sustainability

MDPI

Article

Machine Learning for Leadership in Energy and Environmental Design Credit Targeting: Project Attributes and Climate Analysis Toward Sustainability

Ali Mansouri, Mohsen Naghdi and Abdolmajid Erfani *

Department of Civil, Environmental, and Geospatial Engineering, Michigan Technological University, Houghton, MI 49931, USA; alimanso@mtu.edu (A.M.); mnaghdi@mtu.edu (M.N.)
* Correspondence: aerfani@mtu.edu

Abstract: Achieving Leadership in Energy and Environmental Design (LEED) certification is a key objective for sustainable building projects, yet targeting LEED credit attainment remains a challenge influenced by multiple factors. This study applies machine learning (ML) models to analyze the relationship between project attributes, climate conditions, and LEED certification outcomes. A structured framework was implemented, beginning with data collection from the USGBC (LEED-certified projects) and US NCEI (climate data), followed by preprocessing steps. Three ML models—Decision Tree (DT), Support Vector Regression (SVR), and XGBoost—were evaluated, with XGBoost emerging as the most effective due to its ability to handle large datasets, manage missing values, and provide interpretable feature importance scores. The results highlight the strong influence of the LEED version and project type, demonstrating how certification criteria and project-specific characteristics shape sustainability outcomes. Additionally, climate factors, particularly cooling degree days (CDD) and precipitation (PRCP), play a crucial role in determining LEED credit attainment, underscoring the importance of regional environmental conditions. By leveraging ML techniques, this research offers a data-driven approach to optimizing sustainability strategies and enhancing the LEED certification process. These insights pave the way for more informed decision-making in green building design and policy, with future opportunities to refine predictive models for even greater accuracy and impact.

Keywords: sustainability; Leadership in Energy and Environmental Design credit targeting; machine learning; climate factors

check for updates

Academic Editor: Ali Bahadori-Jahromi

Received: 11 February 2025
Revised: 7 March 2025
Accepted: 11 March 2025
Published: 13 March 2025

Citation: Mansouri, A.; Naghdi, M.; Erfani, A. Machine Learning for Leadership in Energy and Environmental Design Credit Targeting: Project Attributes and Climate Analysis Toward Sustainability. *Sustainability* **2025**, *17*, 2521. https://doi.org/10.3390/su17062521

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The building and construction sectors significantly contribute to global energy consumption and CO_2 emissions. According to the International Energy Agency (IEA), buildings account for approximately 30% of global energy use and 26% of energy-related CO_2 emissions. When factoring in embodied emissions from building materials, their total contribution rises to nearly one-third of global energy-related emissions [1]. Public awareness of green buildings is increasing, driven by both regulatory and market factors [2–6]. In the United States, federal, state, and local authorities either incentivize or mandate sustainable design and construction [7]. Beyond regulatory programs, market demand and environmental concerns have further fueled interest in sustainable buildings and certification systems. Since the 2000s, the Leadership in Energy and Environmental Design (LEED) certification has been widely used in the U.S. and internationally to assess and promote green building practices [8]. On average, LEED-certified buildings consume 25%

less energy than non-LEED-certified buildings, contributing more than USD 29 billion to the U.S. gross domestic product (GDP) [9,10]. LEED, developed by the U.S. Green Building Council (USGBC), is a flexible certification system designed for various project types. Its rating systems include Building Design and Construction (LEED BD + C), Interior Design and Construction (LEED ID + C), Operations and Maintenance (LEED O + M), and Neighborhood Development (LEED ND). These systems apply to new and existing buildings across multiple sectors, including schools, healthcare facilities, data centers, warehouses, hospitality venues, homes, commercial interiors, multifamily mid-rise developments, and retail spaces [11,12].

As constructors plan new buildings or building owners and operators retrofit existing structures for LEED certification, various considerations and challenges emerge. Among these, the selection of target LEED credits remains one of the most critical and complex decisions for LEED managers [13–16]. While achieving LEED certification provides significant benefits, it often necessitates additional funding and presents implementation challenges [17,18]. Initially, constructors were hesitant to participate due to uncertainties in cost estimation [19] and potential productivity losses associated with adopting new sustainable construction methods [20]. For instance, Mansouri et al. [21] demonstrated, using a simulation-based study, that operational inefficiencies in earthmoving equipment can significantly impact greenhouse gas (GHG) emissions, highlighting the need for innovative planning tools to enhance sustainability in construction processes. However, with the growing popularity of green buildings, planning to achieve certification has become essential. Selecting the appropriate LEED credits from the rating system is challenging due to direct factors such as budget constraints, tight schedules, and site-specific geographic and environmental conditions, as well as indirect factors like climate variables, including local precipitation and sunlight availability [13,22]. Therefore, analyzing the relationships between decision-making factors—including project type, other attributes, and geographical factors such as climatic conditions—and LEED credits has become a critical research focus.

Previous research on LEED credit selection [23–27] has employed a range of methodologies, including statistical analysis, multi-criteria decision-making (MCDM) techniques, and, more recently, data mining approaches incorporating artificial intelligence (AI)-based solutions. Pushkar [28], for example, assessed the effects of a project's size on the achievement of LEED for operations and maintenance. The results reveal that medium-sized projects excelled in sustainable structures and energy categories, while large projects performed best in location, transportation, and water efficiency. Medium-to-large projects showed similar outcomes in materials, resources, and indoor environment quality. Wu et al. [27] examined if there are significant differences in LEED credits across various building types and regions. Although these studies offer valuable insights, they are limited in their ability to provide actionable strategies for new projects [18], especially those involving dynamic non-linear relationships between variables. On the other hand, MCDM techniques, such as the Analytical Hierarchy Process (AHP) [29,30] and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [31], have also been utilized. For example, Seyis and Ergen [32] integrated the Delphi method, multi-attribute utility technique, and TOPSIS technique to create a decision support tool and use the relative weights of the attributes that were determined following expert opinions to guide experts in selecting suitable credits based on project delivery attributes. Madanayake et al. [33] proposed a decision-making model that combines the AHP and Monte Carlo simulations [34], considering multiple criteria such as project cost fluctuations, environmental impact, schedule implications, and construction productivity. Similarly, Attallah et al. [35] proposed an MCDM approach for ranking green building certification credits using the ELECTRE III method. Their

selection criteria included factors such as project type and location, client type, and the architect/engineer's experience and familiarity. Although these types of studies are effective in case studies, they often rely on expert experience and knowledge, and the credit selection process remains complex [18].

As projects continue to increase in number and the relationships between factors influencing credit selection grow more complex, AI-based methods have become a promising solution to tackle non-linear challenges and improve overall effectiveness and accuracy. Ma and Cheng [13] developed a framework using the Random Forest (RF) algorithm to select credits based on project and climate factors. Similarly, Jun and Cheng [16] employed and compared three classification algorithms—RF, AdaBoost Decision Tree, and Support Vector Machine (SVM)—for selecting target LEED credits based on the project's information and climatic factors. Cheng and Ma [36] proposed a case-based reasoning (CBR) approach to aid LEED managers in selecting target credits for green building projects by referencing similar certified cases.

According to Table 1, previous studies examining the relationship between LEED credits and affective features can be divided into three categories. The first category employed statistical methods, which were limited in features and ability to capture non-linear relationships. The second category depended on surveys and expert opinions, leading to complexity in scaling and time-consuming processes; these results were often influenced by the expert's experience and lacked overall generalizability. The third category used AI tools, but their datasets were limited and not comprehensive across all regions. In contrast, this paper analyzes project attributes and climate features by examining 67,347 projects across approximately 5500 cities and all U.S. states and territories. This approach promises more accurate and stable results. Additionally, it considers multiple versions of LEED rather than being restricted to just one version.

Table 1. Summary of previous studies on the relation of LEED credits and affective features.

