A machine learning approach based on contract parameters for cost forecasting in construction

BEN TALHA Maroua[1], HOUARI Fatima-Zahrae[1], BENGHABRIT Asmaa[2]
and RHAZI Mohamed Sayf Eddine[3]

[1] INDUSLAB, ONRST, ENSMR Rabat, Morocco
[2] Laboratory of Applied Mathematics and Business Intelligence, ENSMR Rabat, Morocco
[3] JACOBS Engineering S.A, Casablanca, Morocco
maroua.bentalha@enim.ac.ma

Abstract. The construction industry, characterized by its inherent complexity, is a cornerstone of economic growth and development. In the era of Construction 4.0, accurate cost estimation within this sector is essential for successful project planning and execution. This research aims to predict construction project costs using real data from 203 projects, incorporating contract parameters provided by an engineering consulting firm. Among the eleven machine learning algorithms tested, the Partial Least Squares regression model emerged as the most effective. It achieved the highest ranking in the TOPSIS analysis, demonstrating a coefficient of determination of 91% for the test set.

Keywords: Construction industry, Cost forecasting, Contract parameters, Machine learning, TOPSIS.

1 Introduction

Economic growth is heavily influenced by the construction industry, which involves a diverse range of stakeholders such as contractors, clients, and consultants [1]. The inherent complexity and multifaceted nature of construction projects demand effective collaboration and efficient management [2]. Success in this sector is often measured through effective collaboration and efficient management [2], it is often measured through the principles of the Iron Triangle [3], which emphasize the importance of staying within budget, meeting deadlines, and maintaining high-quality standards. The complexity of construction projects makes variations in design or building processes almost unavoidable. Variation orders, or change orders, are challenging elements that require careful management to prevent disputes and claims between contractors and owners, especially concerning contract costs and schedule delays [4]. These orders frequently lead to higher costs and disruptions to ongoing work, ultimately resulting in both cost and time overruns [5]. Therefore, developing robust strategies and approaches to manage risks and uncertainties in construction projects is imperative [6], the construction project management sector primarily consists of several key stages: the initial design phase, detailed design phase, construction cost estimation, project bidding phase, construction phase, and final delivery after completion [7]. Accurate cost estimation is seen as essential for project success across all construction phases [8], it significantly impacts the profitability of projects during the tender phase and is crucial for the overall viability of the project [9]. This process involves the collaborative efforts of a tender manager, who coordinates the cost estimation, and technical experts such as engineers and project managers. Additionally, consultancy firms evaluate various parameters, including the type and scope of projects, to determine the necessary time and effort for project preparation, i.e., consultancy services. This approach differs significantly from cost estimation during construction, which primarily focuses on labor, materials, equipment, and other related expenses [10]. There are numerous estimation methods and techniques available [11]. However, these conventional methods fail to fully utilize the wealth of knowledge from past projects, leading to slow and inconsistent estimates. Advancements in computing power have led to the adoption of more complex methods in recent cost estimation approaches. Artificial Intelligence (AI) methods have been employed to explore multi-dimensional and non-linear relationships between final costs and input variables [12]. In this context, Machine Learning (ML), a subset of AI, has emerged as a promising tool for improving estimation accuracy. Recent studies indicate that ML can be beneficial for predicting both duration and costs [13]. Several researchers emphasize the importance of validating these machine learning techniques using real-world data to ensure their accuracy and reliability [14]. This research focuses on developing a precise machine learning model that utilizes historical data to estimate costs for engineering consultancy firm. To determine the most effective model, we compared various ML regression techniques. Additionally, we used a TOPSIS analysis (Technique for Order of Preference by Similarity to Ideal Solution) to rank the algorithms based on four criteria: coefficient of determination (R-squared), Root Mean Squared Error (RMSE), memory usage, and computational time. Out of the eleven cost forecasting models evaluated, the PLS (Partial Least Squares regression) model stood out for its robust learning capabilities and superior accuracy in cost estimation, leading to its final deployment. In exploring these techniques, we address the question of how ML models can be effectively leveraged within the context of Construction to accurately predict costs using contract parameters in the early stages of Project Management? The remainder of this paper is structured as follows: we begin with a review of relevant literature. Next, we discuss the proposed methodology in detail including data acquisition, description, pre-processing and implementation. This is followed by analyzing the results of the cost estimates, and discuss the main findings. We conclude our study by summarizing the main contributions, limitations, and future research.

