Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3, 2508-2530
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i3.5835

Customer perceived value theory and PSO-LightGBM algorithm-based approach to evaluating satisfaction factors with Net-zero energy building retrofits

Qin Xiaoxiang[1,2], Yin Junjia[1,3*], Nuzul Azam Haron[1], Aidi Hizami Alias[1,3], Law Teik Hua[1], Nabilah Abu Bakar[1]
[1]Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, Serdang, 43300, Malaysia; gs64764@student.upm.edu.my (Y.J.).
[2]School of Architectural Engineering, Nanning Vocational and Technical University, Nanning, 530008, China.
[3]UPM-Bentley BIM Advancement Laboratory, Universiti Putra Malaysia, Serdang, 43300, Malaysia.

Abstract: With the growing global focus on sustainability, net-zero energy building (NZEB) retrofitting has become a meaningful way to improve energy efficiency and reduce carbon emissions. However, research on evaluating residents' satisfaction with these retrofit projects is lacking. This study aims to fill this gap using customer perceived value (CPV) theory and PSO-LightGBM algorithms to evaluate the factors influencing satisfaction with NZEB retrofits. The research framework is based on CPV, with functional, emotional, social, and cost values as critical drivers of satisfaction. Post-retrofit feedback was analyzed using PSO-LightGBM and other machine learning (ML) models like CatBoost, XGBoost, and AdaBoost. The study found that "government subsidies and support," "living comfort," "personalized experience," "social acceptance," and "improved environmental image" are the top five most important factors affecting satisfaction with renovating NZEB. In addition, the PSO-LightGBM algorithm excels in accuracy, precision, and F1 score, outperforming other ML models. The study also suggests several enhancement strategies, such as the use of energy-efficient technologies and environmentally friendly materials, to ensure that the performance of the retrofitted buildings improves significantly.

Keywords: *Building retrofit, Customer perceived value, LightGBM, machine learning, Net-zero energy buildings, Particle swarm optimization (PSO).*

1. Introduction

Under the combined pressures of global climate change, resource shortages, and environmental degradation, the issue of energy consumption in the construction industry has become a key focus for governments and academia worldwide. In Europe, for example, building energy consumption accounts for about 40% of total energy consumption and 36% of greenhouse gas emissions [1]. The residential sector is the most important in total energy consumption in almost all countries, with the EU countries, the USA, and China topping the list [2]. Net-Zero Energy Building (NZEB) retrofitting, as an essential means of energy conservation and emission reduction, aims to balance building energy consumption and energy production by improving the energy efficiency of buildings and introducing renewable energy technologies [3]. For instance, China's Ministry of Housing and Urban-Rural Development (MHURD) plans to complete the renovation of 219,000 urban areas during the 14th Five-Year Plan period [4]. Figure 1 shows the number of renovations of old buildings starting in each province in 2023.

History: Received: 20 January 2025; Revised: 7 March 2025; Accepted: 12 March 2025; Published: 28 March 2025
* Correspondence: gs64764@student.upm.edu.my

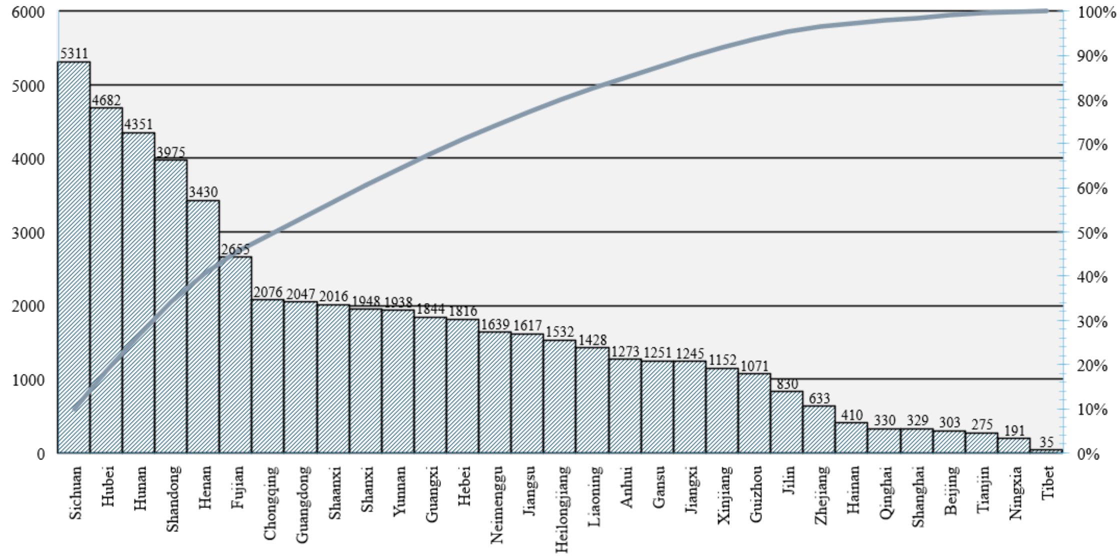

Figure 1.
Number of residential renovations in each province of China in 2023.

Old buildings mostly use HVAC equipment, insulation materials, and lighting technology with low energy efficiency. Therefore, most current research focuses on energy retrofitting measures and technological upgrading programs. Wills, et al. [5] retrofitted a community of 50 older single-family homes from the 1980s. They found that deep retrofits and fuel switching from natural gas to electric heat pump systems could reduce the community's energy demand by 69%. Ohene, et al. [6] analyzed a typical residential building in Ghana through parametric simulation. They found that passive design strategies such as natural ventilation, shading, lighting, and airtightness of the envelope reduced the energy use intensity (EUI) from 136–138 kWh/m²/year to 68–70 kWh/m²/year, thereby reducing total energy demand by 48–50%. Costa, et al. [7] simulated a representative four-story office building in Design Builder and Diva in Rio de Janeiro. They found that adding photovoltaic equipment to the roof and building facade can reduce total energy consumption by 46%. In the past five years, with the rise of artificial intelligence technology, the renovation of zero-energy buildings has not only been limited to the optimization of the external envelope but also considers the application of building-integrated renewable energy, intelligent materials (such as phase change materials in concrete), smart glass, smart buildings and the implementation of the Internet of Things [8]. However, few studies have focused on occupant satisfaction evaluations after NZEB retrofits. User satisfaction affects the social acceptance and long-term use of the project and largely determines the widespread promotion and marketization of such projects [9]. Unlike new buildings, the renovation process of old buildings often involves complex technical, economic, and social issues. Although the technology for renovating zero-energy buildings is becoming increasingly mature, user perceptions, experiences, and expectations of renovation projects vary greatly. For example, one study showed that energy-efficient buildings do not provide satisfactory indoor environmental quality [10]. Some occupants reported difficulties understanding and operating their heating and ventilation systems [11]. In addition, there is a lack of consideration of the thermal satisfaction of occupants in methods based on actual data. Geraldi and Ghisi [12] proposed integrating thermal satisfaction into energy benchmarking. Therefore, understanding the key factors affecting user satisfaction is the basis for driving the success of near-zero energy building retrofits.

Combining the customer perceived value (CPV) theory and the PSO-LightGBM algorithm, this paper constructs a novel analytical framework for evaluating customer satisfaction factors for NZEB retrofits. Unlike previous studies that mainly relied on traditional statistical methods or single-factor analysis, this study systematically classifies customer satisfaction into functional, emotional, social value, and cost value dimensions through CPV theory to comprehensively capture customers' multidimensional perceptions [13] the use of the PSO-LightGBM algorithm to handle nonlinear relationships and high-dimensional data significantly improves the accuracy of model prediction and the explanatory power of feature importance. The functional value measures users' evaluation of the actual energy efficiency and improvement in living comfort after the building renovation; emotional value is related to the sense of pleasure and psychological satisfaction that users feel during their stay; social value reflects users' perception of social identity and personal image; and cost value involves users' financial contribution to the renovation project and the cost of use [14]. These perceived value dimensions affect users' overall satisfaction [15]. Most research is limited to qualitative analysis [16, 17]. It fails to conduct an in-depth quantitative analysis of the perceived value of different dimensions, resulting in an incomplete identification of the critical factors affecting satisfaction.

Furthermore, a series of innovative algorithms and techniques have emerged in machine learning recently, greatly expanding its application scope and performance capability [18]. For example, deep learning, especially in convolutional neural networks (CNN) and recurrent neural networks (RNN), has made revolutionary breakthroughs in image processing, speech recognition, and natural language processing. In addition, reinforcement learning (RL) has been widely used in complex decision-making tasks such as autonomous driving and robot control by continuously optimizing strategies through environmental interaction [19]. Meanwhile, integrated learning methods, such as XGBoost, CatBoost, and LightGBM, have demonstrated powerful performance when dealing with large-scale data, especially in classification and regression tasks, where prediction accuracy is significantly improved by combining the advantages of multiple models [20]. The PSO-LightGBM model proposed in this study further optimizes the hyperparameters of LightGBM with the help of the particle swarm optimization (PSO) algorithm, which is an optimization method that mimics the behavior of groups in nature and exhibits good global search capability in optimization problems. Combining PSO with LightGBM can effectively avoid the local optimal solution problem in traditional model training and improve the prediction accuracy and stability of the model. Compared with conventional machine learning methods, PSO-LightGBM can provide more accurate and robust results when dealing with data with complex features and nonlinear relationships, providing an innovative solution for complex tasks such as building energy efficiency assessment.

The rest of this study is organized: Section 2 describes the factor system, and Section 3 describes the data overview and research methods. Section 4 presents the results of applying multiple ML methods and explains the weighting results. Section 5 presents recommendations for future improvements to the NZEB retrofit. Section 6 summarizes the study and describes its limitations.

2. Literature Review

As indicated in Table 1, We conducted a literature review to construct a multidimensional system of factors based on functional, emotional, social, and cost value. Functional value focuses on the technical performance of the retrofitted building, emotional value pays attention to the user's psychological and emotional identity, social value explores the users' sense of gain in social relationships and identity, and cost value balances the users' economic investment and long-term return. This system of factors provides a scientific basis for studying user satisfaction.

Table 1.
NZEB retrofit satisfaction indicator system.

