Evidence for residential building retrofitting practices using explainable AI and socio-demographic data

Simon Wenninger, Philip Karnebogen, Sven Lehmann, Tristan Menzinger, Michelle Reckstadt

Angaben zur Veröffentlichung / Publication details:

Wenninger, Simon, Philip Karnebogen, Sven Lehmann, Tristan Menzinger, and Michelle Reckstadt. 2022. "Evidence for residential building retrofitting practices using explainable AI and socio-demographic data." *Energy Reports* 8: 13514–28. https://doi.org/10.1016/j.egyr.2022.10.060.

Energy Reports 8 (2022) 13514–13528

Contents lists available at ScienceDirect

Energy Reports

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

Research paper

Evidence for residential building retrofitting practices using explainable AI and socio-demographic data

Simon Wenninger[a,b,*], Philip Karnebogen[a,b], Sven Lehmann[c], Tristan Menzinger[c], Michelle Reckstadt[d]

[a] Branch Business & Information Systems Engineering of the Fraunhofer FIT, Alter Postweg 101, 86159 Augsburg, Germany
[b] FIM Research Center, University of Applied Sciences Augsburg, Alter Postweg 101, 86159 Augsburg, Germany
[c] FIM Research Center, University of Augsburg, Universitätsstr. 12, 86159 Augsburg, Germany
[d] University of Augsburg, Universitätsstr. 12, 86159 Augsburg, Germany

ARTICLE INFO

Article history:
Received 13 June 2022
Received in revised form 12 September 2022
Accepted 5 October 2022
Available online xxxx

Keywords:
Energy performance certificates
Retrofitting
Energy efficiency policy
Explainable AI
Data analytics
Policy implications

ABSTRACT

Extensive retrofits and effective policy measures are needed to meet the ambitious climate goals, particularly in the UK, with the EU's oldest residential building stock. Researchers must investigate the factors influencing retrofits to enable effective and targeted policy measures. To date, however, there is a lack of holistically large-scale quantitative studies accounting for such factors. At the same time, great potential is seen in data-driven solutions and the use of explainable artificial intelligence (XAI). We address this research gap by combining supervised machine learning with XAI employing a three-stage approach: First, we consolidate datasets of Energy Performance Certificates from England and Wales from which we extract conducted retrofits, house prices, and socio-demographic information. Second, we apply an eXtreme Gradient Boosting (XGBoost) model that predicts whether a building has been retrofitted or not. Lastly, we use SHapley Additive exPlanations values (SHAP) as an XAI technique to identify the key factors and relationships that influence the implementation of retrofits. We succeed in substantiating results previously obtained in qualitative or small-scale studies and also find that retrofit-related policies already implemented in regional cases, such as the "Better Homes for Yorkshire" initiative, can successfully achieve large-scale success through replication in other regions. Further, our results suggest the implementation of income-based CO2 taxes as a reasonable and easy-to-implement policy measure.

1. Introduction

The COVID-19 pandemic has stalled the growth of global CO_2 emissions (Forster et al., 2020). Rising international interest in climate change and the ambitious climate goals defined under the Paris Climate Agreement require policy decisions and actions to reduce greenhouse gas emissions (Krausmann et al., 2020). The global buildings sector is responsible for nearly 38% of global greenhouse gas emissions and 39% of global energy consumption and thus holds great potential to progress towards climate goals (Somu et al., 2020).

Therefore, extensive retrofitting of energy-inefficient buildings is necessary to achieve the climate goals (Fylan et al., 2016). Policymakers need to increase the effectiveness and attractiveness of support measures and programs, e.g., subsidies to maximize greenhouse gas savings per monetary value invested in the face of limited financial resources to promote and incentivize retrofitting (Csutora and Zsóka, 2011). Thus, researchers must analyze the circumstances of and existing barriers against retrofitting to design effective support measures for retrofits (Fylan et al., 2016). Existent research on the circumstances and obstacles to implementing energetic retrofits is diverse (Ben and Steemers, 2018; Bertoldi and Mosconi, 2020; Ahlrichs et al., 2022). For instance, Tziogas et al. (2021) identify regional differences in the number and costs of retrofits in Greek. In contrast, Magnani et al. (2020) find that tax incentives for retrofits are ineffective and local intermediaries strongly influence the local retrofit level in Italy using a mixed-methods approach. Further, Gómez-Navarro et al. (2021) analyze survey-based energy poverty in Valencia, Spain, while Ahlrichs et al. (2022) use a data-driven approach to identify the impact of socioeconomic factors on local building energy efficiency in England, Scotland,

* Corresponding author at: Branch Business & Information Systems Engineering of the Fraunhofer FIT, Alter Postweg 101, 86159 Augsburg, Germany.
E-mail address: simon.wenninger@fit.fraunhofer.de (S. Wenninger).

https://doi.org/10.1016/j.egyr.2022.10.060
2352-4847/

and Wales. Most of this research focuses on the energy efficiency gap, which discusses reasons against implementing retrofit investments with seemingly clear economic and environmental benefits (Ahlrichs et al., 2020).

However, research to date is often limited to qualitative studies or only investigates influencing factors on energy efficiency (Ahlrichs et al., 2022) instead of factors influencing retrofitting. Thus, research does not fully exploit the opportunities created by advancing digitization and data availability. In this vein, recent papers highlight two different needs for future research in the field of building energy consumption using data-driven methods. On the one hand side, (Pasichnyi et al., 2019) proposed using the (openly accessible) databases of building-related Energy Performance Certificates (EPC) for data-enabled urban energy policy instruments. They conclude that EPC data might have a broader spectrum of applications than initially intended and are suitable for designing energy-efficiency policy instruments. On the other hand, the literature suggests using explainable artificial intelligence (XAI) in the building sector to derive insights into the relations of different parameters and variables (Golizadeh Akhlaghi et al., 2021). In this context, prior research such as Athey (2017) also encouraged using artificial intelligence (AI) beyond plain predictions to derive data-driven policy implications. In this course, we present an XAI-based approach to decrease the energy efficiency gap and build upon existing research by formulating our research question as follows:

How can XAI approaches based on EPC, house price, and socioeconomic data contribute to deriving policy implications for retrofitting behavior in residential buildings?

To answer our research question, we use the case of the UK's residential building stock, which represents the oldest building stock in western Europe and accounts for more than a quarter of the UK's total energy consumption (Piddington et al., 2020; Dowson et al., 2012; Filippini et al., 2014; Fylan et al., 2016). With more than 82% of buildings constructed before 1991, the building stock reflects loose building regulations and poor insulation, leading to high energy consumption and greenhouse gas emissions (Dowson et al., 2012). Although the UK government recognized the high energy consumption early in 2012 and tried to make retrofits more attractive through the Green Deal policy initiative, the UK missed its greenhouse gas reduction targets and all critical indicators for energetic retrofits in 2018 (Brown, 2018). Additionally, the initiative stopped in 2015 due to low demand and skepticism about possible savings (Dowson et al., 2012; Comerford et al., 2018).

Given the high retrofitting potential and the accessibility of publicly available data, the UK residential building stock qualifies perfectly for our work. We, therefore, use multiple data sources of EPC data from England and Wales, additional house price data, and socio-demographic data. We extract whether a building has been retrofitted and which measures have been carried out from the EPC data. We then apply machine learning to the datasets using an XGBoost (e**X**treme **G**radient **Boost**ing) model to derive the probability that a building is retrofitted. We subsequently use **SH**apley **A**dditive ex**P**lanations values (SHAP) as an XAI technique to identify the most critical factors and relationships that influence the implementation of retrofits. We finally derive policy implications for the effective design of support instruments and programs for retrofits based on the insights of building characteristics, house prices, and socio-demographic data.

With our work, we contribute to existing literature and design effective support instruments and programs toward the climate goals set in several ways. First, to the best of our knowledge, we are the first to introduce and apply the combination of supervised machine learning for classifying building retrofits and using XAI techniques to derive essential insights and relations on how these classifications came to pass. Second, we can confirm existing qualitative research with our quantitative study and reveal new relationships. Third, we present a method to extract building retrofits of the UK EPC data. Fourth, we derive policy implications to effectively design support instruments and programs for retrofits in the residential building stock.

2. Theoretical background and related work

2.1. Explainable artificial intelligence in energy research

Although AI and data-driven methods have gained momentum in energy research (Wenninger et al., 2022), the application of XAI techniques is not yet established. XAI refers to techniques that provide details or rationales that make the workings of AI straightforward to understand for a given audience (Barredo Arrieta et al., 2020). In addition to explainability, which is seen as a significant obstacle to the widespread use of AI, XAI also promises to improve the trustworthiness of AI algorithms, causality, transferability, and confidence (Burkart and Huber, 2021). Moreover, XAI can derive recommendations for action or control from AI models, improve them, and extract new information and insights from data (Adadi and Berrada, 2018).

Despite XAI's numerous advantages and research opportunities, only a few applications and investigations using XAI are present in energy research to date. Applications of XAI often focus on engineering-related topics and disciplines whose findings are not undoubtedly transferable and relevant for policymaking. Recent articles that have already applied XAI techniques in energy research are listed in Table 1. Nevertheless, some authors spotlight the great potential of XAI in energy research (Miller, 2019). For example, identifying different occupant behaviors when using a building could unlock the full potential of data-driven models in the building energy sector (Miller, 2019). In addition, XAI could result in higher energy savings from retrofits by increasing the confidence and understanding of data-driven models. Here, increased confidence in energy savings would significantly contribute to energy retrofitting, as uncertainty is considered an essential criterion in investment decisions (Ahlrichs et al., 2020; Rockstuhl et al., 2021). We directly address this research gap with our study, focusing on applying XAI techniques to provide further insights for energy policymaking than existing non-explainable approaches.

2.2. Research on energy performance certificates

Since 2007, EPCs have been mandatory for almost all residential buildings in the UK, following a directive by the European Parliament and the Council in 2002 establishing the need for EPCs (Watts et al., 2011). EPCs provide owners, occupants, and property developers with information on the energy performance of individual buildings and other information, such as the building's position in an energy performance ranking, which allows the comparison of different structures (Poel et al., 2007; Zhang et al., 2012). Following a Standard Assessment Procedure (SAP), an energy assessor collects information on building characteristics, rates their performance, and estimates the energy costs and consumption (Watts et al., 2011). Based on this information, the energy assessor calculates an energy efficiency score, ranging from 0 to 100 points. A score of 100 points thereby corresponds to zero energy expenses (Zhang et al., 2012). This score is then assigned an energy efficiency label, which rates the building on a scale from A to G, whereby the label A corresponds to the most efficient rating (Fuerst et al., 2015). The Energy Efficiency Label, resulting from the procedure described above to obtain the

Table 1
Recent articles using XAI techniques in energy research.

