A data-driven energy performance gap prediction model using machine learning

A data-driven energy performance gap prediction model using machine learning

Yılmaz, D.[a], Tanyer, A.M.[b,*], Toker Dikmen, I.[c,1]

[a] Department of Architecture, Middle East Technical University, Ankara, 06800, Turkey
[b] Department of Architecture, Research Center for Built Environment, Middle East Technical University, Ankara, 06800, Turkey
[c] Department of Civil Engineering, Middle East Technical University, Ankara, 06800, Turkey

Abstract

The energy performance gap is a significant obstacle to the realization of ambitions to mitigate the environmental impact of buildings. Although extensive research has been conducted on the causes, minimization, or the quantifying of the energy performance gap in buildings, comparatively minimal work has been done on raising decision-makers awareness of a potential gap.

This paper positions project risks at the core of the gap and proposes an innovative performance gap prediction model focusing on heating and electricity demand in buildings by utilizing the machine learning classification. In this research, the performance gap and project risks of 77 buildings was collected via a web-based survey. The predictive performance of the four machine learning algorithms, namely i) Naive Bayes, ii) k-Nearest Neighbors, iii) Support Vector Machine, and iv) Random Forest, were compared to determine the best model.

The results obtained revealed that Naive Bayes was better able to predict the direction of the heating performance gap (72.50%), the negative heating performance gap (71.81%), the positive electricity performance gap (77.08%), and the negative electricity performance gap (83.85%). Furthermore, k-Nearest Neighbors and Support Vector Machine were more accurate to predict the direction of the electricity performance gap (79.00%), and the positive heating performance gap (76.04%).

Highlights

- A performance gap prediction model was proposed based on buildings' risk data.
- The models use machine learning to focus on the electricity and heating gaps.

[*] Corresponding author.
E-mail address: tanyer@metu.edu.tr (A.M.Tanyer)
[1] Present address: School of Construction Management and Engineering, University of Reading, Reading, RG6 6EN, United Kingdom

- The performance of four machine learning algorithms was compared.
- The suggested method can predict the direction of the gap (positive and negative).
- The suggested method can predict the gap in three levels (low, medium, high).

Keywords

Algorithm, building, classification, energy performance gap, machine learning, risk identification

Word Count: 7572

Nomenclature

Variables and parameters

FP	false positives
N	negative instances
P	positive instances
TN	true negatives
TP	true positives

Abbreviations

AUC	the area under the ROC curve
AutoML	automated machine learning
BEG	binary electricity gap
BHG	binary heating gap
BREEAM	Building Research Establishment Environmental Assessment Method
Chi	chi-squared attribute evaluator
ECC	exhaustive correction code
EPG	energy performance gap
KNN	k-nearest neighbor
ML	machine learning
NB	naive bayes
NEG	negative electricity gap
NHG	negative heating gap
OVA	one-vs-all
OVO	one-vs-one
PEG	positive electricity gap
PHG	positive heating gap
RCC	random correction code
RF	random forest
ROC	receiver operating characteristic
SMOTE	synthetic minority oversampling technique
SVM	support vector machines
USGBC	US Green Building Council
Wrapper	wrapper attribute evaluator

1. Introduction

The construction sector is the largest consumer of energy in the world. Buildings account for over 40% of global energy consumption and are a similarly significant origin of carbon emissions [1]. While various standards and rating systems exist that aim to promote resource efficiency and the construction of more environmentally friendly buildings, the literature suggests that the performance of buildings generally fails to achieve the required standards or meet design predictions [2-4]. This phenomenon, which is called the energy performance gap (EPG), denotes the difference between the predicted performance (or anticipated, calculated, designed, etc.) in the design phase, and that measured (or real, actual, achieved, etc.) in the operational phase [5]. The existence of EPG in buildings represents the gulf between reality and government policies designed to reduce energy consumption and greenhouse gas emissions [6], is a cause of considerable increases in energy costs and environmental impact [7], and demonstrates failures in the design of the system, as well as improper usage of capital investment [8]. The energy performance gap also endangers the chance that policymakers will succeed with future strategies [9].

A growing awareness of the significance of EPG has led to a considerable number of studies. In one such work, Janser *et al.* [10] noticed that studies on this issue are generally concerned with one or more topics: defining, explaining, quantifying, and controlling energy performance gaps. In support of the categorization, a review of relevant published literature indicated that it is critical to look at different stages of a building's life cycle to explain the reasons for the gap. The absence of building adaptability [11], design complexity [12], poor workmanship [13], and miscommunication about building performance targets between project stakeholders [14] are just some of the reasons cited. De Wilde [15] explains that the specific causes for a gap differ from one building to another, and that it is usually the case that a gap is caused by a combination of several problems. In addition to this, other scholars have stressed that energy performance gaps resulted from risk factors that occur during different stages of the building life cycle [16, 4, 17]. Doylend [4], for example, assigned the risk factors to four general classes: design and engineering, management and process, external constraints, and operation and maintenance, whereas Topouzi *et al*. [17] concluded that three types of risks appear in different retrofit techniques and work plan stages: assessment, sequence, and communication. Alam *et al*. [16] classified the risk factors of the construction and commissioning stages into six groups: material and equipment, knowledge

and working skills, construction management process, procurement process, design input, and client-related problems.

Researchers have also demonstrated that the magnitude of the EPG could be very different. Even though the performance gap is often connected with increased energy consumption it can, in fact, also mean reduced consumption [18]. However, in the majority of cases, the measured energy use is higher than predictions [19]. A study by Galvin [20] of the domestic heating of three retrofitted apartment buildings demonstrates how profound this variation can be. Galvin found that the energy performance gap ranged from 2% for the first building, 56.8% for the second, and 272.9% for the third. That said, there is no doubt that EPG is of great concern. The Innovative UK and the Zero Carbon Hub study claims that the performance gap is typically 2.6 times worse than the design predictions [21], while Calì *et al*. [22] concluded that values for the gap can reach as high as 287%.

In an attempt to reduce the EPG in buildings, researchers have conducted post-occupancy [23], and pre-occupancy evaluations [24], as well as using monitoring data to calibrate simulation models [25]. Hong *et al.* [26] have suggested using the operational ratings from the assessment method that investigates the actual energy performances of similar buildings with data mining or a machine learning (ML) approach. This latter approach is particularly supported by Hong *et al*. [27], who suggest that technological trends enable the collection and storage of increased amounts of data more cheaply, while the usage of powerful and low-cost computational resources, and the use of advanced ML algorithms, increase the advancement and application of ML in a diverse and extensive range of fields. ML, as a branch of artificial intelligence, uses example data or past experiences to optimize performance criterion [28], and aims to predict future events and scenarios unfamiliar to the computer [29]. ML algorithms currently represent the most contemporary and best effective way of prediction [30]. An extensive summary of applications of ML demonstrates how it is used in many applications, such as generating and evaluating design models, predicting construction costs, detecting construction objects within the image content, and detecting construction defects [31]. ML methods often appear in energy performance prediction studies [32-35].

This research builds upon work such as that of Nižetić & Papadopoulos [36], who proposed a novel but conceptual strategy to predict EPG. In this work, the authors introduced *energy efficiency building dissipation rate* as an essential factor for determining the magnitude of the performance gap. This study positions risk as a core concept of the energy performance gap and proposes an innovative performance gap prediction model for buildings by utilizing the

ML classification technique. No previous study has been located which uses ML methods to predict the energy performance gap in buildings, as so this research is seen as a valuable contribution to the existing body of knowledge by providing a new perspective for the prediction of future scenarios of energy performance gaps in buildings through the utilization of ML applications which benefit from past experience. This study considers buildings where the achieved energy savings are both higher and lower than the designed energy savings. ML has the ability to identify the patterns in the data that humans are often unable to notice, and thus better allow project stakeholders to appreciate the risk of an energy performance gap in terms of its nature and degree. This will, in turn, enable decision-makers to revise the decisions made about their projects and suggest new strategies for better controlling the gap.