Types	ID	Factors	Descriptions
Function	F1	Energy Efficiency Improvement	Whether the energy consumption of the retrofitted building is significantly reduced and whether the goal of zero energy consumption can be achieved.
	F2	Indoor Environment Improvement	Whether the retrofitted building's temperature, humidity, ventilation, and air quality significantly improve the living comfort.
	F3	Equipment reliability and ease of use	Whether the smart home system or energy management device introduced is easy to use and less prone to failure.
	F4	Integration of building functions	Whether various intelligent functions such as energy management and temperature control can be integrated into a convenient system.
	F5	Extended building life	Whether the retrofit has improved the durability and safety of the building and whether the building materials and workmanship have helped extend its life.
Emotion	E1	User engagement	Whether the user feels involved in the decision-making process or can express their needs and see their suggestions adopted.
	E2	Living comfort	Whether the design and decoration of the renovated building enhance the user's daily sense of pleasure and psychological comfort.
	E3	Sense of security	Whether the building gives users a stronger sense of security regarding safety, privacy protection, and equipment reliability.
	E4	Psychological identification and satisfaction	Whether users feel a sense of psychological achievement and belonging through participation in a green and sustainable lifestyle.
	E5	Personalized experience	Whether the user's preferences were considered during the renovation process and whether customized functions and designs can be provided.
Society	S1	Social acceptance	Whether users living in zero-energy buildings feel positively evaluated and recognized by society for their environmentally friendly lifestyles.
	S2	Neighborhood interaction and community atmosphere	Whether the renovation has promoted communication and identification among community neighbors and enhanced the overall community atmosphere.
	S3	Improved environmental image	Whether users feel that the retrofitted building positively impacts their personal image and social identity and whether it helps demonstrate their sense of responsibility for environmental protection.
	S4	Media and social feedback	Whether the retrofit project has received positive media coverage and enhanced the user's sense of social recognition.
	S5	Sense of social responsibility	Whether users have enhanced their sense of responsibility and action towards social and environmental issues through participation in renovating zero-energy buildings.
Cost	C1	Reasonableness of conversion costs	Whether the retrofitting costs paid by the user are within their budget and whether they meet market expectations.
	C2	Long-term energy cost savings	Whether the retrofit can reduce energy costs and alleviate the family's long-term financial burden.
	C3	Controllable maintenance costs	Whether the equipment and systems in the building after the renovation are easy to maintain, whether the maintenance costs are low, and whether there are frequent problems.
	C4	Government subsidies and support	Whether the user can enjoy government subsidies, tax incentives, or other incentives to reduce the economic burden.
	C5	Return on investment period	Whether the user's investment in zero-energy retrofits will be paid back within a reasonable period through energy cost savings or added value.

Source: The indicator system refers to studies [20-25].

3. Methods

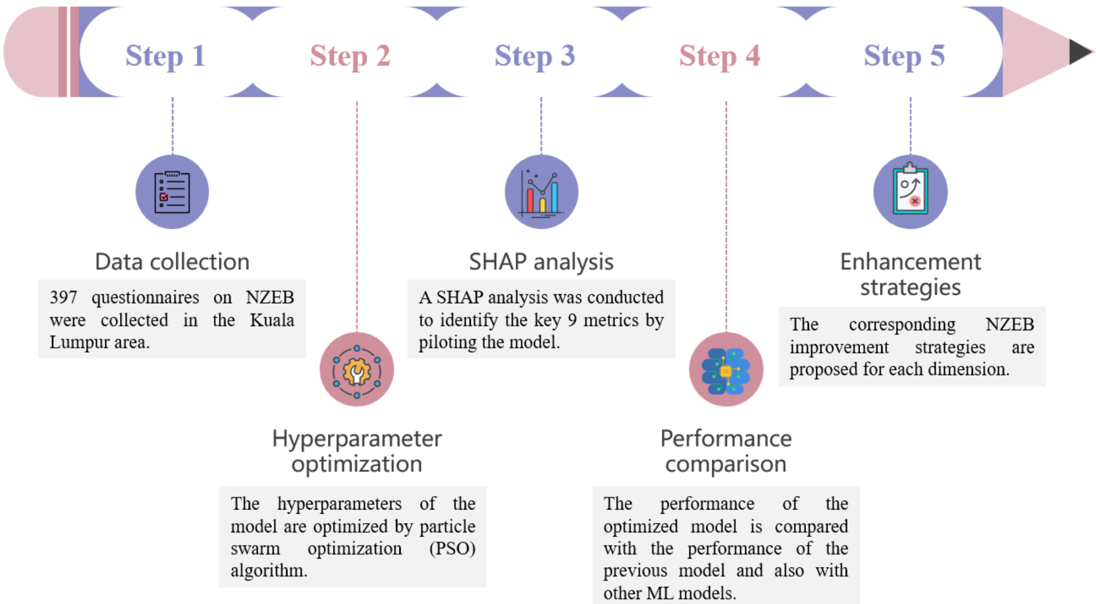

Figure 2.
Research process.

As illustrated in Figure 2, this study is structured into five systematic steps to ensure a comprehensive approach. Step 1 involves collecting questionnaire data on NZEB retrofits within the Kuala Lumpur region, laying the foundation for subsequent analysis. Step 2 focuses on enhancing the LightGBM model by integrating Particle Swarm Optimization (PSO), aiming to improve its prediction accuracy and efficiency. In Step 3, SHAP analysis is employed to filter and streamline the features, enabling the identification of key contributors to NZEB retrofits. Step 4 entails a comparative evaluation, where the improved PSO-LightGBM model is analyzed alongside the original LightGBM and other machine learning models, offering insights into its relative performance. Finally, Step 5 proposes enhancement strategies, leveraging the research findings to provide actionable recommendations for optimizing NZEB transformations.

3.1. Questionnaire Data Collection

We selected Malaysia's Kuala Lumpur area (See Figure 3) to conduct a satisfaction study on the retrofit of zero-energy buildings due to the accelerated urbanization process, growing energy demand, and government policy support for sustainable development in the region. The construction industry in Malaysia consumes 14.3% of total energy, with the residential and commercial sectors consuming 53% of electricity [26]. As Southeast Asia's economic and cultural center, Kuala Lumpur faces enormous challenges and opportunities in building a green city and low-carbon economy. Its climatic conditions, population density, and diverse building types make it an ideal testing ground for studying NEZB retrofit' effects [27]. At the same time, the Malaysian government has actively promoted green building policies in recent years, providing a policy and market foundation for NEZB retrofits, which provides rich data support for studying user satisfaction and the application effects [28].

Figure 3.
Location map of Kuala Lumpur.

Data collection began on 1 March 2024. It is worth noting that the questionnaire for this study has been granted a waiver of ethical review from the affiliation (see Supplementary file 1), as it does not involve sensitive information such as invasive experiments, medical investigations, and commercial confidentiality. We confirm that informed consent was obtained from all respondents. The sample population was selected to include a representative sample of tenants and homeowners directly connected to building environmental issues. Refer to previous research [29, 30] the sample population was chosen from tenants and homeowners directly related to the NEZB renovation issue. It contained representative samples of different ages, genders, occupations, and educational backgrounds to ensure the diversity and representativeness of the data. A random sampling method randomly selected samples from different NEZB retrofit projects. A standardized questionnaire instrument, the 5-point Likert scale, was used to measure participants' importance ratings of the factors. The above factors were listed in the questionnaire, and the respondents were asked to rate each factor using a range of scores (e.g., 1 to 5), indicating its importance to satisfaction. Survey stations were set up at suitable locations, such as centers and entrances of the retrofitted dwellings, and participants were invited to complete the questionnaire through online survey platforms (SurveyMonkey®) to reach a broader range of respondents. All surveys were conducted with the consent of the participants and to ensure the confidentiality of their personal information and response results.

450 questionnaires were sent out, and 397 were returned for a recovery rate of 88.22%. As indicated in Table 2, of those who participated in the survey, 52.39% were male, and 47.61% were female. The vintage of the renovated buildings was mainly 10-20 years, 20-30 years, and 30-40 years, accounting for 26.70%, 29.47%, and 23.68%, respectively. Most people who live in remodeled buildings are over 10 years old but not over 40 years old. The types of retrofitted buildings inhabited by residents are mainly low-rise and multi-story buildings, with proportions of 34.76% and 28.97%, respectively. Low-rise buildings are the primary type of retrofitted building where residents live. The steelwork structure has the lowest selection proportion of 15.08% among the structure types. In comparison, brick-concrete and reinforced-concrete structures have similar selection proportions of 33.52% and 29.97%, respectively. The length of residence after retrofitting was mainly concentrated in the stages of 1-2 and 2-3 years, accounting for 37.28% and 25.94%, respectively, while the proportion of the length of residence within less than 1 year and more than 3 years was relatively low, at 22.17% and 14.61%, respectively.

ISSN: 2576-8484
Vol. 9, No. 3: 2508-2530, 2025
DOI: 10.55214/25768484.v9i3.5835

Table 2.
Profile of the sample population.

Classification	Items	Number	Percentage
Gender	Male	209	52.39%
	Female	189	47.61%
Age of renovated building	10-20	106	26.70%
	20-30	117	29.47%
	30-40	94	23.68%
	Over 40	80	20.15%
Building type	Low-rise	138	34.76%
	Multi-story	115	28.97%
	Middle-rise	81	20.40%
	High-rise	63	15.87%
Structure type	Brick-wood	77	19.40%
	Brick-concrete	141	33.52%
	Reinforced-concrete	119	29.97%
	Steelwork	60	15.11%
Length of residence	0-1	148	37.28%
	1-2	103	25.94%
	2-3	88	22.17%
	Over 3	58	14.61%

The reliability analysis was conducted using SPSS27® software, and the results showed that the Cronbach.α coefficient was 0.968 [31]. This coefficient is greater than the standard value of 0.80, which is usually considered excellent internal consistency, indicating that the questionnaire has a high level of reliability and can reflect the characteristics of the concepts being tested stably. The KMO value in the validity analysis of the questionnaire was 0.983, which is much greater than the generally accepted value of 0.8 or greater, indicating that the sample has reasonable sampling reasonableness [32] the value of Butterball (approximately chi-square) was 6467.919, the df (degrees of freedom) was 210, and the p-value was 0.000 (less than 0.05), indicating that the data are suitable for factor analysis.

