934.

[15] Al-Saadi, S. N., & Zhai, Z. J. (2015). A new validated TRNSYS module for simulating latent heat storage walls. *Energy and Buildings*, 109, 274-290.

[16] Peng, C. (2016). Calculation of a building's life cycle carbon emissions based on Ecotect and building information modeling. *Journal of Cleaner Production*, 112, 453-465.

[17] Yu, S., Cui, Y., Xu, X., & Feng, G. (2015). Impact of civil envelope on energy consumption based on EnergyPlus. *Procedia Engineering*, 121, 1528-1534.

[18] Fumo, N., & Biswas, M. R. (2015). Regression analysis for prediction of residential energyconsumption. *Renewable and sustainable energy reviews*, 47, 332-343.

[19] Bourdeau, M., Zhai, X. Q., Nefzaoui, E., Guo, X., & Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society*, 48, 101533.