A Comparative Study on Machine Learning Algorithms for Assessing Energy Efficiency of Buildings

Christian Nnaemeka Egwim[1[0000-0003-2965-8749]], Oluwapelumi Oluwaseun Egunjobi[2[0000-0002-8799-2887]], Alvaro Gomes[2[0000-0003-1229-6243]], and Hafiz Alaka[1[0000-0003-3863-1046]]

[1]*Big Data Technologies and Innovation Laboratory, University of Hertfordshire, Hatfield, United Kingdom*
[2]*Energy for Sustainability (EFS), MIT-Portugal, Universidade de Coimbra, Coimbra, Portugal*

Abstract

An increase in energy demand in buildings continues to give rise to air pollution with a consequent impact on human health. To curb this trend, energy efficiency assessment plays a crucial role in helping to understand the energy in buildings and to recommend strategies to improve efficiency. Unfortunately, many existing approaches to assessing the energy efficiency of buildings are failing to do it accurately. Hence, the recommended energy efficiency strategies thereafter are failing to achieve the expected result. One approach in recent times uses data-driven predictive analytics techniques like machine learning (ML) algorithms to assess a building's energy efficiency towards improving its performance. However, as many ML algorithms exist, the selection of the right one is important for a successful assessment. Unfortunately, many of the existing works in this regard have simply adopted an ML algorithm without a justified rationale which may result in poor selection of the good performing ML algorithm. Therefore, in this study, a premise to compare the performance of ML algorithms for the assessment of energy efficiency of buildings was proposed. First, consolidated energy efficiency ratings of buildings from different data sources are used to develop predictive models using several ML algorithms. Thereafter, identification of best performing model was done by comparing evaluation metrics like RMSE, R-Squared, and Adjusted R-Squared. From the comparison, Extra Trees predictive model came top with RMSE, R-Squared, and Adjusted R-Squared of 2.79, 93%, and 93% respectively. This approach helps in the initial selection of suitable and better-performing ML algorithms.

Keywords: Buildings, Energy Efficiency, Machine Learning.

1 Introduction

More than 40% of carbon emissions are attributed to the consumption of energy in buildings [1]. According to Penistone [2], this high energy demand is due to the increasing number of building dwellers with corresponding population growth and growing appetite for energy-consuming appliances. Unfortunately, energy-related carbon emissions give rise to indoor and outdoor air pollution with corresponding negative impacts on human health. For example, Rural Affairs Committee [3] reports that in the UK, a considerable number of deaths are caused by poor air quality from carbon emissions. As such, in recent times several collaborations, policies, and strategies have been introduced by many developed countries to meet this goal. Among these policies are the EU's nearly-zero energy building proposal, requiring buildings from 2021 to have high energy performance. Another is the introduction of the issuance of energy certificates to promote energy efficiency awareness [4].

One strategy of enhancing the energy performance of buildings is improving their energy efficiency. Oliver and Peters [5] state that energy efficiency strategies alone have the potential to save 23.6 metric tons of carbon dioxide per year by 2030. However, despite the interventions by the government and other bodies at improving energy performance, many reports [6]–[10] indicate insufficient progress. As such, there is an urgent need to introduce new strategies or complement existing ones if building energy performance goals are to be met timely. A crucial step in improving energy efficiency is its assessment. In the light of this, a contemporary trend in research has emerged in which data-driven predictive analytics approaches are used to assess the energy efficiency of buildings towards making better decisions and choices in improving energy performance [11]. The predictive analysis utilizes Artificial Intelligence (AI)/ Machine Learning (ML) which has been widely adopted across other industries with records of tremendous successes [12], [13]. For example, it has been successfully employed in the healthcare industry for precise diagnosis and to make the best choice of treatment course from several alternatives. Likewise, in the transportation sector, it seats at the center of decisions for autonomous driving.

AI is a collection of state-of-the-art technologies that permit machines or any computer programme to sense, comprehend, act, and learn [14]. ML on the other hand is a branch of AI that allows computers to learn by a direct route from examples, data and experience. ML approaches to replace the traditional methods of programming that relied on hardcoded step by step rules [15]. This is done by giving the system a huge amount of data to learn from as a task, leaving it to decide how best to achieve the task in form of the desired output. Several ML algorithms such as Genetic Algorithm (GA), Artificial Neural Networks (ANN), Linear Regression (LR), Logistic Regression, Nearest-Neighbour Mapping, Decision Trees (DT), K-Means Clustering, Random Forests, Support Vector Machines, Principal Component Analysis, Singular Value Decomposition, among many others exist for implementation. Many research like [11], [16], [17] have already attempted the use of ML algorithms for predicting the energy efficiency of buildings.

The choice of which ML algorithm to use depends on several factors like ease of use, accuracy, the structure of the dataset, training time, among others. Likewise, outcomes and performances of different ML algorithms vary even when used against the same dataset due to several factors. The main influencing factors being the nature of the underlying ML algorithm, characteristics of the dataset regarding its size, resolution and data type, and the number of selected features. For example, Sha [18] comparative study of the performance of several ML algorithms in predicting cooling and consumption in buildings observed significant performance degradation from changing dataset resolution of training data from one (1) hour to six (6) minutes. In general, the LR algorithm which inherently only supports linear model is likely to perform better than DT when the feature set is many on a small dataset. Similarly, DT which employs non-parametric methods is likely to outperform ANN when the large training dataset is made up of categorical values data type. Therefore, considering the dilemma vis-a-vis the performance of ML algorithms, choosing a suitable ML algorithm is a tough and crucial decision towards its successful.

