Journal of Building Engineering 95 (2024) 110279

Contents lists available at ScienceDirect

Journal of Building Engineering

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

Exploring green building certification credit selection: A model based on explainable machine learning

Yixin Li [a], Xiaodong Li [a], Dingyuan Ma [b,*], Wei Gong [c]

[a] Department of Construction Management, School of Civil Engineering, Tsinghua University, Beijing, 100084, China
[b] Department of Construction Management, School of Urban Economics and Management, Beijing University of Civil Engineering and Architecture, Beijing, 100044, China
[c] Center of Science and Technology & Industrialization Development Center of Housing Industrialization, Ministry of Housing and Urban-Rural Development of the PR China, No. 9 Sanlihe Road, Beijing, 100835, China

A R T I C L E I N F O

Keywords:
Green building
Certification credits
Explainable machine learning
XGBoost
Shapley values

A B S T R A C T

Green buildings represent a promising solution for advancing high-quality development in the building sector to combat climate change. Selecting appropriate credits from the rating system based on the distinct characteristics of buildings is a crucial step in the green building certification process. However, in practice, credit selection is a complex and challenging task, often relying on the personal experience of experts. This study develops a model based on explainable machine learning techniques, aiming to aid architects in selecting suitable credits in the early stages of green building design and explore the impact of various factors on credit selection. A case library of 210 green buildings is established to verify the model's performance. The model demonstrated a notable precision rate of 82.38 % in the selection of regular credits. Leveraging SHapley Additive exPlanations (SHAP) technology, the model uncovers a pattern indicating that buildings sharing specific characteristics tend to exhibit similar performance on particular credits, suggesting an inherent preference or avoidance of these credits. The model developed in this study offers practical strategies for architects in credit selection, reducing the reliance on expert opinions and simplifying the credit selection process. The introduction of explainable machine learning techniques enhances the transparency of model decisions and provides targeted insights for architects and standard setters.

Abbreviation

GB	Green building
GBES	Green Building Evaluation Standard
GBT	Green building technology
TC	Target credits
RC	Regular credit
BC	Bonus credit
ML	Machine learning
C&O buildings	Commercial and office buildings
SECH buildings	Science, education, culture and health buildings

* Corresponding author. No.1 Zhanlanguan Road, Xi Cheng, Beijing, 100044, China.
E-mail address: madingyuan@bucea.edu.cn (D. Ma).

https://doi.org/10.1016/j.jobe.2024.110279
Received 24 April 2024; Received in revised form 16 July 2024; Accepted 24 July 2024
Available online 25 July 2024

1. Introduction

Despite the pivotal role the construction industry plays in meeting people's living needs and driving economic development, it is being regarded as a significant contributor to global climate change due to its high energy consumption and emissions [1,2]. In recent years, promoting green building has become a critical strategy in the industry for energy conservation and emissions reduction. Green building (GB) aims to conserve resources throughout the life cycle of a building, minimize negative impacts on the environment, and foster a healthier and more comfortable living environment for occupants [3–5]. GBs often refer to buildings that have been certified by GB evaluation standards and rating systems [6,7].

Currently, the widely used and popular GB evaluation tools include UK's BREEAM (Building Research Establishment Assessment Method), US's LEED (Leadership in Energy and Environmental Design), etc. To promote the green development of the construction industry, China's Ministry of Housing and Urban-Rural Development (MOHURD) introduced the Green Building Evaluation Standard (GBES) in 2006, which is China's most widely used and fundamental national standard for evaluating GBs [8,9]. LEED, BREEAM, and GBES, these three widely-used assessment systems evaluate buildings based on specific credits, with the overall rating of a building determined by its cumulative score. Overall, assessment standards in China, the United States, the United Kingdom, and other countries or regions exhibit substantial similarities. They commonly address aspects such as water conservation, energy efficiency, material conservation, sustainable site development, and indoor and outdoor environmental quality [10–12]. Additionally, the credit allocation and verification of different standards are roughly similar in certain aspects [13].

In the early design phase of a GB project, architects need to establish green design strategies and develop measures for aspects such as energy efficiency, water conservation, and indoor environmental quality, in conjunction with certification requirements and building characteristics. The initial stage of GB certification involves selecting appropriate target credits (TCs) from the rating system, which is crucial for the early design phase of green building projects and significantly impacts the success of GB projects [7,14]. However, in the actual evaluation process, selecting the appropriate TCs that align with the characteristics of a building project from the numerous optional credits within a short period is a challenging and complex task. The same credit can yield varying levels of green or functional benefits depending on the building it is applied to Ref. [7]. It often requires experts to rely on their personal experience and knowledge to make judgments and selections. Currently, there is no established structured guide to assist architects in quickly selecting the appropriate optional credits from rating systems. Consequently, this step is both time-consuming and labor-intensive, potentially hindering the progress of green building projects [5,15].

Some research has begun to focus on credit selection in green building (GB) projects. One approach involves using traditional statistical methods to analyze differences in credit performance under various conditions. For instance, Wu et al. utilized non-parametric statistical methods to investigate whether significant differences exist in LEED credits across different building types and regions [16]. Similarly, Svetlana Pushkar employed similar research methods to discuss LEED-EB V4 projects of varying scales [17]. Although these studies can obtain information based on past projects, they cannot offer credit selection strategies for new GB samples under differing conditions.

Additionally, some studies have utilized multi-criteria decision-making methods to develop decision support tools for GB credit selection. For example, Seyis and Ergen integrated the Delphi method, multi-attribute utility technique, and TOPSIS technique to create a decision support tool aiding GB practitioners in selecting suitable credits based on project team characteristics [5]. While this tool proved effective in case studies, it still requires expert input for attribute weighting allocation and lacks emphasis on building-specific features. Similarly, Attallah et al.'s research is also based on structured interviews for the development of credit selection methods [18]. Overall, this type of research still relies on expert knowledge, and the process of selecting TCs remains cumbersome.

In light of the increasing volume of GB samples and the intricate relationships among various influencing factors, machine learning (ML) methods have gained attention for their significant advantages in handling large datasets, simulating complex relationships, and uncovering latent knowledge [19,20]. Notably, the application of ML in GB credit selection has been explored in some pioneering studies. For instance, Ma and Cheng developed a system for selecting LEED credits based on project information and climate factors using the random forest algorithm [21]. Similarly, the use of ML to scrutinize critical factors affecting credit selection has been noted, even if tools for credit selection weren't developed. For example, Juan and Lee employed association rule mining to understand the effects of building type and grade on the adoption of GB technologies [14]. In summary, there have been studies using various methods to discuss the selection of GB credits, which have provided important insights for us in data processing and framework construction. Also, ML methods have offered promising approaches to solving this problem. Nonetheless, these studies encounter certain limitations. Primarily, they tend to consider a narrow set of influencing factors [16,21]—occasionally focusing on a single one type [5,22]—despite the fact that GB credit selection is influenced by a multitude of factors (such as building characteristics and regional features). Secondly, there is a lack of in-depth discussion on how different factors affect credit selection, which challenges the trust of various stakeholders in the ML decision results. Additionally, many GB assessment systems have introduced bonus credits related to innovative

Table 1
GB assessment system with bonus credits (BCs).

GB assessment system	Chapters about BCs	Maximum points allocated to BCs	Maximum total points for all credits
LEED BD + C: New Construction v4.1	Innovations	5 points	110 possible points
	Regional Priority	4 points	
BREEAM International New Construction 2016	Innovation	10 points	150 possible points
GBES V2014	Enhancement and Innovation	10 points	110 possible points

technologies, as shown in Table 1, a key aspect that previous studies have overlooked. Bonus credits often focus on innovative technologies, which are crucial and cannot be overlooked for the development of GBs [23]. These credits are worthy of focused discussion.

Based on the limitations of previous studies, the work of this paper is as follows: (1) To establish a decision support model for TC selection based on ML, predicting green building credit selection strategies, by considering a series of factors such as building characteristics, urban features, and architecture company attributes. This model aims to help simplify the complexity architects face when selecting appropriate credits in the early design stages, reduce reliance on individual expert knowledge, and assist architects in making more optimized decisions. (2) To analyze the key features that impact credit selection using explainable ML techniques, and to understand how these features influence the selection of TCs and the reasons behind their impact. Addressing the 'black box' nature of machine learning decisions, this model aims to enhancesthe credibility of decisions, making the decision results more acceptable to various stakeholders. (3) To conduct a case study based on China's GB projects to verify the model's effectiveness and to delve into the impact of different features. This study introduces an explainable ML framework that combines Extreme Gradient Boosting (XGBoost) and Shapley Additive Explanations, meeting the needs of practical decision support while deepening the understanding of the influencing factors.

The rest of this paper is organized as follows. Section 2 presents the methodological of the research, which is followed by the case study section. Section 4 illustrates and discusses the results obtained from the study. Section 5 concludes the paper.

