Proceedings of the 6th World Congress on Civil, Structural, and Environmental Engineering (CSEE'21)
~~Lisbon, Portugal~~ Virtual Conference – June 21 – 23, 2021
Paper No. ICSECT 104
DOI: 10.11159/icsect21.lx.104

On the Implementation of Machine Learning Models for Emulating Daily Electricity Consumption in Hotel Facilities

Eslam Mohammed Abdelkader[1], Abobakr Al-Sakkaf[2,3], Nehal Elshaboury[4] and Ghasan Alfalah[5]
[1]Structural Engineering Department, Cairo University
Giza, Egypt
[2]Department of Building, Civil, and Environmental Engineering, Concordia University
Montréal, QC, Canada
[3]Department of Architecture & Environmental Planning/College of Engineering & Petroleum/Hadhramout University
Mukalla, Yemen
[4]Construction and Project Management Research Institute, Housing and Building National Research Center
Giza, Egypt
[5]Deptartment of Architecture and Building Sciences, King Saud University
Riyadh, Saudi Arabia
eslam-ahmed1990@hotmail.com; abobakr.alsakkaf@concordia.ca; nehal_ahmed_2014@hotmail.com;
galfalah@ksu.edu.sa

Abstract **-** The increase in electricity consumption has negative effects on the environment, including greenhouse gas emissions, global warming, and rapid climate change. In an attempt to address this global problem, this research aims at developing machine learning models for the prediction of electricity consumption in hotels. The models are implemented using several techniques, including K-nearest neighbors, radial basis neural network, support vector machines, decision tree, and Gaussian process. The performance of the aforementioned models is evaluated by measuring the mean absolute error, mean absolute percentage error and root-mean squared error using split validation. Finally, two-tailed Student's t-tests are performed to evaluate the statistical significance level of prediction models. The results demonstrate that the K-nearest neighbors model accomplished the highest prediction performance yielding mean absolute error, mean absolute percentage error and root-mean squared error of 130.77, 276.45 and 4.79%, respectively. Support vector machines provided the lowest prediction accuracy accomplishing mean absolute error, mean absolute percentage error and root-mean squared error of 303.39, 382.78 and 11.98%, respectively.

Keywords**:** Electricity consumption; machine learning; K-nearest neighbors; radial basis neural network; two-tailed student's t-test

1. Introduction

The hospitality industry in Canada is generally characterized by high energy use, high cost of energy consumption (bills up to $1.5 billion), and significant environmental footprint (5 million tonnes of greenhouse gas emissions yearly) [1]. These numbers are expected to further escalate due to the rising interest in tourism (25 million travelers in 1950 to 1,035 million in 2012) [2]. As a result, there is a need to implement sustainability initiatives to help reduce the environmental impact of the industry. Such initiatives that focus on the reduction of energy consumption and cost would lead to a rise in revenue for the industry. Hence, the objective of this study is to first evaluate the overall challenges of the hospitality industry with respect to sustainability and then explore the incorporation of proposed initiatives into operating agreements in hotels. To accomplish this, an analysis is performed on the factors and/or activities that influence energy consumption in the industry as well as mitigation procedures to help reduce this consumption. The capacity of different hotel sizes in the industry is also examined to determine their ability to implement these initiatives. The current measures in these hotels to limit energy consumption will also be assessed, especially with respect to their presence in the hotel's operating agreement [3]. Findings will solidify the proposed solutions for hotels in their efforts to reduce energy consumption through the use of operating agreements.

Employment and Energy are the top two cost-incurring areas for hotels [4]. About $2000, on average, is spent on energy for a standard American hotel room [5]. Since the size of hotels differs, their energy consumption is expected to vary based on technical and management factors, amongst others. The implementation of initiatives to reduce energy consumption in hotels has not been so welcomed by the owners due to the additional 'cost' involved [6]. However, this 'cost' is offset by the increase in revenue that follows the implementation of the measures and for many of the initiatives, the major 'cost' investment is that of human involvement and attention [7]. Generally, the implementation of sustainability initiatives is relatively easier to accomplish from the onset of the design of a new building.

