European Journal of
Technology
(EJT)

Data-Driven Approaches for Forecasting Cost Overruns in Infrastructure Projects

Sridhar Reddy Bandaru, Dhuli Shyam, Prabu Manoharan, Muzaffer Hussain Syed, Uday Kumar Ragireddy, Prasanth Varma Addepalli

AJPO

European Journal of Technology
ISSN 2520-0712 (online)
Vol.7, Issue 5, pp 23-34, 2023
www.ajpojournals.org

Data-Driven Approaches for Forecasting Cost Overruns in Infrastructure Projects

Sridhar Reddy Bandaru[1*], Dhuli Shyam[2], Prabu Manoharan[3], Muzaffer Hussain Syed[4], Uday Kumar Ragireddy[5], Prasanth Varma Addepalli[6]

[1]Program Management, IT, Microsoft, Senior ACE Engineer, Redmond, WA
[2]Business Application, IT, Nagase Holdings America Corp, Manager, Application & Software Development, NYC, NY
[3]Information Technology, Bourns Inc, HRIS Manager, California, USA
[4]IT Project Lead, Vdrive IT Solutions Inc
[5]Sr Technical Program Manager, Vdrive IT Solutions, Inc, Richardson, Texas
[6]Lead Data Architect/ Engineer, Federal Motor Carrier Safety Administration, Atlanta, Georgia

Crossref
Article history
Submitted 27.09.2023 Revised Version Received 25.10.2023 Accepted 27.11.2023

Abstract

Purpose: The aim of the study was to assess sustainable development depends on the infrastructure projects that affect the long-term economic, social, and environmental impacts. Cost overruns, however, tend to occur in construction projects.

Methodology: This research study uses the Build Bridges, Not Walls dataset available in the U.S. National Bridge Inventory, which contains over 600,000 bridge records, and builds a predictive model to estimate the number of infrastructure cost overruns. To begin with, it involved extensive preprocessing, including imputing missing values, scaling the features, and analyzing their correlations. The model was assessed using R2, RMSE, MAE, MAPE, and SHAP interpretability indicators.

Findings: The XGBoost model performed very well on the prediction accuracy metric, achieving an R2 of 95.67%, exceeding the ANN benchmark (R2 = 77.3%). The analysis of the Shapley values showed that landform and location were the most influential features, with construction type being a minor factor. Additionally, the strength of XGBoost was further confirmed through evaluations of RMSE, loss curves, and feature-wise performance plots.

Implications to Theory, Practice and Policy: The overall conclusion from the study is that applying ensemble learning methods especially XGBoost was the right approach for plausible costing of large-scale infrastructure projects, leading to better planning and decision-making.

Keywords: *Cost Overruns, Infrastructure Projects, Construction Estimation, Hyperparameter Optimization, SHAP Analysis, Predictive Modeling, Sustainable Development*

INTRODUCTION

The phrases "budget increase," "cost increase," and "cost growth" are sometimes used interchangeably with "cost overrun."[1]. Cost overruns in construction projects are common, can occur under a variety of market and legal circumstances, and, regrettably, can have a detrimental impact on achieving project objectives. The results of many investigations show the extent of this problem. Different construction investment types can be defined and described at different phases of the project's implementation [2]. Different organizational, economic, and technical specificities define these investments. However, one should take into account factors like the proportion of a particular element in the facility's overall cost, the risk of variations in the quantity of work and the vulnerability of a certain kind of work to variations in unit prices, such as the cost of building materials, to evaluate the risk of cost overruns of that specific component.

The unpredictability of building materials is a problem for the construction industry. Key building materials, including steel, cement, asphalt, gravel, sand, and concrete, are subject to daily price fluctuations influenced by a variety of factors, including market conditions, raw material costs, and energy prices. These discrepancies create challenges for bidders and project managers alike, making material cost estimation more complex and prone to error. Data-driven approaches have been developed as a crucial tool for forecasting building material costs to address these issues. These techniques leverage historical data to model and predict future material costs and provide insight into potential price fluctuations [3]. They have been extensively used in pertinent research for the past 20 years and are frequently categorized under "machine learning" and "deep learning." The data-driven forecasting methods used in construction economics nowadays may be roughly divided into two groups: time-series analysis and causal modeling.

The sustainability perspective needs to be carefully considered when developing infrastructure, as it has a significant impact on the economy, society, and the environment. Specifically, intercity transportation facilities promote urban sprawl and aggregation, thereby significantly accelerating regional economic growth [4]. Furthermore, the largest linear infrastructure projects affect biodiversity and natural habitats in several ways and contribute to landscape fragmentation [5]. However, the project's activities also significantly impact the ecosystem, including through infrastructure development.