2. Research methodology

2.1. Research framework

The research framework of this study is illustrated in Fig. 1. **The first step** involves establishing the GB project database. Features that may affect credit selection are identified, and data on features and credits are collected and preprocessed to meet the requirements of ML. The Synthetic Minority Over-Sampling Technique (SMOTE) is introduced to handle highly skewed innovative credit data. Following the establishment of the GB project database, the core aspect of this study commences: ML model construction and the analysis of the influencing factors. **The second step** is the establishment of the ML model. The ML model developed in this study takes four identified feature categories as input and generates TC selection strategies as output. This model aims to offer viable credit selection strategies for architects, facilitating swift and suitable decision-making. During model evaluation, the performance of ML models is compared before and after oversampling of reward credits to validate the effectiveness of the SMOTE method. **The third step** involves a thorough investigation of the factors influencing various credit options, for which separate ML models for each credit are developed. The ML Models with superior classification performance are further analyzed using SHAP to identify crucial influencing factors. An analysis is then conducted to understand how these factors affect credit selection and their underlying reasons.

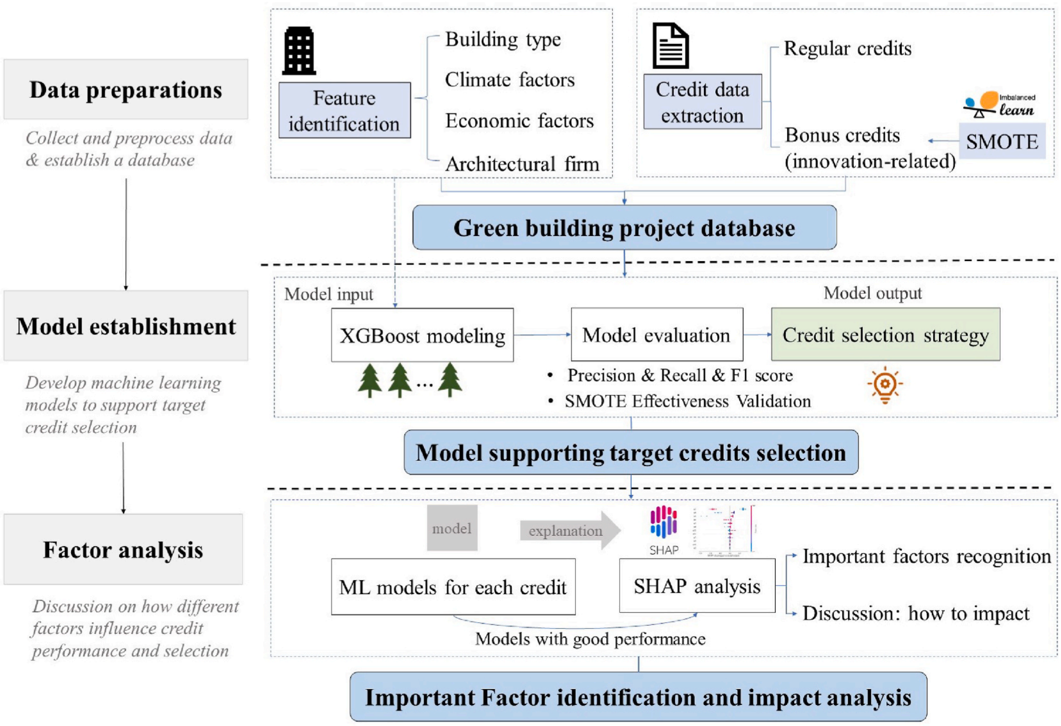

Fig. 1. Research framework.

2.2. Influencing factors

Influencing factors serve as inputs for machine learning models. Drawing from prior research, we've identified four categories encompassing a total of 12 influencing factors. These factors span building type, climate considerations, the economic level of the building's location, and the attributes of companies applying for the GB project.

The first category pertains to building type, where prior studies have established a robust correlation between building type and the adoption of green building technologies [14]. In this study, buildings are categorized into three distinct types: commercial and office buildings (encompassing malls, office spaces, abbreviated as C&O buildings), science, education, culture, and health buildings (encompassing schools, hospitals, kindergartens, abbreviated as SECH buildings), and residential buildings.

The second category revolves around climate factors, recognizing their significant impact on building design. The influence of climate factors on building design is of significance. Identical building technologies may exhibit varying performance outcomes in diverse climatic conditions [24–26]. In addition, the climate of the city in which the building is located has an impact on the prioritization of the use of sustainable technologies in buildings [27]. For instance, arid areas necessitate heightened attention to water efficiency in buildings. Our study incorporates seven commonly used climate factors [20]: average temperature, summer and winter temperatures, temperature differentials, annual precipitation, annual sunshine hours, and average wind speed.

The third category addresses economic factors. Economically developed regions often boast mature GB markets, potentially enjoying superior access to GB professionals and favorable governmental policies [28–30]. Additionally, as living standards rise, residents increasingly demand environmentally conscious and comfortable living spaces [31]. Consequently, buildings in economically prosperous areas may exhibit preferences for certain credits. We employ per capita gross domestic product (GDP) as an economic indicator, reflecting a city's economic health and the specific living standards of its residents [6].

The fourth category delves into company characteristics. Attributes of project teams, such as educational backgrounds, experience levels, and scale, significantly influence project success and credit selection [5,18,32]. Within this category, we consider whether the architectural design firm ranks among the industry's top enterprises as an influencing factor. It is conjectured that industry-leading architectural firms may outperform their counterparts on certain credits requiring higher technical complexity. In summary, these influencing factors are detailed in Table 2.

2.3. Machine learning

In today's era, people are confronted with vast amounts of data, which can serve as invaluable resources for knowledge discovery [33,34]. Machine learning stands as a useful approach, facilitating the transition from data to knowledge by extracting latent patterns or regularities from massive datasets, thereby aiding in knowledge acquisition [35–37].

This study adopts XGBoost algorithm to establish ML models. XGBoost is an improved machine learning algorithm based on Gradient Boosting Decision Tree (GBDT). This algorithm iteratively constructs multiple decision tree models and optimizes the model's predictive ability using gradient descent. It employs a strategy of parallel computation, boosting the training speed of the model. Additionally, unlike GBDT, XGBoost incorporates a regularization term in the objective function to mitigate overfitting [38]. XGBoost has demonstrated outstanding performance in many studies in the field of construction [39,40]. Additionally, the XGBoost algorithm can be effectively combined with SHAP technology to enhance the interpretability of the machine learning model [41]. This advanced algorithm consists of multiple categorical regression trees (CARTs) for joint decision making [38] expressed as an additive function in Equation (1):

$$\widehat{y}_i = \sum_{k=1}^{K} f_k(\mathbf{x}_i), f_k \in \mathscr{F} \tag{1}$$

where, $\widehat{y}_i$ represents the predicted value of the model, $\mathbf{x}_i$ denotes the i_{th} input sample of the model, K signifies the total number of CARTs, $\mathscr{F}$ represents the CART space, and f_k is an independent tree.

Table 2
Influencing factor descriptions.

Category	Influencing factor	Variable encoding	Index	Reference
Building type	Commercial & office building	[1, 0, 0]	x1	[14,16]
	Science, education, culture & health building	[0, 1, 0]	x2	
	Residential building	[0, 0, 1]	x3	
Climate factor	Average temperature	Continuous variable	x4	[21,22,24–27]
	Summer temperature	Continuous variable	x5	
	Winter temperature	Continuous variable	x6	
	Temperature difference	Continuous variable	x7	
	Annual precipitation	Continuous variable	x8	
	Annual sunshine time	Continuous variable	x9	
	Average wind speed	Continuous variable	x10	
Economic factor	GDP per capita	Continuous variable	x11	[6,28–30]
Architectural firm	Whether the architectural design firm is at the top of the industry	Binary variable (0/1)	x12	[5,18,32]

2.4. SHapley additive exPlanations (SHAP) technique

Machine learning models are often criticized for their "black box" nature. In the era of big data, the use of complex models demonstrates superior performance, however, the interpretability of complex machine learning models is often more challenging than simple ones, and has drawn increasing attention in recent research [41–43]. The SHapley Additive exPlanations (SHAP) technique, proposed by Lundberg et al. based on cooperative game theory, serve as a valuable tool to understand the modeling process [42]. SHAP value not only assesses the importance of features, but also presents whether the influence of features is positive or negative.

The SHAP technique employs an additive feature attribution method, represented by Equation (2):

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i x_i' \tag{2}$$

In Equation (4), f is defined as the original predictive model and g is the explanatory model. x' is the simplified input feature and M denotes the number of features. $\phi_0 = E[f(x)]$ denotes the base value of the prediction, and ϕ_i represents the SHAP value of the i_{th} feature. Explanatory models commonly utilize simplified inputs, which are mapped to the original inputs by the mapping function $x = h_x(x')$. Here, $x' \in \{0,1\}^K$, meaning that the simplified feature value is either 1 or 0, which signifies "present" and "absent" respectively. The absence of a feature implies it has no effect on the final prediction and the model will randomly select the value of the feature.

The calculation of the SHAP value is expressed by Equation (3):

$$\phi_i(f,x) = \sum_{z' \subseteq x'} \frac{|z'|!(M - |z'| - 1)!}{M!} [f_x(z') - f_x(z' \setminus i)] \tag{3}$$

where $f_x(z') = f(h_x(z'))$, $z' \setminus i$ denotes the setting $z_i' = 0$, $|z'|$ is the number of non-zero entries in z', and z' is a subset.

