2025 European Conference on Computing in Construction

CIB W78 Conference on IT in Construction

Porto, Portugal
July 14-17, 2025

RENOVATION COST PREDICTION AT AN EARLY STAGE: AN EXPERIMENT USING MACHINE LEARNING REGRESSION AND CLASSIFICATION TECHNIQUES

Svenja Lauble[1], Bin Wu[1], Katharina Eisen[1], and Hendrik Seibel[2]
[1]Karlsruher Institute of Technology, Karlsruhe, Germany
[2]PLAN4 Software GmbH, Freiburg, Germany

Abstract

Accurate early-stage cost forecasting in renovation projects is challenging due to their inherent complexity and lack of high-quality data. This study evaluated 104 cases using regression and classification models. Regression models outperformed classification, achieving a mean absolute percentage error (MAPE) of 7% versus 15%. Incorporating building-specific data such as roof area from sources like Google Earth further improves prediction accuracy. The findings highlight the value of integrating external data with regression models to support data-driven decisions. This work addresses a gap in literature by focusing exclusively on renovation projects and offers strategies to address data limitations in this domain.

Introduction

The significance of sustainability in construction planning and realization, particularly in renovation projects, has increased markedly. Sustainability is assessed from social, ecological, and economic perspectives, which are essential for a comprehensive evaluation of construction projects (Kleine, 2009). Social sustainability faces challenges such as housing shortage and high construction interest rates, leading to a decline in new construction projects. At the same time, around 70% residences in Germany require renovation measures (Fraunhofer Institut für Bauphysik, 2024). Ecologically, renovation projects can significantly reduce CO2 emissions by integrating energy-efficient heating methods, with buildings accounting for 30% of Germany's CO2 emissions in 2021 (Umweltbundesamt, 2023). Economically, reducing CO2 emissions can financially benefit building operators, with early cost forecasting being critical for efficient decision-making and minimizing project costs and risks.

However, the intricacy and unpredictability of renovation projects are considerably greater than those of new constructions, due to the necessity of planning measures based on existing structures. Consequently, the uncertainty of early-stage cost estimation is particularly high, which necessitates robust methods describing the building stock.

To date, existing research has largely focused on cost prediction for new construction (Prasetyono et al., 2021; Alshemosi and Alsaad, 2017) or specialized areas like seismic retrofits (FEMA, 1994; Fung et al., 2017; Mirzaei et al., 2024), often relying on single case studies with limited generalizability (Ali et al., 2018; Islam et al., 2019; Akintoye, 2000; Hatamleh et al., 2018). This leaves a significant gap in the development of generalized and data-driven approaches for predicting renovation costs, particularly in the German context.

This research aims to address this gap by evaluating and applying two distinct methods—regression and classification—for predicting renovation costs as decision support tools during the early phases of construction projects in Germany. Specifically, it addresses the research question: How does the use of classification and regression in predicting renovation costs affect the accuracy and trust in forecasts as a basis for decision-making in renovation measures?

This study makes the following contributions:

- Methodological innovation: It compares regression and classification approaches for renovation cost prediction, emphasizing their impact on accuracy and decision-making trust.
- Practical framework: It develops a conceptual model to support property owners in making informed decisions during the early planning stages.
- Empirical validation: It uses a robust experimental dataset, offering generalizable insights into renovation projects beyond single-case studies.
- Sustainability and economic alignment: It demonstrates how high-quality data and accurate forecasting contribute to reducing CO2 emissions and achieving cost efficiency, aligning with broader environmental and energy goals.

With Germany's aging housing stock and the urgent need for sustainable retrofits, this research provides timely and actionable solutions. By advancing cost prediction methodologies, it supports property owners, consultants, and policymakers in bridging the gap between economic efficiency and sustainability, driving impactful outcomes in renovation planning and realization.

Overall, our contributions are as follows. First, the background and existing literature on cost prediction is reviewed. Second, the experimental setup to introduce the dataset and implementation details as well as the evaluation of the results for both strategies are described. Third, we discuss the results and create a conceptual model for property owners to drive added business-value in an accurate early-phase prediction of retrofit costs.

Background and Related Work

The following section will elaborate on the definition of renovation measures, the relevant features that must be considered in any analysis of renovation measures, and the relevant works predicting construction costs.