2. Research background

2.1. Previous studies

Machine learning is much talked about nowadays, and one of this technology's almost limitless applications is as an essential energy prediction technique [37]. This is due to its superb ability to capture non-linear and complex relationships [38]. The intense activity in the sector can be seen in the number of papers published in the years 2011 and 2019 on the application of the use of ML in buildings: a four-fold increase [31]. In this section, some of these papers will be reviewed.

According to Mocanu *et al.* [39], there are multiple influencing factors involved in the prediction of energy use in a building. These include: the performance and settings of heating and cooling systems, weather conditions, and the number of people present. Mocanu *et al.* utilized the Conditional Restricted Boltzmann Machine method to forecast building electricity consumption by using a dataset that contained seven weeks of hourly resolution electricity consumption obtained from an office building. Paudel *et al.* [33] introduced two prediction modeling approaches for heating consumption based on support vector machines. Amasyali & El-Gohary [40] focused on forecasting the cooling energy consumption of a building by comparing ANN and other ML models. Mohammadiziazi & Bilec [41] developed four ML models to address the challenges of inconsistencies linked to integrating climate change models into energy modeling. Revati *et al.* [42] studied a smart commercial building to predict the electricity consumption profile via Gaussian Process Regression. Mounter *et al.* [37] explained that errors increased significantly beyond short-period energy forecasts, and that most reported energy forecasts relying on machine learning and statistical methods are

within one week. Therefore, the authors presented a detailed study of data processing and machine learning methods to enhance the accuracy of long-term energy forecasts of their building. Anand *et al.* [43] used time-series data of occupant density and energy consumption to develop building and space-wise energy prediction models with different ML algorithms. Ngo *et al.* [44] suggested an ensemble approach that uses artificial neural networks, support vector regression, and M5Rules models to forecast energy consumption in non-residential buildings.

Other areas of investigation have included prediction of the electricity [39, 42], heating [33], and cooling energy consumption [40]; addressing inconsistencies in integrating climate change models into energy modeling [41]; or improving the accuracy of the building's long-term energy forecasts [37] using many different algorithms.

2.2. Classification in machine learning

Classification is one of the most common applications for data mining [45]. The method works under supervision by being arranged according to the actual outcome for each training session. For example, classification learning is sometimes called "supervised learning" [46]. Classification aims to enable a system to predict the unknown output class of a formerly unseen instance, while also demonstrating a good generalization ability [47]. It involves the construction of a classifier, which is a function that attributes a class label to instances characterized by a set of attributes [48]. Classification's primary steps are synthesizing a model, using a learning algorithm, and applying the model to the labeling of new data [29]. If there are only two values that are used as labels to predict future unseen examples, then it is a binary classification problem [49], while having more than two classes to assign instances is a multi-class classification [50].

3. Method and material

The study was conducted as represented in Fig. 1. First, a comprehensive literature review was performed to determine the causes of the energy performance gap. Second, semi-structured interviews were performed with some of the experts working on different energy-efficient buildings to introduce a conceptual *energy performance gap risk framework*. Third, a web-based survey was designed to collect data about the risk and energy performance gap information on buildings. Next, using the gathered data in the EPG dataset, data was preprocessed using data cleaning, integration, and transformation. The dataset was then split

into a training (80%) and a testing set (20%). Using the training set, feature selection was applied to reduce the dimension of data and remove unnecessary inputs. An oversampling technique named Synthetic Minority Oversampling Technique (SMOTE) was then employed to create a balanced training set in the instance of there being a class imbalance problem. The models were also trained by hyperparameter tuning to optimize the model's performance. Subsequently, the performance of the algorithms was tested on unseen data (test data) based on different performance metrics such as accuracy, kappa, precision, recall, F-measure, the area under the receiver operating characteristic (ROC) curve (AUC), and statistical tests. Each of the steps is explained in more detail as follows:

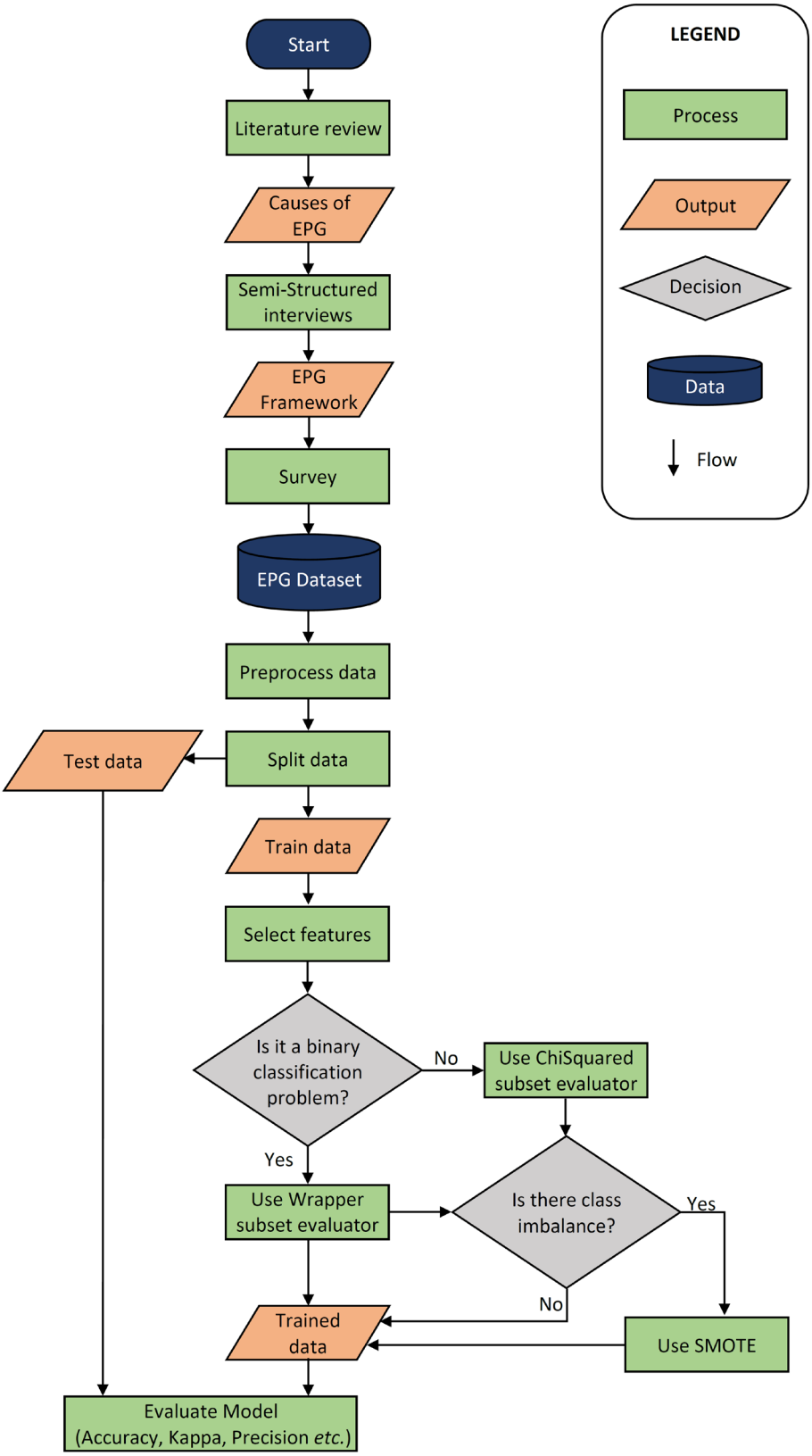

Fig. 1. A flowchart explaining the research process

The findings of the literature review were considered in the semi-structured interviews with experts on the project risks in the design, construction, and operational phases of specific buildings. The interviewees, who had an average of 12 years of experience in energy-efficient buildings, comprised of three project managers, three mechanical engineers, a site manager, an electrical technician, a quality manager, a building commissioning agent, and a CEO (Table 1). Six case studies with different certificates, locations, and project types were selected to observe the diverse risk paths of different buildings, countries, and company conditions.