Study	Method	Features	Dataset Size	Limitation
Pushkar [28]	Statistical analysis	Project size	94 projects	Focuses only on project size, misses other factors and non-linear relationships, limited to LEED-EB V4, small dataset
Wu et al. [27]	Statistical analysis	Building types	3416 projects	Lack of actionable strategies, limited to LEED 2009, small dataset
Seyis and Ergen [32]	Delphi method, MAUT *, TOPSIS **	Project delivery attributes	Expert opinions	Relies on expert opinions, may not generalize well
Madanayake et al. [33]	AHP, Monte Carlo simulation	Project cost fluctuations, environmental impact, schedule implications, construction productivity	Expert opinions	Depends on expert knowledge, time-consuming to implement
Attallah et al. [35]	Statistical analysis, ELECTRE III, Simulation	Project type, location, client type, architect/engineer's experience	Expert opinions	Depends on expert knowledge, limited projects' features and regions
Cheng and Ma [36]	Case-based reasoning (CBR), ANN	Gross sqft, total property area, owner type, project type and location, grade level	1000 projects	Depends on availability and similarity of past cases, limited to LEED-NC v2009, small dataset
Ma and Cheng [13]	Random Forest (RF)	Climate factors	912 projects	Limited to LEED-EB 2009, small dataset
Jun and Cheng [16]	RF, AdaBoost Decision Tree, SVM	Project information, climatic factors	912 projects	Limited to LEED-EB 2009, small dataset

* Multi-attribute utility technique. ** Technique for order of preference by similarity to ideal solution.

Goal and Objectives

This study aims to investigate the relationship between LEED credit achievement and a project's attributes, including its type, size, ownership, and surrounding climatic factors. Our objective is to develop an AI-based model for predicting suitable LEED credit targets using available green building data. This study contributes to the body of knowledge by implementing advanced AI algorithms, including ensemble learning techniques, and leveraging a large-scale green building dataset across the United States. It incorporates a comprehensive set of features, including complete climatic factors (temperature, precipitation, humidity, snowfall, and wind speed). Using a feature importance analysis, this study identifies key contributors to optimizing LEED credit targets for new buildings. Compared to existing AI-based studies, this research utilizes a significantly larger dataset (over 60,000 records), a more comprehensive set of attributes, and, most importantly, incorporates advanced AI techniques to better capture the complex relationships between input features and the target LEED score.

2. Materials and Methods

Advancements in AI, including machine learning, deep learning, computer vision, and generative AI, have increasingly shaped civil engineering research [37–42], particularly in sustainability-focused studies [43–46]. This study benefits from machine learning models to investigate the relationship between the achievements of LEED credits and project attributes and climate variables. The methodology framework is shown in Figure 1. The required data for this study were collected from two public databases.

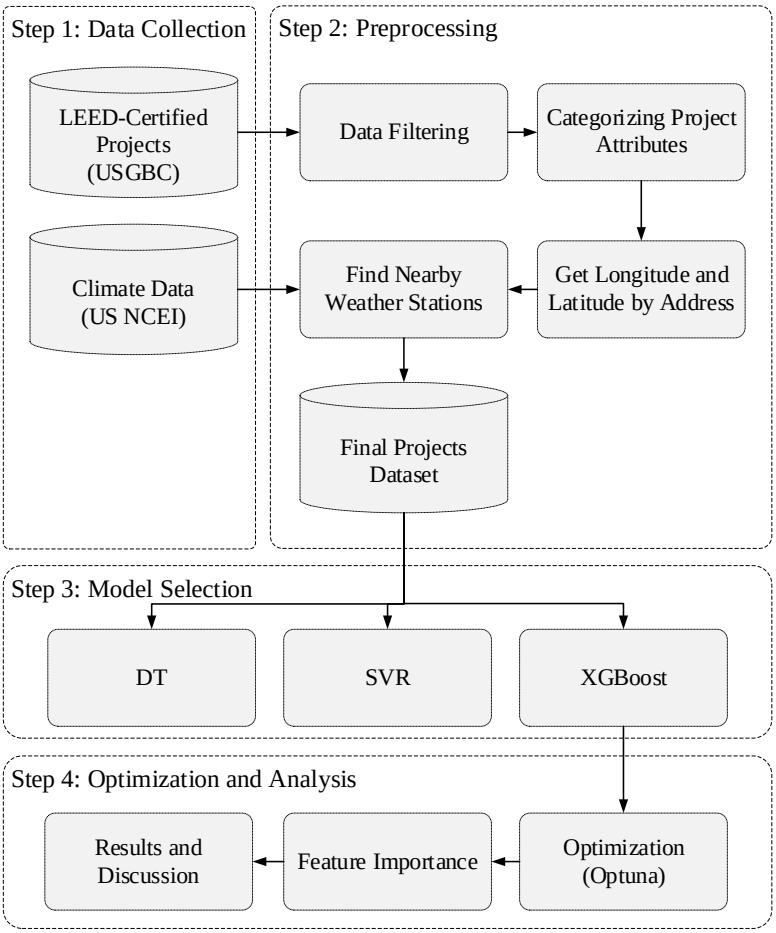


Figure 1. Framework for LEED project credit targeting.

The first is a dataset of LEED projects collected from the USGBC website [8], and the second is the climate variable dataset of weather stations from the US National Centers for Environmental Information (NCEI) website [47]. After preprocessing the data, different ML models were applied, and the best of them, in terms of performance, were selected to develop the final optimized model. Finally, the most important features are identified and discussed.

All experiments were conducted using Google Colab, a cloud-based platform running on a Google Compute Engine backend. The models were implemented in Python 3.11 using the scikit-learn library. The computations were performed on the default Colab environment without GPU or TPU acceleration. The system was equipped with an Intel(R) Xeon(R) CPU @ 2.20GHz (2 vCPUs), approximately 12.7 GB of RAM, and 107.7 GB of disk space.

2.1. Data Collection

The USGBC provides the LEED project directory, which contains LEED-certified projects around the globe on their website publicly [8]. This study utilized the dataset supplied on this website, which included 174,471 projects at the time of the study in 2025. The dataset contains information like address, points achieved (credit score), owner types, project types, gross floor area, etc. Additionally, the National Centers for Environmental Information (NCEI) provides U.S. Climate Normal data, calculated for a uniform 30-year period from almost 15,000 U.S. weather stations, on their website [47]. This study uses the annual dataset, which contains the averages and statistics of temperature, precipitation, and other climatological variables.

2.2. Data Preprocessing

2.2.1. Data Filtering

Data filtering is the first step of preprocessing the data applied to the LEED project dataset. Some projects are confidential in this dataset, and their information is missing, so these projects are dropped. Since this study focuses on U.S. projects, data from other countries has been excluded. Lastly, the LEED credit for some projects is missing, and because this parameter is the target of this study, the missing ones are also removed. By performing these filtrations, the total number of projects reached 68,549. Figure 2 illustrates the geographic distribution of publicly available projects used in this study across the U.S. Green dots represent buildings that have been certified, blue indicates those that have achieved the silver level, yellow represents those that have achieved the gold level, and red denotes buildings that have reached the highest certification level, platinum.

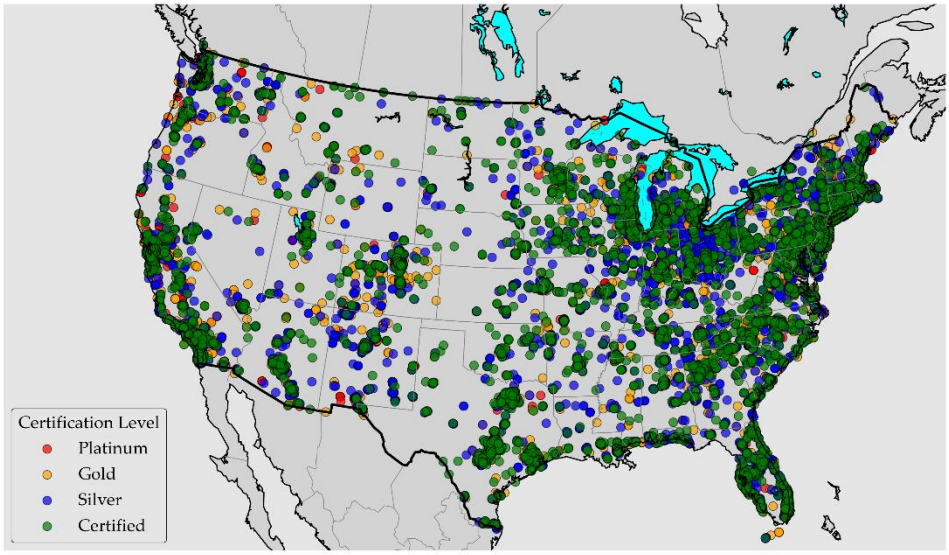


Figure 2. Geographical distribution of selected green buildings of US in this study.

