

# AI-Based Construction Cost Estimation and Budget Forecasting

Amarsinh B. Landage

<sup>1</sup>Assistant Professor, Department of Civil and Infrastructure Engineering, Government College of Engineering, Ratnagiri, 415612, India,

Aniket A. Shinde , Arun R. Survase, Sahil J. Juvalekar

Research Scholar, Department of Civil and Infrastructure Engineering, Government College of Engineering, Ratnagiri, 415612, India

**Abstract** - The increasing complexity of construction projects and rising cost uncertainties have highlighted the limitations of traditional cost estimation methods, which rely heavily on manual calculations and expert judgment. These conventional approaches often fail to capture dynamic market conditions and complex relationships among project variables, leading to inaccurate budget forecasting and cost overruns. This study presents the design and development of an Artificial Intelligence (AI)-based framework for construction project cost estimation and budget forecasting using historical project data.

The proposed system integrates machine learning techniques with data-driven analysis to provide accurate, real-time cost predictions. Key project parameters such as built-up area, location, material quantities, labor costs, and project duration are considered to effectively model cost behavior. The framework employs advanced algorithms, including Linear Regression, Random Forest, and Gradient Boosting, to capture nonlinear relationships and improve prediction accuracy. Additionally, a web-based interface is developed to enable users to input project details and obtain instant cost estimates.

The results demonstrate that AI-based models significantly outperform traditional estimation methods in terms of accuracy, consistency, and efficiency. The integration of predictive analytics with real-time data processing enhances decision-making, reduces human bias, and improves budget reliability. This study emphasizes the potential of AI-driven solutions to transform construction cost management practices and supports the development of intelligent, data-centric systems for sustainable infrastructure planning.

**Keywords:** Construction Cost Estimation, Artificial Intelligence, Machine Learning, Budget Forecasting, Predictive Analytics.

## 1. INTRODUCTION

The rapid growth of urban infrastructure and large-scale construction projects has significantly increased the complexity of cost management and budget planning in the construction industry. As development expands across metropolitan and semi-urban regions, the demand for accurate and reliable cost estimation becomes critical for ensuring project feasibility and financial sustainability. However, conventional cost estimation methods continue to rely heavily on manual calculations, historical averages, and expert judgment, which often fail to capture dynamic market conditions and complex relationships among project variables. This frequently results in cost overruns, inefficient resource allocation, and project delays (Flyvbjerg et al., 2002; Akinosho et al., 2020).

Traditional estimation approaches are limited in their ability to process large and diverse datasets, such as variations in material prices, labour costs, project location, and construction methods. These methods also struggle to adapt to nonlinear relationships and uncertainties inherent in construction projects. As a result, inaccuracies in early-stage budgeting remain a persistent challenge in the Architecture, Engineering, and Construction (AEC) industry. The inability to predict cost variations in real time highlights the need for advanced, data-driven solutions that can improve estimation accuracy and decision-making (Elbeltagi et al., 2014; Juszczak, 2017).

To address these challenges, Artificial Intelligence (AI)-based construction cost estimation has emerged as a transformative approach. By integrating Machine Learning (ML) algorithms with historical project data, AI systems can identify hidden patterns, model complex relationships, and generate accurate cost predictions. The primary objective of this research is to develop an AI-

driven framework that utilizes key project parameters—such as built-up area, location, material quantities, labour costs, and project duration—to forecast construction costs with higher precision. The system also incorporates a web-based interface to enable real-time user interaction and instant cost estimation (Sonmez, 2011; Chou et al., 2015).

Recent advancements in machine learning have demonstrated the effectiveness of algorithms such as Linear Regression, Random Forest, and Gradient Boosting in predicting construction costs. These models are capable of handling large datasets, reducing human bias, and improving estimation consistency. Additionally, the integration of real-time data sources and cost libraries (such as SSR and DSR) enhances the reliability of predictions by reflecting current market conditions (Adeli & Wu, 1998; Cheng et al., 2010; Kim et al., 2019).

Beyond improving estimation accuracy, AI-driven systems also contribute to better project management by enabling predictive analytics and informed decision-making. These systems allow stakeholders to evaluate multiple scenarios, optimize resource allocation, and reduce financial risks. Furthermore, predictive maintenance and cost optimization strategies can be incorporated to improve overall project efficiency (Fayek et al., 2010; Love et al., 2019).