Table 1

Profile of the interviewees

Case	Profession	Job position	Country	Experience (year)	
				Construction sector	**Energy efficient buildings**
I	Architect	Project manager	Turkey	21	8
	Civil engineer	Site manager		39	8
	Mechanical engineer	Mechanical engineer		23	8
	Mechanical engineer*	Commissioning agent		36	12
II	Mechanical engineer	Mechanical engineer		5	2.5
	Electrical technician	Electrical technician		37	10
	Printing educator	Quality manager		8	8
III-IV	Architect & city planner	Project manager		11	9
III-IV	Mechanical engineer	Mechanical engineer		9	8
V	Architect	Project manager	Germany	28	21
VI	Architect & civil engineer	CEO		34	34

*Fourth interviewee participated in the interviews as the commissioning agent of Cases I and II.

A two-stage approach, under which the questions were sent to the interviewees beforehand, was used in the semi-structured interviews which were conducted between December 2020 and March 2021. The first stage aimed to listen to the interviewees' explanations of the problems that might cause the energy performance gap in the buildings. In addition, the interviewees were asked to describe the effects of, and their responses to, the problems. In the second stage, the problems were presented to the interviewees through cognitive maps. This enabled the relationships between the problems to be revised or verified. Later, the factors described on the maps for each building were listed and designed as a questionnaire. A pretest was performed with two researchers working on building energy performance, and one consultant working on energy-efficient buildings. The respondents were asked to give feedback on the questionnaire design, including length, confusing questions, terminology,

etc. Utilizing the pretesting process, some factors describing similar conditions were defined as a broader concept. The factors repeated several times in the questionnaire were recategorized. Accordingly, a conceptual *framework wa*s suggested, as in Fig. 2. This framework helped to design a web-based survey to collect buildings' risk and energy performance gap data to develop ML models.

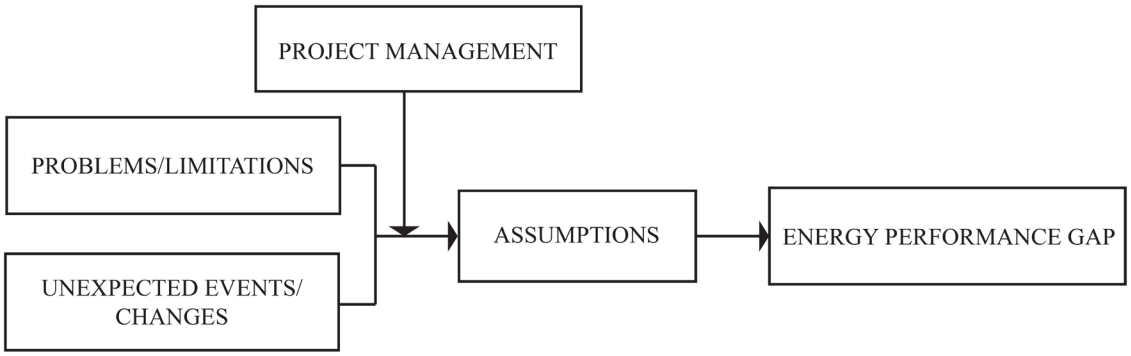

Fig. 2. Conceptual EPG risk framework

In this framework, the risks related to the EPG were classified into four main groups, namely "Problems/limitations", "Unexpected events /changes", "Project management", and "Assumptions". According to this framework, "Problems/ limitations" may cause "Unexpected events /changes" or vice versa, while project management processes may influence the manageability of "Problems/limitations," and "Unexpected events/ changes". These factors may subsequently lead to a change in "Assumptions", and this group refers to the assumptions made during the design phase. Consequently, a combination of the factors identified in this framework may result in an energy performance gap in a building.

3.1. Web-based survey

A Web-based survey consisting of forty-two questions was designed using eSurveyPro® to collect buildings' risk and energy performance gap data. The survey contained open, closed, and mixed question types. The first page of the survey required information on project characteristics, whereas the remainder focused on the risk factors that can emerge at different times throughout the project. The survey questions related to risk factors were evaluated on a 5-point Likert scale, from very low to very high, and included a ‘not applicable’ (N/A) option. Online survey tools, e-mail, face-to-face meetings, and phone interviews were used to deliver the survey to some of the experts who are working on energy-efficient buildings. The target group included professionals responsible for project management, energy consultancy, and commissioning in energy-efficient building projects, while 1,000 surveys were sent out

based on cluster sampling. LinkedIn, USGBC, BREEAM, and Passive House Database websites were used to collect the contact information of the target group.

The survey consists of six sections, with the first containing questions, such as project type, location, and year of construction, concerning general information about a specific building that the respondents worked on. Additionally, respondents were required to provide the energy demand given in simulation software and actual consumption for that building regarding electricity and heating demand. The second section collected information about the problems/limitations encountered during the life cycle stages of the relevant building. While respondents were required to evaluate factors about project management in the third section, the fourth aimed to understand how the design assumptions regarding the building had changed. The fifth section listed unexpected events and changes in the process, while the sixth contained questions about the respondents, such as education level, profession, *etc.* Furthermore, a space is left for the respondents to make any further comments at the end of the survey.

3.2. Survey responses

A total of 72 responses out of 1,000 were received, yielding a response rate of 7.2%. While such a low response to the survey might at first seem disappointing, it could be said that the surveys that were received may be an accurate representation of the population's attitudes. In other words, a low response rate should not be a reason to imagine that the results are uninformative [51]. Some possible reasons for the low response rate could be as follows: some experts explained that the reasons for nonparticipation in the survey were because of a nondisclosure agreement signed with clients, not having the opportunity to obtain in-use metered data for buildings, not having access to the data, only having a couple of projects with information on operational energy consumption as well as energy modeling, and time limitations in obtaining the client's permission to use them.

3.3. Data preprocessing

If they are applied before mining, data processing methods can substantially improve the quality of the patterns mined and the time needed for the actual mining. The major steps in data preprocessing are data cleaning, integration, reduction, and transformation [52]. In this study, data cleaning aimed to handle missing values and noisy data by detecting outliers. When the surveys are examined, it is seen that nine respondents did not answer one or both

questions about the magnitude of the energy performance gap in buildings. Although these records were removed, the missing values were replaced in the others with the field mode in order to not lose any data. In addition, one respondent who participated in the survey twice gave the same answers to all questions. These records were removed as duplicate records caused an overweighting of the data values in those records [53]. Moreover, five buildings in the database showed between 200-410% higher energy consumption than the design predictions. When similar values were organized into groups, these cases fell outside the set of the groups. They were therefore considered outliers and removed.

Integrating data from the greatest possible variety of sources is crucial to effective machine learning [54]. For this reason, this study integrated empirical research articles into data obtained using a web-based survey. Studies by Pegg *et al.* [55], Korjenic & Bednar [56], and Herrando *et al.* [57] were analyzed by manual topic-based text classification utilizing a rule-based approach. After defining a list of words representing each group, tags were assigned considering their content and frequency. This has produced a database consisting of 77 projects, as can be seen in Table 2.

Table 2

Project profile in the dataset

	Category	Number of projects
Construction period	18th - 19th century	10
	20th century	29
	21st century	38
Construction type	Reinforced concrete	43
	Mixed construction	20
	Masonry construction	10
	Timber frame	4
Heated floor area	up to 500 m²	11
	501 - 2,500 m²	33
	2,501 - 6,000 m²	11
	6,001 - 10,000 m²	12
	more than 10,001 m²	10
Project type	Educational building	48
	Office building	15
	Multi-family dwelling	4
	Single-family house	8
	Hospital	1
	Cafe-Restaurant	1
Country	Germany	39
	Spain	18
	Turkey	8
	UK	6
	USA	2
	Austria, Belgium, Iran, Ukraine	4
Total number of projects in the dataset		77

The data mainly contains categorical data. While Likert scale survey questions are ordinal, "project type" (*e.g.*, office building, educational building, *etc.*), "construction type" (*e.g.*, reinforced concrete, steel frame, *etc.*), and "new building or refurbishment" are nominal. Numerical data is also included, such as "the year of construction", "heated floor area," and "energy performance gap", which are examples of interval and ratio data, respectively. In this study, data discretization was applied for data transformation. Numeric attributes were replaced by interval and conceptual labels. In these, the prediction problems were first studied as a binary classification problem, and energy performance gaps in buildings were denoted as being positive or negative. In this denotation: a positive performance gap represents the case where the real performance is more than the predicted performance.

