Proceedings of International Structural Engineering and Construction, 10(1), 2023
Innovative Theory and Practices in Structural Engineering and Construction
Edited by Karatas, A., Iranmanesh, A., Gurgun, A., Yazdani, S., and Singh, A.

ISSN: 2644-108X
www.doi.org/10.14455/ISEC.2023.10(1).ENR-05

PREDICTING ENERGY USE INTENSITY OF US HOTEL BUILDINGS USING CBECS MICRODATA

HEVAR PALANI, JUAN ACOSTA-SEQUEDA, SYBIL DERRIBLE, and ASLIHAN KARATAS

Dept of Civil Engineering, Materials, and Environmental Engineering, University of Illinois at Chicago, Chicago, IL, USA

In the United States, the commercial sector consumes 21% of the total energy use. With 14%, the ~47,000 hotels are considered to be the third main source of energy consumption in the commercial sector. Stakeholders in the hotel industry have shown significant interest in reducing energy consumption in hotel buildings. However, determining the primary factors that contribute to overall energy consumption is important to develop efficient and effective retrofitting strategies. To address this question, this study focuses on identifying the variables in hotels that contribute majorly to their energy consumption. To achieve that, this study utilizes different machine learning approaches for estimating the source of Energy Use Intensity (EUI) for US hotel buildings based on Commercial Building Energy Consumption Survey (CBECS) 2018 microdata. The findings derived from this research can significantly contribute to the optimization of retrofitting strategies and building design in hotel buildings, as well as the development of effective electrification strategies. Ultimately, this knowledge will empower decision makers to make informed choices that enhance energy efficiency and sustainability in hotel buildings.

Keywords: Energy consumption, Electrification, Feature importance, Machine learning, Hotels, Sustainability.

1 INTRODUCTION

Hotel buildings account for 14% of commercial sector's total energy consumption (US DOE 2023). Of the ~47,000 hotel buildings in the US (EnergyStar 2022, US DOE 2023), the retrofitting and electrification of these hotel buildings is recognized as a desirable approach to substitute traditional fossil fuel-based systems with renewable energy sources (Hansen *et al*. 2019). While some hotels have made satisfactory progress in electrifying their operations, the extent of electrification across the entire hotel industry is still not significant. This transition can help improve energy efficiency and lower environmental impacts for hotel buildings.

Additionally, uncertainties about the primary factors that influence energy consumption in hotel buildings have made it more challenging to develop electrification and retrofitting strategies for achieving net-zero buildings and for decarbonization. These uncertainties have added complexity to the ongoing efforts in this area (Connolly 2017, Hong *et al*. 2023). Therefore, the primary objective of this study is to identify the most influential variables for energy consumption (i.e., source energy usage) of the hotel buildings. These findings will subsequently inform the development of effective retrofitting and electrification strategies. Researchers have developed statistical methodologies over the past two decades to better predict total energy consumption of buildings using large data sets (Ahmad *et al*. 2014, Deng *et al.* 2018) and investigate the interrelationships between energy use and the consumption of other resources such as water, gas, and even transport (Movahedi and Derrible 2021, Palani *et al*. 2023). However, these studies tend to look at buildings in aggregate form or to focus on residential or office buildings. Therefore, this study uses the Commercial Building Energy Consumption Survey (CBECS) 2018 microdata (US EIA 2023) to measure the energy use intensity (annual energy use of building in kBtu/total floor area in square footage) of US hotel buildings. To achieve this, various machine learning (ML) techniques—*K*-nearest neighbor (KNN), linear regression (LR), support vector machine (SVM), random forest (RF), and gradient boosting (GB)—were applied to form generalized models for the prediction of source energy usage in hotels.

The findings are expected to help decision makers such as utility companies in developing their electrification strategies to achieve net zero energy buildings and decarbonization goals.

2 METHODOLOGY

The methodology of this study is divided into three sections: (1) data organization, (2) uncertainty analysis, and (3) data analysis.

2.1 Data Organization

The data organization includes developing the list of categorical and continuous variables. Data was collected from the Commercial Building Energy Consumption Survey (CBECS) 2018 microdata governed by the US Department of Energy (DOE) and conducted by the US Energy Information Administration (EIA). The CBECS microdata includes the total energy consumption for six years (2013-2018) in US commercial buildings.