Energy consumption in hotels accounts for between 3% and 6% of the total running costs. The implementation of adequate sustainability initiatives would lead to a reduction of this percentage while ensuring comfort of clients [8]. As the major part of the used energy is produced by gas, coal, and petroleum products, reducing energy consumption would also contribute to decreasing greenhouse gas emissions [9-10]. In hotels, the main energy consuming systems include: heating, air conditioning and ventilation, hot water production, lighting, electricity (for mechanical installations such as lifts), and cooking. Cogeneration is proposed as a sustainability initiative that can be implemented, with pay-back period of about three years for medium-sized or large (not seasonal) hotels. The decision to proceed with a cogeneration project is basically an investment decision. Generally, investment decisions are influenced by both present and future costs. Like any other investment, cogeneration plants involve capital expenditure to gain additional assets. These assets are expected to provide a predetermined minimum attractive rate of return. According to Freeman et al., [11] both electricity and low temperature thermal energy are needed and small combined heat and power (CHP) units can be used for hotels (for cases where there is a continuous and significant heat demand). The potential benefits of cogeneration for hotels include reduced energy cost, more reliable power supply, and improved power supply quality. On a nationwide scale, the benefits of the implementation of cogeneration projects in hotels include fewer electricity shortages, primary fuel savings, reduced or deferred capital expenditures for power plant construction, enhanced efficiency of electric utility service, and better environmental conditions.

2. Literature Review

Several studies were previously carried out to evaluate electricity consumption in buildings. Khalil et al. [12] developed artificial neural network model for predicting heating and cooling loads in buildings based on energy performance dataset. The input variables encompassed relative compactness, roof area, overall height, surface area, glazing area, etc. It was found that the developed model was capable of achieving accuracy of 99.6%. Wang et al. [13] utilized random forest for hourly building energy consumption. The input variables included outdoor temperature, relative humidity, wind speed, solar radiation, number of occupants, workday type, etc. The developed model was validated through comparisons against regression trees and support vector machines. In this regard, it was concluded that the random forest improved the prediction accuracies by 25% and 5.5% with reference to regression tree and support vector regression, respectively.

Mohammed Abdelkader et al. [14] compared five types of machine learning models for the prediction of heating and cooling loads in buildings. The machine learning models comprised back-propagation artificial neural network, radial basis neural network, generalized regression neural network, ANOVA kernel support vector machines and radial kernel support vector machines. It was inferred that radial basis neural network had the highest prediction accuracies achieving mean absolute percentage error, mean absolute error and root-mean squared error of 1.016%, 0.5363 and 0.2133, respectively. It was also found that the radial basis neural network significantly outperformed other prediction models based on the conducted student two-tailed student's t-test. Marzouk and Mohammed Abdelkader [15] introduced a hybrid multi-objective genetic algorithm model for the purpose of selecting the most sustainable materials in construction projects. The most sustainable materials are the ones which yielded least project duration, project life cycle cost, project emissions and total project primary energy. Technique Order Preference by Similarity to Ideal Solution was then utilized to select the best solution among the set of Pareto optimal solutions.

Lee et al. [16] proposed artificial neural network model for the prediction of user-based energy consumption. The conducted study considered 24 hour schedule of 5240 single person households and the corresponding energy consumption was computed using Energyplus V 8.8 software. Six different architectures of artificial neural network were experimented and results demonstrated the best model achieved a correlation coefficient of 0.6 for the testing dataset. Panklib et al. [17] introduced regression and artificial neural network models for forecasting electricity consumption in Thailand. The input variables comprised gross domestic product, number of population, maximum ambient temperature and electricity power demand. Artificial neural network outperformed regression model accomplishing mean absolute percentage error and root-mean squared error of 1.5% and 2162, respectively.

