Comparison of Machine Learning Algorithms for Evaluating Building Energy Efficiency Using Big Data Analytics

Abstract

Purpose – The study aims to compare and evaluate the application of commonly employed Machine Learning (ML) algorithms used to develop models for assessing energy efficiency of buildings.

Design/methodology/approach – The study foremostly combined building energy efficiency ratings from several data sources and utilised them to create predictive models using a variety of ML methods. Secondly, to test the hypothesis of ensemble techniques, this study designed a hybrid stacking ensemble approach based on the best performing bagging and boosting ensemble methods generated from its predictive analytics.

Findings – Based on performance evaluation metrics scores, the Extra Trees model was shown to be the best predictive model. More importantly, it demonstrated that the cumulative result of ensemble ML algorithms is usually always better in terms of predicted accuracy than a single method. Finally, it was discovered that stacking is a superior ensemble approach for analysing building energy efficiency than bagging and boosting.

Research limitations/implications – 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 other than the United Kingdom.

Practical implications – This study aids in the initial selection of appropriate and high-performing machine learning algorithms for future analysis. It also assists building managers, residents, government agencies, and other stakeholders in better understanding contributing factors and making better decisions about building energy performance. Furthermore, it will assist the general public in proactively identifying buildings with high energy demands, potentially lowering energy costs by promoting avoidance behaviour and assisting government agencies in making informed decisions about energy tariffs when this novel model is integrated into an energy monitoring system.

Originality/value – This study fills a gap in the lack of a reason for selecting appropriate machine learning algorithms for assessing building energy efficiency. More importantly, it demonstrated that the cumulative result of ensemble ML algorithms is usually always better in terms of predicted accuracy than a single method.

Keywords: Big Data Analytics, Buildings, Energy Efficiency, Machine Learning, Predictive Modeling

Article Type: Research paper.

1 Introduction

More than 40% of carbon emissions are attributed to the consumption of energy in buildings(Guo *et al.*, 2020; Nazir *et al.*, 2021). According to Penistone (2019), 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, (Affairs Committee, 2021) 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 (Ekins and Lees, 2008).

One strategy of enhancing the energy performance of buildings is improving their energy efficiency. Olivier and Peters (2020) 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 (Brooks, Law and Huang, 2014; Rosenow *et al.*, 2018; Koranteng, Simons and Essel, 2019; Malinauskaite *et al.*, 2019) 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 (Mostafaeipour *et al.*, 2019; Sert and Yazici, 2021; Atmaca, Şenol and Çağlar, 2022; He *et al.*, 2022). The predictive analytics utilizes Artificial Intelligence (AI)/ Machine Learning (ML) which has been widely adopted across other industries with records of tremendous successes (Marks, 2017; Blanco *et al.*, 2018; Ikediashi, Udo and Ofoegbu, 2020; Egwim, Alaka, Toriola-Coker, Balogun, Ajayi, *et al.*, 2021). 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 centre 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 (Goyal, 2019). 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 (Royal Society, 2017; Egwim, Egunjobi, *et al.*, 2021). 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 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 (Goyal, 2019; Benavente-Peces and Ibadah, 2020; Dadzie, Runeson and Ding, 2020; Dowlatshahi, Kuchaki Rafsanjani and Gupta, 2021; Qin and Wang, 2022) 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 (Sha, Moujahed and Qi, 2021) 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 implementation.

Unfortunately, many of the existing studies (Bilous, Deshko and Sukhodub, 2018; Benavente-Peces and Ibadah, 2020; Goyal, Pandey and Thakur, 2020) 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. 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 assess the energy efficiency of buildings. More so, it proved that the cumulative outcome from ensemble ML algorithms is almost always greater in terms of predictive accuracy relative to the use of a standalone algorithm. Finally, it proved stacking to be a better ensemble method than bagging and boosting for assessing energy efficiency of buildings. The outcome of our study will help in the initial choice of suitable ML for further predictive analytics. 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 Literature Review

A key technique to understand building energy efficiency is through accurate energy consumption predictions. This is important for energy conservation, better decision-making on lowering energy use, and the development of buildings that are not energy efficient (Egwim, Egunjobi, *et al.*, 2021). Several lines of evidence from vast body of knowledge suggests that the choices made during a building's lifecycle have a big impact on how energy-efficient it is. For instance, recent evidence from Mulero-Palencia, Álvarez-Díaz and Andrés-Chicote, (2021) who developed a tool for prototype diagnosis during design stage of building renovation for different countries suggests that by choosing the best design from a variety of options based on orientation, type, or shape can cut energy use by over 30% without incurring additional costs. Similarly other notable energy simulation tools such as COMFIE, EnergyPlus, BESTEST etc with comparatively satisfactory results have been used for analysing and modelling energy usage (Tsoka, 2015; Chiesa, Fasano and Grasso, 2021; Kosori *et al.*, 2022). However, along the same lines, Abediniangerabi, Makhmalbaf and Shahandashti, (2021) subsequently argued that there are certain limitations to using energy simulation tools, such as the requirement for a large number of specific building features that are frequently unavailable, like the interior occupancy loads, heating, ventilation and air conditioning system, etc.

These limitations of energy performance evaluation have piqued the interest of academics, who are continually studying new approaches to better comprehend building energy efficiency, resulting in new advancements for estimating energy consumption (Mazzeo *et al.*, 2021; Maltais and Gosselin, 2021; Alduailij *et al.*, 2021). A vital component of such advancements is the use of machine learning for energy contemporary predictive analytics having been widely adopted across different industries such as healthcare: aiding in diagnoses of patients using genetic data (Huang *et al.*, 2021; Malik, Khatana and Kaushik, 2021); manufacturing: use in managing workforces production process and allowing predictive maintenance (Chen *et al.*, 2021); education: virtual lectures (Bajaj and Sharma, 2018; Harmon *et al.*, 2021); finance: fraud detection (Iong-Zong Chen and Lai, 2021; Bao, Hilary and Ke, 2022), and transportation: self-driving autonomous cars (Manoharan, 2019; Ma *et al.*, 2020) among many others. Machine learning is a subset of artificial intelligence that analyses historical data to provide predictions and then utilises those predictions to guide decision-making (Balogun, Alaka and Egwim, 2021b). A broadly similar point has also recently been made by Sulaimon *et al.*, (2022), who argued that those decisions generate results which are used to improve future predictions. Drawing on the work of a wide range of philosophers, Olu-Ajayi *et al.*, (2022b) advances the notion that machine learning can make predictions from huge datasets, optimize utility functions extract hiding patterns and structures from datasets by classifying data thus helps the software program to learn and make predictions in the future.

A comprehensive review of literature showed that the field of machine learning is often classified into three broad categories: reinforcement learning, unsupervised learning, and supervised learning. More precisely, In reinforcement learning, model(s) developed by machine learning algorithm(s) learns by interacting with the environment and learns to take action to maximize the total reward (Su *et al.*, 2018). On the basis of these findings, Mahesh, (2018) proposes that in reinforcement learning, a software agent determines the ideal behaviour in a specific context for a particular problem. This view is supported by Schneckenreither, Haeussler and Peiró, (2022) who writes that the agent takes the input and decides the best action for the problem and then based on the result of the action the agent then receives simple reward feedback to allow it to learn from its behaviour. In unsupervised learning the outputs (labels) aren't known, as the models find patterns and structure from the data without any assistance (Greener *et al.*, 2022). Building on the work of Greener *et al.*, (2022), Tehrani *et al.*, (2022) argues that there is no instructor in unsupervised learning since it is based purely on local information, as such the model uses just the data supplied to its network to recognise emergent traits in the input dataset and then generate patterns from the available information without any pretrained data. In supervised learning, input(s) to the model including the example labels are known, and the model learns to generalise the output from these examples (El-Hasnony *et al.*, 2022).

Furthermore, Auletta *et al.*, (2022) in their findings claims that human instructor's expert is used to inform the model which outputs are correct and which are not. Supervised machine learning is further classified into classification and regression. In classification the output variable is a category while in regression the output variable is a number/value. For instance Amasyali and El-Gohary, (2021) used both classification and regression machine learning to predict energy consumption patterns in buildings while including occupancy behaviour. The approach takes into consideration occupancy patterns to attain better accuracy in predicting energy consumption for the purpose of identifying potentials for energy savings. Over a decade, vast body of literature (Seyedzadeh *et al.*, no date; Tsanas and Xifara, 2012; Goyal, 2019; Sun, Haghighat and Fung, 2020; Dowlatshahi, Kuchaki Rafsanjani and Gupta, 2021; Qin and Wang, 2022) have used one or more machine learning algorithm to predict energy efficiency, consumption, and optimization in buildings. For example, a good summary of regression of Artificial Neural Network (ANN) has been provided in the work of (Mazzeo *et al.*, 2021) who used 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. In the same vein Abediniangerabi, Makhmalbaf and Shahandashti, (2021) helped to establish an explanatory random forest model for the prediction of the energy performance of building façade system. The façade system considered are fibre 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.