2.2.2. Merging the LEED Dataset with Climate Data

Another key data preprocessing step involves merging climatic data with the LEED dataset using geospatial information. The projects' nearby weather stations should be identified to add climate variables to project data. To achieve this, we must know the geographic coordination of the projects, but the only related information in this regard were ones like street, city, state, and zip code. Therefore, as a first step, these location data were merged to demonstrate the full address for a project. We used Google Map API to extract the longitude and latitude by sending the addresses. As a result, we could obtain the location information for more than 98 percent of projects, which narrowed them down to

67,347. On the other hand, in the weather dataset, some stations had missing data for some climate variables. In this regard, assigning one nearest station to the project may have led to missing data for that project. So, a radius of 20 km was considered around each project location, and the stations in this range were assigned to the project. The corresponding climate variables of each project were calculated by averaging the values of the nearby stations. Maria et al. [48] compared the Haversine formula and the Euclidean method for distance calculations. They found that the Haversine formula yielded results more aligned with Google Maps, making it better suited for geographic applications. Therefore, to calculate the distance between projects and stations in this process, the Haversine formula was used, as shown in Equation (1).

$$Distance = 2r \arcsin\left(\sqrt{\frac{1-\cos(\varphi_2-\varphi_1)+\cos\varphi_1\cdot\cos\varphi_2\cdot(1-\cos(\lambda_2-\lambda_1))}{2}}\right), \quad (1)$$

where φ_1 and φ_2 represent the latitude of point 1 and point 2 and λ_1 and λ_2 represent the longitude of point 1 and point 2.

2.2.3. Handling Missing Values

Since some features in the dataset contain missing values, they must be addressed before training the machine learning models. However, certain models, such as XGBoost, can handle missing data under specific conditions. Various techniques are available for managing missing data, depending on the nature and extent of the missing values [49]. One is dropping the feature with missing values when the feature is not important or valuable. Second is to drop the rows which contain missing values. This may lead to loss of the majority of data because unless the dropped datapoint is relatively small. Another method is imputing the missing values. The most common way to impute data is replacing the missing values with the mean for the numerical data and the most frequent values for the categorical data [50]. In this study, the imputation method was used for the DT and SVR models, and, since XGBoost model has built-in methods for handling missing values, the missing values were ignored for this model.

2.2.4. Category Encoding

The project dataset includes three categorical features: LEED version, owner type, and project type. Although their default values are basically categorical, to achieve better results, we labeled them to decrease the number of categories. Figure 3 shows the number of projects per category.

Category encoding is a crucial step in machine learning, especially when dealing with categorical variables [51]. Different methods are used to convert these categories into numerical values that models can interpret. The most common include the following: Label Encoding, which is suitable when there is an ordinal relationship between categories [52]; One-Hot Encoding, which creates a new binary column for each category [53]; Frequency Encoding, which replaces each category with the frequency or count of occurrences in the dataset; Target Encoding, which replaces each category with the mean of the target variable for that category; and Leave-One-Out Encoding, a variant of target encoding where the target value for a given category is the mean of the target variable for all instances of that category, except for the current instance. According to the categorical features in this study, since there are no ordinal relationships between categories, the number of categories is relatively high and the frequency of the category does not provide additional insight for the model; the Label Encoding, One-Hot Encoding, and Frequency Encoding are not suitable for this study. So, Target Encoding and Leave-One-Out Encoding are used for the

categorical features in this study. Figure 4 illustrates these two encoding methods using a hypothetical example.

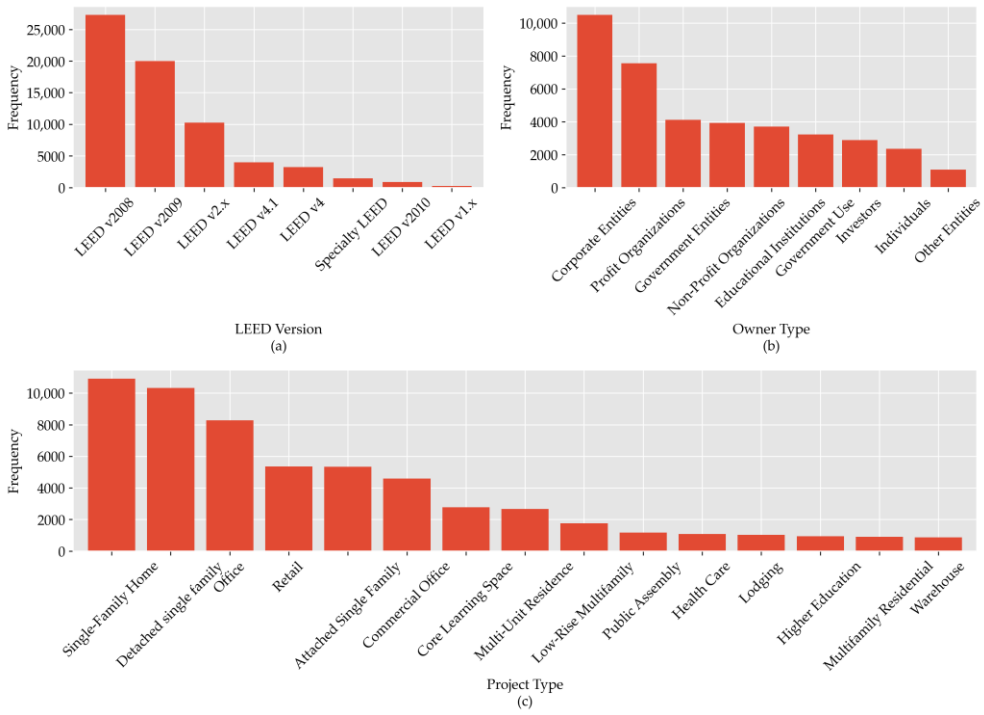


Figure 3. Frequency of projects in the (**a**) LEED version, (**b**) owner type, and (**c**) project type categories.

Project ID	Owner Type	LEED Credit	Target	Leave-One-Out
1	Educational Institutions	50	45	40
2	Educational Institutions	40	45	50
3	Government Entities	90	80	75
5	Government Entities	80	80	80
5	Government Entities	70	80	85
6	Individuals	30	30	30

Average Credits of Projects in the 'Government Entities' Category = (90+80+70)/3 = 80

Average Credits of Projects in the 'Government Entities' Category excluding project 3 = (80+70)/2 = 75

Figure 4. Illustration of target and leave-one-out encoding.

2.2.5. Final Dataset

In the final dataset, the features include four project attributes: LEED version, owner type, project type, and gross floor area. Additionally, there are eight climate variables: average temperature (TAVG), diurnal temperature range (DUTR), cooling degree days (CLDD), heating degree days (HTDD), total precipitation (PRCP), snowfall (SNOW), the average number of days with precipitation exceeding 0.01 inches (PRCP_GE0.01), and the average number of days with snowfall greater than 0.1 inches (SNOW_GE0.1). Table 2 presents the selected project attributes and climate factors used in this study, along with their data types and units. Table 3 presents the descriptive statistics for the numerical input

features. Also, Table 4 provides a few examples of data points from the final dataset after preprocessing, prepared for machine learning algorithms.

Table 2. Final chosen variables in this study.