3. Case study

3.1. Building case collection

A total of 210 GB projects certified by China's Green Building Evaluation Standard V2014 were collected. The standard is aimed at general buildings for civilian use, including residential and public buildings. This study primarily focuses on new buildings because the extensive design choices and potential changes in new building projects significantly impact energy consumption and sustainability goals. Decision-makers require more reliable decision support information during the design process of new buildings. Additionally, many new buildings have been certified by the Chinese GBES, which has accumulated a large number of cases and provides valuable reference. These 210 projects are geographically dispersed across 20 provinces within China, comprising 16 one-star buildings, 166 two-star buildings and 28 three-star buildings. The distribution of building types and star ratings for the 210 projects is illustrated in Fig. 2 below.

Green Building Evaluation Standard divides criteria into two main categories: prerequisites and credits. The prerequisites are mandatory and are not the subject of this study, while credits are optional and are selected by architects based on the characteristics of the building. The credits are subdivided into regular credits (RCs) and bonus credits (BCs). RCs cover aspects such as water conservation, energy saving, material efficiency, and indoor environmental quality. BCs are mostly related to innovative building technologies, which are more challenging to achieve. Each criterion is represented by a code consisting of three digits. As an example, in the code "4.2.5," the first number "4″ represents the fourth chapter. The second number indicates the category of the criterion; for instance, "2″ stands for a credit, whereas "1″ represents a prerequisite. The third number "5″ denotes the fifth credit. Thus, "4.2.5″ signifies the fifth credit in the fourth chapter. The standards permit projects to skip the assessment of certain credits, as not all buildings are equipped with the facilities or specific rooms mandated by the evaluation criteria, such as public bathrooms and subterranean spaces. Therefore, those buildings without specific facilities will directly skip the corresponding credits. This flexibility can lead to a deficiency in valid credit point data when the number of projects not subject to evaluation becomes excessive. To ensure adequate data for

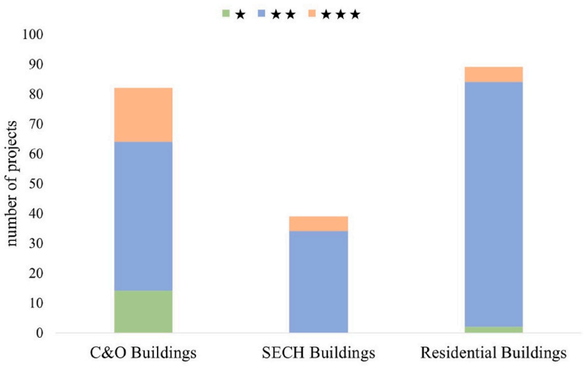

Fig. 2. Building type and star rating distribution of 210 GB projects.

machine learning inputs, we excluded credits skipped by more than 50 building projects, leaving 42 RCs.

3.2. Data processing

In measuring credit performance, two common approaches to handling credit data have been identified in previous studies. The first approach involves calculating the percentage of points earned out of the total possible points [16,44], while the second method categorizes credits by whether their scores are above or below the average [21,22]. This study opts for the latter approach, driven by three key considerations. Firstly, the assessment of some credits in the standards is inherently binary: either full points are awarded if the requirements are met, or zero points if they are not. This evaluation method is very similar to the second approach. Secondly, an analysis of actual scores from GB cases shows that for credits with multiple potential points (e.g., 0, 3, 6, 9), the points tend to cluster around two values (for example, 6 or 9 in about 200 cases). Lastly, scoring above the average suggests not only the selection of a particular credit by a project but also its commendable performance in that criterion, providing meaningful insights for new projects. Additionally, the binary output from the ML model offers a clear recommendation on whether to select a particular credit. If the model output is 1, it is recommended to select the credit; otherwise, it is not recommended to select the credit. Therefore, this methodological choice is considered to align more closely with our analytical objectives.

The category distribution of awarded and non-awarded credits related to innovation significantly deviates from balance, as shown in Fig. 3(a). Data class imbalance may interfere with the classification results of the model, leading to poor generalization ability [45]. To redress the class imbalance challenge, this study employed the Synthetic Minority Oversampling Technique (SMOTE) proposed by Chawla et al., in 2002, one of the most popular and influential oversampling algorithms [46]. The key idea of SMOTE is to introduce synthetic examples by interpolating between neighboring minority classes to foster balance, rather than simply copying the few class instances [46,47]. By rebalancing the data distribution, SMOTE enhances model generalization. In this study, the five nearest neighbors were employed for generating synthetic instances. Notably, due to a dearth of samples (even less than five samples), credit 11.2.7 and 11.2.8 were excluded. The class distribution before and after SMOTE augmentation for the remaining 3 BCs is presented in Fig. 3.

As for the architectural firm factor, it is important to find a way to assess its level. The magazine "Design New Wave," sponsored by the Shanghai Academy of Social Sciences, annually compiles a list of the top 100 Chinese civil building design firms, utilizing various indicators such as building design area and enterprise scale. Given that the majority of the GB cases analyzed were developed around 2017, according to its 2017–2018 ranking [48], architectural design firms ranked within the top 100 are marked with a 1, while those outside this range are given a mark of 0. Additionally, climate data were sourced from the China Meteorological Data Network (https://data.cma.cn/) and the National Centers for Environmental Information (NCEI) at the National Oceanic and Atmospheric Administration (NOAA) (https://www.ncei.noaa.gov/data/) and economic data were derived from the statistical yearbooks of China and its cities, with all data averaged over the years 2015–2017.

3.3. Machine learning model construction

Based on the established GB project database, we use XGBoost to create ML models. 80 % of the samples are used as the training set, and the remaining 20 % as the test set. The input to the models consists of 4 types of features, and the output is the GB credit selection strategy. For RCs, the model's output example would be: [1,0,0,1 … 0], with each digit corresponding to each credit, where 1 indicates the recommendation to select that credit. The performance of ML models is measured using precision, recall and F1 score. Precision gauges the accuracy of a model's positive predictions, while recall assesses its capacity to capture all pertinent instances in the dataset. The F1 score serves as a balanced measure that optimizes the trade-off between precision and recall. We also use the area under the curve (AUC) metric, commonly used in imbalanced binary classification problems [49], to compare performance changes before and after oversampling. The calculation AUC is shown in the following Equation (4), Equation (5) and Fig. 4. The receiver operating characteristic curve (ROC curve) is plotted by the True Positive Rate (TPR) and False Positive Rate (FPR), using classification thresholds ranging from 0 to 1. Each point on the ROC curve represents the classifier's performance at a different threshold. The program is executed in python 3.9 and MultiOutputClassifier is employed to wrap XGBoost.

$$\textit{True positive rate} = \frac{TP}{TP + FN} \tag{4}$$

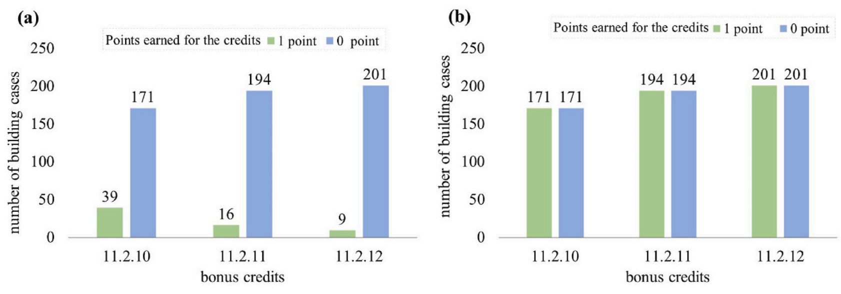

Fig. 3. Comparison of Bonus Credit Points Distribution Before and After SMOTE The vertical axis indicates the number of building cases that earned 0 or 1 point for the credit. (a) and (b) respectively show the scoring situation before and after applying SMOTE oversampling.

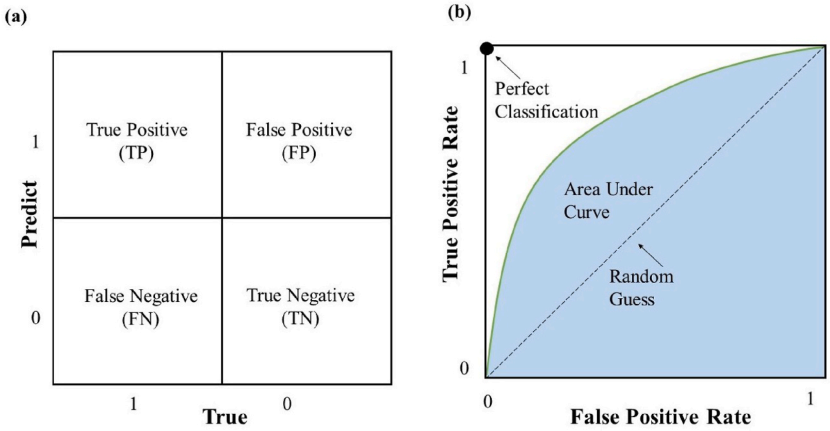

Fig. 4. Schematic representation of AUC calculation.

$$False\ positive\ rate = \frac{TN}{TN + FP} \tag{5}$$

To evaluate the performance of the selected algorithm, we compared several commonly used machine learning algorithms—Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and XGBoost—focusing on precision, recall and runtime. As shown in Table 3, XGBoost outperformed the others, achieving the highest prediction precision and relatively short runtime, demonstrating excellent overall performance.

