Energy & Buildings 258 (2022) 111836

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

A meta-learning classification model for supporting decisions on energy efficiency investments

Elissaios Sarmas [a,*], Evangelos Spiliotis [b], Vangelis Marinakis [a], Themistoklis Koutselis [a], Haris Doukas [a]

[a] Decision Support Systems Laboratory, School of Electrical & Computer Engineering, National Technical University of Athens, Greece
[b] Forecasting and Strategy Unit, School of Electrical & Computer Engineering, National Technical University of Athens, Greece

ARTICLE INFO

Article history:
Received 24 November 2021
Revised 29 December 2021
Accepted 3 January 2022
Available online 12 January 2022

Keywords:
Decision support systems
Energy efficiency
Investment
Machine learning
Meta-learning

ABSTRACT

Energy efficiency is critical for meeting global energy and climate targets, requiring however significant investments. Due to the lack of mature decision-support systems and the utilization of traditional investment mechanisms that focus on the economical aspects of the energy efficiency projects and neglect their environmental impact, such projects can experience difficulties in being funded. In the interim, the impact of the digitization era is more apparent than ever, as algorithms and data availability and quality have significantly improved. This study aspires to bridge the gap in energy efficiency financing with the development of a data-driven methodology that labels energy efficiency investments based on their expected utility in terms of renovation cost and energy savings. Various machine learning classification methods are deployed and combined through a meta-learning model with the objective to improve overall classification performance and determine the funding that each investment should receive according to its particular characteristics. The proposed methodology is evaluated using a set of 312 projects that have been completed in Latvia. Our results indicate that the meta-learner outperforms all baseline classifiers, effectively identifying projects of high and medium potential and successfully distinguishing low from high potential ones.

1. Introduction

Climate change already affects our lives and puts the Earth's ecosystem at risk. Scientists, governments, and institutions alert that we are in an uncharted area full of uncertainty. The latest attempt to reduce CO_2 emissions and impose certain bounds on temperature increasing rate was the Paris Agreement [19]. According to this agreement, the average global temperature increase should be kept below 2°C in the following decades. To achieve this goal, the International Renewable Energy Agency (IRENA) suggests that a complete decarbonization of the energy sector is required by 2050 [30]. The change of energy mixture and the transition to a "green" energy era is not homogeneous throughout the Globe, though; different countries meet different development conditions, i.e. different historical and cultural backgrounds as well as varying economic, social, political, and environmental frameworks [55]. In other words, achieving the targets set globally depends on actions of individual countries or sectors which are not always harmonically acting against greenhouse gas emissions. These barriers in energy transition enhance the importance of supplementary pathways to mitigate the energy footprint, such as the energy consumption reduction ones. This pathway includes the fostering of energy efficiency (EE) investments, especially in buildings, as well as the citizens' willingness to engage in community-based renewable energy projects, which is influenced by many factors according to Kalkbrenner & Roosen [34]. The most important ones are environmental awareness and the prospect of energy self-sufficiency [7], although willingness can also be affected by societal factors, such as trust and social norms [61]. On the other hand, the difficulty of older generations to adopt climate-friendly habits [2], necessitates the adoption of more efficient technologies. In this respect, smart cities and building renovation projects are critical towards a green energy era.

According to the International Energy Agency (IEA), the implementation of EE policies could result in nearly 36% of the avoided greenhouse emissions by 2050 [27,28]. The vast majority of this reduction is expected to come from building renovation projects [59]. However, aside from the theoretical expectations, in practice, EE investments face certain barriers; EE projects are often fragmented and their cost and energy savings are a priori unknown or challenging to estimate accurately. Thus, stakeholders (private financial institutions, investment funds, national and regional authorities, as well as energy solution providers) lack mature

* Corresponding author.
E-mail address: esarmas@epu.ntua.gr (E. Sarmas).

https://doi.org/10.1016/j.enbuild.2022.111836
0378-7788/

decision-making tools that could predict the utility of future EE projects and help them guide their actions more reliably. Moreover, developing countries, which have tremendous potential to increase their efficiency in terms of energy savings, have faced several obstacles, including lack of access to appropriate financing mechanisms [48].

When it comes to financing large-scale EE investments, one of the obstacles present is the potential lack of a sufficient amount of data[1] that could enable assessing their utility but, more importantly, the fact that even when such data exist and are readily available to use, they are not always exploited in a systematic way to effectively support decision making. For instance, financing bodies currently use different methods, standards, and data sources to evaluate the risk and the efficiency of future investments [25], as well as different indicators to account for their economic structure, climate, and geography, among other influencing factors. According to Painuly [47], the barriers to renewable energy may differ across technologies and countries. Especially the EU has recognized this problem, funding research projects which aim to de-risk EE investments, such as *EEnvest*, *Triple-A* and *Quest: 2050*, among others [38]. As a result, EE projects are often perceived to incorporate higher risks and financing institutes tend to focus on other types of projects or traditional investments [53].

In this paper we offer a data-driven perspective to the problem of EE investments assessment. Motivated by the recent advances in the field of machine learning (ML) and its successful use in the area of classification, we propose the use of a meta-learning[2] model that predicts the utility of future EE projects in terms of cost and realized energy savings. Using a rich data set that captures (i) the cost of renovation, (ii) the current energy consumption of the building, (iii) the building's special characteristics, as well as (iv) the expected and (v) realized energy savings of multiple EE projects that have been completed in the past, we train several classification methods to determine the attractiveness of future EE investments (high, medium, or low). Then, a meta-learner is used to define which classifier should be particularly selected for making the prediction according to the particularities of the examined project, thus further improving the accuracy of our results. We demonstrate the merits of our approach and discuss how it can be used in practice to assist stakeholders diversify their EE investments based on their expected utility.

The contributions of our study are summarized as follows:

- We offer a novel ML model that assesses the utility of future EE projects in terms of cost and realized energy savings in a systematic way, thus assisting financing institutes in decision making. Our model is easy to replicate and can therefore be incorporated in the existing systems of the institutes at a low cost.
- Instead of classifying future EE investments using particular indicators that estimate their efficiency theoretically, we base our proposals on the realized utility of multiple, similar EE projects that have been completed in the past, thus empirically learning from their success and limitations and tailoring our suggestions according to the special characteristics of each new investment.
- To mitigate the underlying risks of using a single classifier (e.g. in terms of accuracy and robustness when it comes to small, imbalanced, or noisy data sets) to evaluate future EE investments, we propose the use of a meta-learner which is responsible for identifying the classifier that is expected to result in the most reliable predictions.
- We provide our classification predictions in the form of probabilities to account for uncertainty and use these results to dynamically make recommendations about the percentage of grant financing that future EE investments should receive.

The rest of the paper is organized as follows. In Section 2 the problem setting and literature review are presented. In Section 3 the developed methodology is described. In Section 4 an experimental application based on a real case study in Latvia is presented. Finally, concluding remarks are provided in Section 5.

2. Problem setting and literature review

It is generally acknowledged that energy use in buildings can be significantly reduced through the adoption of existing energy saving technologies. However, EE projects may still experience difficulties in being financed due to the uncertainty and lack of an integrated methodology for assessing and comparing their potential [33]. More specifically, Intrachooto & Horayangkura [29] have identified renovation components interdependency, speculative financial return, and lowest-cost mindset as the main factors from which financial barriers emerge. Currently, EE refurbishments in buildings are assessed by traditional investment delivery mechanisms developed by large financing institutions and banks. Although several projects have been financed through this process, much potential can still be exploited [60].

In recent years, the increasing adoption of information and communication technologies (ICT), such as the internet of things (IoT) and artificial intelligence (AI), have made it possible to obtain large volumes of heterogeneous data which can lead to novel solutions and analyses [40]. More specifically, in the field of EE financing it is now possible to collect data from smart meters after the implementation of refurbishments and obtain insightful information on their actual efficiency. However, EE financing in buildings remains relatively under-researched and under-invested, according to Zhang et al. [67]. Consequently, future studies should focus on the application of AI methods and techniques [15], as well as on the optimal exploitation of available data towards providing tools that accurately assess EE investments [32].

Few research works have attempted to provide a classification prediction[3] framework to support the financing community to identify prosperous future renovation projects [18]. A literature review for EE assessment approaches can be found in Tuominen et al. [63] where the authors concluded that the economic indices are often overlooked or superficially covered when EE is investigated. According to the same study, payback period and cost mitigation are the two most frequent economic objectives considered. In Stocker et al. [57], de Vasconcelos et al. [64], and Ortiz et al. [46] the reader may find studies related to cost-optimal analysis for the energy refurbishment of residential buildings in the cases of Alps, Portugal, and Catalonia, respectively. Apart from the cost and payback period, the financial risk of the projects is also used as a supplementary objective [22]. However, contrary to an intuitive approach, financial efficiency is subset to the EE concept itself; EE usually interplays with other environmental and social aspects [4]. In this context, in the present study we additionally include information about the buildings' current energy consumption and characteristics (e.g. year of construction, heating area, and number of apartments), as well as the CO_2 reductions the EE investments are expected to achieve as fundamental parameters of the EE assessment process.