Category	Feature	Type	Unit
Project attribute	LEED version	Categorical	-
	Owner Type	Categorical	-
	Project Type	Categorical	-
	Gross Floor Area	Numerical	Sq ft
Climate variables	TAVG	Numerical	Fahrenheit
	DUTR	Numerical	Fahrenheit
	CLDD	Numerical	Fahrenheit
	HTDD	Numerical	Fahrenheit
	PRCP	Numerical	Inches
	SNOW	Numerical	Inches
	PRCP_GE0.01	Numerical	Day
	SNOW_GE0.1	Numerical	Day

Table 3. Descriptive statistics.

Feature	Count	Mean	Standard Deviation	Min	25%	50%	75%	Max
Gross Floor Area	67,347	82,482	323,003	0	2026	5669	64,263	37,000,000
TAVG	65,797	59	8	13	52	57	65	82
DUTR	65,797	20	3	8	18	20	22	38
CLDD	65,797	1627	1122	0	768	1301	2612	6289
HTDD	65,797	3725	2077	0	1949	4150	5376	18,846
PRCP	66,945	37	15	2	23	41	47	176
SNOW	61,621	17	24	0	0	4	25	463
PRCP_GE0.01	66,282	102	38	12	77	113	125	285
SNOW_GE0.1	61,621	9	13	0	0	2	13	100

Table 4. Sample of the final dataset.

LEED Version	Owner Types	Project Types	Gross Floor Area	TAVG	DUTR	CLDD	HTDD	PRCP	SNOW	PRCP_GE0.01	SNOW_GE0.1	Points Achieved
LEED v4	Investors	Office	163,887	68.2	23.0	2948.1	1741.3	38.0	0.2	86.7	0.4	61
LEED v4	Corporate Entities	Industrial	135,000	50.3	20.8	649.2	5980.0	50.7	36.5	131.0	12.1	43
LEED v2009	Investors	Office	438,580	61.9	21.2	1777.0	2877.8	52.7	0.5	117.5	0.3	64
LEED v2009	Investors	Warehouse	1,331,763	65.5	29.1	1767.1	1566.6	11.0	0.0	36.5	0.0	61
LEED v4	Educational Institutions	Other	34,000	53.2	20.4	950.4	5217.8	48.6	25.2	125.7	13.0	51

2.3. Model Selection

We employed machine learning models to analyze the impact of various parameters on the LEED credits achieved by the projects. To do so, we first identified the target variable

and its associated features. In this study, the target variable is the LEED points earned by a project, represented as a continuous numerical value, making regression models the appropriate choice. Figure 5 illustrates the distribution of data based on the target variable.

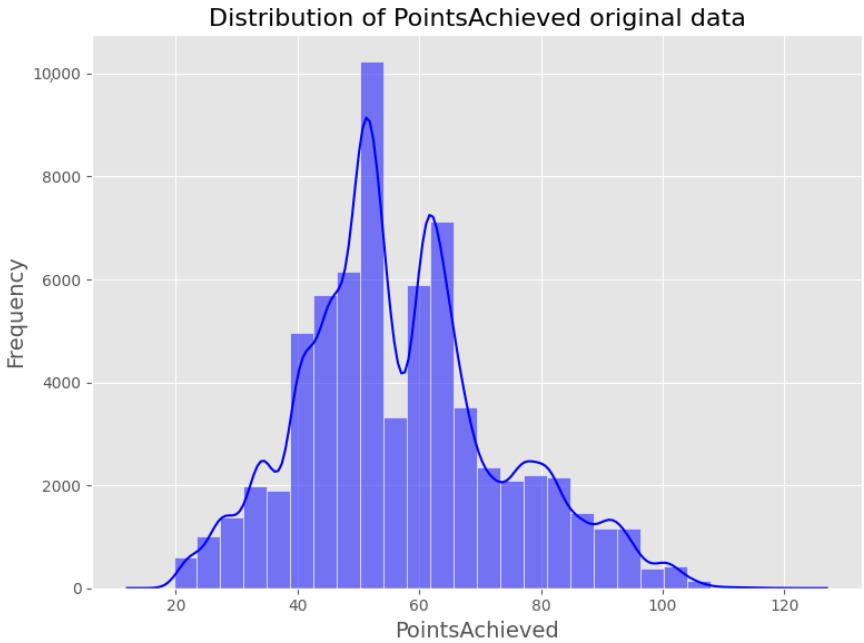


Figure 5. The distribution of data based on the LEED points achieved by projects.

For the regression analysis aimed at the achieved predicting points, as the first step, we selected three models to choose the best one for the final model in case of performance. The models employed include the following: (1) Decision Tree (DT), (2) Support Vector Regression (SVR), and (3) XGBoost. These models were chosen to balance interpretability, computational efficiency, and predictive accuracy, ensuring a robust analysis of LEED credit achievement [54–58]. The final model will be selected based on a performance benchmarking [59] comparison of the tested models.

2.3.1. Decision Tree (DT)

The DT is a widely used machine learning algorithm for both classification and regression tasks, favored for its high interpretability and versatility in handling both numerical and categorical data. A key advantage of this model is its visual representation of decision paths, which simplifies the understanding of how predictions are derived from the input variables [60]. Its ability to model complex non-linear relationships also makes it well-suited for predicting LEED credits. However, the challenge of overfitting requires the careful tuning of hyperparameters to ensure that the model remains generalized to new data [61].

2.3.2. Support Vector Machine (SVM)

The SVM is a robust machine learning model used primarily for classification, but which is also effective in regression [62]. It works by finding the hyperplane that best separates different classes or predicts values, maximizing the margin between data points. SVM is favored for its effectiveness in high-dimensional spaces and its flexibility in handling various types of data through different kernel functions. SVR is an extension of SVM that is used for predicting numerical values.

2.3.3. XGBoost

XGBoost, short for eXtreme Gradient Boosting, is an advanced implementation of gradient-boosting algorithms that known for its speed and performance. This algorithm is designed to be highly efficient, scalable, and flexible. It works by sequentially building a series of decision trees, where each tree attempts to correct the residuals or errors of the previous trees [63,64]. The strength of XGBoost lies in its ability to handle large datasets with a robust framework for handling missing data, regularizing to prevent overfitting, and tuning a variety of hyperparameters to optimize performance [65].

2.3.4. Performance Metrics

To evaluate the predictive performance of each model, this study employed standard statistical metrics commonly used in regression analysis [66]. The selected metrics—RMSE, R-squared, MAE, and MAPE—are defined in Equations (2)–(5). These equations represent the mathematical calculations used to evaluate the accuracy of regression models by comparing the predicted values $\hat{y_k}$ to the actual values y_k across n observations, where $\overline{y}$ is the meaning of the observed data.

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n} y_k - \hat{y_k}}, \tag{2}$$

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(y_k - \hat{y_k})^2}{\sum_{k=1}^{n}(y_k - \overline{y})^2} \tag{3}$$

$$MAE = \frac{1}{n}\sum_{k=1}^{n}|y_k - \hat{y_k}| \tag{4}$$

$$MAPE = \frac{100\%}{n}\left|\frac{y_k - \hat{y_k}}{y_k}\right| \tag{5}$$

2.3.5. Hyper-Tuning Parameters

To improve the model's results, machine learning models need to be optimized by identifying the best parameters that provide the best performance. In this study, Optuna is used for automatic hyperparameter optimization [67]. It efficiently searches for optimal hyperparameter values using an iterative trial-and-error process, aiming to improve the model's performance. It employs a Bayesian optimization algorithm called Tree-structured Parzen Estimator to estimate the promising areas for hyperparameter values. Table 5 presents the optimum values for the parameters of the XGBoost model.

Table 5. Optimum values for the parameters of the XGBoost model.

Parameter	Optimum Value
n_estimators	300
max_depth	10
learning_rate	0.054
subsample	0.930
colsample_bytree	0.709
min_child_weight	2

3. Results

Table 6 presents the performance metrics of the three models evaluated in this study. Since the DT and SVR models do not inherently handle missing values, data imputation was performed to address any missing entries before training these models. Additionally, the hyperparameters of the XGBoost model were optimized using Optuna to enhance its

performance. XGBoost outperformed the other models across all performance metrics, achieving the lowest Mean Squared Error (MSE) of 50.76, Mean Absolute Error (MAE) of 4.70, and Root Mean Squared Error (RMSE) of 7.12. Additionally, XGBoost demonstrated the highest R-squared value of 0.81, indicating the best fit to the data. The Mean Absolute Percentage Error (MAPE) for XGBoost was also the lowest at 9.05%, showing its superior predictive accuracy.

