Estimations for Energy Consumption in Domestic and Public Buildings in London Using Machine Learning

Fatihah Shah
20th September 2024

Outline

1

Introduction

- Motivation
- Objectives
- Advantages and Disadvantages of Machine Learning (ML) models
- Previous study using ML models

2

Methodology

- Flowchart process
- Feature Importance
- Model Evaluation

3

Results

- Impact of EPC, DEC, and LBSM Data Integration on Model Performance
- Comparative Analysis of Energy Consumption Patterns in Domestic and Public Buildings
- Analyzing Model Effectiveness Using Solely LBSM Data
- A Comparative Analysis of Estimated Energy Consumption with CO2 Emissions

4

Conclusion

- Future work

INTRODUCTION

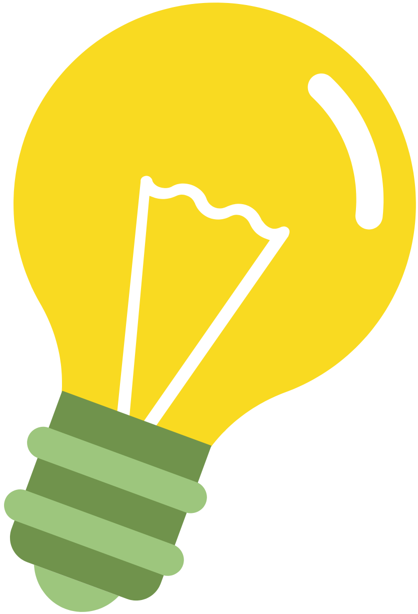
MOTIVATION

- Most of the years of construction for domestic buildings in London are built in age band 1900 to 1949 [1].

- Domestic sector has the highest demand for total electricity in United Kingdom (UK) which is 29.8% compared to industry, commercial and energy industry [2].

- It is predicted almost 68% of the world's population will move to the urban area by 2050 [3]

- In 2022, 84% of population in UK lives in urban areas and this leads to higher energy consumption [4,5].



OBJECTIVES

- To determine the best Machine Learning models using spatial and socioeconomic factors from Energy Performance Certificate (EPC), Display Energy Certificate (DEC) and London Building Stock Model (LBSM) for domestic and public buildings

- To investigate the difference in estimated energy consumption between domestic and public buildings based on the features selected from EPC,DEC, and LBSM data

- To do a comparative analysis of the impact of feature selection from EPC,DEC, and LBSM, compared to using only LBSM data, on the accuracy of energy consumption predictions for domestic and public buildings in London

- To study the correlation of estimated energy consumption with CO2 emissions from EPC and DEC

Table 1: Advantages and Disadvantages of the Machine Learning (ML) models

References	Models	Advantages	Disadvantages
[6]	**Decision Tree (DT)**	• Useful in Data Exploration • Minimal Data Cleaning Required	• Overfitting • Not Ideal for Continuous variables
[7]	**Gradient Boosting (GB)**	• Best in handling complexity between linear and non-linear features	• More prone to overfitting if not properly tuned
[7]	**Extreme Gradient Boosting (XGBoost)**	• Highly scalable and fast to train with regularization • Efficient handling of sparse datasets	• Requires careful tuning of hyperparameters
[7,8]	**Random Forest (RF)**	• Independent from hyperparameter tuning	• Requires higher computational time
[9]	**Deep Neural Network (DNN)**	• Can scale with large datasets • Capable of modeling complex and non-linear relationships	• Requires significant computational resources and overfitting
[10,11]	**Support Vector Machine (SVM)**	• Robust to noise • Perform well with smaller datasets	• Hyperparameter Sensitivity • Lack of Interpretability

Table 2: Previous study estimates Energy consumption using ML

References	Models	Building type	Type of data	Sample size
[11]	**DNN** GB DT SVM Stacking Multilinear Regression(MLR) K Nearest Neighbours (KNN) Artificial Neural Network (ANN)	Domestic Building	Annually -Environmental and building features	5000 buildings
[12]	**DNN** ANN XGBoost SVM MLR	Domestic Building	Hourly -Occupants and seasonality	6 buildings
[13]	**XGBoost** GB RF SVM MLR Elastic Net (ELN)	Public Building	Daily -Environmental and building features	1 building