Abed and Milosavljevic [18] analyzed power electricity consumption of a single home using multi-layer perceptron. The performance of the developed model was investigated using real data of power consumption per minute measured over four years for a single home. Results showed that the developed prediction model of one day a head power consumption, yielded mean absolute relative percentage error of 25.12%. Yuan et al. [19] proposed artificial neural network model to forecast the seasonal hourly electricity consumption of three areas in the campus namely, science and technology area, humanities college area and old liberal arts area. The input variables of the model were hourly relative humidity, hourly dry-bulb temperature, hourly global irradiance, etc. Levenberg-Marquardt algorithm was adopted to train the developed artificial neural networks. Results demonstrated that the developed model provided correlation coefficient ranging from 96% to 99% and 95% to 99% for the training and testing cases, respectively. It can be interpreted from the previous studies that there is lack of models which dealt with electricity consumption of hotels. These models are highly required to better manage and forecast the electricity consumption in hotel facilities.

3. Model Development

The ultimate objective of the present study is to develop a K-nearest neighbors (KNN) model for the prediction of electricity consumption in hotels in an attempt to enhance energy efficiency measures in hotel facilities. The framework of the developed model is composed of four main modules (see Figure 1). The dataset utilized in the present study are adopted from Eras et al. [20]. They are based on monthly electricity consumption between January 2011 and December 2014. The input parameters utilized to predict the daily electricity consumption (DEC) encompass the number of occupied rooms per day (ORD), outdoor temperature (OT), cooling degree day (CDD) and room degree day (RDD). The developed daily electricity consumption prediction model is validated through performance comparisons against a set of four widely-recognized machine learning models namely, radial basis neural network ($RBNN$), support vector machines (SVM), decision tree (DT) and Gaussian process (GP). More information about KNN, $RBNN$, SVM, DT and GP can be adopted from Gholami et al. [21], He et al. [22], El-Zahab et al. [23] and Kutyłowska [24]. The performance comparisons are carried out using split validation based on mean absolute error (MAE), mean absolute percentage error ($MAPE$) and root-mean squared error ($RMSE$). Finally, two-tailed Student’s t-tests are performed to evaluate the statistical significance level of prediction models [25].

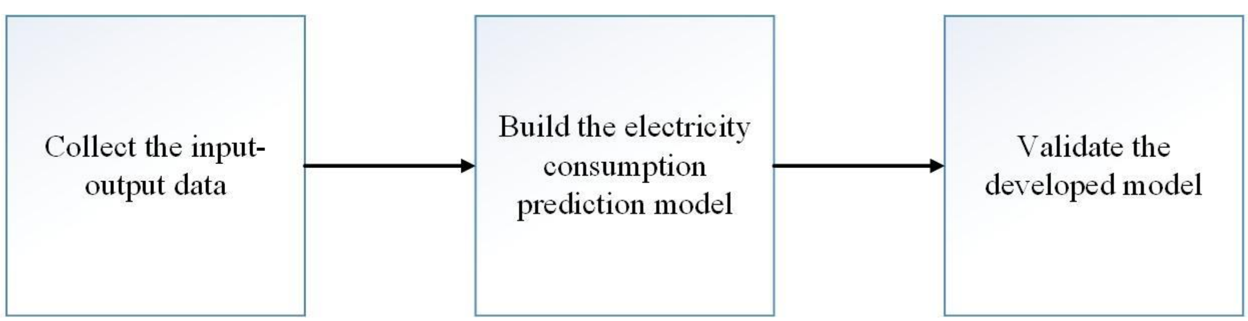

Figure 1: Framework of the developed model

4. Performance Evaluation

As mentioned earlier, the present stud, the present study utilizes $RMSE$, MAE and $MAPE$ to evaluate the developed machine learning models as presented in Eq.s (1) - (3), respectively [27-28].

$$RMSE = \sqrt{\frac{1}{K}\sum_{i=1}^{K} (O_i - P_i)^2} \quad (1)$$

$$MAE = \frac{1}{K}\sum_{i=1}^{K} |(O_i - P_i)| \quad (2)$$

$$MAPE = \frac{100}{k} \times \sum_{i=1}^{K} \frac{|P_i - O_i|}{O_i} \quad (3)$$

Where;

P_i and O_i denote the predicted and observed daily electricity consumption. K stands for number of days.

