Building energy consumption prediction using deep learning

Razak Olu-Ajayi
Big Data Technologies and Innovation Laboratory, University of Hertfordshire, Hatfield, AL10 9AB, UK
r.olu-ajayi@herts.ac.uk

Hafiz Alaka
Big Data Technologies and Innovation Laboratory, University of Hertfordshire, Hatfield, AL10 9AB, UK
h.alaka@herts.ac.uk

Abstract: The consumption of energy in buildings has elicited the occurrence of many environmental problems such as air pollution. Building energy consumption prediction is fundamental for improved decision-making towards regulating or decreasing energy usage. There have been several applications of Machine Learning (ML) algorithms for predicting the energy consumption of operational buildings without much exploration into forecasting the potential building energy consumption at the early design stage. On the topic of reducing energy inefficient buildings, it is essential to address the root of the problem, the essentiality of predicting energy use before construction to alleviate futuristic problems of constructing new buildings that are harmful to the environment. At the early design stage, the customary model utilised for predicting energy consumption is the forward model, based on building energy modelling tools, which is stated to be mundane and time consuming. In contrast, the Machine Learning (ML) model is recognized as the most contemporary and best technique for prediction. To address this gap, this paper (1) presents the utilization of deep learning for predicting annual energy consumption of buildings, and (2) conduct a comparative analysis of the prediction performance of the models. The originality of this paper is to build a model trained by a dataset of multiple buildings that enables building designers to input key features of a building design and forecast the annual average energy consumption at the early stages of development. The ANN method outperforms SVM and DT for predicting annual energy consumption.

Keywords: Building energy consumption prediction, Deep learning, Data-driven model, Machine Learning.

1. INTRODUCTION

In recent years, the forecasting of building energy consumption has gained more attention by researchers and specialists (Amasyali and El-Gohary, 2017). This emanates from the report that the amount of energy consumed in buildings deploys a severe impact on the existence of mankind through the cause of major environmental problems such as climate change, air pollution, among others (Dandotiya, 2020). Energy inefficient buildings are recognised as the main contributors to global energy consumption and Green House Gas (GHG) emission (Pham et al., 2020; United Nations Environment Programme, 2017). According to the United Kingdom (UK) Building Energy Efficiency Survey (BEES), the energy consumed in buildings is accountable for 70 per cent of the total consumption (Building Energy Efficiency Survey, 2016). Therefore, the prediction of building energy use is essential to provide building owners and facility managers with the capacity to make informed decisions towards reducing energy consumption. However, accurate building energy forecasting remains a complex task due to certain factors that cannot be easily determined or obtained, such as occupant energy-use behaviour, physical properties of the building, among others (Hamed and Nada, 2019).

Building energy simulation tools, such as TRNSYS, DOE-2, and EnergyPlus are extensively utilized for forecasting energy consumption of operational buildings and buildings at the design stage. However, these tools are very detailed and elaborate, often requiring a large number of input parameters about the building and its environment (e.g., HVAC (Heating, Ventilation and Air Conditioning) system, Physical properties, internal occupancy loads, solar information and so on) which are not generally available to users (Runge and Zmeureanu, 2019). In many cases, the inability to provide the required input parameters leads to poor estimation performance (Amasyali and El-Gohary, 2017). In contrast, data driven models use Machine Learning (ML) algorithms such as Support Vector Machine (SVM), Decision Tree (DT), among others for energy use prediction. These models do not require a significant number of input parameters and make predictions based on historical data obtained from Building management systems and smart meters (Tardioli et al., 2015). However, in machine learning, the achievement of good prediction performance is predicated on these three factors: model selected, quantity and quality of the data (Runge and Zmeureanu, 2019).