METHODOLOGY

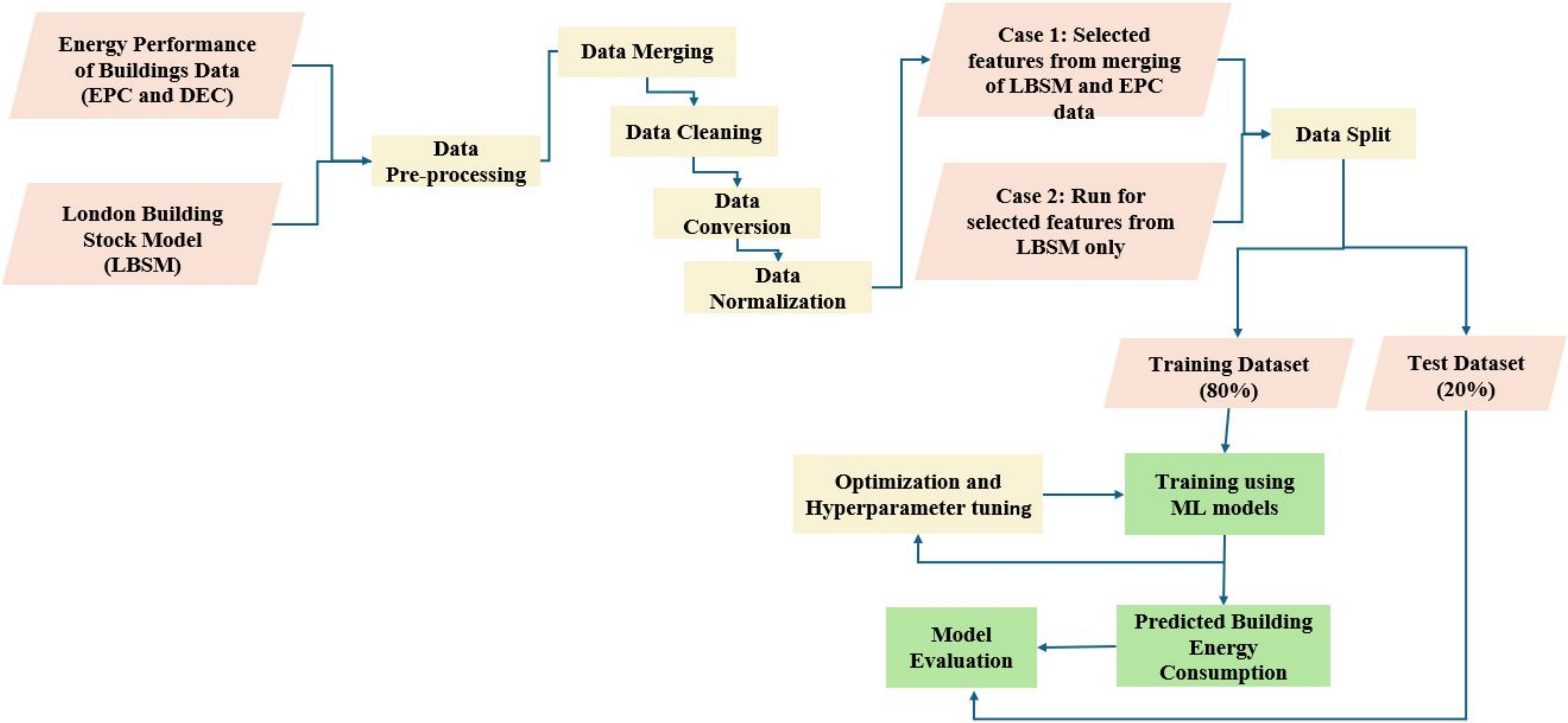


Figure 1: Flowchart process of estimating energy consumption for Domestic and Public Buildings in London

Table 3: Model performance for feature importance (Domestic Buildings)

Model	Training Time (s)	R-squared
XGBoost	1.800	0.890
GB	**126.930**	**0.900**
RF	80.690	0.830
DT	0.180	0.490

Figure 2: Feature Importance using Gradient Boosting (Domestic Building)

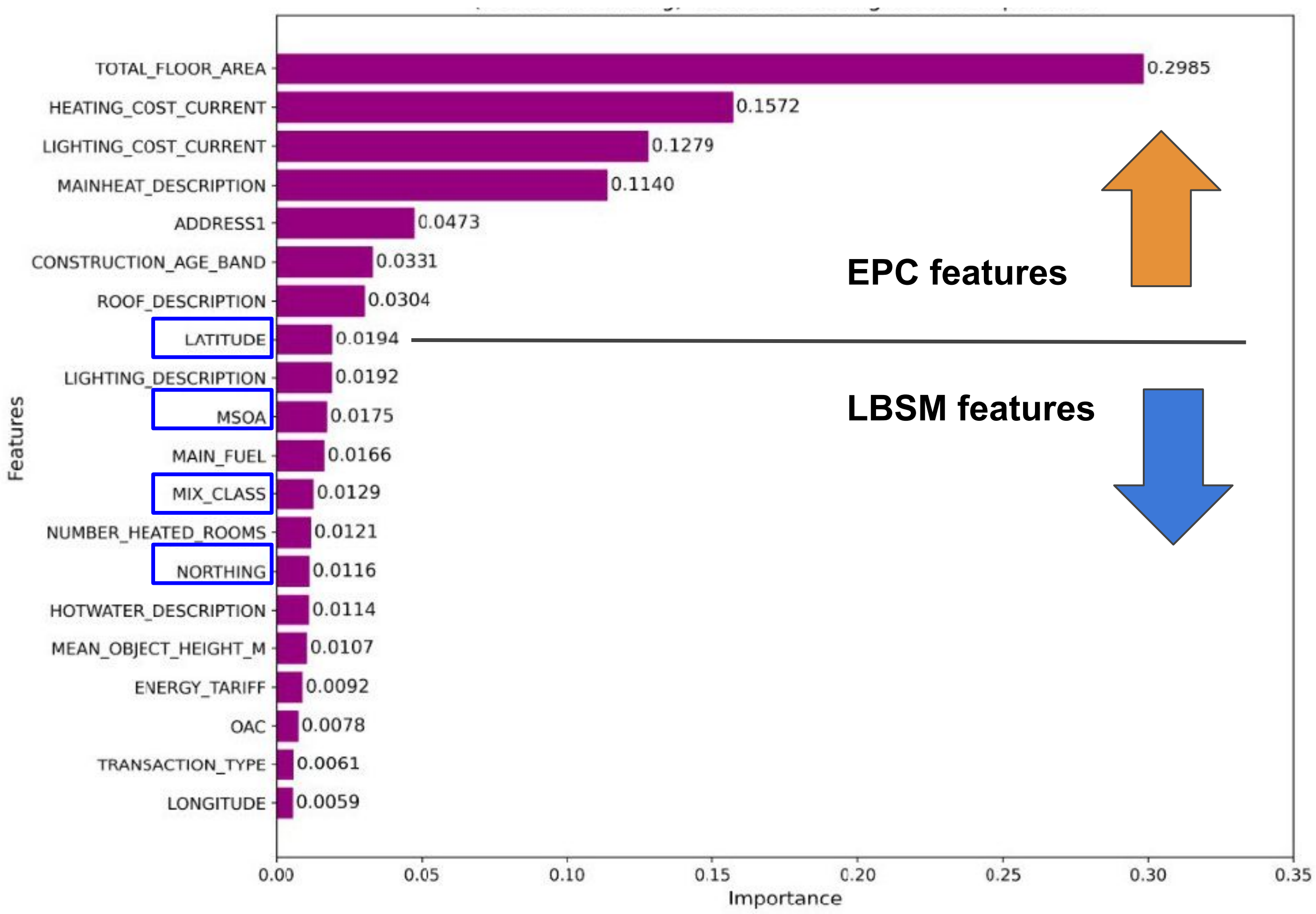


Table 5: Features from EPC and LBSM (Domestic Buildings) [14,15]