3.2. LightGBM Algorithm

LightGBM is an efficient machine learning algorithm based on gradient-boosted decision trees (GBDT), which is mainly used to handle learning tasks with large-scale data and high-dimensional features [33]. It uses a series of optimization strategies to make training faster, use less memory, and be able to handle large amounts of data and features. The algorithm steps are as follows [34]: (1) initialization, constructing an initial learner (tree) as the base model; (2) iterative training, constructing more learners in turn through iteration, each learner trying to correct the errors of the previous learner; (3) gradient optimization, each iteration optimizing the model, so that the loss function on the training set is minimized; (4) leaf node splitting, which gradually generates more complex decision tree structures by selecting the optimal features and split points based on split gain. The leaf-wise algorithm is used to find the leaf with the largest splitting gain from all the current leaves for splitting; (5) boosting learning, which improves the prediction ability of the overall model by accumulating the prediction results of multiple simple models.

The objective function of LightGBM consists of a loss function and a regularization term [35]:

$$\mathrm{Obj} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{j=1}^{T} \Omega(f_j) \quad (1)$$

where l is the loss function, y_i is the true value, $\hat{y}_i$ is the predicted value, Ω is the regularisation term, and T is the number of trees.

LightGBM selects the leaf node with the greatest split gain at each split. The split gain is calculated using the following formula [36]:

$$w_j = -\frac{G_j}{H_j+\lambda} \quad (2)$$

The weight of each leaf node is calculated as follows [37]:

$$w_j = -\frac{G_j}{H_j+\lambda} \quad (3)$$

where G_j is the gradient of the leaf node and H_j is the second derivative of the leaf node, and λ is the regularisation parameter.

4. Results

4.1. Spearman Correlation Analysis

Spearman correlation analysis is used to study the relationship between quantitative data, including whether there is a relationship and the degree of closeness of the relationship [38]. Its arithmetic rule is as in Equation 4. We used SPSS27® software to correlate the relationship between each factor and satisfaction level (SL); the results are shown in Table 3; all the factors strongly correlate with SL, which can be used for the subsequent analysis.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (4)$$

Where, ρ is the Spearman correlation coefficient; d_i is the rank difference for each pair of observations; and n is the number of observations.

Table 3.
Result of Spearman correlation analysis.

ID	Mean	Std.	Correlation coefficient (math.)
F1	3.718	1.270	0.565**
F2	3.688	1.171	0.554**
F3	3.670	1.279	0.565**
F4	3.645	1.201	0.584**
F5	3.715	1.282	0.649**
E1	3.577	1.264	0.600**
E2	3.587	1.274	0.657**
E3	3.642	1.263	0.634**
E4	3.615	1.241	0.671**
E5	3.539	1.276	0.662**
S1	3.685	1.253	0.660**
S2	3.630	1.266	0.631**
S3	3.594	1.275	0.673**
S4	3.615	1.245	0.690**
S5	3.688	1.234	0.674**
C1	3.547	1.244	0.580**
C2	3.678	1.194	0.550**
C3	3.685	1.245	0.643**
C4	3.670	1.271	0.562**
C5	3.773	1.199	0.562**

Note: * $p<0.05$ ** $p<0.01$; Dependent variable is SL.

4.2. Construction of the Initial Prediction Model

Nine ML models such as AdaBoost, CatBoost, XGBoost, RF, MLP, and LightGBM, are constructed, and the experimental environment is Nvidia GeForce MX550 and 12th Gen Intel (R) Core (TM) i5-1235U. The code editor used in this study is Visual Studio Code®, which is widely used to develop and run ML algorithms, supports a wide range of programming languages and has a rich set of

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 2508-2530, 2025
DOI: 10.55214/25768484.v9i3.5835

extension libraries such as Python®, Jupyter Notebook®, and TensorFlow®. The selection of specific ML algorithms in this study is mainly based on their advantages in dealing with nonlinear feature relationships, high-dimensional data, and feature significance interpretation while considering their computational efficiency and applicability.

The model evaluation indicators are selected in Equations 4 to 7 [39]. Accuracy is the proportion of samples with correct prediction results to the total samples; the higher the accuracy, the better. Precision is the proportion of results with optimistic predictions that are positive samples; the higher, the better for this metric. Recall is the proportion of positive samples that are positive. The higher the recall, the better; the F1-score is a comprehensive evaluation index that combines precision and recall; it is the harmonic mean of precision and recall.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

Where, TP stands for True Positives, TN stands for True Negatives, FP stands for False Positives and FN stands for False Negatives.

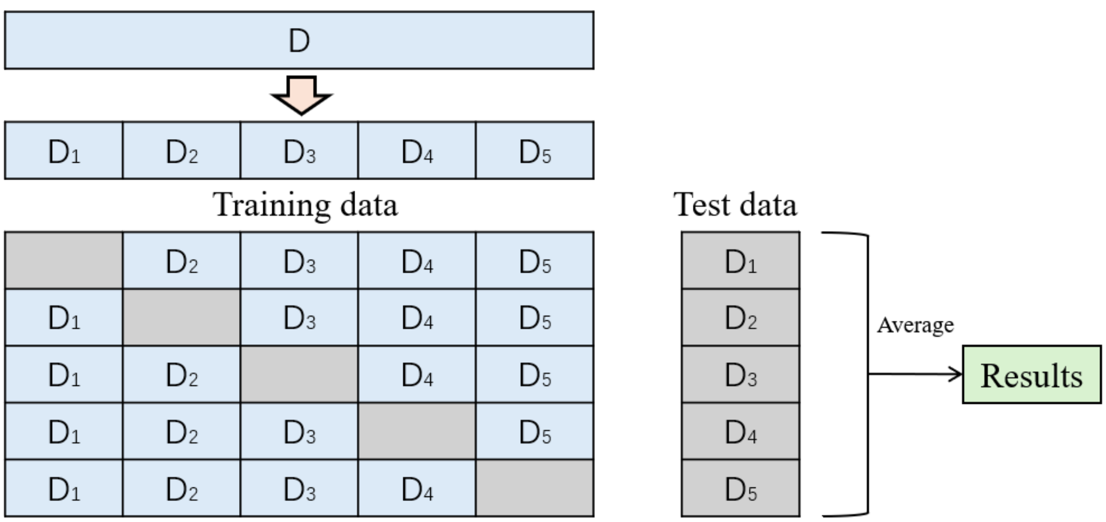

Figure 4.
Five-fold cross-validation.

The optimal hyperparameters of each ML model are obtained (see Figure 4) [40]. In the pre-processing of the ML algorithms, all ML algorithms were "iterative" by selecting similar datasets; then, the hyperparameters were adjusted to obtain the best prediction results for each ML method chosen. Finally, Table 4 illustrates the information about the hyperparameters used to regulate the ML algorithm.

Particle Swarm Optimization (PSO) is an optimization algorithm based on group intelligence, proposed by Marini and Walczak [41]. It simulates the behavior of groups of birds, fish, etc., when

searching for food and searches for optimal solutions through information sharing among individuals. The basic formula is as follows [42]:
(1) Speed update formula

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot \left(p_i^{\text{best}} - x_i(t)\right) + c_2 \cdot r_2 \cdot \left(g^{\text{best}} - x_i(t)\right) \tag{9}$$

(2) Position update formula

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{10}$$

Where, $v_i(t)$ is the velocity of particle i at time t; $x_i(t)$ is the position of particle i at time t; w is the inertia weight; c_1 and c_2 are the acceleration constants; r_1 and r_2 are random numbers between [0, 1]; p_i^{best} is the historical best position of particle i; g^{best} is the global best position.

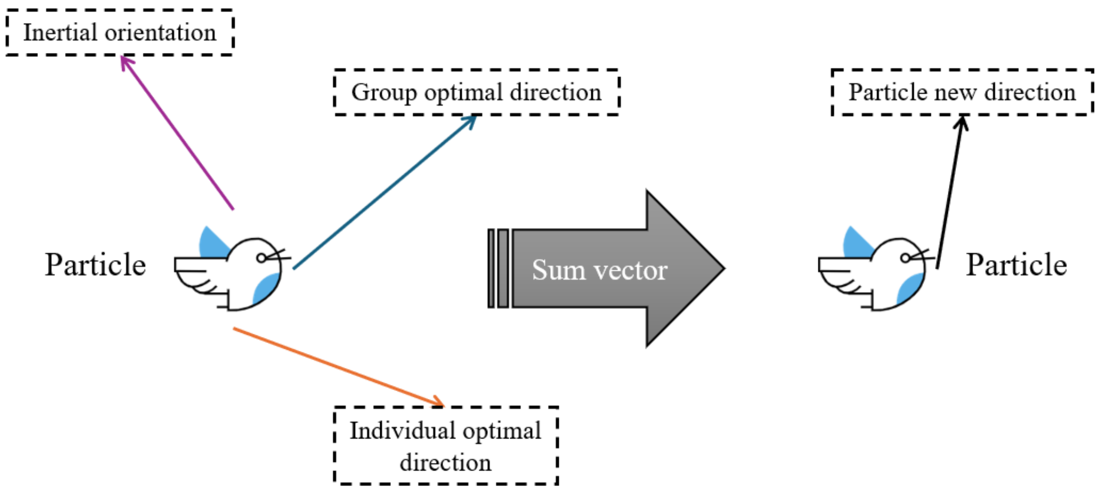

Figure 5.
Particle swarm optimization schematic.

As indicated in Figure 5, it works as follows:
(1). Randomly initialize the position and velocity of the particles.
(2). Calculate the fitness value of each particle.
(3). Update each particle's historical (individual best position).
(4). Update the global (population best) position.
(5). Update the velocity and position of each particle according to the velocity and position update formula.
(6). Stop when the maximum number of iterations is reached, or accuracy requirements are met.

The flock is composed of multiple particles, each representing a potential solution. Each particle has a position and velocity for moving in the search space, and the fitness function evaluates the merit of each particle's position (solution).

ISSN: 2576-8484
Vol. 9, No. 3: 2508-2530, 2025
DOI: 10.55214/25768484.v9i3.5835

Table 4.
Information of the hyper-parameters used for regulating the ML algorithms.