Unfortunately, many of the existing studies [11], [16], [17] have arbitrarily utilized or simply adopted various ML algorithms from previous research without rationale, resulting in poor performance, bad selection of good performing models or unenhanced generalizability of models developed from these ML algorithms across other regions. As a result, these studies have produced a knowledge vacuum that must be filled. Hence the need for a comparative study that will consolidate and evaluate the application of several ML algorithms in developing predictive models for assessing the energy efficiency of buildings. Thus, this study, therefore, aims to compare and evaluate the application of commonly employed ML algorithms used to develop models for assessing the energy efficiency of buildings. The following objectives will be:

1. Consolidate energy efficiency ratings of domestic and non-domestic buildings from different data sources into one database to establish the most applicable factors affecting the energy efficiency of buildings.
2. Utilize established factors in objective 1 as independent variables for all ML algorithms to develop predictive models.
3. Compare the performance of all ML algorithms against their respective predictive models.

The contribution of this study is therefore to fill the gap in the lack of a rationale in the selection of suitable ML algorithms for assessing the energy efficiency of buildings. For this work, due to availability and ease of access, energy data from the UK is utilized. Consequently, this is novel because the thorough review of the existing body of knowledge indicated that this is the first-time robust ML methods are employed to predict the energy efficiency of buildings in the UK. The same approach can be utilized for energy data from other countries. The outcome of our study will help in the initial choice of suitable ML for further predictive analysis. Furthermore, it will help to guide the decision of building construction managers, building dwellers, government bodies, and other concerned stakeholders in implementing strategies and employing measures for buildings energy performance improvement towards reduced carbon emissions and improved air quality.

2 Related Work

This section examines the aim, methodology, result, and analysis of the most recent related vast body of literature by numerous authors from across the world in the subject of energy consumption and optimization in buildings, as indicated in Table 1, using various or combined individual ML Algorithms.

Table 1. Survey of related literature

Author	ML Model	Methodology	Result
Mazzeo [19]	Artificial Neural Network (ANN) with Gargon Algorithm	ANN for flexible power system design to forecast energy performance of an energy community. In the approach, ANN is applied to large data set with dimensionless input variables to estimate energy performance indicators and grid indicator factors for the energy community.	The optimized ANN with 20 neurons produced the highest prediction accuracy with a global R of 0.9958 and P of 0.0004 in comparison with lower neurons.
Abediniangerabi, Makhmalbaf and Shahandashti [20]	Deep Learning Models, Gradient Boosting Machine, Random Forest, Generalized Linear Regression	ML models for the prediction of the energy performance of building façade system. The façade system considered are fiber reinforced concrete and conventional panels for making decisions to support energy efficient building vis-a-vis energy savings during early design stages. The accuracy of the result obtained was compared with other common prediction models.	The Deep Learning models in comparison to others had the best accuracy with MAE of 1.59 and RMSE of 3.48.
Maltais and Gosselin [21]	Artificial Neural Networks (ANN) with optimized parameters	ANN for the prediction of domestic hot water usage. The approach attempts to improve the accuracy of prediction of load demands from domestic water heating systems for the purpose of improving energy efficiency.	The ANN models which were tested with data from a 40-unit residential condominium of varying family sizes yielded good results with R2 of 0.88 but produced uncertainties for families with smaller water heating systems.
Alishahi, Nik-Bakht and Ouf [22]	Poisson Regression	Poisson Regression to study occupancy behavior in building using WIFI count data. The method attempts to provide an alternative approach as opposed to using sensor information from devices (like heat and ventilation systems) to obtain and integrate occupancy information to adapt to building operation for the purpose of increasing energy savings.	The system was validated using data obtained from an academic building in Canada. It produced a good prediction pattern with R2 of 0.98 during the week and 0.81 during weekends.

Sha, Moujahed and Qi [18]	Gradient Tree Boosting (GTB), Linear Regression, Rid Regression, Elastic Net (ELN), Multilayer Perceptron MLP), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN)	ML for predicting cooling loads and energy consumption in buildings. The work aimed at developing an approach for controlling and evaluating the performance of mechanical ventilators for reducing building cooling loads.	Data obtained from Building Automation System (BAS) from a high-rise building in Canada was applied to the ML models developed of which GTB produced the best accuracy with RMSE of 12.3%, 12.4% and 12.7% in 1 hour, 30 minutes and 6 minutes, respectively.
Mulero-Palencia, Álvarez-Díaz and Andrés-Chicote [23]	Decision Tree	Developed a tool for prototype diagnosis during design stage of building renovation. The system which is aimed at reducing emissions during building renovation helps to make critical decision and select better renovation alternatives.	The tool developed was tested using renovation buildings for different countries. The result obtained varied from country to country as the building state and government regulations were different from country to country.
Yigit [24]	Evolutionary Algorithm (EA), Gradient Boosting Machine (GBM)	ML to develop an energy simulation tool for optimized thermal design in residential buildings. The tool attempts to shorten the time required in optimization simulation so that simulation for larger buildings can be done faster. GBM was used as a surrogate model and DEAP, an evolutionary algorithm was used for optimization.	The surrogate model on test yielded R2 of 0.992 on cross-validation and 0.991 on testing. The result helps to make decision in selecting an alternative optimal energy design approach.
Alduailij [25]	Linear Regression, Dynamic Regression, ARIMA Time Series, Exponential Smoothing Time Series, Artificial Neural Network, Deep Neural Network	ML is used to detect consumption peaks in buildings. The system uses historic load demand curves to provides potential insights for making decisions towards energy saving, efficient use of appliances, and identification of demand response possibilities.	Data energy and weather data obtained from five (5) government buildings collected over 1 week uninterrupted were applied to the ML models. ARIMA yielded the highest accuracy of 98.91%.