Source	Research topic
Miller (2019)	Interpretability of machine learning-based classification for smart meter data from non-residential buildings
Arjunan et al. (2020)	Identification of factors that most influence a building's energy use relative to other buildings, using SHAP
Papadopoulos and Kontokosta (2019)	Interpreting the importance of individual features and understanding drivers of energy use intensity of residential properties in New York City, using SHAP
Feng et al. (2021)	Interpretation of space cooling energy predictions and feature interactions for residential buildings with SHAP
Ke et al. (2020-2020)	Analyzing feature importance for building power consumption in the context of data-driven building control strategies using SHAP
Seyrfar et al. (2021)	Interpretation of variable importance on building energy consumption of multifamily residential buildings in Chicago using SHAP
Coma-Puig and Carmona (2019)	Detection of essential features for non-technical losses and fraud in electricity consumption using SHAP
Setyantho and Chang (2020)	Identification of factors influencing patterns of energy consumption in commercial and residential buildings using SHAP
Wenninger and Wiethe (2021)	Identification of essential variables influencing residential heating energy consumption using permutation feature importance

score, is not influenced by the occupants' usage behavior but is based purely on the building information (Comerford et al., 2018). In addition to energy efficiency, the EPC lists recommendations for energetic retrofit measures (Department for Communities and Local Government, 2017). Further, the UK's EPCs also reflect the aforementioned poor building conditions. The average energy efficiency in the UK has an energy efficient label of D for the year 2017 and corresponds to an SAP rating between 55 and 68, which is significantly below the optimum of 100 (Piddington et al., 2020, p. 21). Although the EPC stock is only a sample of all buildings, it covers more than 50% of them (Office for National Statistics, 2020a, p. 10).[1]

Both literature and practice discuss different aspects of EPCs manifold (Li et al., 2019) due to their growing number and richness of information (Jenkins et al., 2017; Pasichnyi et al., 2019). In addition to investigations of the EPC data, literature examines the extent to which EPCs influence the real estate market and the influence of EPCs on retrofit and purchase decisions (Pasichnyi et al., 2019). Research on EPC data often focuses on the energy performance gap, which describes the phenomenon that the actual metered energy consumption differs significantly from the calculated energy demand illustrated in EPCs (Burman et al., 2014; Herrando et al., 2016; Menezes et al., 2012). According to Wilde (2014), the energy performance gap may result, for example, from the neglect of external factors, which are consequently not included in EPCs. These external factors are either not easy to determine, e.g., construction defects of the building, or factors that are not included in the creation of the EPCs due to their high national comparability, e.g., weather conditions or the occupants' behavior concerning energy consumption. Prior work such as Jenkins et al. (2017) also investigates the quality of the EPC data regarding the inspection necessary for preparing EPCs and found that the assessment of various building parts varies greatly. Further, Hardy and Glew (2019) offer an overview of general data quality problems and inconsistencies considered in this research paper to ensure high data quality in our study (cf. Section 4). However, while many studies deal with the errors in EPC data and show which weaknesses they have, few studies deal with the value of EPCs and the information they contain (Pasichnyi et al., 2019).

According to Fuerst et al. (2015), who studied the impact of energy efficiency ratings on real estate values in the private sector, EPCs significantly impact real estate rents and values. They found a significant positive correlation between the energy efficiency rating and price per m^2, reflected in a price premium for energy-efficient houses. These results were confirmed and extended by Fuerst et al. (2020) and Khazal and Sønstebø (2020), who showed that energy efficiency influences the purchase price of real estate and the respective rents. For example, buildings with a B rating exhibit a rent premium of 4% compared to those with a D rating. These results primarily stem from the increase in property value achieved by investing in a zero-emission, greener building and are priced in by higher rents.

Even though EPC data do not intend initially to enable the analysis of retrofitting practices, several papers address this topic (Pasichnyi et al., 2019). Gupta and Gregg (2018) explore the application of an approach developed in co-operation with the UK government (Local Energy Mapping for Urban Retrofit) to plan and implement better and more targeted local energy efficiency programs. However, their process does not take into account other causes that lead to not retrofitting. Further, utilizing a microeconomic model and simple regression, Adan and Fuerst (2015) examined the fundamental drivers of energy retrofit measures. However, their analysis was limited to two retrofit measures: the cavity wall and loft insulation. One of their findings was the positive correlation between the vacancy rate and retrofitting. In contrast, a real estate market with worse conditions meaning a high vacancy rate, shows a higher number of retrofits. Adan and Fuerst (2015) also provided two possible explanations for this relationship: on the one hand, (major) retrofits need a certain vacancy period to be carried out without disruption. On the other hand, a real estate market with high vacancies provides a higher need to increase the attractiveness of a property. In addition, they found that income influences retrofitting: The lower the income, the greater the share of ancillary costs, and the greater the incentive to increase energy efficiency. However, Adan and Fuerst (2015) focus on two specific measures and do not derive any additional policy implications from increasing the number of retrofits. Beyond EPC data, researchers also studied socio-demographic factors and barriers influencing retrofitting through survey data and literature reviews (Achtnicht and Madlener, 2014; Kastner and Stern, 2015). Homeowners' education level, age, income, and employment correlate with the energy efficiency of a building

[1] Focusing on England and Wales, the countries considered in this paper. Both countries account for more than 89% of the total building stock in UK Piddington et al. (2020).

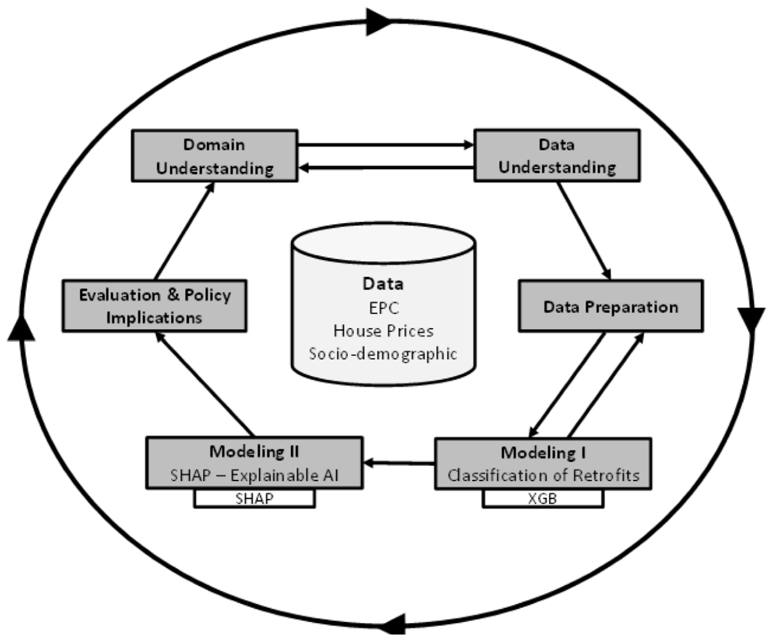

Fig. 1. Research design derived from the CRISP DM (Shearer, 2000).

and the frequency of retrofits (Achtnicht and Madlener, 2014; Jakob, 2007; Wilson et al., 2015). Researchers such as Druckman and Jackson (2008) and Ma and Cheng (2016) also found differences in regional energy consumption resulting from various socio-demographic factors. However, while there is some current work regarding EPCs and retrofitting, it is not analyzed in the desired depth with the capabilities of data-driven methods. We address this research vacuum by conducting an empirical study of the influence of socioeconomic factors on retrofitting practices using extensive real-world datasets.

3. Methodology

3.1. Research design

We implemented a suitable method and research design to address our research question and show how XAI may lead to more sophisticated data-driven policy implications. Therefore, we derived a six-step process illustrated in Fig. 1, based on the established Cross Industry Standard Process for Data Mining (CRISP-DM) and the guidelines for conducting extensive data analysis by Müller et al. (2016). We adapted the CRISP-DM as we aim to guide policymakers by spotlighting how XAI techniques contribute to unraveling the interdependencies between socio-demographic factors, EPC data, house prices, and the installation of retrofitting measures. In the following, we first introduce the research design before presenting details on the XGBoost model for classification and SHAP values as the XAI technique to ensure the reproducibility and replicability of our work.

The first process step within our research design is "Domain Understanding", which extends the initial first stage of "Business Understanding" by our primary objective of analyzing EPCs concerning energy retrofits in residential buildings. In addition, we modify the intention of the business understanding to collect domain-specific knowledge about energetic building retrofitting, which is necessary to derive practical policy implications. We motivate and set the context for the need of our study in the Introduction and present domain-specific knowledge in Section 2. We continue with "Data Understanding" as the second step, which we do not modify compared to the original CRISP-DM. However, we skip the initial data collection step, as we have publicly available datasets, and start with an initial data review and identification of data quality issues. Analogously, we have also not modified the third step, "Data Preparation" where we prepare the data by applying common steps of data cleaning, data transformation, preselecting variables, and checking errors in the EPC dataset. We proceed analogously to Hardy and Glew (2019) to construct a high-quality dataset for the EPC dataset. We provide details on the datasets, data preparation, and the method developed to extract whether a building has been retrofitted and what measures have been implemented from the EPC data in Section 4. To adapt the research design to our specific research question, we divide the modeling step into two sequential modeling steps: In the first modeling step, we fit the classification algorithm XGBoost to predict whether buildings have undergone retrofitting measures based on the available socio-demographic, house price, and EPC data (see Section 3.2 for details). In the second modeling step, we apply SHAP to investigate the reasons and relations behind the classifications (see Section 3.3 for more information). The second modeling step thus forms the basis and enables the necessary understanding for deriving data-based policy implications as the final step of our research design. All steps were iterated several times and continuously evaluated to ensure an optimal understanding of the data and the AI approaches used.

3.2. Classification of retrofitted buildings with XGBoost

The classification algorithm applied in this work to predict retrofits is XGBoost, which has recently found increasing application and popularity. According to the data science community and competition platform Kaggle, XGBoost was part of the winning solution 17 out of 29 times in 2015 (Chen and Guestrin, 2016) and showed convincing results in EPC research (Wenninger and Wiethe, 2021). In our work, XGBoost is advantageous for several reasons: XGBoost is known for its strong performance using parallel processing and out-of-core computing. Consequently, XGBoost can easily handle models that contain large amounts of data, as is the case in our study – after data processing, there are still 11,519,036 observations and 72 features left. Moreover, for classification problems like ours, the algorithm achieves excellent and accurate predictions with few false orders (Dhaliwal

Table 2
General structure of a confusion matrix (Ciaburro and Venkateswaran, 2017).