Moreso, a significant analysis and discussion on the subject was presented by Alishahi, Nik-Bakht and Ouf, (2021) who used a regression model to study occupancy behaviour 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. Additionally, a key study comparing Extreme Gradient Boosting and Back Propagation Neural Network is that of Seyrfar *et al.*, (2021), in which they combined 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. However, a potential source of bias for these studies is the influence the researchers had upon the choice of the machine learning being employed. The lack of a formal method in the selection of the machine learning technique used was, specifically, the main limitation of the analysis performed in these studies. This is crucial because, abundant evidence from vast body of literature (Zhou *et al.*, 2017; Alaka *et al.*, 2018a; Jayatilake and Ganegoda, 2021; Priyadharshini *et al.*, 2022) has shown that, the type of machine learning algorithm affects the accuracy of its prediction as the performance of the various machine learning algorithms depends on the dataset and input features made available, thereby affecting the generalizability of its results. Given the uncertainty surrounding the performance of machine learning algorithms, selecting a suitable one for building energy efficiency prediction is a difficult but vital decision for its successful implementation. Thus, the knowledge gap that has to be addressed as a result of these investigations.

3 Research Methodology

Vast body of international literature ((Ding, Fan and Liu, 2021; Li and Yao, 2021; Zhou *et al.*, 2021; Olu-Ajayi *et al.*, 2022a)) have considered features such as CO_2 emissions, lightning, wall types, heating, floor level, among many others as crucial independent variables for energy use and efficiency assessment. 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 2011 were used because it enables independent research into energy efficiency issues such as fuel poverty and climate change. Furthermore, the data contains the unique property reference number: a unique identifier for every addressable location in Great Britain. These data contain information on the energy efficiency ratings (see Table 1) 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 2011 and April 2021. As shown in table 1, the building energy rating adheres to the UK standard rating and is given based on the building's level of energy usage. Table 2 describes the major features of this dataset used in this study.

Table 1. Energy performance ratings and limits

S/N	Energy Performance Limit	Energy Efficiency Rating	Remarks
1	92+	A	Very energy efficient
2	81 – 91	B	
3	69 – 80	C	
4	55 – 68	D	Moderate energy efficient
5	39 – 54	E	
6	21 – 38	F	
7	1 – 20	G	Least energy efficient

Table 2. Dataset description

Feature ID	Features	Description	Feature Type
F1	Energy consumption	Current estimated total energy consumption for the property in a 12-month period (kWh/m2). Displayed on EPC as the current primary energy use per square metre of floor area	Independent variable
F2	CO_2 Emissions	Carbon dioxide emissions per year in tonnes/year	Independent variable
F3	Lightning cost	Current estimated annual energy costs for lighting the property	Independent variable
F4	Heating cost	Current estimated annual energy costs for heating the property	Independent variable
F5	Hot water cost	Current estimated annual energy costs for hot water	Independent variable
F6	Total floor area	Total of all enclosed spaces measured to the internal face of the external walls, i.e., the gross floor area as measured in accordance with the guidance issued from time to time by the Royal Institute of Chartered Surveyors or by a body replacing that institution. (m^2)	Independent variable

F7	Floor level	Floor level relative to the lowest level of the property	Independent variable
F8	CO_2 emissions per floor area	Carbon dioxide emissions per square metre floor area per year in kg/m^2	Independent variable
F9	Number of habitable rooms	Habitable rooms include any living room, sitting room, dining room, bedroom, study and similar; and a non-separated conservatory	Independent variable
F10	Number of heated rooms	The number of heated rooms in the property if more than half of the habitable rooms are not heated	Independent variable
F11	Hot water energy efficiency	Hot water energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one-to-five-star rating	Independent variable
F12	Hot water environmental efficiency	Environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one-to-five-star rating	Independent variable
F13	Windows energy efficiency	Windows energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one-to-five-star rating	Independent variable
F14	Windows environmental efficiency	Windows environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one-to-five-star rating	Independent variable
F15	Walls energy efficiency	Walls energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one-to-five-star rating	Independent variable
F16	Walls environmental efficiency	Walls environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one-to-five-star rating	Independent variable
F17	Main heat energy efficiency	Main heat energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one-to-five-star rating	Independent variable
F18	Main heat environmental efficiency	Main heat environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one-to-five-star rating	Independent variable
F19	Lighting energy efficiency	Lighting energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one-to-five-star rating	Independent variable
F20	Lighting environmental efficiency	Lighting environmental efficiency rating. One of: very good; good; average; poor; very poor. On actual	Independent variable

		energy certificate shown as one-to-five-star rating	
F21	**Energy efficiency of buildings**	Energy required for space heating, water heating and lighting [in kWh/year] multiplied by fuel costs. (£/m²/year where cost is derived from kWh)	**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 techniques. 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 to eliminate or significantly minimise bias in training data for machine learning models. Additionally, splitting datasets into train and test sets creates a training dataset for the model to learn an effective mapping of inputs that produces good and effective outputs, while the test set effectively evaluates model performance. 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 (Pedregosa *et al.*, 2011). Furthermore, the experimentation was performed on an Apple MacBook Pro macOS Monterey version 12.4 with an Apple M1 chip, 16 gigabyte random access memory and 8 cores hardware.

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) (Kuhn and Johnson, 2013). 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 the developed models. 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 as cross-validation technique, 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.

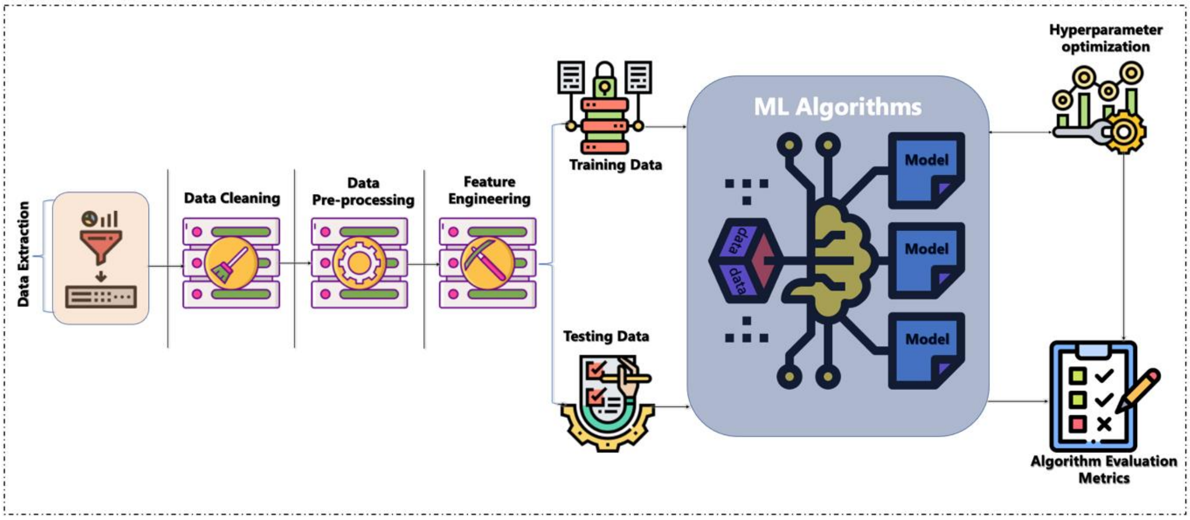

Fig. 1. ML Assessment Architecture

4 Analysis and Results

4.1 Data pre-processing

An initial investigation on the relatively big data used for predictive analytics in this study through Exploratory Data Analysis (EDA) showed that the data is a two - dimensional array with one million three hundred nineteen thousand one hundred twenty-two rows and twenty-one 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 (see Table 2). Also, by conducting descriptive statistics it was observed that some of the columns (columns F11 – F20) contain discrete categorical data with values ranging from Very Poor, Poor, Average, Good, to Very Good.

Furthermore, using their relative correlation coefficient values, correlation analysis was used to find multicollinearity among independent variables and the dependent variable (See Fig. 2). Figure 2 displays the cross correlation between each independent variable (F1 – F20) and the dependent variable (F21) in a correlation matrix plot. F14 and F15, for example, has a 0.50 positive correlation with F21, while F1 has a -0.81 negative correlation with F21, and so on. Multicollinearity is known as an absolute correlation coefficient of > 0.7 between two or more predictors (independent variables vs dependent variable) (Egwim, Alaka, Toriola-Coker, Balogun and Sunmola, 2021). As a result, multicollinearity thus exists between F8 and F1, F10 and F9, F17 and F11, among many others.