Table 6. Performance metrics comparison of ML models.

Performance Metric	DT	SVR	XGBoost
Mean Squared Error (MSE)	70.79	110.88	50.76
Mean Absolute Error (MAE)	5.32	7.67	4.70
Root Mean Squared Error (RMSE)	8.41	10.53	7.12
R-squared	0.74	0.59	0.81
Mean Absolute Percentage Error (MAPE)	10.17%	13.68%	9.05%
Training Time (Seconds)	0.55	211.73	9.58
Test Time (Seconds)	0.01	47.60	0.32
Total Execution Time (Seconds)	0.56	259.33	9.91

The Decision Tree (DT) model performed moderately well, having an MSE of 70.79, MAE of 5.32, and RMSE of 8.41. It achieved an R-squared value of 0.74, which is lower than XGBoost but significantly higher than SVR. The MAPE for DT was 10.17%, indicating a slightly higher prediction error than XGBoost. In contrast, the SVR model exhibited the weakest performance, with the highest MSE (110.88), MAE (7.67), and RMSE (10.53). Its R-squared value was the lowest at 0.59, suggesting a relatively poor fit to the data. Furthermore, the MAPE for SVR was the highest at 13.68%, reflecting the largest percentage error among the models.

Overall, the results indicate that XGBoost is the most effective model for this dataset, providing the most accurate predictions with the lowest error values and highest R-squared score. The Decision Tree model performs reasonably well, but falls short compared to XGBoost. SVR, however, struggles to capture the underlying patterns in the data, resulting in the highest errors and lowest predictive accuracy.

Based on the performance metrics presented in Table 6, XGBoost was selected as the best model for predicting LEED credit due to its superior accuracy and lower error rates compared to DT and SVR. The best parameters obtained using this optimization process were n_estimators = 300, max_depth = 10, learning_rate = 0.0536, subsample = 0.9296, colsample_bytree = 0.7086, and min_child_weight = 2. These optimized settings contributed to the improved efficiency and accuracy of the XGBoost model in this study.

To further evaluate the predictive performance of the optimized XGBoost model, a residual analysis was conducted by plotting the absolute residuals against the actual LEED credit values (Figure 6). Residuals, which are the absolute differences between predicted and actual values, reveal the extent of prediction errors. The plot shows that residuals are relatively lower for smaller LEED credit values, suggesting that the model performs with higher accuracy in this range. As the actual LEED credit values increase, the spread of residuals does expand slightly, but this could be attributed to the lower volume of data in the higher credit ranges, as indicated by the credit distribution. Although there are some outliers with larger residuals, they are relatively infrequent and do not significantly impact the overall performance.

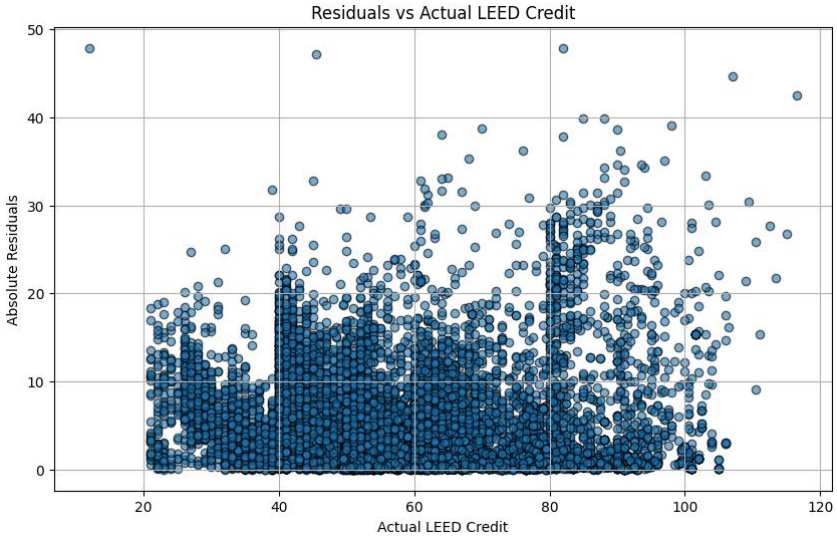


Figure 6. The scatter plot of absolute residuals versus actual LEED credit values.

Figure 7 illustrates the correlation between actual and predicted values. The data points are dispersed along a regression line, ideally representing a perfect fit where actual values equal predicted values. The degree of scatter around this line indicates the model's accuracy; a tighter concentration suggests higher predictive power. Notably, the plot reveals a cluster of data points at lower value ranges, suggesting higher accuracy in this segment of the data.

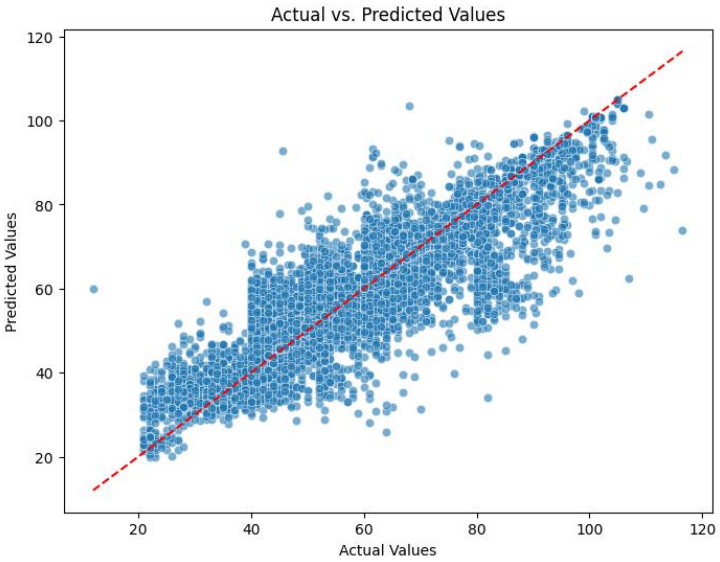


Figure 7. The scatter plot of the correlation between actual and predicted values.

Figure 8 presents the learning curve of RMSE (Root Mean Squared Error) vs. training samples for both the training and validation datasets. The RMSE for the validation dataset is slightly higher than that of the training dataset, suggesting that the model performs very well on the training data while encountering only a minor challenge when generalized to unseen data. The convergence of the training and validation curves suggests a reduction in overfitting as the training set size grows, implying that the model generalizes well to unseen data. The relatively small gap between the two RMSE values indicates that the model generalizes effectively and maintains strong performance on new data. As expected, the training RMSE increases as the dataset grows, reflecting larger datasets' natural tendency to introduce more complexity and variance, while smaller datasets are more easily memorized,

leading to lower RMSE. These observations suggest that the model is learning effectively and generalizing well to unseen data.

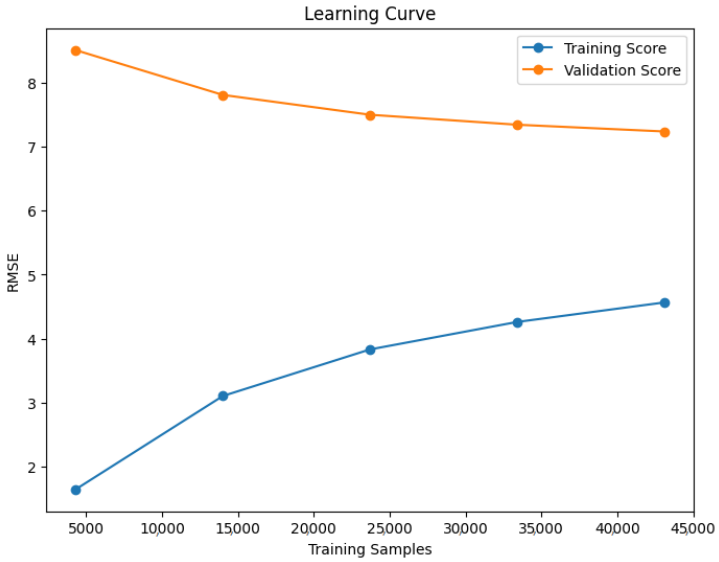


Figure 8. The learning curve of RMSE vs. training samples for both the training and validation datasets.