The behavior of steel rebar-concrete systems can be predicted using machine learning (ML) techniques. Still, several obstacles remain, such as the scarcity of reliable beam test data, which makes it challenging to develop a trustworthy bond-strength model using hyperparameter-optimized ML algorithms [6]. It has always been challenging to integrate theoretical understanding with machine learning techniques for nonlinear material characteristics in structural mechanics. This paper proposes a method for determining bond strength between steel rebars and the surrounding concrete using machine learning techniques to overcome the drawbacks of traditional empirical formulae and data-driven models. By combining Bayesian optimization with machine learning, the model's accuracy can be improved. It has not been common practice to forecast bond strength using ML-based models in conjunction with the SHAP technique. This work provides crucial insights into the complex nonlinear behavior of bond strength by integrating SHAP with ML algorithms. It also shows how different aspects contribute to bond strength.

Motivation and Contribution of Paper

In building projects, cost overruns are inevitable and result from unpredictable material quantities, fluctuating material prices, and the difficult nature of construction work. Infrastructure development has its pros and cons: it not only brings economic benefits but also harms the environment. Therefore, valid prediction devices are increasingly becoming necessary. Machine learning has also introduced tremendous opportunities for predicting the prices and behaviors of materials. Still, issues such as the lack of high-quality data and the need to incorporate nonlinear modeling remain. The development of machine learning tools, such as Bayesian optimization and SHAP-based interpretability, can lead to models that are not only more accurate and transparent but also enable cost prediction and support environmentally aware, intelligent decisions in construction operations. The important contributions of the study are as follows:

- The work used was based on the effective collection of a build bridges, not walls database that Kaggle provided.
- The study put in place a solid data pre-processing pipeline involving the imputation of missing values, and scaling of data through a Standard Scaler to make features standardized so that the models contribute fairly.
- It provided input in the form of applying and describing the use of the XGBoost algorithm.
- It provided a thorough analysis of how the model performs in terms of significant metrics: R2, RMSE, MAE, and SHAP highlighting that lower values of errors and high values of R2 are necessary to successfully predict data.

Significance of the Study

The significance of this research is that it demonstrates the capacity to forecast cost overruns in mega infrastructure projects with high accuracy using a vast dataset of actual bridge statistics. The work provides a credible framework for identifying the very small and large variables that significantly impact project costs, thereby making it easier to assess and support evidence-based decision-making. The outcomes shed some light on the engineers, planners, and policymakers and give them the chance to better allocate resources, schedule projects, and even accommodate the possibility of future cost increases in infrastructure development, in a more optimistic way.

Structure of Paper

The structure of this paper is organized as follows: Section II a literature review of the available literature on predicting cost overruns in infrastructure projects. Section III presents the proposed methodology, data description, and model implementation. Section IV presents the experimental results and key findings. Finally, Section V concludes the study, discusses its limitations, and provides ideas for future research.

LITERATURE REVIEW

This study of the literature offers a thorough summary of previous research and conclusions on predicting cost overruns in infrastructure projects.

Bataev (2019) concluded that additive technologies would be useful only in the mass building industry. Additionally, depending on the methods employed, the cost of a single square meter of additive technology-built dwelling in the Russian market is 2.5 to 3.1 times lower than that of traditional building methods [7].

Bao et al. (2019) consider that estimating ego-motion and position change for robots using on-board cameras and SLAM algorithms is not only less expensive but also more reliable and broadly applicable. They conducted trials at construction sites and mounted many consumer cameras on construction equipment to test the viability of the concept. Their studies yield encouraging results, with a localisation accuracy of around 0.06 meters [8].

Bajjou and Chafi (2018) A total of 330 responses were examined using the Statistical Package for Social Sciences (SPSS V20.0 for Windows). LC is a top priority for Moroccan construction projects, and more than half of the respondents (172 professionals; 52.1%) believe the Lean Construction approach applies to companies of any size. The respondents agreed that implementing Lean Construction would have positive effects on Moroccan construction projects, with an overall mean value of 3.84 [9].

Arage and Dharwadkar (2017). The Ordinary Least Squares (OLS) approach, a type of simple linear regression model, is recommended for forecasting future building project expenses. When the dataset is small, the Ordinary Least Squares approach works well and can yield the optimal solution. To evaluate the model's accuracy, experiments were conducted using the Pune region's District Schedule Rates over 12 years. The results show that the proposed model achieves 91% to 97% prediction accuracy [10].