5. Model Implementation

The dataset utilized in the present study is composed of 606 observations. In this regard, 482 and 124 instances are utilized for training and testing the developed models, respectively. As described earlier, five machine learning models are developed to predict the daily electricity consumption in hotel facilities. Their hype parameters are as described in the following lines. In the radial basis neural network, the maximum number of neurons in the hidden layer is assumed 10, and the spread of the Gaussian activation function is assumed 1. In the Gaussian process regression, radial basis function is selected as the kernel function and the kernel length scale is assumed three. In the support vector machines, radial basis function is chosen as the kernel function. In this context, the gamma and convergence epsilon of the kernel function are 1 and 0.001, respectively. For the decision tree, the minimum leaf size and minimum parent size are assumed 1 and 10, respectively. A sensitivity analysis is carried out to select the optimum number of neighbors in the K-nearest neighbors model (see Figure 2). Figure 2 describes the relationship between the number of neighbors and their respective mean absolute percentage error. As can be seen, one and ten neighbors provide the lowest and highest mean absolute percentage error, respectively. One and ten neighbors accomplish mean absolute percentage error of 4.79% and 20.06%, respectively. As such, the number of neighbors in the K-nearest neighbors is assumed one.

A sample of 30 instances of the actual and predicted daily electricity consumption using K-nearest neighbors, decision tree are depicted in Figures 3 and 4, respectively. It can be inferred that K-nearest neighbors yielded very promising results in the prediction of daily electricity consumption, whereas it accomplished very close prediction to the actual values of daily electricity consumption. Decision tree attained satisfactory results. However, it predicted very far values from the actual values at some instances.

A comparative analysis between the developed five machine learning models is depicted in Table 1. In this context, K-nearest neighbors accomplished the highest prediction performance followed by decision tree while support vector machines provided the least prediction accuracy. K-nearest neighbors attained $MAPE$, $RMSE$ and MAE of 11.79%, 377.42 and 309.52, respectively. Decision tree provided $MAPE$, $RMSE$ and MAE of 8.63%, 302.61 and 231.74, respectively. Support vector machines accomplished $MAPE$, $RMSE$ and MAE of 11.98%, 382.78 and 303.39, respectively. With regards to each performance indicator separately, K-nearest neighbors obtained the least $MAPE$ and $RMSE$ followed by decision tree while support vector machines attained the highest $MAPE$ and $RMSE$. In terms of MAE, K-nearest neighbors yielded the lowest value and then decision tree while Gaussian process achieved the highest MAE.