The investigation of the applicability of a considerable number of machine learning algorithms for predicting energy consumption have been conducted in several studies (Aversa et al., 2016; Dong et al., 2005; Q. Li et al., 2009a; Pham et al., 2020; Tardioli et al., 2015). Among them, the most utilized for energy prediction are Decision Tree (DT), Artificial Neural Networks (ANN) and Support Vector Machine (SVM). These ML algorithms have been applied for predicting the energy consumption of operational buildings without much exploration into forecasting the potential energy consumption at the early design stage. Energy use prediction at the design stage is often implemented using building energy modelling tools, which is quantified to be resource demanding and time consuming (Zhu, 2006). Contrary to the utilization of ML models, which is recognized as the most contemporary and best technique for prediction (Canales, 2016; Vorobeychik and Wallrabenstein, 2013). To significantly decrease the construction of energy inefficient buildings, it is essential to optimize and identify the best model for energy consumption prediction during the design period. The development of a model with good prediction performance would enable building designers to check the energy consumption level at the early design stage, to detect if there is a need for design adjustment. In this study, the major ML algorithms namely DT, multi-layer perceptron ANN and SVM are applied to forecast the annual energy consumption of residential buildings. This paper focuses on developing a model that enables building designers to input parameters that are readily accessible at the early design stage to forecast the annual average energy consumption of the building before construction.

The remainder of the paper is structured as follows: Section 2 delivers a brief background on the most utilized supervised machine learning techniques for forecasting building energy consumption. Section 3 presents the research methodology, which consists of the description of the data collected, data pre-processing, model development and performance measures. Section 4 provides and examines the result, while Section 5 tenders the conclusions and future work.

2. LITERATURE REVIEW

Energy use predicting plays a vital role in energy conservation, financial cost reduction and enabling facility managers to make informed decisions towards reducing the energy consumed. In the application of ML algorithm for energy prediction, majority of the research aims to obtain the highest level of accuracy (García-Martín et al., 2019). A large variety of supervised machine learning algorithms have been utilized for energy prediction, although ANN, SVM and DT are recognised as the most utilized (Amasyali and El-Gohary, 2017). These algorithms comprise of both advantages and disadvantages in various situations. For instance, ANN and SVM often produce more accurate results than DT. However, DT techniques are elementary and easier to implement (Tso and Yau, 2007).

Artificial Neural Networks (ANN) is a non-linear computational algorithm that emulates the functional concepts of the human brain (Alaka et al., 2018; Amasyali and El-Gohary, 2018). A basic form of ANN consists of three consecutive layers: input, hidden and output layer. ANN is identified as a recurrent model and has gained more attention due to its good prediction performance. It is known to be dominant with big datasets, which enable the neural network sufficient data to train the model (Bourhnane et al., 2020). ANN algorithm is also known to produce good accuracy in energy load prediction (K. Li et al., 2018). Multi-layer Perceptron (MLP) is a function of a deep neural network that utilizes a feed forward propagation process with one hidden layer where latent and abstract features are learned (Donoghue and Roantree, 2015). In research by Khantach et al., Multi-layer Perceptron ANN produced the most accurate result between support vector machine (SVM), Gaussian process and radial basis function (RBF) with a Mean Absolute Percentage Error (MAPE) of 0.96 (Khantach et al., 2019). In 2019, Runge and Zmeureanu concluded that ANN produces good results when applied to single and multi-step ahead forecasting (Runge and Zmeureanu, 2019). Neto et al conducted a comparative analysis of the prediction performance of ANN and an energy simulation tool called EnergyPlus using a single dataset. The results revealed that data driven methods (ANN) are better suited for predicting building energy load (Aversa et al., 2016). The study by Bagnasco et al concludes that ANN performs better in the winter season when forecasting electrical consumption based on meteorological data and time/day variation (Bagnasco et al., 2015).