In contrast, a negative performance gap represents the case where the actual performance is less than the design expectations. In order to develop multi-class classification models to

predict the magnitude of the gap in percentages, the raw values were replaced by interval labels (*e.g.*, 0-15%, 15.1-40%, 40.1-90%) corresponding to low, medium, and high for both positive and negative performance gap classes. An increase in the number of class labels often decreases the accuracy of the classifier in multi-class classification [58], so the number of classes was limited to three. García *et al.* [59] explain that binning techniques are helpful not only in reducing the dimensionality and complexity of the dataset, but also in improving the predictive power of a variable. At this stage, equal-width binning created a significant class imbalance problem. It is for this reason that the intervals of the categories were decided by aiming for equal frequency binning as possible given six different classification problems. Furthermore, the year of construction was partitioned into three bins: 18^{th} - 19^{th} century, 20^{th} century, and 21^{st}-century buildings. In contrast, the heated floor area was partitioned into five bins: very small, small, medium, large, and very large.

This study used an open-source software named Weka 3.8.5 for data visualization, data preprocessing, attribute selection, and classification. While the Explorer interface was used for data preprocessing, classification, attribute selection, and data visualization, the experimenter interface was used to perform experiments, compare a variety of classification algorithms, and conduct Paired T-Tester statistical tests.

3.4. Feature selection

Datasets may include hundreds of attributes, which may be unimportant to the mining task or unnecessary [52]. Feature selection has been a valuable method to reduce the complexity of machine learning and data mining applications [60], evaluate the informative features, and reduce the dimension of data [61]. There are three groups of feature selection methods: filter, wrapper, and embedded methods [62]. While an independent assessment is made according to the general characteristics of the data in filter methods [46], an ML algorithm is used in wrapper methods to choose the best subset of features [63].

Due to their ability to create better predictive models [64], WrapperSubsetEval was applied for binary classification problems since it is the starting point for the whole classification procedure. There are several models in multi-class classification problems, so the preference was a filter method since they are more practical and much faster [65]. In this section, the chi-squared attribute evaluator was used as a filter method. Witten *et al.* [46] explain that the chi-squared statistic of each attribute is computed concerning the class in this method, and Table 3 illustrates the configuration of the attribute selection methods using Weka. Four

algorithms were chosen, one after the other, and three direction alternatives were tried during wrapper subset evaluation configuration. Values between 1 to 10 were tested and feature subsets delivering higher performance were stored. In terms of the use of ChiSquaredAttributeEvaluator, after the "Ranker" search method was selected, 3 to 10 features were retained to find the best performers.

Table 3

Configuration of the attribute selection

Evaluation method	**Wrapper Subset Evaluator**
Classifier	i) Naive Bayes, ii) KNN, iii) SVM iv) Random Forest
Folds	10
Threshold	-1
Search method	Best first
Direction	backward, forward, bi-dimensional
Search termination	1 to 10
Evaluation method	**ChiSquaredAttributeEvaluator**
Search method	Ranker
numToSelect	3 to 10

3.5. Synthetic minority oversampling technique

If the classes are not almost equally represented, a dataset becomes imbalanced [66]. Minority class instances are more often miscategorized in imbalanced datasets [47]. It is for this reason, in the case of a class imbalance problem, that an oversampling technique called SMOTE is used to create a balanced training set. This technique involves new data being added to the minority class of an imbalanced training set.

3.6. Selection of the algorithms

The configuration of ML tools is usually performed manually to achieve better predictive performance. A recently much discussed technique is a new sub-field of ML, called automated machine learning (AutoML) [67]. AutoML aims to select, compose, and parametrize ML algorithms automatically [68] to conserve effort and time on repetitive work in ML pipelines [69], and to close the gap for inexperienced ML users by undertaking the role of the field expert [70]. AutoML makes ML available to everyone, and it appears to be promising [71]. However, a study comparing the capability of AutoML tools concluded that although some tools performed better than others, these were subject to poor performance in

either binary or multi-class classification [69]. In addition, despite the increased efforts to confront the challenges of AutoML, numerous challenges are still available [70]. Therefore this study applied a manual approach during the algorithm selection and other ML tasks.

After testing the performance of 20 classification algorithms on the training sets, four algorithms were selected, namely Naive Bayes, k-Nearest Neighbor, Support Vector Machines, and Random Forest, due to their better performance on the dataset. The scatter plots showed that the data was not linearly separable. The improved performance of these algorithms could be due to their ability to perform well with non-linearly separable data.

Naive Bayes (NB) is one of the most effective classifiers in its predictive performance, despite its assumptions about independence [48], and uses probability theory to determine the most likely possible classifications [45]. It is also easy to construct and use [29]. Moreover, NB is efficient and robust for both small and normal size of datasets [72]. Besides, Stribos [73] showed that NB was more robust against data noise in the training data when compared to Random Forest.

The k-Nearest Neighbor (KNN) algorithm has yet to attract considerable interest in the building energy prediction field [30]. It is a simple instance-based learner that applies the class of the nearest k training instances for the class of the test instances [74] and is able to manage binary and multi-class data classification problems [8].

Support Vector Machines (SVM) combine instance-based and linear modeling [46]. SVM is skilled at solving non-linear problems even if the training data amount is small [40], and has demonstrated outstanding performance in binary classification tasks [75]. Nevertheless, methods such as One-Vs-All (OVA) and One-Vs-One (OVO), *etc.,* are necessary to be used externally for multi-class classification problems [8].

Random Forests (RF) are among the popular decision tree methods in predicting building energy consumption [40]. This is an ensemble learning technique that creates a forest of random trees with controlled variance [74]. It is fast [65] and provides good accuracy, even though a substantial amount of data is missing [29].

3.7. Configuration of the algorithms

Optimization of model parameters, referred to as tuning, plays a key role in the accuracy of ML model predictions [34]. Configuration of the algorithms is necessary to tune the hyperparameters as they influence the learning and prediction procedure and affect the

performance of ML models. Table 4 provides the configuration settings of the studied algorithms for binary and multi-class classification problems.

Breaking the problem into binary components is one of the ways to manage multi-class issues. There are several methods to transform a multi-class problem into binary ones, including OVO, OVA, Random Correction Code (RCC), and the Exhaustive Correction Code (ECC). Each decomposition method was set using Multi-Class Classifier to discover the optimal performance method in Weka. Furthermore, CV Parameter Selection was performed to identify an optimum C parameter by cross-validation for SVM. In addition, the grid search method was used with Random Forest to test each set of possible combinations of the parameters and select the one with the highest accuracy as the final value.