ML	Hyper-parameter	Value	ML	Hyper-parameter	Value
PSO-LightGBM	n_estimators	100	SVM	random_state	2
	learning_rate	0.001		loss	epsilon insensitive
	max_depth	15		tolerance	1e-04
	boosting type	gbdt		max_iter	2000
	number of leaves	32		kernel	rbf
AdaBoost	n_estimators	100	MLP	loss	mse
	learning_rate	0.001		hidden_layer 1	10
	min_samples_leaf	4		hidden_layer 2	10
	min_samples_split	6		hidden_layer 3	2
	max_depth	25		activation	relu
CatBoost	n_estimators	500		optimizer	adam
	learning_rate	0.02		max_iter	2000
	max_depth	16	KNN	n_neighbors	8
	min_samples_leaf	10		weights	distance
XGBoost	n_estimators	3500		neighborhood	global search
	learning_rate	0.01		leaf_size	25
	min_samples_leaf	20		p_value	2
	min_samples_split	10	GBDT	n_estimators	1000
	max_depth	8		learning_rate	0.01
RF	min_samples_leaf	15		min_samples_leaf	10
	min_samples_split	6		min_samples_split	12
	max_depth	25		max_depth	8

Note: The training ratios are all 0.8.

4.3. SHapley Additive exPlanation

SHAP (SHapley Additive exPlanation) is a method for interpreting the predictions of ML models. It is based on Shapley values, an impartial method used in game theory to distribute the benefits of cooperation [43]. The SHAP value formula is as follows [44]:

$$\phi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)] \tag{11}$$

Where, $\phi_i(v)$ is the SHAP value of the feature i; N is the set of all features; S is a subset of N without feature i; $v(S)$ is the model output of the subset of features S.

SHAP values provide each feature with the degree of its contribution to the prediction outcome, thus helping to understand the model's decision-making process. Figure 6 illustrates the SHAP value contribution of different features in the model to the predictions of the five categories (Type 1 – Type 5), thus revealing the importance and directionality of the features. The horizontal axis indicates the cumulative contribution of SHAP values, the colors distinguish different categories, and the length of the bars indicates the total contribution of features to the predictions of each category. Top-ranked features (e.g., S3, S5, E4, etc.) have a more significant impact on the projections of the classification model, with S3 contributing mainly to the type 5 category. In contrast, S5 and E4 contribute more to the type 4 and 3 categories. The total contribution of SHAP values for these features decreases as the ranking decreases (e.g., C1, F4), indicating that they are relatively less critical for model prediction.

According to the weighting results of SHAP analysis, the social dimensions are ranked as S3>S5>S1>S4>S2; the emotional dimensions are ranked as E4>E1>E2>E5>E3; the functional dimensions are ranked as F1>F5>F3>F2>F4; and the cost dimensions are ranked as C4>C2>C5>C3>C1.

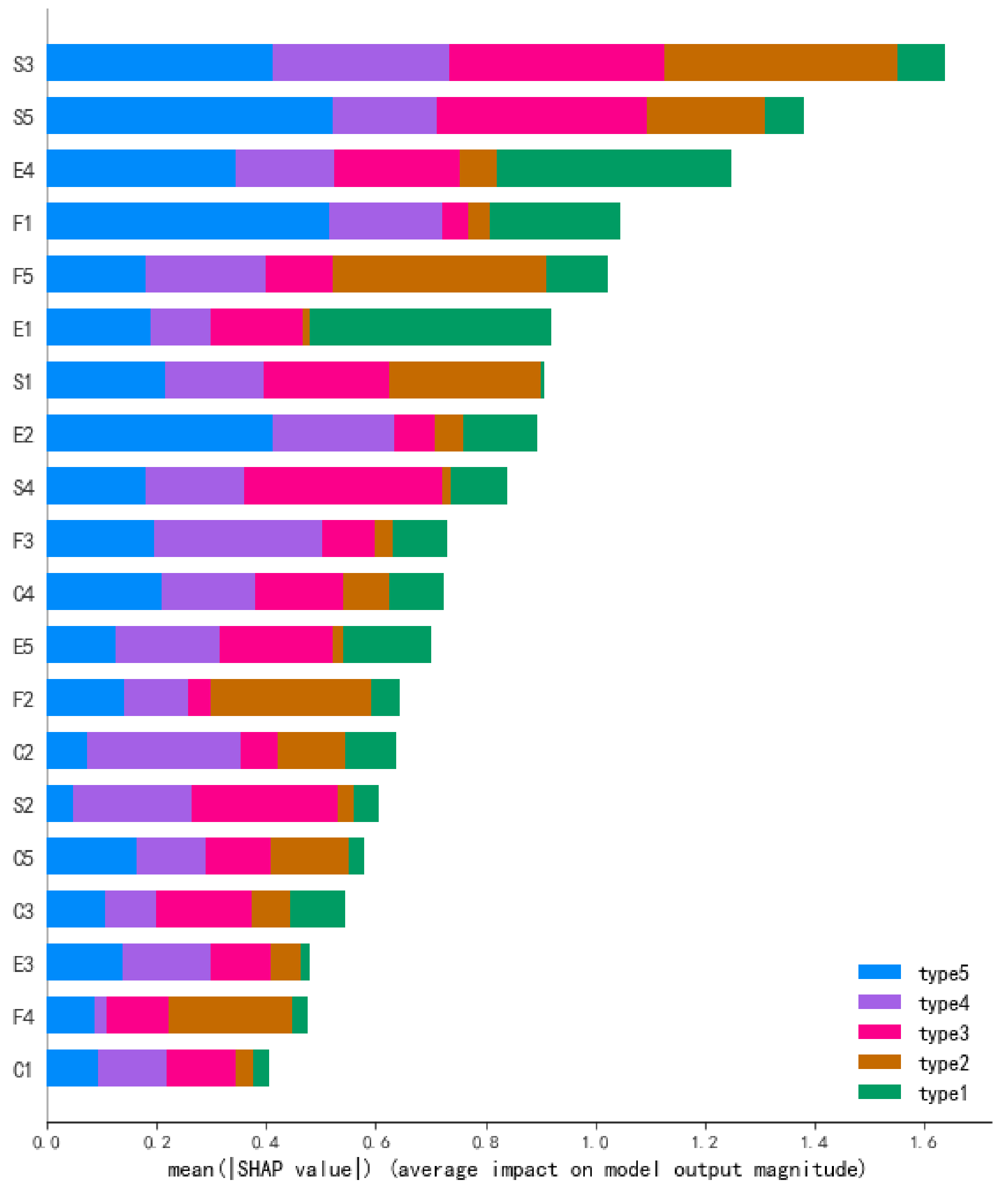

Figure 6.
The weights of the features SHAP values.

Among the social dimensions, improved environmental image (S3) has the highest ranking, indicating that users highly value demonstrating their ecological responsibility and social image through retrofitting, and this external recognition directly affects users' satisfaction. Sense of social responsibility (S5) follows, reflecting the intrinsic drive of users to fulfill their social responsibility by participating in net-zero energy retrofits. Social acceptance (S1) ranks third, indicating that users also care about the overall acceptance of their neighborhood and society. Still, it is slightly less critical because users focus on factors directly relating to their identity and image. In contrast, Media and social feedback (S4) and Neighborhood Interaction (S2) ranked lower because the influence of these factors on

ISSN: 2576-8484
Vol. 9, No. 3: 2508-2530, 2025
DOI: 10.55214/25768484.v9i3.5835

user satisfaction is more dependent on specific socio-cultural and environmental contexts, and their effects are more indirect.

Among the emotion dimensions, psychological identity and satisfaction (E4) ranked first, highlighting the deep emotional needs of users to participate in a green lifestyle and gain a sense of fulfillment. User participation (E1) ranked second, indicating that users want to express their wishes and see tangible results in the remodeling process, reinforcing satisfaction. Living comfort (E2) is third, reflecting a concern for indoor quality of life, but is slightly lower because users consider comfort a basic expectation rather than a core priority in decision-making. Personalized experience (E5) and Sense of security (E3) ranked lower due to their more limited specific impacts, with users' expectations of these factors tending to be more of an icing on the cake than a decisive influence.

Energy efficiency improvement (F1) is the highest ranked of the functional Dimensions, indicating that when evaluating a net-zero energy retrofit, users are most concerned with whether the retrofit significantly reduces energy consumption and achieves sustainability goals. Building life extension (F5) was a close second, reflecting a strong concern about whether the retrofit would improve the overall durability and safety of the building. Facility reliability (F3) comes in third, with users concerned about the stability of retrofitted smart devices, but this is slightly less important than overall building performance. In contrast, indoor environment improvement (F2) and feature integration (F4) ranked lower, as users already have a basic expectation of indoor environment improvement, while the demand for smart feature integration is mainly an added value rather than a core need.

In the cost dimension, government subsidies and support (C4) rank the highest, reflecting that policy incentives are an essential factor in lowering the threshold of user participation, which directly affects their motivation for retrofitting. Long-term energy cost savings (C2) is ranked second, indicating that users are interested in the initial investment and the long-term savings benefits. The payback period (C5) is third, indicating that customers want to see a return on their investment in a reasonable amount of time. Still, it is less important than direct subsidies and cost savings. Controllable maintenance costs (C3) and Reasonable retrofit costs (C1) ranked lower because users view these factors as essential safeguards and are more concerned with returns and support policies in their decision-making. It is worth noting that the reason for the low sensitivity of the population to the value of the cost is that the Malaysian government provides a variety of subsidies and financial incentives to encourage energy-efficient building retrofits. For example, the government has introduced the Green Building Index (GBI) certification program, which provides tax incentives for certified buildings. There are also specific subsidy programs such as the Solaris Photovoltaic Incentive Program, which provides a subsidy of up to RM4,000 per kilowatt. In addition, Malaysia's relatively low cost of energy as an oil-exporting country makes energy efficiency retrofits less financially stressful.