True values	Predicted values	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

et al., 2018). XGBoost is an optimized decision-tree-based machine learning algorithm from the field of supervised machine learning (Chen and Guestrin, 2016). The core idea of the algorithm is the ensemble technique (gradient) boosting. Boosting refers to the sequential combination of different simple models (= single decision trees), called weak learners, aiming to create a strong learner (Hastie et al., 2009). In this process, new models that use the information of the prior models are added iteratively and then try to correct the preceding models' errors (Mitchell and Frank, 2017; Chen and Guestrin, 2016). XGBoost achieves this by calculating an objective function (see Eq. (1)) which measures the performance of the model by considering the differences between actual and predicted values as well as the complexity of the model:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t\left(x_i\right)\right) + \Omega\left(f_t\right) \tag{1}$$

where $\mathcal{L}^{(t)}$ is the objective function in iteration t, l is the loss function,[2] y_i is the actual value for observation i the model tries to predict, $\hat{y}_i^{(t-1)} + f_t(x_i)$ is the predicted value for observation i, which is the prediction of the previous iterations $t-1$ adjusted by the current one in iteration t, and $\Omega(f_t)$ is the penalty function which penalizes the underlying model of iteration t for its complexity. This function is then optimized using a gradient descent algorithm that changes the weighting of the observations. Correctly classified observations are weighted lower, and incorrectly classified observations are weighted higher in the next iteration (Chen and Guestrin, 2016; Mitchell and Frank, 2017).

Performance measures are necessary (Kaymakci et al., 2021) to evaluate and compare the prediction performance of (classification) models. Confusion matrices for binary classification problems like in this work are particularly suitable. These generally indicate whether the predicted value corresponds to the actual given output value or not (Ciaburro and Venkateswaran, 2017). We classify the predictions into one of four categories according to Table 2:

Using the confusion matrix defined in Table 2, we calculate three performance measures to assess the performance of the model in this work (Ciaburro and Venkateswaran, 2017):

$$\boldsymbol{Sensitivity} = \frac{TP}{TP + FN} \tag{2}$$

$$\boldsymbol{Specificity} = \frac{TN}{TN + FP} \tag{3}$$

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

In Table 3, we further summarize the technical details of the first modeling step – the classification of retrofitted buildings – with a brief problem statement, information about the data and several data preparation procedures, and details on the XGBoost. Table 3 thereby represents an adapted version of the "Supervised Machine Learning report card" initially introduced by Kühl et al. (2020) to document and structure machine learning projects and ensure the reproducibility of our work.

[2] Any arbitrary differentiable loss function can be applied for this purpose.

3.3. Identifying critical factors and relationships that influence the implementation of retrofits with SHAP

Since we do not search for explanations and relationships in the data itself but instead do so based on our model's predictions, understanding the model's prediction is critical for our research. In our study, we use the established XAI technique SHAP (Lundberg and Lee, 2017) to identify important factors and relationships that influence the implementation of retrofits predicted by the XGBoost model.

SHAP combines Shapley values and local interpretable model-agnostic explanations (LIME). SHAP builds upon Shapley values and LIME in that the Shapley values are additive (e.g., the sum of Shapley values for a single data point results in the prediction's deviation from the average prediction). A single observation's SHAP value for a single feature does not indicate the model's prediction value directly, but instead, the value with which the feature moves the prediction from the model's average for that specific observation's set of values (Lundberg and Lee, 2017).

As Shapley values are calculated via marginalization, non-independent features affect the resulting values as the order of consideration matters. SHAP mitigates this by averaging the feature's SHAP values for all possible orderings. Our EPC data possesses several features we suspect to be correlated and various algorithmically derived features (e.g., the energy rating itself, ecological impact scores) that are, therefore, not independent. SHAP also introduces optimizations to reduce computational complexity, as Shapley values are computationally challenging. Specifically, in this work, we use TreeSHAP (Lundberg et al., 2018). TreeSHAP uses the conditional expectation instead of the model's marginal expectation for the Shapely values. For XGBoost, an ensemble model, the SHAP values are calculated from the weighted average of the Shapely values of the individual trees. We refer to the literature for further details on the method used (Lundberg and Lee, 2017; Lundberg et al., 2018).

4. Data and extraction of retrofitting measures

4.1. Data

For our study, we dispose of three real-world and publicly available datasets containing information on EPCs, house prices, and socio-demographic data to address our guiding research question. The EPC data stems from the Ministry of Housing, Communities & Local Government, containing 19,732,107 certificates issued by all 339 Local Authorities in England and Wales between December 2007 (the Introduction of EPCs) and September 2020 (Ministry of Housing, Communities & Local Government, 2020). As mentioned above, the EPC data contains a rating of energy efficiency, which allows potential buyers or tenants to quickly compare the energy consumption of different houses and further information. We use 28 of the 91 columns in the data set for our study, which can be categorized according to Pasichnyi et al. (2019), as illustrated in Table 4. The selected subset of data is necessary to extract performed retrofits following the procedure in Section 4.3, analyze the status quo of retrofits in the UK in Section 5.1, and derive insights on retrofitting behavior in combination with house prices socio-demographic data by applying SHAP in Section 5.2.

The house price data is obtained from the price Paid data set provided by HM Land Registry and includes details of 25,667,093 registered real estate sales in England and Wales since 1995 (Government Digital Service, 2021). In addition to the purchase price, the data also contains information regarding the address, local authority, and purchase date (HM Land Registry, 2016). Because we could only identify a house price for 10% of the

Table 3
Supervised Machine Learning report card for rigor documentation of the retrofit classification problem adapted from Kühl et al. (2020).

Problem statement	Predicting whether buildings have undergone retrofits or not for the England and Wales building stock with EPC data, price data, and socio–demographic data
Data	
Data gathering	The data originates from publicly available sources: – EPC data from England and Wales (Ministry of Housing, Communities & Local Government, 2020) – House price data (Government Digital Service, 2021) – Sociodemographic data (Welsh Government, 2020; Office for National Statistics, 2021, 2020c,b) For details, we refer to Section 4
Sampling	Creation of a "balanced dataset", as retrofitted buildings are highly underrepresented in the EPC dataset
Data quality	Generally high, partly missing, or the incorrect values
Data pre-processing methods	Application of several pre-processing methods/steps. For details, we refer to Section 4
Feature engineering and vectorizing	One-hot encoding and derivation of variables. For details, we refer to Section 4
Algorithm & performance estimation	
Algorithm	Xgboost python package (Chen and Guestrin, 2016)
Parameters	max_depth: 18; objective: binary:logistic; eval_metric: mae, error, logloss; gamma: 3; lambda: 3; alpha: 3; min_child_weight: 6; eta: 0.02; subsample: 0.8; colsample_bytree: 0.9; colsample_bylevel: 0.9; colsample_bynode: 0.9; iterations: 200 For details of the specific parameters, we refer to the official python package (Chen and Guestrin, 2016)
Data split	70% training and 30% test data
Performance metric	Mean absolute error, error, log-loss (for their definition, we refer to the python package)
Performance evaluation measure	Sensitivity, specificity, accuracy

Table 4
Clustering of relevant columns in the EPC data set.

Category	Related columns
Building Reference	Building ID, Local Authority, Tenure, Construction Age
Building Geometry	Building Form and Type, Total Floor Area
Certificate Methodology	Certificate ID, Date of Inspection and Lodgement, Transaction Type
Energy Performance	Current Energy and Environment Rating, Energy Consumption, CO_2 emissions per m^2 floor area, Costs regarding Lighting, Heating, and Hot Water
Energy System Installations	Fuel Type, Type and Proportion of Glazing, Efficiency of central Heating System, Hot Water Systems, Wall, Floor, Roof, and Lighting
Recommendations	Retrofit suggestions

houses in the EPC data, we calculated an annual house price at the local authority level, which we appended to each EPC that remained after the data preparation process (see Section 4.2). As already listed in Section 2, Adan and Fuerst (2015) found a (significant) relation between house prices and investment in a specific retrofit measure (e.g., building insulation). We assume that this observed relationship between a particular measure and house prices can be generalized in this work. Consequently, we consider house prices to impact the general circumstance of whether someone retrofits or not, regardless of the implemented measure.

We further enrich the EPC data with socio-demographic data, contributing to whether to invest in retrofitting or not (Adan and Fuerst, 2015). We dispose of annual distributions of age groups (Welsh Government, 2020), the employment rate (Office for National Statistics, 2021), gross disposable household income (GDHI) (Office for National Statistics, 2020c), and education (Office for National Statistics, 2020b) on the regional level, as we do not dispose of more detailed data on local authority level.

4.2. EPC data preparation

To ensure high data quality, we apply two sequential data preparation steps (cf. Fig. 2), especially for EPC data, as these are not yet available in the desired data format and usually require more elaborate preparation than house price and socio-demographic data, which, on the other hand, require hardly any adjustments. First, we remove duplicates and EPCs that refer to the "Building Type" "Flat" or "Maisonette", since the focus of this work is to analyze retrofits of houses in the UK. Furthermore, we remove EPCs with no assignable local authority since enriching EPC data with socio-demographic data provides value only through a regional assignment. In addition, we remove EPCs created before 01.01.2008 or after 31.12.2019 to consider only full years. Second, we apply the guidelines introduced by Hardy and Glew (2019), especially those dealing with lodgement errors, to remove incorrect EPCs. Therefore, we only keep the last (corrected) EPC if several identical EPCs exist within a short time (usually within a week) for the same house. We further use the validation levels 0 and 1 proposed by Pasichnyi et al. (2019) to consistently identify missing values and check the plausibility of individual metrics (such as the number of m^2, which must not be negative) in addition to correcting spelling and typing errors in string variables. Despite the restriction to only consider EPCs for houses from 2008 to 2019, almost 60% of the data remain, corresponding to more than 11 million certificates. This number only slightly changed after cleaning the data set with the help of the error codes of Hardy and Glew (2019), resulting in 58% EPCs of the initial dataset.

4.3. Extraction of conducted retrofits

Since our research included predicting whether a building is retrofitted or not as a basis for further XAI approaches to derive policy implications for retrofitting behavior in residential buildings, we needed to extract historical information on whether a building has been retrofitted. As EPC data does not explicitly record this historical information, we had to apply an approximation approach. Therefore, we draw on a method

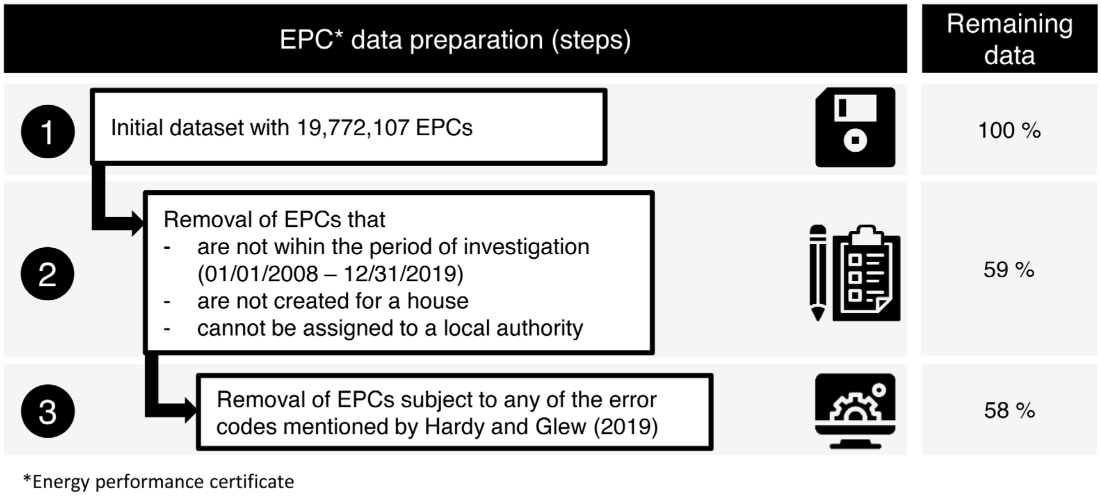

Fig. 2. EPC data preparation steps and the remaining share of EPC data.