Table 4

Configuration setting of algorithms

Classifier	Setting	Binary classification	Multi-class classification
1.weka.classifiers.bayes.NaiveBayes		√ (default)	√ (default)
2.weka.classifiers.meta.MultiClassClassifier		x	√ (default)
	2.1 classifier	x	Naive Bayes
	2.2 method	x	OVA, OVO, RCC, ECC
	2.3 usePairwiseCoupling	x	True for the OVO technique
3.weka.classifiers.lazy.Ibk	3.1 k	1 to 10	default
	3.2 cross-validate	true	true
	3.3 distanceWeighting	No distance weighting, Weight by 1/distance	No distance weighting, Weight by 1/distance
	3.4 distanceFunction	Manhattan Distance	Manhattan Distance
4.weka.classifiers.meta.MultiClassClassifier	4.1 classifier	x	CVParameterSelection (IBk)
	4.2 CVParameters	x	K 1 35 1
	4.3 numFolds	x	4
	4.4 method	x	OVA, OVO, RCC, ECC
	4.5 usePairwiseCoupling	x	True for the OVO technique
5.weka.classifiers.meta.CVParameterSelection	5.1 classifier	SMO	SMO
	5.2 CVParameters	C 1 1000 1	C 1 1000 1
	5.3 numFolds	10	4
	5.4 filterType	No normalization/standardization	No normalization/standardization
	5.5 kernel	Polykernel,RBFkernel,NormalizedPolyKernel	Polykernel,RBFkernel,NormalizedPolyKernel
	5.6 method	x	OVA, OVO, RCC, ECC
	5.7 usePairwiseCoupling	x	True for the OVO technique
6.weka.classifiers.meta.GridSearch	6.1 Classifier	Random Forest	Random Forest
	6.2 Xproperty	numFeatures	numFeatures
	6.3 Xbase,Xmax,Xmin,Xstep	2,3,-3,1	2,3,-3,1
	6.4 Yproperty	numIterations	numIterations
	6.5 Ybase, Ymax, Ymin, Ystep	10,2,0,1	10,2,1,1
	6.6 Evaluation	Accuracy	Accuracy
7.weka.classifiers.meta.MultiClassClassifier	6.7 Classifier	x	GridSearch
	6.8 method	x	OVA, OVO, RCC, ECC
	6.9 usePairwiseCoupling	x	True for the OVO technique

3.8. Model evaluation

Performance measures and statistical significance testing are some of the main factors that are essential to evaluate learning algorithms [49]. The performance of machine learning algorithms was assessed in this study through the following six metrics: accuracy, precision, recall, F-measure, kappa statistic, and AUC. The following equations are used to calculate the performance metrics [45]:

$$\text{Accuracy} = \frac{TP+TN}{P+N} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = TP/P \quad (3)$$

$$\text{F-measure} = 2x\frac{PrecisionxRecall}{Precision+Recall} \quad (4)$$

where *TP*, *TN*, *FP*, *P*, and *N* are the number of true positives, true negatives, false positives, and positive and negative instances.

In addition, the kappa statistic shows the prediction agreement with the true class, with 1.0 signifying complete agreement [76]. AUC can be explained as the probability that a randomly chosen positive example will be ranked higher by the model than a randomly chosen negative example [77]. The model is better if the area under the curve is larger [46]. Furthermore, the differences between any pairwise algorithm performance comparisons were analyzed to learn if there is a statistical significance, with a confidence of 95%, using Paired T-Tester statistical test. A statistical significance test can be used to evaluate whether the accuracy between two classifiers is different due to chance [52].

Moreover, the K-fold cross-validation resampling method is applied to estimate the performance of the models. It enables the achievement of statistically valid results when the original sample size is not large [78]. Dataset X is split randomly with K equal- cross-validation sized pieces, X_i, i = 1,..., K in K-fold cross-validation. One of the K parts is kept out to generate each pair as the validation set, and the unused K-1 parts are combined to constitute the training set [28]. As the standard procedure, repeating the cross-validation process ten times and averaging the results [46] can overcome limitations such as the small size of training and validation sets, availability of noise and outliers in the dataset, or sources of randomness in the learning method [28]. The 10-fold cross-validation with ten repetitions

was applied in binary classification problems. As the datasets were smaller in multi-class classification problems, 4-fold cross-validation with four repetitions was used in these groups.

4. Results

4.1. Selected features using wrapper and filter methods

There are 33 attributes used in ML classification problems in this study. Considering the electricity performance gap in buildings, the following features were selected at least twice by Weka in binary and multi-class classification problems:

- The shortcomings of modeling, software, or calculation methodology
- Problems with the quality of workmanship
- Problems with simulation inputs
- Problems with the quality of materials
- Inconsistencies in design projects and construction
- Problems with occupant behavior
- Problems with commissioning
- The motivation of the project parties
- Effective communication between the project parties
- Simplicity of detailing
- Unexpected events and changes in policy, legislation, or regulations

Considering the heating performance gap in buildings, the following features were selected at least twice by Weka in binary and multi-class classification problems:

- Problems with the quality of workmanship
- Problems with design
- Shortcomings of modeling, software, or calculation methodology
- Problems with simulation inputs
- Heated floor area

- Construction type

Although the features selected by the wrapper or chi-squared attribute evaluator are different in number and type for each classification model, the above features appeared more often in energy performance gap prediction problems. Table 5 represents the selected features/attributes by different feature selection methods.

Table 5

Selected attributes with Wrapper Subset evaluator and ChiSquared attribute evaluator

			BEG				PEG	NEG	BHG				PHG	NHG
			Wrapper				Filter	Filter	Wrapper				Filter	Filter
Group	**No**	**Attributes**	**NB**	**KNN**	**SVM**	**RF**	**Chi**	**Chi**	**NB**	**KNN**	**SVM**	**RF**	**Chi**	**Chi**
General info	**1**	Project type											√	
	2	Year of construction												
	3	New or refurbished								√				
	4	Heated floor area			√							√	√	
	5	Construction type		√							√	√		
Problems/limitations	**6**	Modeling		√	√	√	√	√					√	√
	7	Inputs	√					√					√	√
	8	Design					√		√		√	√		
	9	Project budget												
	10	Quality of workmanship		√	√			√	√	√	√	√		
	11	Quality of materials					√	√						
	12	Inconsistencies in project			√		√							
	13	Bankruptcy												
	14	Occupants' behavior			√			√						
	15	Commissioning		√	√									√
	16	Building management					√							
	17	Regular maintenance												
	18	Quality of measured data												
Project management	**19**	Experience of the stakeholders					√							
	20	Motivation of the stakeholders					√	√			√			
	21	Effective communication					√	√					√	
	22	Training of the stakeholders												
	23	Design flexibility							√					
	24	Occupant surveys		√								√		
	25	Applying passive measures		√										
	26	Simplicity of detailing	√		√									√
Unexpected events/changes	**27**	Client/ user expectations												
	28	Project stakeholders												
	29	County conditions			√									√
	30	Policy/legislation/regulations			√	√								
	31	Public sector building process												
	32	Climate				√								
	33	Force majeure events												

BEG: Binary electricity gap
PEG: Positive electricity gap
NEG: Negative electricity gap
Chi: Chi-Squared attribute evaluator
BHG: Binary heating gap
PHG: Positive heating gap
NHG: Negative heating gap
Wrapper: Wrapper attribute evaluator

4.2. Electricity performance gap prediction

Datasets were initially studied as binary classification problems to predict the electricity performance gap and to recognize cases that might demonstrate a negative or positive performance gap. 10-fold cross-validation was repeated ten times, and the average value of the following performance metrics was obtained (accuracy, kappa statistic, precision, recall, F measure, and AUC). The results indicate that KNN and SVM are the two best performers regarding accuracy, kappa statistic, precision, and F-measure, with KNN and SVM achieving accuracies of 79% and 75% on unseen (test) data, respectively (Table 6).

A statistical significance test was carried out using Paired T-Tester statistical test on all classification problems. The results in bold font illustrate the classifiers which are significantly better, and the results in italics show the classifiers which are considerably worse than the base classifier. Regular font indicates doubt over whether there is a statistically significant difference or not. At this stage, ZeroR, which always classifies to the largest class, was selected as the baseline classifier in all classification problems. Since KNN exhibited the best accuracy among all of the classification algorithms, it was used as the suggested model.

Table 6

Performance of the algorithms of the electricity performance gap prediction

	Binary electricity gap			
Algorithms	**Naive Bayes**	**KNN**	**SVM**	**Random Forest**
Accuracy (%)	66.00	**79.00**	**75.00**	73.50
Kappa	0.53	**0.65**	**0.61**	**0.59**
Precision	**0.81**	**0.82**	**0.83**	**0.81**
Recall	*0.68*	*0.86*	*0.79*	*0.81*
F-measure	**0.90**	**0.89**	**0.91**	**0.87**
AUC	**0.80**	0.62	**0.72**	**0.67**

Bold font illustrates the classifiers significantly better than the base classifier.
Italic font shows the classifiers which are significantly worse than the base classifier.
Regular font indicates doubt over whether there is a statistically significant difference or not.