2 Literature Review

Previous research has explored various cost estimation techniques, dividing them into qualitative and quantitative methods. Qualitative methods rely on the estimator's knowledge of the project [15], and include expert judgment and heuristic rules. Expert judgment is shaped by the success of prior estimations and often involves consulting experienced professionals to enhance accuracy [16]. Heuristic rules are intuitive, using rules of thumb based on comparisons with similar past projects to streamline the estimation process. Quantitative methods rely on collecting and analyzing historical data, and using quantitative models, techniques, and tools to estimate project costs. Among these approaches, parametric and simulation methods are prominent. Parametric methods estimate project costs by identifying causal relationships with specific parameters, resulting in a mathematical function of the related variables. This method is particularly efficient during the early stages of a project when limited information is available [17]. Stochastic Budget Simulation (SBS) enhances cost estimation by combining Monte Carlo simulation with the successive principle. It uses triple estimates (minimum, most likely, maximum) to address uncertainties, improving reliability. This method produces a probabilistic cost range, making it ideal for early-stage, large-scale projects [18].

Advanced methods utilizing AI and ML have also been employed for cost estimation. Among these, Artificial Neural Networks (ANNs) are frequently used for estimating the duration and costs of construction projects during the preliminary design stage [19]. The effectiveness of ANN and SVM (Support Vector Machine) models in estimating the cost and duration of road projects was evaluated [20], with SVM models outperforming ANN models in terms of accuracy and error reduction. Another study highlighted the effectiveness of SVM in estimating the cost of bridge construction during the initial stages of project development [21]. In [22], a comparison of twenty AI techniques for estimating construction costs in field canal improvement projects revealed that the XGBoost (eXtreme Gradient Boosting) algorithm provided the most accurate predictions, achieving an R-squared value of 0.929. These models can significantly aid decision-makers in substation projects. The study in [23] applied three machine learning algorithms—MLP (Multi-Layer Perceptron), GRNN (General Regression Neural Network), and RBFNN (Radial Basis Function Neural Network) along with a process-based method for early-stage cost prediction in project management. It confirmed that the GRNN algorithm produced superior results compared to other models, assisting project managers in predicting construction costs during the contracting phase. Our work builds on this foundation to forecast construction project costs by analyzing contract parameters and using a quantitative approach based on machine learning techniques. This method helps predict potential cost overruns early, offering insights for more effective project management.

We overview the eleven regression algorithms used in our study, divided into linear models (Ridge [24], Lasso [24], Elastic Net [25], PLS Regression [26]) and non-linear models; Random Forest [27], ETR (Extra Tree Regression) [28], AdaBoost (Adaptive Boosting) [29], CatBoost [30], LightGBM (Light Gradient Boosting Machine) [31], Gradient Boosting [32], XGBoost [33]. This categorization stems from our literature review, which shows these groups are commonly used in cost estimation for their specific benefits and applications. Linear methods are foundational in regression analysis due to their simplicity and interpretability. These models assume a linear relationship between input features and the target variable, making them effective for problems where this assumption holds true. Ridge Regression adds an L2 penalty to the regression model, stabilizing it by shrinking the coefficients of correlated predictors. Lasso Regression introduces an L1 penalty, which can shrink some coefficients to zero, effectively performing variable selection and simplifying the model [24]. Elastic Net Regression, a hybrid of Lasso andRidge, combines their penalties, making it particularly effective in the presence of highmulticollinearity among predictors [25]. PLS Regression reduces predictors to a smaller set of uncorrelated components while preserving as muchvariance in the response variable as possible, making it useful in situations with many correlated predictors [26]. The selected linear models offer substantial flexibility in hyperparameter tuning. This flexibility is critical, enabling precise adjustments to the model parameters, such as regularization strength in Ridge and Lasso, or the mixing ratio in Elastic Net. These adjustments allow the models to better fit the specific characteristics of the dataset, thereby enhancing both predictive accuracy and model generalizability.