Another novelty of our study is the utilisation of machine learning (ML) and meta-learning methods [39]. Most of past research works assessed investments using frameworks based on multi-

[1] As an example, many of the projects being currently financed are of small scale or at a pilot-level stage, rendering benchmarking a challenging task.

[2] Meta-learning involves the utilization of ML methods with the objective to learn how to best select or combine the predictions made by other ML methods.

[3] The prediction can be either binary (invest vs. do not invest) or multi-class (e.g. the potential of the investment is low, medium, or high).

criteria decision analysis (MCDA) [54] (either exploiting discrete MCDA methods [49,43], or continuous multiobjective optimization [1]), cost-benefit analysis (CBA) [44], and cost-effectiveness analysis (CEA) [63,21] methods. However, these frameworks require in depth knowledge of the economic, social, and climate factors that affect each investment. Moreover, their performance is strongly influenced by the assumptions made by the respective methods. In this context, ML methods could be used as alternative, data-driven approaches to traditional EE assessment frameworks [66]. A recent review on ML applications in the energy economics and finance area [24] showed that ML methods are mostly used in crude oil and power price forecasting. Despite the constantly increasing utilization of ML theory in the energy sector, the classification of EE investments based on their expected utility is a task that could be further investigated.

This research work attempts to deploy ML methods and data analysis techniques in the EE investments area. In Doukas et al. [20], traditional, statistical classification methods, such as the ordinal logit, the ordinal probit, and the linear discriminant analysis methods along with a limited number of ML methods, such as the k-nearest neighbours and support vector machines were applied. In this regard, the present study additionally considers the Gaussian naive Bayes, extreme gradient boosted trees, and random forest methods in an attempt to improve the accuracy of the predictions made through methods of higher learning capacity. In addition, the proposed framework utilizes a meta-learner with the objective to identify the most accurate classification method for each investment based on its particular characteristics, thus allowing further optimization and increasing the potential value added by the proposed classification process. The contribution of this study can be summarized in the utilization of a meta-learning model with the aim to reduce the inherent risk of using a single classifier, especially with small or imbalanced data sets. Contrary to previous studies, which have proposed tailor-made models for specific data sets, this study provides a generalized framework that generates reliable predictions in terms of accuracy and robustness regardless of the selected features and the amount of available data.

3. Proposed methodology for assessing future energy efficiency investments

The presented methodology aspires to provide a grant financing recommendation for future EE projects by learning from a pool of already implemented projects with similar characteristics. The basis of the proposed approach is a stacking ensemble classification model that builds on five baseline ML classification methods, namely *k-Nearest Neighbors*, *Gaussian Naive Bayes*, *Extreme Gradient Boosted Trees*, *Random Forest*, and *Support Vector Machine*. The train set of the classifiers is essentially a set of projects that involves information about the renovations performed in a variety of buildings in the past. This includes data that was originally available before implementing the renovation (used as features) and data recorder after the investment has been completed (used for labeling). In practice, and to allow for automation, this information should be stored in the database of the information system used by the financing institute to evaluate future EE investments.

Specifically, the features of the investment involve specific attributes of the projects that, in our case, are the renovated building's age, number of apartments, total heating area, and current energy consumption, the expected CO_2 decrease, and the renovation cost. The output of the baseline classifiers are probabilities that the investment belongs to each discrete class among the following three:

- *Class A*: The project should be financed.
- *Class B*: The project should be partially financed.
- *Class C*: The project should not be financed.

In order to label past investments, the realized utility of each project is computed using a measure of choice. In this study we consider the investment efficiency ranking, as described in Section 4.1. Consequently, the top performing projects are labeled as *Class A*, the worst performing projects as *Class C*, while the rest as *Class B*. Having labeled the investments, the baseline classification methods are used for classifying past projects and their predictions are stored. Using these predictions as input along with the actual labels of the investments and the features originally used by the baseline classifiers, a meta-learner, namely a logistic regression classifier, is trained with the objective to predict which of the five baseline methods is the most appropriate for predicting the class of each project. Given this insight, the "optimal" classifier is used for each investment to derive the probability of each class. These probabilities are finally combined in order to provide a recommendation on the percentage of grant financing for a future investment, as described in Section 3.3. The complete recommendation process is summarized in Fig. 1.

Note that the framework of the proposed methodology is very flexible in terms of investment features, classes, methods, and utility measures used. For instance, decision makers can adjust the input variables used by the classifiers depending on their preferences or data availability. Similarly, they can modify the set of baseline classifiers, considering different number and types of methods. Finally, they can define the number of labels considered by the classification methods, as well as the approach/measures used for their determination. As such, the choices made in our study for setting up the framework of the described methodology and demonstrating its use should be considered indicative, although they do utilize a meaningful set of parameters.

Note also that the flexibility offered by the proposed model can in general be counterbalanced effectively in terms of classification performance through the meta-learner used. Depending on their algorithmic nature, different classification methods may be more appropriate for handling small or imbalanced data sets, as well as for making accurate predictions when multiple features are provided as input, especially in cases where not all features are of critical or of the same importance. For instance, decision-tree based methods (e.g. XGBoost and RF) can naturally select the most relevant features for making predictions based on the contribution of each feature in the overall reduction of the forecast error when fitting the method. Moreover, some methods (e.g. RF) are particularly effective in handling multiple feature, while others in dealing with data randomness (e.g. RF and SVM). This is in contrast to methods like k-NN and Gaussian naive Bayes that, although more effective with small train sets (less "data-hungry" algorithms), they cannot automatically select the most relevant features and may therefore be negatively affected when multiple variables are provided as input. It is exactly the purpose of the meta-learner to account for such issues and particularities, mitigate the underlying risks of using a single classifier to evaluate future EE investments, and provide as reliable predictions as possible given the data set available for training and the key parameters of the examined classification problem.

3.1. Baseline classification methods

The proposed methodology builds on five baseline classification methods, selected so that the meta-learner selects the most appropriate classifier from a truly diverse[4] set of methods, each making

[4] Diversity refers to the different algorithmic nature of the selected ML methods and the strengths and weaknesses that each of these methods is expected to display by design in terms of accuracy and robustness when it comes to small, imbalanced, or noisy data sets, as well as to handling multiple features. These properties are summarized in the introductory part of Section 3 and described in detail in the following subsections.

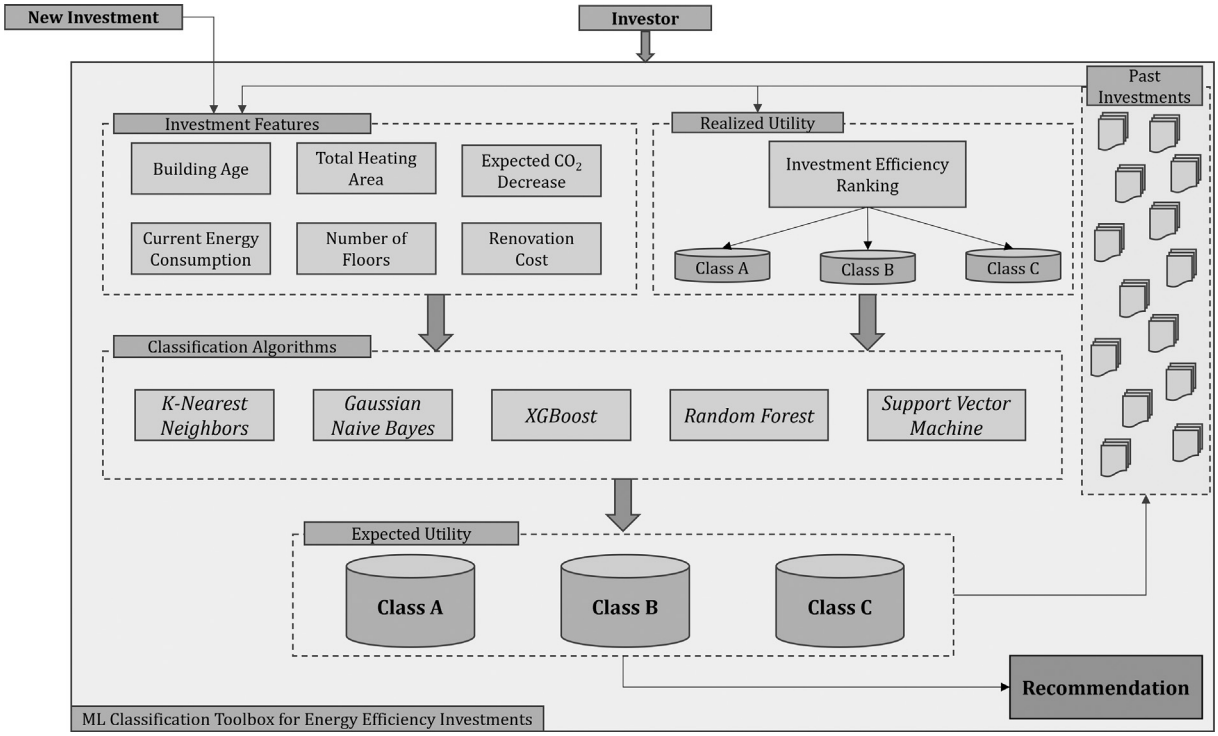


Fig. 1. Overview of the proposed methodology. The recommendation on the grant financing percentage of future investments is based on a data-driven approach that exploits useful insights from past projects, labeled based on the achieved energy consumption reduction and the investment cost (realized utility). The output of the baseline classification methods, as well as the meta-learner, provides the percentage of recommended grant financing.

different assumptions about the prediction task. This approach has been proven effective for improving overall accuracy [58]. The baseline classification methods are shortly presented below.