The CBECS 2018 microdata reported annual energy consumption of 6,436 buildings in the US, representing over five million total commercial buildings in the country. It has 510 variables. As part of data cleaning process, variables were eliminated that had (1) a significant number of missing values (more than 80% per variable), (2) no relevance to standard hotel buildings (e.g., office, restaurant), (3) no relation to building energy consumption, (4) little to no variation (<10%), and (5) similar or duplicate information. After completing the cleaning process, the total number of input variables was 79 where 57 were categorical and 22 were continuous; see Tables 1 and 2. Out of the 6,436 commercial buildings in the CBECS 2018 microdata, 272 were hotels (4.2%).

To evaluate the energy performance of hotel buildings in the US, the Energy Use Intensity (EUI) metric was utilized, which is commonly used for this purpose (Bauer and Scartezzini 1998 Chung *et al*. 2006). EUI is calculated by dividing the annual energy consumption of the hotel building (in kilo British thermal unit, or kBtu) by the hotel's total floor area. Figure 1 shows the distribution of the calculated source EUI of the hotel population after incorporating the weights available in the survey. From the distribution we can observe that the EUI ranges from 16 to 265 kBtu/ft^2. The hotels at the 95th percentile use almost 6 times the energy oat the 5th percentile. The source EUI distribution presents a negative skew, meaning that the most efficient hotels are closer to the median than the most energy intensive.

Table 1. List of categorical variables for source EUI prediction.

#	Categorical Variables	Variable Label	No. of Classes	#	Categorical Variables	Variable Label	No. of Classes
1	Census Region	REGION	4	30	Elec. used for cooking	ELCOOK	2
2	Census Division	CENDIV	9	31	Energy management plan	ENRGYP LN	2
3	Wall cons. material	WLCNS	9	32	Fast food/small restaurant	FASTFD	2

4	Roof cons. material	RFCNS	9	33	Cafeteria/large restaurant	CAF	2
5	Roof tilt	RFTILT	3	34	Ext. overhangs or awnings	AWN	2
6	Building shape	BLDSHP	11	35	Conference or event space	CONFSP	2
7	Building owner type	OWNTYPE	10	36	Indoor pool	POOL	2
8	Main heating equipment	MAINHT	8	37	Full-size residential type refrigerators/freezers	RFGRES	2
9	Main cooling equipment	MAINCL	9	38	Half-size/compact refrigerators/freezers	RFGCOMP	2
10	How reduce heating during 24 h period	HWRDHT	5	39	Walk-in refrigerators/freezers	RFGWI	2
11	How reduce cooling during 24 h period	HWRDCL	5	40	Open refrigerator/freezer cases or cabinets	RFGOP	2
12	Window glass type	WINTYP	4	41	Closed refrigerator/freezer cases or cabinets	RFGCL	2
13	Laundry onsite	LAUNDR	3	42	Fluorescent lighting	FLUOR	2
14	Cool roof material	RFCOOL	2	43	CFL bulbs	CFLR	2
15	Escalators	ESCLTR	2	44	Incandescent light bulbs	BULB	2
16	Attic	ATTIC	2	45	Halogen bulbs	HALO	2
17	Owner responsible for operat. and maint. of energy system	OWNOCC	2	46	High intensity discharge (HID) lighting	HID	2
18	Window replacement	RENWIN	2	47	LED lighting	LED	2
19	HVAC equip. upgrade	RENHVC	2	48	Light scheduling	SCHED	2
20	Lighting upgrade	RENLGT	2	49	Occupancy sensors	OCSN	2
21	Energy for secondary heating	HT2	2	50	Multi-level lighting/ dimming	DIM	2
22	Insulation upgrade	RENINS	2	51	Daylight harvesting	DAYHARV	2
23	On a multibuilding complex	FACIL	2	52	Electricity for water heating	ELWATR	2
24	Roof replacement	RENRFF	2	53	Tinted window glass	TINT	2
25	Electric upgrade	RENELC	2	54	Reflective window glass	REFL	2
26	Building automation system (BAS)	EMCS	2	55	Large, commercial kitchen prep area	FDPREP	2
27	Economizer cycle	ECN	2	56	Skylight/atriums designed	SKYLT	2
28	Regular HVAC maint.	MAINT	2	57	Electricity for cooling	ELCOOL	2
29	Light parking area	PKLT	2				