No.	Features	Abbreviation	Unit	Label
1	Energy Consumption	Energy Consumption	kWh/m²/ yr	Output
2	Total Floor Area	Total Floor Area	m²	Input
3	Heating Cost Current	Heating Cost Current	NA	
4	Lighting Cost Current	Lighting Cost Current		
5	Mainheat Description	Mainheat Description		
6	Detail address of a building	Address1		
7	Age of a building	Construction Age Band		
8	Roof Description	Roof Description		
9	A value of Latitude location based on ETRS89 coordinate reference system from OS AddressBase	Latitude		

Table 5: Features from EPC and LBSM (Domestic Buildings) [14,15]

No.	Features	Abbreviation	Unit	Label
10	Census Middle Layer Super Output Area	MSOA	NA	Input
11	Lighting Description	Lighting Description	NA	Input
12	Main Fuel	Main Fuel	NA	Input
13	Mix class	Mix class	NA	Input
14	Number of heated rooms	Number Heated Rooms	NA	Input
15	Hotwater Description	Hotwater Description	NA	Input
16	The x location based on OS AddressBase	Northing	NA	Input

Table 4: Model performance for feature importance (Public Buildings)

Model	**Training Time (s)**	**R-squared**
XGBoost	0.330	0.580
GB	25.53	0.640
RF	**32.74**	**0.690**
DT	0.060	0.550

Figure 3: Feature Importance using Random Forest (Public Building)

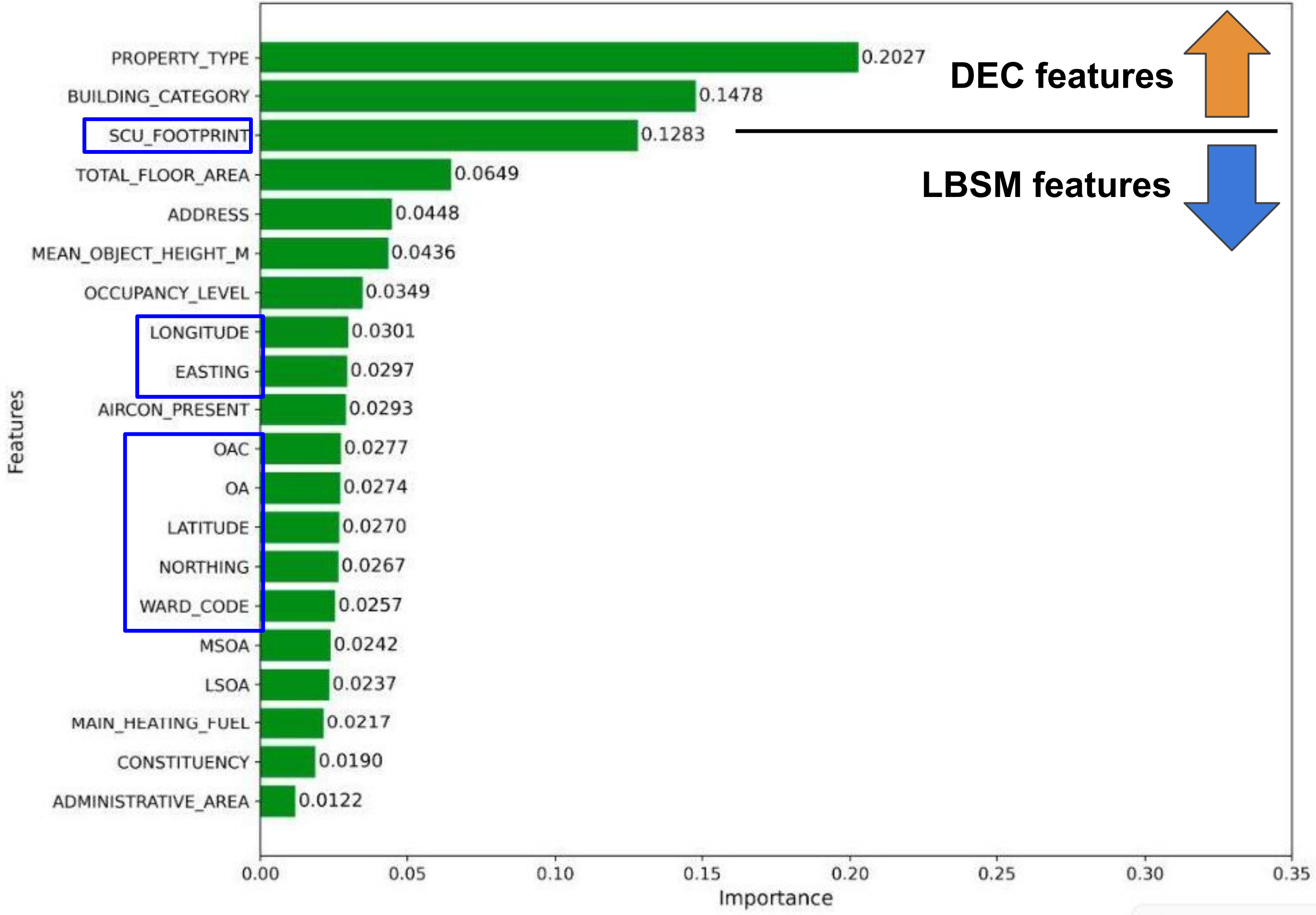


Table 6: Features from DEC and LBSM (Public Buildings) [14,15]

No.	Features	Abbreviation	Unit	Label
1	Energy Consumption	Energy Consumption	$kWh/m^2/yr$	Output
2	Total Floor Area	Total Floor Area	m^2	Input
3	Gross External area of a self-contained units (SCU)	SCU Footprint	m^2	
4	Height of a building	Mean Object Height M	m	
5	Type of property	Property Type	NA	
6	Mix uses in a building (retail store, office space, etc)	Building Category		
7	Detail address of a building	Address		
8	Total number of occupants	Occupancy Level		
9	A value of Longitude location based on ETRS89 coordinate reference system from OS AddressBase	Longitude		