3.1.1. k-Nearest neighbors

The k-nearest neighbors (k-NN) is a non-parametric classification method [14]. The input of the method consists of n features that characterize each observation (in our case an EE investment), while the output is a label that determines its class (in our case the realized utility of the investment). Effectively, the method classifies future unlabeled observations by identifying the k most similar labeled observations (nearest neighbours) to the examined one and considering the plurality of their classes.

Specifically, the method consists of two phases; the train phase and the test phase. During the train phase the labeled observations, which generally are vectors in a multidimensional space, each mapped to a class, are stored and the number of neighbors k is defined. During the test phase unlabeled vectors are classified by mapping them to the class which most frequently appears among the k labeled observations nearest to the test vector. The term "nearest" can be interpreted and calculated using several distance measures according to the type of the features used as input. For continuous variables the Euclidean distance is the most commonly used measure, while for discrete variables the Hamming distance prevails in the literature [45]. Another strategy is the employment of correlation coefficients, such as Pearson's correlation [36]. Since the method is based on calculating distances among the input features, normalizing the values of the features is recommended for improving accuracy [62].

Defining k, essentially the only hyperparameter of k-NN, is critical because it directly affects the performance of the labeling process. If k is set to unity, then unlabeled observations are classified according to the label of their nearest neighbor. If k is set equal to the number of observations included in the train set, then unlabeled observations are classified according to the most popular label in the complete train set. Since the "optimal" value of k is typically subject to the particularities of the train set and the classification task, it is usually defined through heuristic cross-validation techniques, widely used in the ML literature for hyperparameter fine-tuning.

In the context of this study, the features considered are continuous variables. As a result, the Euclidean distance was used to determine the nearest neighbours. Given the Cartesian coordinates of two points p and q in an n-dimensional Euclidean space, the distance is computed as follows:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2}. \quad (1)$$

3.1.2. Gaussian Naive Bayes

The family of probabilistic classifiers includes classifiers that are able to predict the probability distribution over an available set of classes, instead of predicting a single class prediction that the observation should belong to [42]. Naive Bayes classifiers are a sub-category of probabilistic classifiers which rely on Bayes' theorem, assuming strong independence between the features. Gaussian naive Bayes has been utilized in many areas, including tasks related with the assessment of investment performance [13], forecasting of photovoltaic production [5], and analysis of buildings' energy efficiency [52].

According to Zhang [68], given a feature vector x_1 to x_n and a class variable y, the following relationship is stated by the Bayes' theorem.

$$p(y|x_1,\ldots,x_n) = \frac{p(y)\ p(x_1,\ldots,x_n|y)}{p(x_1,\ldots,x_n)}. \quad (2)$$

Using the naive conditional independence assumption for all i, this relationship can be transformed to the following equation:

$$p(y|x_1,\ldots,x_n) = \frac{p(y)\ \prod_{i=1}^{n} p(x_i|y)}{p(x_1,\ldots,x_n)}. \tag{3}$$

The classification rule given by the following equation derives since $p(x_1,\ldots,x_n)$ is constant given the input:

$$\hat{y} = \underset{y}{\operatorname{argmax}}\, p(y)\prod_{i=1}^{n} p(x_i|y). \tag{4}$$

There are different naive Bayes classifiers depending on the assumptions made regarding the distribution of $p(x_i|y)$. The inspected problem is set on continuous data, thus we can make the assumption that the continuous values associated with each class follow the Gaussian distribution [31]. Therefore, the Gaussian naive Bayes is utilized, where the likelihood of the features is assumed to be Gaussian:

$$p(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}}\, e^{-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}}. \tag{5}$$

3.1.3. Extreme gradient boosted trees

Classification trees is a ML method that recursively partitions the data space through rules to classify a set of observations [8]. Since partition rules are constructed sequentially, the tree starts from a root, where the complete train set is included, and splits the observations available to branches, each containing the observations that satisfy the first rule of the tree. Then, each branch can be split further using additional rules. Branches that do not split are called leaves and include the final predictions of the tree. In order to determine which leaf should be used for predicting each observation, the conjunctions of rules must be considered. The rules constructed are build on the features available for training the method and are defined automatically using various criteria, such as the Gini impurity and information gain. Also, more often than not, additional criteria (e.g. learning rate, maximum number of leaves, maximum tree depth, minimum sum of instance weight needed in a child, and minimum loss reduction required to make a further partition on a leaf node of the tree) are used in the form of hyperparameters to specify when the growth of the tree should be terminated, thus avoiding overfitting and reducing computational cost.

Gradient boosting (GB) is a technique that can be equipped to any ML method to reduce the bias and variance of its predictions and, as a result, convert weak learners to strong ones. Consequently, GB combines multiple weak learners sequentially to create a single strong learner of higher learning capacity and improved accuracy [23]. This combination is performed so that each new weak learner is specialized on improving the predictions of the previous ones, i.e. minimize previous prediction errors. In this context, GB can also be used to enhance the performance of classification trees and construct more powerful classifiers.

Extreme gradient boosting (XGBoost) is probably the most popular implementation of gradient boosted trees [11]. It exploits memory and hardware resources for tree boosting methods, thus resulting in superior speed and performance [16]. Also, XGBoost incorporates three well-known GB techniques, namely gradient, regularized, and stochastic boosting. In this study, we implemented XGBoost using gradient boosting.

3.1.4. Random forest

Random forest (RF) is an ensemble ML classification method which is based on developing multiple classification trees and considering the plurality of their votes to make a final prediction [9]. In order for the constructed trees to exploit the benefits of combining, reducing both the bias and variance of the predictions, each classification tree is trained on different observations and using a different number of features, randomly sampled from those originally available. This technique, widely known as bagging, ensures that the created trees will be truly diverse, having access to different information and being less sensitive to changes made in the train set.

In the energy and building sector, many studies have used RF to forecast energy consumption in buildings [3], develop personalized conditioning systems in offices [37], and detect faults in distribution networks [10], among others.

3.1.5. Support vector machines

Support vector machines (SVMs) rely on the construction of a hyper-plane or set of hyper-planes in a high or infinite dimensional space which, among others, can be used for solving binary and multi-class classification tasks [56]. The hyper-plane with the largest functional margin (distance to the nearest data points of the train set) achieves lower generalization errors and therefore provides a better separation of the data. In such settings, SVMs are typically called support vector classifiers (SVCs) [41,6].

Given the vectors $x_i \in \mathbb{R}^p,\ i = 1,\ldots,n$and a vector $y \in \{1,-1\}$, SVC calculates the values $w \in \mathbb{R}^p$ and $b \in \mathbb{R}$ in order to generate correct predictions for any observation using the value of $\text{sign}(w^T\phi(x)+b)$. The primal problem solved by SVCs can be formulated a follows:

$$\begin{aligned} \min_{w,b,z} \quad & \tfrac{1}{2}w^T w + C\sum_{i=1}^{N} z_i, \\ \text{s.t.} \quad & y_i(w^T\phi(x_i)+b) \geqslant 1 - z_i, \\ & z_i \geq 0. \end{aligned} \tag{6}$$

This mathematical formula describes that SVCs attempt to maximize the margin, while also imposing a penalty C for observations that are misclassified or fall within the margin boundary. Finally, the distance factor z_i for each vector X_i allows an observation to be at a certain distance from the correct margin boundary, because the majority of problems are not separable by a hyper-plane.