Renovation Measures

Renovation is often understood as construction measures on existing buildings, without specifying the type or complexity (Bundesministerium des Innern, für Bau und Heimat, 2019). Neddermann (2007) defines renovation as "sustainable repair and comprehensive modernization," involving extensive measures like repairs and improvements. Modernizations aim to increase market value (HOAI, 2021, p. 5), aligning with renovation as an enhancement. Renovations may also involve compliance with regulations such as the Building Energy Act (GEG). While often overlapping with modernization, the term "building renewal measure" (Neddermann, 2007) includes renovation, expansion, and other related concepts. This study defines renovation as comprehensive construction measures that restore or significantly alter a building's functions, often including energy-related improvements.

Influences on Renovation Costs

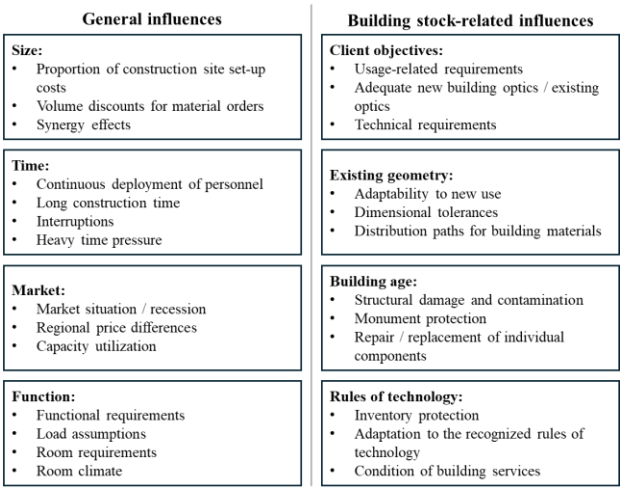

Figure 1: Influences on the costs of renovation projects

Unlike new constructions, renovations must account for varying building conditions, undocumented modifications, and unforeseen structural issues. Therefor renovation cost prediction models struggle with data availability due to the unique challenges of renovation projects. Factors such as structural and material conditions or regulatory considerations significantly impact the necessary measures and consequently the renovation costs (Friedrichsen, 2024, pp. 40-41). Bielefeld and Wirths (2010, p. 234) categorize cost-influencing factors into general and building-specific influences (see Figure 1). While the list is not exhaustive, it represents numerous influences that particularly affect existing buildings, thereby highlighting the increased complexity and uncertainty of renovation projects.

Therefore, accurately assessing the condition of the building is crucial for economic planning (Friedrichsen, 2024, p. 42). According to Neddermann (2007, pp. 54-55), this assessment is particularly challenging and requires high expertise. Fundamental damages often become apparent only as the renovation project progresses and are not visible at the outset. Since damaging the building fabric is generally undesirable, specific cost risks emerge that can only be minimized later in the project (Friedrichsen, 2024, p. 43).

Analytical Models for Construction Cost Predictions

Analytics, as defined by Davenport and Harris (2007), involves using data, statistical analysis, models, and fact-based management to guide decisions. Due to complexity and uncertainty, integrating analytical models with professional expertise is crucial for forecasting renovation costs. O'Neill (2022) identifies three AI interaction levels:

1. Decision Support: AI provides visualizations and hints, while the expert makes predictions.
2. Decision Augmentation: AI offers predictive forecasts, with both expert and AI contributing to decision-making.
3. Decision Automation: AI autonomously makes decisions, with expert intervention for risks.

The following literature review aims to classify existing scientific models for cost estimation in renovation projects into these categories.

The literature review used Booth et al.'s (2012) method: defining search terms for renovation cost prediction and planning methods, searching databases for 56 relevant sources, assessing quality based on as "title," "year of publication," and "key sources/authors", conducted a forward and backward citation search to add 10 sources, and categorizing literature into analytical and predictive methods. This approach ensured a comprehensive exploration of methods for subsequent analysis.

For decision support the review detailed various statistical analyses commonly used to identify factors influencing renovation costs. Methods such as the Spearman correlation coefficient (Ali et al., 2018; Islam et al., 2019), ANOVA (Akintoye, 2000), and Kendall rank correlation (Hatamleh et al., 2018) were highlighted for their utility in discerning significant differences among cost factors. Studies like those by Ali et al. (2018) emphasized the impact of operational and comfort-related elements on cost deviations in Malaysian renovation projects, using tools like the Relative Importance Index (RII) to prioritize cost factors (Islam et al., 2019). Other studies, such as

those by Hassanain et al. (2013, 2019; Ibrahim & Elshwadfy, 2021), explored factors such as building age and size, identifying moderate influences on project costs across different project types (e.g., office buildings, schools, hospitals).