Table 6: Features from DEC and LBSM (Public Buildings) [14,15]

No.	Features	Abbreviation	Unit	Label
10	The y location based on OS AddressBase	Easting	NA	Input
11	Air condition system	Aircon present		
12	Census Output Area classification	OAC		
13	Census Output Area	OA		
14	A value of Latitude location based on ETRS89 coordinate reference system from OS AddressBase	Latitude		
15	The x location based on OS AddressBase	Northing		
16	The code for the current borough based ONS Postcode Directory.	Ward Code		

Table 7 : Model Evaluation [16]

R-squared	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)
$R^2 = 1 - \sum_{m=1}^{n} \frac{(y_m - \widehat{y_m})}{(y_m - \bar{y}_m)^2}^2$	$MAE = \frac{1}{n} \sum_{m=1}^{n} \lvert y_m - \widehat{y_m} \rvert$	$MSE = \frac{1}{n} \sum_{k=1}^{n} (y_m - \widehat{y_m})^2$	$RMSE = \sqrt{\frac{1}{n} \sum_{m=1}^{n} (y_m - \widehat{y_m})^2}$
Measure how well the regression of the predicted values with the actual values	Measure the average magnitude of the target variable without considering the direction	The difference between predicted and actual values are squared before averaging the values	It brings the error metric back to the same unit as the predicted values

RESULTS

Impact of EPC, DEC, and LBSM Data Integration on Model Performance

Table 7: Model performance based on selected features (Domestic Buildings)

Model	Training Time (s)	R^2	MAE	MSE	RMSE
GB	**0.307**	**0.898**	**0.132**	**0.063**	**0.251**
XGBoost	0.270	0.886	0.147	0.070	0.265
DNN	567.600	0.846	0.138	0.094	0.308
RF	22.900	0.826	0.156	0.107	0.327
SVM	176.620	0.827	0.159	0.106	0.326
DT	6.195	0.743	0.210	0.158	0.397

Table 8: Model performance based on selected features (Public Buildings)

Model	Training Time (s)	R^2	MAE	MSE	RMSE
RF	**6.330**	**0.726**	**0.277**	**0.250**	**0.500**
GB	21.640	0.657	0.309	0.313	0.560
XGBoost	0.390	0.635	0.340	0.333	0.577
DT	0.028	0.598	0.321	0.367	0.606
SVM	19.310	0.563	0.400	0.399	0.632
DNN	228.600	0.420	0.470	0.520	0.720

Figure 4: Comparison of Actual versus Estimated Energy Consumption for Public and Domestic Buildings (2017–2023) Using RF

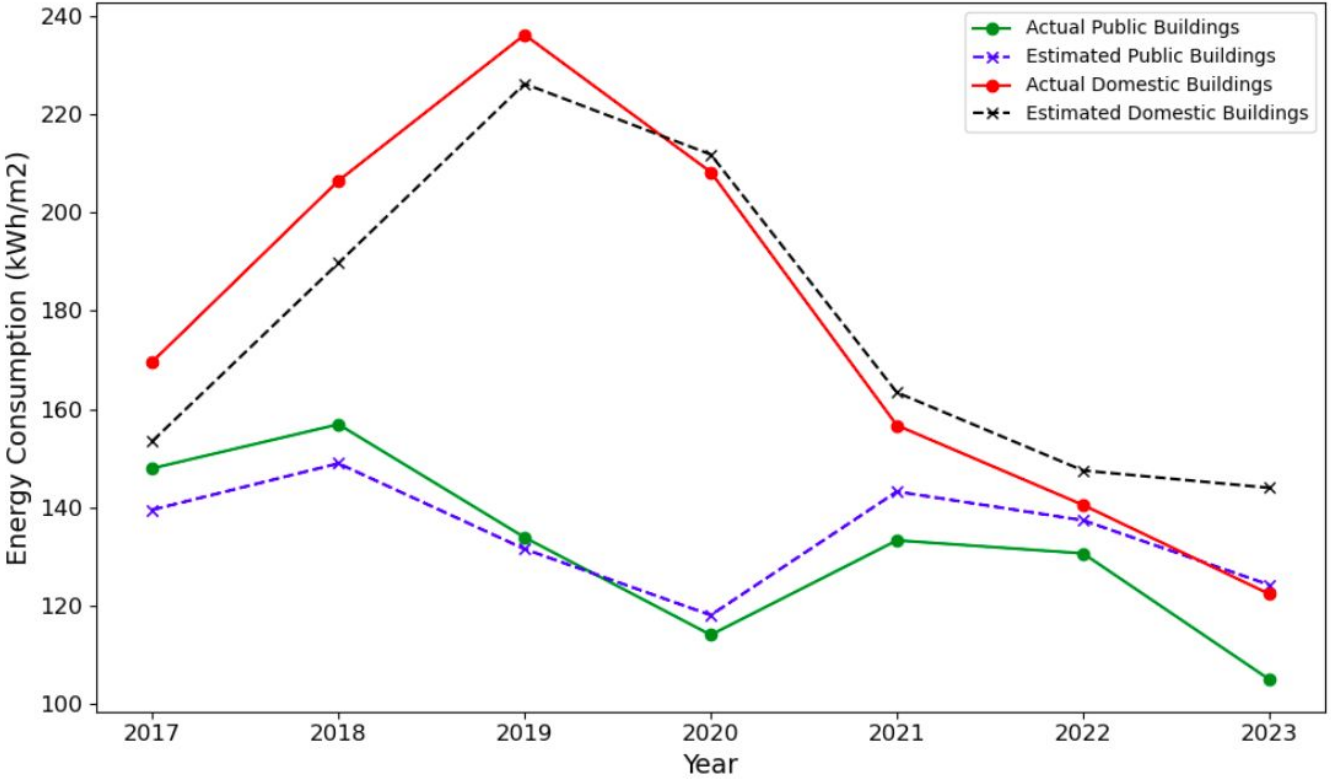
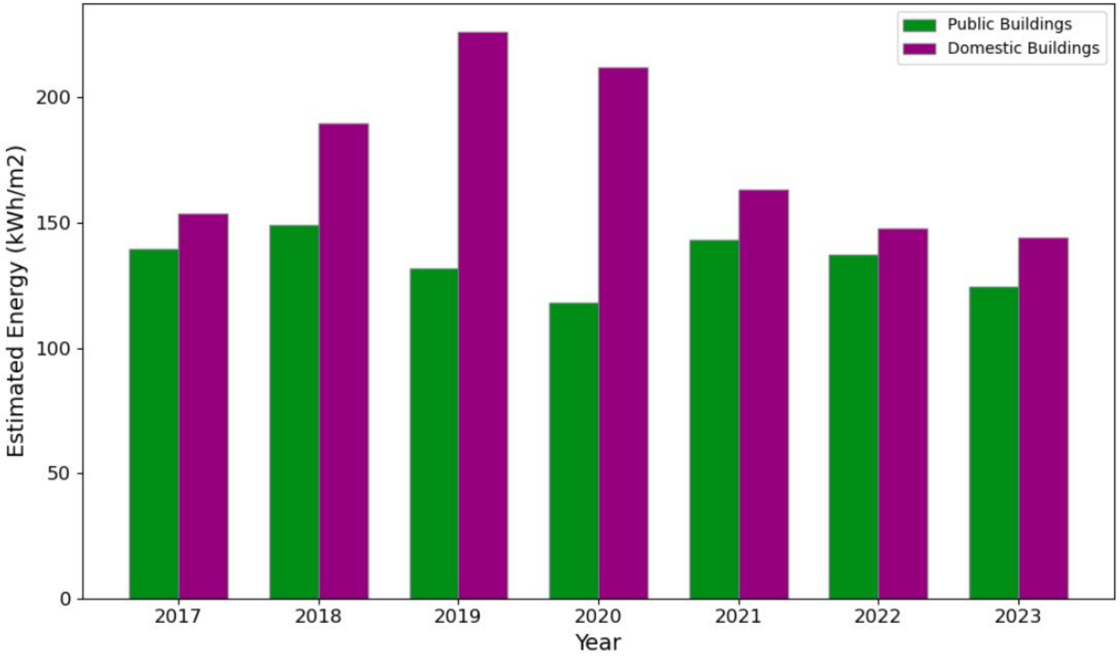