Non-linear methods are chosen for their ability to model complex, non-linear relationships between variables, capabilities that linear models lack. These methods do not assume a linear relationship and can capture intricate patterns in the data. Random Forest, introduced by Leo Breiman, constructs multiple decision trees using subsets of randomly selected variables, improving predictiveaccuracy and reducing overfitting by averaging these trees' results [27]. ETR, proposed by Geurts et al., builds unpruned decision or regression trees,introducing more randomness in tree construction, which enhances its performance oncomplex datasets [28]. AdaBoost introduced by Freund and Schapire, is an ensemble method that combines multiple weak learners, adjusting the weights of incorrectly predicted instances to focus more on challenging cases [29]. CatBoost, developed by Yandex, is a gradient boosting algorithm designed to handle categorical features efficiently without extensive preprocessing, significantly enhancing model performance on categorical data [30]. LightGBM is a highly efficient gradient boosting framework that uses a histogram-based approach to speed up training and reduce memory usage [31]. Gradient Boosting builds models sequentially, each correcting the errors of its predecessor, thus improving predictive accuracy incrementally [32]. XGBoost, a scalable and high-performance implementation of gradient boosting, is designed for speed and performance, making it highly effective for large datasets [33].

3 Proposed Methodology

3.1 Workflow

This research utilizes a systematic methodology to develop and evaluate machine learning models, as illustrated in Figure 1. The process begins with data acquisition, followed by data preprocessing to organize and clean the data for effective modeling. Key predictors are then identified during the feature selection phase. Machine learning algorithms are applied within a Jupyter Notebook, incorporating hyperparameter tuning and testing, with Cross Validation to ensure robustness. The TOPSIS method ranks the algorithms, selecting the highest-performing one based on predefined criteria. The final step involves deploying the top model for practical use.

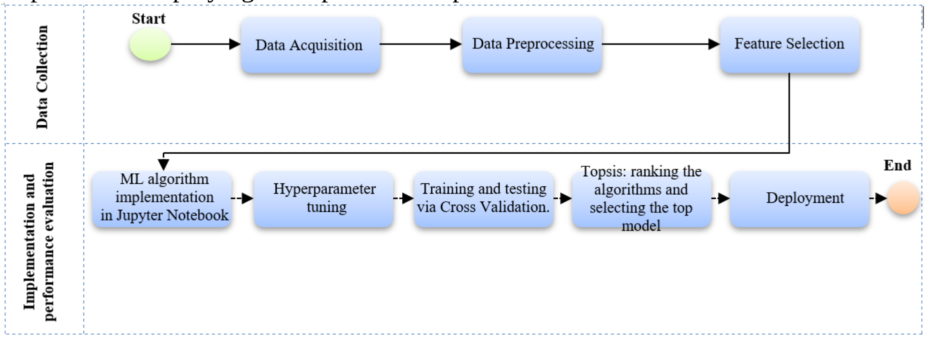

Data Collection
Start
Data Acquisition
Data Preprocessing
Feature Selection
Implementation and performance evaluation
ML algorithm implementation in Jupyter Notebook
Hyperparameter tuning
Training and testing via Cross Validation.
Topsis: ranking the algorithms and selecting the top model
Deployment
End

Fig. 1. Research Workflow.

3.2 Data acquisition and description

The dataset utilized in this research was furnished by the consulting firm JESA SA and includes data on 203 real-world projects from sectors such as building, transportation, and ports, covering a temporary margin from the year 2013, offering a longitudinal perspective that enhances the analysis of trends and patterns over time. This dataset is structured into 11 columns, divided into numerical and categorical variables:

- Numerical Variables: Contract Value (Revenue); indicates the financial magnitude of the projects in MAD. Final Cost: represents the ultimate expenditure incurred onthe projects in MAD.
- Categorical Variables: Project Number: a unique identifier for each project, ProjectDescription: detailed narrative of the project scope, Project Size: categorized into Small, Medium, and Large, this variable assesses the scale of the project, Business Area: encompasses 14 distinct sectors within the business, Service Type: includes 12 different services offered, detailing the variety of services engaged, Legal Entity:identifies one of four legal entities involved and Geography: divided into Africa andNational (Morocco), indicating the geographical location of the project execution.

3.3 Data preprocessing

Data preprocessing is a critical phase in our research, essential for ensuring data quality and integrity before applying machine learning models. The initial step in data cleaning involved addressing missing values and managing outliers, both of which can significantly impact the accuracy of machine learning models. For categorical variables such as Business Area and Service Type, missing entries were imputed with the mode of the respective variable, ensuring that the most frequently occurring category was used to fill the gaps. Outliers, which can distort predictive modeling and lead to mis- leading results, were addressed using the Interquartile Range (IQR) method. This method of capping helps in reducing the skewness of the distribution and maintains the integrity of the data set.