Dong et al 2005, first proposed the utilization of Support Vector Machine (SVM) for building energy use prediction and used four buildings dataset for forecasting monthly electricity consumption. Dong et al proclaimed that SVM outperforms other related research using neural networks with a coefficient of Determination R^2 higher than 0.99 (Dong, Cao and Lee, 2005). Similarly, Li et al explored the utilization of SVM for forecasting hourly cooling load in buildings and concluded that SVM produces a good result with a Root Mean Square Error (RMSE) of 1.17% (Li et al., 2009). On the other hand, Decision Tree (DT) does not outperform neural networks for non-linear data. However, its popularity can be attributed to its ease of use and ability to produce predictive models with interpretable structures. In Hong Kong, Tso and Yau conducted a comparative analysis of the prediction performance of decision tree, neural network and regression method in predicting weekly electricity consumption. Tso and Yau gathered that decision trees and neural networks perform slightly better than the regression method with a root of average squared error (RASE) of 39.36 (Tso and Yau, 2007).

3. RESEARCH METHODS

In this research, a prediction approach is explored based on supervised machine learning regression algorithms to forecast the annual energy consumption of residential buildings. The dataset is collected for this research is contains residential buildings data from various cities in the United Kingdom (UK) from January 2020 to December 2020. The prediction method will employ three machine learning algorithms which are multi-layer perceptron ANN, SVM and DT. The model development will be implemented in jupyter notebook utilizing the python programming language. Prior to training and testing the model, the raw data will first be analysed and cleaned or pre-processed to avoid any possible complexity, such as missing data during the training stage. Lastly, the performance of each model will be evaluated using performance metrics. Furthermore, the energy use prediction framework will consist of four sections: Data collection, Data pre-processing, Model development and Model evaluation.

3.1 Data collection

The building database was gathered from an online repository called Ministry of Housing Communities and Local Government (MHCLG) Repository. The project recorded the metadata of 300 real residential buildings around the UK. Each budling data encompasses the annual average energy consumption data. Additionally, the metadata includes several characteristics that are accessible at the design stage, such as floor area, wall type, roof type and so on. There are two types of data: meteorological data and building data. The meteorological data was obtained from an online repository called Meteostat repository, and it was recorded each instance which consist of temperature, wind speed and humidity. The monthly meteorological dataset was obtained from January 2020 to December 2020.

All variables utilized in this study are listed in Table 1. The building metadata is labelled as internal variables, while the meteorological data is labelled as external variables.

Table 1: List of variables selected

ID	Variable	Internal/External	Type	Unit	Label
1	Temperature	External	Continuous	^{o}c	Independent
2	Wind speed			km/h	
3	Pressure			Hg	
4	Total Floor Area	Internal		m^2	
5	Floor Description		Categorical	0, 1, 2, 3	
6	Walls Description			0, 1, 2, 3	
7	Windows Description			0, 1, 2, 3	
8	Number of Heated Rooms		Discrete	0, 1, 2, 3	
9	Number of Habitable Rooms			0, 1, 2, 3	
10	Energy Consumption		Continuous	kWh/m^2	Dependent

3.2 DATA PRE-PROCESSING

The preliminary procedure in machine learning involves data pre-processing for the preparation of data, and it often expends a significant amount of time and computational power (Shapi et al., 2021). This process is requisite as the dataset sometimes contains missing values and a few outliers. In this study, the data was pre-processed by employing the method of the imputation of missing data and normalization. The meteorological data had few missing values that were managed by utilizing the mean value imputation as presented in by Newgard and Lewis, 2015. In the raw data, the metadata consisted of specific variables that were descriptive, such as wall description, windows description and so on. The affected variables in the metadata were assigned values and filled using python library. The metadata was merged with meteorological annual average data using each location postcode as the primary key.

At this point, the data was normalized using the friedman of the sklearn python package. Data normalization is a very common technique of data pre-processing as various features have dissimilar dimensions, it is requisite to normalize different data to eradicate the influence of the dimension (Liu et al., 2020).

3.3 Model development (training)

This research utilized a supervised machine learning approach to forecasting the annual energy consumption of buildings. Subsequent to data preparation, it was fed into the learning algorithms. The data was then partitioned into two sets namely train set and test set at a ratio of

7:3. The predictive modelling approach utilized the regression model to forecast energy consumption. The three machine learning algorithms developed in this research are multi-layer perceptron ANN, SVM, and DT. The partitioned training set of 70% was used to train each machine learning algorithm to develop a predictive model that would estimate energy consumption based on the specified input. Consequently, the other 30% was utilized to test the trained prediction model.