The XGBoost model includes a built-in function that calculates the importance scores of features related to the target variable. Figure 9 displays these scores in descending order. It shows that the LEED version is by far the most important factor. The project type follows as the second most crucial factor. The next four features pertain to climate variables, highlighting the significant role that climate plays in earning LEED points. Conversely, the Gross Floor Area, one of the main project attributes, is identified as the least important feature for achieving LEED credits.

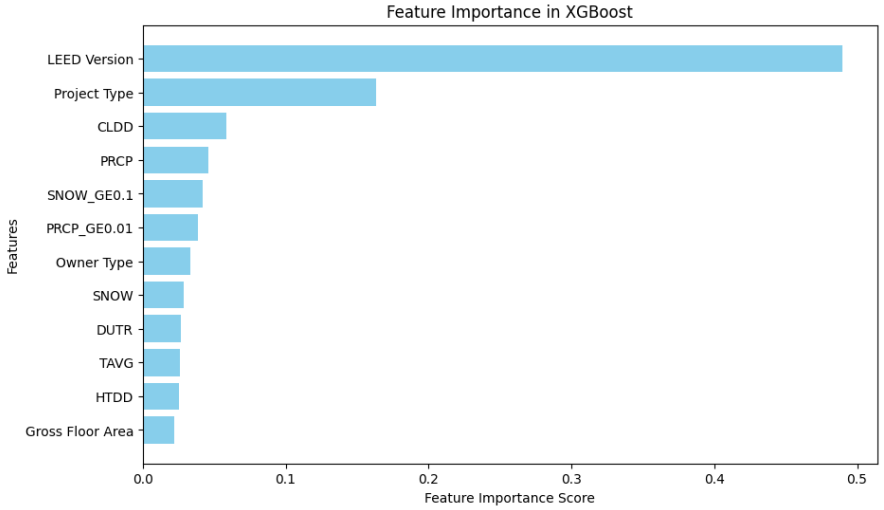


Figure 9. The feature importance scores in XGBoost.

4. Discussion

In this paper, we implemented the three ML models and found that two of them, including DT and XGBoost, which is also a decision-tree-based model, gave better results than the SVR model, which struggles with large datasets due to high computational cost. Besides the performance metrics, which showed the better performance of the XGBoost, the other features of this model, like handling missing values and providing inherent

feature importance scores, which make it more interpretable, led us to continue with this model for further analysis. Hypertuning the parameters of the XGBoost model resulted in a well-fitted model for our case, which showed relatively low errors.

The LEED version emerges as the most important factor, suggesting that variations in scoring criteria and policies across different versions significantly impact the points achieved by projects. Although the points required for each certification level (Certified, Silver, Gold, Platinum) are defined within each version, differences in credit requirements and weighting across versions may contribute to substantial variations in project scores. This highlights the evolving nature of LEED standards and their influence on certification outcomes. The second most important factor is the project type, which demonstrates that different types of projects do not face the same conditions in their pursuit of LEED points. This may be because of differences in the overall budget dedicated to various types or differences in the requirements they must meet to achieve certification. The next four features are climate variables, which underscore the climate's significant role in obtaining LEED points. Interestingly, CDD is the most effective climate variable, while TAVG, another variable related to temperature, does not seem very important relatively. It shows that, in regions with hot summers and high CDD, some LEED credit categories have become more critical and more challenging to achieve. PRCP as the second important climate factor also seems reasonable because some credits are related directly to the amount of precipitation, like credit associated with controlling the stormwater, which makes more sense to be achieved in regions with high PRCP [68].

Compared to previous studies, this paper benefits from a large dataset encompassing a broader range of project types, regions, and climates. This, indeed, makes the results more reliable. However, this study still has some limitations. First, the budget or cost parameter is not considered a feature in this study, which absolutely is a significant factor affecting the achievement of LEED credits. Second, this study did not investigate LEED credit categories independently as target variables, which would have made the results more interpretable. Future research could expand the model presented in this study to target LEED credit categories. For instance, one key category in the LEED rating system is Energy and Atmosphere, which emphasizes optimizing energy performance [12]. This presents opportunities to utilize energy optimization and sustainability tools, such as load forecasting techniques [69–71] and prescriptive maintenance systems [72], to enhance the selection and prediction of LEED credits.

5. Conclusions

This study demonstrates the effectiveness of machine learning in predicting LEED credit attainment, with XGBoost outperforming DT and SVR due to its superior handling of large datasets, missing values, and interpretation of feature importance. This analysis reveals that LEED version and project type were the most influential factors, highlighting variations in certification criteria and project-specific challenges. Additionally, climate variables, particularly CDD and PRCP, played a significant role, emphasizing the impact of regional climate conditions on sustainability outcomes.

These insights can inform project planners and policymakers in tailoring sustainability strategies based on project attributes and environmental conditions. Moreover, the model can assist decision-makers in optimizing investments by identifying the most impactful strategies for achieving LEED certification, whether by constructing new sustainable buildings or by enhancing existing ones. Beyond guiding investment decisions, this research can benefit sustainability consultants and developers by providing data-driven recommendations on the most feasible LEED credits under specific project and climate conditions.

Additionally, it can contribute to developing targeted financial incentives and regulatory policies that promote green building practices more effectively.

Despite the robustness of the model and the comprehensive dataset used, certain limitations remain. The absence of budget-related factors may overlook critical financial constraints affecting LEED credit attainment. Moreover, analyzing LEED credit categories separately as target variables could provide deeper insights into how specific credits are influenced by project and climate attributes. Future research should incorporate economic considerations and a more granular examination of credit categories to enhance the interpretability and applicability of machine learning models in sustainable building certification.

Author Contributions: Conceptualization, A.E.; methodology, A.M.; formal analysis, A.M.; resources, A.E.; data curation, A.M.; writing—original draft preparation, A.M., M.N. and A.E.; writing—review and editing, A.E.; visualization, M.N. and A.M.; supervision, A.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data for this study were collected from publicly available data sources. The LEED project directory, provided by the U.S. Green Building Council (USGBC), includes information on LEED-certified projects worldwide and was accessed from their official website. Climate data were obtained from the National Centers for Environmental Information (NCEI), which provides U.S. Climate Normal based on a uniform 30-year period from approximately 15,000 U.S. weather stations.

Acknowledgments: The authors would like to acknowledge the support of the Department of Civil, Environmental, and Geospatial Engineering (CEGE) at Michigan Technological University for providing resources to conduct this study.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. International Energy Agency (IEA). Buildings and Construction: Tracking Clean Energy Progress. Available online: https://www.iea.org/energy-system/buildings (accessed on 2 February 2025).
2. Atmoko, D.; Susilawati, C.; Devi, B.; Wilkinson, S.; Puspitarini, I.; Lukito, J.A.; Goonetilleke, A. Challenges and Opportunities for Promoting Sustainability in Public Buildings. *Sustainability* **2025**, *17*, 403. [CrossRef]
3. Tămaş-Gavrea, D.R.; Iştoan, R.; Tiuc, A.E.; Dénes, T.O.; Manea, D.L.; Ciorîță, A.; Vasile, O. Sustainable Approach to Achieve New Green Solutions for the Construction Industry. *Sustainability* **2024**, *16*, 10556. [CrossRef]
4. AuOlu-Ajayi, R.; Alaka, H.; Egwim, C.; Grishikashvili, K. Comprehensive Analysis of Influencing Factors on Building Energy Performance and Strategic Insights for Sustainable Development: A Systematic Literature Review. *Sustainability* **2024**, *16*, 5170. [CrossRef]
5. Darko, A.; Zhang, C.; Chan, A.P. Drivers for Green Building: A Review of Empirical Studies. *Habitat Int.* **2017**, *60*, 34–49. [CrossRef]
6. Voland, N.; Saad, M.M.; Eicker, U. Public Policy and Incentives for Socially Responsible New Business Models in Market-Driven Real Estate to Build Green Projects. *Sustainability* **2022**, *14*, 7071. [CrossRef]
7. Gurgun, A.P.; Arditi, D. Assessment of Energy Credits in LEED-Certified Buildings Based on Certification Levels and Project Ownership. *Buildings* **2018**, *8*, 29. [CrossRef]
8. USGBC. Benefits of Green Building. Available online: http://www.usgbc.org/projects (accessed on 2 February 2025).
9. Fowler, K.M.; Rauch, E.M.; Henderson, J.W.; Kora, A.R. *Re-Assessing Green Building Performance: A Post Occupancy Evaluation of 22 GSA Buildings*; No. PNNL-19369; Pacific Northwest National Laboratory (PNNL): Richland, WA, USA, 2010.
10. Rastogi, A.; Choi, J.K.; Hong, T.; Lee, M. Impact of Different LEED Versions for Green Building Certification and Energy Efficiency Rating System: A Multifamily Midrise Case Study. *Appl. Energy* **2017**, *205*, 732–740. [CrossRef]