Decision Augmentation models include extrapolation techniques progressed from statistical analyses to predictive models aiming to forecast future scenarios. Regression analysis, encompassing methods like simple linear, multiple, or polynomial regression (Mittag & Schüller, 2020). Studies such as Prasetyono et al. (2021), Alshemosi & Alsaad (2017), Alshamrani (2017) or Alshboul (2022) concentrate on predicting costs for new buildings. Further studies concentrate on seismic renovations. These studies integrated factors such as existing building characteristics and seismicity (FEMA, 1994; Fung et al., 2017; Mirzaei et al., 2024), demonstrating varied approaches to model complexity and data interpretation. Machine learning techniques, including Artificial Neural Networks (ANN) (Elhag & Boussabaine, 1998), Decision Trees (Khodabakhshian et al., 2023), and Case Based Reasoning (Craw, 2017), were explored under predictive analyses. For instance, Elhag and Boussabaine (1998) used ANNs to predict construction costs, incorporating factors such as site slope, access conditions, and contract type, achieving an accuracy of up to 82.2%. Similarly, Khodabakhshian et al. (2023) compared decision trees with ANNs for seismic renovation projects, suggesting that ANNs perform better with large datasets, while decision trees provide more interpretable results. However, these studies have focused on either new buildings or seismic renovation, leaving a gap in predictive modeling specifically tailored to general renovation cost estimation.

Decision augmentation models can enable decision automation, depending on forecast accuracy. For cost forecasting and optimization, models like ant colony algorithms, genetic algorithms (Dorigo & Birattari, 2017; Sammut, 2017), and simulated annealing (Aarts et al., 2005) optimize costs and project durations through iterative improvement, though they are not suited for early project phases.

Further on, practical tools can be named to predict renovation costs, primarily focusing on condition assessments. These methods typically involve assigning condition classes to elements, which are then statistically correlated with cost metrics based on quantity measurements. Notable examples include the Swiss Impulse Program (IP Bau, 1995) and the EPIQR software (Fraunhofer Institute for Building Physics, 1999), each employing distinct approaches tailored to high-cost components.

To summarize, existing models in renovation cost prediction struggle with data availability due to the unique challenges of renovation projects. Unlike new constructions, renovations must account for varying building conditions, undocumented modifications, and unforeseen structural issues. While new buildings and seismic renovations have been studied, general renovations lack structured datasets, leading to inconsistent predictions. While decision trees have been applied in some cost prediction studies, their potential for early-stage renovation forecasting remains underexplored. Existing research, such as Fung et al. (2020), highlights decision trees' balance between accuracy and interpretability but primarily in seismic renovations. Our approach addresses these gaps by integrating external data sources, such as Google Earth, to supplement missing parameters and improve prediction accuracy. By applying decision trees, we enhance explainability and adaptability (Lauble, 2022), making cost forecasts more transparent and reliable for diverse renovation scenarios.

Experiment

The following section describes the experimental setup, including the introduction of the dataset and details about the implementation. Finally, the experimental results for the classification and regression are presented.

Problem Description

The study addresses the problem of accurately predicting renovation costs at an early project phase, focusing on comparing regression and classification models. Regression predicts continuous renovation costs, while classification categorizes projects into cost ranges. This dual approach aims to explore which approach can offer useful insights for strategic decision-making.

Data Description

The data consists of 104 renovation projects from PLAN4 Software GmbH (2024), including cost-related features at different levels of granularity. The data was segmented into three datasets (I: 24 projects, II: 53 projects, III: 101 projects) to analyze the value of incremental data collection. Each project includes 14 features:

- Project ID (which is eliminated after calculating the total retrofit costs of all measures)
- Postal code
- Year of construction
- Cost group (according to DIN 276)
- Element group (according to 276; e.g., room doors, flooring)
- Building element position (according to 276; e.g., coated doors, ceramic flooring)
- Condition (physical condition of the examined element group; A (best) to D(worst))
- Standard service book (StLB), describes standardized construction services for tendering, awarding and invoicing.
- Measure group (according to StLB; e.g., heating surfaces, surface coverings)
- Measure (according to StLB; e.g., panel radiators, linoleum flooring)
- Quantity
- Gross floor area [m²]
- Living space [m²]
- Unit price [€]

In assessing the condition, the inspector's role is neutral and overarching, based on extensive experience of numerous refurbishment projects, ensuring an objective assessment independent of financial or ownership interests. Unlike asset owners or capital analysts, who may have biases based on investment goals, inspectors rely on standardized evaluation criteria derived from prior projects to provide consistent and impartial assessments.