3.4 Model evaluation (test)

To assess estimation performance of each model, the following performance measure were utilized: R-Squared (R2), Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), Among all stated measures, the most frequently employed are MSE and RMSE (Dong et al., 2005; Q. Li et al., 2009a; Pham et al., 2020)

1. Mean Absolute Error (MAE) is a method of computing the variation between the predicted values and true or actual values. MAE scores close to zero indicates better performance while a score greater than zero signifies a bad performance.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |AE_i - PE_i| \quad (1)$$

2. Mean Squared Error (MSE) is the evaluation of the squared variance between the estimated values and the actual values. MSE is the calculation of the quality of a prediction model. MSE score closer to zero have good performance.

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (AE_i - PE_i)^2 \quad (2)$$

3. Root Mean Squared Error (RMSE) is also a measure of calculating the variances between the estimated value and the true value. It is calculated by the square root of the Mean Square Error (MSE).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (AE_i - PE_i)^2} \quad (3)$$

4. R-Squared (R^2) is an evaluation method that determines the proportion of the variance in the target variable that can be justified by the independent variables. It exhibits the degree to which the data fits the model. R^2 can generate a negative result, but the best performance of R^2 is 1.0.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_{pred,i} - y_{data,i})^2}{\sum_{i=1}^{n} (y_{data,i} - y_{data})^2} \quad (4)$$

The range of experimental task conducted from data pre-processing to model evaluation was implemented using python. The computation was performed on a MacBook BigSur Os version 11.3 with M1 chip and 16gb RAM.

4. RESULT AND DISCUSSION

The prediction performance of the ANN and SVM models are comparable. However, DT evidently has the lower performance. The four performance metrics utilized to evaluate the performance were computed for the three models as presented in table 2 below. Based on the accuracy and computational efficiency, The result reveals which model best predicts annual energy consumption. Each model produced satisfactory results except the decision tree model. The ANN based model slightly outperforms the SVM based model with better MAE, RMSE, MSE and R^2, However, during training, the SVM based model consumed a shorter time than

ANN. Contrarily, the DT model performed worse than both ANN and SVM model but trained faster. The best models are presented in bold font in Table 2 below. This affirms ANN based model as a potentially good model to consider for forecasting annual energy consumption.

Table 2: Performance result for all model

Model	Training Time	R-squared	MAE	RMSE	MSE
Artificial Neural Network	1.2s	**0.66**	**2.20**	**2.80**	**7.85**
Support Vector Machines	1.0s	0.59	2.40	3.07	9.41
Decision Tree	**877ms**	0.27	3.17	4.10	16.83

Figures 1a-1c displays a visual representation of actual and predicted annual energy consumption for each test instance. This presents the difference between the actual and predicted values which corroborates ANN’s performance result as the better model. The performance of ANN corroborates the proclaimed theory in several studies that ANN is the better model for building energy prediction (Khantach et al., 2019; K. Li et al., 2018; Runge and Zmeureanu, 2019).