After binary classification, smaller datasets were studied to forecast the level of performance gap in buildings based on three classes, low (0-15%), medium (15.1-40%), and high (40.1-90%). By repeating the 4-fold CV four times, the top two performers regarding accuracy and kappa statistic in predicting a positive performance gap for electricity demand are Naive Bayes (OVO) and Random Forest (ECC) for which accuracy of 77.08% and 73.96% on unseen data was recorded respectively (Table 7). Naive Bayes is the suggested classifier for

the solving of multi-class classification for positive electricity gap prediction due to its superior prediction accuracy.

Table 7

Performance of the algorithms of the positive electricity performance gap prediction

Multi-class electricity gap: Positive				
Algorithms	**Naive Bayes**	**KNN**	**SVM**	**Random Forest**
Accuracy (%)	**77.08**	**60.42**	**65.63**	**73.96**
Kappa	**0.65**	0.42	0.49	**0.62**
Precision	**1.00**	0.43	0.67	0.67
Recall	**0.67**	**0.67**	**0.67**	**0.75**
F-measure	**1.00**	0.83	**1.00**	**1.00**
AUC	0.83	0.75	0.90	**0.92**

Bold font illustrates the classifiers significantly better than the base classifier.
Italic font shows the classifiers which are significantly worse than the base classifier.
Regular font indicates doubt over whether there is a statistically significant difference or not.

Examining multi-class classification of the negative performance gap prediction for electricity demand, Naive Bayes (RCC) and KNN (OVA) are seen to be the top two performers concerning accuracy, kappa statistic, recall, and AUC. At this stage, precision and F- Measure was indeterminable when the model was evaluated on unseen data, and so the results are based on repeating 4-fold CV 4 times on the whole data set, including unseen test data. Naive Bayes (RCC) and KNN (OVA) achieved an accuracy of 83.85% and 82.29%, respectively (Table 8). The suggested classifier for solving multi-class classification for negative electricity gap prediction is Naive Bayes due to its superior prediction accuracy.

Table 8

Performance of the algorithms of the negative electricity performance gap prediction

Multi-class electricity gap: Negative				
Algorithms	**Naive Bayes**	**KNN**	**SVM**	**Random Forest**
Accuracy (%)	**83.85**	**82.29**	**78.13**	**79.17**
Kappa	**0.76**	**0.73**	**0.67**	**0.69**
Precision	**1.00**	**0.90**	**0.92**	**0.92**
Recall	0.90	**0.94**	0.84	**0.84**
F-measure	**0.90**	**0.94**	**0.92**	**0.92**
AUC	**0.98**	**0.98**	**0.97**	**0.97**

Bold font illustrates the classifiers significantly better than the base classifier.
Italic font shows the classifiers which are significantly worse than the base classifier.
Regular font indicates doubt over whether there is a statistically significant difference or not.

4.3. Heating performance gap prediction

A similar process was conducted to predict the heating performance gap. In this process, datasets were initially studied as binary classification problems to recognize cases that might demonstrate a negative or positive performance gap. Naive Bayes and SVM were seen to be the top two performers for binary classification of the heating gap with a 10-fold CV, providing an accuracy of 72.5% and 68.5% on unseen data, respectively (Table 9). Naive Bayes was seen to outperform other classifiers in five performance metrics, and so is the suggested classifier to solve binary classification for heating gap prediction due to its superior prediction accuracy.

Table 9

Performance of the algorithms of the heating performance gap prediction

	Binary heating gap			
Algorithms	**Naive Bayes**	**KNN**	**SVM**	**Random Forest**
Accuracy (%)	**72.50**	68.00	**68.50**	67.50
Kappa	**0.62**	**0.54**	**0.55**	**0.56**
Precision	**0.74**	**0.70**	**0.70**	**0.72**
Recall	*0.90*	*0.87*	*0.87*	*0.84*
F-measure	**0.90**	0.86	0.86	**0.89**
AUC	0.64	**0.67**	**0.66**	**0.69**

Bold font illustrates the classifiers significantly better than the base classifier.
Italic font shows the classifiers which are significantly worse than the base classifier.
Regular font indicates doubt over whether there is a statistically significant difference or not.

Following binary classification, smaller datasets were considered to predict the level of performance gap in buildings based on three classes, low (0-15%), medium (15.1-40%), and high (40.1-90%). SVM (ECC) and Random Forest (OVA) are the top two performers concerning accuracy and kappa statistics for the positive heating performance gap prediction with a 4-fold CV (Table 10). SVM (ECC) and Random Forest (OVA) provided an accuracy of 76.04 % and 69.79% on unseen data, respectively. SVM outperforms Random Forest in accuracy, kappa statistics, precision, and recall, and so is the suggested classifier in this classification problem.

Table 10

Performance of the algorithms of the positive heating performance gap prediction

Multi-class heating gap: Positive				
Algorithms	**Naive Bayes**	**KNN**	**SVM**	**Random Forest**
Accuracy (%)	**59.38**	**55.21**	**76.04**	**69.79**
Kappa	**0.45**	**0.42**	**0.68**	**0.57**
Precision	0.44	0.44	0.50	0.43
Recall	**0.83**	**0.83**	**0.92**	**0.83**
F-measure	**0.92**	**0.92**	**0.92**	**1.00**
AUC	**1.00**	**1.00**	**0.96**	**1.00**

Bold font illustrates the classifiers significantly better than the base classifier.
Italic font shows the classifiers which are significantly worse than the base classifier.
Regular font indicates doubt over whether there is a statistically significant difference or not.

Finally, multi-class classification models were studied for the negative performance gap of heating demand. Naive Bayes (OVA) and Random Forest (OVO) provided an accuracy of 71.81% and 65.14%, respectively, and so were seen to be the top two performers regarding accuracy and precision. (Table 11). However, as with the multi-classification problem of negative performance gap prediction for electricity demand, precision, and F-measure were indeterminable when evaluating the model on test data. The results are therefore based on repeating 4-fold CV 4 times on the whole data set, including test data. Due to its superior prediction accuracy, Naive Bayes is the suggested classifier to solve multi-class classification for negative heating gap prediction.

Table 11

Performance of the algorithms of the negative heating performance gap prediction

Multi-class heating gap: Negative				
Algorithms	**Naive Bayes**	**KNN**	**SVM**	**Random Forest**
Accuracy (%)	**71.81**	**60.69**	**64.58**	**65.14**
Kappa	**0.58**	**0.41**	**0.47**	**0.47**
Precision	**0.77**	**0.59**	**0.63**	**0.64**
Recall	*0.45*	*0.56*	*0.59*	*0.60*
F-measure	0.61	0.58	0.57	0.56
AUC	**0.81**	**0.76**	**0.76**	**0.76**

Bold font illustrates the classifiers significantly better than the base classifier.
Italic font shows the classifiers which are significantly worse than the base classifier.
Regular font indicates doubt over whether there is a statistically significant difference or not.

5. Conclusion

This research has explored the energy performance gap in buildings through the perspective of project risks. The main contribution of the present research is to indicate the potential application of ML classification in the energy performance gap prediction of buildings. This study can therefore be said to provide a new perspective by using project risk data to predict the direction and magnitude of EPG with ML. If project stakeholders know the direction (positive or negative) and the magnitude of a performance gap (low, medium, high) beforehand, their strategies and decisions might change to control the gap.

A web-based survey collected risk and energy performance gap information on buildings worldwide in this study. The gathered data from the survey was then studied as ML classification problems to predict EPG in buildings regarding heating and electricity demand. Binary classification was the starting point for the whole prediction procedure. This step aimed to predict whether the buildings might consume less (negative performance gap) or more (positive performance gap) than design expectations. Multi-class classification problems subsequently aimed to predict the magnitude of the performance gap as percentages in buildings (low - 0-15%, medium - 15.1-40%, and high - 40.1-90%).

Unnecessary attributes were eliminated using wrapper and filter methods. Wrapper Subset Evaluator was initially used for binary classification problems due to their ability to create better predictive models. Chi-Squared Attribute Evaluator was then used as a filter method in multi-class classification problems as they are faster and more practical than the former. The performance of four machine learning algorithms (Naive Bayes, SVM, KNN, and Random Forest) was compared using Weka to find the best prediction model based upon six performance metrics and the Paired T-Tester statistical test.