For multi-class classification the SVC problem setting differs significantly from the binary classification approach [26,17]. There are two approaches used to generalize from binary to multi-class classification. The direct method suggests that one binary classifier is generalized in order to create multi-class predictions, while the most common approach, the indirect one, proposes combining N independent binary classifiers to produce a single multi-class one. In the latter approach, binary tasks can be defined in the form of a matrix of size $M \times N$, where M represents the number of classes and N the number of tasks. The values of each matrix element $R_{ij} \in \{-1, 0, 1\}$ generated by the binary classifiers f are used to predict the class label according to the following equation:

$$\hat{y} = \arg\min_{y\in 1,\ldots,n}\left\{\sum_{k=1}^{N} R_{yk} f^k(x)\right\}. \tag{7}$$

3.2. Stacking ensemble model

Stacking (or stacked) ensemble is a scheme used for minimizing the generalization error of multiple prediction methods considered in an ensemble [65]. Typically, stacking ensembles utilize meta-learners which are responsible for combining the predictions from multiple baseline ML methods in an optimal way or simply selecting the prediction that is expected to be the most accurate for a particular case. Consequently, this approach can exploit various well-performing methods that display diverse characteristics, have

different structures, and make different assumptions about the data.

The utilization of stacked ensemble modeling was inspired because of the low correlation of the prediction errors usually made by different baseline methods. As a result, this technique is appropriate when the prediction errors of the baseline classifiers are uncorrelated, signifying that each method captures different attributes of the problem and, therefore, is more skillful in different perspectives. When this assumption is met, tacked ensembles generate more accurate predictions than any single baseline method by reducing the biases of the individual methods with respect to the provided train set.

The developed stacked ensemble model has a simple architecture, composed of two levels of classification methods. The first layer includes the five baseline classifiers described in Section 3.1, to be called "level-0" methods. The selected meta-learner, to be called "level-1" method, is a logistic regression method. Our choice was based on the interpretability this method offers, providing a straightforward reasoning for its predictions, i.e. the factors that led the meta-learner select a baseline classifier over others. Several studies have proposed the utilization of relatively simple classifiers as level-1 models to optimally integrate the predictions of the baseline methods [12,50]. The overall architecture of the stacked generalization model is summarized in Fig. 2.

The overall procedure of creating stacking model can be divided in three phases; the ensemble set up, the training, and the prediction on new data. These phases are described below:

Phase 1 – Ensemble Set Up: In this phase, the baseline and meta-learning classifiers are selected. The baseline methods should incorporate different characteristics and derive from a different category of algorithms (e.g. they should not include only tree-based methods). The meta-learner should be a relatively simple and interpretable classifier.

Phase 2 – Training: Initially, each of the baseline methods are trained using an appropriate set of hyperparameters, defined through a cross-validation process (e.g. k-fold cross-validation). Then, the meta-learner is trained using the predictions and the input vectors of the level-0 methods as input and the actual labels of the observations as output.

Phase 3 – Prediction: The stacked ensemble is used to generate predictions on test data. To do so, each baseline classifier is first used to make an individual prediction. Then, these predictions are fed to the meta-learner to determine, based on its input vector, which baseline predictions should be used.

3.3. Recommendation

The proposed methodology aspires to provide recommendations about the percentage of grant financing that future investments should receive. If the predicted class of the meta-learner was used for determining this percentage, then the suggestions of the decision-support system would essentially be discrete numbers, getting one the three possible percentages assigned to each class. For instance, if we assumed that *Class A* suggested 100% financing, *Class B* 50% financing, and *Class C* 0% financing, then all projects would be financed by a factor of $f_A = 1, f_B = 0.5$, or $f_C = 0$, respectively. It becomes evident that this approach significantly limits the value added by the methodology, ignoring also the uncertainty around the predictions. For example, assume a scenario where an investment is labeled with a probability of 0.55 to *Class A*, 0.35 to *Class B*, and 0.10 to *Class C*. Using the aforementioned approach, the project would be 100% financed, although there is evidence against that decision.

To mitigate those issues and account for uncertainty, we propose building financing recommendations on the probability of each class instead of the dominant class itself, summarising the likelihood that a future investment belongs to each class. Given the variables p_A, p_B, and p_C, which represent the probability that an investment belongs to classes A, B and C, respectively, as defined by the meta-leaner, as well as the grant financing factors f_A, f_B, f_C, the financing formula of a future investment is given by the following equation:

$$\text{Financing Percentage} = \sum_{i \in A,B,C} f_i \times p_i. \tag{8}$$

According to the equation above, in our previous example the decision-support system would recommend financing the new project by $0.55 \times 1.00 + 0.35 \times 0.50 + 0.10 * 0.00 = 72.5\%$, thus effectively reflecting the insights of the proposed data-driven methodology.

4. Case study

This section provides an extensive experimental application of the proposed methodology on data stemming from real renovation projects. The training process of the five baseline classification methods is described in detail, including the fine-tuning of their hyperparameters, followed by the training of the meta-learning model. Finally, the results of the application are presented and discussed.

4.1. Data set

The experimental application of the proposed methodology has been applied on data collected from the Latvian Environmental Investment Fund (LEIF), an organization owned by The Ministry of Environmental Protection and Regional Development of Latvia, established in 1997. Starting from 2009, LEIF has supervised the implementation and post-implementation monitoring of several projects co-financed by climate change financial instruments, thus being the only institution in Latvia that possess reliable data on the actual performance of various EE investments in terms of energy savings.

Although our approach is based on assessing single-building investments, many of the collected records concerned investments that were applied on more than one buildings. Therefore, these records were excluded from the data set, including a final sample of 312 EE projects. The selected ML methods were trained using a randomly selected 80% of the total sample, i.e. 249 investments, while the remaining 63 investments were utilized to test the performance of our methodology.

The above-mentioned investments were classified using three labels: *Class A* including high potential investment, *Class B* including medium potential investments, and *Class C* including low potential investments. The labelling of the investments was based on the investment efficiency ranking. This is a simple ranking method using two criteria; the investment cost of the project and the energy consumption reduction on the building after the refurbishment took place. These criteria were combined using the following steps: First, the values for each criterion were normalized to a $(0, 1)$ scale. Second, the weighted average of the normalized values was calculated for all the available investments. This process was performed using equal weights for the selected criteria; nevertheless the selection of the weighting factors can be customized according to the preferences of the investor. Finally, the values were ranked from best to worst. The top 20% of the projects were labelled as *Class A* investments, while the bottom 20% of the projects were labelled as *Class C* investments. The remaining 60% of the investments were labelled as *Class B*. A representation of the investments' cost against the achieved energy consumption reduc-

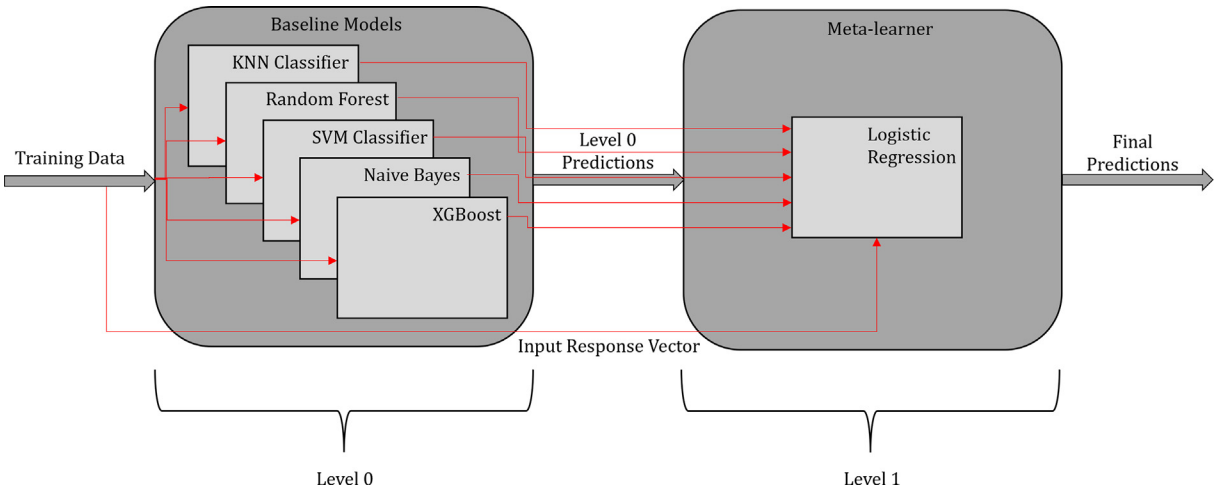


Fig. 2. Overview of the proposed stacked ensemble model. The baseline classification methods (level-0) are trained following the process described in Fig. 1 and provide predictions for future investments. These predictions are then fed to the meta-learner along with the input vectors of the baseline methods. The output of the meta-learner is the final prediction of the stacked ensemble model, i.e. the prediction of the most appropriate baseline classifier.

tion is shown in Fig. 3. Projects with green color are *Class A* investments, yellow-colored projects are *Class B* investments, while red-colored ones are *Class C* investments.