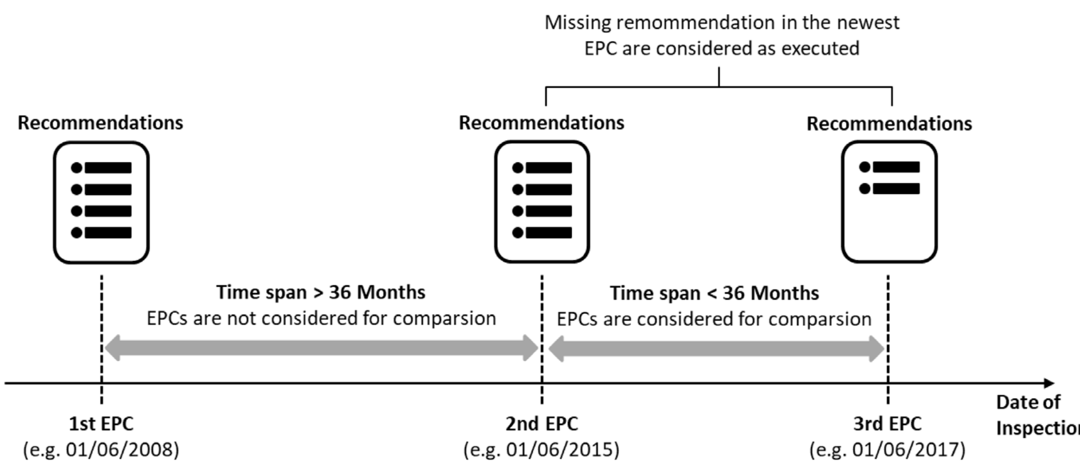

Fig. 3. Procedure of conducted retrofit extraction.

initially introduced by Prieler et al. (2017), who estimated the number of executed retrofits in Austria by comparing the changes in retrofit recommendations in EPCs of a specific building over time. Following this approach, we extracted the automatically generated retrofit recommendations (Gonzalez Caceres, 2018) of two EPCs of a house and compared them against each other. We assume these retrofit recommendations will only differ from the first to the last EPC if specific retrofit measurements have been conducted and are consequently not further recommended. Since retrofit measures typically have payback periods of more than ten years (Salata et al., 2017), retrofits tend to take place at longer intervals. Therefore, we define a span of three years in which two consecutive EPCs must take place. In this way, we avoid individual retrofit recommendations being added or removed over time due to technical progress and errors resulting, for example, from the software change for issuing EPCs. Fig. 3 illustrates the applied extraction of conducted retrofits.

This approach led to four different scenarios within the extraction process. First, if buildings had only one EPC, we classified the included retrofit recommendations as not executed. Second, if more than one EPC of a building exists, but the EPCs are so far apart in time that they do not exceed the required maximum span of three years, we classified these buildings as not retrofitted. Third, multiple EPCs for a building exist, of which at least two are within the defined period, yet the comparison reveals that the recommendations proposed do not differ. Accordingly, we classified such buildings as not retrofitted. Fourth, there are multiple EPCs of a building, and at least two are within the defined period. Additionally, comparing these EPCs shows that their proposed measures differ, and at least one measure of the previous EPC does not appear in the subsequent EPC. Such buildings are then classified as retrofitted, and the measures implemented can be explicitly stated. Using this approach, we extracted 1,708,409 cases of specific retrofitting measurements. Further, we identified 9,810,627 cases with no conducted retrofits.

5. Results and evaluation

5.1. Data insights and descriptive statistics

Before presenting the results using XGBoost and SHAP, we provide a brief overview of the status quo of the building stock in England and Wales by applying descriptive statistics on the remaining prepared data—in full 11,519,036 houses. The remaining EPC database covers more than half of all residential houses in England and Wales, enabling us to draw fairly general conclusions on the overall building stock (Piddington et al., 2020). The Energy Efficiency Rating in Fig. 4a reflects that the UK has an old building stock compared to the EU and poor building conditions (Dowson et al., 2012; Piddington et al., 2020). We see that the regional differences are not too pronounced, and most local authorities exhibit an average rating of C or D. The only significant deviation occurs in Wales, with an average rating of predominantly E.

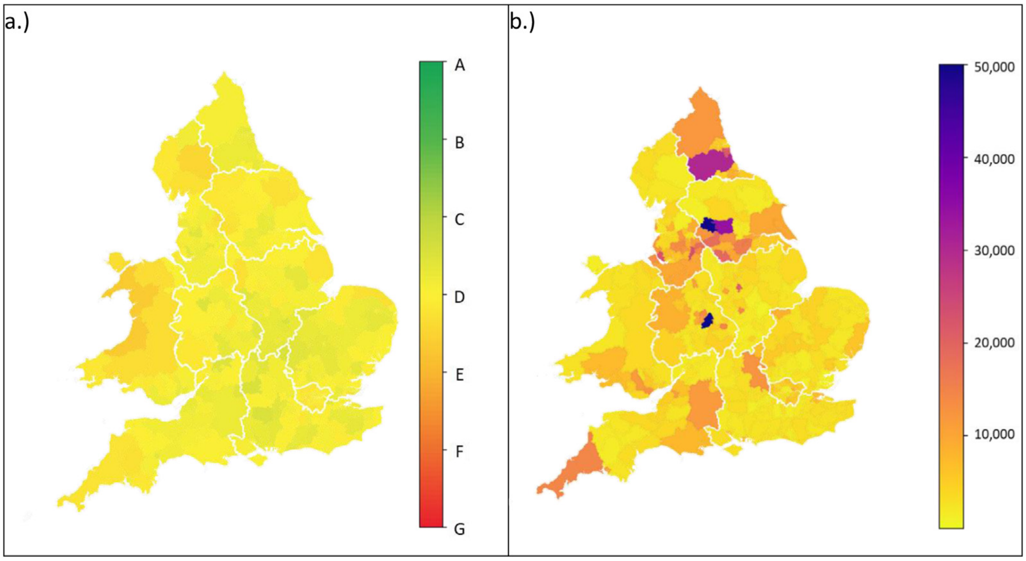

Fig. 4. Average energy efficiency rating for each local authority in (a.) and the number of conducted retrofits for each local authority in (b.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Consequently, there is a high need for retrofitting in England and Wales to meet climate goals, which are barely met as depictable in the number of conducted retrofits per local authority (cf. Fig. 4b). Most local authorities reported 15,000 or fewer retrofits between 2008 and 2019. Here we find an inverse trend: local authorities with averagely low energy efficiency ratings tend to have fewer retrofits, such as Wales, with numbers of retrofits often below 10,000, than those with above-average C ratings, such as for the northeast and some local authorities in the southwest. Beyond four local authorities – County Durham, Bradford, Leeds, and Birmingham – where more than twice as many retrofits have occurred, as indicated by the purple and dark blue colored areas in Fig. 4b. These high numbers of retrofits carried out might result from policymakers actively addressing climate change and the impact of building conditions. Bradford and Leeds, for example, are part of the "Better Homes for Yorkshire" initiative, which was launched in 2015 and aimed to increase retrofits through funding and education about the need to improve the energy efficiency of buildings (Adam, 2018). County Durham and Birmingham have also worked hard for years to address climate change in various areas of public life and regularly publish reports on the subject (Shaw, 2019; Birmingham City Council, 2010).

Apparent differences can be observed concerning the building age classes, as depicted in Fig. 5, by shifting the focus from the regional distribution of energy efficiency ratings to the retrofit measures carried out from 2008 to 2019. First, focusing on the most popular retrofit measure per building age class, we notice that floor insulation was most common for homes built before the first building standards were introduced in 1976 (Dowson et al., 2012). This pattern shifts towards replacing the condensing boiler and using low-energy lighting as the building age decreases. In addition, insulation retrofits (floor, attic, cavity, and interior or exterior wall) are more prominent in houses built before 2002. The picture changes for newer houses (built-in 2003 or later), with heating control upgrades, installation of low-energy lighting, or a PV system is most popular. Consequently, this building age class mainly conducts measures that enhance improvements and the existing energy efficiency standard. Furthermore, we see that installing a condensing boiler is one of the top two measures in the age groups up to 2002 and de facto no longer occurs in houses built in 2003 and later. This is because condensing boilers have been mandatory since 2005 (The Office of the Deputy Prime Minister, 2005). The popularity of all these measures may also be related to the Carbon Emissions Reduction Target Program conducted from 2008–2012 and the subsequent Energy Company Obligations. They aim to reduce carbon emissions in the private sector by requiring certain energy suppliers to promote energy efficiency measures in buildings (Fawcett et al., 2019). With the help of these programs, many of these measures have been carried out, e.g., more than 3.4 million cavity wall insulations (Office of Gas and Electricity Markets, 2013; Department for Business, Energy & Industrial Strategy, 2020c). In summary, we find that the proportions of measures implemented in the building age classes do not differ significantly until 2002, resulting from a less strong energy efficiency presence in the building regulations. However, the Introduction of Part L1 A and the associated new (energy efficiency-related) building standards have led to significant differences in the measures implemented.

Table 5
Confusion Matrix and performance measures for the test data.

True values	Predicted values	
	Positive (Retrofit)	Negative (No Retrofit)
Positive (Retrofit)	34%	14.7%
Negative (No Retrofit)	12%	39.3%
Performance measure	Performance	
Sensitivity	66.1%	
Specificity	76.2%	
Accuracy	73.3%	

5.2. Factors and relationships influencing the implementation of retrofits

Before presenting the results, factors, and relationships that influence retrofits' implementation as the basis for deriving policy implications, we first evaluate the XGBoost model's predictive performance. Following the procedure presented in Section 3.2, we obtain the out-of-sample goodness for the final XGBoost model illustrated with a confusion matrix and performance measures in Table 5.