Szul, Tabor and Pancerz [26]	BORUTA on Rough Set Theory (RST)	ML for features selection to forecast the heating energy demand rate of a building. The works aim to emphasize the need for care in the selection of model features. It also aims at providing insight for developing diverse approaches to improving energy efficiency in buildings.	The model which was tested using data from 109 multi-family buildings produced a satisfactory result with R2 between 0.81 and 0.85. 14 features were selected by the BORUTA algorithm and a further decrease in the number of features selected yielded no significant difference, hence confirming the feature selection.
Amasyali and El-Gohary [27]	Classification and Regression trees (CART), Ensemble Bagging trees (EBT), Artificial Neural Networks (ANN), and Deep Neural Networks (DNN)	ML for predicting energy consumption patterns in buildings while including occupancy behavior. The approach takes into consideration occupancy patterns to attain better accuracy in predicting energy consumption for the purpose of identifying potentials for energy savings.	A simulation of the model on EnergyPlus using 3 months of energy, building, weather, occupancy data with reliable performance and high accuracy emphasized the importance of the occupancy variable in the prediction algorithm.
Seyrfar [28]	Back-Propagation Neural Network (BPNN), Extreme Gradient Boosting (XGBoost), Random Forest (RF)	Combines energy, demographic, and socio-economic data to predict energy consumption in buildings. The approach aims at attaining higher accuracy and identify consumption patterns toward implementing energy efficiency measures and reducing carbon emission.	The ML models were simulated using data obtained from the United States (US) Consensus Bureau of which XGBoost had better performance with 68% accuracy.

3 Research Methodology

To consolidate energy efficiency ratings of domestic and non-domestic buildings as an approach to data collection, this study uses open data from the department of the energy performance of buildings data: England and Wales. Energy Performance Certificates (EPCs) for domestic and non-domestic buildings built, sold, or rented since 2008 were used. These data contain information on the energy efficiency ratings of domestic and non-domestic buildings during the energy assessment process. More precisely, this study uses all datasets from every constituency under the city of London local authority, consisting of property types: flat, bungalow, maisonette, house, and park-home; property total floor area ranging from one meter squared and hundred and ten meters squared; and finally with current EPC rating from A to G, (where A is very efficient, and G is the least efficient) lodged between April 2018 and April 2021. Table 2 describes the major features of this dataset used in this study.

Table 2. Dataset description

Feature ID	Features	Feature Type
F1	Energy consumption	Independent variable
F2	CO2 Emissions	Independent variable
F3	Lightning cost	Independent variable
F4	Heating cost	Independent variable
F5	Hot water cost	Independent variable
F6	Total floor area	Independent variable
F7	Floor level	Independent variable
F8	CO2 emissions per floor area	Independent variable
F9	Number of habitable rooms	Independent variable
F10	Number of heated rooms	Independent variable
F11	Hot water energy efficiency	Independent variable
F12	Hot water environmental efficiency	Independent variable
F13	Windows energy efficiency	Independent variable
F14	Windows environmental efficiency	Independent variable
F15	Walls energy efficiency	Independent variable
F16	Walls environmental efficiency	Independent variable
F17	Main heat energy efficiency	Independent variable
F18	Main heat environmental efficiency	Independent variable
F19	Lighting energy efficiency	Independent variable
F20	Lighting environmental efficiency	Independent variable
F21	**Energy efficiency of buildings**	**Dependent variable**

The raw dataset was extracted and downloaded as a comma-separated values file. To achieve the second objective of this study, this raw dataset was pre-processed into a clean dataset and analyzed by carrying out data imputation and outlier detection. Scaling and encoding feature engineering techniques were implemented to enable the selection of features or independent variables (see Table 2) to increase the predictive power (hyperparameter optimization) of the ML algorithms. The resulting clean and pre-processed dataset was split randomly into two in a ratio of 60% to 40% of the training dataset and testing dataset, respectively. Several ML algorithms were imported into a running instance of Jupiter Notebook using Scikit-learn - an integral Python programming language module with a broad spectrum of state-of-the-art algorithms for supervised and unsupervised medium-scale problems [29].

Since these ML algorithms fit independent variables (features) to a known dependent variable (target), supervised modeling taxonomy was undoubtedly chosen in this study. Additionally, because the target contains numerical data, regression analysis was used. Regression analysis is a type of predictive modeling approach that examines the connection between a target and feature(s) [30]. This is especially useful as it can express the degree to which one or more features have an influence on a target during ML predictions. There are a variety of regression algorithms that can be used to develop predictive models when experimenting with regression analysis. Which one to employ primarily depends on three factors – number of features, type of target, and shape of the regression line. Therefore, to mitigate any form of bias, we rather employed all regression algorithms that are available in scikit-learn version 0.23.1 at the time of this study for experimentations without any constraints on the previously mentioned factors. In concrete, a total of 42 regression algorithms available in this version was employed to develop the individual models using the training dataset (60% of the total dataset). This resulted in

42 developed regression models. Afterward, the unseen test dataset (40% of the total dataset) was used to evaluate the performance of these models that were developed. As the 42 models are all regressors, stratified k-fold, a variant of k-fold that returns stratified folds containing about the same proportion of target class as the initial dataset was used for cross-validation, where k=10, in order to avoid individual model overfitting on the dataset. Finally, Root Mean Square Error (RMSE), Coefficient of Determination (R-Squared) Adjusted Coefficient of Determination (Adjusted R-Squared) modeling evaluation metrics were employed to measure the several model performances on the testing dataset as shown in Fig. 1.