Figure 5: Estimated Energy Consumption for Public and Domestic Buildings (2017 – 2023) (Random Forest)



- Energy consumption for public buildings decreased slightly between 2018 to 2020 and keep below 150 kWh/m² after lockdown in 2020
- Meanwhile energy consumption for domestic building reached its peak in 2019 and reduced abruptly from more than 200 kWh/m² in 2019 to less than 200 kWh/m² starting in 2021

Analyzing Model Effectiveness Using Solely LBSM Data

Figure 6: Metrics comparison with different dataset (Domestic Building)

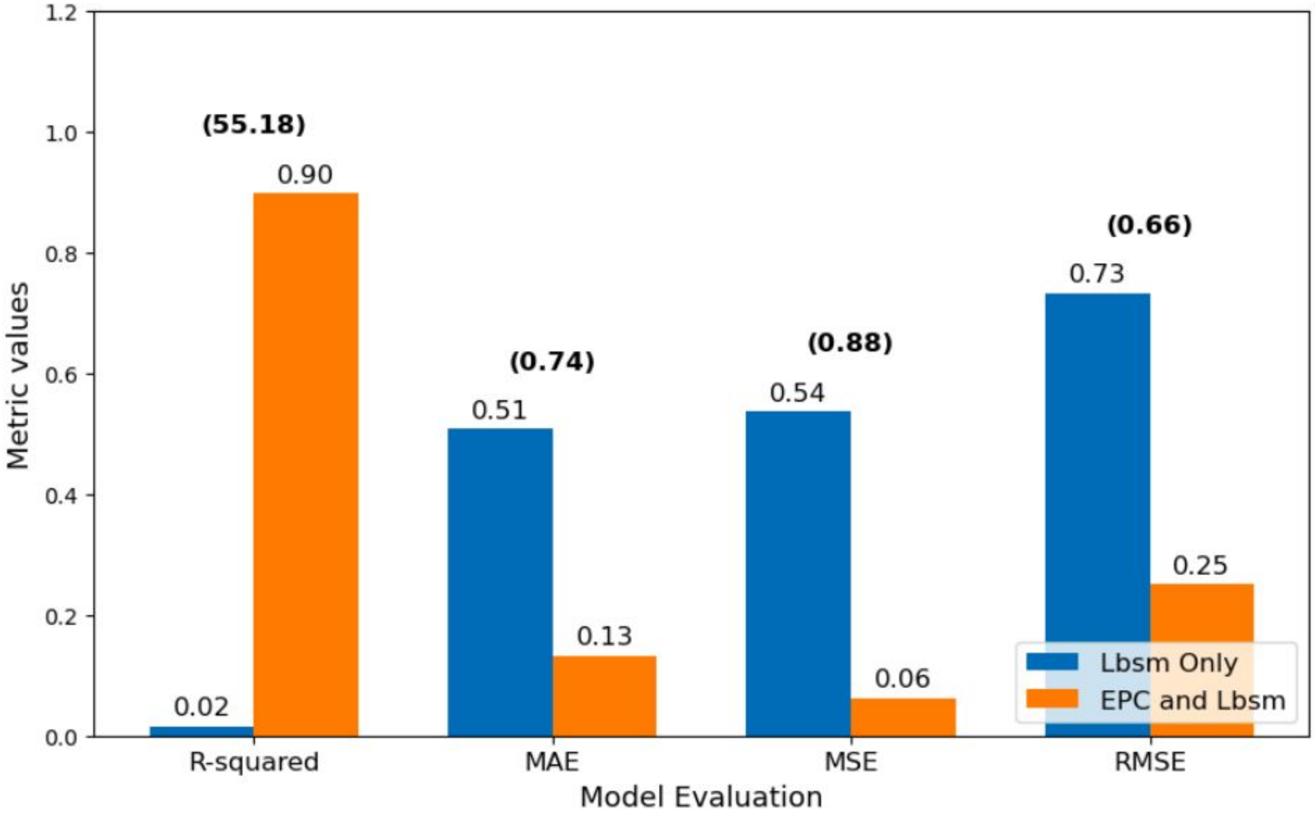
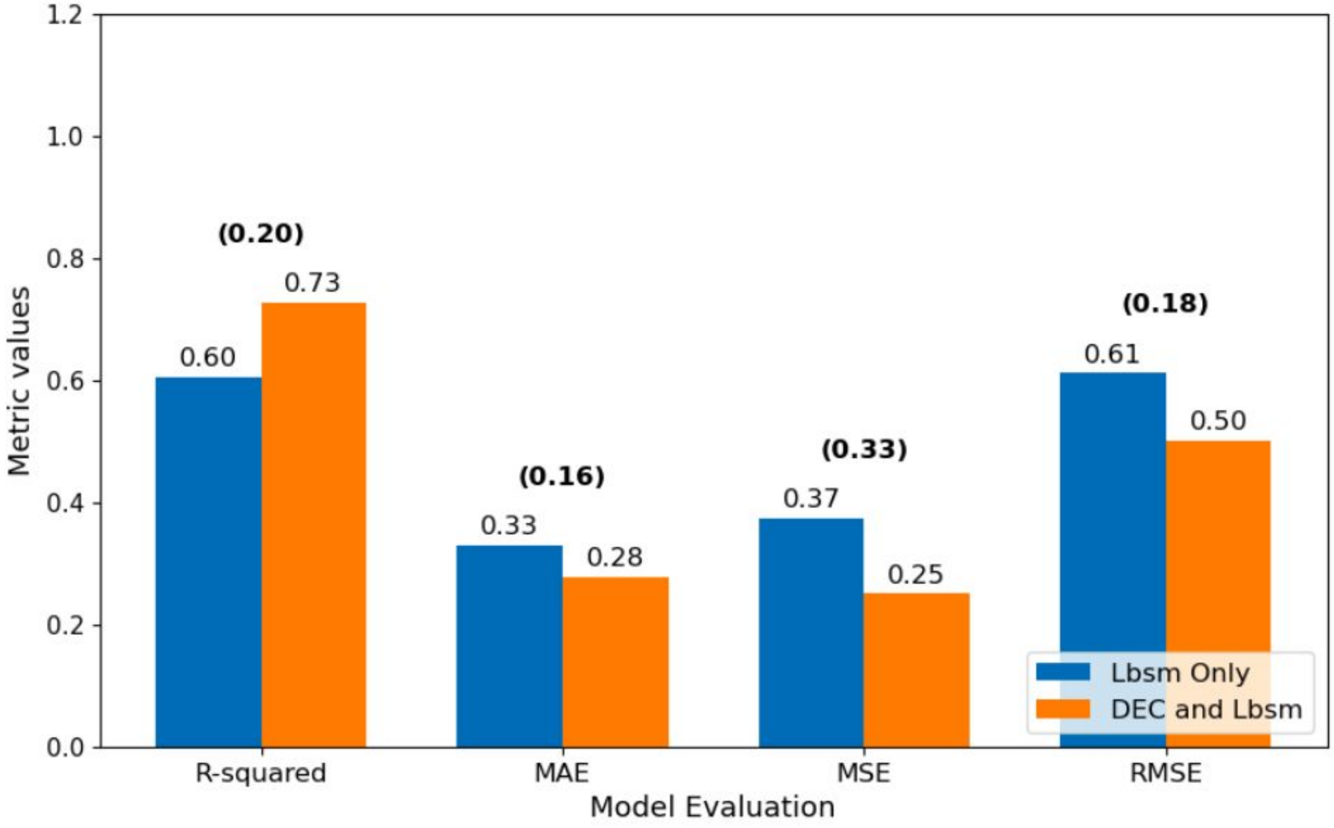
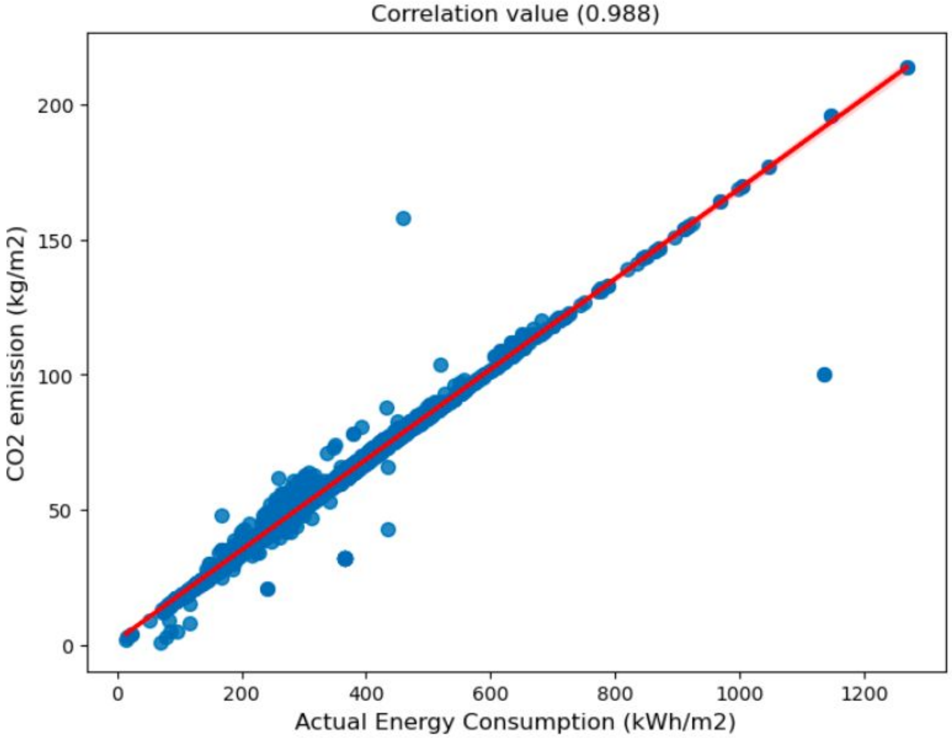