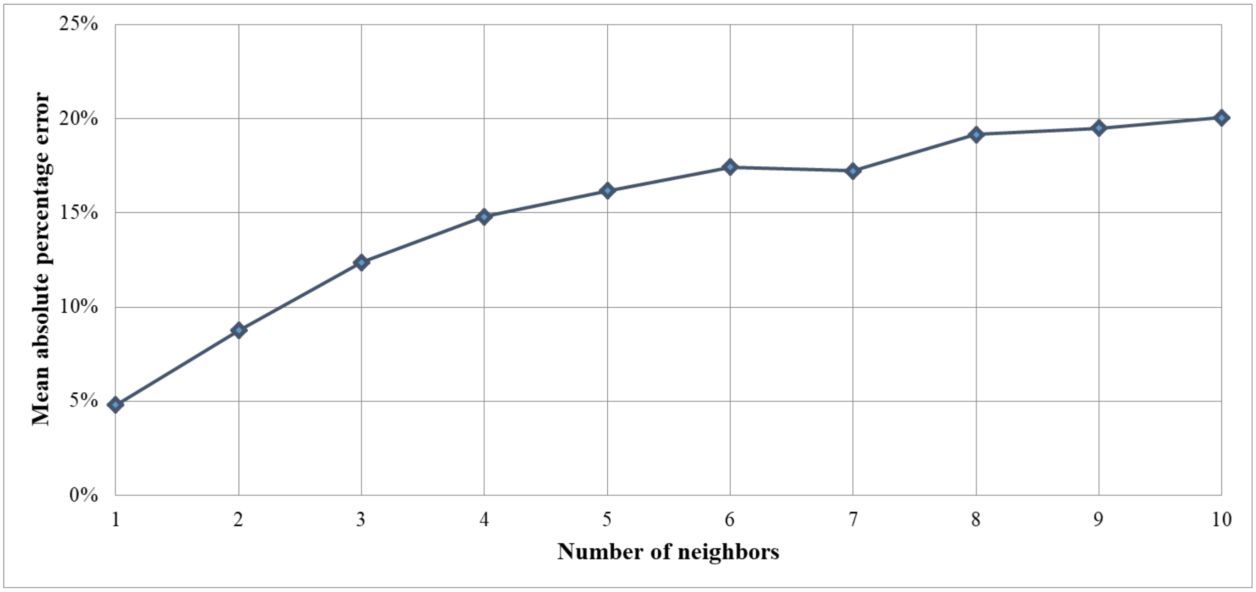

Figure 2: Plot of the mean absolute percentage with respect to the number of neighbors

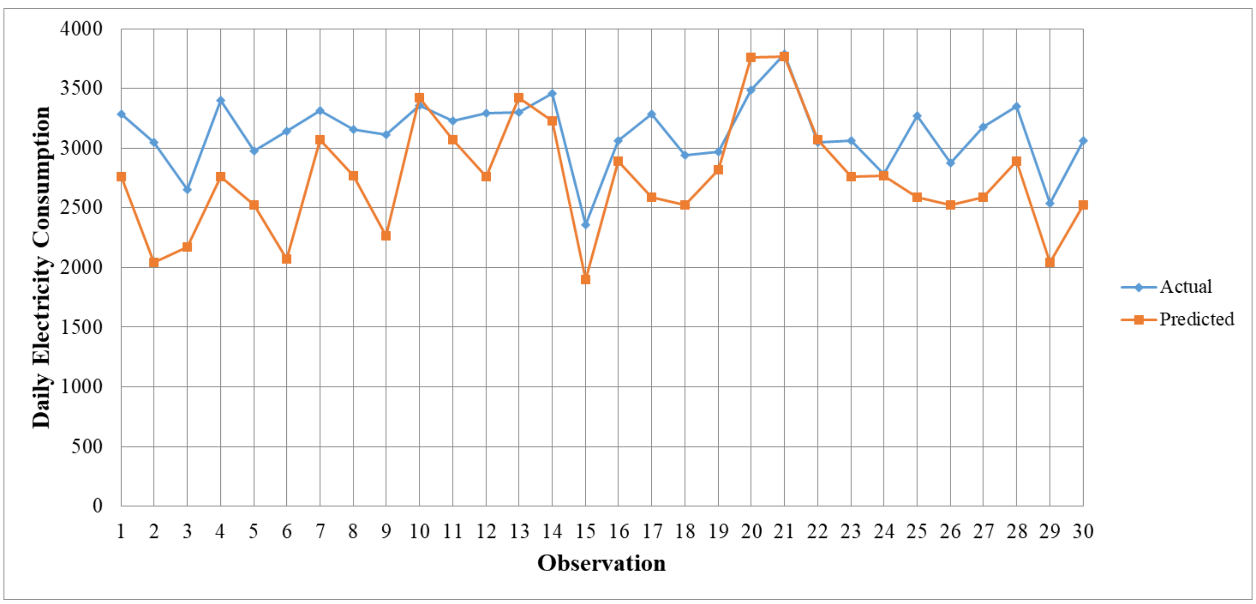

Figure 3: Actual and predicted daily electricity consumption using K-nearest neighbors