Table 5 depicts that retrofits are incorrectly predicted in 12% and not detected in 14.7%, while our approach correctly predicts

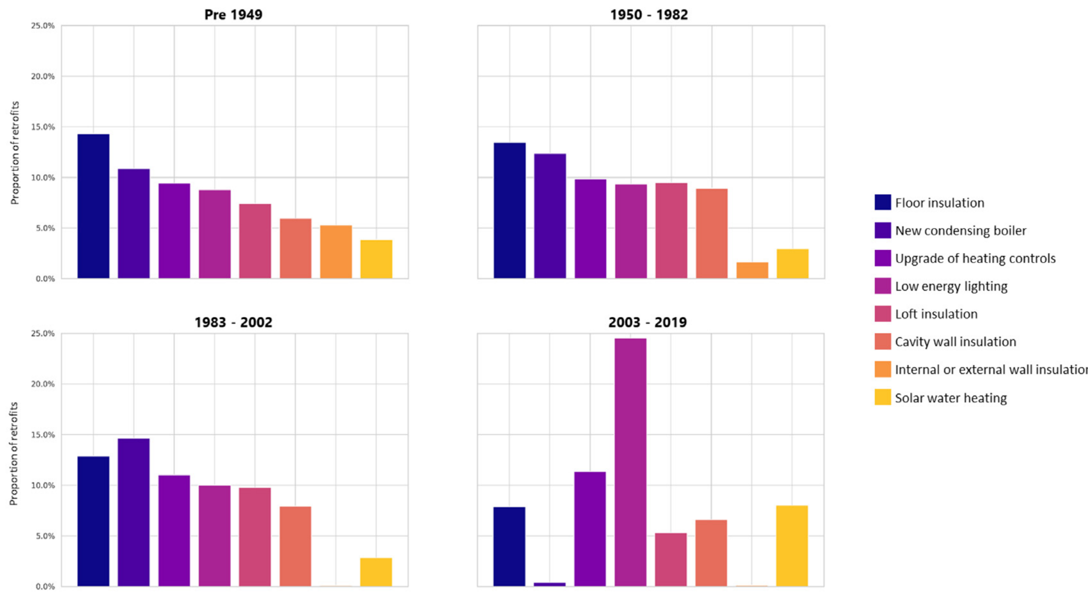

Fig. 5. Percentage of the eight most frequently carried out measures in the total number of retrofits per building age class.

the remaining 73.3%. The values indicate that our model recognizes a retrofit in 66% of cases (Sensitivity), while in 76% of cases, non-retrofits were also correctly classified in the model (Specificity). Our approach assigned 73% of the observations to the correct class (Accuracy). Consequently, the XGBoost model provides a similarly good prediction for both classes.

Subsequently, as the XGBoost model shows high predictive performance and thus is suitable to predict (non-)retrofits, we present the factors and relationships influencing (non-)retrofits utilizing SHAP values in more detail to derive policy implications from these findings. We, therefore, illustrate the results in SHAP summary plots combining feature importance and feature effects for the most critical factors. SHAP summary plots display features listed along the y-axis and the SHAP value on the x-axis. Interpreting the figures, the SHAP value of an individual data point represents the force and direction in which it "pushes" the overall prediction of the data point. For each feature, the data points in the dataset are colored by their respective value of that feature. Positive SHAP values indicate that a retrofit has been carried out, while negative SHAP values indicate the opposite effect.

We provide a first overview of the top features ranked by their importance as measured by the sum of the SHAP value's magnitude per feature in Fig. 6. Essential features represent a mixture of building properties (e.g., heating cost current, glazing proportion or wall, and roof energy efficiency), economic properties (median and mean house prices), and socio-demographic properties (e.g., a population from 0 to 15, employment rate, GDHI). Thus, as a first indication, the characteristics of all data sources used to enrich the EPC data appear to be important in predicting retrofits. In the following, we dive deeper into the analysis of individual features to derive details and further insights that allow the derivation of policy implications.

First, we analyze the effect of heating costs influencing retrofit measures conducted. The SHAP summary plot in Fig. 7 shows that retrofits are more likely when high heating costs are high. While not surprising, it is worth noting that while meager heating costs are responsible for a very negative impact on the likelihood of retrofits, very high heating costs do not have the same effect.

Fuel poverty is common in England & Wales (Department for Business, Energy & Industrial Strategy, 2020a). Fuel poverty typically occurs in households where "[…] the fuel costs are above average, and their disposable income (after housing and fuel costs) is below the poverty line" (Department for Business, Energy & Industrial Strategy, 2020b). The main drivers for this phenomenon resulting partly from the definition are low income (represented by a low GDHI), high fuel costs, and high fuel consumption, which often results from the poor energy efficiency of the building (Department for Business, Energy & Industrial Strategy, 2020a; Moore, 2012). Besides these drivers, fuel poverty is also indirectly caused by unemployment (the low employment rate). Households where the person responsible for the accommodation is unemployed, are three times more likely to suffer from fuel poverty (Department for Business, Energy & Industrial Strategy, 2020a). Consequently, retrofits would be necessary for households with low income living in areas with low employment rates. This circumstance also explains the non-symmetrical effects of very high and meager heating costs. The diminishing impact of retrofitting for low heating costs compared to high heating costs may represent this phenomenon, as some households may not be able to afford a retrofit. Since high heating costs indicate poor technical equipment and thus poor energy efficiency, it is essential to improve the energy efficiency considering the above points regarding fuel poverty. Supporting the people suffering from this phenomenon and initiating selected programs is necessary.

Second, we analyze the effect of the employment rate, the GDHI, and the education score on retrofit measures. For this purpose, we present Fig. 8 using a SHAP summary plot focusing on the three central features and a Spearman rank correlation matrix. Here, the education score is a synthetic value obtained by multiplying the percentage of a region's population per level of education by the level of education. We denote education levels as values between zero (no qualification) and four (bachelor's degree and higher). From the SHAP summary plot, we can infer that a high employment rate hampers retrofits, a high GDHI tends to impact retrofits negatively, and higher education score values positively impact retrofits.

Interestingly, while the education score has a (strong) positive correlation with employment rate and GDHI, their effects are (almost) contrary. While a low education score (i.e., more unqualified persons) impedes retrofits, a low employment rate, and GDHI positively affect retrofits' conduct. Since the employment rate and wage level (represented by the GDHI) are positively correlated (see spearman rank correlation matrix), one could

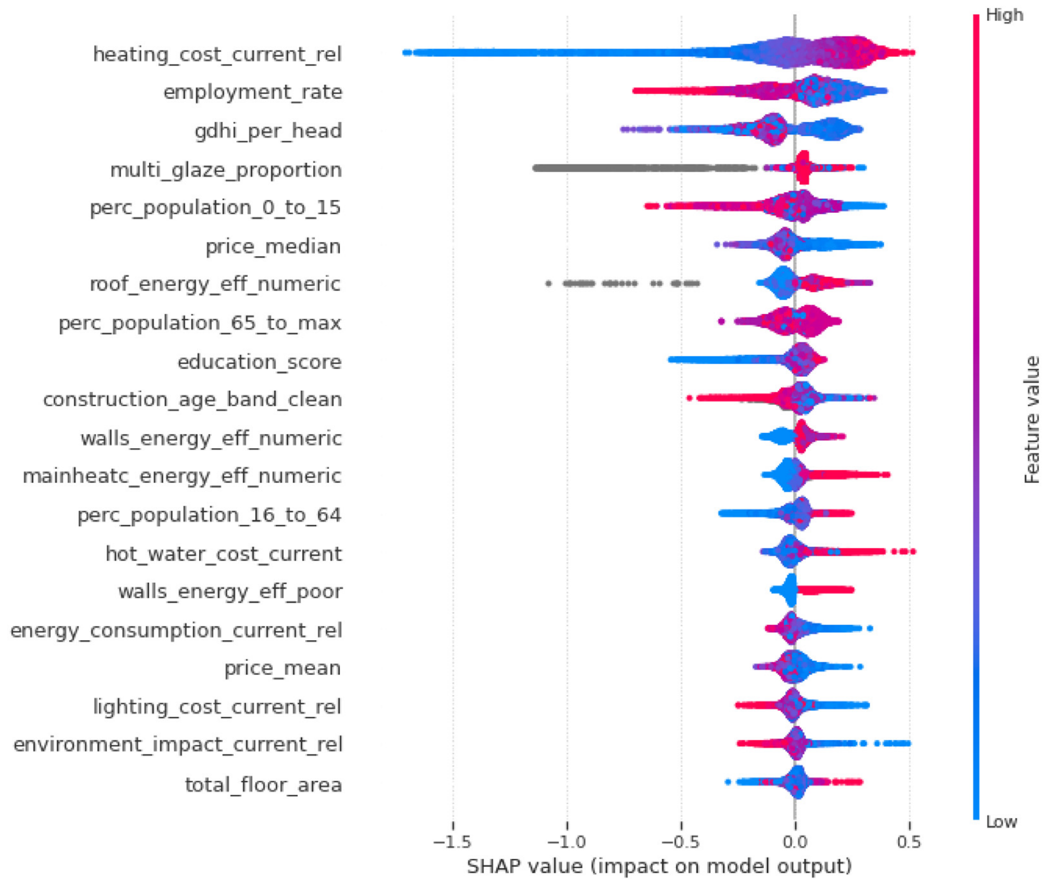

Fig. 6. SHAP values for top features.

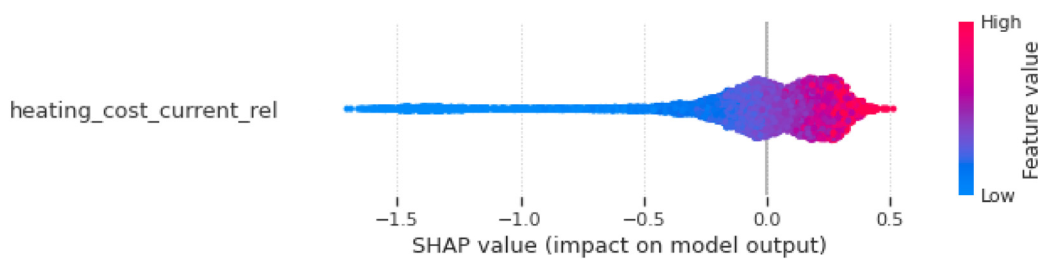

Fig. 7. SHAP values for heating costs.

conclude that high employment rates cause a higher wage level, which makes fuel expenses less relevant and thus retrofits less critical. Regarding SHAP values for a low GDHI and a low employment rate, opposing trends are evident regarding retrofitting. While we could assume that a low GDHI reduces the likelihood of retrofitting, there are also several cases in which households with lower incomes or households in regions with low employment rates are more likely to retrofit. The phenomenon of fuel poverty might explain the latter finding, as already discussed above. Local governments in the UK have recognized the problem of often unaffordable retrofits and have already adopted measures to support affected households (e.g., the scheme proposed by Gupta and Gregg (2018) or action plans of various local authorities such as Birmingham—see Section 5.1). Consequently, these implemented government programs may start to yield results and be responsible for this unusual relationship between low income and a high likelihood of retrofitting. Interestingly, we find slight differences when comparing our results with a recent and thematically related study by Hall et al. (2021) on innovative utility business models. Prior research such as Hall et al. (2021) concludes that lower income, homeownership, and education lead to lower preferences for new (innovative) utility business models and risk exacerbating existing social inequalities, which is not entirely consistent with our findings. Reasons for the differences in this regard might stem from differences in the timing of data collection and the subject of analysis. For example, our research focuses on past years, whereas Hall et al. (2021) examine present-day decisions. Consequently, they do not consider related government programs and their effects.