Hemal, Waidyasekara and Ekanayake (2017) found that the two main stakeholders influencing design modifications are the customer and the consultant. Choosing a skilled, knowledgeable design team, allowing enough time for design development, and considering other important factors are key to putting a DF into practice. Additionally, by proposing tactics to maximise the design percentage, the respondents approved the idea of a partial design freeze, following the study's finding that a 100% design freeze was not feasible in large-scale construction projects [11].

Morad and El-Sayegh (2016) reported results based on 53 replies. According to the report, several businesses are using EVM in their cost control systems. However, a large portion of the United Arab Emirates' construction companies do not employ EVM. The study found that 87% of respondents felt that EVM deployment is essential as a cost management tool, even though many businesses are not using it [12]

The Comparative Analysis Of Recent Studies Based On Findings, Challenges, Future Study Present In Table I

Table 1: Summary of Recent Studies on Forecasting Cost Overruns in Infrastructure Projects

Author	Methodology	Dataset	Key Findings	Challenges	Future Scope
Bataev (2019)	Simulation analysis of additive construction technologies	Cost comparison of additive vs traditional construction in Russian market	Additive technologies effective mainly for mass construction; cost per m² is 2.5–3.1 times lower than traditional methods	Limited effectiveness in small-scale construction; technology adoption barriers	Expand evaluation across diverse construction types; integrate economic and lifecycle assessments
Bao et al. (2019)	SLAM-based localization using on-board consumer cameras	Real construction machines equipped with multiple cameras at active sites	Achieved 0.06 m localization accuracy; low-cost and robust alternative to traditional sensors	Environmental variability; camera calibration and stability issues	Improve robustness in harsh environments; integrate with autonomous machinery systems
Bajjou & Chafi (2018)	Questionnaire-based survey analyzed with SPSS (V20)	330 responses from Moroccan construction professionals	Lean Construction perceived as highly beneficial; widely applicable across company sizes; mean acceptance level 3.84	Resistance to cultural and organizational change; low maturity in Lean practices	Develop Lean training programs; assess long-term impact of Lean adoption
Arage & Dharwadkar (2017)	Ordinary Least Squares (OLS) simple regression	12 years of District Schedule Rates (Pune, India)	OLS achieved 91–97% accuracy in forecasting construction costs; effective for small datasets	Limited capability for complex or nonlinear patterns	Explore advanced ML/AI methods; incorporate additional cost-influencing variables
Hemal, Waidyasekara & Ekanayake (2017)	Interviews and qualitative analysis on design freeze practices	Feedback from stakeholders in large-scale construction projects	Clients and consultants major drivers of design changes; full design freeze impractical; partial design freeze strategies recommended	Difficulty coordinating design teams; evolving client requirements	Develop structured partial design-freeze guidelines; evaluate impacts on project performance
Morad & El-Sayegh (2016)	Survey-based assessment of Earned Value Management (EVM) usage	53 responses from UAE construction companies	Some use EVM, but many do not; 87% agree EVM is necessary for cost control	Limited awareness and training; low adoption rates	Promote EVM training; integrate digital cost-control tools (BIM/EVM systems)

METHODOLOGY

Forecasting Cost Overruns in Infrastructure projects is a critical aspect of project management, aiming to predict and mitigate financial deviations from initial estimates. The process begins with collecting construction estimation data from Kaggle, followed by a crucial data pre-processing phase that involves data cleansing, imputation of missing values, and data scaling using a standard scaler. The data is preprocessed and then split into training (80%) and test (20%) sets. The performance of an XGBoost model is then assessed using metrics like R2, RMSE, MAE, and SHAP Analysis after it has been used for classification. Hyperparameter optimization is also integrated into the workflow to refine the XGBoost model, ultimately leading to the final results. The provided flowchart in Figure. One outlines a machine learning workflow for construction estimation using XGBoost.

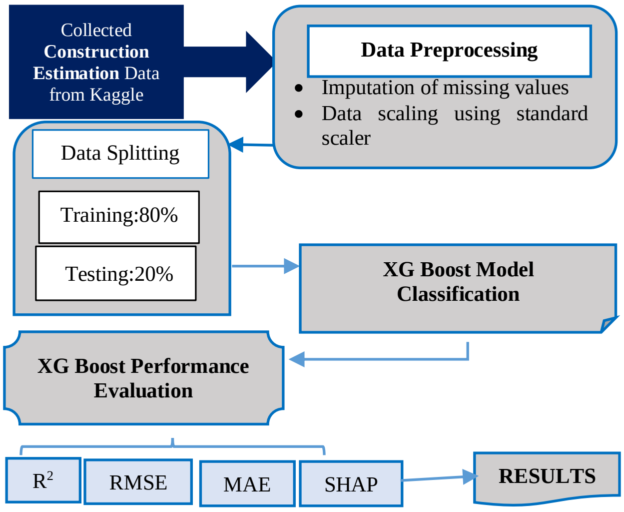

Figure 1: Flowchart Diagram of Forecasting Cost Overruns in Infrastructure

Each step and process of a flowchart is explained below:

Data Collection

The current research has used the "Build Bridges, Not Walls" dataset, which is from the U.S. National Bridge Inventory and consists of long-term records for roughly 600,000 U.S. bridges. The database contains 607070 rows and 135 columns. The provided information about the structure, its condition, and the inspections carried out makes the dataset significant for infrastructure assessment, risk analysis and predictive modeling.

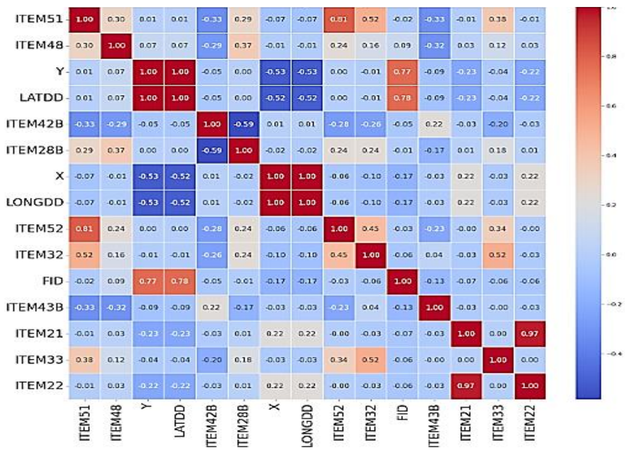

Figure 2: Correlation Coefficients Among the Features

A correlation heat map of the dataset's numerical variables is shown in Figure 2, where warm colors indicate strong positive correlations and cool colors indicate weaker or negative correlations. The plot emphasized the strong relations among several structural-condition measures, whereas the geographic variables showed weak or mixed associations.

Data Preparation

Data preparation, which entails converting raw data into a clear, organised, and useable format, is an essential stage in the data analysis and ML pipeline. The preprocessing pipeline involved in this study includes imputing missing values and data scaling using standard scaler as defined below:

- **Imputation of Missing Values:** Missing values were deleted or replaced based on the training data. Distribution of missing values per feature in the training data.
- **Data Scaling using Standard Scaler:** Scaling was performed to match the range of the data using the Standard. To standardize attributes, this approach removes the mean and scales to unit variance. This ensures that each feature contributes equally to the learning process of the model, which is essential for algorithms that depend on the amount of the input data in Equation (1) describes it:

$$z = \frac{x-\mu}{\sigma} \qquad (1)$$

This transformation results in a distribution where each feature has a mean of 0 and a standard deviation of 1.

Data Splitting

The dataset was split randomly into training and test sets at 70:30.

Model classifcation of XGBoost Model

In supervised learning, XGBoost uses training data to predict target variables [13]. A decision tree serves as the base learner for XGBoost. The difference between the objective and prediction values is lessened by adding additional base learners. The final predictive values are equivalent when all base learners are combined.

The XGBoost method may be thought of as an additive model made up of M DT [14], which are determined by Equation (2):

$$Y_i = \sum_{m=1}^{M} f_m(x_i), f_m \in F \qquad (2)$$

where, f stands for a decision tree, and F stands for each decision tree's function.

Model Performance Assessment

The correlation coefficient (R2), MAE, RMSE, MAPE, and SHAP analysis metrics are used to assess the models' performance. A model should ideally have greater R2 scores and lower MAE and RMSE values. Below is an outline of the statistical metrics.

Coefficient of Determination (R2)

The coefficient of determination, or R2, measures how closely a model's predictions match actual values. An algorithm that explains a higher percentage of the variation in the target variable is said to have a higher R2 score. Mathematically, it is represented by the equation. (3):

$$\mathrm{R}^2 = 1 - \frac{\sum(y_{\text{true}} - y_{\text{pred}})^2}{\sum(y_{\text{true}} - y_{\text{mean}})^2} \qquad (3)$$

Root Mean Square Error (RMSE)

RMSE highlights larger errors more than smaller ones and provides a sense of the error scale by showing the average error between actual and anticipated values. Equation. (4) depicts the formula:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(y_{\text{true}} - y_{\text{pred}})^2} \qquad (4)$$

Mean Absolute Error (MAE)

A simple indicator of prediction accuracy, MAE computes the average absolute deviation between anticipated and observed values. This information is useful for comparing error sizes to real values. The Equation. (5) formulates it:

$$MAE = \frac{1}{n}\sum(y_{\text{true}} - y_{pred})^2 \qquad (5)$$

SHAP Analysis

The SHAP technique is a tool for reading ML models that allows one to examine and comprehend the complex procedures involved in training black-box models. When used in conjunction with ML models, SHAP weighs each feature for each training data sample, performs a weighted fit to determine each feature's relevance, and assigns a value.