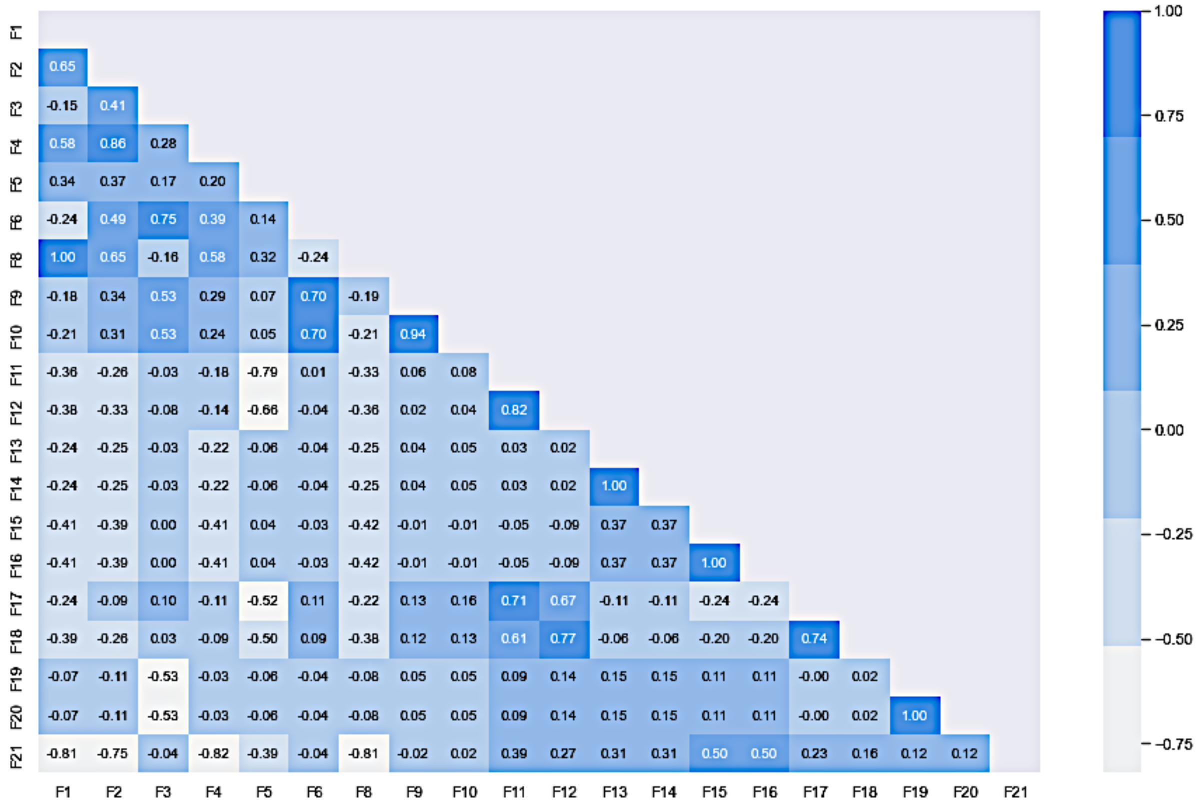

Fig. 2. Predictors Correlation Matrix Plot

4.2 Feature engineering

With a firm understanding of this big dataset obtained through EDA, it was observed that the raw data is unsuitable for model training as it is not normally distributed, it contains category features, outliers (noise) and have missing values. As a result, a transformation on the dataset was implemented using One-hot encoding (k-1 variant) a categorical encoding technique to transform all categorical datasets into ordinal values between one and five. More precisely, category columns F11 – F20 (see Table 2) were encoded into 1 (Very Poor), 2 (Poor), 3 (Average), 4 (Good), and 5 (Very Good). This dataset contains different building types (see Fig. 3) owned/ rented including detached, semi-detached, end-terrace, mid-terrace among many others.

Outliers and missing values (see Fig. 3) were detected and dropped thus resulting in a final one million two hundred seventy-nine thousand seven hundred ninety-three rows and twenty-one columns of dataset. Furthermore, due to their underlying assumptions that any given dataset is normally distributed with zero mean and unit variance, most ML estimators make this a standard condition (Pedregosa *et al.*, 2011; Alaka *et al.*, 2018b; Balogun, Alaka and Egwim, 2021a). Thus, as a final transformation on the dataset, standardization feature scaling method was employed to normalise the dataset. 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, centred its mean at 0, and with a varying min-max value.

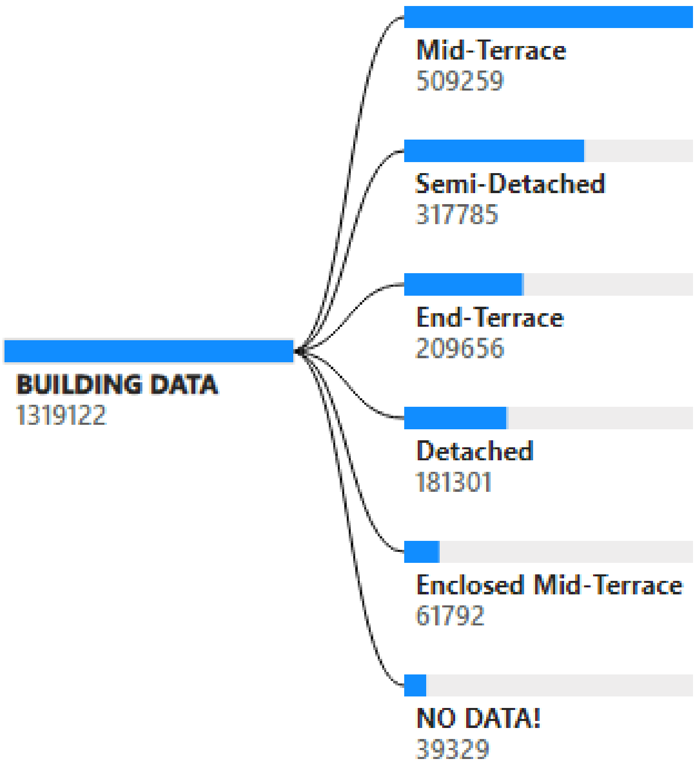

Fig. 3. A description of building types used and their respective quantities.

4.3 Feature selection

Given the availability of the relatively large/ high volume data extracted for this study, the significant number of features collected have the potential to influence the accuracy of the ML models to be developed in a harmful way or require a large amount of computation resources for their training thus needs careful consideration. For this reason, this study considered a multivariate filter-based feature selection method called Spearman's rank correlation coefficient 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 ranges between +1 and 1, which happens when one feature is a perfect monotone function of the other. The outcome of the entire feature space ranking by this feature selection method is as shown in Fig. 4 below. As the coefficient values vary from +1 to -1, with +1 denoting a perfect positive association, -1 denoting a perfect negative relationship, and 0 denoting no relationship, this study therefore selected only the features that tends closely towards a perfect relationship with the dependent variable (F21).

More precisely only the first five features (F16, F15, F13, F14, F19) with a relatively perfect positive relationship and the first five features (F8, F1, F4, F2, F5) with a relatively perfect negative relationship making a total of ten features out of the twenty features were finally considered to be used to develop the robust ML models in this study.

4.4 Model Development

Consequently, the resulting cleaned, pre-processed and feature engineered dataset (1279793 datapoints) was split randomly at a ratio of 60:40 for training and testing, respectively. 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).