Regarding policy implications, we interpret our results by recommending that government programs should be similarly divided, as low-income households feel the (financial) need for retrofitting but often cannot afford it, and high-income households could afford it but have no financial need. The most obvious option is to force high-income households to retrofit (e.g., when buying or renting a property), while lower-income households must receive targeted assistance and consultations to highlight the positive financial effects of retrofitting, which is in line with Hammitt (2021) proposals.

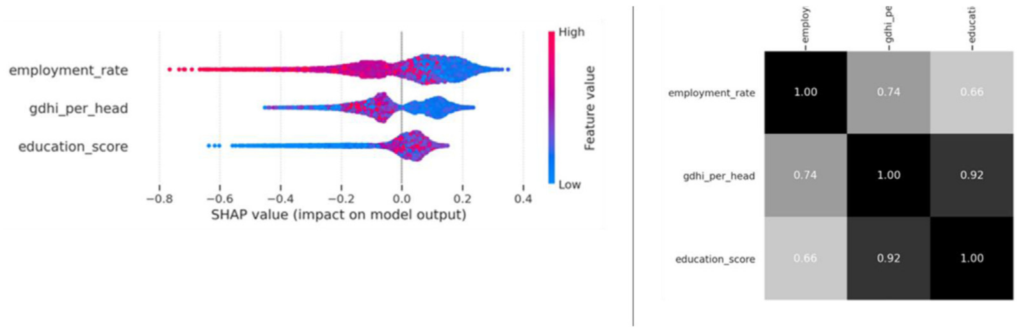

Fig. 8. SHAP values and spearman rank correlation matrix for employment rate, GDHI and, education score.

While there are already ongoing and effective political programs for low-income households, as described before, forcing high-income households to retrofit is likely an unpopular measure and not feasible from a policy perspective. Several schemes, e.g., the Green Deal, aiding households to retrofit, were previously in place, but the desired effect was never achieved (Dowson et al., 2012; Comerford et al., 2018).

With the implementation of CO_2 taxes, a possible approach would be to raise retrofitting-related CO_2 taxes for higher-income households with high energy consumption, incentivizing them to retrofit. We recommend not imposing taxes on lower-income households and an accumulating grant allowance depending on the heating costs given for each year. Therefore, a low-income household with a high heating bill would receive an (accumulating) grant each year for retrofitting. To mitigate the landlord-tenant dilemma, the landlord should be required to foot the CO2-related taxes (depending on her household). Due to numerous advantages, we propose this scheme as a specific policy measure:

First, Her Majesty Revenue & Customs (HMRC) already collects income taxes and is already informed about every household's income. Second, the only document required for the HMRC to compute an additional CO_2 tax or grant towards retrofitting is an energy bill. Third, lower-income households would automatically receive their grant amount with their yearly tax documents, which requires no further action from their side. Clearly stating an amount should help with uncertainty regarding the amount of funding. Fourth, the HMRC could either pay out the grant with the tax return or directly pay craftsmen, removing the need to foot a potentially large bill temporarily. Fifth, the proposal can be fiscally neutral, e.g., the additional taxes collected can fully offset the cost of the grants.

A similar scheme was proposed as a Variable Council Tax by the UK Association for the Conservation of Energy (UK Green Building Council, 2013) and found to be a suitable alternative as a future retrofit policy (Miu et al., 2018), which should now be forced based on our quantitative analysis.

Third, we analyze the effect of house prices using the mean and median house prices, again with a SHAP summary plot in Fig. 9. We see that retrofits are more likely where house prices are comparatively low, the market is not as tight, and energy efficiency might be more relevant in purchase decisions. In contrast, retrofits are less relevant when a tight market depicts higher prices (Adan and Fuerst, 2015). A potential hypothesis from a buyer perspective is that the high level of competition in the market leads people to refrain from their demands on the house (especially concerning energy efficiency) to increase the chance of a purchase. From the seller's perspective, there is hardly any need to improve the attractiveness of his house through retrofits, and consequently, there is little or no retrofitting.

We see several potential policy implications regarding the effect of house prices. Firstly, as tight markets negatively influence retrofits, relaxing the property market could lead to more retrofits – e.g., motivated through state-subsidized housing. Secondly, where the property is (highly) expensive, retrofits could be made mandatory. This is especially relevant for London. Properties in the British capital are hugely expensive, but London is worse, not better, than other Local Authorities in terms of average EPCs ratings. We propose retrofits up to a certain standard as mandatory for new rentals and sales in some high-priced regions. As the effects of such a policy would probably drive prices up further, researchers need to examine its possible consequences closely. Irrespective of this, we believe it is necessary to enforce at least some comparatively cost-effective retrofitting measures. One measure could be so-called green mortgages, as proposed by the UK Green Building Council (Miu et al., 2018). These offer the customer beneficial mortgaging conditions (e.g., discount on the interest rate) when the customer has either already paid attention to a good rating at the time of purchase or wants to improve the house's rating after purchase through suitable measures. In addition, governments could offer favorable follow-up loans (possibly in combination with state subsidies) for further retrofit measures. Furthermore, mortgages with poorer conditions (e.g., premium on the interest rate) can be granted to people who do not attach importance to the energy efficiency of their building and consequently do not want to improve their rating through retrofit measures. However, governments must consider separately households that suffer from fuel poverty and thus do not have the necessary means for retrofitting (Miu et al., 2018).

Fourth, we examine the effect of the population's age distribution in Fig. 10. The SHAP summary plot shows that higher shares of the population of children (ages 0 to 15) lead to fewer retrofits. The opposite is true with a more significant representation of the population aged 16–64, this large category, with a mean share per region of 63% of the total population and a standard deviation of only 1.7%, does not allow for more accurate conclusions. Where the population is comparatively old, we can make no statement as we could not identify a clear trend.

Given the concerns above, we focus on the youngest age group and the corresponding household type of families to discuss possible policy implications. There are two conceivable reasons families with children aged 15 or younger do fewer retrofits than families with older children. On the one hand, they may lack the monetary resources to carry out retrofits. On the other hand, there may also be non-monetary reasons, such as a lack of time or comfort and the additional administrative work that comes with retrofitting (Comerford et al., 2018). Findings of the Annual Fuel Poverty Statistics Report reinforce the former by stating that in households where the youngest member is aged below 15, fuel poverty is more likely to exist (Department for Business, Energy & Industrial Strategy, 2020a). Our analysis reinforces the consecutive hypotheses that fuel-poor households are

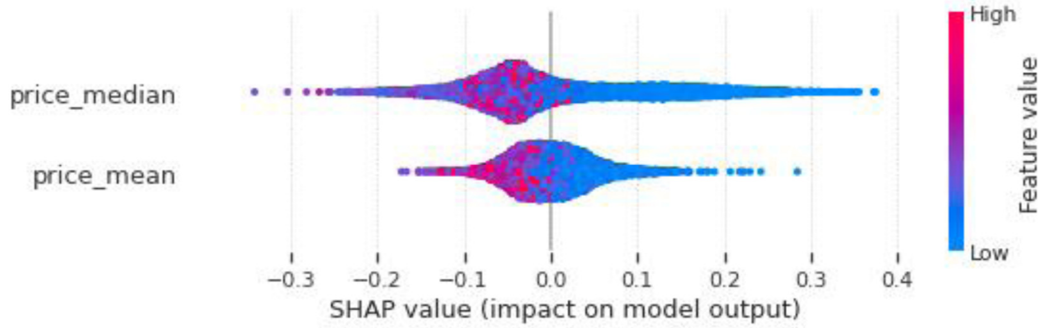

Fig. 9. SHAP values for median and mean house prices.

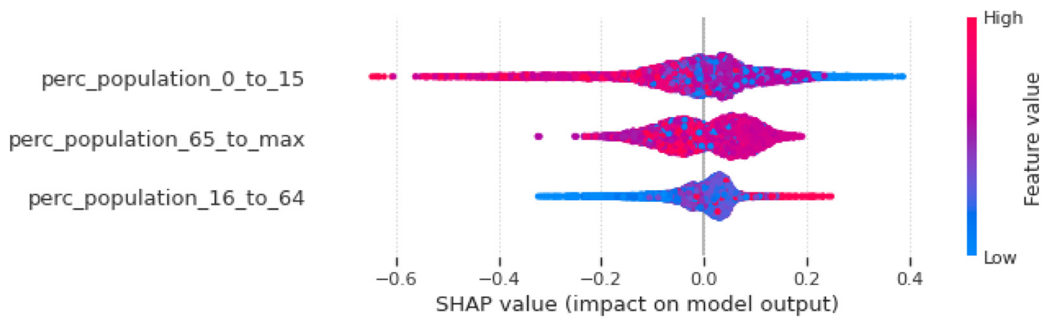

Fig. 10. SHAP values for the age distribution in the population.

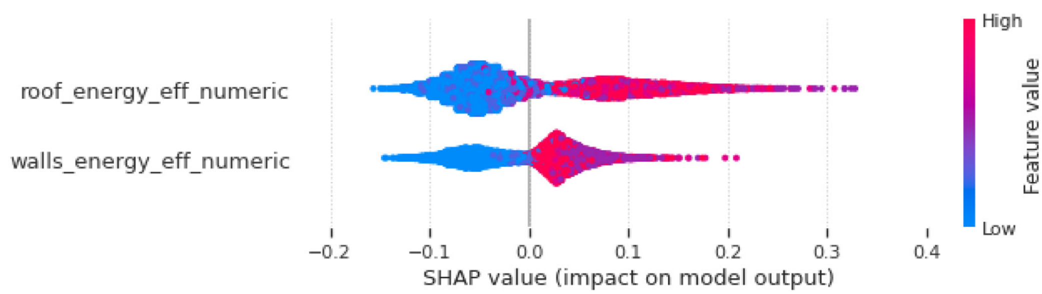

Fig. 11. SHAP values for roof and wall energy efficiency.

less likely to retrofit. Governments must create special programs for families with children to counteract these reasons. These programs should be characterized by monetary support and offer transparent information and fast, uncomplicated help in implementing the retrofits. A blueprint for this is the "Better Homes for Yorkshire" initiative, which might be replicated in other regions.