RESULTS ANALYSIS AND DISCUSSIONS

The local machine, which was an experimental setup, was equipped with an AMD Ryzen 5 3600 6-Core Processor that operates at 3.60 GHz, 16.0 GB of RAM, and is capable of providing the computational power required to run the processing tasks needed for the experiment. The proposed model, trained on an established dataset, achieved a high coefficient of determination (R2) of 95.67 in predicting cost overruns in infrastructure projects. In this benchmark data, the XGBoost model demonstrated strong predictive performance. As shown in Table II, the evaluation results demonstrate the model's strength across several performance criteria, highlighting the significance of ML, particularly the application of ensemble methods such as XGBoost, for accurately predicting the costs of outcomes in infrastructure design and execution.

Table 2: Results of XGBoost Forecasting Cost Overruns in Infrastructure

Models	XGBoost
R2	95.67
RMSE	86.90
MAE	48.75

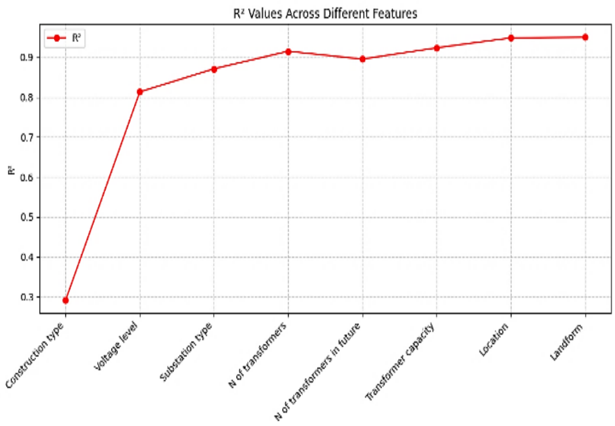

Figure 3: R2 Plot of the XGBoost Model

Figure 3 indicates the R² values of each feature in the XGBoost model. The highest values of 0.92 or higher are observed for landform, location, and transformer capacity, indicating strong predictive value, whereas construction type has the lowest value of 0.29. The trend toward higher feature counts indicates that adding high-impact variables vastly enhances model accuracy, underscoring the importance of effective feature selection for cost forecasting.

European Journal of Technology
ISSN 2520-0712 (online)
Vol.7, Issue 5, pp 23-34, 2023

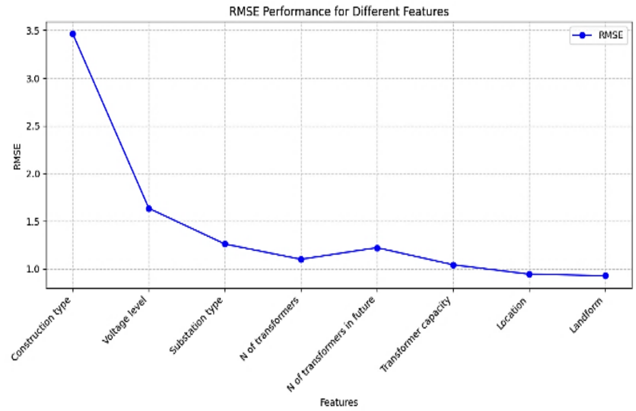

Figure 4: RMSE Plot of XGBoost Model

The RMSE performances of individual features in Figure 4 show the XGBoost model. The type of construction leads to the highest error, whereas the place and the landform together account for the least error confirming their stronger predictive value. It is shown that adding more relevant inputs makes predictions more accurate.

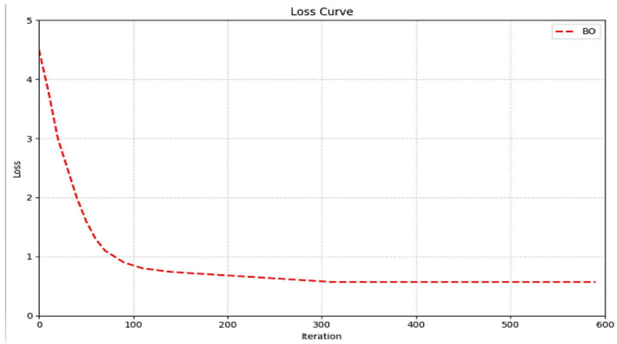

Figure 5: Loss Curve XGBoost Model

The XGBoost loss curve over training iterations is shown in Figure 5. The loss decreases quickly during the first phase and then slowly, finally reaching around 0.6, indicating the model has converged. This gradual decline is a sign of learning and, at the same time, points to the effectiveness of Bayesian Optimization in guiding the model to the best solution.