Key attributes such as "cost group" and "element group" were originally assigned at the object level within buildings rather than at the project level. To ensure consistency in modeling, we aggregated these attributes for entire projects by computing frequency distributions and dominant classifications across building elements.

Additional external data, such as roof area in square meters and building height in meters, were incorporated for enhanced predictive accuracy, extracted using Google Earth (2024) and Google Solar API (2024). As this data was not available for all buildings, this results in a reduced dataset B. On contrast dataset A describes the original dataset without additional external data.

Renovation costs as target variable are calculated by the product of the unit price and quantity. The study's focus is on the early prediction of renovation costs, defined as total building costs (sum of cost groups 300 and 400 according to DIN 276:2018-12, containing costs about building structures and technical installations) necessary for all renovation measures. Possible outliers have been removed.

Key parameters describing the datasets are shown in Table 1. Detailed descriptions pertain to Dataset III A, encompassing projects across various building types and ages (1908-2016). Of these, 39 can be allocated to school renovation projects, 14 to (sports) hall renovation projects, 25 to housing projects and nine to fire stations, while 17 cannot be allocated.

Table 1: Overview of the datasets

	I A	I B	II A	II B	III A	III B
Number of projects	24	17	53	31	101	41
Number of features	14	16	14	16	14	16
Avg. number of renovation measures	234	256	190	202	121	167
Avg. costs per renovation project [Mio. €]	1.400	1.232	0.895	0.926	0.491	0.768
Standard deviation of renovation costs [Mio. €]	0.731	0.433	0.751	0.503	0.596	0.532

Data Processing

Data preparation is crucial for effective modeling, encompassing several key steps:

- Outlier Handling: Identify and possibly remove outliers to improve modeling accuracy (Géron, 2022).
- Missing Values: Remove rows with missing values to ensure robust modeling (Géron, 2022).
- One-Hot-Encoding & Label Encoding: Convert categorical data into numeric form. Label Encoding assigns each category a unique number, while One-Hot-Encoding creates new columns for each category (Müller & Guido, 2017). Mapping is used for ordinal variables like the "Condition" feature.
- Feature Correlation: A correlation analysis revealed limited linear relationships among features, justifying the inclusion of all variables to capture potential non-linear dependencies.
- Standardization: Standardization of selected data can enhance results in linear regression by adjusting feature scales to be standard normally distributed, with a mean of 0 and standard deviation of 1 (Scikit-Learn Developers, 2024b). This adjustment ensures that high numerical values do not disproportionately influence the model compared to lower values. In contrast, decision tree models typically do not require standardization (Filho, 2023).

Algorithm Implementation

Decision trees were chosen for their high explainability, accuracy and ability to model non-linear relationships. For modeling, Python is employed using Microsoft Visual Studio Code as the code editor, supporting multiple programming languages (Van Rossum & Drake, 1995). Essential libraries include Pandas (McKinney, 2010), Numpy (Harris et al., 2020), Matplotlib (Hunter, 2007), Scikit-learn (Pedregosa et al., 2011), Seaborn (Waskom, 2021), and Optuna (Akiba et al., 2019), which facilitates hyperparameter optimization and optimal dataset selection considering outliers and model choices.

Further on regression and classification trees were pruned using minimum samples per leaf and maximum depth. Hyperparameters (e.g., max depth, split criteria) were tuned using Optuna for optimal performance. The pruning strategy differed slightly: classification models used smaller bucket sizes to enhance class separation, while regression models prioritized minimizing residual errors.

The Mean Absolute Percentage Error (MAPE) is used for ensuring the comparability of all models across projects. Data is split with 70% for training and 30% for testing across all models, with cross-validation (k = 5) performed for robust estimation.

Post Processing

Feature importance scores were extracted to interpret the relative influence of each variable. Further on, the decision tree structures were visualized to assess explainability and compare the interpretability of regression and classification models.

Results of the Regression

In the experiment, renovation costs for individual projects are predicted using linear regression, Random Forest, CatBoost, and Gradient Boosting models. These machine

learning models were selected based on the experience and findings in the study by Lauble (2022), which highlighted their effectiveness in predicting renovation costs. The datasets are enriched with external data and utilized for modeling. The outlier-excluded results are detailed in Table 2 which generally yields better average results.