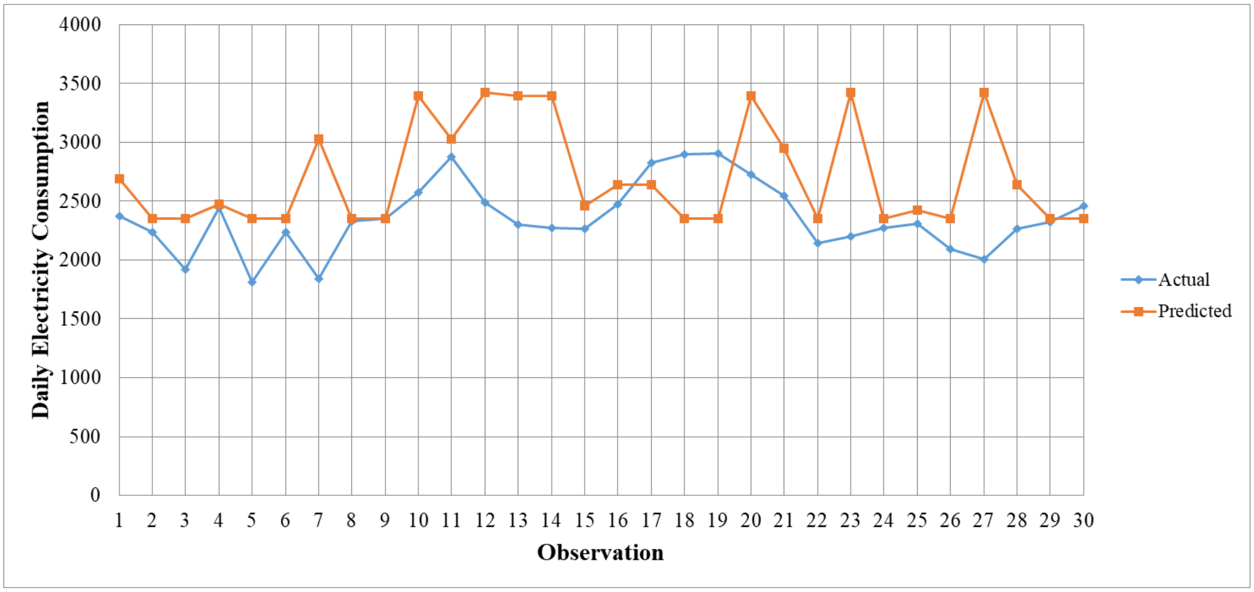

Figure 4: Actual and predicted daily electricity consumption using decision tree

Table 1: Comparative analysis of the machine learning models for predicting daily electricity consumption.

Prediction model	$MAPE$	$RMSE$	MAE
K-nearest neighbors	4.79%	276.45	130.77
Radial basis neural network	10.33%	374.49	245.85
Support vector machines	11.98%	382.78	303.39
Decision tree	8.63%	302.61	231.74
Gaussian process	11.79%	377.42	309.52

Two-tailed Student’s t-tests are carried out to test the statistical significance levels of the outcome of the developed machine learning models at significance level 5%. They examine the null hypothesis (H_0), which implies that there are no significant differences between the capacities of the machine learning models. On the other hand, the alternative hypothesis (H_1) implies that there are no significant difference between the prediction capacities of the developed machine learning models. If the $P-value$ is less than the significance level, then the null hypothesis is rejected in favor of the alternative hypothesis. However, if the $P-value$ is more than the significance level, therefore the null hypothesis is accepted. As can be seen in Table 2, there are no statistical significant differences between the prediction capacities of the developed prediction models except for the pair (RBNN, SVM). In it, the $P-value$ is 1.54×10-2. As such, radial basis neural network significantly outperforms support vector machines in the prediction of daily electricity consumption.

Table 2: Statistical comparison between the developed prediction models based on two-tailed Student's t-test.