4.4. Reconstruction of the Final Prediction Model

According to the results of the SHAP analysis, we finally selected 9 indices (S3, S5, E4, F1, F5, E1, S1, E2, S4) to be used as input metrics for the final prediction model. SL was used as a dependent variable. Figures 7 and 8 show the model's performance. The initial test model obtained an accuracy of 86.25%, a precision of 84.28%, a recall of 86.25%, and an F1-score of 0.842. The final test model obtained an accuracy of 89.64%, a precision of 93.68%, a recall of 89.64%, and an F1-score of 0.909. The optimization of model parameters by combining PSO improves the training accuracy from 99.36% to 99.68% and the test set accuracy from 86.25% to 91.25%. This result indicates that the combination of feature streamlining and PSO optimization not only improves the model's ability to fit the training data but also significantly enhances its generalization performance on the test data, thus capturing the key factors of NZEB retrofit satisfaction more efficiently.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 3: 2508-2530, 2025
DOI: 10.55214/25768484.v9i3.5835

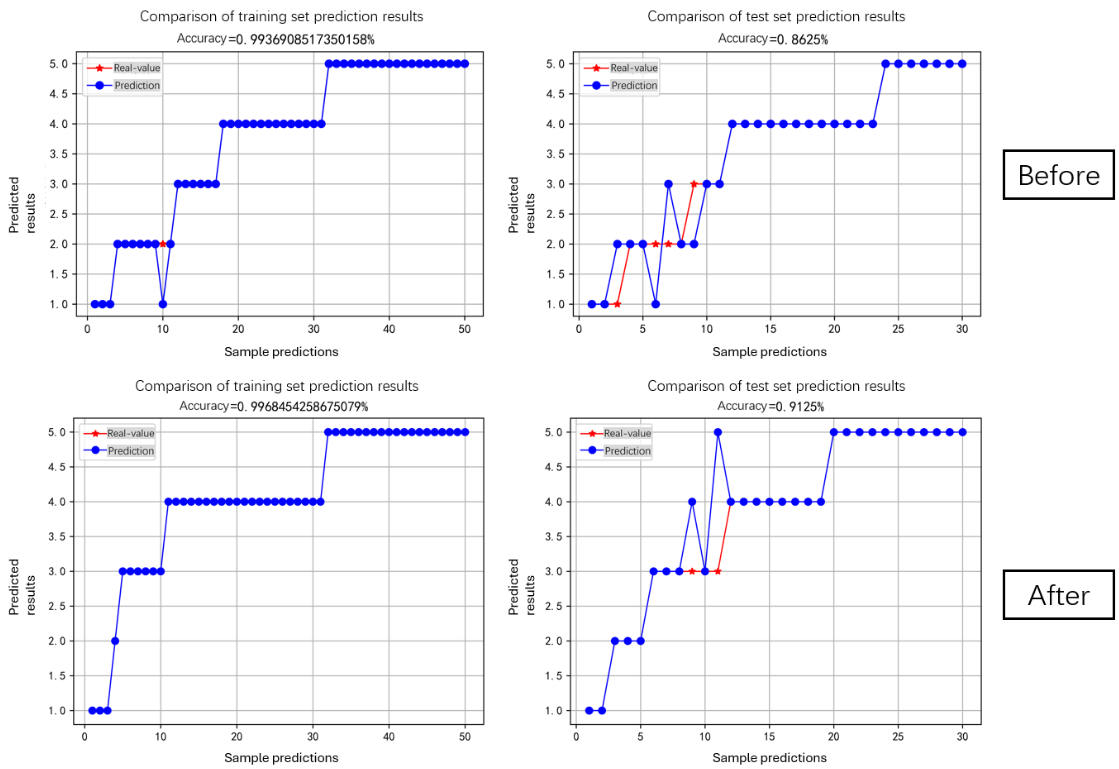

Figure 7.
Accuracy of PSO-LightGBM model before and after tuning.

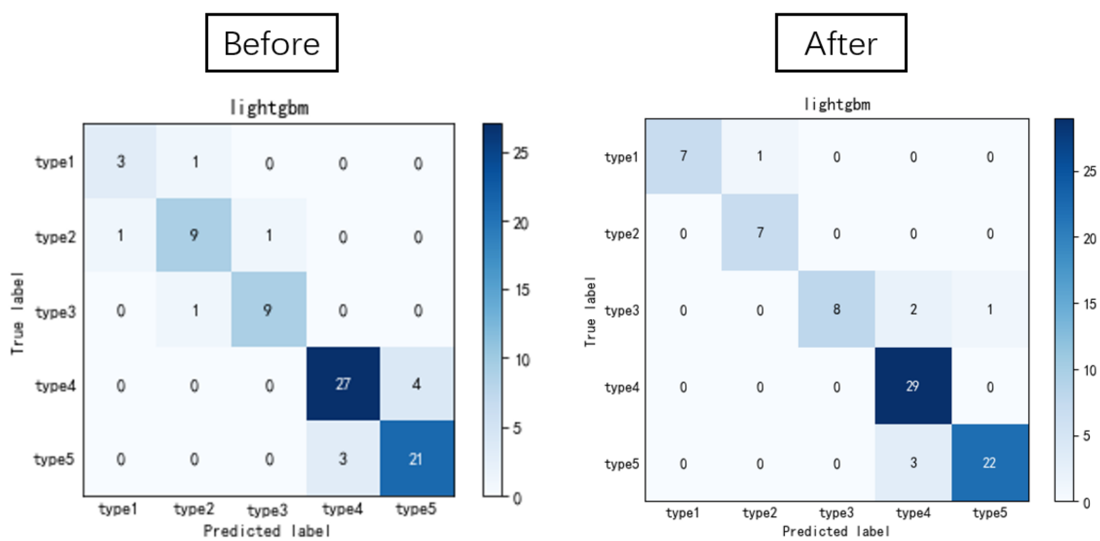

Figure 8.
Confusion matrix of the PSO-LightGBM model before and after tuning.

Figure 9 illustrates that the area's value under the ROC (Receiver Operating Characteristic) curve is a key metric. The area of a perfect classifier is 1, while the area of a random classifier is 0.5 [45].

Therefore, a good model, with the ROC curve closer to the upper left corner, has a higher area value (close to 1). The PR Curve (Precision-Recall Curve) is particularly useful when dealing with unbalanced datasets (e.g., when there are fewer positives than negatives.) The area under the PR Curve is also a key metric. Good models with high PR curves (close to 1) indicate high precision and recall across thresholds [46].

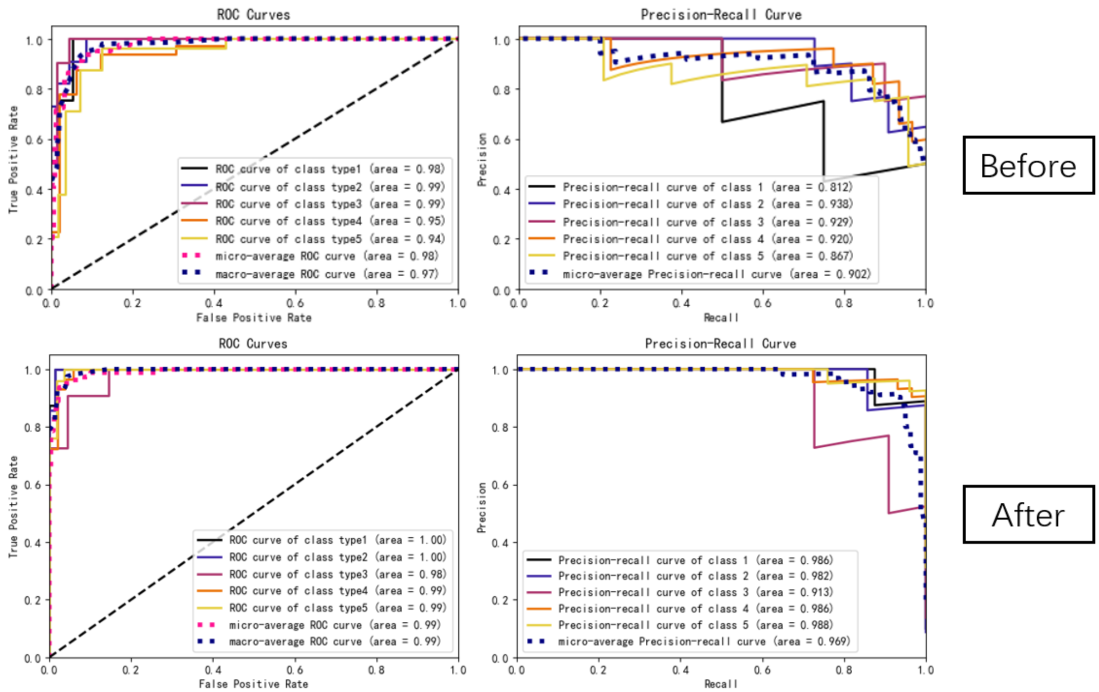

Figure 9.
ROC and Precision-Recall curves before and after PSO-LightGBM model adjustment.

The classification performance is significantly improved by reducing the number of features in the POS-LightGBM model from 20 to 9. Before the improvement, the ROC curves of each category performed better, with AUC values between 0.94 and 0.99, indicating that the model has strong differentiation ability. Still, the PR-AUC of some categories (e.g., Class 1) in the Precision-Recall curves was low at 0.812, which affected the overall prediction effect. After improvement, the ROC curve of the model is further optimized, and the AUC values of all categories are raised to 0.99 or 1.00. The AUC values of micro-averaging and macro-averaging also reach 0.99, close to the perfect classification performance. Meanwhile, the overall performance of the Precision-Recall curve is significantly improved, with the PR-AUC of Class 1 improved from 0.812 to 0.986 and the micro-average PR-AUC improved from 0.902 to 0.969, which fully demonstrates the balanced improvement of the model in precision and recall. This improvement is attributed to applying the feature refinement strategy; by removing redundant features and focusing on key features, the model avoids overfitting while enhancing its prediction ability for low-performing categories. This improves the model's classification accuracy and optimizes the overall stability and efficiency.

As illustrated in Figure 10, the Partial Dependence Plot (PDP) shows the effect of each feature in the model on the prediction of different target categories. The horizontal axis of each plot indicates the range of feature values, and the vertical axis is the predicted value (Partial Dependence), with different colored curves corresponding to distinct categories. By looking at these curves, it is possible to analyze

ISSN: 2576-8484
Vol. 9, No. 3: 2508-2530, 2025
DOI: 10.55214/25768484.v9i3.5835

the marginal contribution of features to the model's predictive output and identify which features are more critical in predicting specific categories. Some features (e.g., F1, S1) significantly impact model predictions, with a significant curve variation. For example, in F1, Class 3 (purple color) rises considerably with increasing feature values, indicating that F1 contributes more to the prediction of Class 3. In contrast, the curves change more gently for features like E4 and S4, suggesting they have a weaker effect on model prediction. In addition, curve crossovers (e.g., red and purple in the S1 plot) indicate that the differentiation between classes is weakened at certain eigenvalues and that there is a class-switching phenomenon.