Particularly, results from Random Forest and CatBoost models show close similarity, with Gradient Boosting slightly lagging, notably visible in Datasets I to III A. The inclusion of external data also generally improves outcomes, especially evident in Random Forest and CatBoost models, albeit with a reduced number of projects compared to original datasets (see Table 1).

Table 2: MAPE values of the datasets for regression (in percentage, best values in bold)

	I A	I B	II A	II B	III A	III B
Linear regressions	153	364	371	255	275	1,010
Random Forest	**129**	**7**	48	**25**	**27**	24
CatBoost	133	13	48	42	44	**23**
Gradient Boosting	168	26	**42**	45	51	32

Further on, Dataset III results stand out with only slight improvements in accuracy while increasing the number of included projects. This phenomenon may be attributed to the relatively low average costs and the nearly constant standard deviation (see Table 1).

Results of the Classification

To analyze the data based on classification, the k-nearest neighbor algorithm (kNN) was employed. The kNN was employed to compare ten randomly selected projects with others based on similar parameters, such as gross floor area and measures. For each selected project, the most similar one is identified, and the corresponding error values are calculated using MAPE across all ten projects. To enhance accuracy, only school projects were initially selected from the available data to increase the similarity of the projects and reduce the MAPE. The following Table 3 presents the results of this algorithm. In a subsequent step, the projects can be categorized into cost classes and combined with regression to further refine and validate the forecast results.

Table 3: MAPE values of the datasets for classification (best value in bold, in percentage)

	I A	I B	II A	II B	III A	III B
kNN	30	**15**	57	32	79	77

Discussion

The performance of regression and classification models highlights significant differences, with regression models consistently outperforming classification models in terms of MAPE. The lowest MAPE (7%) was achieved using regression, compared to 15% for classification. These results demonstrate the suitability of regression for early-stage cost prediction. The choice of MAPE as the primary metric is justified by its ability to provide a relative measure of prediction accuracy, which is critical for comparing projects of varying scales. However, for classification models, the use of MAPE may not fully capture performance nuances, and future studies could incorporate metrics like accuracy or F1-score for a more comprehensive evaluation.

Errors in the results can be attributed to several factors. The limited size of the datasets, particularly when external data is integrated, restricts the generalizability of the models of Dataset B. Additionally, the lack of clear project type assignments introduces variability, complicating the classification task. For regression models, the nearly constant standard deviation in Dataset III likely explains the minimal improvement in accuracy despite the larger dataset. Gradient Boosting's slight underperformance compared to Random Forest and CatBoost may stem from its higher sensitivity to hyperparameter tuning, indicating the need for further optimization.

Despite the strengths of our approach, several limitations should be acknowledged. First, the reliance on external data results in a reduced dataset size, which may affect the generalizability of the model. In addition, the limited quality of the available data, coupled with the absence of key numerical features, poses a challenge to predictive accuracy.

The classification approach used in this study, while promising for identifying similar projects, also faces problems due to the inherent variability of the data. This variability, together with the higher MAPE associated with bucket sizes, affects the overall reliability of the model. However, with an additional hyperparameter tuning process, classification models might still offer value by improving their predictive performance and adaptability to renovation cost forecasting.

In addition, the regional focus of the study on Germany is a potential limitation in terms of the global applicability of our findings. Refurbishment costs and their drivers can vary significantly between countries and regulatory environments, which may affect the applicability of the results to other regions.

The analysis in this study was also conducted within a static framework that did not consider dynamic factors such as inflation, material availability and labor costs. These factors are known to influence refurbishment costs but were excluded due to data limitations and the broader focus of the study. Future research could improve the robustness of cost models by incorporating these dynamic variables, possibly through real-time economic data or by modelling cost fluctuations over time. Such an approach would provide a more nuanced understanding of how these factors interact to shape refurbishment costs under different economic conditions.

Finally, while decision trees provide interpretability, industry professionals may need more detailed model

explanations. Techniques such as feature importance analysis or SHAP scores could add clarity.

Despite these challenges, the results underscore the importance of data quality and feature integration. The inclusion of external features like roof area and building height demonstrates potential for improved accuracy, though at the cost of reduced dataset size. These findings suggest that enhancing data collection practices and implementing robust data management systems are critical for advancing predictive modeling in renovation cost forecasting. By pre-filtering projects based on types, feature collection, analysis efficiency, and comparability can be significantly improved. Further on, collecting more numerical features can also support improving the data quality.