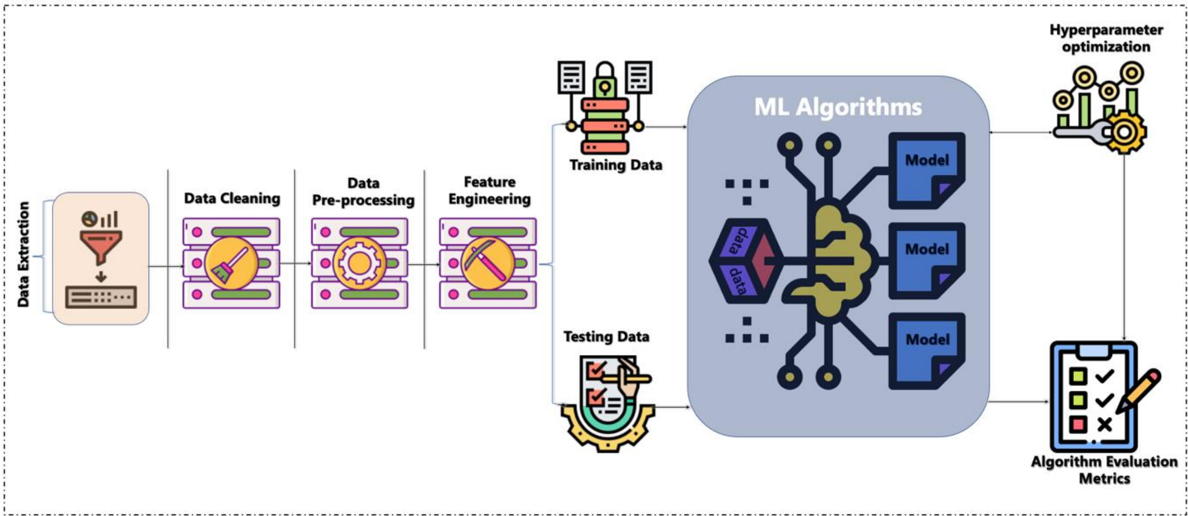

Fig. 1. ML Prediction Architecture

4 Analysis and Results

An initial investigation on the data through Exploratory Data Analysis (EDA) showed that the data is a two - dimensional array with 826 rows and 21 columns where the 1st to the 20th columns (F1 – F20 factor IDs) represent the features/independent variables and the 21st column (F21) represent the target/dependent variable. Outliers and missing values were detected and dropped thus resulting in a final 772 rows and 21 columns. Category columns F11 – F20 (see Table 2) were encoded into 1 (Very Poor), 2 (Poor), 3 (Average), 4 (Good), and 5 (Very Good). Furthermore, as a final transformation on the dataset, One-hot encoding (k-1 variant) a categorical encoding technique was used to transform all categorical datasets into a set of binary results (0 or 1). As most ML algorithms assume that any given dataset is normally distributed, with zero mean and unit variance, this study used the standardization feature scaling method to meet this requirement [29], [31]. This method involves subtracting the mean from each feature observation and dividing by the standard deviation as shown in the equation below:

$$X' = \frac{X - \bar{\mathbf{x}}}{\sigma} \qquad (1)$$

Where X' represents the standardized value; X a given feature observation; $\bar{\mathbf{x}}$ the mean and σ the standard deviation. Hence our resulting feature scaled dataset has its variance at 1, centered its mean at 0, and with a varying min-max value. Ultimately, a multivariate filter-based feature selection method called Spearman's rank correlation coefficient was implemented to evaluate the entire feature space, and eliminate obsolete, redundant, and noisy features, boost model accuracy, improve model interpretability, lower computational complexity and enhance generalizability. This Spearman's correlation coefficient is a non-parametric test used to determine the degree of connection between two or more features with a monotonic function, indicating a growing or decreasing relationship. The calculated strength between the features using Spearman's correlation coefficient fluctuates between +1 and 1, which happens when one feature is a perfect monotone function of the other. Thereafter, the dataset was split using the "train_test_split" function of Scikit-learn at a ratio of 60:40 for training and testing, respectively.

Consequently, after the described pre-processing, encoding, and standardization steps were implemented, the resulting training dataset (60% of the entire dataset) was utilized to train individual models in this study by fitting 42 ML algorithms (all regression algorithms available in scikit-learn version 0.23.1) to their respective models using their respective Scikit-learn libraries (see Table 3). This resulted in 42 developed models. Afterward, we used the test dataset (40% of the entire dataset) to evaluate the performance of these models that were developed. To mitigate the potential of these models' overfitting on the test dataset, a stratified10-fold cross-validation resampling technique was used to evaluate the performance of all the ML models developed using the 42 ML algorithms employed. The main parameters for each model used for hyperparameter optimization are alpha and lambda of values 100 and 10 respectively. These parameters were chosen to control the learning process as a way to apply regularization on each model for the bias-variance trade-off (low bias and low variance). The outcome of these assessments implemented on the test dataset is given as performance evaluation metrics for all models developed in this study (see Table 3). More precisely, it reveals the RMSE, R-Squared, and Adjusted R-Squared computed using the stratified10-fold cross-validation for the ML algorithms.

RMSE (see Equation 2) represents the standard deviation of the differences between the model predictions and the true values (training data). The closer the RSME value is to 0 the better the model.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})} \qquad (2)$$

R-Squared (see Equation 3) on the other hand represents the proportion of variance of target (dependent variable) that has been explained by the independent variables in the model. Its values range between 0 and 1 where 1 represent a perfect model and 0 a poor model.

$$\text{R} - \text{Squared} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (3)$$

Adjusted R-Squared (see Equation 4) is a modified and better version of R-Squared that considers the number of predictors (independent variables) in a given model.

$$\text{R} - \text{Squared}_{adjusted} = 1 - \left[\frac{(1 - R^2)(n - 1)}{n - k - 1}\right] \qquad (4)$$

Table 3. Algorithms, Models and Their Respective Performance Evaluation Metrics Implemented on the Test Dataset.