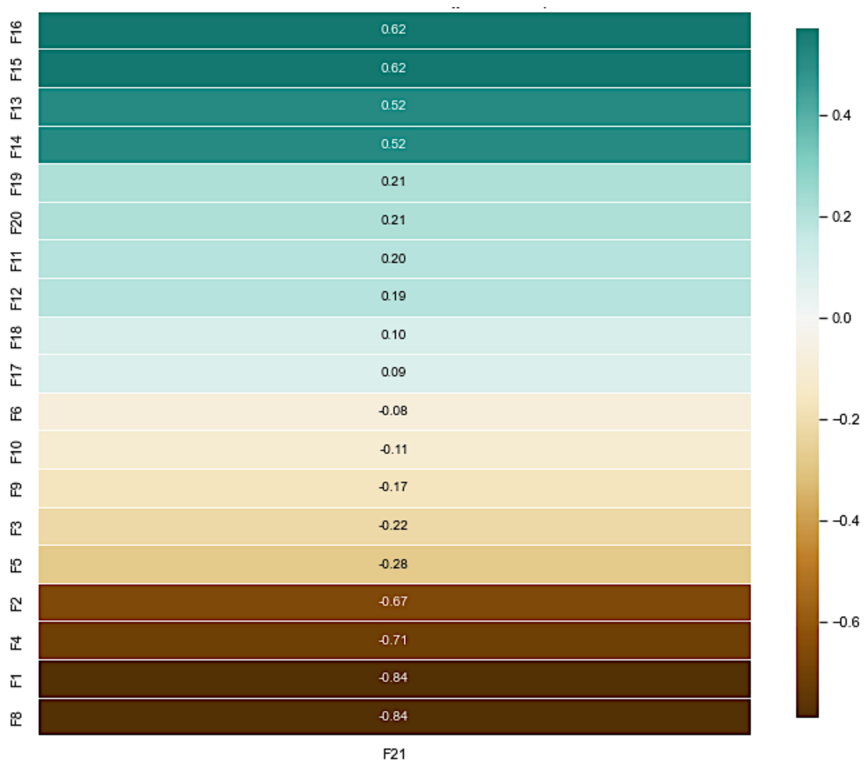

Fig. 4. Feature selection ranking according to Spearman

This resulted in 42 developed models. To mitigate the potential of these models' overfitting on the training 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. Then we proceeded the experimentation by tuning the hyperparameters to gain stability for each model. Afterward, we used the test dataset (40% of the entire dataset) to evaluate the performance of these models that were developed. The main parameters for each model used for hyperparameter optimization are as shown in column 4 of Table 3 below.

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 training and test dataset is given as performance evaluation metrics for all models developed in this study (see Table 3). More precisely, it reveals the Time Taken, RMSE, R-Squared, and Adjusted R-Squared computed for the ML algorithms during predictive modelling each on training and test dataset.

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.

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

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.

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

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.

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

Table 3. Algorithms, models, and their respective performance evaluation metrics

				Performance Evaluation Metrics							
				Training Dataset				Test Dataset			
S/N	Algorithms	Model	Hyperparameter	AR^2	R^2	RMSE	Time	AR^2	R^2	RMSE	Time
1	Extra-trees	ExtraTreesRegressor	Maximum depth = 6 Cost-complexity pruning = 3.4	0.95	0.95	2.77	0.09s	0.93	0.93	2.79	0.22s
2	Gradient boosting	GradientBoostingRegressor	Learning rate = 100.0 Maximum depth = 6	0.94	0.94	2.86	1.06s	0.91	0.92	3.05	0.11s
3	Extreme gradient boosting	XGBRegressor	Gamma = 10 Bootstrap = true	0.94	0.94	2.88	0.33s	0.91	0.92	3.07	0.05s
4	Histogram-based gradient boosting	HistGradientBoostingRegressor	L2 regularization = 1.8 Learning rate = 100.0	0.94	0.94	2.93	3.05s	0.91	0.91	3.15	0.40s
5	Transformed target	TransformedTargetRegressor	Regressor = linear regressor Check inverse = true	0.94	0.94	2.95	0.87s	0.90	0.91	3.16	0.01s
6	Ordinary least square linear regression	LinearRegression	Fit intercept = true Normalize = false	0.94	0.94	3.05	1.24s	0.90	0.91	3.16	0.01s
7	Linear least squares (with l2 regularization)	Ridge	Alpha = 2.6 Solver = auto	0.93	0.93	3.11	0.43s	0.90	0.91	3.18	0.01s
8	Lasso linear model (with iterative fitting along a regularization path)	LassoCV	Length of path = 0.007 Cross validation = 5	0.93	0.93	3.18	0.15s	0.90	0.91	3.19	0.08s
9	Bayesian ridge regression	BayesianRidge	Lambda = 0.000004 Maximum no of iterations = 56	0.91	0.91	3.59	0.08s	0.90	0.91	3.19	0.01s
10	Light Gradient Boosted Machine	LGBMRegressor	Learning rate = 100.0 Maximum depth = 6	0.89	0.89	4.10	0.01s	0.90	0.91	3.19	0.06s
11	Elastic Net model (with iterative fitting	ElasticNetCV	Precompute = auto Cross validation = 5	0.89	0.89	4.10	0.01s	0.90	0.91	3.22	0.08s

	along a regularization path)										
12	Generalized Linear Model (with a Poisson distribution)	PoissonRegressor	Alpha = 2.6 Maximum no of iterations = 600	0.89	0.89	4.10	0.01s	0.90	0.91	3.26	0.01s
13	Ridge regression (with built-in cross-validation)	RidgeCV	Normalize = False Cross validation = 5	0.89	0.89	4.10	0.03s	0.90	0.90	3.27	0.01s
14	Stochastic Gradient Descent	SGDRegressor	Shuffle = true Epsilon = 1.8	0.89	0.89	4.10	0.03s	0.90	0.90	3.28	0.02s
15	Random Forest	RandomForestRegressor	Maximum depth = 6 Minimum sample split = 4	0.89	0.89	4.10	0.01s	0.88	0.89	3.49	0.24s
16	Huber Linear regression model	HuberRegressor	Alpha = 2.6 Epsilon = 1.8	0.89	0.89	4.10	0.01s	0.88	0.89	3.59	0.04
17	Lasso Lars Information Criterion	LassoLarsIC	Normalize = False Precompute = auto	0.89	0.89	4.10	0.01s	0.87	0.88	3.65	0.02s
18	Least Angle Regression model (cross-validated (CV))	LarsCV	Cross validation = 5 Normalize = False Precompute = auto	0.89	0.89	4.10	0.01s	0.87	0.88	3.72	0.04s
19	Orthogonal Matching Pursuit model (OMP-CV)	OrthogonalMatchingPursuitCV	Fit intercept = true Cross validation = 5 Normalize = False	0.88	0.89	4.13	0.11s	0.86	0.87	3.83	0.01s
20	Linear Support Vector Regression	LinearSVR	Tolerance = 0.0006 Intercept scaling = 2.5	0.88	0.88	4.18	0.03s	0.86	0.87	3.87	0.01s
21	Lasso Lars (CV)	LassoLarsCV	Maximum no of iterations = 600 Cross validation = 5	0.88	0.88	4.18	0.07s	0.85	0.86	3.97	0.03s
22	AdaBoost	AdaBoostRegressor	Learning rate = 100.0 Loss function = linear	0.88	0.88	4.21	0.02s	0.85	0.86	4.02	0.12s
23	Bagging	BaggingRegressor	Bootstrap = true Maximum features = 10	0.88	0.88	4.27	0.02s	0.84	0.85	4.07	0.04s
24	RANdom SAmple Consensus	RANSACRegressor	Residual threshold = 3	0.87	0.87	4.31	0.26s	0.84	0.85	4.12	0.03s
25	Lasso linear model	Lasso	Alpha = 2.6 Normalize = False	0.87	0.87	4.37	0.05s	0.81	0.82	4.47	0.01s
26	Decision Tree	DecisionTreeRegressor	Splitter = random Cost-complexity pruning = 3.4	0.87	0.87	4.43	0.03s	0.79	0.80	4.70	0.01s

27	Linear regression (with combined L1 and L2 priors as regularizer)	ElasticNet	Precompute = auto Tolerance = 0.0006 Normalize = False Alpha = 2.6	0.87	0.87	4.46	0.02s	0.79	0.80	4.74	0.02s
28	K-Nearest Neighbors	KNeighborsRegressor	Number of neighbors = 20 Power = 0 Leaf size = 100	0.86	0.86	4.49	0.93s	0.78	0.80	4.76	0.02s
29	Generalized Linear Model (with a Gamma distribution)	GammaRegressor	Stopping criterion = 0.0002 Alpha = 2.6	0.86	0.86	4.50	1.04s	0.78	0.79	4.81	0.01s
30	Generalized Linear Model	GeneralizedLinearRegressor	Fit intercept = true Maximum no of iterations = 600	0.85	0.85	4.74	0.01s	0.77	0.78	4.94	0.01s
31	Generalized Linear Model (with a Tweedie distribution)	TweedieRegressor	Alpha = 2.6 Power = 0	0.83	0.83	5.06	0.01s	0.77	0.78	4.94	0.01s
32	Least Angle Regression model	Lars	Precompute = auto Normalize = False	0.82	0.82	5.11	0.02s	0.73	0.75	5.29	0.02s
33	Passive Aggressive Machine	PassiveAggressiveRegressor	Maximum step size = 8 Shuffle = true	0.82	0.82	5.16	0.01s	0.71	0.73	5.53	0.01s
34	Extremely Randomized Tree	ExtraTreeRegressor	Splitter = random Maximum depth = 6	0.81	0.81	5.33	0.01s	0.71	0.73	5.53	0.01s
35	Epsilon-Support Vector Machine	SVR	Gamma = auto Tolerance = 0.0006 Epsilon = 1.8	0.81	0.81	5.33	0.02s	0.70	0.72	5.59	0.03s
36	Orthogonal Matching Pursuit model (OMP)	OrthogonalMatchingPursuit	Tolerance = 0.0006	0.73	0.73	6.35	0.01s	0.67	0.70	5.84	0.01s
37	Nu Support Vector Machine	NuSVR	Shrinking = true Gamma = auto Cache size = 250	0.73	0.73	6.73	0.01s	0.67	0.69	5.86	0.02s
38	Dummy Estimator	DummyRegressor	Strategy = mean	-0.00	-0.00	12.18	0.01s	-0.07	-0.00	10.59	0.01s
39	Lasso Lars	LassoLars	Maximum no of iterations = 600 Precompute = auto Positive = true	-0.00	-0.00	12.18	0.01s	-0.07	-0.00	10.59	0.01s