Path to Deployment

To assess the reliability and relevance of the results, the project team consulted PLAN4 experts, who specialize in advising building owners on the planning of renovation measures and associated costs. Through a series of workshops, a target of no more than a 20% deviation between predicted and actual costs was established. The regression model was close to achieving this goal, but further steps were identified to improve both the accuracy of the predictions and the confidence in the results.

In response to these findings, it was concluded that a combination of approaches was needed to support expert decision making. Specifically, the classification approach was integrated to assist in the evaluation and refinement of predicted values. This approach serves as an essential tool for filtering and evaluating potential remediation projects. Users can customize the display by selecting a specific number of comparable projects, sorting by accuracy metrics (e.g., MAPE), and filtering based on project types to identify the most relevant projects for comparison.

A key component of the deployment process is to further improve data quality to ensure that forecasts become more reliable over time. To achieve this, a mock-up was designed combining both regression and classification methods, as shown in Figure 2. This mock-up aims to translate the theoretical framework and data-driven insights into a practical, user-friendly interface. The goal is to incorporate this mock-up into the software, allowing building owners and professionals to make more informed decisions regarding renovation planning. The mock-up serves as an initial step in realizing the system's deployment within the software, ensuring that the final implementation meets user needs and enhances decision-making. The classification approach is primarily used to evaluate the predicted values and works in conjunction with the regression model to ensure more accurate and contextually relevant predictions.

The next steps in the deployment path focus on analyzing the implementation of this integrated pipeline and identifying potential adjustments to further improve forecast accuracy. These actions are intended to improve both the performance of the model and the confidence that building owners and professionals have in the system, ultimately supporting more informed decisions in renovation planning.

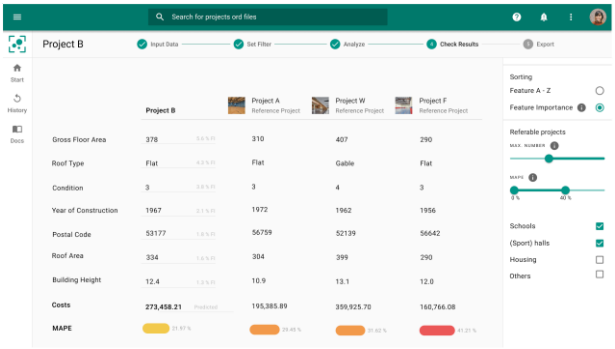

Figure 2: Combined view of classification and regression to predict renovation costs.

Conclusions

This paper evaluates two approaches to predict renovation costs for building construction projects at an early stage. The first approach uses regression analysis to predict costs, while the second involves identifying referable projects through classification. Our experiments demonstrate that the regression approach achieves a better mean absolute percentage error (MAPE) value, though data quality and quantity limit its effectiveness. This study addresses a gap in literature by exclusively focusing on renovation costs, distinguishing it from broader literature that includes new construction and seismic renovations.

We propose a pipeline to integrate both approaches, enhancing prediction accuracy and trust as a decision-making aid. This contribution is particularly valuable for building owners and supports experts aiming to generate data-driven business value by forecasting renovation costs at an early stage. By addressing the named limitations and exploring these avenues, this study lays a foundation for advancing predictive modeling in the construction domain, contributing to more informed and data-driven decision-making processes.

Future research should focus on improving data quality through advanced imputation methods and increasing dataset size via collaborations or public data sources. Refining classification approaches, potentially combining them with regression, could enhance predictive accuracy. Adopting standardized data aggregation processes and exploring hybrid models, along with incorporating dynamic cost factors like inflation or material availability, may significantly improve model applicability and insights across different construction domains.

Acknowledgements

We express our gratitude to PLAN4 Software GmbH for facilitating and supporting this research. The authors acknowledge the financial support by the Federal Ministry for Economic Affairs and Climate Action of Germany in the project "Nachhaltige intelligente Sanierungsmaßnahmen" (project number 01MN23005).

References

Aarts, E., Korst, J., & Michiels, W. (2005) Simulated annealing. In: Search methodologies: introductory tutorials in optimization and decision support techniques, Springer, pp.187–210.

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019) Optuna: A Next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp.2623–2631.

Akintoye, A. (2000) Analysis of factors influencing project cost estimating practice. In: Construction Management & Economics, 18(1), pp.77–89.

Ali, A. S., Azmi, N. F., & Baaki, T. K. (2018) Cost performance of building refurbishment works: The case of Malaysia. In: International Journal of Building Pathology and Adaptation, 36(1), pp.41–62.