Figure 7: Metrics comparison with different dataset (Public Building)



A Comparative Analysis of Estimated Energy Consumption with CO2 Emissions

Figure 8: Correlation between actual energy consumption with CO2 emission of domestic building using regression plot



A Comparative Analysis of Estimated Energy Consumption with CO_2 Emissions

Figure 9: Correlation between estimated energy consumption with CO_2 emission of domestic building using regression plot

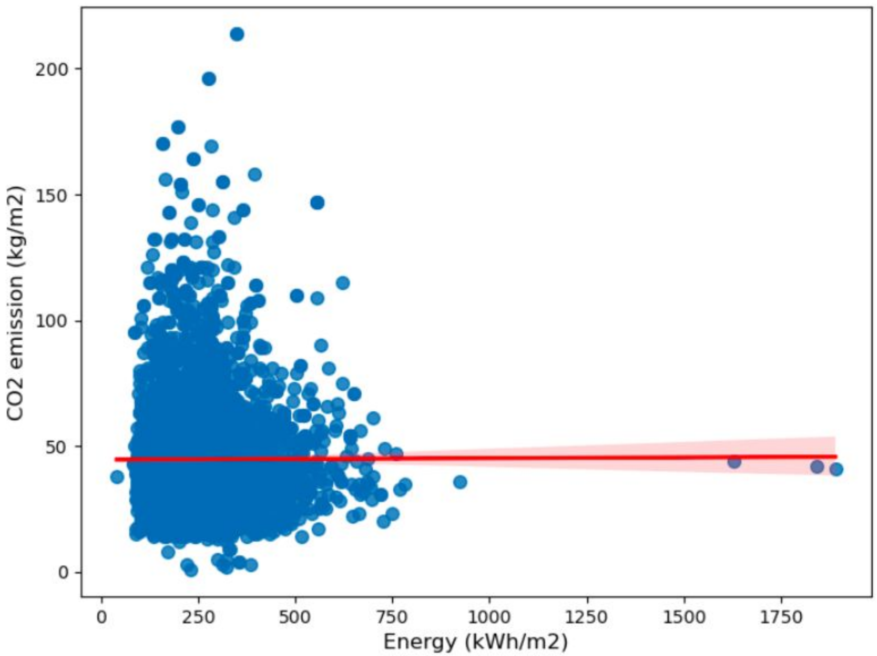


Figure 10: Correlation between estimated energy consumption with CO_2 emission of the domestic building using polynomial plot

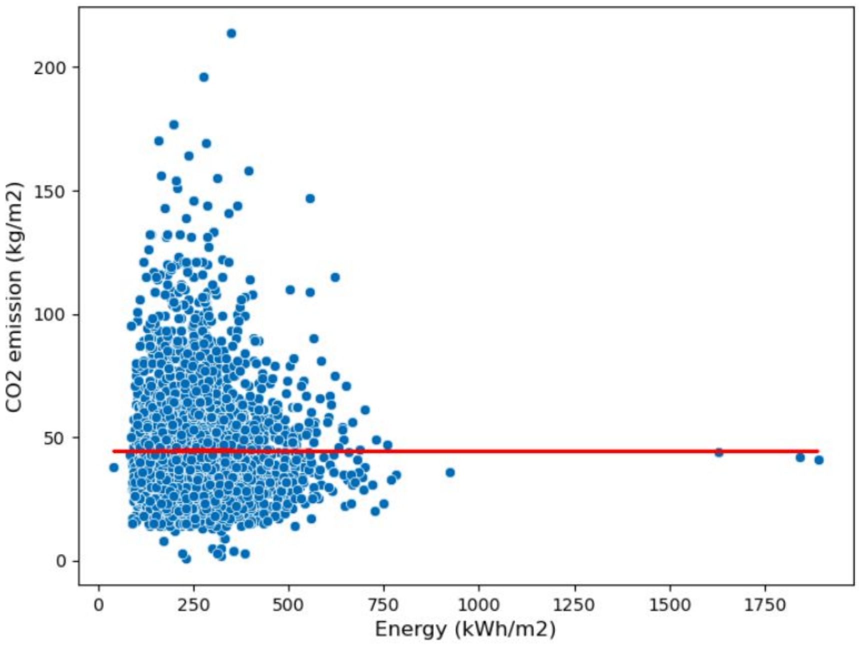


Figure 11: Correlation between estimated energy consumption with CO2 emission from Heating of public building