3.4 Feature selection

In the feature selection phase of our study, we have considered all available variables within the dataset to forecast project costs accurately. Our aim is to establish a statistically significant relationship between these variables and the output cost, thereby validating their inclusion in the predictive model. This process begins with a comprehensive evaluation of the independence of categorical variables through the Chi-square test [34].

Table 1. Chi-square Test Results for Independence.

Variable 1	Variable 2	P-value
Contract Type	Size	0.09990
Contract Type	Geography	0.37234
Size	Geography	0.89744

Table 1 shows Chi-square test results for independence between Contract Type and both Size and Geography, and between Size and Geography. All pairs listed have p-values above 0.05, indicating no significant associations and thus they are independent. Other variable pairs not shown in the table have significant p-values below 0.05, indicating strong dependencies between these variables. Following the assessment of categorical variables, we employ the Kruskal-Wallis test, that tests the null hypothesis that the medians of all groups are equal. This allows us to determine if there are statistically significant differences between the groups' medians [35].

Table 2. Kruskal-Wallis results.

Variable	P-value
Size	0.0000001
Contract Type	0.0009505
Business Area	0.0000008
Service Type	0.0000431
Legal Entity	0.0090484
Geography	0.0175414

The results from the Kruskal-Wallis tests, as summarized in Table 2, confirm the statistical significance of several variables influencing the output cost. The findings confirm that these factors significantly impact the dependent variable, thus justifying their selection based on the rigorous criteria set by the Kruskal-Wallis test.

3.5 Algorithm implementation

In this section, we will detail the implementation of machine learning algorithms as guided by the previously discussed flowchart Figure 1. Once the data is prepared, we employ a grid search strategy to fine-tune the hyperparameters for each model, leveraging the GridSearchCV function. For instance, in the Ridge regression model, different alpha values ranging from 0.01 to 100.0 were tested to identify the optimal setting. Similarly, for models like Elastic Net, both alpha and l1_ratio parameters were explored to balance the penalty terms, while the Random Forest model's hyperparameters and boosting models, including n_estimators, max_depth, and min_samples_split, were tuned to enhance model performance. The machine learning models were implemented and trained in a Jupyter Notebook environment (version 7.0.8) using Python 3.9, chosen for its support of scientific computing libraries like Scikit-learn and Pandas. Following the training phase, each model underwent rigorous evaluation using the designated test set. Key performance metrics, including RMSE and R², were calculated to assess both accuracy and generalizability, ensuring the models' applicability to real-world data.

4 Experimental results

4.1 Performance of ML techniques

In the results section of our study, we evaluated the performance of various machine learning models, the evaluation focused on two key metrics: R² and RMSE. The variation in R² and RMSE scores across different models can be attributed to the distinct algorithms they employ and their specific approaches to handling data complexities. For example, Ridge Regression effectively manages multicollinearity, while PLS Regression simplifies the dataset by reducing predictors to a smaller set of uncorrelated components. These models have demonstrated strong performance, achieving an R² of 91% for PLS Regression and 90% for Ridge Regression, showcasing their aptness for such data structures. To assess the robustness of our models, we employed c
Cross-Validation, analyzing the variability and consistency of model performance through R² scores from various data splits. Figures 2 and 3 display the results, identifying models with consistent performance and those prone to variations across different splits. Specifically, models like Lasso and PLS Regression showed narrow distributions in their R-squared scores, indicating stable and reliable performance. Conversely, models such as Gradient Boosting displayed wider score distributions, suggesting they may be sensitive to specific data characteristics.

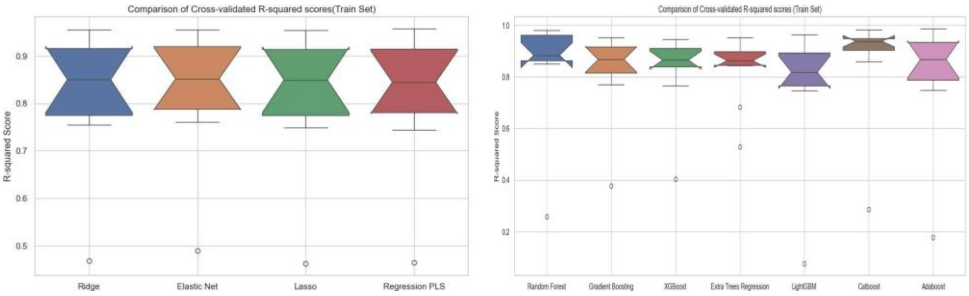


Fig. 2. Cross-validated R-squared Scores for train set.