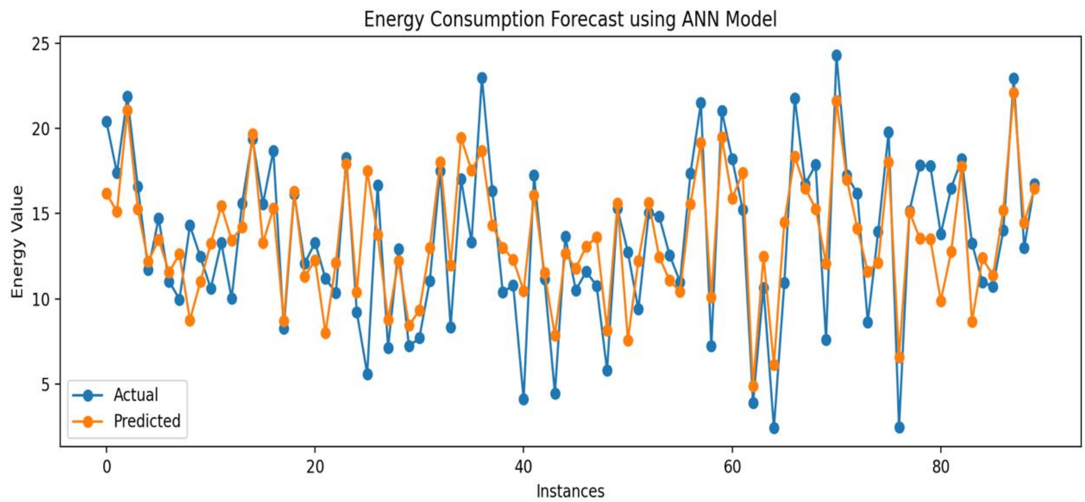

Figure 1a: Visualization of ANN based model performance

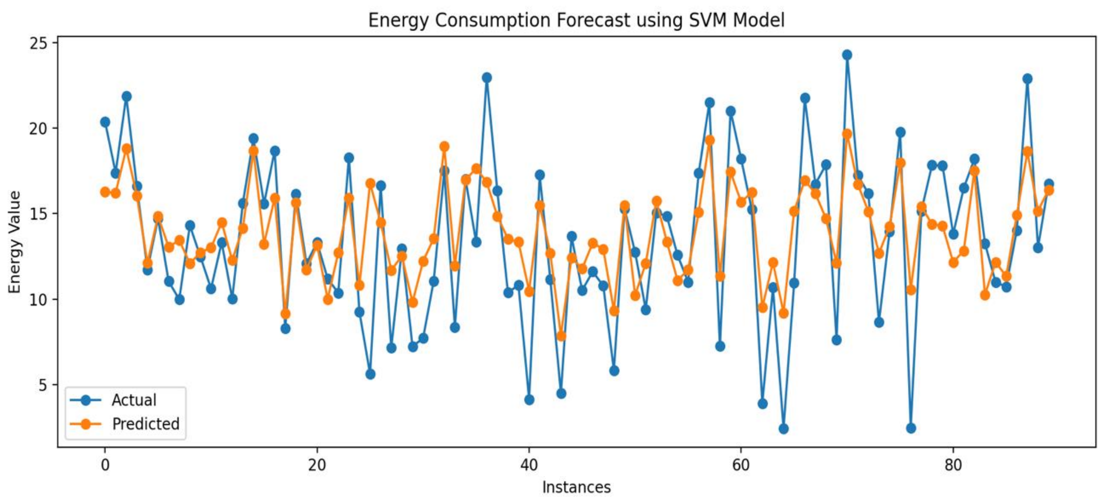

Figure 1b: Visualization of ANN based model performance

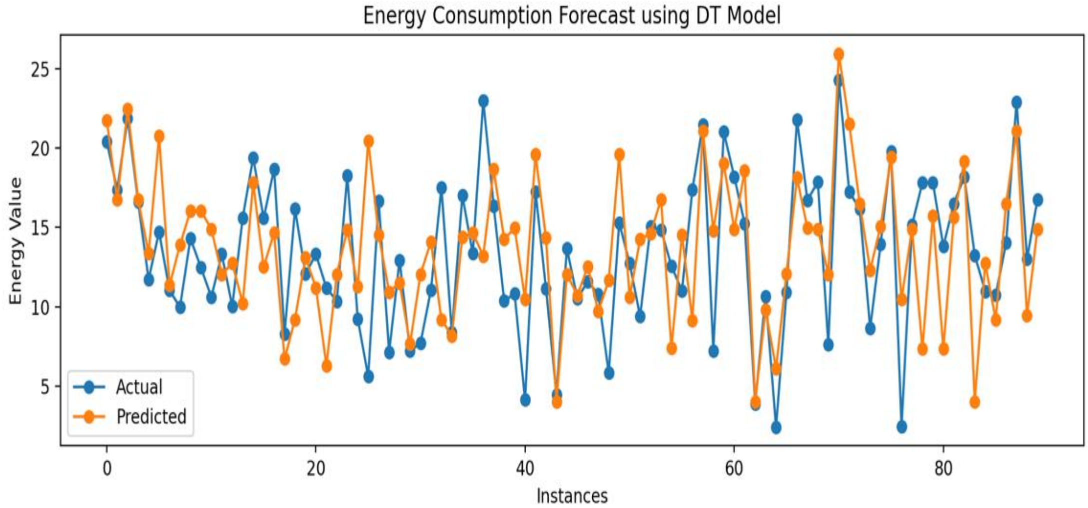

Figure 1c: Visualization of actual and predicted values of DT

The predictive performance of ANN states it as a better model than SVM in an application on the same data and same situation. This presents a motivation for utilization of ANN for energy prediction before construction which would reduce the development of energy in efficient buildings. In related works by Dong et al 2005, SVM performed better than ANN using a dataset of 4 buildings, as opposed the utilization of 300 buildings in this research. The performance of ANN can be attributed to its dominance in big datasets to enable the neural network sufficient data to train the model (Bourhnane et al., 2020).

5. CONCLUSIONS AND RECOMMENDATIONS

The course of the research focused on the annual energy prediction approach of residential buildings using the most utilised data driven based algorithms. The framework of this study affirms the possibility of a model that enable building designers to make informed decisions at the predesign stage of development. In this study, the models utilised, namely ANN, SVM and DT were trained, tested, and evaluated based on computational efficiency and accuracy. The performance result conveys multi-layer perceptron ANN has the best predictive model with an R^2 value of 0.66 and MAE value of 2.20. Similarly, SVM achieved good performance with an R^2 value of 0.59 and MAE value of 2.40. On the other hand, the DT model had the fastest training time of 877ms but produced the worst result for annual energy prediction. The