KNN was first used to impute the missing data in the sample set of 272 hotel buildings from the CBECS 2018 dataset. The KNN value was set to 10 since it worked best. Five nearest neighbors with KNN values of 3, 5, 6, 7, and 10 were considered. Following this, hot encoding (i.e., a process for converting categorical variables to numerical, non-binary ones) was applied on all ML techniques with more than 2 classes for better model prediction. Then, the categorical variables with only two classes (yes, no questions) as shown in Table 1 were converted to '0' and '1' instead of '1' and '2' as we had in the original data. Finally, the continuous variables (see Table 2) were normalized to make the training model less sensitive to the scale of variables and prevent certain variables from dominating over other to create more accurate model.

Table 2. List of continuous variables for source EUI prediction.

#	Continuous Variables	Variable Label	#	Continuous Variables	Variable Label
1	Number of floors	NFLOOR	12	Number of desktop computers	PCTERMN
2	Number of underground floors	BASEMNT	13	Number of laptops	LAPTPN
3	Floor to ceiling height	FLCEILHT	14	Number of TV or video displays	TVVIDEON
4	Number of elevators	NELVTR	15	Percent lit when open	LTOHRP
5	Year of Construction	YRCONC	16	Percent lit off hours	LTNHRP

6	Number of guest rooms	LODGRM	17	Percent of exterior lighted	LTEXPC
7	Number of businesses	NOCC	18	Percent daylight	DAYLTP
8	Lodging room percent occupancy	LODOCCP	19	Percent exterior glass	GLSSPC
9	Number of employees	NWKER	20	Heating degree days	HDD65
10	Percent heated	HEATP	21	Cooling degree days	CDD65
11	Percent cooled	COOLP	22	Annual major fuels consumption (thous Btu)	MFBTU

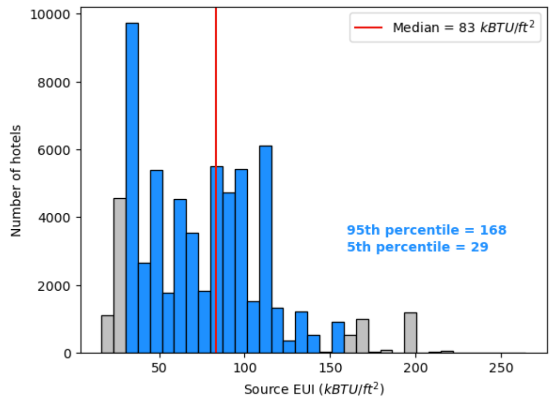

Figure 1. Distribution of source EUI for hotel population.

2.2 Uncertainty Analysis

Uncertainty analysis was performed by using cross-validation to evaluate the model performance of the proposed regression and ML analytical techniques. The K-fold method was adopted, where the data is divided into K partitions. The CBECS 2018 microdata was randomly split into three portions, with 64% assigned to the training set, 16% to the validation set, and 20% to the testing set. The validation set was considered a part of the training set and used to tune the parameters of each proposed model. Grid search (GridSearchCV) was used for tuning the parameters and selecting optimal ones with mean squared error (MSE). Finally, all the developed models were evaluated on the testing set (20%) to compare their generalization capabilities. To evaluate each model's performance, mean absolute error (MAE) and root mean squared error (RMSE) were used to measure the deviation between actual and predicted energy usage in US hotel buildings.

2.3 Data Analysis

Table 3 shows the MAE and the RMSE for each of the applied prediction models. Normally, the errors in the training set and validation set are indicating the goodness-of-fit of the developed models while the errors in the testing set indicate the willingness of the developed model to generalize to new and unseen data (Deng *et al.* 2018). The results show that overall GB has the best performance when taking into account both error metrics and all three datasets. For this reason, from this point all the presented results use this model to interpret the data.

Table 3. Comparison of MAE and RMSE among different sets for total source EUI prediction.