Pair of prediction models	KNN	$RBNN$	SVM	DT	GP
KNN	($P-value$ =1)	($P-value$ =4.2×10^{-1})	($P-value$ =3.16×10^{-1})	($P-value$ =4.67×10^{-1})	($P-value$ =2.74×10^{-1})
$RBNN$	($P-value$ =4.2×10^{-1})	($P-value$ =1)	($P-value$ =1.54×10^{-2})	($P-value$ =9.11×10^{-2})	($P-value$ =1.75×10^{-2})
SVM	($P-value$ =3.16×10^{-1})	($P-value$ =1.54×10^{-2})	($P-value$ =1)	($P-value$ =5.27×10^{-1})	($P-value$ =3.26×10^{-1})
DT	($P-value$ =4.67×10^{-1})	($P-value$ =9.11×10^{-2})	($P-value$ =5.27×10^{-1})	($P-value$ =1)	($P-value$ =3.81×10^{-1})
GP	($P-value$ =2.74×10^{-1})	($P-value$ =1.75×10^{-2})	($P-value$ =3.26×10^{-1})	($P-value$ =3.81×10^{-1})	($P-value$ =1)

6. Conclusion

The electricity efficiency in hotel facilities is crucial to lessen the hotel operating costs and the associated environmental impacts. Therefore, it is essential to predict the daily electricity consumption in hotel facilities. It is imperative to establish a relationship between the electricity consumption and the significant input factors, including the number of occupied rooms per day, outdoor temperature, cooling degree day and room degree day. The machine learning models have been recently applied for predicting the electricity consumption in hotel facilities. This research presented a methodology to forecast the electricity consumption in hotel facilities using K-nearest neighbors, radial basis neural network, support vector machines, decision tree, and Gaussian process. The performance of the five models was evaluated using the mean absolute error , mean absolute percentage error and root-mean squared error grounded on split validation. The results of these models were evaluated using two-tailed Student's t-tests. The results showed that the KNN model exhibited better performance and more accurate predictions than the remaining models. The developed model can assist the specialists in developing control measures for an efficient energy management system.

References

[1] Upadhyay, A., and Vadam, C. "The Role of Energy Consumption in Hotel Operations." *22nd International Annual EurOMA Conference*, pp. 1-10, 2015.

[2] Chadha, H., and Onkar, P. "Sustainable urban tourism: A case of Omkareshwar." *Oida International Journal of Sustainable Development*, vol. *7*, no.2, pp. 55-70, 2014.

[3] Emmett, R. S., and Nye, D. E. *The environmental humanities: A critical introduction*. MIT Press, 2017.

[4] Song, X., Zhu, T., Liu, L., and Cao, Z. "Study on optimal ice storage capacity of ice thermal storage system and its influence factors." *Energy Conversion and Management*, vol. 164, pp. 288-300, 2018.

[5] Stipanuk, D. M. "Energy management in 2001 and beyond: operational options that reduce use and cost." The *Cornell Hotel and Restaurant Administration Quarterly*, vol. 42, no.3, pp. 57-70, 2001.

[6] D'Agostino, D., Cuniberti, B., and Bertoldi, P. "Energy consumption and efficiency technology measures in European non-residential buildings." *Energy and Buildings*, vol. 153, pp. 72-86, 2017.

[7] Tanaka, K. "Review of policies and measures for energy efficiency in industry sector." *Energy policy*, vol. 39, no. (10), pp. 6532-6550, 2011.

[8] Becken, S., Frampton, C., and Simmons, D. "Energy consumption patterns in the accommodation sector—the New Zealand case." *Ecological economics*, vol. 39, no. 3, pp. 371-386, 2001.

[9] Al-Sakkaf, A., Zayed, T., Bagchi, A., Mahmoud, S., and Pickup, D. "Development of a sustainability rating tool for heritage buildings: future implications." *Smart and Sustainable Built Environment*, pp. 1-17, 2020.