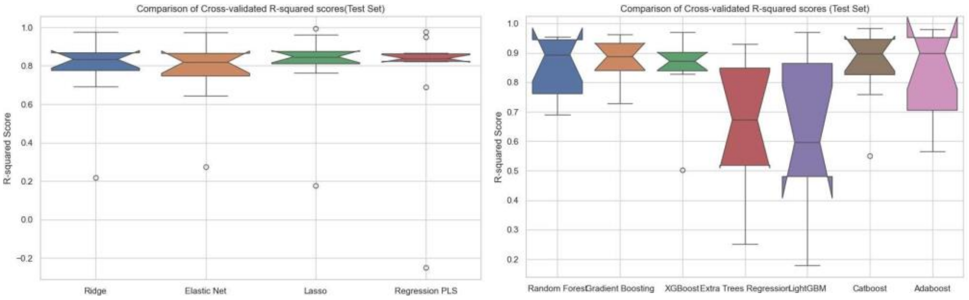


Fig. 3. Cross-validated R-squared Scores for test set.

4.2 TOPSIS analysis

To rank the algorithms, we used the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), developed by Hwang & Yoon [36]. This method evaluates alternatives based on their proximity to an ideal solution, favoring those closest to the positive ideal and farthest from the negative ideal. We considered four criteria: R², RMSE, Usage Memory, and Computational Time.

Table 3. Score Matrix for TOPSIS.

ML Technique	R²	RMSE	Usage Memory (MiB)	Computational time (Seconds)
PLS Regression	0.9118800	309182.6	0.08593	0.3181
Ridge	0.8983940	332007.7	0.28515	0.2207
Elastic Net	0.8929560	340775.6	0.02343	0.3939
Lasso	0.8983090	332145.9	0.00000	0.2312

Extra Trees Regressor	0.7515880	519128.4	0.43359	1.0290
Random Forest	0.8880313	348527.5	0.07812	16.283
Gradient Boosting	0.8403190	416212.8	0.09760	22.492
XGBoost	0.8426780	413127.0	0.52734	16.687
LightGBM	0.8021900	463245.4	0.24218	6.7239
CatBoost	0.6913820	578627.2	18.4179	44.553
AdaBoost	0.6616910	605822.0	0.00000	5.2532

In our evaluation methodology, we focused on optimizing the R-squared value to gauge model accuracy and assigned lesser weights to RMSE, memory usage, and com- putational time (0.4 for R-squared and RMSE, 0.1 for memory and time). Sensitivity analysis validated this approach, consistently highlighting the PLS Regressor as the best model. This confirms its effectiveness in various scenarios, leading to its imple- mentation in a Streamlit interface for dynamic user interaction.

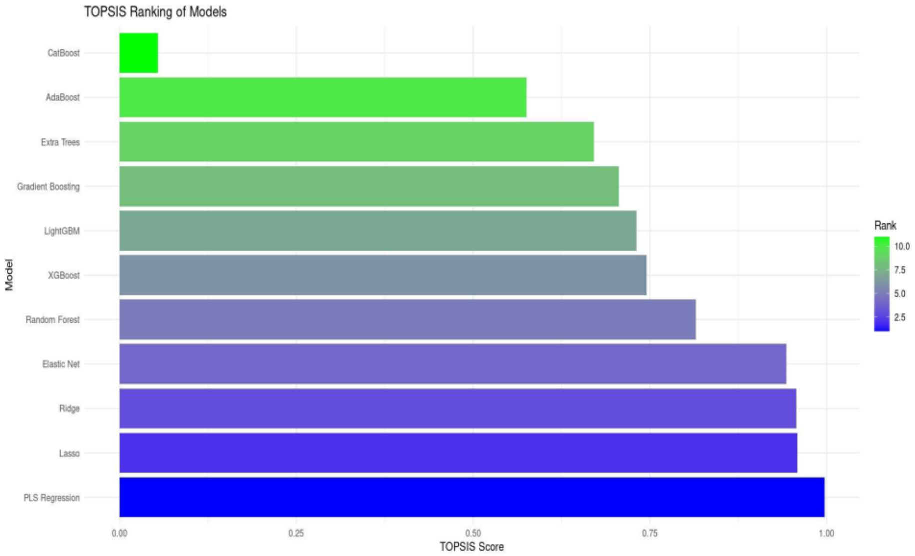
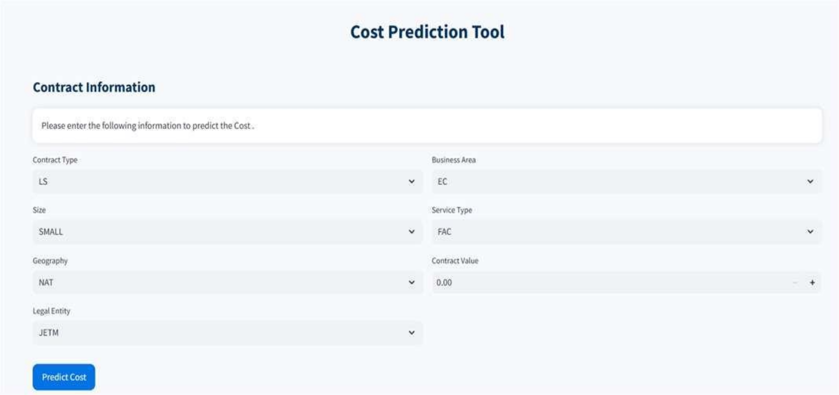