S/N	Algorithms	Model	Performance Evaluation Metrics		
			Adjusted R-Squared	R-Squared	RMSE
1	Extra-trees	ExtraTreesRegressor	0.93	0.93	2.79
2	Gradient boosting	GradientBoostingRegressor	0.91	0.92	3.05
3	Extreme gradient boosting	XGBRegressor	0.91	0.92	3.07
4	Histogram-based gradient boosting	HistGradientBoostingRegressor	0.91	0.91	3.15
5	Transformed target	TransformedTargetRegressor	0.90	0.91	3.16
6	Ordinary least square linear regression	LinearRegression	0.90	0.91	3.16
7	Linear least squares (with l2 regularization)	Ridge	0.90	0.91	3.18
8	Lasso linear model (with iterative fitting along a regularization path)	LassoCV	0.90	0.91	3.19
9	Bayesian ridge regression	BayesianRidge	0.90	0.91	3.19
10	Light Gradient Boosted Machine	LGBMRegressor	0.90	0.91	3.19
11	Elastic Net model (with iterative fitting along a regularization path)	ElasticNetCV	0.90	0.91	3.22
12	Generalized Linear Model (with a Poisson distribution)	PoissonRegressor	0.90	0.91	3.26
13	Ridge regression (with built-in cross-validation)	RidgeCV	0.90	0.90	3.27
14	Stochastic Gradient Descent	SGDRegressor	0.90	0.90	3.28
15	Random Forest	RandomForestRegressor	0.88	0.89	3.49
16	Huber Linear regression model	HuberRegressor	0.88	0.89	3.59
17	Lasso Lars Information Criterion	LassoLarsIC	0.87	0.88	3.65
18	Least Angle Regression model (cross-validated (CV))	LarsCV	0.87	0.88	3.72
19	Orthogonal Matching Pursuit model (OMP-CV)	OrthogonalMatchingPursuitCV	0.86	0.87	3.83
20	Linear Support Vector Regression	LinearSVR	0.86	0.87	3.87
21	Lasso Lars (CV)	LassoLarsCV	0.85	0.86	3.97
22	AdaBoost	AdaBoostRegressor	0.85	0.86	4.02

23	Bagging	BaggingRegressor	0.84	0.85	4.07
24	RANdom SAmple Consensus	RANSACRegressor	0.84	0.85	4.12
25	Lasso linear model	Lasso	0.81	0.82	4.47
26	Decision Tree	DecisionTreeRegressor	0.79	0.80	4.70
27	Linear regression (with combined L1 and L2 priors as regularizer)	ElasticNet	0.79	0.80	4.74
28	K-Nearest Neighbors	KNeighborsRegressor	0.78	0.80	4.76
29	Generalized Linear Model (with a Gamma distribution)	GammaRegressor	0.78	0.79	4.81
30	Generalized Linear Model	GeneralizedLinearRegressor	0.77	0.78	4.94
31	Generalized Linear Model (with a Tweedie distribution)	TweedieRegressor	0.77	0.78	4.94
32	Least Angle Regression model	Lars	0.73	0.75	5.29
33	Passive Aggressive Machine	PassiveAggressiveRegressor	0.71	0.73	5.53
34	Extremely Randomized Tree	ExtraTreeRegressor	0.71	0.73	5.53
35	Epsilon-Support Vector Machine	SVR	0.70	0.72	5.59
36	Orthogonal Matching Pursuit model (OMP)	OrthogonalMatchingPursuit	0.67	0.70	5.84
37	Nu Support Vector Machine	NuSVR	0.67	0.69	5.86
38	Dummy Estimator	DummyRegressor	-0.07	-0.00	10.59
39	Lasso Lars	LassoLars	-0.07	-0.00	10.59
40	Multi-layer Perceptron	MLPRegressor	-0.89	-0.77	14.08
41	Gaussian Process	GaussianProcessRegressor	-12.3	-11.43	37.33
42	Kernel ridge regression	KernelRidge	-44.25	-41.31	68.86