The performance result of Artificial Neural Network (ANN) shows promising potential for the estimation of annual energy consumption, especially on a larger dataset. Future research should focus on the application of ANN on a larger dataset and exploration into the suitability of other data driven regression models for annual energy use prediction.

6. REFERENCES

Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., and Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, *94*, 164–184. https://doi.org/10.1016/j.eswa.2017.10.040

Amasyali, K., and El-Gohary, N. (2017, May 31). *Deep Learning for Building Energy Consumption Prediction*.

Amasyali, K., and El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, *81*, 1192–1205. https://doi.org/10.1016/j.rser.2017.04.095

Aversa, P., Donatelli, A., Piccoli, G., and Luprano, V. A. M. (2016). Improved Thermal Transmittance Measurement with HFM Technique on Building Envelopes in the Mediterranean Area. *Selected Scientific Papers - Journal of Civil Engineering*, *11*. https://doi.org/10.1515/sspjce-2016-0017

Bagnasco, A., Fresi, F., Saviozzi, M., Silvestro, F., and Vinci, A. (2015). Electrical consumption forecasting in hospital facilities: An application case. *Energy and Buildings*, *103*, 261–270. https://doi.org/10.1016/j.enbuild.2015.05.056

Bourhnane, S., Abid, M. R., Lghoul, R., Zine-Dine, K., Elkamoun, N., and Benhaddou, D. (2020). Machine learning for energy consumption prediction and scheduling in smart buildings. *SN Applied Sciences*, *2*(2), 297. https://doi.org/10.1007/s42452-020-2024-9

Building Energy Efficiency Survey. (2016). *Building Energy Efficiency Survey (BEES)*. GOV.UK. https://www.gov.uk/government/collections/non-domestic-buildings-energy-use-project

Canales, V. R. (2016). *Using a Supervised Learning Model: Two-Class Boosted Decision Tree Algorithm for Income Prediction*. 7.