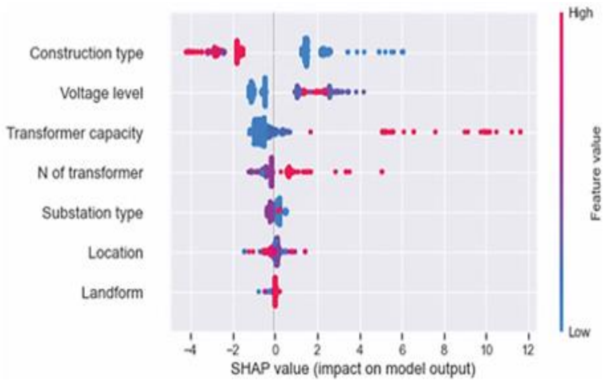

Figure 6: SHAP Explanation on the XGBoost Model

The SHAP summary graphic in Figure 6 shows how each attribute contributes to the model's predictions. The features act as rows, and the SHAP values indicate whether each feature had a positive or negative effect on the predicted target value. The colors of the points indicate the feature values, with low values shown in blue and high values in red. The range of points demonstrates the influence of various feature values on the output and, at the same time, highlights the features with the most significant overall effect on the model.

Comparative Analysis

Table III shows the evaluation of the two models' predictive abilities—ANN and XGBoost—both of which predict infrastructure cost overruns. The outcome indicates that XGBoost provides much better predictive accuracy, achieving an R^2 of 95.67%, while the ANN model achieves an R^2 of 77.3%. It indicates that XGBoost accounts for a much larger portion of the variance in cost overruns and is thus the more reliable model for predicting them.

Table 3: Comparison for Forecasting Cost Overruns in Infrastructure

Models	R2
ANN [15]	77.3
XGBoost	95.67

XGBoost is significantly advantageous for cost overrun prediction in infrastructure projects, as it can handle diverse data types, design versatile models, and regularize them. In addition, the rapid training and scaling of this approach correspond to its application in very large and complex datasets, thus producing high-quality predictions that are effective and robust.

Conclusion

This is a paramount undertaking in forecasting Cost Overruns in Infrastructure projects to ensure financial stability and deliver projects successfully. The paper has shown that XGBoost might be the most suitable option to forecast cost overruns in infrastructure projects based on the massive Build Bridges, Not Walls huge dataset. Following data preparation, feature grouping, and model trials, the XGBoost achieved impressive predictive accuracy, exceeding the boundary of the ANNs. The model boasts an excellent R^2 coefficient of 95.67, minimal error values, and a transparent SHAP justification, all of which appear to support its strength and reliability in real-time decision-making. The terrain, location, and transformer power were found to be the main factors that determine the accuracy of the prediction. Overall, the findings not only recognize the contribution of machine-learning-based approaches in simplifying the cost estimation process, making risk management smarter, and planning large-scale infrastructure more efficiently, but also in cost management, risk assessment, and conventional project planning.

In the future, research could successfully integrate larger infrastructure datasets, test hybrid and deep learning models, and introduce temporal patterns, among other things, to achieve higher forecasting precision. Conversely, a further exploration of SHAP-based interpretability, where real-time prediction systems and model validation across different project types are tested, will most certainly make it even more powerful when used for large-scale infrastructure planning and cost management.

REFERENCES

[1] E. Plebankiewicz, “Model of Predicting Cost Overrun in Construction Projects,” *Sustainability*, vol. 10, no. 12, p. 4387, Nov. 2018, doi: 10.3390/su10124387.

[2] S. Garg, “AI/ML Driven Proactive Performance Monitoring, Resource Allocation and Effective Cost Management in SaaS Operations,” *Int. J. Core Eng. Manag.*, vol. 6, no. 6, pp. 263–273, 2019.

[3] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, “A review on time series forecasting techniques for building energy consumption,” *Renew. Sustain. Energy Rev.*, vol. 74, pp. 902–924, Jul. 2017, doi: 10.1016/j.rser.2017.02.085.