Observe that, by design, the labeling system used results in an imbalanced train set (biased sampling), with *Class B* investments having a higher probability for being identified as the correct class. Although this issue can be tackled using resampling techniques, such as methods that randomly under-sample the majority class (*Class B*) or generate synthetic data (over-sample) for the minority classes (*Class A* and *Class C*; [35]), the meta-learner should still be able to deal with class imbalance by selecting the baseline classification method that best predicts each class of investment. Moreover, since imbalanced data sets are typical in practice, with some classes being naturally more frequently observed than others, this kind of imbalance serves as an additional stress-test for evaluating the value added by the proposed approach.

The investment features used in our case study were partially subject to data availability. In this regard, we selected six key features as input to the classification methods, including the construction year of the building, the total heating area (m^2), the expected CO_2 decrease due to the refurbishment actions ($kgCO_2$), the current energy consumption of the building (MWh), the number of floors of the building that will be renovated, and the amount of the requested grant financing for the renovation (€). Since all investments were made in the same country, we did not include geographical variables (e.g. town and city identifiers) as additional feature. The pairwise relationships between the selected investment features are shown in Fig. 4.

4.2. Training and hyperparameter fine-tuning

The optimal hyperparameters for the five baseline classification methods were configured through a stratified k-fold cross-validation process. The term stratified indicates a variation of the traditional k-fold cross-validation process, where the folds are made by preserving the percentage of samples for each class. The reason that stratified k-fold was preferred over its standard counterpart is that we are facing a classification task with imbalanced class distributions, consisting of 20% *Class A* investments, 60% *Class B* investments, and 20% *Class C* investments. The number of splits was set to 10, while the number of repeats for stratified k-fold was set to 3. The selected hyperparameter values for each method

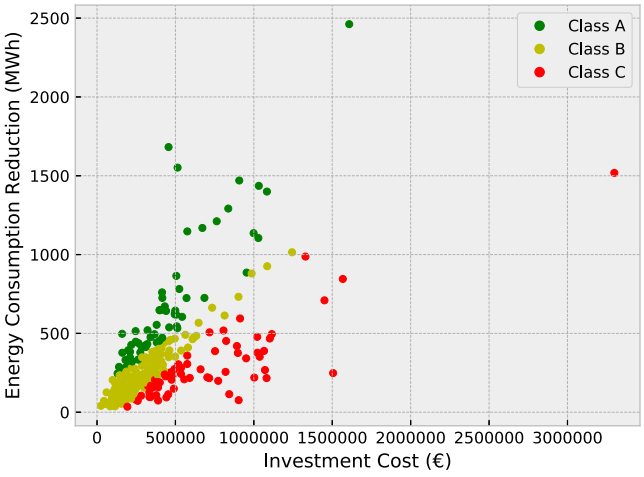


Fig. 3. Investment labeling of the examined investments based on the investment efficiency ranking.

and the functions used for their implementation are summarized below.

- *k-Nearest Neighbors*: The most critical hyperparameters in this method are the number of neighbors used for running the queries, set equal to 6, and the weight function employed for making the predictions, set to the uniform one. The *neighbors* function of the *sklearn* library for Python was used for implementing the method [51].
- *Gaussian Naive Bayes*: The method does not require fine-tuning its hyperparameters. The *naive_bayes* function of the *sklearn* library for Python was used for implementing the method.
- *Extreme Gradient Boosted Trees*: XGBoost involves various hyperparameters which are crucial for its performance. The most important ones are the *learning rate*, the decision tree's *number of leaves*, the *bagging fraction*, allowing to decrease the variance in the prediction, and the *feature fraction*, allowing the random selection of a subset of features on each iteration. Having implemented the cross-validation process, the following values were selected: Learning rate = 0.95; maximum tree depth = 10; minimum sum of instance weight needed in a child = 1; number of

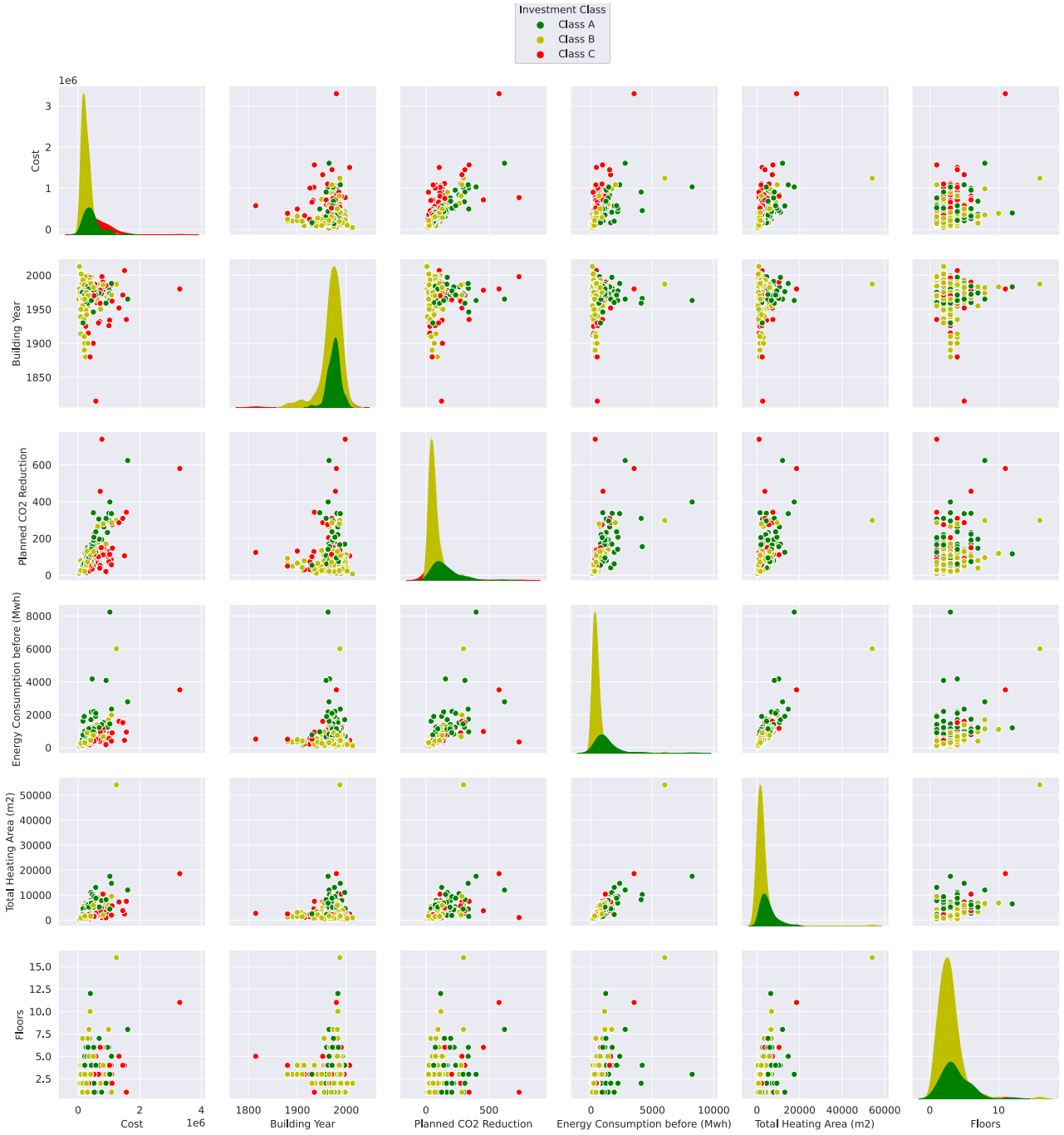


Fig. 4. Pairplot of the investment variables considered by the proposed classification methods in the examined case-study: Cost (€), Building Construction Year, Planned CO_2 Reduction ($kgCO_2$), Energy Consumption Before (MWh), Total Heating Area (m^2), Floors.

gradient boosted trees = 1000; all other hyperparameters were set to their default value. The *Classifier* function of the *XGBoost* library for Python was used for implementing the method [11].

- *Random Forest*: The number of trees (number of estimators) in the RF method was set to 500, the factor of weights associated with classes was set to balanced (using the values of the target variable to adjust weights inversely proportional to class frequencies in the feature variables) and the maximum number of features used in each iteration was set to log_2. The *ensemble* function of the *sklearn* library for Python was used for implementing the method.
- *Support Vector Machine*: The SVM employed the linear kernel, while the regularization parameter C was set to 1. The *svm* function of the *sklearn* library for Python was used for implementing the method.

Following the hyperparameter configuration for the baseline classification methods, the final step of the training process involves the training of the level-1 model on the same train set. In this study, the meta-learner was a logistic regression classifier due to its simplicity, which is recommended for level-1 predictors. Since the selected meta-learner does not have any hyperparameters to tune, it was trained directly on the level-1 training set

which consists of the level-0 predictions from the baseline methods and the real labels. The baseline methods predictions which formed the level-1 training set were made through a 3-fold cross validation process on the training set.