To identify the key factors affecting the selection and performance of various credits, ML models were constructed separately for each credit. The models' hyperparameters were adjusted using Bayesian optimization. SHAP analysis was conducted on models with superior performance. Good predictive performance demonstrates that the model can effectively learn from the input features and accurately capture patterns within the data, leading to reliable classification results. Further analysis can yield more effective results.

4. Results and discussion

4.1. Machine learning results

As shown in Table 4 for the 42 RCs, the prediction performance of the ML decision support model reaches over 0.8 in terms of precision, recall, and F1 score, and over 0.9 for the three BCs, which shows high reliability. For the 3 BCs, the AUC values before and after oversampling, as illustrated in Table 5, show improvement across the board after SMOTE processing, with AUC values exceeding 0.9. This demonstrates the effectiveness of SMOTE processing. Overall, the ML models demonstrate good performance in supporting credit selection.

4.2. Discussion of influencing factors

This section focuses on the factors affecting each credit and discusses their impact. Table 6 presents the credits corresponding to ML models with an AUC greater than 0.85 and the precision exceeding 0.8, and further SHAP analysis reveals the following findings.

(1) Lower Performance in Light Pollution Credits in Sunny, Cooler Regions.

Credit 4.2.4 focuses on light pollution stemming from nocturnal lighting. According to Fig. 5, the important factors affecting this credit are sunshine and temperature (especially summer temperature), and the performance of this credit is poor in areas with long light hours and low summer temperatures. Fig. 5(b) and (c) offer a clear visual insight. Non-parametric tests were also performed to demonstrate statistical differences in the points across conditions. The Shapiro-Wilk test indicated that the credit data did not conform to a normal distribution. Consequently, the Kruskal-Wallis one-way ANOVA was employed to assess the difference in scores under the different conditions depicted in Fig. 5(b) and (c). The null hypothesis posits that there is no significant difference in credit performance across various conditions. A p-value below 0.05 would lead us to reject this hypothesis. Following the non-parametric analysis, significant statistical differences were observed in credit performance under varying conditions of light ($p = 0.0000$) and temperature (p

Table 3
Performance comparison of different machine learning algorithms.

Algorithm	Final indicators	Training Time (s)	Prediction Time (s)	Precision	Recall
RF	n_estimators = 100	1.4918	1.4660	0.8218	0.8541
XGBoost	n_estimators = 100	4.6973	0.8474	0.8238	0.8576
LR	max_iter = 1000	1387.6773	0.2022	0.6936	0.6742
SVM	kernel = 'rbf'	3.9870	1.0708	0.7628	0.7921

Table 4
Performance of the model supporting selection of regular credits and bonus credits.

Model	Precision	Recall	F1 Score
ML model for RCs	0.8238	0.8576	0.8403
ML model for BCs	0.9675	0.9520	0.9597

Table 5
Performance of the model for BCs before and after oversampling.

Credit	AUC		Percentage increase
	Before SMOTE	After SMOTE	
Bonus credit 1#	0.5938	0.9367	57.75 %
Bonus credit 2#	0.8947	0.9888	10.52 %
Bonus credit 3#	0.7436	0.9933	33.58 %

Table 6
Credits to be Discussed.

Credit	Content	Category
4.2.4	Architectural and lighting design avoid creating light pollution	Light pollution
4.2.5	The environmental noise within the site complies with current standards.	Noise control
8.2.3	Measures are taken to reduce noise interference.	
4.2.6	The wind environment on the site is conducive to comfortable outdoor walking and activities, as well as natural ventilation for buildings.	Building ventilation
5.2.2	The operable sections of exterior windows and glass curtain walls enable buildings to achieve good ventilation.	
8.2.10	Optimize the architectural space, layout, and structural design to enhance the effect of natural ventilation.	
4.2.15	Select greening methods rationally and scientifically arrange the landscaping plants.	Architectural greening
8.2.8	Adopt adjustable shading measures to reduce heat gain from solar radiation in the summer.	Adjustable sunshade
6.2.11	Cooling water makeup uses non-traditional water sources.	Water conservation
4.2.8	The site has convenient connections to public transportation facilities.	Convenience of public transportation

= 0.0000).

The introduction of artificial light at night to change the level of natural light can create light pollution [50] With the acceleration of urbanization, billboard lights, spotlights, streetlights, and car lights may have become luminous polluters [51] In recent years, many large cities in China have been focusing on the construction of nightscape. Urban centers have progressively embraced nocturnal illumination as a testament to economic prosperity, thus fostering prosperous nightscape environments. Notably, some studies have shown that the lighting of public space has a great impact on people's sense of security and comfort [52]. For instance, research by Svechkina et al., in 2020 [53] explored the link between lighting in public spaces and residents' sense of security, particularly highlighting that desert city residents require heightened illumination for perceived safety. This difference may be due to the fact that residents are accustomed to stronger daylight, and therefore may prefer stronger nighttime lighting. This explains why cities with long hours of sunshine are not too keen on controlling light pollution at night. Moreover, areas with high temperatures, especially in the summer months, got high points for credit 4.2.4. Thermal comfort is an important factor influencing outdoor space utilization [54], with overheated weather reducing residents' outdoor activities [55]. Due to the long summer season in most parts of China, residents prefer to spend time in cool outdoor thermal environments [56], such as evenings. Cities with lower summer temperatures tend to witness heightened nocturnal public space activities, thus leading to stronger public space lighting need.

(2) Residences Favor Comfort-Oriented Credits

Credits 4.2.5 and 8.2.3 focus on noise control. As shown in Fig. 6, building type is the most significant influencing factor. Residential buildings received more points for both credits, and there was a statistically significant difference in the performance of the credits for different types of buildings ($p = 0.0000$ for both credits).

A similar pattern emerges for credits related to natural ventilation (Credits 5.2.2, 4.2.6, and 8.2.10) and adjustable shading (Credit 8.2.8). SHAP analysis indicates that residential buildings positively impact credit performance, as shown in Fig. 7 (a) and 7(d), while commercial offices negatively influence it, as shown in Fig. 7 (b) and 7(c). Simultaneously, it is noteworthy that their content, such as noise control, natural ventilation, and shading, is closely related to resident comfort [57,58]. Residential buildings exhibit a positive preference for these credits, possibly because the occupants of residential buildings prioritize these comfort conditions. These credits can enhance residential comfort, thereby increasing the willingness of potential home buyers.

The greening of a city can enhance the well-being of its residents. Credit 4.2.15 specifies the rate and type of greening of buildings. As shown in Fig. 8, residential buildings are clearly more focused on this credit ($p = 0.0000$). The greenery of a residential area is one of the most important criteria for measuring the quality of the living environment, and developers prefer the perfection of greenery in order to attract potential home buyers. In addition, it's worth noting that specifications of GBES for this credit vary based on building

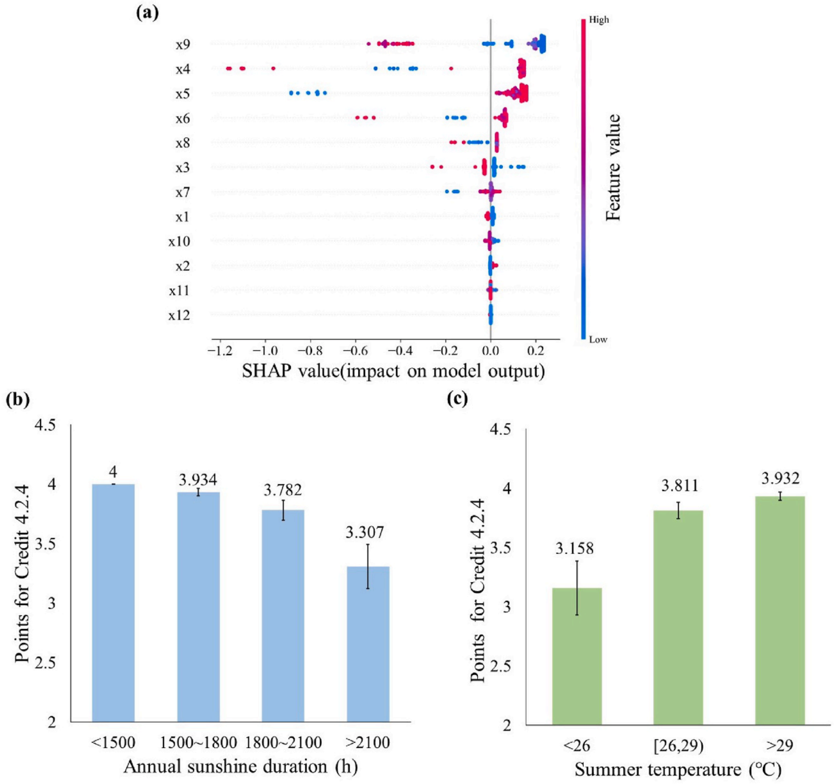

Fig. 5. Credit 4.2.4 Analysis. Figure (a) represents the SHAP analysis for Credit 4.2.4. Figure (b) illustrates the relationship between the points for Credit 4.2.4 and the annual sunshine duration. Figure (c) displays the relationship between the points for Credit 4.2.4 and the summer temperature.