[4] X. Jiang, L. Zhang, C. Xiong, and R. Wang, “Transportation and Regional Economic Development: Analysis of Spatial Spillovers in China Provincial Regions,” *Networks Spat. Econ.*, vol. 16, no. 3, pp. 769–790, Sep. 2016, doi: 10.1007/s11067-015-9298-2.

[5] J. H. Min, W. Jang, S. H. Han, D. Kim, and Y. H. Kwak, “How Conflict Occurs and What Causes Conflict: Conflict Analysis Framework for Public Infrastructure Projects,” *J. Manag. Eng.*, vol. 34, no. 4, Jul. 2018, doi: 10.1061/(ASCE)ME.1943-5479.0000625.

[6] S. Garg, “Predictive Analytics and Auto Remediation using Artificial Intelligence and Machine learning in Cloud Computing Operations,” *Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci.*, vol. 7, no. 2, 2019, doi: 10.5281/zenodo.15362327.

[7] A. V Bataev, “Efficiency Estimation Model of 3D Technology in the Construction Industry,” in *2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, 2019, pp. 1376–1381. doi: 10.1109/EIConRus . 2019.8657181.

[8] R. Bao, R. Komatsu, R. Miyagusuku, M. Chino, A. Yamashita, and H. Asama, “Cost-effective and Robust Visual-Based Localization with Consumer-level Cameras at Construction Sites,” in *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, 2019, pp. 983–985. doi: 10.1109/GCCE46687.2019.9015417.

[9] M. S. Bajjou and A. Chafi, “Towards implementing lean construction in the Moroccan construction industry: Survey study,” in *2018 4th International Conference on Optimization and Applications (ICOA)*, 2018, pp. 1–5. doi: 10.1109/ICOA.2018.8370556.

[10] S. S. Arage and N. V. Dharwadkar, “Cost estimation of civil construction projects using machine learning paradigm,” in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2017, pp. 594–599. doi: 10.1109/I-SMAC.2017.8058249.

[11] R. M. N. Hemal, K. G. A. S. Waidyasekara, and E. M. A. C. Ekanayake, “Integrating design freeze into large-scale construction projects in Sri Lanka,” in *2017 Moratuwa Engineering Research Conference (MERCon)*, 2017, pp. 259–264. doi: 10.1109/MERCon . 2017.7980492.

[12] M. Morad and S. M. El-Sayegh, “Use of Earned Value Management in the UAE Construction Industry,” in *2016 International Conference on Industrial Engineering, Management Science and Application (ICIMSA)*, 2016, pp. 1–5. doi: 10.1109/ICIMSA.2016.7504044.

[13] L. Pang, J. Wang, L. Zhao, C. Wang, and H. Zhan, “A Novel Protein Subcellular Localization Method With CNN-XGBoost Model for Alzheimer’s Disease,” vol. 9, no. January, pp. 1–7, 2019, doi: 10.3389/fgene . 2018.00751.

[14] H. Zheng and Y. Wu, “A XGBoost Model with Weather Similarity Analysis and Feature Engineering for Short-Term Wind Power Forecasting,” *Appl. Sci.*, vol. 9, no. 15, p. 3019, Jul. 2019, doi: 10.3390/app9153019.

[15] F. M. S. Al-zwainy and I. A. Aidan, “Forecasting the Cost of Structure of Infrastructure Projects Utilizing Artificial Neural Network Model (Highway Projects as Case Study),” vol. 10, no. May, 2017, doi: 10.17485/ijst/2017/v10i20/108567.

Attipalli, A., BITKURI, V., KURMA, J., Enokkaren, S., Kendyala, R., & Mamidala, J. V. (2021). A Survey of Artificial Intelligence Methods in Liquidity Risk Management: Challenges and Future Directions. *Available at SSRN 5741342*.

Attipalli, A., BITKURI, V., Mamidala, J. V., Kendyala, R., & KURMA, J. (2022). Empowering Cloud Security with Artificial Intelligence: Detecting Threats Using Advanced Machine learning Technologies. *Available at SSRN 5741263*.

Attipalli, A., Enokkaren, S., BITKURI, V., Kendyala, R., KURMA, J., & Mamidala, J. V. (2021). Enhancing Cloud Infrastructure Security Through AI-Powered Big Data Anomaly Detection. *Available at SSRN 5741305*.

Bitkuri, V., Kendyala, R., Kurma, J., Mamidala, V., Enokkaren, S. J., & Attipalli, A. (2021). Systematic Review of Artificial Intelligence Techniques for Enhancing Financial Reporting and Regulatory Compliance. *International Journal of Emerging Trends in Computer Science and Information Technology*, *2*(4), 73-80.