The performance of the baseline methods and the stacking classifier in the stratified k-fold validation process (10 folds $\times$ 3 repeats = 30 validations) is summarized in Table 1. From the baseline classifiers, the SVM method reports the best performance according to the F1 score, followed by the RF and XGBoost methods. As discussed in Section 3, this can be attributed to the ability of the XGBoost and RF methods to select the most relevant features from a large set of explanatory variables, as well as the fact that RF and SVM are generally better in dealing with data randomness. Moreover, we find that the stacking model outperforms all baseline methods according to both the accuracy and the F1 score measures, being slightly worse than the SVM in terms of precision. Note that although the differences between the meta-learner and the SVM classifier are small according to the classification measures used, the former is more robust, having a tendency to classify most of the investments more precisely, while avoiding extreme errors. This becomes evident by observing Fig. 5, which provides a visual comparison of the classifiers' performance in the form of box-plots. As explained, the median of the F1 score and the accuracy measures is higher for the meta-leaner compared to the SVM (more investments are classified correctly), while its interquartile range is significantly smaller (better robustness).

4.3. Prediction for future investments

After training the stacking model, the performance of the proposed methodology is evaluated on the test data set which, as explained earlier, is composed of 63 randomly selected EE investments. The results of this application are summarized in the confusion matrix of Fig. 6, presenting the number of cases the meta-learner has predicted each class compared to their actual labels. We observe that 32 out of 36 *Class B* investments (88.8%) have been correctly predicted. For *Class A*, the respective number is 10 out of 14 (71.4%), while for *Class C*, 7 out of 13 (53.2%). Interestingly, the classifier has never predicted *Class A* investment as *Class C* or the opposite. This finding is particularly encouraging, suggesting that the risk of using the proposed meta-learner for recommending financing (or not financing) an investment of low (or high) potential is insignificant.

The classification performance of the meta-learner is summarized in Table 2. This includes the precision, recall, and F1 score of each class, as well as the macro and weighted average. The macro average is the simple arithmetic mean of each measure across all classes. Since this measure assigns equal weights to all classes, it is appropriate for balanced classification tasks. On the contrary, the weighted average incorporates class imbalance, computing the average of binary measures by considering the number of samples of each class in the target. Therefore, the weighted average is more appropriate for summarizing the performance of the meta-learner in the present case-study.

It is evident that the model performs slightly better with *Class B* investments, i.e. the most popular class, resulting in 0.76 precision and 0.82 F1 score. The respective values for *Class A* are 0.83 and 0.77, while for *Class C* 0.78 and 0.64. Therefore, we find that it is highly unlikely for the meta-learner not to correctly identify *Class A* and *Class B* investments when present, in contrast to *Class C* investments where it is likely for the meta-learner to classify them as *Class B*. As a result, we can conclude that the meta-learner can be effectively used in practice for identifying investments of high and medium potential and supporting the financing of large-scale EE projects. This is confirmed by the overall accuracy of the model, being 0.78 across all classes. Observe also that although the data set used for training the baseline classification methods was imbalanced, the recall scores reported in Table 2 suggest that the meta-leaner has effectively tackled this issue, providing unbiased forecasts, especially when it comes to predicting *Class A* and *Class B* investments.

Table 1
Classification performance (mean and standard deviation) of the baseline methods and the stacking model on the stratified 10-fold validation process. Column-wise best values are displayed in bold.

Classifier	Accuracy	Precision	F1 Score
k-Nearest Neighbors	0.77 (0.08)	0.78 (0.10)	0.74 (0.09)
Gaussian naive Bayes	0.70 (0.10)	0.66 (0.14)	0.67 (0.12)
XGBoost	0.78 (0.10)	0.79 (0.10)	0.78 (0.10)
Random Forest	0.80 (0.09)	0.82 (0.08)	0.79 (0.09)
Support Vector Machine	0.82 (0.09)	**0.85 (0.07)**	0.82 (0.09)
Stacking Model	**0.83 (0.09)**	0.84 (0.09)	**0.83 (0.09)**

Since the accuracy measure does not integrate class imbalance, which is a very significant factor for this problem (60% of the observations are in the same class), we exploit Cohen's kappa coefficient (κ). This coefficient is a statistical index used to measure inter-rater reliability for categorical data, being a more robust measure than simple percent agreement calculation, due to the parameter κ which incorporates the possibility of the agreement occurring accidentally. The formula for Cohen's kappa coefficient is given as follows:

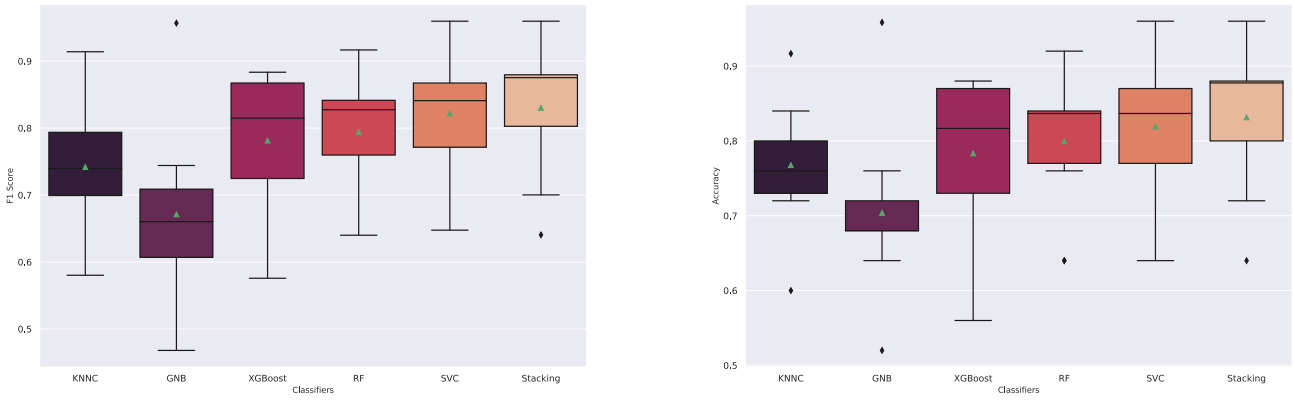
$$\kappa = \frac{p_0 - p_e}{1 - p_e}, \tag{9}$$

where p_0 is described as the observed proportional agreement between actual and predicted values and p_e is the expected agreement when both annotators assign labels randomly. The Cohen's kappa score for the test set is 0.594, which denotes substantial agreement for the stacking classifier.

However, the proposed methodology is based on classifiers that can generate class membership probabilities. Therefore, the abovementioned metrics fail to describe the whole picture. The model's ability to distinguish an instance between each class can be described by the area under the receiver operating characteristic curve (ROC AUC). ROC AUC is calculated, generating predictions for a range of the decision threshold values $[0, 1]$. For each value the true positive rate (TPR) - also known as recall- and false positive rate (FPR) are calculated. In the case of multi-class classification, the measure is calculated separately for each class following the one-vs-rest (OVR) classification strategy, assuming a different classifier per class. The ROC curve for each class is shown in Fig. 7. A large area in the ROC AUC means that there is great distinction between the classes. The value of the area under the ROC curve for the stacking classifier overall is 0.883, which is a good score given that AUC values lie between 0.5 to 1 (where 0.5 denotes a bad classifier and 1 denotes an excellent classifier).

5. Conclusions

This study focuses on the problem of assessing the potential of future EE investments in terms of renovation costs and realized energy savings. Given the lack of mature information systems to support such assessments, several renovation projects are currently straggling to receive financial support by funding institutions, thus putting the global environmental targets set for the next decades at risk. In this respect, a data-driven methodology is proposed which aspires to pave the way towards accurately identifying attractive EE projects and determining the funds that should be invested per case. To do so, critical investment features of projects that have been completed in the past are considered



(a) F1 Score (b) Accuracy

Fig. 5. Comparison of the baseline classifiers and the stacking on the stratified 10-fold validation process. The left sub-figure shows the F1 score of the classifiers, while the right one their accuracy.

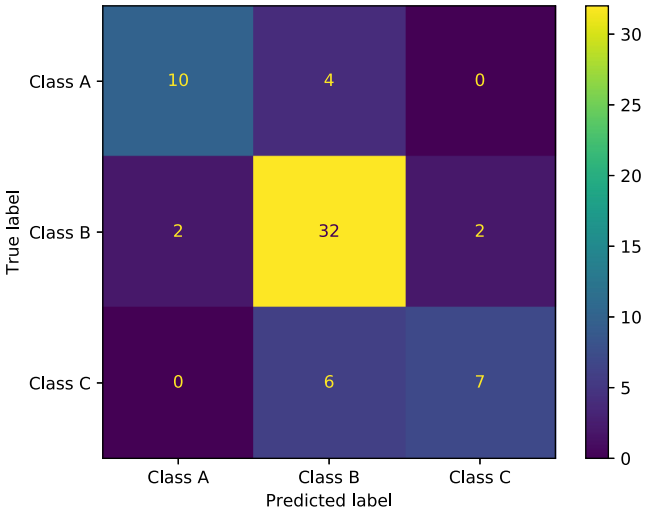


Fig. 6. Confusion matrix for the stacking model on the test set.