Error on source EUI	**MAE**			**RMSE**		
	Training	**Validation**	**Testing**	**Training Set**	**Validation Set**	**Testing Set**
LR	8.58	N/A	52.22	11.69	N/A	66.99
SVM	23.46 ± 1.00	28.88 ± 3.56	33.12	35.69 ± 1.076	38.32 ± 5.78	42.17
RF	23.27 ± 0.95	28.79 ± 2.43	31.48	29.12 ± 0.96	37.82 ± 4.65	40.61
GB	21.72 ± 0.62	28.00 ± 2.80	31.78	28.7 ± 0.92	36.82 ± 4.62	38.56

3 RESULTS AND DISCUSSION

Using GB, the 20 most important features were obtained (see Figure 2). The important variables are the percent of the building lit when open (i.e., LTOHRP), followed by the existence of a cafeteria or large restaurant (i.e., CAF), the window glass type (i.e., WINTYP), the percent of daylight (i.e., DAYLTP), and the percentage building heated to at least 50 degrees Fahrenheit (i.e., HEATP). Remarkably, the variables of number of guest rooms (i.e., LODGRM) and number of floors (i.e., NFLOOR) were found to be relatively less important than the others. Even though they rank within the top 20 out of the 79 variables considered, their relative lower significance suggests that they do not exert a substantial influence on the source EUI in hotels. Further, the importance of each variable was also investigated based on the distribution of the observations using SHapley Additive exPlanations (SHAP). In Figure 3, the features that are at the top are the ones that impact the model's output the most. On the plot, each observation is presented for each feature as a single point that is positioned on the horizontal axis according to its influence on the output EUI value. In addition, the piling up of the points are meant to represent the density of observations in a distribution of values of the corresponding feature. Figure 3 indicates that the lower the percentage of buildings lit when open (i.e., LTOHRP), the more likely the building is to be more energy efficient. In addition, the existence of a cafeteria or large restaurant (i.e., CAF) has a significant impact on the likelihood of a hotel being less energy efficient, but the absence of it has a low impact in making it more efficient. One additional thing worth noticing is the presence of hot encoded features. BLDSHP_2, HWRHDT_2, and RFCNS_3, which determine whether a building has a wide rectangle shape, whether the temperature is set manually, and whether it has wooden roof, respectively. According to Figure 3, having a wide rectangle shape and setting the temperature manually to reduce heating make a hotel more likely to be more energy efficient, while having a wooden roof makes it less likely to be efficient. In conclusion, building efficiency relates more to design and control choices such as the additional facilities added to the hotel (e.g., restaurants, cafeteria), window glass types, temperature adjustment mechanisms, and lighting and daylight allowance. The MAE and RMSE errors for the testing set of the LR model are higher than other models, indicating that the data is not linear, which aligns with the findings from Deng *et al.* (2018) that applied ML techniques to CBECS 2012 microdata for office buildings. Furthermore, the accuracy (R^2) of the used ML techniques (LR, SVM, RF, GB) was <10%, making it unsuitable for comparing model performance given the data's complexity. Instead, the MAE and RMSE were used as metrics for estimating model performance.

4 CONCLUSION

The retrofitting and electrification of hotel buildings in the US is recognized as a beneficial strategy to replace fossil fuel with renewable energy sources. This transition offers advantages in terms of energy efficiency and environmental impact for the hotel industry. Hence, the main aim of this study was to determine the key factors that have the greatest impact on source EUI in hotels. The study utilized the CBECS 2018 microdata to assess the source EUI of the US hotel buildings. To do that, several ML techniques were used. The CBECS microdata was selected since it represents over 47,000 hotels in the US. The outcome of this study can provide valuable insights to decision-makers (e.g., utility companies) in formulating effective electrification strategies that align with the goals of achieving net zero energy buildings and decarbonization. These findings are anticipated to inform and support the development of strategies aimed at reducing energy consumption and advancing sustainability in the hotel industry.

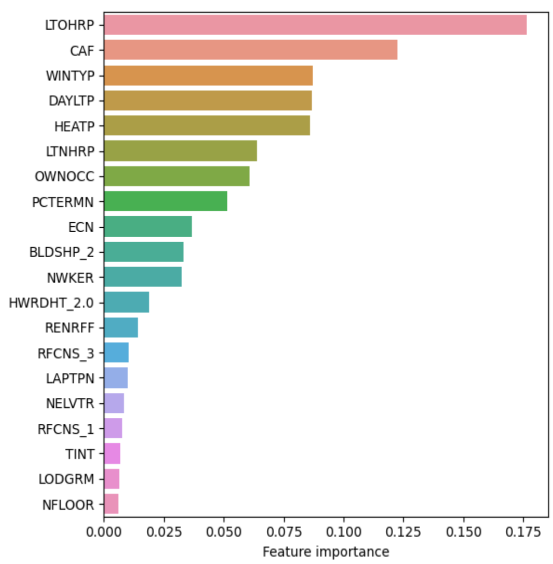

Figure 2. Source EUI feature importance under the GB model.