Dandotiya, B. (2020). *Climate-Change-and-Its-Impact-on-Terrestrial-Ecosystems*. https://doi.org/10.4018/978-1-7998-3343-7.ch007

Dong, B., Cao, C., and Lee, S. E. (2005). Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings*, *37*(5), 545–553. https://doi.org/10.1016/j.enbuild.2004.09.009

Donoghue, J. O., and Roantree, M. (2015). A Framework for Selecting Deep Learning Hyper-parameters. In S. Maneth (Ed.), *Data Science* (Vol. 9147, pp. 120–132). Springer International Publishing. https://doi.org/10.1007/978-3-319-20424-6_12

García-Martín, E., Rodrigues, C. F., Riley, G., and Grahn, H. (2019). Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing*, *134*, 75–88. https://doi.org/10.1016/j.jpdc.2019.07.007

Hamed, M., and Nada, S. (2019). Statistical Analysis for Economics of the Energy Development in North Zone of Cairo. *International Journal of Finance and Economics*, *5*, 140–160.

Khantach, A. E., Hamlich, M., Belbounaguia, N. eddine, Khantach, A. E., Hamlich, M., and Belbounaguia, N. eddine. (2019). Short-term load forecasting using machine learning and periodicity decomposition. *AIMS Energy*, *7*(3), 382–394. https://doi.org/10.3934/energy.2019.3.382

Li, K., Xie, X., Xue, W., Dai, X., Chen, X., and Yang, X. (2018). A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction. *Energy and Buildings*, *174*, 323–334. https://doi.org/10.1016/j.enbuild.2018.06.017

Li, Q., Meng, Q., Cai, J., Yoshino, H., and Mochida, A. (2009a). Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. *Energy Conversion and Management*, *50*(1), 90–96. https://doi.org/10.1016/j.enconman.2008.08.033

Li, Q., Meng, Q., Cai, J., Yoshino, H., and Mochida, A. (2009b). Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*, *86*(10), 2249–2256. https://doi.org/10.1016/j.apenergy.2008.11.035

Liu, Y., Chen, H., Zhang, L., Wu, X., and Wang, X. (2020). Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in China. *Journal of Cleaner Production*, *272*, 122542. https://doi.org/10.1016/j.jclepro.2020.122542

Newgard, C. D., and Lewis, R. J. (2015). Missing Data: How to Best Account for What Is Not Known. *JAMA*, *314*(9), 940–941. https://doi.org/10.1001/jama.2015.10516

Pham, A.-D., Ngo, N.-T., Ha Truong, T. T., Huynh, N.-T., and Truong, N.-S. (2020). Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. *Journal of Cleaner Production*, *260*, 121082. https://doi.org/10.1016/j.jclepro.2020.121082

Runge, J., and Zmeureanu, R. (2019). Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review. *Energies*, *12*(17), 3254. https://doi.org/10.3390/en12173254

Shapi, M. K. M., Ramli, N. A., and Awalin, L. J. (2021). Energy consumption prediction by using machine learning for smart building: Case study in Malaysia. *Developments in the Built Environment*, *5*, 100037. https://doi.org/10.1016/j.dibe.2020.100037

Tardioli, G., Kerrigan, R., Oates, M., O'Donnell, J., and Finn, D. (2015). Data Driven Approaches for Prediction of Building Energy Consumption at Urban Level. *Energy Procedia*, *78*, 3378–3383. https://doi.org/10.1016/j.egypro.2015.11.754

Tso, G. K. F., and Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, *32*(9), 1761–1768. https://doi.org/10.1016/j.energy.2006.11.010

United Nations Environment Programme, U. N. (2017, September 26). *Sustainable buildings*. UNEP - UN Environment Programme. http://www.unep.org/explore-topics/resource-efficiency/what-we-do/cities/sustainable-buildings

Vorobeychik, Y., and Wallrabenstein, J. R. (2013). *Using Machine Learning for Operational Decisions in Adversarial Environments*. 9.

Zhu, Y. (2006). Applying computer-based simulation to energy auditing: A case study. *Energy and Buildings*, *38*(5), 421–428. https://doi.org/10.1016/j.enbuild.2005.07.007