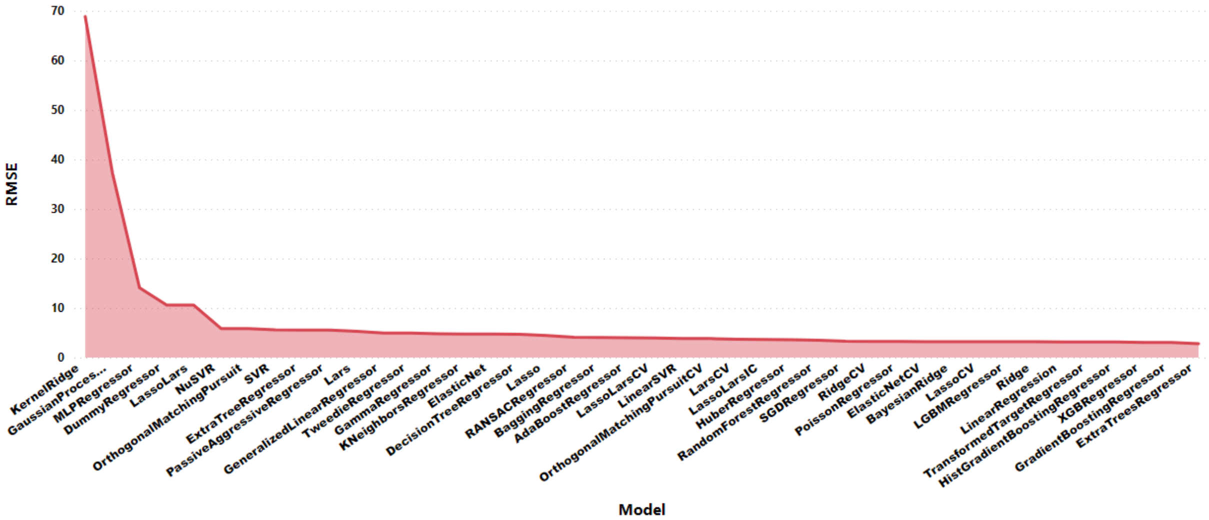

Fig. 2. Predictive Models by RSME

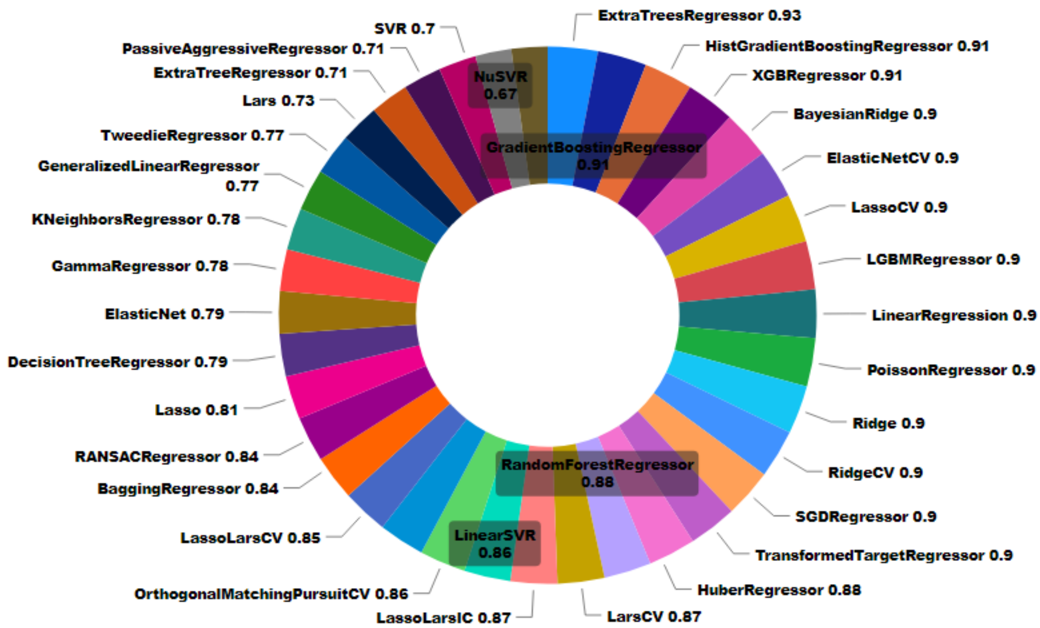

Fig. 3. Predictive Models by Adjusted R-Squared

5 Discussion of Results

Comparatively, looking through Table 3 and Fig. 3, Extra-trees predictive model came out as the top performant model having achieved an Adjusted R-Squared and R-Squared 0.93, and 0.93 respectively higher than the rest of the models. This implies a high correlation between the independent variable (F1 to F20) and the dependent variable F21. Also, Extra-trees's RSME value of 2.79 is the closest value to 0.00 (see Fig. 2), thus still making it the best performant model. This is an excellent performance level that is tending towards perfection and unachieved in any previous study, well justifying the need for trailing multiple algorithms when developing forecasting/predictive models. It was discovered that Dummy Estimator, Lasso Lars, Multi-layer Perceptron, Gaussian Process, and Kernel ridge regression models had Adjusted R-Squared and R-Squared values less than 0, hence are referred to as worst models in their descending order in this study. More so, these predictive models all had RSME values greater than other predictive models (see Fig. 2) and tending above value 0.00, well justifying them as worst models and therefore should be the least considered for predicting energy efficiency of a building. As Random Forest is built upon Decision Tree, it is no surprise based on the results of this study that

Random Forest was a better model than Decision Tree having achieved an Adjusted R-Squared and R-Squared 0.88, and 0.89 respectively better than Decision Tree.

Surprisingly, this study's top four performant predictive models (Extra-trees, Gradient boosting, Extreme gradient boosting, and Histogram-based gradient boosting) are all ensemble methods, which are machine learning methods that involve the use of multiple algorithms where the cumulative outcome from them is almost always greater in terms of predictive accuracy relative to the use of a single algorithm as they integrate decisions from different algorithms to maximize their overall performances [32]. Extra-trees belongs to the family of bagging ensemble method where multiple models of the same algorithm are used, however with different subsets of data selected randomly [33]. Gradient boosting, Extreme gradient boosting, and Histogram-based gradient boosting, on the other hand, belongs to the family of boosting ensemble method which is known as a repetitive technique that adapts the weight of the observation to the last grading. If an observation has been falsely categorized, the weight of this observation would be raised and conversely [34]. Thus, this study also proved these assertions. Also, interestingly, Multi-layer Perceptron, a class of feedforward artificial neural network (ANN) and one of the well-known and widely used algorithms by researchers were found among the four non-performing algorithms assessed in this study as the least models to consider for forecasting building energy efficiency. This is arguably due to the fact that although neural networks have been shown to approximate every continuously differentiable function, there is no assurance that a given network would ever learn this approximation given a specific weights initialization since, for example, the independent variables and the dependent variable used in this study are mostly continuous variables (see section 4). Thus, making their weight matrices susceptible to initial randomization.