Alshamrani, O. S. (2017) Construction cost prediction model for conventional and sustainable college buildings in North America. In: Journal of Taibah University for Science, 11(2), pp.317.

Alshemosi, A. M. B., & Alsaad, H. S. H. (2017) Cost estimation process for construction residential projects by using multifactor linear regression technique. In: International Journal of Science and Research, 6(6), pp.151–156.

Alshboul, O. et al. (2022) 'Extreme Gradient Boosting-Based Machine Learning Approach for Green Building Cost Prediction', Sustainability, 14(11), p. 6651. Available at: https://doi.org/10.3390/su14116651.

Bielefeld, B., & Wirths, M. (2010) Development and implementation of construction projects in existing buildings [Entwicklung und Durchführung von Bauprojekten im Bestand]. Wiesbaden: Vieweg+Teubner Verlag, pp.234.

Booth, A., Papaioannou, D., & Sutton, A. (2012) Systematic approaches to a successful literature review. London: SAGE.

Bundesministerium des Innern, für Bau und Heimat (2019) Guidelines for sustainable building [Leitfaden Nachhaltiges Bauen]. Available 26 February 2024 at: https://www.nachhaltigesbauen.de/fileadmin/publikationen/BBSR_LFNB_D_190125, pp.134–135.

Craw, S. (2017) Case-based reasoning. In: C. Sammut & G.I. Webb, eds. Encyclopedia of machine learning and data mining, Springer US, pp.180–188.

Davenport, T. H., & Harris, J. G. (2007) Competing on analytics: The new science of winning. Boston, MA: Harvard Business School Press, pp.15(217).

Dorigo, M., & Birattari, M. (2017) Ant colony optimization. In: C. Sammut & G.I. Webb, eds. Encyclopedia of machine learning and data mining, Springer US, pp.37–40.

Elhag, T., & Boussabaine, A. (1998) An artificial neural system for cost estimation of construction projects. In: 14th Annual ARCOM Conference, 1, pp.219–226.

FEMA (1994) Typical costs for seismic rehabilitation of existing buildings.

Filho, M. (2023) sklearn.ensemble.RandomForestRegressor. Available 26 May 2024 at: https://forecastegy.com/posts/do-decision-trees-need-feature-scaling-or-normalization/.

Fraunhofer Institut für Bauphysik (2024) From private single-family houses to large public buildings [Von privaten Einfamilienhäusern bis zu großen öffentlichen Gebäuden]. Available 25 May 2024 at: https://www.ibp.fraunhofer.de/de/kompetenzen/energieeffizienz-und-raumklima/gebaeude-quartier-stadt/sanierungskonzepte.html.

Fraunhofer Institut für Bauphysik (1999) Handbook epiqr.

Friedrichsen, S. (2024) Sustainable planning, building, and living - criteria for new construction and building in existing structures [Nachhaltiges Planen, Bauen und Wohnen - Kriterien für Neubau und Bauen im Bestand]. Wiesbaden: Springer Vieweg, 3rd updated and expanded ed., pp.40–43.

Fung, J. F., Butry, D. T., Sattar, S., & McCabe, S. L. (2017) A methodology for estimating seismic retrofit costs. Gaithersburg, MD, USA: US Department of Commerce, National Institute of Standards and Technology.

Géron, A. (2022) Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, Inc.

Google Earth. (2024) The most accurate globe in the world. Available 25 May 2024 at: https://www.google.com/intl/de_in/earth/.

Google Solar API. (2024) Request to create statistics on buildings. Available 25 May 2024 at: https://developers.google.com/maps/documentation/solar/building-insights?hl=de.

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020) Array programming with NumPy. In: Nature, 585(7825), pp.357–362.

Hassanain, M., Assaf, S., Al-Ofi, K., & Al-Abdullah, A. (2013) Factors affecting maintenance cost of hospital facilities in Saudi Arabia. In: Property Management, 31(4), pp.297–310.

Hassanain, M., Al-Zahrani, M., Abdallah, A., & Sayed, A. M. (2019) Assessment of factors affecting maintenance cost of public school facilities. In:

International Journal of Building Pathology and Adaptation, 37(5), pp.528–546.

Hatamleh, M. T., Hiyassat, M., Sweis, G. J., & Sweis, R. J. (2018) Factors affecting the accuracy of cost estimate: Case of Jordan. In: Engineering, Construction and Architectural Management, 25(1), pp.113–131.