The results revealed that while different algorithms provided the highest prediction accuracy for each EPG prediction problem, Naive Bayes was the best, and Random Forest was the second-best, overall performing algorithm. The success of the Naive Bayes algorithm can be explained by its robustness and efficiency for both small and normal size of datasets. In addition, the better performance of Naive Bayes and Random Forest algorithms on EPG prediction can be explained by their ability to deal with noisy data.

Moreover, the feature selection step results revealed that different subsets of features were selected in each classification problem. Nevertheless, regarding performance gap prediction

considering electricity demand in buildings, "problems with modeling, software or calculation methodology" and "problems with the quality of workmanship" were often selected. Additionally, regarding performance gap prediction considering heating demand in buildings, "problems with design" and "problems with the quality of workmanship" were usually selected.

Prediction of the energy performance gap is critical for deciding on possible investment in buildings, eliminating unreasonable design of the system, reducing energy cost and environmental impact increases, and, most importantly, for the success of policymakers' plans on future strategies. Testing the occurrence of a performance gap in buildings using the suggested method can be a starting point for tackling the adverse outcomes of the gap for many different project stakeholders and the environment.

6. Limitations

ML is suggested for use as a powerful computational method when sufficient data is available. The main limitation of this study was the limited amount of data gathered through surveys and empirical articles used in solving of classification problems. In the study, the energy performance gap prediction problems were therefore divided into steps. First, the problems were studied as binary classification problems and then as multi-class classification problems. This study demonstrates, despite the limited sample size, a generic way to predict EPG in buildings. It is suggested that the recommended method be applied on a larger pool of data since it is critical to have access to a representative sample and quality data for the target population.

Acknowledgment

This work was supported by the Scientific and Technological Research Council of Turkey [grant number 1059B142000267, 2021].

We extend our gratitude to the staff at WG Energy Team and the Chair of Building Physics at Bauhaus University Weimar for their support of the study.

References

[1] Alencastro J, Fuertes A, de Wilde P. The relationship between quality defects and the thermal performance of buildings. Renew Sustain Energy Rev 2018;81:883–94.

[2] Borgstein EH, Lamberts R, Hensen JLM. Evaluating energy performance in non-domestic buildings: a review. Energy Build 2016; 128:734–55.

[3] Corry E, Pauwels P, Hu S, Keane M, O'Donnell J. A performance assessment ontology for the environmental and energy management of buildings. Autom Constr 2015; 57:249–59.

[4] Doylend N. Evaluating building energy performance: a life-cycle risk management methodology [dissertation]. Loughborough (UK): Loughborough University; 2015.

[5] Shi X, Si B, Zhao J, Tian Z, Wang C, Jin X, et al. Magnitude, causes, and solutions of the performance gap of buildings: a review. Sustain 2019; 11: 937.

[6] Imam S, Coley DA, Walker I. The building performance gap: are modellers literate? Build Serv Eng Res Technol 2017; 38 (3):351–75.

[7] Birchall SJ. An appraisal of the performance of a "green" office building [dissertation]. Leeds (UK): University of Leeds; 2011.

[8] Harrison S, Jiang L. An investigation into the energy performance gap between the predicted and measured output of photovoltaic systems using dynamic simulation modelling software-a case study. Int J Low-Carbon Technol 2018; 13(1):23–9.

[9] Li Y, Kubicki S, Guerriero A, Rezgui Y. Review of building energy performance certification schemes towards future improvement. Renew Sustain Energy Rev 2019; 113:109244.

[10] Janser M, Hubbuch M, Windlinger L. Call for a definition and paradigm shift in energy performance gap research. IOP Conf Ser Earth Environ Sci 2020; 588(5).

[11] Montazami A, Gaterell M, Nicol F. A comprehensive review of environmental design in UK schools: history, conflicts and solutions. Renew Sustain Energy Rev 2015; 46: 249–64.

[12] Jradi M, Arendt K, Sangogboye FC, Mattera CG, Markoska E, Kjærgaard MB, et al. ObepME: an online building energy performance monitoring and evaluation tool to reduce energy performance gaps. Energy Build 2018; 166:196–209.

[13] Delzendeh E, Wu S, Lee A, Zhou Y. The impact of occupants' behaviours on building energy analysis: a research review. Renew Sustain Energy Rev 2017; 80:1061–71.

[14] Zou PXW, Xu X, Sanjayan J, Wang J. Review of 10 years research on building energy performance gap: life-cycle and stakeholder perspectives. Energy Build 2018; 178:165–81.

[15] De Wilde P. The gap between predicted and measured energy performance of buildings: a framework for investigation. Autom Constr 2014; 41:40–9.

[16] Alam M, Phung VM, Zou PXW, Sanjayan J. Risk identification and assessment for construction and commissioning stages of building energy retrofit projects. In:

Proceedings of the 22nd International Conference on Advancement of Construction Management and Real Estate; 2017 Nov 20-23; Melbourne, Australia.

[17] Topouzi M, Killip G, Fawcett T, Owen A. Deep retrofit approaches: managing risks to minimise the energy performance gap. In: Proceedings of the Eceee 2019 Summer Study on energy efficiency; 2019 June 3-8; Presqu'ile de Giens, France.

[18] Mojic I, Lehmann M, Van Velsen S, Haller M. ImmoGap - Analysis of the performance gap of apartment buildings. In: Proceedings of the E3S Web Conferences CLIMA 2019 Congress; 2019 May 26-29; Bucharest, Romania.

[19] De Wilde P. Building performance analysis. Hoboken: Wiley Blackwell; 2018.

[20] Galvin R. Making the "rebound effect" more useful for performance evaluation of thermal retrofits of existing homes: defining the "energy savings deficit" and the "energy performance gap".Energy Build 2014; 69: 515–24.

[21] Dollard T. Designed to perform: an illustrated guide to delivering energy efficient homes. London: RIBA Publishing; 2018.

[22] Calì D, Osterhage T, Streblow R, Müller D. Energy performance gap in refurbished German dwellings: lesson learned from a field test. Energy Build 2016; 127:1146–58.

[23] Menezes AC, Cripps A, Bouchlaghem D, Buswell R. Predicted vs. actual energy performance of non-domestic buildings: using post-occupancy evaluation data to reduce the performance gap. Appl Energy 2012; 97:355–64.

[24] Niu S, Pan W, Zhao Y. A virtual reality integrated design approach to improving occupancy information integrity for closing the building energy performance gap. Sustain Cities Soc 2016; 27:275–86.

[25] Cuerda E, Guerra-Santin O, Sendra JJ, Neila FJ. Understanding the performance gap in energy retrofitting: measured input data for adjusting building simulation models. Energy Build 2020; 209:109688.

[26] Hong T, Koo C, Kim J, Lee M, Jeong K. A review on sustainable construction management strategies for monitoring, diagnosing, and retrofitting the building's dynamic energy performance: focused on the operation and maintenance phase. Appl Energy 2015; 155:671–707.

[27] Hong J, Kang H, Hong T. Oversampling-based prediction of environmental complaints related to construction projects with imbalanced empirical-data learning. Renew Sustain Energy Rev 2020; 134:110402.

[28] Alpaydin E. Introduction to machine learning. 2nd ed. Cambridge: MIT Press; 2010.

[29] Awad M, Khanna R. Efficient learning machines: theories, concepts, and applications for engineers and system designers. New York: Apress; 2015.

[30] Olu-Ajayi R, Alaka H, Sulaimon I, Sunmola F, Ajayi S. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. J Build Eng 2022; 45:103406.

[31] Hong T, Wang Z, Luo X, Zhang W. State-of-the-art on research and applications of machine learning in the building life cycle. Energy Build 2020; 212:109831.

[32] Arjunan P, Poolla K, Miller C. EnergyStar++: towards more accurate and explanatory building energy benchmarking. Appl Energy 2020; 276:115413.