Fig. 4. TOPSIS ranking of models.

5 Deployment

This section introduces the Cost Prediction Tool, a user-friendly interface built with Streamlit that utilizes the PLS Regression model. Users can input key contract parameters like type, size, and geography to obtain precise cost estimates, demonstrating the practical application of our research findings.



Cost Prediction Tool

Contract Information

Please enter the following information to predict the Cost .

Contract Type: LS

Business Area: EC

Size: SMALL

Service Type: FAC

Geography: NAT

Contract Value: 0.00

Legal Entity: JETM

Fig. 5. Interface for cost predicting using contract parameters.

The following table provides a detailed specification sheet for the Cost As Sold Prediction Tool, summarizing its key features.

Table 4. Specification sheet of cost prediction tool.

Specification	Details
Overview	Web-based tool for predicting construction project costs using contract parameters.
Key features	Machine Learning Model: PLS Regression. User Inputs: Contract Type, Size, Geography, Legal Entity, Business Area, Service Type, Contract Value. Prediction Output: Estimated Cost in MAD.
Supported Platforms	Web-Based: Chrome, Firefox, Edge.

6 Conclusion and discussion

Accurate construction cost estimation is crucial for ensuring the financial success and operational efficiency of projects. Our research leveraged the PLS regression model, which proved to be highly effective in forecasting costs based on contract parameters. While the results are promising, it is important to acknowledge certain areas where further refinement and exploration could enhance the robustness of our findings.

One consideration is the scope of the dataset, which could benefit from further expansion to include a broader range of project types and conditions. This would not only enhance the model's generalizability but also its adaptability to different contexts within the construction industry. Additionally, although the PLS regression model performed exceptionally well, exploring more advanced modeling techniques, such as Artificial

Neural Networks (ANNs), could further capture the complex, nonlinear relationships that may exist within larger and more varied datasets.