HOAI (2021) Fee structure for architects and engineers from 01 January 2021 [Honorarordnung für Architekten und Ingenieure vom 01. Januar 2021]. Wiesbaden: Springer Vieweg, 5th ed.

Hunter, J. D. (2007) Matplotlib: A 2D graphics environment. In: Computing in Science & Engineering, 9(03), pp.90–95.

Ibrahim, A. H., & Elshwadfy, L. M. (2021) Factors affecting the accuracy of construction project cost estimation in Egypt. In: Jordan Journal of Civil Engineering, 15(3).

Impulsprogramm IP Bau. (1995) Rough diagnosis: Condition assessment and cost estimation of buildings – method [Grobdiagnose: Zustandserfassung und Kostenschätzung von Gebäuden – Methode]. Bern: Federal Office for Economic Affairs.

Islam, R., Nazifa, T. H., & Mohamed, S. F. (2019) Factors influencing facilities management cost performance in building projects. In: Journal of Performance of Constructed Facilities, 33(3), 04019036.

Khodabakhshian, A., Rampini, L., Vasapollo, C., Panarelli, G., & Cecconi, F. R. (2023) Application of machine learning to estimate retrofitting cost of school buildings. In: Resilient and Responsible Smart Cities: The Path to Future Resiliency, Cham: Springer International Publishing, pp.215–228.

Kleine, A. (2009) The three dimensions of sustainable development [Die drei Dimensionen einer Nachhaltigen Entwicklung]. In: Operationalizing a sustainability strategy: Integrating ecology, economy, and social issues [Operationalisierung einer Nachhaltigkeitsstrategie: Ökologie, Ökonomie und Soziales integrieren], pp.10–12.

Lauble, S., Haghsheno, S., Franz. M. (2023) Predicting the construction duration in the predesign phase with decision trees. In: Proceedings of the 14th European Conference on Product and Process Modelling (ECPPM 2022), September 14-16, 2022, Trondheim, Norway

McKinney, W. (2010) Data structures for statistical computing in Python. In: Proceedings of the 9th Python in Science Conference, pp.56–61.

Mirzaei, J., Hanzaei, H. A., Khaleghi, H., & Kashani, H. (2024) Cost estimation models for seismic retrofit of masonry school buildings. In: International Journal of Construction Management, 24(8), pp.843–853.

Mittag, H.-J., & Schüller, K. (2020) Statistics: An introduction with interactive elements [Statistik: Eine Einführung mit interaktiven Elementen]. Berlin: Springer Berlin Heidelberg, 6th fully revised and expanded ed.

Müller, A. C., & Guido, S. (2017) Introduction to machine learning with Python: Practical knowledge in data science [Einführung in Machine learning mit Python: Praxiswissen data science]. Heidelberg: O'Reilly.

Neddermann, R. (2007) Cost estimation in existing buildings [Kostenermittlung im Altbau] (4th ed.). Cologne: Werner Verlag, pp.53–55.

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2022) Human–autonomy teaming: A review and analysis of the empirical literature. In: Human Factors, 64(5), pp.904–938.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011) Scikit-learn: Machine learning in Python. In: Journal of Machine Learning Research, 12, pp.2825–2830.

PLAN4 Software GmbH (2024) Intelligent asset evaluation and agile construction planning: Mobile and simple. Available 25 May 2024 at: https://plan4software.de/

Prasetyono, P. N., Suryanto, M. & Dani, H. (2021) Predicting construction cost using regression techniques for residential building. In: Journal of Physics: Conference Series, 1899(1), p. 012120.

Van Rossum, G. & Drake, F. L. Jr. (1995) Python tutorial (Vol. 620). Amsterdam, The Netherlands, Centrum voor Wiskunde en Informatica.

Sammut, C. (2017) Genetic and Evolutionary Algorithms. In: Sammut, C. & Webb, G. I., eds. Encyclopedia of Machine Learning and Data Mining, Springer US, p.pp. 456–457.

Scikit-Learn Developers. (2024b) StandardScaler. Available 21 May 2024 at: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

Umweltbundesamt. (2023) Energy efficiency in numbers [Energieeffizienz in Zahlen]. Available 29 April 2024 at: https://www.bmwk.de/Redaktion/DE/Publikationen/Energie/energieeffizienz-in-zahlen-2022.pdf?__blob=publicationFile&v=7

Waskom, M. L. (2021) Seaborn: Statistical data visualization. Journal of Open Source Software, 6(60).