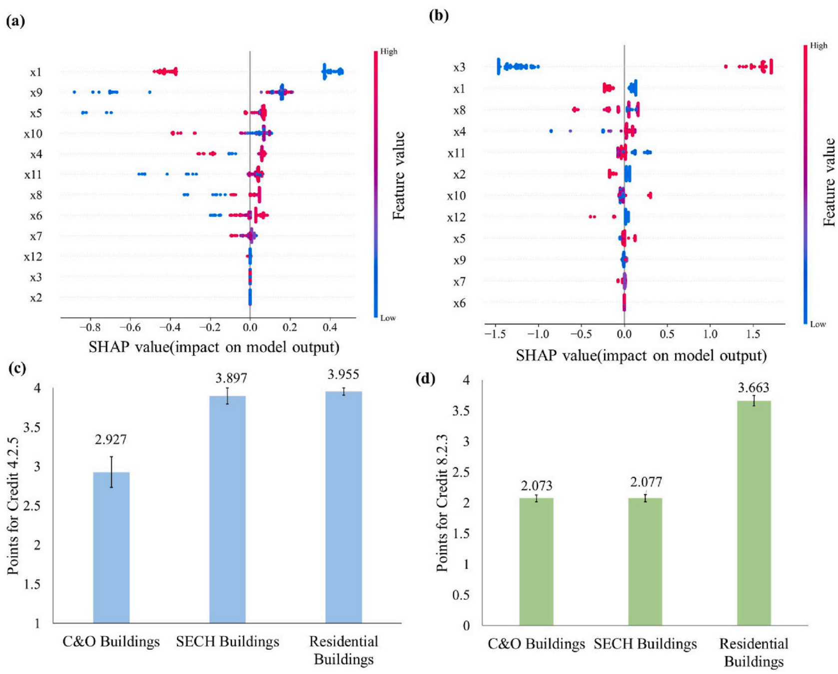

Fig. 6. Analysis of credits related to noise control. Figures (a) and (b) represent the SHAP analysis for Credits 4.2.5 and 8.2.3, respectively. Figures (c) and (d) illustrate the points for Credits 4.2.5 and 8.2.3 for three types of buildings.

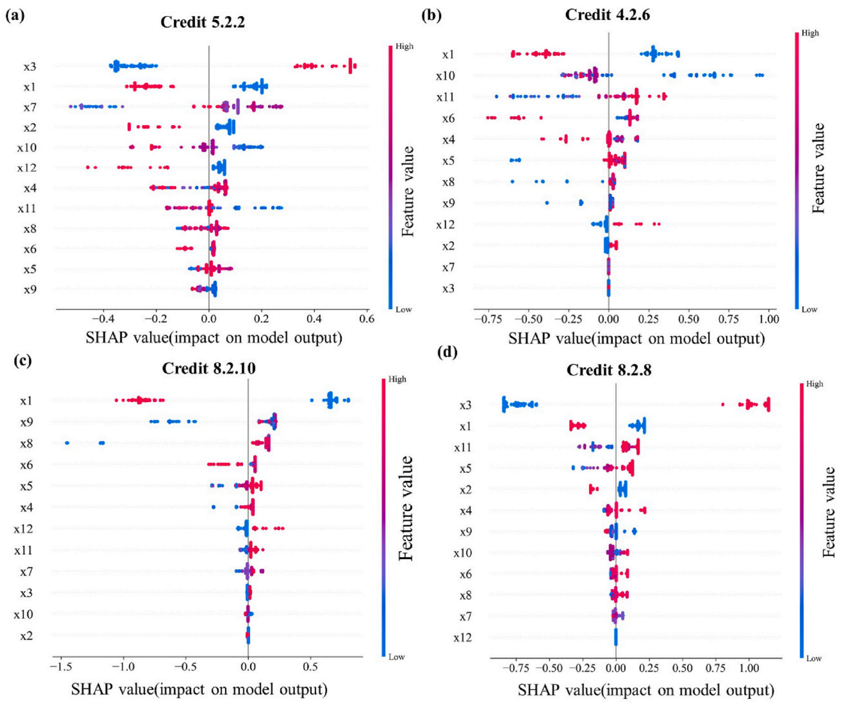

Fig. 7. SHAP Analysis of four credits about natural ventilation and shading. Figures (a), (b), (c), and (d) represent the SHAP analysis for Credits 5.2.2, 4.2.6, 8.2.10, and 8.2.8, respectively.

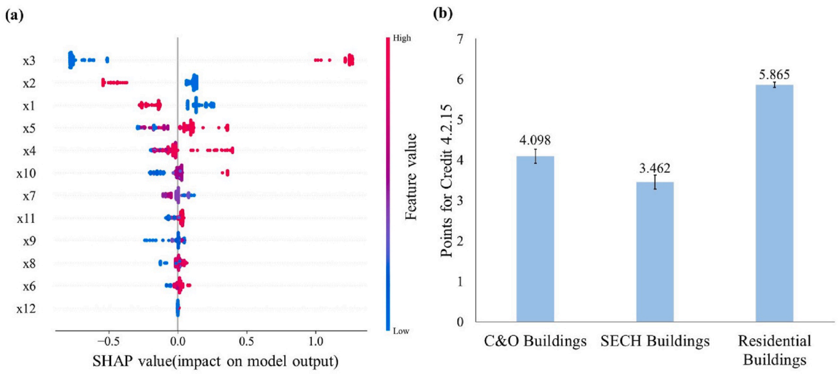

Fig. 8. Analysis of greening-related Credit 4.2.15. Figure (a) depicts the SHAP analysis for Credit 4.2.15. Figure (b) illustrates the credit points for three building types in Credit 4.2.15.

type, with public buildings mandated to adopt green roofs or vertical greening. Green roofs are a viable strategy in densely urbanized areas with land scarcity, particularly for commercial buildings [59] Some studies have shown that there is a great potential for green roofs in Chinese cities [60,61]. However, the low score for public buildings in this credit suggests that the actual promotion of green roofs is limited.

(3) Poor performance of water conservation credits in arid areas

Credit 6.2.11 pertains to the utilization of non-traditional water sources for cooling water systems. While the initial assumption was that precipitation would play a significant role in this credit, the results of the SHAP analysis show that building type (x1) is instead the key factor., as shown in Fig. 9(a). According to the GBES, buildings that lack cooling water systems are automatically awarded full points for this particular credit. This prevalent absence of cooling systems in most residential and SECH buildings could potentially skew the analysis of this credit. To further explore whether buildings equipped with cooling water systems utilize relevant water-saving technologies, a study was conducted on 82 commercial and office buildings. Fig. 9(b) presents the correlation between the scores of these buildings and the level of precipitation. The findings indicate that buildings in regions with higher precipitation tend to score higher than those in areas with lower precipitation. The K-W test results show a p-value of 0.0656. This indicates that at a

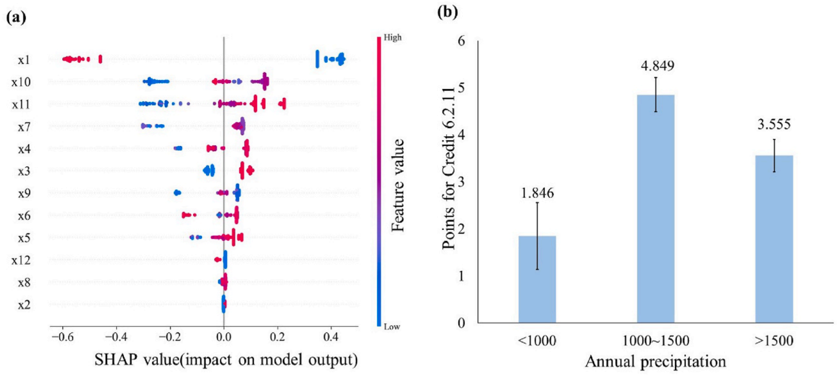

Fig. 9. Credit 6.2.11 Analysis. Figure (a) presents the SHAP analysis. Figure (b) illustrates the relationship between credit points for Credit 6.2.11 and annual precipitation, specifically analyzing data for commercial office buildings.

significance level of 0.05, the differences in scores for C&O buildings under varying precipitation conditions are not statistically significant; however, when the significance level is set to 0.1, these differences become significant, which suggests that buildings in areas with less precipitation perform poorly on this particular credit. Due to the small sample size leading to wider confidence intervals [62], these statistical results are not sufficient to conclusively determine that precipitation is unrelated to the scores. Nevertheless, both results suggest that regions with lower precipitation levels do not exhibit outstanding performance in this credit and may even perform relatively poorly. This could potentially be ascribed to the use of rainwater to supplement cooling water systems, particularly in areas blessed with abundant precipitation. In fact, the Chinese government is concerned about the utilization of non-conventional water sources in water-scarce areas [63]. However, the performance of this credits has not been satisfactory. These areas do not have sufficient rainwater to utilize, but recycled water could be considered in the future. To promote the development of regional priorities based on local characteristics, a priority reward mechanism is suggested. Similar to the regional priority credits set by LEED, this mechanism can encourage architects to focus on local key issues. For example, in areas with water scarcity, implementing water-saving technologies would earn additional points.