References

1. Ingle, P. V., & Mahesh, G.: Construction project performance areas for Indian construction projects. International Journal of Construction Management 22(8), 1443–1454 (2020).
2. Jing, W., Naji, H. I., Zehawi, R. N., Ali, Z. H., Al-Ansari, N., & Yaseen, Z. M.: System dynamics modeling strategy for civil construction projects: the concept of successive legislation periods. Symmetry 11(5), 677 (2019).
3. Pollack, J., Helm, J., & Adler, D.: What is the Iron Triangle, and how has it changed? International Journal of Managing Projects in Business 11(2), 527– 547 (2018).
4. Charoenngam, C., Coquinco, S. T., & Hadikusumo, B. H. W.: Web-based ap- plication for managing change orders in construction projects. Construction In- novation 3(4), 197-215 (2003).
5. Chan, A. P., & Y. C. M.: A comparison of strategies for reducing variations. Construction Management and Economics 13(6), 467-473 (1995).
6. Jaafari, A.: Management of risks, uncertainties and opportunities on projects: time for a fundamental shift. International Journal of Project Management 19(2), 89–101 (2001).
7. Kim, S., Chang, S., & Castro-Lacouture, D.: Dynamic modeling for analyzing impacts of skilled labor shortage on construction project management. Journal of Management in Engineering 36(1) (2020).
8. ElSawy, I., Hosny, H., & Razek, M. A.: A neural network model for construc- tion projects site overhead cost estimating in Egypt. arXiv Prepr. (2011).
9. Ahn, J., Ji, S. H., Ahn, S. J., et al.: Performance evaluation of normalization- based CBR models for improving construction cost estimation. Automation in Construction 119, Article ID 103329 (2020).
10. Zwaving, J.: Probabilistic estimating of engineering costs. Delft University of Technology, Delft, The Netherlands (2014).
11. Burke, R.: Project management - planning and control techniques. 4th edn. Wiley, India, 251–266 (2009).
12. Günaydin, H. M., & Doğan, S. Z.: A neural network approach for early cost estimation of structural systems of buildings. International Journal of Project Management 22(7), 595–602 (2004).
13. Pellerin, R., & Perrier, N.: A Review of Methods, Techniques and Tools for Project Planning and Control. International Journal of Production Research 57(7), 2160–2178 (2018).
14. Aramali, V., Sanboskani, H., Gibson, G. E., Jr., El Asmar, M., & Cho, N.: For- ward-Looking State-of-the-Art Review on Earned Value Management Systems: The Disconnect Between Academia and Industry. Journal of Management in Engineering 38(3) (2022).
15. Hashemi, S. T., Ebadati, O. M., & Kaur, H.: Cost estimation and prediction in construction projects: a systematic review on machine learning techniques. SN Applied Sciences 2, Article 1703 (2020).
16. PF: Project estimating and cost management. Management Concepts Incorpo- rated, Virginia (2002).
17. Hegazy, T., & Ayed, A.: Neural network model for parametric cost estimation of highway projects. Journal of Construction Engineering and Management 124(3), 210-218 (1998).

18. Elkjaer, M.: Stochastic budget simulation. International Journal of Project Man- agement 18(2), 139–147 (2000).
19. Lowe, D. J., Emsley, M. W., & Harding, A.: Predicting construction cost using multiple regression techniques. Journal of Construction Engineering and Man- agement 132(7), 750–758 (2006).
20. Peško, I., Mucenski, V., Seslija, M., et al.: Estimation of costs and durations of construction of urban roads using ANN and SVM. Complexity 13 (2017).
21. Juszczyk, M.: On the search of models for early cost estimates of bridges: an SVM-based approach. Buildings 10(1), 2 (2019).
22. Elmousalami, H. H.: Artificial intelligence and parametric construction cost es-timate modeling: state-of-the-art review. Journal of Construction Engineering and Management 146(1) (2020).
23. Car-Pusic, D., Petruseva, S., Zileska Pancovska, V., & Zafirovski, Z.: Neural network-based model for predicting preliminary construction cost as part of cost predicting system. Advances in Civil Engineering (2020).
24. Melkumova, L. E., & Shatskikh, S. Ya.: Comparing Ridge and LASSO estimators for data analysis. Procedia Engineering 201, 746–755 (2017).
25. Zou, H., & Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(2), 301-320 (2005).
26. Chun, H., & Keles, S.: Sparse partial least squares regression for simultaneous dimension reduction and variable selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72(1), 3–25 (2010).
27. Breiman, L.: Bagging Predictors. Machine learning 24(2), 123-140 (1996).
28. Lu, S., Wang, X., Zhang, G., & Zhou, X.: Effective algorithms of the Moore Penrose inverse matrices for extreme learning machine. Intelligent Data Analysis 19(4), 743–760 (2015).
29. Shrestha, D. L., & Solomatine, D. P.: AdaBoost.RT: A boosting algorithm for regression problems. Proceedings of the IEEE International Joint Conference on Neural Networks, 1163–1168 (2004).
30. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A.: CatBoost: unbiased boosting with categorical features. arXiv preprint arXiv:1706.09516v5 (2019). https://arxiv.org/abs/1706.09516
31. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.- Y.: LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems 30, 3149–3157 (2017).
32. Friedman, J. H.: Greedy function approximation: a gradient boosting machine. Annals of Statistics, 1189–1232 (2001).
33. Chen, T., & Guestrin, C.: XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794 (2016).
34. Chicco, D., Warrens, M. J., & Jurman, G.: The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science (2021).
35. Rana, R., & Singhal, R.: Chi-square test and its application in hypothesis testing. Journal of the Practice of Cardiovascular Sciences 1(1), 69-71 (2015).
36. Hwang, C. L., & Yoon, K. P.: Multiple attributes decision making methods and applications. Springer-Verlag, Berlin (1981).