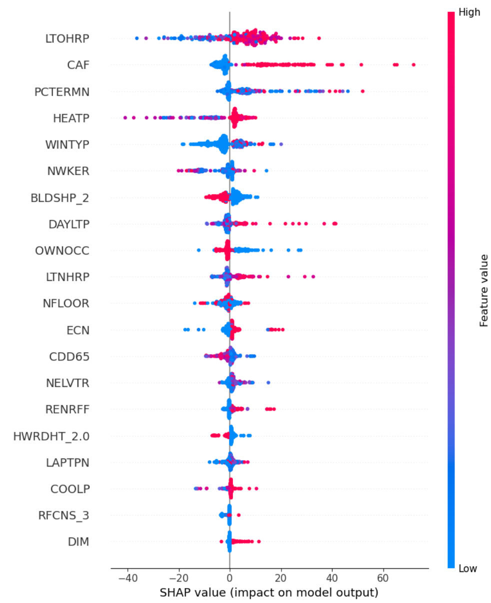

Figure 3. SHAP summary plot obtained from GB model for source EUI.

References

Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H., and Saidur, R., *A Review on Applications of ANN and SVM for Building Electrical Energy Consumption Forecasting,* Renew. Sust. Energ. Rev., Elsevier, 33, 102–109, May, 2014.

Bauer, M., and Scartezzini, J. L., *A Simplified Correlation Method Accounting for Heating and Cooling Loads in Energy-efficient Buildings,* Energy Build., Elsevier, 27(2), 147–154, April, 1998.

Chung, W., Hui, Y. V., and Lam, Y. M., *Benchmarking the Energy Efficiency of Commercial Buildings,* Appl. Energy, Elsevier, 83(1), 1–14, January, 2006.

Connolly, D., *Heat Roadmap Europe: Quantitative Comparison between the Electricity, Heating, and Cooling Sectors for Different European Countries,* Energy, Elsevier, 139, 580–593, November, 2017.

Deng, H., Fannon, D., and Eckelman, M. J., *Predictive Modeling for US Commercial Building Energy Use: A Comparison of Existing Statistical and Machine Learning Algorithms Using CBECS Microdata,* Energy Build., Elsevier, 163, 34–43, 2018.

EnergyStar, *Hotels: An Overview of Energy Use and Energy Efficiency,* www.energystar.gov, 2022. Retrieved from https://www.energystar.gov on May 10, 2023.

Hansen, K., Mathiesen, B. V., and Skov, I. R., *Full Energy System Transition Towards 100% Renewable Energy in Germany in 2050,* Renew. Sust. Energ. Rev., Elsevier, 102, 1–13, March, 2019.

Hong, T., Lee, S. H., Zhang, W., Sun, K., Hooper, B., and Kim, J., *Nexus of Electrification and Energy Efficiency Retrofit of Commercial Buildings at the District Scale,* Sustain. Cities Soc., Elsevier, 95, 104608, August, 2023.

Movahedi, A., and Derrible, S., *Interrelationships between Electricity, Gas, and Water Consumption in Large-scale Buildings*, J. Ind. Ecol., John Wiley & Sons, 25(4), 932–947, August, 2021.

Palani, H., Acosta-Sequeda, J., Karatas, A., and Derrible, S., *The Role of Socio-Demographic and Economic Characteristics on Energy-Related Occupant Behavior*, J. Build. Eng., Elsevier, 31, 106875, September, 2023.

US DOE, *Hospitality*, betterbuildingssolutioncenter.energy.gov, 2023. Retrieved from https://betterbuildingssolutioncenter.energy.gov/sectors/hospitality on May 3, 2023.

US EIA, *2018 CBECS Survey Data*, www.eia.gov, 2023. Retrieved from https://www.eia.gov/consumption/commercial/data/2018/index.php?view=microdata on March 7, 2023.