[33] Paudel S, Elmitri M, Couturier S, Nguyen PH, Kamphuis R, Lacarrière B, et al. A relevant data selection method for energy consumption prediction of low energy building based on support vector machine. Energy Build 2017;138: 240–56.

[34] Seyedzadeh S, Pour Rahimian F, Oliver S, Rodriguez S, Glesk I. Machine learning modelling for predicting non-domestic buildings energy performance: a model to support deep energy retrofit decision-making. Appl Energy 2020; 279:115908.

[35] Walker S, Khan W, Katic K, Maassen W, Zeiler W. Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. Energy Build 2020; 209:109705.

[36] Nižetić S, Papadopoulos AM. Concept of building evaluation methodology for gap estimation between designed and achieved energy savings. Procedia Environ Sci 2017; 38:538–45.

[37] Mounter W, Ogwumike C, Dawood H, Dawood N. Machine learning and data segmentation for building energy use prediction—a comparative study. Energies 2021; 14(18):5947.

[38] Fan C, Xiao F, Yan C, Liu C, Li Z, Wang J. A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning. Appl Energy 2019; 235:1551–60.

[39] Mocanu E, Nguyen PH, Gibescu M, Kling WL. Comparison of machine learning methods for estimating energy consumption in buildings. In: Proceedings of the 2014 International Conference on Probabilistic Methods Applied to Power Systems; 2014 July 7-10; Durham, United Kingdom.

[40] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 2018; 81:1192–205.

[41] Mohammadiziazi R, Bilec MM. Application of machine learning for predicting building energy use at different temporal and spatial resolution under climate change in USA. Buildings 2020; 10(8):139.

[42] Revati G, Hozefa J, Shadab S, Sheikh A, Wagh SR, Singh NM. Smart building energy management: load profile prediction using machine learning. In: Proceedings of the 2021 29th Mediterranean Conference on Control and Automation; 2021 June 22-25; Puglia, Italy.

[43] Anand P, Deb C, Yan K, Yang J, Cheong D, Sekhar C. Occupancy-based energy consumption modelling using machine learning algorithms for institutional buildings. Energy Build 2021; 252: 111478.

[44] Ngo NT, Pham AD, Truong TTH, Truong NS, Huynh NT, Pham TM. An ensemble machine learning model for enhancing the prediction accuracy of energy consumption in buildings. Arab J Sci Eng 2022; 47(4):4105–17.

[45] Bramer M. Principles of data mining. 4th ed. London: Springer; 2020.

[46] Witten IH, Frank E, Hall MA. Data mining: practical machine learning tools and techniques. 3rd ed. Burlington: Morgan Kaufmann; 2011.

[47] Galar M, Fern A, Barrenechea E, Bustince H. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans Syst Man Cybern 2011; 42 (4): 463-84.

[48] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. Mach Learn 1997; 29:131-63.

[49] Japkowicz N, Shah M. Evaluating learning algorithms: a classification perspective. Cambridge: Cambridge University Press; 2011.

[50] Ayhan M. Development of dispute prediction and resolution method selection models for construction disputes [dissertation]. Ankara (TR): Middle East Technical University; 2019.

[51] Meterko M, Restuccia JD, Stolzmann K, Mohr D, Brennan C, Glasgow J, et al. Response rates, nonresponse bias, and data quality: results from a national survey of senior healthcare leaders. Public Opin Q 2015; 79(1):130–44.

[52] Han J, Kamber M, Pei J. Data mining concepts and techniques. 3rd ed. Waltham: Morgan Kaufmann; 2012.

[53] Larose TD, Larose DC. Discovering knowledge in data: an introduction to data mining. 2nd ed. New Jersey: Wiley; 2014.

[54] Dong XL, Rekatsinas T. Data integration and machine learning: a natural synergy. Proc VLDB 2018; 11(12): 2094-7.

[55] Pegg IM, Cripps A, Kolokotroni M. Post-occupancy performance of five low-energy schools in the UK. ASHRAE Trans 2007; 113 (2):3–13.

[56] Korjenic A, Bednar T. Validation and evaluation of total energy use in office buildings: a case study. Autom Constr 2012; 23: 64–70.

[57] Herrando M, Cambra D, Navarro M, de la Cruz L, Millán G, Zabalza I. Energy performance certification of faculty buildings in Spain: the gap between estimated and real energy consumption. Energy Convers Manag 2016; 125:141–53.

[58] Ram VSS, Kayastha N, Sha K. OFES: Optimal feature evaluation and selection for multi-class classification. Data Knowl Eng 2022;139:102007.

[59] García S, Luengo J, Herrera F. Data preprocessing in data mining. Switzerland: Springer; 2015.

[60] Zhang S. Cost-sensitive KNN classification. Neurocomputing 2020; 391:234–42.

[61] Trivedi SK. A study on credit scoring modeling with different feature selection and machine learning approaches. Technol Soc 2020;63:101413.

[62] Ma Y, Guo G. Support vector machines applications. Switzerland: Springer; 2014.

[63] Rachman A, Ratnayake RMC. Machine learning approach for risk-based inspection screening assessment. Reliab Eng Syst Saf 2019; 185:518–32.

[64] Magoulès F, Zhao HX. Machine learning in building energy analysis. London: Wiley; 2016.

[65] Rokach L, Maimon ZO. Data mining with decision trees: theory and applications.2nd ed.Vol 69. Singapore: World Scientific; 2008.

[66] Chawla NV., Bowyer KW, Hall LO, Kegelmeyer WP.SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 2002; 16: 321-57.

[67] Yao Q, Wang M, Chen Y, Dai W, Li Y-F, Tu W-W, et al. Taking human out of learning applications: a survey on automated machine learning. arXiv: 1810.13306v4 [Preprint]. 2019 [cited 2023 Apr 7]: [20 p.].Available from: https://arxiv.org/abs/1810.13306

[68] Mohr F, Wever M, Hüllermeier E. ML-Plan: automated machine learning via hierarchical planning. Mach Learn 2018; 107(8–10):1495–515.

[69] Truong A, Walters A, Goodsitt J, Hines K, Bruss CB, Farivar R. Towards automated machine learning: evaluation and comparison of AutoML approaches and tools. arXiv: 1908.05557v2 [Preprint]. 2019 [cited 2023 Apr 7]: [9 p.].Available from: https://arxiv.org/abs/1908.05557

[70] Elshawi R, Maher M, Sakr S. Automated machine learning: state-of-the-art and open challenges. arXiv: 1906.02287v2[Preprint]. 2019 [cited 2023 Apr 7]: [23 p.].Available from: https://arxiv.org/abs/1906.02287

[71] Krzywanski J. Advanced AI applications in energy and environmental engineering systems. Energies 2022; 15(15):15–7.

[72] Sangounpao K, Muenchaisri P. Ontology-based naive bayes short text classification method for a small dataset. In: Proceedings of the 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing; 2019 July 8-11; Toyama, Japan.

[73] Stribos RH. The impact of data noise on a naive bayes classifier. In: Proceedings of the 34th Twente Student Conference on IT; 2021 Jan 29; Enschede, Netherlands.

[74] González-Vidal A, Jiménez F, Gómez-Skarmeta AF. A methodology for energy multivariate time series forecasting in smart buildings based on feature selection. Energy Build 2019; 196:71–82.

[75] Li T, Zhang C, Ogihara M. A comparative study of feature selection and multi-class classification methods for tissue classification based on gene expression. Bioinformatics 2004; 20(15):2429–37.

[76] Bouckaert RR, Frank E, Kirkby R, Reutemann P, Seewald A, Scuse D. WEKA manual for version 3-7-8 [Internet]. Hamilton (NZ): The University of Waikato; 2014. Available from: https://statweb.stanford.edu/~lpekelis/13_datafest_cart/WekaManual-3-7-8.pdf

[77] Flach PA, Lachiche N. Naive bayesian classification of structured data. Mach Learn 2004; 57(3):233–69.

[78] Kadyrova NO, Pavlova LV. An analysis of methods for tuning a support-vector machine for binary classification. Biophys 2018; 63(6):994–1003.