Fifth and last, we take a closer look at the effects of building characteristics—in detail, the roof and wall energy efficiency depicted in EPCs. Unsurprisingly, the SHAP summary plot in Fig. 11 shows that the low energy efficiency of walls and roofs positively impacts retrofitting. Note that low feature values denote a high efficiency. Although roof and wall renovations are expensive measures and severe interventions in the building envelope, they also significantly impact energy efficiency and savings. Additionally, the measures are well known, and the awareness of their positive effects is high (Kaveh et al., 2018).

As households that receive an EPC and recognize their low roof and wall energy efficiency are more likely to retrofit, our research suggests prioritizing recommendations that focus on wall and roof efficiency ratings. Thereby, decision-makers must consider the plethora of problems related to the wall and roof efficiency evaluation process and, if necessary, introduce higher quality requirements in creating EPCs.

6. Conclusions and policy implications

This study addressed the research gap of missing large-scale quantitative studies on influencing factors for or against energy retrofits. We showed which factors correlate with the implementation of retrofits and how to derive practical policy implications. Our contribution to the theoretical body of knowledge and identified policy implications can be divided into several points: First, our approach allows us to corroborate findings previously obtained in qualitative or small-scale studies with a quantitative approach using real-world data and identify further influencing factors. Second, we present a method for extracting building retrofits from the UK EPC data that is reproducible for further studies. In addition, the method can extract and analyze buildings' changes over time from EPCs. Third, we confirm existing studies on the influence of house prices on retrofit behaviors. Since higher house prices correlate with a lower likelihood of retrofits, we propose to make at least some comparatively low-cost retrofit measures mandatory for new rentals and sales in local authorities with high average house prices (e.g., London). Further, green mortgages offering the customer beneficial mortgaging conditions might be approaches to improve retrofitting rates in high-prices areas. Fourth, since families with children aged 15 or younger are less likely to carry out retrofits, special programs for families with children might be helpful considering their specific needs. Financial support, transparent information, and quick, straightforward help in implementing retrofits and minimizing family burdens might characterize these programs. Local authorities where the "Better Homes for Yorkshire" initiative was effective with corresponding characteristics showed many retrofits so that other regions might replicate the initiative. Fifth, since low energy efficiency of walls and roofs, is an essential criterion for energy efficiency and has a positive impact on retrofits, we propose to incentivize retrofits for buildings with EPCs, that exhibit poor levels of energy efficiency in walls and roofs.

Moreover, the abundance of problems related to the wall and roof efficiency evaluation process must be considered, which may require the Introduction of higher quality standards in preparing EPCs. Sixth, with fuel poverty being a common phenomenon in England and Wales, we consider retrofitting-related CO2 taxes as reasonable. For higher-income households with a high energy consumption raising retrofitting-related CO2 taxes would be an incentive to retrofit. Governments could cut taxes to low-income households, and a cumulative subsidy could be granted depending on the heating costs for each year. This approach could be implemented through Her Majesty Revenue & Customs (UK Government department responsible for collecting taxes) without much additional effort, as it already collects income taxes and is aware of each household's income. Thus, the only other document needed to calculate an extra CO2 tax or retrofit grant is an energy bill. In summary, our findings confirm existing studies with the help of data-driven approaches and that policy measures already found in individual cases should be rolled out broadly.

Naturally, our study has some limitations but likewise gives rise to new research potential. First, the data used in this study might be a limiting factor regarding data quality, level of detail, and regional and building type focus. Additional data sources could be used, for example, to allow investigations regarding the relationship between retrofits and health perceptions (Feng et al., 2021) or the comparison across building types and sectors to identify even more targeted policy measures. Second, different algorithms might be tested and compared with our XGBoost, and SHAP approach. Third, our study only focused on whether retrofits had been conducted but neglected what part of a building was retrofitted. Despite these limitations, this work provides important insights into better-retrofitting practices and thus assists policymakers in the UK in developing more effective measures to increase retrofits in the domestic sector to achieve climate goals.

CRediT authorship contribution statement

Simon Wenninger: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing, Project administration. **Philip Karnebogen:** Formal analysis, Writing – original draft, Writing – review & editing. **Sven Lehmann:** Writing – original draft, Data curation. **Tristan Menzinger:** Software, Data curation, Writing – original draft, Visualization. **Michelle Reckstadt:** Writing – original draft, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used in this work is cited and is open access available.

References

Achtnicht, M., Madlener, R., 2014. Factors influencing German house owners' preferences on energy retrofits. Energy Policy 68, 254–263.

Adadi, A., Berrada, M., 2018. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access 6, 52138–52160.

Adam, K.L., 2018. The Role of Collaboration in Realising Local Authority Energy Objectives: An Institutional and Stakeholder Perspective. University of Leeds.

Adan, H., Fuerst, F., 2015. Modelling energy retrofit investments in the UK housing market. Smart Sustain. Built Environ. 4 (3), 251–267.

Ahlrichs, J., Rockstuhl, S., Tränkler, T., Wenninger, S., 2020. The impact of political instruments on building energy retrofits: A risk-integrated thermal Energy Hub approach. Energy Policy 147, 111851.

Ahlrichs, J., Wenninger, S., Wiethe, C., Häckel, B., 2022. Impact of socio-economic factors on local energetic retrofitting needs - A data analytics approach. Energy Policy 160, 112646.

Arjunan, P., Poolla, K., Miller, C., 2020. EnergyStar++: Towards more accurate and explanatory building energy benchmarking. Appl. Energy 276, 115413.

Athey, S., 2017. Beyond prediction: Using big data for policy problems (in eng). Science 355 (6324), 483–485.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115.

Ben, H., Steemers, K., 2018. Household archetypes and behavioural patterns in UK domestic energy use. Energy Efficiency 11 (3), 761–771.

Bertoldi, P., Mosconi, R., 2020. Do energy efficiency policies save energy? A new approach based on energy policy indicators (in the EU member states). Energy Policy 139, 111320.

Birmingham City Council, 2010. Climate change action plan 2010+. Climate change and sustainability team. https://www.birmingham.gov.uk/download/downloads/id/1889/climate_change_action_plan_2010.pdf (accessed 24.04.2021).

Brown, D., 2018. Business models for residential retrofit in the UK: a critical assessment of five key archetypes. Energy Efficiency 11 (6), 1497–1517.

Burkart, N., Huber, M.F., 2021. A survey on the explainability of supervised machine learning. Jair 70, 245–317.

Burman, E., Mumovic, D., Kimpian, J., 2014. Towards measurement and verification of energy performance under the framework of the European directive for energy performance of buildings. Energy 77, 153–163.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Balaji, K., Mohak, S. (Eds.), KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA, pp. 785–794.

Ciaburro, G., Venkateswaran, B., 2017. Neural Networks with R: Smart Models using CNN, RNN, Deep Learning, and Artificial Intelligence Principles. Packt Publishing, Birmingham, UK.

Coma-Puig, B., Carmona, J., 2019. Bridging the gap between energy consumption and distribution through non-technical loss detection. Energies 12 (9), 1748.

Comerford, D.A., Lange, I., Moro, M., 2018. Proof of concept that requiring energy labels for dwellings can induce retrofitting. Energy Econ. 69, 204–212.

Csutora, M., Zsóka, Á., 2011. Maximizing the efficiency of greenhouse gas related consumer policy. J Consum Policy 34 (1), 67–90.

Department for Business, Energy & Industrial Strategy, 2020a. Annual fuel poverty statistics in England 2020 (2018 data). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/882404/annual-fuel-poverty-statistics-report-2020-2018-data.pdf (accessed 02.02.2021).

Department for Business, Energy & Industrial Strategy, 2020b. Fuel poverty factsheet: England, 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/882159/fuel-poverty-factsheet-2020-2018-data.pdf (accessed 30.01.2021).

Department for Business, Energy & Industrial Strategy, 2020c. Household energy efficiency detailed release: Great britain data to 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/873663/Detailed_Release_-_HEE_stats_19_March_2020.pdf (accessed 07.02.2021).

Department for Communities and Local Government, 2017. A guide to energy performance certificates for the marketing, sale and let of dwellings: Improving the energy efficiency of our buildings. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/671018/A_guide_to_energy_performance_certificates_for_the_marketing__sale_{and}_let_of_dwellings.pdf (accessed 08.01.2021).

Dhaliwal, S., Nahid, A., Abbas, R., 2018. Effective intrusion detection system using XGBoost. Information 9 (7), 1–24.

Dowson, M., Poole, A., Harrison, D., Susman, G., 2012. Domestic UK retrofit challenge: Barriers, incentives and current performance leading into the Green Deal. Energy Policy 50, 294–305.

Druckman, A., Jackson, T., 2008. Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model. Energy Policy 36 (8), 3177–3192.