40	Multi-layer Perceptron	MLPRegressor	Hidden layer size = 300 Alpha = 2.6 Learning rate = 100.0	-6.21	-5.83	16.86	0.03s	-0.89	-0.77	14.08	0.34s
41	Gaussian Process	GaussianProcessRegressor	Optimizer = callable Alpha = 2.6	-31.30	-31.24	68.79	0.55s	-12.3	-11.43	37.33	0.07s
42	Kernel ridge regression	KernelRidge	Degree = 7 Kernel = linear Alpha = 2.6	-316.62	-316.01	215.72	1.59s	-44.25	-41.31	68.86	0.02s

Key: **RMSE** = Root Mean Square Error, **R^2**= Coefficient of Determination (R-Squared), **AR^2** = Adjusted Coefficient of Determination (Adjusted R-Squared)

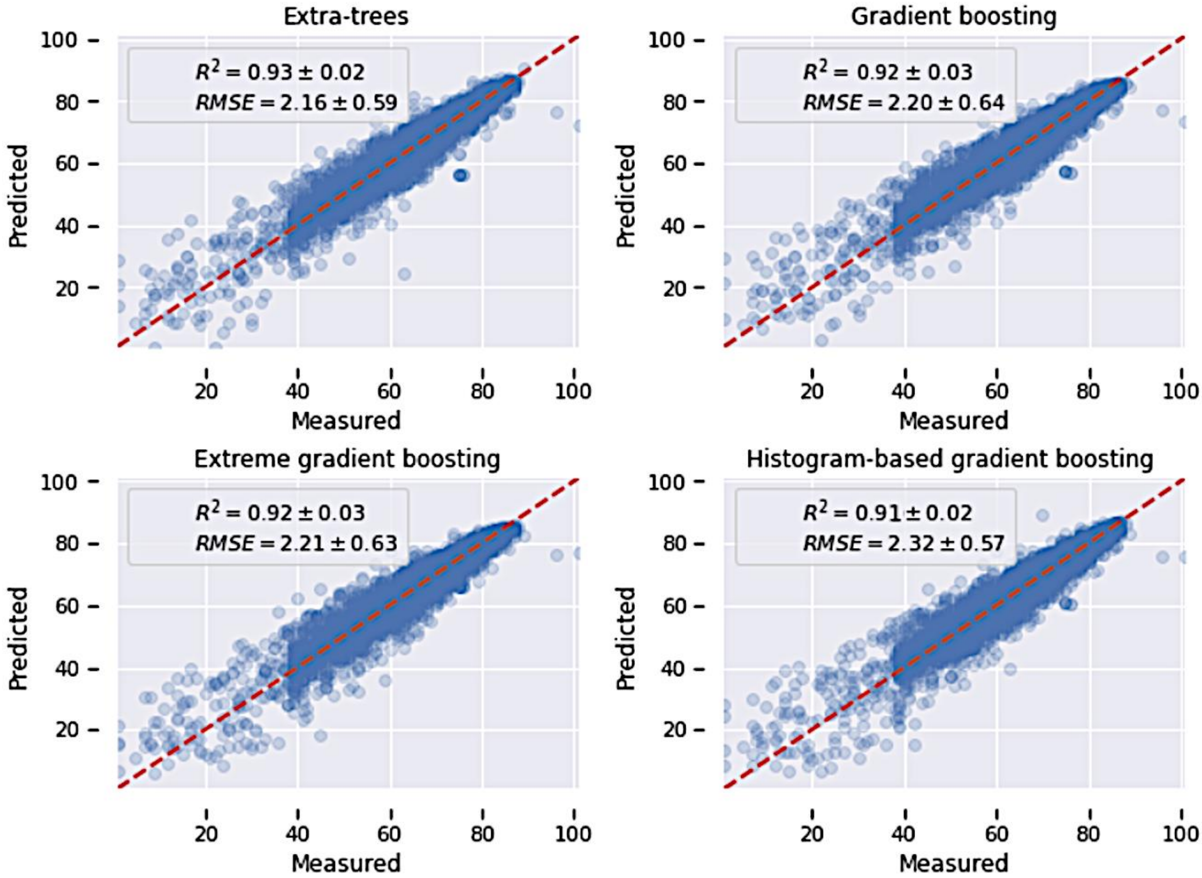

Fig. 5. Prediction plot of the top four performant predictive models

Given that the top four performant predictive models as shown in Fig. 5 (**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, this study further proceeded to test this hypothesis. Additionally, Fig. 5 shows there's a strong correlation between the individual model's predictions and its measured results respectively. More precisely it displays a direct (positive) relationship between the dependent variable and each of the ten independent variables considered for experimentation in this study. For a start we categorised them based on their respective ensemble method group. Extra-trees belong to the family of bagging ensemble method while Gradient boosting, Extreme gradient boosting, and Histogram-based gradient boosting belongs to the family of boosting ensemble method. **Bagging** (see Equation 5) is mathematically expressed by the following formula:

$$f_{bag} = f_1(x) + f_2(x) + \ldots + f_b(x) \qquad (5)$$

where the term on the left, f_{bag} is the bagged prediction, and $f_1(x)$ to $f_b(x)$ the actual learners (Extra-trees used in this study) are the term on the right. b represents the cumulative number of learners. Three key steps were used to experiment the **boosting** ensemble technique (see Equation 6 and 7). First, the target variable (energy efficiency of buildings) is predicted using an initial model f_0 with a residual (y – f_0). Secondly, a new model h_1 is fit to the previous step's residuals. Finally, f_0 and h_1 are merged to produce f_1, the boosted variant of f_0 as shown below:

$$f_1(x) < -f_0(x) + h_1(x) \qquad (6)$$

To boost f_1's results, we built a new model f_m based on f_1's residuals repeated for 'm' iterations until the residuals are as low as possible as shown below:

$$f_m(x) < -f_{m-1}(x) + h_m(x) \qquad (7)$$

The challenge, however, was how to identify the best boosting ensemble method since they all had the same **Adjusted R-Squared (0.91)**. To avoid bias and to enhance generalizability, we casted a vote amongst them using the hard and soft voting rule in Scikit-learn's VotingRegressor (See Fig. 6). Interestingly, Extreme gradient boosting emerged as the best boosting ensemble method estimator. Finally, we conclusively proceeded with the hypothesis testing by combining this Extreme gradient boosting ensemble method with the Extra-trees bagging ensemble method using stacked generalization (stacking) algorithm via scikit-learn's StackingRegressor whose base estimator algorithm used in this study are trained on heterogeneous ensemble machine learning algorithms such that base estimator's outputs are combined using a meta-classifier as shown below:

$$\min_{f} \sum_{i=1}^{n} l(f(x_i), y_i) + \lambda r(f) \qquad (8)$$

Where the first term in equation 8 is the empirical risk which is defined by a loss function S, that evaluates the effectiveness of the function f. The second term is the regularization term, and it evaluates the complexity of the function f, which is normally a norm of function f or its derivatives. Ultimately, this yielded a new high performant hyperparameter optimized hybrid stacking ensemble method that is higher in accuracy (**Adjusted R-Squared and R-Squared of 0.9487**) than the bagging and boosting ensemble methods based on its performance evaluation metrics as shown in Fig. 7 below. Looking through the learning curve of this multilayer high performant hyperparameter optimized hybrid stacking ensemble method in Fig. 7, we can observe a note of caution as regards its evaluation time. More specifically, although this novel hybrid stacking ensemble method have been found to have more predictive power over bagging and boosting ensemble methods for assessing energy efficiency of buildings, there is a trade-off for its time complexity.