11. Wu, P.; Song, Y.; Shou, W.; Chi, H.; Chong, H.Y.; Sutrisna, M. A Comprehensive Analysis of the Credits Obtained by LEED 2009 Certified Green Buildings. *Renew. Sustain. Energy Rev.* **2017**, *68*, 370–379. [CrossRef]
12. Guide to LEED Certification. Available online: https://www.usgbc.org/guide-LEED-certification (accessed on 2 February 2025).
13. Cheng, J.C.; Ma, L.J. A Data-Driven Study of Important Climate Factors on the Achievement of LEED-EB Credits. *Build. Environ.* **2015**, *90*, 232–244. [CrossRef]
14. Alothaimeen, I.; Arditi, D.; Türkakın, O.H. Multi-Objective Optimization for LEED-New Construction Using BIM and Genetic Algorithms. *Autom. Constr.* **2023**, *149*, 104807. [CrossRef]
15. Ma, J.; Cheng, J.C. Data-Driven Study on the Achievement of LEED Credits Using Percentage of Average Score and Association Rule Analysis. *Build. Environ.* **2016**, *98*, 121–132. [CrossRef]
16. Jun, M.A.; Cheng, J.C. Selection of Target LEED Credits Based on Project Information and Climatic Factors Using Data Mining Techniques. *Adv. Eng. Inf.* **2017**, *32*, 224–236. [CrossRef]
17. Leite Ribeiro, L.M.; Piccinini Scolaro, T.; Ghisi, E. LEED Certification in Building Energy Efficiency: A Review of Its Performance Efficacy and Global Applicability. *Sustainability* **2025**, *17*, 1876. [CrossRef]
18. Li, Y.; Li, X.; Ma, D.; Gong, W. Exploring green building certification credit selection: A model based on explainable machine learning. *J. Build. Eng.* **2024**, *95*, 110279. [CrossRef]
19. Altomonte, S.; Schiavon, S.; Kent, M.G.; Brager, G. Indoor environmental quality and occupant satisfaction in green-certified buildings. *Build. Res. Inf.* **2019**, *47*, 255–274. [CrossRef]
20. Madanayake, H.; Ruwanpura, J. Special purpose simulation for sustainable development: The LEED cost simulation model. In Proceedings of the 2nd International Construction Specialty Conference, St. John's, NL, Canada, 27–30 May 2009. CSCE.
21. Mansouri, A.; Taghaddos, H.; Tak, A.N.; Sadatnya, A.; Aghajamali, K. Simulation-Based Planning of Earthmoving Equipment for Reducing Greenhouse Gas (GHG) Emissions. *Autom. Constr.* **2024**, *168*, 105841. [CrossRef]
22. Syal, M.; Mago, S.; Moody, D. Impact of LEED-NC projects on constructors. *J. Archit. Eng.* **2007**, *13*, 174–179. [CrossRef]
23. Pham, D.H.; Kim, B.; Lee, J.; Ahn, A.C.; Ahn, Y. A Comprehensive Analysis: Sustainable Trends and Awarded LEED 2009 Credits in Vietnam. *Sustainability* **2020**, *12*, 852. [CrossRef]
24. Choi, J.O.; Bhatla, A.; Stoppel, C.M.; Shane, J.S. LEED Credit Review System and Optimization Model for Pursuing LEED Certification. *Sustainability* **2015**, *7*, 13351–13377. [CrossRef]
25. Madson, K.; Franz, B.; Leicht, R.; Nelson, J. Evaluating the Sustainability of New Construction Projects over Time by Examining the Evolution of the LEED Rating System. *Sustainability* **2022**, *14*, 15422. [CrossRef]
26. Rebelatto, B.G.; Salvia, A.L.; Brandli, L.L.; Leal Filho, W. Examining Energy Efficiency Practices in Office Buildings through the Lens of LEED, BREEAM, and DGNB Certifications. *Sustainability* **2024**, *16*, 4345. [CrossRef]
27. Wu, P.; Mao, C.; Wang, J.; Song, Y.; Wang, X. A Decade Review of the Credits Obtained by LEED v2.2 Certified Green Building Projects. *Build. Environ.* **2016**, *102*, 167–178. [CrossRef]
28. Pushkar, S. Impact of project size on LEED-EB V4 credit achievement in the United States. *J. Archit. Eng.* **2021**, *27*, 04021012. [CrossRef]
29. Navaratnam, S. Selecting a Suitable Sustainable Construction Method for Australian High-Rise Buildings: A Multi-Criteria Analysis. *Sustainability* **2022**, *14*, 7435. [CrossRef]
30. Darko, A.; Chan, A.P.C.; Ameyaw, E.E.; Owusu, E.K.; Pärn, E.; Edwards, D.J. Review of Application of Analytic Hierarchy Process (AHP) in Construction. *Int. J. Constr. Manag.* **2019**, *19*, 436–452. [CrossRef]
31. Liu, X.; Wang, W.; Ding, Y.; Wang, K.; Li, J.; Cha, H.; Saierpeng, Y. Research on the Design Strategy of Double-Skin Facade in Cold and Frigid Regions—Using Xinjiang Public Buildings as an Example. *Sustainability* **2024**, *16*, 4766. [CrossRef]
32. Seyis, S.; Ergen, E. A decision-making support tool for selecting green building certification credits based on project delivery attributes. *Build. Environ.* **2017**, *126*, 107–118. [CrossRef]
33. Madanayake, H.L.; Ruwanpura, J. Multi-Criteria Decision Making to Improve Performance in Construction Projects with LEED Certification. In Proceedings of the Construction Research Congress 2012: Construction Challenges in a Flat World, West Palm Beach, FL, USA, 21–23 May 2012; pp. 1899–1909. [CrossRef]
34. Erfani, A.; Tavakolan, M. Risk Evaluation Model of Wind Energy Investment Projects Using Modified Fuzzy Group Decision-Making and Monte Carlo Simulation. *Arthaniti: J. Econ. Theory Pract.* **2023**, *22*, 7–33. [CrossRef]
35. Attallah, S.O.; Kandil, A.; Senouci, A.; Al-Derham, H. Multicriteria decision-making methodology for credit selection in building sustainability rating systems. *J. Archit. Eng.* **2017**, *23*, 04017004. [CrossRef]
36. Cheng, J.C.P.; Ma, L.J. A non-linear case-based reasoning approach for retrieval of similar cases and selection of target credits in LEED projects. *Build. Environ.* **2015**, *93*, 349–361. [CrossRef]
37. Pan, Y.; Zhang, L. Roles of Artificial Intelligence in Construction Engineering and Management: A Critical Review and Future Trends. *Autom. Constr.* **2021**, *122*, 103517. [CrossRef]
38. Erfani, A.; Cui, Q. Predictive Risk Modeling for Major Transportation Projects Using Historical Data. *Autom. Constr.* **2022**, *139*, 104301. [CrossRef]