(4) The impact of regional economy and building type on the performance of public transportation credits

Credit 4.2.8 focuses on the convenient connectivity between sites and public transportation. As shown in Fig. 10(a), building type and per capita GDP exhibit prominent SHAP values. C&O buildings exhibit a stronger preference for this credit compared to residential buildings. Indeed, the relationship between building type and accessibility to public transportation carries nuanced implications. Commercial buildings, particularly shopping centers and business complexes, are strategically situated in close proximity to public transportation stations. This placement is driven by the desire to draw in a larger customer base, bringing convenience to shoppers and employees alike. However, residential buildings entail additional considerations. Some studies suggest that for certain households, particularly high-income ones, the allure of a quieter, less congested living environment outweighs the convenience of immediate public transportation access [64].

Fig. 10(b) illustrates the relationship between per capita GDP and SHAP values. As per capita GDP increases, credit performance tends to decrease. This trend suggests that highly developed cities might paradoxically face challenges in ensuring efficient public transportation systems. It may be due to the intricate balance between rapid urbanization and infrastructure expansion. Excessive urbanization may have some negative impacts on the accessibility and livability of cities [65]. Consequently, more economically developed cities exhibit reduced preferences for public transportation convenience in building sites. Even when the pursuit of public transportation convenience is prioritized, practical considerations in large cities can be challenging to achieve or require higher costs. Governments should consider the practical challenges associated with improving public transportation accessibility in large cities.

(5) Innovation Credits Predominantly Earned by Urban and Major Design Firms

The three BCs are all about the application of innovative technologies and they all have AUC exceeding 0.85. Further analysis unveils commonalities among these BCs: economic conditions and applicants are key factors. Per capita GDP exhibits high SHAP values in all three BCs. Fig. 11 depicts the relationship between SHAP values and per capita GDP, notably reflecting similar distribution patterns with prominent SHAP values around a per capita GDP of approximately 12,000. Further investigation reveals that these high-SHAP-value green building projects are predominantly situated in Shanghai, indicating a good performance in metropolitan areas. Economic development stands as a significant driver of green innovation [66], and government direct investments often exert a substantial positive influence on innovation activities within the construction industry [67].

While architectural firms typically occupy a lower SHAP value position across RCs, the narrative shifts within credits 11.2.11 and 11.2.12. In these credits, the SHAP values of design firms are higher, as shown in Fig. 12, representing that established and large architectural firms command more BCs. This suggests that stronger architectural companies positively favor innovation-related credits and tend to embrace the challenges of implementing new technologies. Therefore, in addition to the government's economic incentives, the efforts of construction companies are also significant in promoting the adoption of innovative green building technologies

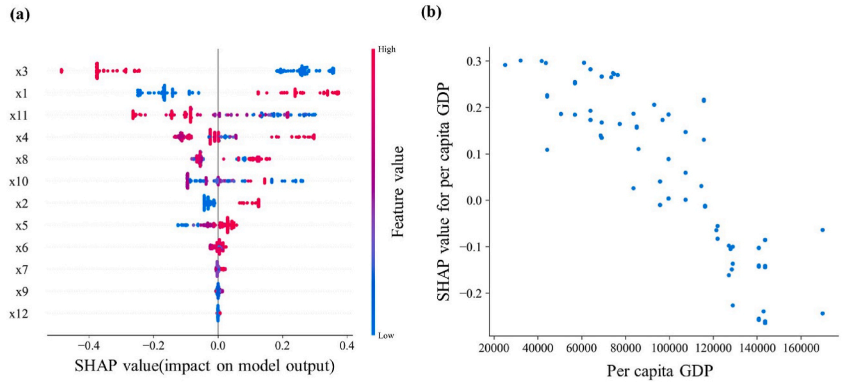

Fig. 10. Credit 4.2.8 Analysis. Figure (a) is the SHAP analysis. Figure (b) depicts the variation of SHAP values with per capita GDP.

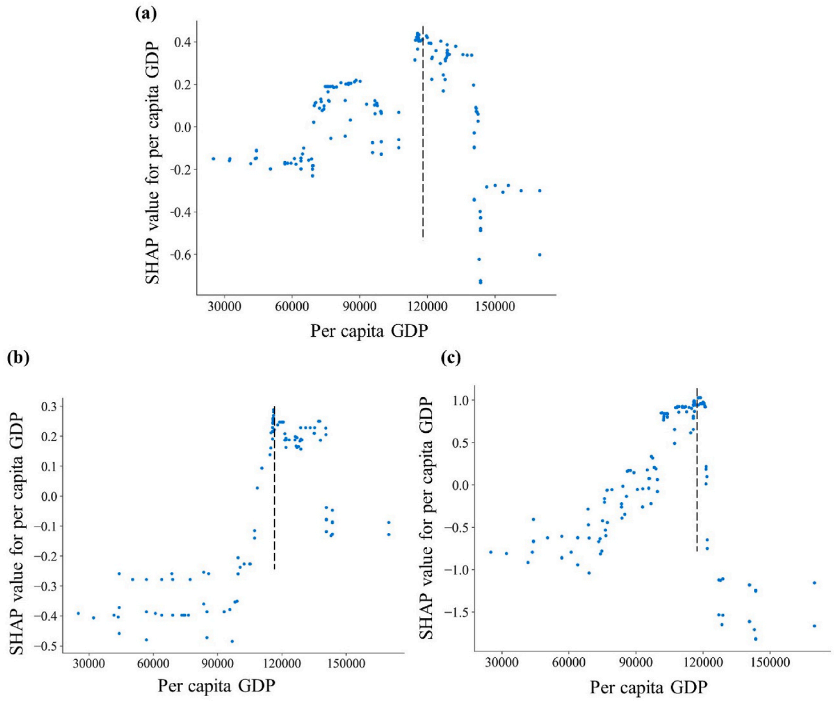

Fig. 11. The relationship between SHAP values of BCs and x11 (per capita GDP). Figures (a), (b), and (c) correspond to credits 11.2.10, 11.2.11, and 11.2.12, respectively.

(GBTs). A “strong leading the weak” approach could be encouraged, where large companies share their innovative experiences and management models to help small and medium-sized enterprises enhance their green innovation capabilities.

5. Conclusion

For GB projects, selecting the appropriate credits is critical yet challenging, influenced by multiple factors. Current research has not provided an adequate solution that considers a variety of influencing factors and multiple credit categories. Additionally, there is limited discussion on the impact of different features on credit selection. This study tackles the issue through the following work. **First**, a database of 210 GB cases is established for training the model. Four types of features–building type, city economy, climate conditions, and architectural company—are identified, and two categories of credits–regular credits and bonus credits—are included. **Second**, an XGBoost-based machine ML model is developed to support the decision-making process for selecting GB credits. The model demonstrates a precision of 82.38 % in selecting RCs. Additionally, the SMOTE method is employed to address class imbalance in innovative credits, resulting in a performance improvement ranging from 10.52 % to 57.75 % in the selection of innovative credits after oversampling. **Third**, a credit selection analysis based on explainable machine learning is conducted. Using SHAP technology to explain model outputs helps to pinpoint crucial factors that influence credit selection and performance, offering insights into how these factors

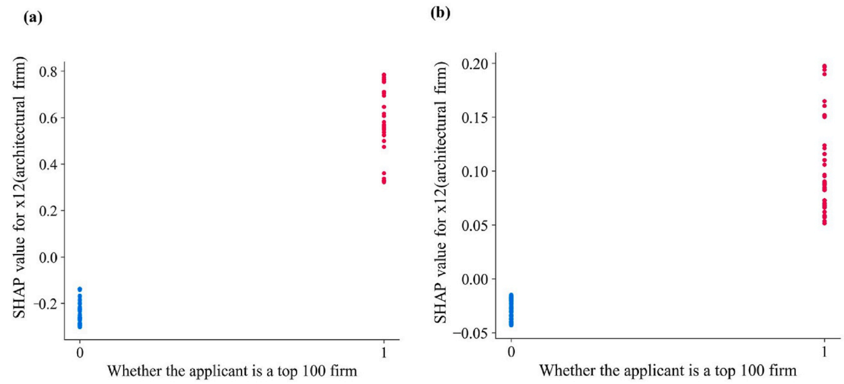

Fig. 12. The relationship between SHAP values of BCs and x12 (applicants). Figures (a) and (b) correspond to credits 11.2.11, and 11.2.12, respectively.

impact credit choices and the underlying reasons. This analysis uncovers a pattern where buildings with certain characteristics demonstrate similar performance on specific credits, suggesting the presence of preferences or avoidance of certain credits. For instance, residential buildings tend to excel in credits related to comfort, while innovative credits are often secured by building projects in major metropolitan areas.