Table 2
Classification performance of the stacking model on the test set.

	Precision	Recall	F1 Score
Class A	0.83	0.71	0.77
Class B	0.76	0.89	0.82
Class C	0.78	0.54	0.64
Macro Average	0.79	0.71	0.74
Weighted Average	0.78	0.78	0.77

and ML methods are used to learn from their successes and mistakes. The recommended approach aims to enhance EE financing procedures and facilitate different types of stakeholders (financiers, bankers, investors, etc.) in comparing and labeling EE projects in a standardised and less uncertain way.

The proposed methodology involves a stacking ensemble model that builds on several ML baseline classification methods with the objective of further improving their performance. Data containing significant attributes of EE investments (e.g. building age and total heating area, expected CO_2 emissions, and renovation cost) are collected and used for training the classification model. The methodology is evaluated considering a real case study in Latvia. Our results suggest that the stacking ensemble model outperforms all the baseline ML classification methods considered, achieving high accuracy when used to assess future EE investments. Moreover, we find that the proposed model can effectively identify projects of high and medium investments, being also excellent in distinguishing low from high potential ones. This finding indicates that stakeholders can exploit the presented methodology to reduce their risk of investments and maximize their returns.

Given the limitations of the present study, future work on the field of EE investments assessment should focus on incorporating other types of investments coming from the manufacturing, transportation, and outdoor lighting sector, thus widening the projects considered by the methodology and increasing its economic and environmental potential. The developed meta-learning model can be generalized for such applications provided that sufficient data are available for supporting its learning process. Another point of interest that could be further investigated is how the geographic topology of the investments affects their efficiency in terms of energy savings. In this respect, EE investments implemented in multiple countries with miscellaneous characteristics could be integrated into the model, posing some additional geographic and climate-related variables to generalize its use and broaden the type of stakeholders that would be interested in its utilization. Alternative measures to the examined ones could also be considered for assessing the potential of future EE projects to make the predictions of the proposed method better reflect the ground truth. More importantly, and in order for the proposed approach to become more relevant for large-scale applications, alternative ways for retrieving the energy- and investment-related data required for training the classifiers should be identified, including data sources that public authorities and financial institutions can easily access. For instance, one alternative would be to exploit the information provided by energy certificates of renovated buildings, i.e., track the EE of multiple buildings before and after particular renovations have taken place, as well as the costs of these renovations. This type of information is standardized, interpretable, and easier to share with interested parties, thus facilitating the utilization of the proposed assessment method on a wider scale.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared

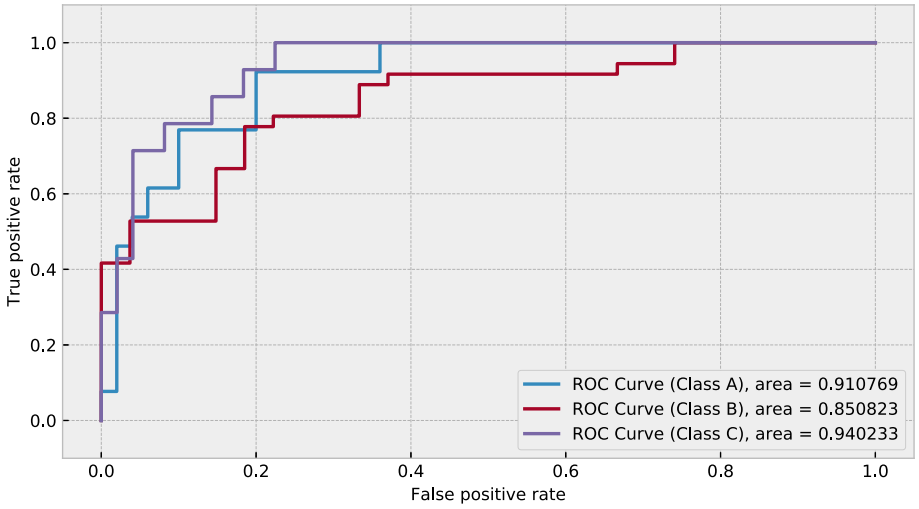


Fig. 7. ROC curve for the three classes based on the stacking classifier's generated probabilistic predictions.

to influence the work reported in this paper. We also confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

Acknowledgments

The work presented is based on research conducted within the framework of the H2020 European Commission project BD4NRG (Grant Agreement No. 872613). The authors wish to thank Aija Zučika and Gints Kārkliņš from the Latvian Environmental Investment Fund (LEIF), whose contribution, helpful remarks and fruitful observations were invaluable for the development of this work. The content of the paper is the sole responsibility of its author and does not necessary reflect the views of the EC.

References

[1] N. Abdou, Y.E. Mghouchi, S. Hamdaoui, N.E. Asri, M. Mouqallid, Multi-objective optimization of passive energy efficiency measures for net-zero energy building in morocco, Building and Environment 204 (2021) 108141.
[2] M.I. Abreu, R.A. de Oliveira, J. Lopes, Younger vs. older homeowners in building energy-related renovations: Learning from the portuguese case, Energy Reports 6 (2020) 159–164.
[3] M.W. Ahmad, M. Mourshed, Y. Rezgui, Trees vs neurons: Comparison between random forest and ann for high-resolution prediction of building energy consumption, Energy and Buildings 147 (2017) 77–89.
[4] A. Arsenopoulos, E. Sarmas, A. Stavrakaki, I. Giannouli, J. Psarras, A data-driven decision support tool at the service of energy suppliers and utilities for tackling energy poverty: A case study in greece, in: 2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA), IEEE, 2021, pp. 1–6.
[5] R. Bayindir, M. Yesilbudak, M. Colak, N. Genc, A novel application of naive bayes classifier in photovoltaic energy prediction, in: 2017 16th IEEE international conference on machine learning and applications (ICMLA), IEEE, 2017, pp. 523–527.
[6] G. Bo, H. Xianwu, Svm multi-class classification, Journal of Data Acquisition & Processing 21 (2006) 334–339.
[7] F.P. Boon, C. Dieperink, Local civil society based renewable energy organisations in the netherlands: Exploring the factors that stimulate their emergence and development, Energy Policy 69 (2014) 297–307.
[8] L. Breiman, Classification and Regression Trees, Chapman & Hall, Boca Raton, FL, 1993.
[9] L. Breiman, Bagging predictors, Machine Learning 24 (1996) 123–140.
[10] D. Chakraborty, U. Sur, P.K. Banerjee, Random forest based fault classification technique for active power system networks, in: 2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE), IEEE, 2019, pp. 1–4.
[11] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[12] Z. Chen, P. Zhao, F. Li, T.T. Marquez-Lago, A. Leier, J. Revote, Y. Zhu, D.R. Powell, T. Akutsu, G.I. Webb, et al., ilearn: an integrated platform and meta-learner for feature engineering, machine-learning analysis and modeling of dna, rna and protein sequence data, Briefings in Bioinformatics 21 (2020) 1047–1057.
[13] C. Chia-Cheng, Y. Liu, T.-H. Hsu, An analysis on investment performance of machine learning: an empirical examination on taiwan stock market, International Journal of Economics and Financial Issues 9 (2019) 1.
[14] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory 13 (1967) 21–27.
[15] C. Debrah, A.P.C. Chan, A. Darko, Green finance gap in green buildings: A scoping review and future research needs, Building and Environment (2021) 108443.
[16] S.S. Dhaliwal, A.-A. Nahid, R. Abbas, Effective intrusion detection system using xgboost, Information 9 (2018) 149.
[17] D.R. Don, I.E. Iacob, Dcsvm: fast multi-class classification using support vector machines, International Journal of Machine Learning and Cybernetics 11 (2020) 433–447.
[18] H. Doukas, On the appraisal of triple-a energy efficiency investments, Energy Sources, Part B: Economics, Planning, and Policy 13 (2018) 320–327.
[19] H. Doukas, A. Nikas, M. González-Eguino, I. Arto, A. Anger-Kraavi, From integrated to integrative: Delivering on the paris agreement, Sustainability 10 (2018) 2299.
[20] H. Doukas, P. Xidonas, N. Mastromichalakis, How successful are energy efficiency investments? a comparative analysis for classification & performance prediction, Computational Economics (2021) 1–20.
[21] E. Foda, A. El-Hamalawi, J. Le Dréau, Computational analysis of energy and cost efficient retrofitting measures for the french house, Building and Environment 175 (2020) 106792.
[22] A. Forouli, N. Gkonis, A. Nikas, E. Siskos, H. Doukas, C. Tourkolias, Energy efficiency promotion in greece in light of risk: Evaluating policies as portfolio assets, Energy 170 (2019) 818–831.
[23] J.H. Friedman, Stochastic gradient boosting, Computational Statistics & Data Analysis 38 (2002) 367–378.
[24] H. Ghoddusi, G.G. Creamer, N. Rafizadeh, Machine learning in energy economics and finance: A review, Energy Economics 81 (2019) 709–727.
[25] S. Hayes, S. Nadel, C. Granda, K. Hottel, What have we learned from energy efficiency financing programs, in: American Council for an Energy-Efficient Economy, 2011.
[26] S. Herrero-Lopez, J.R. Williams, A. Sanchez, Parallel multiclass classification using svms on gpus, in: Proceedings of the 3rd Workshop on General-purpose Computation on Graphics Processing Units, 2010, pp. 2–11.
[27] IEA, Implementing Energy Efficiency Policies: are IEA Member Countries on Track? 2009, https://www.oecd.org/publications/implementing-energy-efficiency-policies-are-iea-member-countries-on-track-9789264075696-en.htm..
[28] IEA, Net Zero by 2050: A Roadmap for the Global Energy Sector, 2021, https://iea.blob.core.windows.net/assets/4719e321-6d3d-41a2-bd6b-461ad2f850a8/NetZeroby2050-ARoadmapfortheGlobalEnergySector.pdf..
[29] S. Intrachooto, V. Horayangkura, Energy efficient innovation: Overcoming financial barriers, Building and Environment 42 (2007) 599–604.
[30] IRENA, Renewable Energy Statistics, 2015, https://www.irena.org/publications/2015/Jun/Renewable-Energy-Target-Setting..
[31] A.H. Jahromi, M. Taheri, A non-parametric mixture of gaussian naive bayes classifiers based on local independent features, in: 2017 Artificial Intelligence and Signal Processing Conference (AISP), IEEE, 2017, pp. 209–212.