39. Koh, Y.; Ceylan, H.; Kim, S.; Cho, I.H. An Artificial-Intelligence-Based Approach for Predicting Structural Damages of Paved-Road Systems under Superloads. *Constr. Build. Mater.* **2024**, *411*, 134257. [CrossRef]
40. Mansouri, A.; Erfani, A. Gender Disparity and Self-Presentation on Social Media among AEC Industry Leaders. *J. Manag. Eng.* **2024**, *40*, 04024042. [CrossRef]
41. Erfani, A.; Cui, Q.; Cavanaugh, I. An Empirical Analysis of Risk Similarity among Major Transportation Projects Using Natural Language Processing. *J. Constr. Eng. Manag.* **2021**, *147*, 04021175. [CrossRef]
42. Lee, G.; Chang, T.; Chi, S. Data-Driven Bridge Maintenance Cost Estimation Framework for Annual Expenditure Planning. *J. Manag. Eng.* **2024**, *40*, 04023068. [CrossRef]
43. Debrah, C.; Chan, A.P.; Darko, A. Artificial Intelligence in Green Building. *Autom. Constr.* **2022**, *137*, 104192. [CrossRef]
44. Li, J.; Liu, Z.; Han, G.; Demian, P.; Osmani, M. The Relationship Between Artificial Intelligence (AI) and Building Information Modeling (BIM) Technologies for Sustainable Building in the Context of Smart Cities. *Sustainability* **2024**, *16*, 10848. [CrossRef]
45. Huang, H.; Dai, D.; Guo, L.; Xue, S.; Wu, H. AI and Big Data-Empowered Low-Carbon Buildings: Challenges and Prospects. *Sustainability* **2023**, *15*, 12332. [CrossRef]
46. Alghamdi, S.; Tang, W.; Kanjanabootra, S.; Alterman, D. Optimising Building Energy and Comfort Predictions with Intelligent Computational Model. *Sustainability* **2024**, *16*, 3432. [CrossRef]
47. National Centers for Environmental Information (NCEI). Available online: https://www.ncei.noaa.gov/cdo-web/ (accessed on 2 February 2025).
48. Maria, E.; Budiman, E.; Taruk, M. Measure Distance Locating Nearest Public Facilities Using Haversine and Euclidean Methods. *J. Phys. Conf. Ser.* **2020**, *1450*, 012080. [CrossRef]
49. Emmanuel, T.; Maupong, T.; Mpoeleng, D.; Semong, T.; Mphago, B.; Tabona, O. A Survey on Missing Data in Machine Learning. *J. Big Data* **2021**, *8*, 140. [CrossRef]
50. Lee, K.J.; Simpson, J.A. Introduction to Multiple Imputation for Dealing with Missing Data. *Respirology* **2014**, *19*, 162–167. [CrossRef] [PubMed]
51. Hancock, J.T.; Khoshgoftaar, T.M. Survey on Categorical Data for Neural Networks. *J. Big Data* **2020**, *7*, 28. [CrossRef]
52. Jia, B.B.; Zhang, M.L. Multi-Dimensional Classification via Sparse Label Encoding. In Proceedings of the International Conference on Machine Learning, PMLR, online, 18–24 July 2021; pp. 4917–4926.
53. Hickey, P.J.; Erfani, A.; Cui, Q. Use of LinkedIn Data and Machine Learning to Analyze Gender Differences in Construction Career Paths. *J. Manag. Eng.* **2022**, *38*, 04022060. [CrossRef]
54. Jha, K.K.; Jha, R.; Jha, A.K.; Hassan, M.A.M.; Yadav, S.K.; Mahesh, T. A Brief Comparison on Machine Learning Algorithms Based on Various Applications: A Comprehensive Survey. In Proceedings of the 2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 16–18 December 2021; pp. 1–5, IEEE.
55. Erfani, A.; Cui, Q. Natural Language Processing Application in Construction Domain: An Integrative Review and Algorithms Comparison. *Comput. Civ. Eng.* **2021**, 26–33.
56. Li, Y.; Yang, L.; Yang, B.; Wang, N.; Wu, T. Application of Interpretable Machine Learning Models for Intelligent Decision. *Neurocomputing* **2019**, *333*, 273–283. [CrossRef]
57. Erfani, A.; Frias-Martinez, V. A Fairness Assessment of Mobility-Based COVID-19 Case Prediction Models. *PLoS ONE* **2023**, *18*, e0292090. [CrossRef]
58. Wang, Z.; Wang, Y.; Srinivasan, R.S. A Novel Ensemble Learning Approach to Support Building Energy Use Prediction. *Energy Build.* **2018**, *159*, 109–122. [CrossRef]
59. Zhang, K.; Erfani, A.; Beydoun, O.; Cui, Q. Procurement Benchmarks for Major Transportation Projects. *Transp. Res. Rec.* **2022**, *2676*, 363–376. [CrossRef]
60. Mohammadi, P.; Rashidi, A.; Malekzadeh, M.; Tiwari, S. Evaluating Various Machine Learning Algorithms for Automated Inspection of Culverts. Eng. *Anal. Bound. Elem.* **2023**, *148*, 366–375. [CrossRef]
61. Probst, P.; Boulesteix, A.L.; Bischl, B. Tunability: Importance of Hyperparameters of Machine Learning Algorithms. *J. Mach. Learn. Res.* **2019**, *20*, 1–32.
62. Roy, A.; Chakraborty, S. Support Vector Machine in Structural Reliability Analysis: A Review. *Reliab. Eng. Syst. Saf.* **2023**, *233*, 109126. [CrossRef]
63. Hou, Y.; Liu, S. Predictive Modeling and Validation of Carbon Emissions from China's Coastal Construction Industry: A BO-XGBoost Ensemble Approach. *Sustainability* **2024**, *16*, 4215. [CrossRef]
64. Gou, L.; Zhang, J.; Wen, L.; Fan, Y. State Reliability of Wind Turbines Based on XGBoost–LSTM and Their Application in Northeast China. *Sustainability* **2024**, *16*, 4099. [CrossRef]
65. Rusdah, D.A.; Murfi, H. XGBoost in Handling Missing Values for Life Insurance Risk Prediction. *SN Appl. Sci.* **2020**, *2*, 1336. [CrossRef]

66. Plevris, V.; Solorzano, G.; Bakas, N.P.; Ben Seghier, M.E.A. Investigation of Performance Metrics in Regression Analysis and Machine Learning-Based Prediction Models. In Proceedings of the 8th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS Congress 2022), Oslo, Norway, 5–9 June 2022. European Community on Computational Methods in Applied Sciences. [CrossRef]
67. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-Generation Hyperparameter Optimization Framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631. [CrossRef]
68. Eissa, R.; El-Adaway, I.H. Circular Economy Integration in LEED: A Decade of Materials and Resources Credit Achievement Patterns. *J. Manag. Eng.* **2024**, *40*, 04024047. [CrossRef]
69. Ahmad, N.; Ghadi, Y.; Adnan, M.; Ali, M. Load Forecasting Techniques for Power System: Research Challenges and Survey. *IEEE Access* **2022**, *10*, 71054–71090. [CrossRef]
70. Zhang, G.; Guo, J. A Novel Ensemble Method for Residential Electricity Demand Forecasting Based on a Novel Sample Simulation Strategy. *Energy* **2020**, *207*, 118265. [CrossRef]
71. Nateghi, S.; Mansoori, A.; Moghaddam, M.A.E.; Kaczmarczyk, J. Multi-Objective Optimization of a Multi-Story Hotel's Energy Demand and Investing the Money Saved in Energy Supply with Solar Energy Production. *Energy Sustain. Dev.* **2023**, *72*, 33–41. [CrossRef]
72. Papaioannou, A.; Dimara, A.; Krinidis, S.; Anagnostopoulos, C.N.; Ioannidis, D.; Tzovaras, D. Advanced Proactive Anomaly Detection in Multi-Pattern Home Appliances for Energy Optimization. *Internet Things* **2024**, *26*, 101175. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.