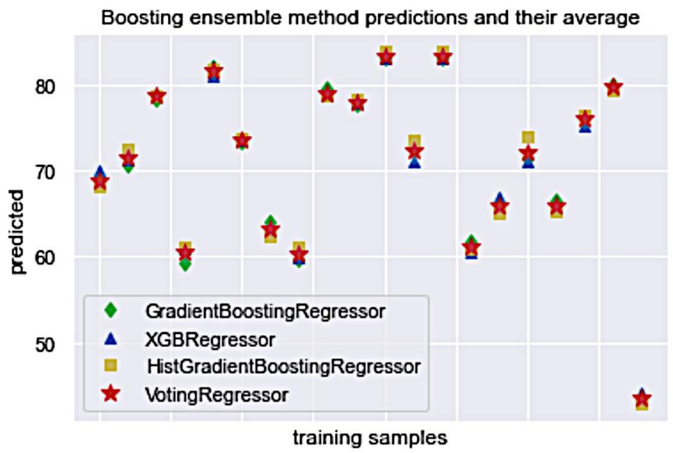

Fig. 6. Boosting ensemble method decision boundaries

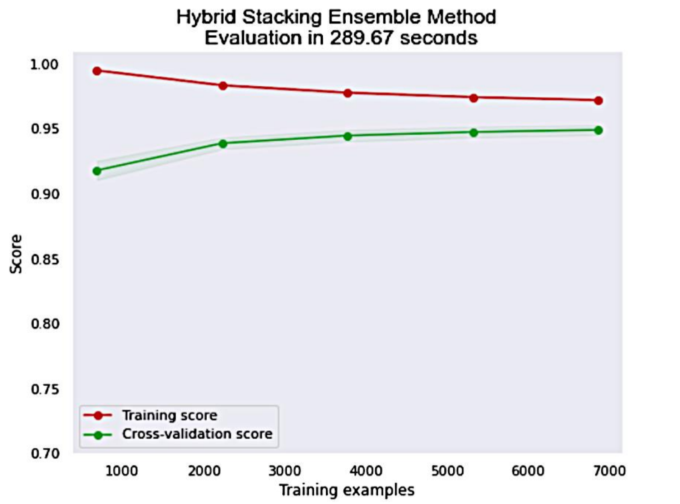

Fig. 7. Hybrid stacking ensemble method learning curve

5 Discussion

Comparatively, looking through Table 3 and Fig. 5, 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 compared to 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. 5), 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. 5) 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 (Badawi *et al.*, 2019). 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 (Opitz and Maclin, 1999). 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 (Dietterich, 2000). This gave the rationale to test this ensemble method hypothesis. Results from the hypothesis test analysis indeed yielded a new hybrid stacking ensemble method with a higher accuracy relative to individual bagging and boosting ensemble methods. 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.

More so, column 8 and 12 of Table 3 compares the training and testing stages of the 42 models developed in terms of computing time. In general, it can be observed that the calculation speeds in the training phase were slower than in the testing stages which is in line with the findings of Zhou (Zhou *et al.*, 2021). Interestingly, it can also be observed by looking through Fig. 7 that the novel multilayer high performant hyperparameter optimized hybrid stacking ensemble method calculation speed both at the training and validation stage were way too slow when benchmarked against any of the 42 individual models. This arguably suggest a trad-off between its predictive power and its training time complexity. Therefore, when adopting this unique paradigm, and as advised by Egwim (Egwim, 2017) researchers are highly encouraged to use the power of on-demand cloud computing platforms with a range of powerful clustered computers deployed across various datacenters throughout the world to decrease computational complexity. Given the relatively high volume datapoints (one million three hundred nineteen thousand one hundred) used in this study we can find that based on the results of this study, large amounts of data can result in lower estimation variance and, as a result, contribute to the predictive power of ML models. This fact is in line with vast body of knowledge. For instance, it is the viewpoint of Owolabi (Owolabi *et al.*, 2018) and Balogun (Balogun, Alaka and Egwim, 2021b) that more data means there's a better chance it'll include relevant information, which is beneficial as there is a natural desire to use these data assets by businesses to improve decision-making. Also, as briefly mentioned above, before testing the ensemble method hypothesis, this study addressed missing data and outliers and chose certain features from the feature space. All of this, along with the findings of other researches (Sharma and Garg, 2020; H. Sayadnavard, Toroghi Haghighat and Rahmani, 2022; Li, Chen and Shang, 2022), shows that merely having more data isn't a one-size-fits-all solution. More specifically, it's not only big data that helps us build high-performing predictive ML models; it's also high (quality) data. Thus, before bigdata may be useful in assessing energy efficiency of buildings, it is strongly advised to undertake exploratory data analysis to discover missing values and outliers, feature engineering, feature selection, and employ various algorithms on it.

Furthermore, in this study, only ten 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 (Abediniangerabi, Makhmalbaf and Shahandashti, 2021) considered only six (6) feature classes which included weather and occupancy data asides from heating and cooling data. 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 behaviour 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 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 (42 in number) algorithms to develop predictive models and evaluation metrics were computed to validate the results. Furthermore, this study tested the ensemble method hypothesis by using the best performing bagging and boosting ensemble methods derived from its predictive analytics as input estimators to develop a novel hyperparameter optimized hybrid stacking ensemble method. From the comparison of metrics for the different ML algorithms, the Extra Trees predictive model came out the top having achieved an RMSE, R-Squared, and Adjusted R-Squared of 2.79, 93%, and 93% while the hybrid stacking ensemble method predictive accuracy was higher relative to individual bagging and boosting ensemble methods respectively.

Thus, this study highly recommends the need for initial predictive analytics 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. The findings of this study will aid in the initial selection of appropriate ensemble ML algorithms for future predictive analytics. Additionally, this innovative approach can be applied by building designers as well as academics to anticipate the energy efficiency level (as shown in Table 1) and produce correct energy efficiency ratings. Furthermore, when this novel model is integrated into an energy monitoring system, it can assist the general public in proactively identifying buildings with high energy demands, potentially lowering energy costs by promoting avoidance behaviour and assisting government agencies in making informed decisions about energy tariffs. Overall, the result from a study of this kind can help 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 in future research. Similarly, future studies should be targeted at extending the algorithms or optimizing already considered one and validate the novel model with a real-world case scenario in order to comprehend its effectiveness.

References

Abediniangerabi, B., Makhmalbaf, A. and Shahandashti, M. (2021) 'Deep learning for estimating energy savings of early-stage facade design decisions', *Energy and AI*, 5. doi:10.1016/j.egyai.2021.100077.
Affairs Committee, R. (2021) *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*. Available at: www.parliament.uk. (Accessed: 29 June 2021).
Alaka, H.A. *et al.* (2018a) 'Systematic review of bankruptcy prediction models: Towards a framework for tool selection', *Expert Systems with Applications*, 94, pp. 164–184. doi:10.1016/j.eswa.2017.10.040.
Alaka, H.A. *et al.* (2018b) 'Systematic review of bankruptcy prediction models: Towards a framework for tool selection', *Expert Systems with Applications*. Elsevier Ltd, pp. 164–184. doi:10.1016/j.eswa.2017.10.040.
Alduailij, Mona A. *et al.* (2021) 'Forecasting peak energy demand for smart buildings', *Journal of Supercomputing*, 77(6), pp. 6356–6380. doi:10.1007/s11227-020-03540-3.
Alishahi, N., Nik-Bakht, M. and Ouf, M.M. (2021) 'A framework to identify key occupancy indicators for optimizing building operation using WiFi connection count data', *Building and Environment*, 200, p. 107936. doi:10.1016/j.buildenv.2021.107936.
Amasyali, K. and El-Gohary, N. (2021) 'Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings', *Renewable and Sustainable Energy Reviews*, 142, p. 110714. doi:10.1016/j.rser.2021.110714.
Atmaca, İ., Şenol, A. and Çağlar, A. (2022) 'Performance testing and optimization of a split-type air conditioner with evaporatively-cooled condenser', *Engineering Science and Technology, an International Journal*, 32, p. 101064. doi:10.1016/J.JESTCH.2021.09.010.
Auletta, F. *et al.* (2022) 'PREDICTING AND UNDERSTANDING HUMAN ACTION DECISIONS DURING SKILLFUL JOINT-ACTION VIA MACHINE LEARNING AND EXPLAINABLE-AI'.
Badawi, H. *et al.* (2019) 'Use of ensemble methods for indirect test of RF circuits: Can it bring benefits?', *LATS 2019 - 20th IEEE Latin American Test Symposium* [Preprint], (1). doi:10.1109/LATW.2019.8704641.
Bajaj, R. and Sharma, V. (2018) 'Smart Education with artificial intelligence based determination of learning styles', *Procedia Computer Science*, 132, pp. 834–842. doi:10.1016/J.PROCS.2018.05.095.
Balogun, H., Alaka, H. and Egwim, C.N. (2021a) 'Boruta-grid-search least square support vector machine for NO2 pollution prediction using big data analytics and IoT emission sensors', *Applied Computing and Informatics*, ahead-of-p(ahead-of-print). doi:10.1108/ACI-04-2021-0092.
Balogun, H., Alaka, H. and Egwim, C.N. (2021b) 'Boruta-grid-search least square support vector machine for NO2 pollution prediction using big data analytics and IoT emission sensors', *Applied Computing and Informatics*, ahead-of-p(ahead-of-print). doi:10.1108/ACI-04-2021-0092.
Bao, Y., Hilary, G. and Ke, B. (2022) 'Artificial Intelligence and Fraud Detection', in. Springer, Cham, pp. 223–247. doi:10.1007/978-3-030-75729-8_8.
Benavente-Peces, C. and Ibadah, N. (2020) 'Buildings energy efficiency analysis and classification using various machine learning technique classifiers', *Energies*, 13(13), pp. 1–24. doi:10.3390/en13133497.
Bilous, I., Deshko, V. and Sukhodub, I. (2018) 'Parametric analysis of external and internal factors influence on building energy performance using non-linear multivariate regression models', *Journal of Building Engineering*, 20, pp. 327–336. doi:10.1016/j.jobe.2018.07.021.
Blanco, J.L. *et al.* (2018) 'Artificial intelligence : Construction technology ' s next frontier', *Mckinsey & Company*, (April), pp. 1–8. Available at: https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/artificial-intelligence-construction-technologys-next-frontier.
Brooks, E., Law, A. and Huang, L. (2014) 'A comparative analysis of retrofitting historic buildings for energy efficiency in the UK and China', *DISP*. Routledge, pp. 66–75. doi:10.1080/02513625.2014.979044.
Chen, J. *et al.* (2021) 'Artificial intelligence-based human-centric decision support framework: an application to predictive maintenance in asset management under pandemic environments', *Annals of Operations Research*, pp. 1–24. doi:10.1007/S10479-021-04373-W/TABLES/11.
Chiesa, G., Fasano, F. and Grasso, P. (2021) 'A New Tool for Building Energy Optimization: First Round of Successful Dynamic Model Simulations', *Energies 2021, Vol. 14, Page 6429*, 14(19), p. 6429. doi:10.3390/EN14196429.