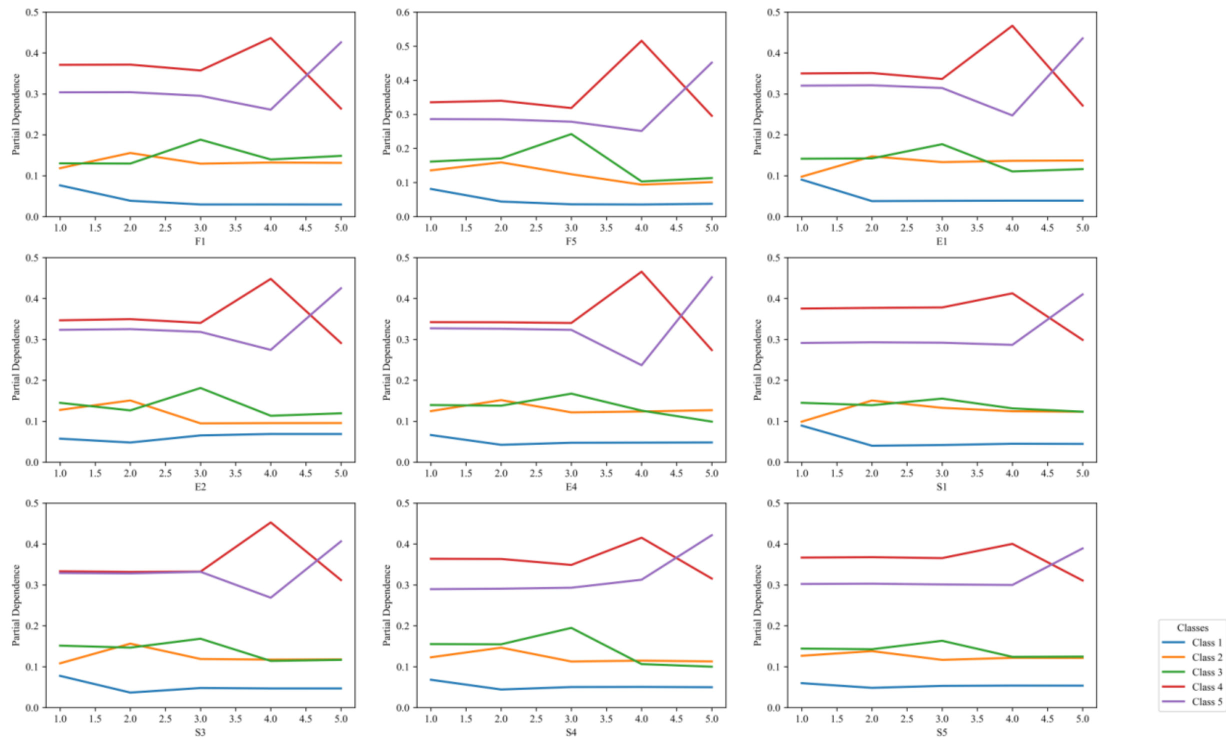

Figure 10.
Partial dependence plot.

5. Discussion

5.1. Comparison with Other Machine Learning Models

From the comparison results in Figure 11, PSO-LightGBM performs well when processing questionnaire data, with accuracy, precision, recall, and F1-score, all higher than most other models, indicating that it can balance accuracy while also capturing the pattern of the questionnaire data. The excellent performance of PSO-LightGBM can be attributed to its efficient processing of large-scale datasets and refined feature selection. PSO-LightGBM uses a leaf-splitting algorithm based on an improvement of gradient-boosted trees (GBDT), which significantly speeds up training while ensuring model accuracy [47]. It is especially suitable for datasets with high-dimensional feature spaces. Questionnaire data often has many feature variables, and PSO-LightGBM can quickly determine which features significantly impact the prediction results [48]. In contrast, SVM excels with small data sets and linearly separable problems, but its performance is susceptible to data noise in high-dimensional features and complex nonlinear issues [49]. More specifically, SVM requires nonlinear mapping by kernel function in high-dimensional feature space. Although the kernel can capture nonlinear relationships, choosing kernel parameters in high-dimensional data significantly impacts the model's performance. In our experiments, the standard kernel function (RBF kernel) fails to model these complex relationships effectively, resulting in degraded classification performance. KNN, which is sensitive to the size of the data and the distribution of the feature space because it relies on a distance metric, performed relatively poorly on this dataset, indicating that it could not capture the patterns in the questionnaire data effectively. RF performs well in dealing with missing data but is relatively weak in feature importance interpretability; MLP performs well in data with large-scale nonlinear and continuous features but requires extensive parameter tuning and is susceptible to overfitting.

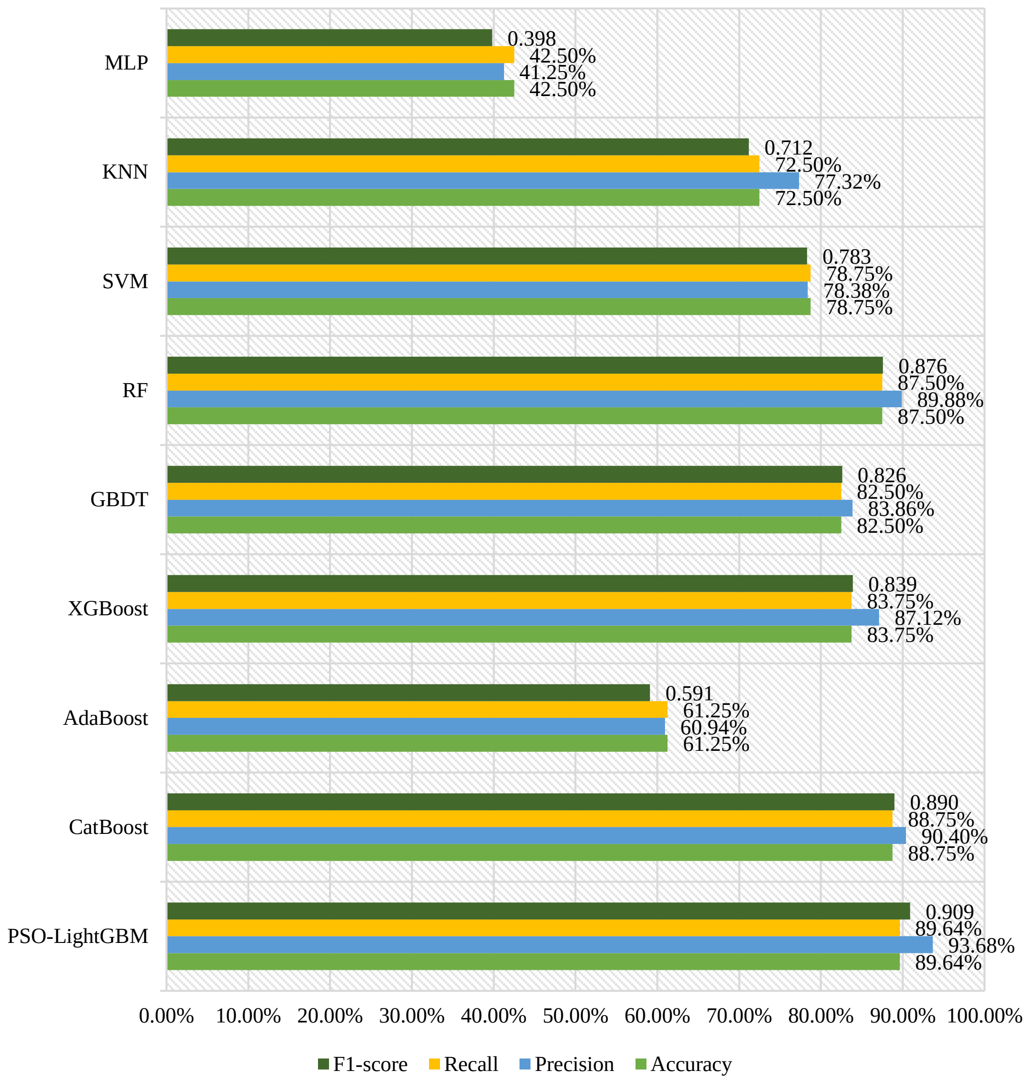

Figure 11.
ML models performance comparison.

XGBoost and GBDT are also based on gradient boosting algorithms, but XGBoost uses a sequential weighted and cumulative optimization strategy [50]. Although its performance was good, it was slightly inferior to PSO-LightGBM, probably because its splitting method consumed more computing resources and was less efficient than PSO-LightGBM at learning complex non-linear relationships. LightGBM has fewer parameters, and it is relatively easier to find the global optimal parameters through PSO. CatBoost has a more complex parameter space (e.g., internal mechanisms to support automatic coding of category variables, depth control, etc.). Adaboost performed poorly, reflecting that its approach of combining weak learners was ineffective at improving model performance when dealing with high-dimensional, non-linear, and noisy data [51].

5.2. Satisfaction Improvement Strategy

In terms of functionality, the key to improving residents' satisfaction is to ensure that the renovated building meets expectations in terms of energy efficiency, comfort, and ease of use of equipment. Smart home systems and efficient energy management equipment can reduce energy consumption while improving the living experience [52]. In addition, it is essential to ensure the equipment's stability and ease of operation, reduce the threshold for users to use and the difficulty of maintenance, and enhance users' acceptance of intelligent functions. For example, retrofitting measures undertaken by Andreas et al. included installing an HVAC system, covering the building envelope with external insulation, replacing lighting with LED fittings, installing a photovoltaic system and solar panels, and replacing external openings with aluminum windows. The energy consumption of the renovated building in Cyprus was reduced from 468 $kWh/m^2{\cdot}yr$ to 218 $kWh/m^2{\cdot}yr$, with renewable energy sources (RES) contributing 177 $kWh/m^2{\cdot}yr$ and emissions of carbon dioxide from 136.73 $kg/m^2{\cdot}yr$ to 11.5 $kg/m^2{\cdot}yr$. Each sector reduced energy consumption, ranging from 25% for lighting to 83% for hot water [53]. We also recommend a parametric analysis to select the proper technical and financial criteria to determine the optimal technology mix [52]. Integrating building functions and extending building life should also be considered to improve residents' confidence in the renovation results.

In terms of emotions, as reflected in Figure 12, residents' sense of participation in the decision-making process should be enhanced, and their opinions and needs should be fully considered to ensure that they can see personalized customization, including thermal comfort and indoor air quality [54]. At the same time, users' sense of psychological belonging and accomplishment should be enhanced by improving the residence's sense of security through design, such as optimizing privacy and security equipment. Most occupants lack knowledge of ventilation and manual controls, leading them to alleviate discomfort by blocking diffusers or disconnecting devices when the machinery malfunctions [55]. Therefore, new retrofit practices should pay special attention to the user-friendliness of the technical facilities, clearly communicate technical information to residents, and closely monitor the performance of the facilities.