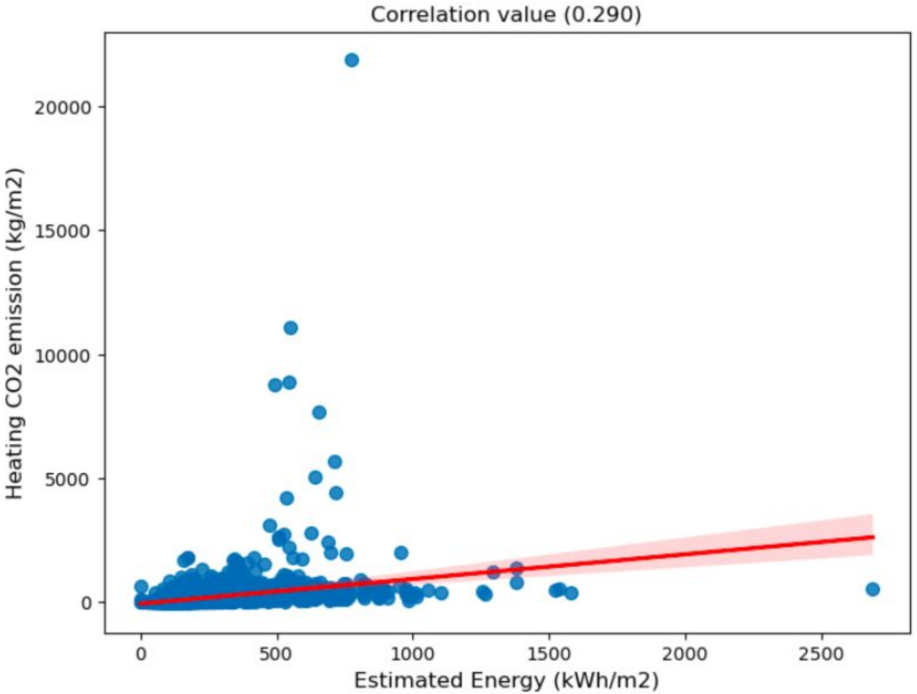
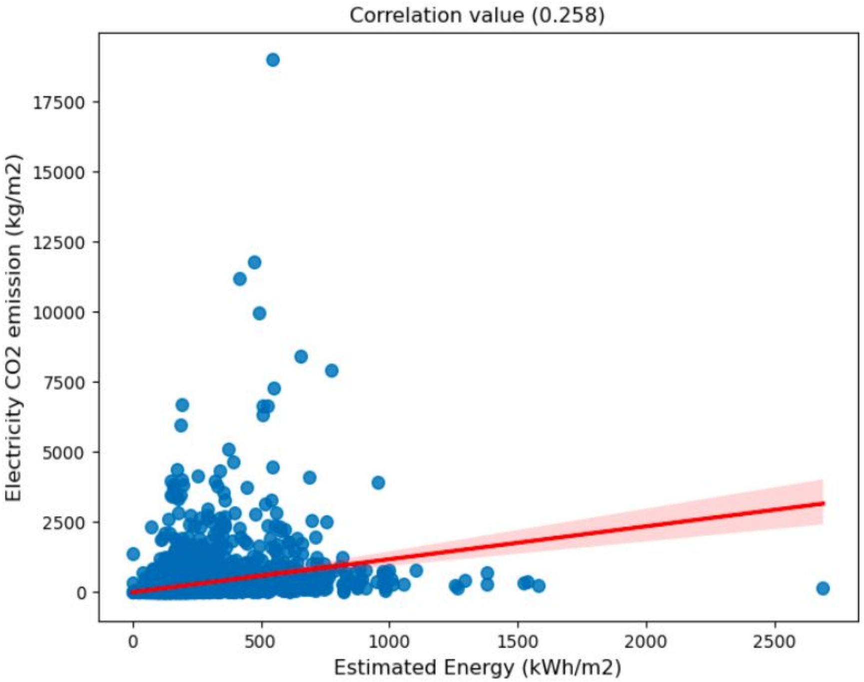


Figure 12: Correlation between estimated energy consumption with CO2 emission from Electricity of public building



CONCLUSION

- Due to population growth and urbanization, the estimated energy consumption for domestic building is relatively high between the year of 2017 until 2020.

- For future work, include environmental and behavioral factors may improve correlation between estimated energy consumption with the CO2 emission .

REFERENCES

[1] Tian, W., Song, J., & Li, Z. (2014). Spatial regression analysis of domestic energy in urban areas. Energy, 76, 629-640. https://doi.org/10.1016/j.energy.2014.08.057

[2] Department for Energy Security and Net Zero. (2024). Digest of UK Energy Statistics (DUKES) 2024: Chapters 1-7. UK Government.https://assets.publishing.service.gov.uk/media/66a7e14da3c2a28abb50d922/DUKES_2024_Chapters_1-7.pdf

[3] Poggi, F., & Amado, M. (2024). The spatial dimension of energy consumption in cities. *Energy Policy, 187*, 114023. https://doi.org/10.1016/j.enpol.2024.114023

[4] Jones, G. A., & Warner, K. J. (2016). The 21st century population-energy-climate nexus. *Energy Policy, 93*, 206-212. https://doi.org/10.1016/j.enpol.2016.02.044

[5] Ritchie, H., & Roser, M. (2023). Urbanization. *Our World in Data*. https://ourworldindata.org/urbanization

[6] Malikov, A. B. (2022). Regression based on decision tree algorithm. Bukhara State University, Republic of Uzbekistan.

[7] Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2020). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54, 1937–1967. https://doi.org/10.1007/s10462-020-09896-5:contentReference[oaicite:0]{index=0}

[8] Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, *15*, 3133-3181.

[9] Yang, H., Zhu, Y., & Liu, J. (2019). ECC: Platform-independent energy-constrained deep neural network compression via a bilinear regression model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11206–11215

[10] Yue, N., Caini, M., Li, L., Zhao, Y., & Li, Y. (2023). A comparison of six metamodeling techniques applied to multi building performance vectors prediction on gymnasiums under multiple climate conditions. *Applied Energy, 332*, 120481. https://doi.org/10.1016/j.apenergy.2022.120481.

[11] Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., & Ajayi, S. (2022). Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. Journal of Building Engineering, 45, 103406-. https://doi.org/10.1016/j.jobe.2021.103406

[12] Truong, L. H. M., Chow, K. H. K., Luevisadpaibul, R., Thirunavukkarasu, G. S., Seyedmahmoudian, M., Horan, B., Mekhilef, S., & Stojcevski, A. (2021). Accurate prediction of hourly energy consumption in a residential building based on the occupancy rate using machine learning approaches. Applied Sciences, 11(5), 2229. https://doi.org/10.3390/app11052229

[13] C. Fan, F. Xiao, Y. Zhao, A short-term building cooling load prediction method using deep learning algorithms, Appl. Energy 195 (2017) 222–233, https://doi.org/10.1016/j.apenergy.2017.03.064.

[14] Energy Performance Certificate Register. (n.d.). Domestic Energy Performance Certificates search. Open Data Communities. Retrieved [August 30 2024], from https://epc.opendatacommunities.org/domestic/search

[15] Steadman, P., Evans, S., Liddiard, R., Godoy-Shimizu, D., Ruyssevelt, P., & Humphrey, D. (2020). Building stock energy modelling in the UK: The 3DStock method and the London Building Stock Model. *Buildings and Cities, 1*(1), 100–119. https://doi.org/10.5334/bc.52

[16] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

Thank You