Fawcett, T., Rosenow, J., Bertoldi, P., 2019. Energy efficiency obligation schemes: their future in the EU. Energy Efficiency 12 (1), 57–71.
Feng, Y., Duan, Q., Chen, X., Yakkali, S.S., Wang, J., 2021. Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods. Appl. Energy 291, 116814.
Filippini, M., Hunt, L.C., Zorić, J., 2014. Impact of energy policy instruments on the estimated level of underlying energy efficiency in the EU residential sector. Energy Policy 69, 73–81.
Forster, P.M., Forster, H.I., Evans, M.J., Gidden, M.J., Jones, C.D., Keller, C.A., Lamboll, R.D., Quéré, C.Le., Rogelj, J., Rosen, D., Schleussner, C.-F., Richardson, T.B., Smith, C.J., Turnock, S.T., 2020. Current and future global climate impacts resulting from COVID-19. Nat. Clim. Change. 10 (10), 913–919.
Fuerst, F., Haddad, M.F.C., Adan, H., 2020. Is there an economic case for energy-efficient dwellings in the UK private rental market? J. Cleaner Prod. 245, 1–12.
Fuerst, F., McAllister, P., Nanda, A., Wyatt, P., 2015. Does energy efficiency matter to home-buyers? An investigation of EPC ratings and transaction prices in England. Energy Econ. 48, 145–156.
Fylan, F., Glew, D., Smith, M., Johnston, D., Brooke-Peat, M., Miles-Shenton, D., Fletcher, M., Aloise-Young, P., Gorse, C., 2016. Reflections on retrofits: Overcoming barriers to energy efficiency among the fuel poor in the United Kingdom. Energy Res. Soc. Sci. 21, 190–198.
Golizadeh Akhlaghi, Y., Aslansefat, K., Zhao, X., Sadati, S., Badiei, A., Xiao, X., Shittu, S., Fan, Y., Ma, X., 2021. Hourly performance forecast of a dew point cooler using explainable artificial intelligence and evolutionary optimisations by 2050. Appl. Energy 281, 116062.
Gómez-Navarro, T., Calero-Pastor, M., Pellicer-Sifres, V., Lillo-Rodrigo, P., Alfonso-Solar, D., Pérez-Navarro, Á., 2021. Fuel poverty map of Valencia (Spain): Results of a direct survey to citizens and recommendations for policy making. Energy Policy 151, 112162.
Gonzalez Caceres, A., 2018. Shortcomings and suggestions to the EPC recommendation list of measures: In-depth interviews in six countries. Energies 11 (10), 2516.
Government Digital Service, 2021. Statistical data set - price paid data. https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads (accessed 24.04.2021).
Gupta, R., Gregg, M., 2018. Targeting and modelling urban energy retrofits using a city-scale energy mapping approach. J. Cleaner Prod. 174, 401–412.
Hall, S., Anable, J., Hardy, J., Workman, M., Mazur, C., Matthews, Y., 2021. Matching consumer segments to innovative utility business models. Nat. Energy 6 (4), 349–361.
Hammitt, J.K., 2021. The future costs of methane emissions (in eng). Nature 592 (7855), 514–515.
Hardy, A., Glew, D., 2019. An analysis of errors in the energy performance certificate database. Energy Policy 129, 1168–1178.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed. Springer, New York.
Herrando, M., Cambra, D., Navarro, M., La Cruz, L., de Millán, G., Zabalza, I., 2016. Energy performance certification of faculty buildings in Spain: The gap between estimated and real energy consumption. Energy Convers. Manage. 125, 141–153.
HM Land Registry, 2016. Guidance: How to access price paid data - information about price paid data and how you can get it. https://www.gov.uk/guidance/about-the-price-paid-data (accessed 06.01.2021).
Jakob, M., 2007. The drivers of and barriers to energy efficiency in renovation decisions of single-family home-owners. In: CEPE Center for Energy Policy and Economics. ETH Zürich, CEPE Working paper series.
Jenkins, D., Simpson, S., Peacock, A., 2017. Investigating the consistency and quality of EPC ratings and assessments. Energy 138, 480–489.
Kastner, I., Stern, P.C., 2015. Examining the decision-making processes behind household energy investments: A review. Energy Res. Soc. Sci. 10, 72–89.
Kaveh, B., Mazhar, M.U., Simmonite, B., Sarshar, M., Sertyesilisik, B., 2018. An investigation into retrofitting the pre-1919 owner-occupied UK housing stock to reduce carbon emissions. Energy Build. 176, 33–44.
Kaymakci, C., Wenninger, S., Sauer, A., 2021. A Holistic Framework for AI Systems in Industrial Applications, Vol. 16. Internationale Tagung Wirtschaftsinformatik.
Ke, J., Qin, Y., Wang, B., 2020-2020. Optimizing and controlling building electric energy using cat boost under the energy internet of things. In: 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2). 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2). IEEE, Wuhan, China, pp. 553–558.
Khazal, A., Sønstebø, O.J., 2020. Valuation of energy performance certificates in the rental market – professionals vs. nonprofessionals. Energy Policy 147, 111830.
Krausmann, F., Wiedenhofer, D., Haberl, H., 2020. Growing stocks of buildings, infrastructures and machinery as key challenge for compliance with climate targets. Global Environ. Change 61, 102034.
Kühl, N., Hirt, R., Baier, L., Schmitz, B., Satzger, G., 2020. How to conduct rigorous supervised machine learning in information systems research: The supervised machine learning reportcard. Commun. Assoc. Inf. Syst. https://aisel.aisnet.org/cais/vol48/iss1/46/.
Li, Y., Kubicki, S., Guerriero, A., Rezgui, Y., 2019. Review of building energy performance certification schemes towards future improvement. Renew. Sustain. Energy Rev. 113, 109244.
Lundberg, S.M., Erion, G.G., Lee, S., 2018. Consistent individualized feature attribution for tree ensembles.
Lundberg, S., Lee, S., 2017. A Unified Approach to Interpreting Model Predictions. USA.
Ma, J., Cheng, J.C., 2016. Identifying the influential features on the regional energy use intensity of residential buildings based on Random Forests. Appl. Energy 183, 193–201.
Magnani, N., Carrosio, G., Osti, G., 2020. Energy retrofitting of urban buildings: A socio-spatial analysis of three mid-sized Italian cities. Energy Policy 139, 111341.
Menezes, A.C., Cripps, A., Bouchlaghem, D., Buswell, R., 2012. Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. Appl. Energy 97, 355–364.
Miller, C., 2019. What's in the box?! towards explainable machine learning applied to non-residential building smart meter classification. Energy Build. 199, 523–536.
Ministry of Housing, Communities & Local Government, 2020. Energy performance of buildings data england and Wales: Guidance. https://epc.opendatacommunities.org/docs/guidance (accessed 13.12.2020).
Mitchell, R., Frank, E., 2017. Accelerating the xgboost algorithm using GPU computing. PeerJ Comput. Sci. 3, 1–37.
Miu, L., Wisniewska, N., Mazur, C., Hardy, J., Hawkes, A., 2018. A simple assessment of housing retrofit policies for the UK: What should succeed the energy company obligation? Energies 11 (8), 1–22.
Moore, R., 2012. Definitions of fuel poverty: Implications for policy. Energy Policy 49, 19–26.
Müller, O., Junglas, I., vom Brocke, J., Debortoli, S., 2016. Utilizing big data analytics for information systems research: Challenges, promises and guidelines. Eur. J. Inf. Syst. 25 (4), 289–302.
Office for National Statistics, 2020a. Energy efficiency of housing in England and Wales: Analysis of the energy efficiency, estimated carbon dioxide emissions and energy cost of dwellings in England and Wales with an energy performance certificate. https://www.ons.gov.uk/peoplepopulation{and}community/housing/articles/energyefficiencyofhousinginengl{and}andwales/2020-09-23/pdf (accessed 10.01.2021).
Office for National Statistics, 2020b. Qualifications of working age population (NVQ), borough. https://data.london.gov.uk/dataset/qualifications-working-age-population-nvq-borough (accessed 09.05.2021).
Office for National Statistics, 2020c. Regional gross disposable household income: all NUTS level regions. https://www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome/datasets/regionalgrossdisposablehouseholdincomegdhi (accessed 09.05.2021).
Office for National Statistics, 2021. Employment rates. https://data.london.gov.uk/dataset/employment-rates (accessed 09.05.2021).
Office of Gas and Electricity Markets, 2013. The final report of the carbon emissions reduction target (CERT) 2008–2012. https://www.ofgem.gov.uk/ofgem-publications/58425/certfinalreport2013300413pdf (accessed 07.02.2021).
Papadopoulos, S., Kontokosta, C.E., 2019. Grading buildings on energy performance using city benchmarking data. Appl. Energy 233–234, 244–253.
Pasichnyi, O., Wallin, J., Levihn, F., Shahrokni, H., Kordas, O., 2019. Energy performance certificates — New opportunities for data-enabled urban energy policy instruments? Energy Policy 127, 486–499.
Piddington, J., Nicol, S., Garrett, H., Custard, M., 2020. The housing stock of the United Kingdom. https://files.bregroup.com/bretrust/The-Housing-Stock-of-the-United-Kingdom_Report_BRE-Trust.pdf (accessed 23.12.2020).
Poel, B., van Cruchten, G., Balaras, C.A., 2007. Energy performance assessment of existing dwellings. Energy Build. 39 (4), 393–403.
Prieler, M., Leeb, M., Reiter, T., 2017. Characteristics of a database for energy performance certificates. Energy Procedia 132, 1000–1005.
Rockstuhl, S., Wenninger, S., Wiethe, C., Häckel, B., 2021. Understanding the risk perception of energy efficiency investments: Investment perspective vs. energy bill perspective. Energy Policy 159, 112616.
Salata, Ferdinando, et al., 2017. Heading towards the nZEB through CHP+HP systems. A comparison between retrofit solutions able to increase the energy performance for the heating and domestic hot water production in residential buildings. Energy Convers. Manage. 61–76. http://dx.doi.org/10.1016/j.enconman.2017.01.062.
Setyantho, G.R., Chang, S., 2020. Identification of primary factors influencing energy consumption patterns of commercial and residential buildings. Kieae 20 (6), 21–30.
Seyrfar, A., Ataei, H., Movahedi, A., Derrible, S., 2021. Data-driven approach for evaluating the energy efficiency in multifamily residential buildings. Pract. Period. Struct. Des. Constr. 26 (2), 4020074.

Shaw, K., 2019. County durham housing strategy. https://www.durham.gov.uk/media/31966/Housing-strategy-County-Durham/pdf/HousingStrategyCountyDurhamJuly19.pdf?m=637147016851070000 (accessed 24.04.2021).

Shearer, C., 2000. The CRISP-DM model: The new blueprint for data mining. J. Data Warehousing 5 (4), 13–22.

Somu, N., M.R., G.R., Ramamritham, K., 2020. A hybrid model for building energy consumption forecasting using long short term memory networks. Appl. Energy 261, 114131.

The Office of the Deputy Prime Minister, 2005. Guide to the condensing boiler installation assessment procedure for dwellings. https://www.gov.je/SiteCollectionDocuments/Planning%20{and}%20building/ID%20AssessmentGuidetoCondensingBoilerInstallationinDwellings%2020100910%20mm.pdf (accessed 07.02.2021).

Tziogas, C., Papadopoulos, A., Georgiadis, P., 2021. Policy implementation and energy-saving strategies for the residential sector: The case of the Greek Energy Refurbishment program. Energy Policy 149, 112100.

UK Green Building Council, 2013. Retrofit incentives: Boosting take-up of energy efficiency measures in domestic properties. https://www.ukgbc.org/wp-content/uploads/2017/09/13070520Retrofit20Incentives20Task20Group20-20Report20FINAL_1.pdf (accessed 02.05.2021).

Watts, C., Jentsch, M.F., James, P.A.B., 2011. Evaluation of domestic energy performance certificates in use. Building Services Engineering Research and Technology 32 (4), 361–376.

Welsh Government, 2020. National level population estimates by year, age and UK country. https://statswales.gov.wales/Catalogue/Population-and-Migration/Population/Estimates/nationallevelpopulationestimates-by-year-age-ukcountry (accessed 09.05.2021).

Wenninger, S., Kaymakci, C., Wiethe, C., 2022. Explainable long-term building energy consumption prediction using qlattice. Appl. Energy 308, 118300.

Wenninger, S., Wiethe, C., 2021. Benchmarking energy quantification methods to predict heating energy performance of residential buildings in Germany. Business & Information Systems Engineering.

Wilde, P.de, 2014. The gap between predicted and measured energy performance of buildings: A framework for investigation. Autom. Constr. 41, 40–49.

Wilson, C., Crane, L., Chryssochoidis, G., 2015. Why do homeowners renovate energy efficiently? Contrasting perspectives and implications for policy. Energy Res. Soc. Sci. 7, 12–22.

Zhang, T., Siebers, P., Aickelin, U., 2012. A three-dimensional model of residential energy consumer archetypes for local energy policy design in the UK. Energy Policy 47, 102–110.