In terms of society, the focus should be on enhancing their sense of identity in the community and society. Strengthening community interaction can promote communication and consensus among neighbors, creating a positive community atmosphere. At the same time, positive media publicity and social recognition can further enhance the social image of users so that residents can gain social recognition for their environmentally friendly lifestyles. The UK's Green Deal, which was intended to improve energy efficiency (the target was 1 million households), ultimately failed because of its complex and bureaucratic procedures, interest rates higher than those for mortgages, and the sole aim of financial savings rather than improving social well-being [56].

Figure 12.
Decision-making process for NZEB retrofit.

In terms of cost, retrofitting projects should ensure cost transparency and reasonableness. Given Malaysia's high energy costs, we recommend that renewable energy incentives abandon feed-in tariffs and subsidies for direct energy use, storage, and load matching. Future optimization analyses should also quantify the costs of thermal discomfort, energy poverty, and grid mismatches [57]. At the same time, the retrofitted buildings should reflect long-term energy cost savings and bring substantial economic returns to users. The potential of ML algorithms in this field has been demonstrated in this study. In the future, attempts could be made to use a combination of energy simulation and more intelligent algorithms to determine the optimal cost plan. For example, Heravi, et al. [58] combined energy simulation with a non-dominated sorting genetic algorithm to determine the optimal cost plan for designing a nearly zero-energy residential building in Kabul, Afghanistan's capital and largest city. The plan has a payback period of two years and a total energy reduction of 83%. Using a tabu search optimization algorithm, Munguba, et al. [59] determined a solution that minimizes electricity intensity and life cycle costs over 25 years. This method integrates thermal modeling and economic analysis to determine synergistic retrofits and PV sizing configurations, reducing consumption by 45 MWh/year and increasing net present value (NPV) by over $170,000 without additional investment (relative to baseline performance).

ISSN: 2576-8484
Vol. 9, No. 3: 2508-2530, 2025
DOI: 10.55214/25768484.v9i3.5835

6. Conclusion

In the current global energy crisis and climate change, renovating zero-energy buildings has become essential to improving energy efficiency and reducing carbon emissions. However, residents' satisfaction with renovation projects is crucial to their promotion and implementation. Therefore, based on the combination of customer perceived value theory and multiple machine learning (ML) algorithms, we systematically analyzed the factor system that affects the satisfaction of zero-energy building renovations, including four dimensions: functional, emotional, social, and cost. Multiple indicators in these dimensions, such as energy efficiency improvement, living comfort, personalized experience, and economic returns, directly affect residents' acceptance and evaluation of retrofit projects. At the same time, by comparing the performance of different ML models (AdaBoost, XGBoost, etc.), it is proved that PSO-LightGBM has superior performance when processing questionnaire data, especially in terms of accuracy, precision, and F1 score, which provides strong support for the subsequent satisfaction prediction of related projects.

It is worth noting that the data collection for this study was limited to Malaysia, a geographical limitation that restricts the generalizability of the findings to other countries and regions. Malaysia's national cultural characteristics (e.g., high collectivism and substantial power distance) and economic context (e.g., lower energy costs and government policy support for sustainable buildings) uniquely influence the prioritization of customer perceived value and satisfaction factors. These cultural and economic factors result in findings more reflective of localized needs and preferences and cannot be directly generalized to countries with different cultural or economic backgrounds. In addition, referring to Sougkakis, et al. [60] we plan to introduce multi-region data and conduct long-term tracking analysis in our follow-up study, as well as try to optimize the performance using a hybrid ML model to validate the generalizability of the study's conclusions and deeply explore the dynamic change patterns of satisfaction.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Authors' Contributions:

Made substantial contributions to the conception and design of the study and performed data analysis and interpretation: Qin X, Yin J.

Performed data acquisition and provided administrative, technical, and material support: Nuzul AH, Aidi HA, Law TH, Nabilah AB.

Copyright:

References

[1] M. Panagiotidou, L. Aye, and B. Rismanchi, "Optimisation of multi-residential building retrofit, cost-optimal and net-zero emission targets," *Energy and buildings*, vol. 252, p. 111385, 2021. https://doi.org/10.1016/j.enbuild.2021.111385

[2] L. Belussi *et al.*, "A review of performance of zero energy buildings and energy efficiency solutions," *Journal of Building Engineering*, vol. 25, p. 100772, 2019. https://doi.org/10.1016/j.jobe.2019.100772

[3] P. Moran, J. O'Connell, and J. Goggins, "Sustainable energy efficiency retrofits as residenial buildings move towards nearly zero energy building (NZEB) standards," *Energy and Buildings*, vol. 211, p. 109816, 2020. https://doi.org/10.1016/j.enbuild.2020.109816

[4] Ministry of Housing and Urban-Rural Development, "Official announcement on housing and urban development policies," Retrieved: https://www.mohurd.gov.cn/xinwen/gzdt/202402/20240201_776526.html. [Accessed September 1, 2024], 2024.
[5] A. D. Wills, I. Beausoleil-Morrison, and V. I. Ugursal, "A modelling approach and a case study to answer the question: What does it take to retrofit a community to net-zero energy?," *Journal of Building Engineering*, vol. 40, p. 102296, 2021. https://doi.org/10.1016/j.jobe.2021.102296
[6] E. Ohene, S.-C. Hsu, and A. P. Chan, "Feasibility and retrofit guidelines towards net-zero energy buildings in tropical climates: A case of Ghana," *Energy and Buildings*, vol. 269, p. 112252, 2022. https://doi.org/10.1016/j.enbuild.2022.112252
[7] J. F. W. Costa, C. N. D. Amorim, and J. C. R. Silva, "Retrofit guidelines towards the achievement of net zero energy buildings for office buildings in Brasilia," *Journal of Building Engineering*, vol. 32, p. 101680, 2020. https://doi.org/10.1016/j.jobe.2020.101680
[8] F. Emil and A. Diab, "Energy rationalization for an educational building in Egypt: Towards a zero energy building," *Journal of Building Engineering*, vol. 44, p. 103247, 2021. https://doi.org/10.1016/j.jobe.2021.103247
[9] S. Roumi, F. Zhang, R. A. Stewart, and M. Santamouris, "Weighting of indoor environment quality parameters for occupant satisfaction and energy efficiency," *Building and Environment*, vol. 228, p. 109898, 2023. https://doi.org/10.1016/j.buildenv.2022.109898
[10] L. Asere and A. Blumberga, "Energy efficiency–indoor air quality dilemma in public buildings," *Energy Procedia*, vol. 147, pp. 445-451, 2018. https://doi.org/10.1016/j.egypro.2018.07.115
[11] L. Sarran, C. A. Hviid, and C. Rode, "Correlation between perceived usability of building services and indoor environmental satisfaction in retrofitted low-energy homes," *Building and Environment*, vol. 179, p. 106946, 2020. https://doi.org/10.1016/j.buildenv.2020.106946
[12] M. S. Geraldi and E. Ghisi, "Integrating evidence-based thermal satisfaction in energy benchmarking: A data-driven approach for a whole-building evaluation," *Energy*, vol. 244, p. 123161, 2022. https://doi.org/10.1016/j.energy.2022.123161
[13] A. Samudro, U. Sumarwan, M. Simanjuntak, and E. Z. Yusuf, "Assessing the effects of perceived quality and perceived value on customer satisfaction," *Management Science Letters*, vol. 10, no. 5, pp. 1077-1084, 2020. https://doi.org/10.5267/j.msl.2019.11.001
[14] M. I. El-Adly, "Modelling the relationship between hotel perceived value, customer satisfaction, and customer loyalty," *Journal of Retailing and Consumer Services*, vol. 50, pp. 322-332, 2019. https://doi.org/10.1016/j.jretconser.2018.07.007
[15] E. Maghsoudi Nia, Q. K. Qian, and H. J. Visscher, "Analysis of occupant behaviours in energy efficiency retrofitting projects," *Land*, vol. 11, no. 11, p. 1944, 2022. https://doi.org/10.3390/land11111944
[16] S. Moeller and A. Bauer, "Energy (in) efficient comfort practices: How building retrofits influence energy behaviours in multi-apartment buildings," *Energy Policy*, vol. 168, p. 113123, 2022. https://doi.org/10.1016/j.enpol.2022.113123
[17] M. Oostra and N. Nelis, "Concerns of owner-occupants in realising the aims of energy transition," *Urban Planning*, vol. 7, no. 2, pp. 45-57, 2022. https://doi.org/10.17645/up.v7i2.5043
[18] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, pp. 1-21, 2021. https://doi.org/10.1007/s42979-021-00592-x
[19] Y. Junjia, A. H. Alias, N. A. Haron, and N. A. Bakar, "Machine learning algorithms for safer construction sites: Critical review," *Building Engineering*, vol. 2, no. 1, pp. 544-544, 2024. https://doi.org/10.59400/be.v2i1.544
[20] W. Zhang, C. Wu, L. Tang, X. Gu, and L. Wang, "Efficient time-variant reliability analysis of Bazimen landslide in the three Gorges Reservoir area using XGBoost and LightGBM algorithms," *Gondwana Research*, vol. 123, pp. 41-53, 2023. https://doi.org/10.1016/j.gr.2022.10.004
[21] U. Haverinen-Shaughnessy *et al.*, "Occupant satisfaction with indoor environmental quality and health after energy retrofits of multi-family buildings: Results from INSULAtE-project," *International Journal of Hygiene and Environmental Health*, vol. 221, no. 6, pp. 921-928, 2018. https://doi.org/10.1016/j.ijheh.2018.05.009
[22] V. Leivo, M. Turunen, A. Aaltonen, M. Kiviste, L. Du, and U. Haverinen-Shaughnessy, "Impacts of energy retrofits on ventilation rates, CO2-levels and occupants' satisfaction with indoor air quality," *Energy Procedia*, vol. 96, pp. 260-265, 2016. https://doi.org/10.1016/j.egypro.2016.09.148
[23] W. Loengbudnark, K. Khalilpour, G. Bharathy, A. Voinov, and L. Thomas, "Impact of occupant autonomy on satisfaction and building energy efficiency," *Energy and Built Environment*, vol. 4, no. 4, pp. 377-385, 2023. https://doi.org/10.1016/j.enbenv.2022.02.007
[24] H. Rau, P. Moran, R. Manton, and J. Goggins, "Changing energy cultures? Household energy use before and after a building energy efficiency retrofit," *Sustainable Cities and Society*, vol. 54, p. 101983, 2020. https://doi.org/10.1016/j.scs.2019.101983