The contributions of this study to green building engineering practice are as follows. **First,** in the early stages of GB design, the decision support model can provide architects with practical credit selection strategies. Without relying on expert knowledge, it helps architects quickly identify suitable credits, reducing the complexity of the credit selection process and enhancing architects' willingness to certify GB projects. **Second,** explainable ML techniques reveal the logic behind model decisions, enabling architects and managers to clearly understand why certain credits are selected. This increases decision transparency, making it easier for various stakeholders to accept the decision outcomes. **Third,** the study reveals the impact of different characteristics on credit selection, providing valuable insights for architects and standard setters. On one hand, architects can make more targeted adjustments to credit selection based on specific needs. On the other hand, standard setters can optimize and update standards accordingly, ensuring they better meet practical requirements. Additionally, this research employs building projects in China as case studies to validate the model's effectiveness and applicability. The findings are particularly relevant for countries and regions that span multiple climate zones. Additionally, the model established in this study has a wide range of applicability. For standard structure, the model is applicable to similar standards composed of prerequisite items, optional credits, and bonus credits. For technical implementation, the model can be trained on other standards using accumulated data and building cases. For result interpretation, the explainable ML model can help offer targeted explanations for new credits.

Moreover, there are some aspects that are worth further discussion. While this study has validated the model's effectiveness using Chinese GB projects, it's important to note that the model and framework are applicable to green buildings across different countries and evaluation standards. Thus, future research could aim to affirm its applicability and effectiveness across a diverse array of international contexts and green building standards. Additionally, with ongoing improvements in statistical data, future discussions can incorporate more quantitative analysis. For instance, guidance could be provided based on the actual environmental benefits associated with each credit.

CRediT authorship contribution statement

Yixin Li: Writing – original draft, Validation, Methodology, Formal analysis, Data curation. **Xiaodong Li:** Writing – review & editing, Supervision, Conceptualization. **Dingyuan Ma:** Writing – review & editing, Writing – original draft, Validation, Methodology. **Wei Gong:** Writing – review & editing, 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

Data will be made available on request.

Acknowledgement

The authors would like to thank the National Natural Science Foundation of China (Grant No. 72071120) and Beijing Building Research Institute Corporation Limited of CSCEC (Grant No. 20232001906) for the financial support for this study.

References

[1] L. Wang, D.W.M. Chan, A. Darko, B.I. Oluleye, A-state-of-the-art review of risk management process of green building projects, J. Build. Eng. 86 (2024), https://doi.org/10.1016/j.jobe.2024.108738.
[2] S. Kaashi, A. Vilventhan, Development of a building information modelling based decision-making framework for green retrofitting of existing buildings, J. Build. Eng. 80 (2023), https://doi.org/10.1016/j.jobe.2023.108128.
[3] J. Zuo, Z.-Y. Zhao, Green building research–current status and future agenda: a review, Renew. Sustain. Energy Rev. 30 (2014) 271–281, https://doi.org/10.1016/j.rser.2013.10.021.
[4] W. Wang, Z. Tian, W. Xi, Y.R. Tan, Y. Deng, The influencing factors of China's green building development: an analysis using RBF-WINGS method, Build. Environ. 188 (2021) 107425, https://doi.org/10.1016/j.buildenv.2020.107425.
[5] S. Seyis, E. Ergen, A decision making support tool for selecting green building certification credits based on project delivery attributes, Build. Environ. 126 (2017) 107–118, https://doi.org/10.1016/j.buildenv.2017.09.028.
[6] Y. Zou, Certifying green buildings in China: LEED vs. 3-star, J. Clean. Prod. 208 (2019) 880–888, https://doi.org/10.1016/j.jclepro.2018.10.204.
[7] S. Altomonte, S. Schiavon, M.G. Kent, G. Brager, Indoor environmental quality and occupant satisfaction in green-certified buildings, Build. Res. Inf. 47 (3) (2019) 255–274, https://doi.org/10.1080/09613218.2018.1383715.
[8] Y. Cao, C. Xu, S.N. Kamaruzzaman, N.M. Aziz, A systematic review of green building development in China: advantages, challenges and future directions, Sustainability 14 (19) (2022) 12293, https://doi.org/10.3390/su141912293.
[9] Z. Wu, H. Li, Y. Feng, X. Luo, Q. Chen, Developing a green building evaluation standard for interior decoration: a case study of China, Build. Environ. 152 (2019) 50–58, https://doi.org/10.1016/j.buildenv.2019.02.010.
[10] Y. Zhang, J. Wang, F. Hu, Y. Wang, Comparison of evaluation standards for green building in China, Britain, United States, Renewable Sustainable Energy Rev. 68 (2017) 262–271, https://doi.org/10.1016/j.rser.2016.09.139.
[11] I.C.S. Illankoon, V.W. Tam, K.N. Le, L. Shen, Key credit criteria among international green building rating tools, J. Clean. Prod. 164 (2017) 209–220, https://doi.org/10.1016/j.jclepro.2017.06.206.
[12] X. Li, W. Feng, X. Liu, Y. Yang, A comparative analysis of green building rating systems in China and the United States, Sustain. Cities Soc. 93 (2023), https://doi.org/10.1016/j.scs.2023.104520.
[13] W.H. Ko, M.G. Kent, S. Schiavon, B. Levitt, G. Betti, A window view quality assessment framework, Leukos 18 (3) (2022) 268–293, https://doi.org/10.1080/15502724.2021.1965889.
[14] Y.-K. Juan, P.-H. Lee, Applying data mining techniques to explore technology adoptions, grades and costs of green building projects, J. Build. Eng. 45 (2022) 103669, https://doi.org/10.1016/j.jobe.2021.103669.
[15] D.H. Pham, B. Kim, J. Lee, Y. Ahn, An investigation of the selection of LEED version 4 credits for sustainable building projects, Appl. Sci. 10 (20) (2020) 7081, https://doi.org/10.3390/app10207081.
[16] P. Wu, Y. Song, W. Shou, H. Chi, H.-Y. Chong, M. Sutrisna, A comprehensive analysis of the credits obtained by LEED 2009 certified green buildings, Renew. Sustain. Energy Rev. 68 (2017) 370–379, https://doi.org/10.1016/j.rser.2016.10.007.
[17] S. Pushkar, Impact of project size on LEED-EB V4 credit achievement in the United States, J. Architect. Eng. 27 (2) (2021) 04021012, https://doi.org/10.1061/(ASCE)AE.1943-5568.0000467.
[18] S.O. Attallah, A. Kandil, A. Senouci, H. Al-Derham, Multicriteria decision-making methodology for credit selection in building sustainability rating systems, J. Architect. Eng. 23 (2) (2017) 04017004, https://doi.org/10.1061/(asce)ae.1943-5568.0000244.
[19] J. Ma, Y. Ding, J.C. Cheng, F. Jiang, Y. Tan, V.J. Gan, Z. Wan, Identification of high impact factors of air quality on a national scale using big data and machine learning techniques, J. Clean. Prod. 244 (2020) 118955, https://doi.org/10.1016/j.jclepro.2019.118955.
[20] D. Ma, X. Li, B. Lin, Y. Zhu, S. Yue, A dynamic intelligent building retrofit decision-making model in response to climate change, Energy Build. 284 (2023), https://doi.org/10.1016/j.enbuild.2023.112832.
[21] M. Jun, J.C. Cheng, Selection of target LEED credits based on project information and climatic factors using data mining techniques, Adv. Eng. Inf. 32 (2017) 224–236, https://doi.org/10.1016/j.aei.2017.03.004.
[22] J.C.P. Cheng, L.J. Ma, A data-driven study of important climate factors on the achievement of LEED-EB credits, Build. Environ. 90 (2015), https://doi.org/10.1016/j.buildenv.2014.11.029.
[23] F.J. Kong, L.H. He, Impacts of supply-sided and demand-sided policies on innovation in green building technologies: a case study of China, J. Clean. Prod. 294 (2021), https://doi.org/10.1016/j.jclepro.2021.126279.
[24] T. Susca, Green roofs to reduce building energy use? A review on key structural factors of green roofs and their effects on urban climate, Build. Environ. (2019) 162, https://doi.org/10.1016/j.buildenv.2019.106273.
[25] F. Kharvari, S. Azimi, W. O'Brien, A comprehensive simulation-based assessment of office building performance adaptability to teleworking scenarios in different Canadian climate zones, Build. Simulat. 15 (6) (2022) 995–1014, https://doi.org/10.1007/s12273-021-0864-x.
[26] P. Engelmann, D. Kalz, G. Salvalai, Cooling concepts for non-residential buildings: a comparison of cooling concepts in different climate zones, Energy Build. 82 (2014) 447–456, https://doi.org/10.1016/j.enbuild.2014.07.011.
[27] S. Mehrdad, N. Reza, B. Kourosh, S. Amiradel, T.M. Reza, M. Reza, Customisation of green buildings assessment tools based on climatic zoning and experts judgement using K-means clustering and fuzzy AHP, Build. Environ. 223 (2022), https://doi.org/10.1016/j.buildenv.2022.109473.
[28] Y. Song, C. Li, L. Zhou, X. Huang, Y. Chen, H. Zhang, Factors affecting green building development at the municipal level: a cross-sectional study in China, Energy Build. 231 (2021), https://doi.org/10.1016/j.enbuild.2020.110560.
[29] Y. Zou, W. Zhao, R. Zhong, The spatial distribution of green buildings in China: regional imbalance, economic fundamentals, and policy incentives, Appl. Geogr. 88 (2017) 38–47, https://doi.org/10.1016/j.apgeog.2017.08.022.
[30] L. Zhang, J. Wu, H. Liu, Policies to enhance the drivers of green housing development in China, Energy Pol. 121 (2018) 225–235, https://doi.org/10.1016/j.enpol.2018.06.029.
[31] W. Zhu, J. Zhang, D. Wang, C. Ma, J. Zhang, P. Chen, Study on the critical factors influencing high-quality development of green buildings for carbon peaking and carbon neutrality goals of China, Sustainability 15 (6) (2023), https://doi.org/10.3390/su15065035.
[32] Y. Han, T. He, R. Chang, R. Xue, Development trend and segmentation of the US green building market: corporate perspective on green contractors and design firms, J. Construct. Eng. Manag. 146 (11) (2020), https://doi.org/10.1061/(asce)co.1943-7862.0001924.
[33] C. Fan, F. Xiao, Z. Li, J. Wang, Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: a review, Energy Build. 159 (2018) 296–308, https://doi.org/10.1016/j.enbuild.2017.11.008.
[34] M.B. Awan, K. Li, Z. Li, Z. Ma, A data driven performance assessment strategy for centralized chiller systems using data mining techniques and domain knowledge, J. Build. Eng. 41 (2021), https://doi.org/10.1016/j.jobe.2021.102751.
[35] S.M.R. Khani, F. Haghighat, K. Panchabikesan, M. Ashouri, Extracting energy-related knowledge from mining occupants' behavioral data in residential buildings, J. Build. Eng. 39 (2021), https://doi.org/10.1016/j.jobe.2021.102319.
[36] D. Ma, X. Li, B. Lin, Y. Zhu, An intelligent retrofit decision-making model for building program planning considering tacit knowledge and multiple objectives, Energy 263 (2023), https://doi.org/10.1016/j.energy.2022.125704.
[37] M. Kent, T. Parkinson, J. Kim, S. Schiavon, A data-driven analysis of occupant workspace dissatisfaction, Build. Environ. 205 (2021) 108270, https://doi.org/10.1016/j.buildenv.2021.108270.
[38] 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, https://doi.org/10.1145/2939672.2939785.
[39] Z. Qiu, J. Wang, B. Yu, L. Liao, J. Li, Identification of passive solar design determinants in office building envelopes in hot and humid climates using data mining techniques, Build. Environ. 196 (2021) 107566, https://doi.org/10.1016/j.buildenv.2020.107566.