Furthermore, in this study, only 20 features from a list of non-exhaustive features were used in the assessment of the energy efficiency of buildings. There are other existing works considering lesser or more features and even other sets of features. For example, Abediniangerabi, Makhmalbaf and Shahandashti [20] considered only six (6) feature classes which included weather and occupancy data asides from heating and cooling data. Feature selection and data representation play a key role in enhancing the performance of ML algorithms and it is being widely explored in representation learning. Similarly, in many cases, the choice of features is dependent on the available dataset obtained directly or computed from sensor data installed in buildings. However, since sensor installation and integration come with a cost, there is usually a trade-off between the number of sensors installed in buildings and the number of classes of data to be obtained. It will be of interest to have additional features like the comfort level indicator of building dwellers since building energy efficiency can only be sustained in a long term within the limits of these comfort levels. Unfortunately, in many buildings, this kind of data is difficult to obtain directly and accurately from sensors because of the ever-dynamic behavior of building occupants. There are several literatures exploring the use of long-term data from building appliances to obtain accurate results in this regard. However, in this study, it can be argued that this data (comfort level indicator) is indirectly encoded in some of the already selected 20-features. Give for instance, a building dweller will adjust operations of heating and cooling appliances to meet needs until at least comfort levels are met. While this argument may seem rational, it will still be important to carry-out comparative studies to evaluate how this feature or other features not included in this study impacts the performance of ML algorithms.

6 Conclusion and Recommendations

The rise in carbon emissions from the caused increase in energy demand from buildings is a major concern as it has continued to cause poor air quality with a consequent negative impact on human health across the globe. More so, efforts at curbing this trend have yielded insufficient results hence necessitating the need for more effective strategies. One of such contemporary strategies employs data-driven predictive analytics techniques to assess building energy efficiency to better explain contributing factors influencing its performance. In this approach that employs ML, the choice of ML algorithm is crucial to obtaining a good result. However, many existing research randomly selects a ML algorithm without justification.

In this study, therefore, a premise to compare the performance of machine learning algorithms for assessing the energy efficiency of buildings was proposed. To begin, this study consolidates energy efficiency ratings of domestic and non-domestic buildings from different data sources into one database as an approach to its quantitative data. The dataset in turn was used to train several ML (43 in number) algorithms to develop predictive models and evaluation metrics were computed. From the comparison of metrics for the different ML algorithms, the Extra Trees predictive model came out top having achieved an RMSE, R-Squared, and Adjusted R-Squared of 2.79, 93%, and 93% respectively.

Thus, this study highly recommends the need for initial predictive analysis for the selection of good performing model and better still the use of ensemble methods in predicting the energy efficiency of buildings. For example, from the result obtained, a choice of Extra Trees predictive model is justified being the best performing algorithm amongst others considered and as such may be further explored for even better result and implementation. Overall, the result from a study of this kind helps to build construction managers, building dwellers, government bodies, and other stakeholders to make better decisions towards improving the energy performance of buildings. However, while the proposed contemporary method of analysis is assumed to be applicable in assessing energy efficiency of buildings within the sector, the unique data transformation employed in this study may not, as typical of any data driven model, be transferable to the data from other regions. Furthermore, to obtain improved outcomes, asides including more features in the selection, representation learning can be employed for features extraction. Similarly, future studies should be targeted at extending the algorithms or optimizing already considered one.

References

[1] IEA and UNEP, "Global Status Report for Buildings and Construction," 2019.

[2] A. Penistone, "UK greenhouse gas emissions, provisional figures," *Natl. Stat.*, no. March, p. 46, 2019, Accessed: Jun. 29, 2021. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/790626/2018-provisional-emissions-statistics-report.pdf.

[3] R. Affairs Committee, "Air Quality and coronavirus: a glimpse of a different future or business as usual Fifth Report of Session 2019-21 Report, together with formal minutes relating to the report," 2021. Accessed: Jun. 29, 2021. [Online]. Available: www.parliament.uk.

[4] P. Ekins and E. Lees, "The impact of EU policies on energy use in and the evolution of the UK built environment," *Energy Policy*, vol. 36, no. 12, pp. 4580–4583, Dec. 2008, doi: 10.1016/j.enpol.2008.09.006.

[5] J. G. J. Olivier and J. A. H. W. Peters, "TRENDS IN GLOBAL CO 2 AND TOTAL GREENHOUSE GAS EMISSIONS 2019 Report," 2020. Accessed: Jun. 30, 2021. [Online]. Available: https://www.pbl.nl/sites/default/files/downloads/pbl-2020-trends-in-global-.

[6] J. Malinauskaite, H. Jouhara, L. Ahmad, M. Milani, L. Montorsi, and M. Venturelli, "Energy efficiency in industry: EU and national policies in Italy and the UK," *Energy*, vol. 172, pp. 255–269, Apr. 2019, doi: 10.1016/j.energy.2019.01.130.

[7] E. Brooks, A. Law, and L. Huang, "A comparative analysis of retrofitting historic buildings for energy efficiency in the UK and China," *DISP*, vol. 50, no. 3. Routledge, pp. 66–75, Jul. 03, 2014, doi: 10.1080/02513625.2014.979044.

[8] E. Marshall, J. K. Steinberger, V. Dupont, and T. J. Foxon, "Combining energy efficiency measure approaches and occupancy patterns in building modelling in the UK residential context," *Energy Build.*, vol. 111, pp. 98–108, Jan. 2016, doi: 10.1016/j.enbuild.2015.11.039.

[9] J. Rosenow, P. Guertler, S. Sorrell, and N. Eyre, "The remaining potential for energy savings in UK households," *Energy Policy*, vol. 121, pp. 542–552, Oct. 2018, doi: 10.1016/j.enpol.2018.06.033.

[10] O. Broad, G. Hawker, and P. E. Dodds, "Decarbonising the UK residential sector: The dependence of national abatement on flexible and local views of the future," *Energy Policy*, vol. 140, p. 111321, May 2020, doi: 10.1016/j.enpol.2020.111321.

[11] C. Benavente-Peces and N. Ibadah, "Buildings energy efficiency analysis and classification using various machine learning technique classifiers," *Energies*, vol. 13, no. 13, pp. 1–24, 2020, doi: 10.3390/en13133497.