[25] M. Ornetzeder, M. Wicher, and J. Suschek-Berger, "User satisfaction and well-being in energy efficient office buildings: Evidence from cutting-edge projects in Austria," *Energy and Buildings*, vol. 118, pp. 18-26, 2016. https://doi.org/10.1016/j.enbuild.2016.02.036

[26] P. H. Shaikh, N. B. M. Nor, A. A. Sahito, P. Nallagownden, I. Elamvazuthi, and M. Shaikh, "Building energy for sustainable development in Malaysia: A review," *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 1392-1403, 2017. https://doi.org/10.1016/j.rser.2016.11.128

[27] N. K. Mustaffa and S. A. Kudus, "Challenges and way forward towards best practices of energy efficient building in Malaysia," *Energy*, vol. 259, p. 124839, 2022. https://doi.org/10.1016/j.energy.2022.124839

[28] E. K. Kermanshahi, M. B. M. Tahir, N. H. A. S. Lim, A. T. Balasbaneh, and S. Roshanghalb, "Sustainable residential building retrofit for improving energy efficiency: The owners' perspective in Malaysia," in *IOP Conference Series: Earth and Environmental Science*, 2020, vol. 476, no. 1: IOP Publishing, p. 012008.

[29] A. N. Sadeghifam, M. S. Joo, N. N. M. Pauzi, and S. S. B. Namini, "Energy retrofitting and thermal comfort improvement of the public buildings in Malaysia," in *2021 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*, 2021: IEEE, pp. 1-5.

[30] C. C. Ohueri, W. I. Enegbuma, and R. Kenley, "Energy efficiency practices for Malaysian green office building occupants," *Built Environment Project and Asset Management*, vol. 8, no. 2, pp. 134-146, 2018. https://doi.org/10.1108/bepam-10-2017-0091

[31] D. George and P. Mallery, *IBM SPSS statistics 26 step by step: A simple guide and reference*. New York: Routledge, 2020.

[32] W. E. Wagner, *Using IBM SPSS statistics for research methods and social science statistics*, 7th ed. Thousand Oaks, CA: Sage Publications, Inc, 2020.

[33] Y. Ju, G. Sun, Q. Chen, M. Zhang, H. Zhu, and M. U. Rehman, "A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting," *Ieee Access*, vol. 7, pp. 28309-28318, 2019. https://doi.org/10.1109/access.2019.2901920

[34] W. Jing, B. Qian, and L. Yannian, "Study on food safety risk based on LightGBM model: A review," *Food Science and Technology*, vol. 42, p. e42021, 2022. https://doi.org/10.1590/fst.42021

[35] W. Liang, S. Luo, G. Zhao, and H. Wu, "Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms," *Mathematics*, vol. 8, no. 5, p. 765, 2020. https://doi.org/10.3390/math8050765

[36] G. Gao, H. Wang, and P. Gao, "Establishing a credit risk evaluation system for smes using the soft voting fusion model," *Risks*, vol. 9, no. 11, p. 202, 2021. https://doi.org/10.3390/risks9110202

[37] H. Yang, Z. Chen, H. Yang, and M. Tian, "Predicting coronary heart disease using an improved LightGBM model: Performance analysis and comparison," *IEEE Access*, vol. 11, pp. 23366-23380, 2023. https://doi.org/10.1109/ACCESS.2023.3253885

[38] J. Zhou, K. Xu, T. Xiang, and Y. Yu, "Composition analysis and identification of ancient glass objects based on Spearman correlation analysis model," *Highlights in Science, Engineering and Technology*, vol. 33, pp. 80-87, 2023. https://doi.org/10.54097/hset.v33i.5269

[39] T. B. Alakus and I. Turkoglu, "Comparison of deep learning approaches to predict COVID-19 infection," *Chaos, Solitons & Fractals*, vol. 140, p. 110120, 2020. https://doi.org/10.1016/j.chaos.2020.110120

[40] D.-n. Wang, L. Li, and D. Zhao, "Corporate finance risk prediction based on LightGBM," *Information Sciences*, vol. 602, pp. 259-268, 2022. https://doi.org/10.1016/j.ins.2022.04.058

[41] F. Marini and B. Walczak, "Particle swarm optimization (PSO). A tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 149, 2015. https://doi.org/10.1016/j.chemolab.2015.08.020

[42] H. Norouzi and J. Bazargan, "Investigation of river water pollution using Muskingum method and particle swarm optimization (PSO) algorithm," *Applied Water Science*, vol. 14, no. 4, p. 68, 2024. https://doi.org/10.1007/s13201-024-02127-0

[43] I. Ekanayake, D. Meddage, and U. Rathnayake, "A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP)," *Case Studies in Construction Materials*, vol. 16, p. e01059, 2022. https://doi.org/10.1016/j.cscm.2022.e01059

[44] Y. Nohara, K. Matsumoto, H. Soejima, and N. Nakashima, "Explanation of machine learning models using shapley additive explanation and application for real data in hospital," *Computer Methods and Programs in Biomedicine*, vol. 214, p. 106584, 2022. https://doi.org/10.1016/j.cmpb.2021.106584

[45] J. Fan, S. Upadhye, and A. Worster, "Understanding receiver operating characteristic (ROC) curves," *Canadian Journal of Emergency Medicine*, vol. 8, no. 1, pp. 19-20, 2006. https://doi.org/10.1017/s1481803500013336

[46] J. Miao and W. Zhu, "Precision–recall curve (PRC) classification trees," *Evolutionary intelligence*, vol. 15, no. 3, pp. 1545-1569, 2022. https://doi.org/10.1007/s12065-021-00565-2

[47] H. Liao, X. Zhang, C. Zhao, Y. Chen, X. Zeng, and H. Li, "LightGBM: an efficient and accurate method for predicting pregnancy diseases," *Journal of Obstetrics and Gynaecology*, vol. 42, no. 4, pp. 620-629, 2022. https://doi.org/10.1080/01443615.2021.1945006

[48] L. Chai, X. Xie, C. Wang, G. Tang, and Z. Song, "Ground subsidence risk assessment method using PS-InSAR and LightGBM: A case study of Shanghai metro network," *International Journal of Digital Earth*, vol. 17, no. 1, p. 2297842, 2024. https://doi.org/10.1080/17538947.2023.2297842

[49] Y. Zhou, Y. Liu, D. Wang, and X. Liu, "Comparison of machine-learning models for predicting short-term building heating load using operational parameters," *Energy and buildings*, vol. 253, p. 111505, 2021. https://doi.org/10.1016/j.enbuild.2021.111505

[50] M. Najafzadeh and M. Mahmoudi-Rad, "Residual energy evaluation in vortex structures: On the application of machine learning models," *Results in Engineering*, vol. 23, p. 102792, 2024. https://doi.org/10.1016/j.rineng.2024.102792

[51] S. Rahman, M. Irfan, M. Raza, K. Moyeezullah Ghori, S. Yaqoob, and M. Awais, "Performance analysis of boosting classifiers in recognizing activities of daily living," *International Journal of Environmental Research and Public Health*, vol. 17, no. 3, p. 1082, 2020. https://doi.org/10.3390/ijerph17031082

[52] Q. Xiaoxiang, Y. Junjia, N. A. Haron, A. H. Alias, T. H. Law, and N. Abu Bakar, "Status, challenges and future directions in the evaluation of net-zero energy building retrofits: A bibliometrics-based systematic review," *Energies*, vol. 17, no. 15, p. 3826, 2024. https://doi.org/10.3390/en17153826

[53] A. Constantinides, M. Katafygiotou, T. Dimopoulos, and I. Kapellakis, "Retrofitting of an existing cultural hall into a net-zero energy building," *Energies*, vol. 17, no. 7, p. 1602, 2024. https://doi.org/10.3390/en17071602

[54] C. Ochoa and B. Massey, "Use of adaptable solutions to achieve near-zero energy buildings through a combined thermal and comfort performance approach," *Building Services Engineering Research and Technology*, vol. 41, no. 3, pp. 361-373, 2020. https://doi.org/10.1177/0143624419887020

[55] L. Sarran, C. A. Hviid, and C. Rode, "How to ensure occupant comfort and satisfaction through deep building retrofit? Lessons from a Danish case study," *Science and Technology for the Built Environment*, vol. 29, no. 7, pp. 663-677, 2023. https://doi.org/10.1080/23744731.2023.2194196

[56] S. Ebrahimigharehbaghi, Q. K. Qian, G. de Vries, and H. J. Visscher, "Identification of the behavioural factors in the decision-making processes of the energy efficiency renovations: Dutch homeowners," *Building Research & Information*, vol. 50, no. 4, pp. 369-393, 2022. https://doi.org/10.1080/09613218.2021.1929808

[57] D. Gatt, C. Caruana, and C. Yousif, "Building energy renovation and smart integration of renewables in a social housing block toward nearly-zero energy status," *Frontiers in Energy Research*, vol. 8, p. 560892, 2020. https://doi.org/10.3389/fenrg.2020.560892

[58] G. Heravi, M. M. Salehi, and M. Rostami, "Identifying cost-optimal options for a typical residential nearly zero energy building's design in developing countries," *Clean Technologies and Environmental Policy*, vol. 22, pp. 2107-2128, 2020. https://doi.org/10.1007/s10098-020-01962-4

[59] C. Munguba *et al.*, "Enhancing cost-efficiency in achieving near-zero energy performance through integrated photovoltaic retrofit solutions," *Applied Energy*, vol. 367, p. 123307, 2024. https://doi.org/10.1016/j.apenergy.2024.123307

[60] V. Sougkakis, K. Lymperopoulos, N. Nikolopoulos, N. Margaritis, P. Giourka, and K. Angelakoglou, "An investigation on the feasibility of near-zero and positive energy communities in the Greek context," *Smart Cities*, vol. 3, no. 2, pp. 362-384, 2020. https://doi.org/10.3390/smartcities3020019