Dadzie, J., Runeson, G. and Ding, G. (2020) 'Assessing determinants of sustainable upgrade of existing buildings: The case of sustainable technologies for energy efficiency', *Journal of Engineering, Design and Technology*, 18(1), pp. 270–292. doi:10.1108/JEDT-09-2018-0148/FULL/PDF.
Dietterich, T.G. (2000) 'Ensemble methods in machine learning', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag, pp. 1–15. doi:10.1007/3-540-45014-9_1.
Ding, Y., Fan, L. and Liu, X. (2021) 'Analysis of feature matrix in machine learning algorithms to predict energy consumption of public buildings', *Energy and Buildings*, 249, p. 111208. doi:10.1016/J.ENBUILD.2021.111208.
Dowlatshahi, M.B., Kuchaki Rafsanjani, M. and Gupta, B.B. (2021) 'An energy aware grouping memetic algorithm to schedule the sensing activity in WSNs-based IoT for smart cities', *Applied Soft Computing*, 108, p. 107473. doi:10.1016/J.ASOC.2021.107473.
Egwim, C.N. (2017) 'A Cloud-Based C/C++ Compiler for Smart Devices'. Available at: http://repository.aust.edu.ng/xmlui/handle/123456789/4884 (Accessed: 24 November 2021).
Egwim, C.N., Egunjobi, O.O., *et al.* (2021) 'A Comparative Study on Machine Learning Algorithms for Assessing Energy Efficiency of Buildings', in *Communications in Computer and Information Science*. Springer, Cham, pp. 546–566. doi:10.1007/978-3-030-93733-1_41.
Egwim, C.N., Alaka, H., Toriola-Coker, L.O., Balogun, H. and Sunmola, F. (2021) 'Applied artificial intelligence for predicting construction projects delay', *Machine Learning with Applications*, 6, p. 100166. doi:10.1016/j.mlwa.2021.100166.
Egwim, C.N., Alaka, H., Toriola-Coker, L.O., Balogun, H., Ajayi, S., *et al.* (2021) 'Extraction of underlying factors causing construction projects delay in Nigeria', *Journal of Engineering, Design and Technology*, ahead-of-p(ahead-of-print). doi:10.1108/jedt-04-2021-0211.
Ekins, P. and Lees, E. (2008) 'The impact of EU policies on energy use in and the evolution of the UK built environment', *Energy Policy*, 36(12), pp. 4580–4583. doi:10.1016/j.enpol.2008.09.006.
El-Hasnony, I.M. *et al.* (2022) 'Multi-Label Active Learning-Based Machine Learning Model for Heart Disease Prediction', *Sensors 2022, Vol. 22, Page 1184*, 22(3), p. 1184. doi:10.3390/S22031184.
Goyal, M. (2019) 'ARTIFICIAL INTELLIGENCE : A TOOL FOR HYPER PERSONALIZATION', *International Journal of 360 Management Review*, 07, pp. 2320–7132.
Goyal, M., Pandey, M. and Thakur, R. (2020) '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)*. Institute of Electrical and Electronics Engineers Inc., pp. 1033–1037. doi:10.1109/ICRITO48877.2020.9197976.
Greener, J.G. *et al.* (2022) 'A guide to machine learning for biologists', *Nature Reviews Molecular Cell Biology*, pp. 40–55. doi:10.1038/s41580-021-00407-0.
Guo, J. *et al.* (2020) 'Extended TODIM method for CCUS storage site selection under probabilistic hesitant fuzzy environment', *Applied Soft Computing*, 93, p. 106381. doi:10.1016/J.ASOC.2020.106381.
H. Sayadnavard, M., Toroghi Haghighat, A. and Rahmani, A.M. (2022) 'A multi-objective approach for energy-efficient and reliable dynamic VM consolidation in cloud data centers', *Engineering Science and Technology, an International Journal*, 26, p. 100995. doi:10.1016/J.JESTCH.2021.04.014.
Harmon, J. *et al.* (2021) 'Use of artificial intelligence and virtual reality within clinical simulation for nursing pain education: A scoping review', *Nurse Education Today*, 97, p. 104700. doi:10.1016/J.NEDT.2020.104700.
He, L. *et al.* (2022) 'Optimization of energy-efficient open shop scheduling with an adaptive multi-objective differential evolution algorithm', *Applied Soft Computing*, p. 108459. doi:10.1016/J.ASOC.2022.108459.
Huang, S. *et al.* (2021) 'Artificial intelligence in the diagnosis of covid-19: Challenges and perspectives', *International Journal of Biological Sciences*, 17(6), pp. 1581–1587. doi:10.7150/IJBS.58855.
Ikediashi, D., Udo, G. and Ofoegbu, M. (2020) 'Post-occupancy evaluation of University of Uyo buildings', *Journal of Engineering, Design and Technology*, 18(6), pp. 1711–1730. doi:10.1108/JEDT-11-2019-0303/FULL/PDF.
Iong-Zong Chen, J. and Lai, K.-L. (2021) 'Deep Convolution Neural Network Model for Credit-Card Fraud Detection and Alert', *Journal of Artificial Intelligence and Capsule Networks*, 03(02), pp. 101–112. doi:10.36548/jaicn.2021.2.003.
Jayatilake, S.M.D.A.C. and Ganegoda, G.U. (2021) 'Involvement of Machine Learning Tools in Healthcare Decision Making', *Journal of Healthcare Engineering*, 2021. doi:10.1155/2021/6679512.
Koranteng, C., Simons, B. and Essel, C. (2019) 'Climate responsive buildings: a comfort assessment of buildings on KNUST campus, Kumasi', *Journal of Engineering, Design and Technology*, 17(5), pp.