[40] H. Mo, H. Sun, J. Liu, S. Wei, Developing window behavior models for residential buildings using XGBoost algorithm, Energy Build. 205 (2019), https://doi.org/10.1016/j.enbuild.2019.109564.
[41] H. Lan, H.C. Hou, Z. Gou, A machine learning led investigation to understand individual difference and the human-environment interactive effect on classroom thermal comfort, Build. Environ. 236 (2023) 110259, https://doi.org/10.1016/j.buildenv.2023.110259.
[42] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems, 2017.
[43] Y. Zhang, B.K. Teoh, M. Wu, J. Chen, L. Zhang, Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence, Energy 262 (2023) 125468, https://doi.org/10.1016/j.energy.2022.125468.
[44] P. Wu, C. Mao, J. Wang, Y. Song, X. Wang, A decade review of the credits obtained by LEED v2.2 certified green building projects, Build. Environ. 102 (2016) 167–178, https://doi.org/10.1016/j.buildenv.2016.03.026.
[45] N. Somu, A. Sriram, A. Kowli, K. Ramamritham, A hybrid deep transfer learning strategy for thermal comfort prediction in buildings, Build. Environ. 204 (2021) 108133, https://doi.org/10.1016/j.buildenv.2021.108133.
[46] A. Fernandez, S. Garcia, F. Herrera, N.V. Chawla, SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary, J. Artif. Intell. Res. 61 (2018) 863–905, https://doi.org/10.1613/jair.1.11192.
[47] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357, https://doi.org/10.1613/jair.953.
[48] Di Editorial Board, di 2017-2018 China Architectural Design Market Rankings, China Commerce and Trade Press, Beijing, 2018 (in Chinese).
[49] R.B. Pereira, A. Plastino, B. Zadrozny, L.H. Merschmann, Correlation analysis of performance measures for multi-label classification, Inf. Process. Manag. 54 (3) (2018) 359–369, https://doi.org/10.1016/j.ipm.2018.01.002.
[50] T. Raap, R. Pinxten, M. Eens, Light pollution disrupts sleep in free-living animals, Sci. Rep. 5 (1) (2015) 13557, https://doi.org/10.1038/srep13557.
[51] F. Bashiri, Light pollution and its effect on the environment, International Journal of Fundamental Physical Sciences (IJFPS) 4 (1) (2014) 8–12, https://doi.org/10.14331/ijfps.2013.330061.
[52] T. Trop, S. Shoshany Tavory, B.A. Portnov, Factors affecting pedestrians' perceptions of safety, comfort, and pleasantness induced by public space lighting: a systematic literature review, Environ. Behav. 55 (1–2) (2023) 3–46, https://doi.org/10.1177/00139165231163550.
[53] A. Svechkina, T. Trop, B.A. Portnov, How much lighting is required to feel safe when walking through the streets at night? Sustainability 12 (8) (2020) 3133, https://doi.org/10.3390/su12083133.
[54] D. Lai, C. Zhou, J. Huang, Y. Jiang, Z. Long, Q. Chen, Outdoor space quality: a field study in an urban residential community in central China, Energy Build. 68 (2014) 713–720, https://doi.org/10.1016/j.enbuild.2013.02.051.
[55] N. Obradovich, J.H. Fowler, Climate change may alter human physical activity patterns, Nat. Human Behav. 1 (5) (2017) 97, https://doi.org/10.1038/s41562-017-0097.
[56] Y. Deng, D. Gan, N. Tang, Z. Cai, X. Li, S. Chen, X. Li, Research on outdoor thermal comfort and activities in residential areas in subtropical China, Atmosphere 13 (9) (2022) 1357, https://doi.org/10.3390/atmos13091357.
[57] S. Vosoughkhosravi, L. Dixon-Grasso, A. Jafari, The impact of LEED certification on energy performance and occupant satisfaction: a case study of residential college buildings, J. Build. Eng. 59 (2022) 105097, https://doi.org/10.1016/j.jobe.2022.105097.
[58] A. Lai, K.W. Mui, L.T. Wong, L. Law, An evaluation model for indoor environmental quality (IEQ) acceptance in residential buildings, Energy Build. 41 (9) (2009) 930–936, https://doi.org/10.1016/j.enbuild.2009.03.016.
[59] N. Xu, J. Luo, J. Zuo, X. Hu, J. Dong, T. Wu, S. Wu, H. Liu, Accurate suitability evaluation of large-scale roof greening based on RS and GIS methods, Sustainability 12 (11) (2020) 4375, https://doi.org/10.3390/su12114375.
[60] W. Liu, Y. Qian, L. Yao, Q. Feng, B.A. Engel, W. Chen, T. Yu, Identifying city-scale potential and priority areas for retrofitting green roofs and assessing their runoff reduction effectiveness in urban functional zones, J. Clean. Prod. 332 (2022) 130064, https://doi.org/10.1016/j.jclepro.2021.130064.
[61] W. Hong, R. Guo, H. Tang, Potential assessment and implementation strategy for roof greening in highly urbanized areas: a case study in Shenzhen, China, Cities 95 (2019) 102468, https://doi.org/10.1016/j.cities.2019.102468.
[62] R.A. Betensky, The p-value requires context, not a threshold, Am. Statistician 73 (2019) 115–117, https://doi.org/10.1080/00031305.2018.1529624.
[63] PRC National Development and Reform Commission, Opinions of the national development and reform commission and other departments on further strengthening the conservation and intensive utilization of water resources. https://www.gov.cn/zhengce/zhengceku/202309/content_6906203.htm, 2023.
[64] W. Seo, H.K. Nam, Trade-off relationship between public transportation accessibility and household economy: analysis of subway access values by housing size, Cities 87 (2019) 247–258, https://doi.org/10.1016/j.cities.2018.11.004.
[65] L. Engelfriet, E. Koomen, The impact of urban form on commuting in large Chinese cities, Transportation 45 (5) (2018) 1269–1295, https://doi.org/10.1007/s11116-017-9762-6.
[66] B. Wang, H. Chen, Y. Ao, F. Liao, Spatiotemporal differentiation and influencing factors of green technology innovation efficiency in the construction industry: a case study of chengdu–chongqing urban agglomeration, Buildings 13 (1) (2022) 73, https://doi.org/10.3390/buildings13010073.
[67] X. Li, Y. Huang, J. Li, X. Liu, J. He, J. Dai, The mechanism of influencing green technology innovation behavior: evidence from Chinese construction enterprises, Buildings 12 (2) (2022) 237, https://doi.org/10.3390/buildings12020237.