[12] J. L. Blanco, S. Fuchs, M. Parsons, and M. J. Ribeirinho, "Artificial intelligence : Construction technology ' s next frontier," *Mckinsey Co.*, no. April, pp. 1–8, 2018, [Online]. Available: https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/artificial-intelligence-construction-technologys-next-frontier.

[13] M. Marks, "Construction: The next great tech transformation Voices Michael Marks," 2017.

[14] M. Goyal, "ARTIFICIAL INTELLIGENCE : A TOOL FOR HYPER PERSONALIZATION," *Int. J. 360 Manag. Rev.*, vol. 07, pp. 2320–7132, 2019.

[15] T. Royal Society, *MACHINE LEARNING: THE POWER AND PROMISE OF COMPUTERS THAT LEARN BY EXAMPLE*. 2017.

[16] I. Bilous, V. Deshko, and I. Sukhodub, "Parametric analysis of external and internal factors influence on building energy performance using non-linear multivariate regression models," *J. Build. Eng.*, vol. 20, pp. 327–336, Nov. 2018, doi: 10.1016/j.jobe.2018.07.021.

[17] M. Goyal, M. Pandey, and R. Thakur, "Exploratory Analysis of Machine Learning Techniques to predict Energy Efficiency in Buildings," in *ICRITO 2020 - IEEE 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*, Jun. 2020, pp. 1033–1037, doi: 10.1109/ICRITO48877.2020.9197976.

[18] H. Sha, M. Moujahed, and D. Qi, "Machine learning-based cooling load prediction and optimal control for mechanical ventilative cooling in high-rise buildings," *Energy Build.*, vol. 242, p. 110980, Jul. 2021,

doi: 10.1016/j.enbuild.2021.110980.
[19] D. Mazzeo *et al.*, "Artificial intelligence application for the performance prediction of a clean energy community," *Energy*, vol. 232, p. 120999, Oct. 2021, doi: 10.1016/j.energy.2021.120999.
[20] B. Abediniangerabi, A. Makhmalbaf, and M. Shahandashti, "Deep learning for estimating energy savings of early-stage facade design decisions," *Energy AI*, vol. 5, Sep. 2021, doi: 10.1016/j.egyai.2021.100077.
[21] L. G. Maltais and L. Gosselin, "Predictability analysis of domestic hot water consumption with neural networks: From single units to large residential buildings," *Energy*, vol. 229, p. 120658, Aug. 2021, doi: 10.1016/j.energy.2021.120658.
[22] N. Alishahi, M. Nik-Bakht, and M. M. Ouf, "A framework to identify key occupancy indicators for optimizing building operation using WiFi connection count data," *Build. Environ.*, vol. 200, p. 107936, Aug. 2021, doi: 10.1016/j.buildenv.2021.107936.
[23] S. Mulero-Palencia, S. Álvarez-Díaz, and M. Andrés-Chicote, "Machine Learning for the Improvement of Deep Renovation Building Projects Using As-Built BIM Models," *Sustainability*, vol. 13, no. 12, p. 6576, Jun. 2021, doi: 10.3390/su13126576.
[24] S. Yigit, "A machine-learning-based method for thermal design optimization of residential buildings in highly urbanized areas of Turkey," *J. Build. Eng.*, vol. 38, Jun. 2021, doi: 10.1016/j.jobe.2021.102225.
[25] M. A. Alduailij, I. Petri, O. Rana, M. A. Alduailij, and A. S. Aldawood, "Forecasting peak energy demand for smart buildings," *J. Supercomput.*, vol. 77, no. 6, pp. 6356–6380, Jun. 2021, doi: 10.1007/s11227-020-03540-3.
[26] T. Szul, S. Tabor, and K. Pancerz, "Application of the BORUTA algorithm to input data selection for a model based on rough set theory (RST) to prediction energy consumption for building heating," *Energies*, vol. 14, no. 10, p. 2779, May 2021, doi: 10.3390/en14102779.
[27] K. Amasyali and N. El-Gohary, "Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings," *Renew. Sustain. Energy Rev.*, vol. 142, p. 110714, May 2021, doi: 10.1016/j.rser.2021.110714.
[28] A. Seyrfar, H. Ataei, A. Movahedi, and S. Derrible, "Data-Driven Approach for Evaluating the Energy Efficiency in Multifamily Residential Buildings," *Pract. Period. Struct. Des. Constr.*, vol. 26, no. 2, p. 04020074, May 2021, doi: 10.1061/(asce)sc.1943-5576.0000555.
[29] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011, Accessed: Jan. 07, 2021. [Online]. Available: http://scikit-learn.sourceforge.net.
[30] M. Kuhn and K. Johnson, *Applied predictive modeling*. 2013.
[31] H. A. Alaka *et al.*, "Systematic review of bankruptcy prediction models: Towards a framework for tool selection," *Expert Systems with Applications*, vol. 94. Elsevier Ltd, pp. 164–184, Mar. 15, 2018, doi: 10.1016/j.eswa.2017.10.040.
[32] H. Badawi, F. Azais, S. Bernard, M. Comte, V. Kerzerho, and F. Lefevre, "Use of ensemble methods for indirect test of RF circuits: Can it bring benefits?," *LATS 2019 - 20th IEEE Lat. Am. Test Symp.*, no. 1, 2019, doi: 10.1109/LATW.2019.8704641.
[33] D. Opitz and R. Maclin, "Popular Ensemble Methods: An Empirical Study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, Aug. 1999, doi: 10.1613/jair.614.
[34] T. G. Dietterich, "Ensemble methods in machine learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2000, vol. 1857 LNCS, pp. 1–15, doi: 10.1007/3-540-45014-9_1.