862–877. doi:10.1108/JEDT-03-2019-0054/FULL/PDF.
Kosori, V. *et al.* (2022) 'Methodology to Determine Energy Efficiency Strategies in Buildings Sited in Tropical Climatic Zones; Case Study, Buildings of the Tertiary Sector in the Dominican Republic', *Energies 2022, Vol. 15, Page 4715*, 15(13), p. 4715. doi:10.3390/EN15134715.
Kuhn, M. and Johnson, K. (2013) *Applied predictive modeling*, *Applied Predictive Modeling*. doi:10.1007/978-1-4614-6849-3.
Li, C., Chen, Y. and Shang, Y. (2022) 'A review of industrial big data for decision making in intelligent manufacturing', *Engineering Science and Technology, an International Journal*, 29, p. 101021. doi:10.1016/J.JESTCH.2021.06.001.
Li, X. and Yao, R. (2021) 'Modelling heating and cooling energy demand for building stock using a hybrid approach', *Energy and Buildings*, 235, p. 110740. doi:10.1016/J.ENBUILD.2021.110740.
Ma, Y. *et al.* (2020) 'Artificial intelligence applications in the development of autonomous vehicles: A survey', *IEEE/CAA Journal of Automatica Sinica*, 7(2), pp. 315–329. doi:10.1109/JAS.2020.1003021.
Mahesh, B. (2018) 'Machine Learning Algorithms-A Review', *International Journal of Science and Research* [Preprint]. doi:10.21275/ART20203995.
Malik, M., Khatana, R. and Kaushik, A. (2021) 'Machine learning with health care: A perspective', in Jain, V. and Chatterjee, J.M. (eds) *Journal of Physics: Conference Series*. Cham: Springer International Publishing (Learning and Analytics in Intelligent Systems). doi:10.1088/1742-6596/2040/1/012022.
Malinauskaite, J. *et al.* (2019) 'Energy efficiency in industry: EU and national policies in Italy and the UK', *Energy*, 172, pp. 255–269. doi:10.1016/j.energy.2019.01.130.
Maltais, L.G. and Gosselin, L. (2021) 'Predictability analysis of domestic hot water consumption with neural networks: From single units to large residential buildings', *Energy*, 229, p. 120658. doi:10.1016/j.energy.2021.120658.
Manoharan, S. (2019) 'AN IMPROVED SAFETY ALGORITHM FOR ARTIFICIAL INTELLIGENCE ENABLED PROCESSORS IN SELF DRIVING CARS', *Journal of Artificial Intelligence and Capsule Networks*, 01, pp. 95–104. doi:10.36548/jaicn.2019.2.005.
Marks, M. (2017) *Construction: The next great tech transformation Voices Michael Marks*.
Mazzeo, D. *et al.* (2021) 'Artificial intelligence application for the performance prediction of a clean energy community', *Energy*, 232, p. 120999. doi:10.1016/j.energy.2021.120999.
Mostafaeipour, A. *et al.* (2019) 'Energy efficiency for cooling buildings in hot and dry regions using sol-air temperature and ground temperature effects', *Journal of Engineering, Design and Technology*, 17(3), pp. 613–628. doi:10.1108/JEDT-12-2018-0216/FULL/PDF.
Mulero-Palencia, S., Álvarez-Díaz, S. and Andrés-Chicote, M. (2021) 'Machine Learning for the Improvement of Deep Renovation Building Projects Using As-Built BIM Models', *Sustainability*, 13(12), p. 6576. doi:10.3390/su13126576.
Nazir, F.A. *et al.* (2021) 'Comparison of modular and traditional UK housing construction: a bibliometric analysis', *Journal of Engineering, Design and Technology*, 19(1), pp. 164–186. doi:10.1108/JEDT-05-2020-0193/FULL/PDF.
Olivier, J.G.J. and Peters, J.A.H.W. (2020) *TRENDS IN GLOBAL CO 2 AND TOTAL GREENHOUSE GAS EMISSIONS 2019 Report*. Available at: https://www.pbl.nl/sites/default/files/downloads/pbl-2020-trends-in-global- (Accessed: 30 June 2021).
Olu-Ajayi, R. *et al.* (2022a) 'Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques', *Journal of Building Engineering*, 45, p. 103406. doi:10.1016/J.JOBE.2021.103406.
Olu-Ajayi, R. *et al.* (2022b) 'Machine learning for energy performance prediction at the design stage of buildings', *Energy for Sustainable Development*, 66, pp. 12–25. doi:10.1016/J.ESD.2021.11.002.
Opitz, D. and Maclin, R. (1999) 'Popular Ensemble Methods: An Empirical Study', *Journal of Artificial Intelligence Research*, 11, pp. 169–198. doi:10.1613/jair.614.
Owolabi, H.O. *et al.* (2018) 'Predicting Completion Risk in PPP Projects using Big Data Analytics', *IEEE Transactions on Engineering Management* [Preprint]. doi:10.1109/TEM.2018.2876321.
Pedregosa, F. *et al.* (2011) 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, 12(85), pp. 2825–2830. Available at: http://scikit-learn.sourceforge.net. (Accessed: 7 January 2021).
Penistone, A. (2019) 'UK greenhouse gas emissions, provisional figures', *National Statistics*, (March), p. 46. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/790626/2018-provisional-emissions-statistics-report.pdf (Accessed: 29 June 2021).
Priyadharshini, T. *et al.* (2022) 'Machine learning prediction of SCOBY cellulose yield from Kombucha tea fermentation', *Bioresource Technology Reports*, 18, p. 101027.

doi:10.1016/J.BITEB.2022.101027.
Qin, H. and Wang, X. (2022) 'A multi-discipline predictive intelligent control method for maintaining the thermal comfort on indoor environment', *Applied Soft Computing*, 116, p. 108299. doi:10.1016/J.ASOC.2021.108299.
Rosenow, J. *et al.* (2018) 'The remaining potential for energy savings in UK households', *Energy Policy*, 121, pp. 542–552. doi:10.1016/j.enpol.2018.06.033.
Royal Society, T. (2017) *MACHINE LEARNING: THE POWER AND PROMISE OF COMPUTERS THAT LEARN BY EXAMPLE*.
Schneckenreither, M., Haeussler, S. and Peiró, J. (2022) 'Average reward adjusted deep reinforcement learning for order release planning in manufacturing', *Knowledge-Based Systems*, 247, p. 108765. doi:10.1016/J.KNOSYS.2022.108765.
Sert, S.A. and Yazici, A. (2021) 'Increasing energy efficiency of rule-based fuzzy clustering algorithms using CLONALG-M for wireless sensor networks', *Applied Soft Computing*, 109, p. 107510. doi:10.1016/J.ASOC.2021.107510.
Seyedzadeh, S. *et al.* (no date) 'Multi-Objective Optimisation for Tuning Building Heating and Cooling Loads Forecasting Models'.
Seyrfar, A. *et al.* (2021) 'Data-Driven Approach for Evaluating the Energy Efficiency in Multifamily Residential Buildings', *Practice Periodical on Structural Design and Construction*, 26(2), p. 04020074. doi:10.1061/(asce)sc.1943-5576.0000555.
Sha, H., Moujahed, M. and Qi, D. (2021) 'Machine learning-based cooling load prediction and optimal control for mechanical ventilative cooling in high-rise buildings', *Energy and Buildings*, 242, p. 110980. doi:10.1016/j.enbuild.2021.110980.
Sharma, M. and Garg, R. (2020) 'HIGA: Harmony-inspired genetic algorithm for rack-aware energy-efficient task scheduling in cloud data centers', *Engineering Science and Technology, an International Journal*, 23(1), pp. 211–224. doi:10.1016/J.JESTCH.2019.03.009.
Su, M. *et al.* (2018) 'Machine Learning from Theory to Algorithms: An Overview You may also like Classification of Rock Mineral in Field X based on Spectral Data (SWIR & TIR) using Supervised Machine Learning Methods S A Pane and F M H Sihombing-Automatic Machine Learning Method for Hyper-parameter Search Machine Learning from Theory to Algorithms: An Overview', *J. Phys*, p. 12012. doi:10.1088/1742-6596/1142/1/012012.
Sulaimon, I.A. *et al.* (2022) 'Effect of traffic data set on various machine-learning algorithms when forecasting air quality', *Journal of Engineering, Design and Technology*, ahead-of-print(ahead-of-print). doi:10.1108/JEDT-10-2021-0554/FULL/PDF.
Sun, Y., Haghighat, F. and Fung, B.C.M. (2020) 'A review of the-state-of-the-art in data-driven approaches for building energy prediction', *Energy and Buildings*, 221, p. 110022. doi:10.1016/J.ENBUILD.2020.110022.
Tehrani, F.S. *et al.* (2022) 'Machine learning and landslide studies: recent advances and applications', *Natural Hazards 2022*, pp. 1–49. doi:10.1007/S11069-022-05423-7.
Tsanas, A. and Xifara, A. (2012) 'Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools', *Energy and Buildings*, 49, pp. 560–567. doi:10.1016/J.ENBUILD.2012.03.003.
Tsoka, S. (2015) 'Optimizing Indoor Climate Conditions in a Sports Building Located in Continental Europe', *Energy Procedia*, 78, pp. 2802–2807. doi:10.1016/J.EGYPRO.2015.11.630.
Zhou, L. *et al.* (2017) 'Machine learning on big data: Opportunities and challenges', *Neurocomputing*, 237, pp. 350–361. doi:10.1016/J.NEUCOM.2017.01.026.
Zhou, Y. *et al.* (2021) 'Comparison of machine-learning models for predicting short-term building heating load using operational parameters', *Energy and Buildings*, 253, p. 111505. doi:10.1016/J.ENBUILD.2021.111505.