[32] P.A. Jensen, E. Maslesa, Value based building renovation–a tool for decision-making and evaluation, Building and Environment 92 (2015) 1–9.
[33] P.A. Jensen, E. Maslesa, J.B. Berg, C. Thuesen, 10 questions concerning sustainable building renovation, Building and Environment 143 (2018) 130–137.
[34] B.J. Kalkbrenner, J. Roosen, Citizens' willingness to participate in local renewable energy projects: The role of community and trust in germany, Energy Research & Social Science 13 (2016) 60–70.
[35] B. Krawczyk, Learning from imbalanced data: open challenges and future directions, Progress in Artificial Intelligence 5 (2016) 221–232.
[36] J.Y. Lee, M.P. Styczynski, Ns-knn: a modified k-nearest neighbors approach for imputing metabolomics data, Metabolomics 14 (2018) 1–12.
[37] Q.Y. Li, J. Han, L. Lu, A random forest classification algorithm based personal thermal sensation model for personalized conditioning system in office buildings, The Computer Journal 64 (2021) 500–508.
[38] T. Loureiro, M. Gil, R. Desmaris, A. Andaloro, C. Karakosta, S. Plesser, De-risking energy efficiency investments through innovation. In Multidisciplinary Digital Publishing Institute Proceedings, vol. 65, 2020, p. 3..
[39] S. Makridakis, E. Spiliotis, V. Assimakopoulos, Statistical and machine learning forecasting methods: Concerns and ways forward, PLOS One 13 (2018) 1–26.
[40] V. Marinakis, Big data for energy management and energy-efficient buildings, Energies 13 (2020) 1555.
[41] A. Mathur, G.M. Foody, Multiclass and binary svm classification: Implications for training and classification users, IEEE Geoscience and Remote Sensing Letters 5 (2008) 241–245.
[42] D. Mena, E. Montañés, J.R. Quevedo, J.J. del Coz, An overview of inference methods in probabilistic classifier chains for multilabel classification, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 6 (2016) 215–230.
[43] F.D. Mexis, A. Papapostolou, C. Karakosta, E. Sarmas, D. Koutsandreas, H. Doukas, Leveraging energy efficiency investments: An innovative web-based benchmarking tool, Advances in Science, Technology and Engineering Systems Journal 6 (2021) 237–248.
[44] J. Morrissey, B. Meyrick, D. Sivaraman, R. Horne, M. Berry, Cost-benefit assessment of energy efficiency investments: Accounting for future resources, savings and risks in the australian residential sector, Energy Policy 54 (2013) 148–159.
[45] M. Norouzi, D.J. Fleet, R.R. Salakhutdinov, Hamming distance metric learning, in: Advances in Neural Information Processing Systems, 2012, pp. 1061–1069..
[46] J. Ortiz, A.F. i Casas, J. Salom, N.G. Soriano, P.F. i Casas, Cost-effective analysis for selecting energy efficiency measures for refurbishment of residential buildings in catalonia, Energy and Buildings 128 (2016) 442–457.
[47] J.P. Painuly, Barriers to renewable energy penetration; a framework for analysis, Renewable Energy 24 (2001) 73–89.
[48] J.P. Painuly, H. Park, M.-K. Lee, J. Noh, Promoting energy efficiency financing and escos in developing countries: mechanisms and barriers, Journal of Cleaner Production 11 (2003) 659–665.
[49] A. Papapostolou, F.D. Mexis, E. Sarmas, C. Karakosta, J. Psarras, Web-based application for screening energy efficiency investments: A mcda approach, in: 2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA, IEEE, 2020, pp. 1–7.
[50] B. Pavlyshenko, Using stacking approaches for machine learning models, in: 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), IEEE, 2018, pp. 255–258.
[51] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.
[52] B. Prasetiyo, M. Muslim, et al., Analysis of building energy efficiency dataset using naive bayes classification classifier, Journal of Physics: Conference Series (2019) 032016, IOP Publishing volume 1321.
[53] A. Sarkar, J. Singh, Financing energy efficiency in developing countries'lessons learned and remaining challenges, Energy Policy 38 (2010) 5560–5571.
[54] E. Sarmas, P. Xidonas, H. Doukas, Multicriteria Portfolio Construction with Python, Springer, 2020.
[55] M. Sarrica, S. Brondi, P. Cottone, B.M. Mazzara, One, no one, one hundred thousand energy transitions in europe: The quest for a cultural approach, Energy Research & Social Science 13 (2016) 1–14.
[56] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Statistics and Computing 14 (2004) 199–222.
[57] E. Stocker, M. Tschurtschenthaler, L. Schrott, Cost-optimal renovation and energy performance: Evidence from existing school buildings in the alps, Energy and Buildings 100 (2015) 20–26.
[58] I. Syarif, E. Zaluska, A. Prugel-Bennett, G. Wills, Application of bagging, boosting and stacking to intrusion detection, in: International Workshop on Machine Learning and Data Mining in Pattern Recognition, Springer, 2012, pp. 593–602.
[59] P.G. Taylor, O.L. d'Ortigue, M. Francoeur, N. Trudeau, Final energy use in iea countries: The role of energy efficiency, Energy Policy 38 (2010) 6463–6474.
[60] R.P. Taylor, C. Govindarajalu, J. Levin, A.S. Meyer, W.A. Ward, Financing energy efficiency: lessons from Brazil, China, India, and beyond, World Bank (2008), Publications.
[61] J. Thøgersen, A. Grønhøj, Electricity saving in households'a social cognitive approach, Energy Policy 38 (2010) 7732–7743.
[62] G. Toussaint, Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and data mining, International Journal of Computational Geometry & Applications 15 (2005) 101–150.
[63] P. Tuominen, F. Reda, W. Dawoud, B. Elboshy, G. Elshafei, A. Negm, Economic appraisal of energy efficiency in buildings using cost-effectiveness assessment, Procedia Economics and Finance 21 (2015) 422–430.
[64] A.B. de Vasconcelos, M.D. Pinheiro, A. Manso, A. Cabaço, Epbd cost-optimal methodology: Application to the thermal rehabilitation of the building envelope of a portuguese residential reference building, Energy and Buildings 111 (2016) 12–25.
[65] D.H. Wolpert, Stacked generalization, Neural Networks 5 (1992) 241–259.
[66] M. Zekić-Sušac, S. Mitrović, A. Has, Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities, International Journal of Information Management 58 (2021) 102074.
[67] D. Zhang, Z. Zhang, S. Managi, A bibliometric analysis on green finance: Current status, development, and future directions, Finance Research Letters 29 (2019) 425–430.
[68] H. Zhang, The optimality of naive bayes, AA 1 (2004) 3.