Kothamaram, R. R., Rajendran, D., Namburi, V. D., Singh, A. A. S., Tamilmani, V., & Maniar, V. (2021). A Survey of Adoption Challenges and Barriers in Implementing Digital Payroll Management Systems in Across Organizations. *International Journal of Emerging Research in Engineering and Technology*, *2*(2), 64-72.

Padur, S. K. R. (2018). Autonomous cloud economics: AI driven right sizing and cost optimization in hybrid infrastructures. *International Journal of Scientific Research in Science and Technology*, *4*(5), 2090-2097.

Padur, S. K. R. (2019). Machine learning for predictive capacity planning: Evolution from analytical modeling to autonomous infrastructure. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *5*(5), 285-293.

Padur, S. K. R. (2020). AI augmented disaster recovery simulations: From chaos engineering to autonomous resilience orchestration. *International Journal of Scientific Research in Science, Engineering and Technology*, *7*(6), 367-378.

Padur, S. K. R. (2020). From centralized control to democratized insights: Migrating enterprise reporting from IBM Cognos to Microsoft Power BI. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol*, *6*(1), 218-225.

Padur, S. K. R. (2021). Bridging Human, System, and Cloud Integration through RESTful Automation and Governance. *the International Journal of Science, Engineering and Technology*, *9*(6).

Padur, S. K. R. (2021). Deep learning and process mining for ERP anomaly detection: Toward predictive and self-monitoring enterprise platforms. *Available at SSRN 5605531*.

Padur, S. K. R. (2021). From Control to Code: Governance Models for Multi-Cloud ERP Modernization. *International Journal of Scientific Research & Engineering Trends*, *7*(3).

Padur, S. K. R. (2022). AI augmented platform engineering, transforming developer experience through intelligent automation and self optimizing internal platforms. *International Journal of Science, Engineering and Technology*, *10*(5), 10-5281.

Padur, S. K. R. (2022). Intelligent resource management: AI methods for predictive workload forecasting in cloud data centers. *J. Artif. Intell. Mach. Learn. & Data Sci*, *1*(1), 2936-2941.

Polam, R. M., Kamarthapu, B., Kakani, A. B., Nandiraju, S. K. K., Chundru, S. K., & Vangala, S. R. (2022). Data Security in Cloud Computing: Encryption, Zero Trust, and Homomorphic Encryption. *International Journal of Emerging Trends in Computer Science and Information Technology*, *3*(4), 31-41.

Polu, A. R., Buddula, D. V. K. R., Narra, B., Gupta, A., Vattikonda, N., & Patchipulusu, H. (2021). Evolution of AI in Software Development and Cybersecurity: Unifying Automation, Innovation, and Protection in the Digital Age. *Available at SSRN 5266517*.

Rajendran, D., Namburi, V. D., Singh, A. A. S., Tamilmani, V., Maniar, V., & Kothamaram, R. R. (2021). Anomaly Identification in IoT-Networks Using Artificial Intelligence-Based Data-Driven Techniques in Cloud Environmen. *International Journal of Emerging Trends in Computer Science and Information Technology*, *2*(2), 83-91.

Reddy Padur, S. K. (2021). From Scripts to Platforms-as-Code: The Role of Terraform and Ansible in Declarative Infrastructure Rollouts. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 621-628.

Routhu, K. K. (2021). Harnessing AI Dashboards in Oracle Cloud HCM: Advancing Predictive Workforce Intelligence and Managerial Agility. *International Journal of Scientific Research & Engineering Trends*, *7*(6).

Routhu, K. K. (2022). From Case Management to Conversational HR: Redefining Help Desks with Oracle's AI and NLP Framework. *International Journal of Science, Engineering and Technology*, *10*(6).

Routhu, K. K. (2022). From RFID to Geofencing: IoT-Enabled Smart Time Tracking in Oracle HCM Cloud. *International Journal of Science, Engineering and Technology*, *10*(4).

Singh, A. A. S., Tamilmani, V., Maniar, V., Kothamaram, R. R., Rajendran, D., & Namburi, V. D. (2021). Predictive Modeling for Classification of SMS Spam Using NLP and ML Techniques. *International Journal of Artificial Intelligence, Data Science, and Machine Learning*, *2*(4), 60-69.

Vattikonda, N., Gupta, A. K., Polu, A. R., Narra, B., Buddula, D. V. K. R., & Patchipulusu, H. H. S. (2022). Blockchain Technology in Supply Chain and Logistics: A Comprehensive Review of Applications, Challenges, and Innovations. *International Journal of Emerging Trends in Computer Science and Information Technology*, *3*(3), 72-80.

License