[10] Al-Sakkaf, A., Zayed, T., Bagchi, A. and Mohammed Abdelkader, E. “Sustainability rating tool and rehabilitation model for heritage buildings”, CSCE Annual Conference, Laval, Canada, 2019.
[11] Freeman, J., Hellgardt, K., and Markides, C. N. “Working fluid selection and electrical performance optimisation of a domestic solar-ORC combined heat and power system for year-round operation in the UK.” *Applied Energy*, vol. 186, pp. 291-303, 2017.
[12] Khalil, A. J., Barhoom, A. M., Abu-Nasser, B. S., Musleh, M. M., and Abu-Naser, S. S. “Energy Efficiency Predicting using Artificial Neural Network.” *International Journal of Academic Pedagogical Research*, vol. 3, no. 9, pp. 1–8.
[13] Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., and Ahrentzen, S. “Random Forest based hourly building energy prediction.” *Energy and Buildings*, vol. 171, pp. 11–25, 2018.
[14] Mohammed Abdelkader, E., Al-Sakkaf, A., and Ahmed, R. (2020). “A comprehensive comparative analysis of machine learning models for predicting heating and cooling loads.” *Decision Science Letters*, vol. 9, no. 3, pp. 409–420, 2020.
[15] Marzouk, M., and Mohammed Abdelakder, E. “A hybrid fuzzy-optimization method for modeling construction emissions.” *Decision Science Letters*, vol. 9, no. 1, pp. 1–20, 2020.
[16] Lee, S., Jung, S., and Lee, J. “Prediction model based on an artificial neural network for user-based building energy consumption in South Korea.” *Energies*, vol. 12, no. 4, pp. 1-18, 2019.
[17] Panklib, K., Prakasvudhisarn, C., and Khummongkol, D. “Electricity Consumption Forecasting in Thailand Using an Artificial Neural Network and Multiple Linear Regression.” *Energy Sources, Part B: Economics, Planning and Policy*, vol. 10, no. 4, pp. 427–434, 2015.
[18] Abed, F. N., and Milosavljević, M. “On the Persistence Forecast of Single Home Electricity Power Consumption.” *International Journal of Innovative Science, Engineering & Technology*, vol. 3, no. 1, pp. 183–188, 2016.
[19] Yuan, J., Farnham, C., Azuma, C., and Emura, K. “Predictive artificial neural network models to forecast the seasonal hourly electricity consumption for a University Campus.” *Sustainable Cities and Society*, vol. 42, pp. 82–92, 2018.
[20] Eras, J. C., Santos, V. S., Gutierrez, A. S., and Vandecasteele, C. “Data supporting the improvement of forecasting and control of electricity consumption in hotels.” *Data in Brief*, vol. 25, pp. 1–8, 2019.
[21] Gholami, A., Bonakdari, H., Zaji, A. H., and Akhtari, A. A. “An efficient classified radial basis neural network for prediction of flow variables in sharp open-channel bends.” *Applied Water Science*, vol. 9, no. 6, pp. 1–17, 2019.
[22] He, P., Li, S. cai, Xiao, J., Zhang, Q. qing, Xu, F., and Zhang, J. (2018). “Shallow Sliding Failure Prediction Model of Expansive Soil Slope based on Gaussian Process Theory and Its Engineering Application.” *KSCE Journal of Civil Engineering*, vol. 22, no. 5, pp. 1709–1719, 2018.
[23] El-Zahab, S., Mohammed Abdelkader, E. and Zayed, T. “An Accelerometer-Based Leak Detection System”. Mechanical Systems and Signal Processing, vol. 108, pp. 276-271, 2018.
[24] Kutyłowska, M. (2018). “K-nearest neighbours method as a tool for failure rate prediction.” *Periodica Polytechnica Civil Engineering*, vol. 62, no. 2, pp. 318–322, 2018.
[25] Mohammed Abdelkader, E., Marzouk, M., and Zayed, T. “An Optimization-Based Methodology for the Definition of Amplitude Thresholds of the Ground Penetrating”. *Soft Computing*, vol. 23, pp. 12063-12086, 2019.
[26] Mohammed Abdelkader, E., Moselhi, O., Marzouk, M. and Zayed, T. (2021). “Entropy-Based Automated Method for Detection and Assessment of Spalling Severities in Reinforced Concrete Bridges”. *Journal of Performance of Constructed Facilities*, vol. 35, no. 1, pp. 1-25, 2021.
[27] Nazari, A., Rajeev, P., and Sanjayan, J. G. “Offshore pipeline performance evaluation by different artificial neural networks approaches.” *Measurement*, vol. 76, pp. 117–128, 2015.