Despite these advancements, a significant gap remains in the integration of these technologies into a unified and practical system. Many existing studies focus on individual machine learning models without providing a complete framework that combines data processing, model development, real-time prediction, and user interaction. This fragmentation limits the practical application of AI in construction cost management (Hegazy & Ayed, 1998; Kim et al., 2019).

The present study addresses this gap by proposing a comprehensive AI-based construction cost estimation and budget forecasting system. The framework integrates data pre-processing, feature engineering, machine learning models, and a web-based interface into a single platform. This unified approach enhances prediction accuracy, reduces dependency on traditional methods, and supports real-time decision-making. Ultimately, the proposed system aims to improve cost reliability, minimize budget overruns, and contribute to the development of intelligent and sustainable construction practices (Elbeltagi et al., 2014; Chou & Lin, 2013).

## 2. METHODOLOGY

The methodology of this study presents a systematic framework for the design, development, and evaluation of an AI-based construction cost estimation and budget forecasting system using historical project data. A hybrid research approach is adopted, combining both qualitative

and quantitative methods. The qualitative component includes literature review, identification of key cost-driving factors, and analysis of limitations in traditional estimation practices. The quantitative component focuses on data collection, pre-processing, machine learning model development, prediction accuracy evaluation, and system implementation through a web-based interface. This integrated approach establishes a structured workflow that combines data analytics, artificial intelligence, and decision-support mechanisms into a unified cost estimation system.

The framework begins with the collection of historical construction project data from multiple reliable sources, including past project records, cost databases, and publicly available datasets. The collected dataset includes key parameters such as project location, built-up area, material quantities, labour costs, project duration, and total project cost. Additional inputs such as SSR (Schedule of Rates) and DSR (Delhi Schedule of Rates) are incorporated to ensure alignment with standardized costing practices and real-world pricing trends.

**Table 1: AI Based Construction Project Cost Forecasting**

Step No.	Process Stage	Description
1	Data Acquisition	Collection of historical construction project data
2	Data Preprocessing	Feature cleaning, normalization, and selection
3	Model Selection	Random Forest, XG Boost, Neural Networks
4	Model Training	Training using historical project data
5	Model Evaluation	Evaluation using RMSE, MAE, R <sup>2</sup> Score
6	Prediction & Forecasting	Cost prediction for new projects
7	Deployment & Monitoring	Continuous learning and system monitoring

Following data collection, a comprehensive data pre processing stage is carried out to improve data quality and consistency. This includes handling missing values through imputation techniques, removing duplicate or inconsistent records, and normalizing numerical features such as area and cost. Categorical variables, such as project location, are converted into machine-readable formats using encoding techniques like label encoding and one-hot encoding. Outlier detection methods, including Interquartile Range (IQR) and Z-score

analysis, are applied to eliminate abnormal data points that could negatively impact model performance.

In the feature engineering stage, both primary and derived features are developed to enhance the predictive capability of the model. Primary features include built-up area, material quantities, and location index, while derived features such as cost per unit area, material cost ratios, and regional cost indices are generated to capture complex relationships within the data. These engineered features enable the model to better understand cost variations across different project conditions.

The model development phase involves implementing and evaluating multiple machine learning algorithms for cost prediction. Algorithms such as Linear Regression, Random Forest Regression, and Gradient Boosting (e.g., XG Boost) are trained using the processed dataset. The dataset is divided into training and testing sets to validate model performance. Ensemble learning techniques are particularly emphasized due to their ability to handle nonlinear relationships and improve prediction accuracy. Hyper parameter tuning is performed using grid search and cross-validation techniques to optimize model performance.

**Table 2 : Feature Importance Ranking**

Feature	Importance (%)
Material Cost	35%
Labor Cost	25%
Built-up Area	20%
Location	12%
Equipment Cost	8%

Model evaluation is conducted using standard regression metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared ( $R^2$ ) score. These metrics provide insights into prediction accuracy, error distribution, and model reliability. The best-performing model is selected based on its ability to minimize prediction errors and generalize well to unseen data.

To enhance usability, the selected model is integrated into a web-based application. The frontend interface is developed using web technologies such as HTML, CSS, and JavaScript, allowing users to input project parameters and receive instant cost estimates. The backend system processes user inputs, applies the trained machine learning model, and generates real-time predictions. This integration ensures that complex AI processes remain accessible to non-technical users.

Furthermore, the system incorporates dynamic cost integration by linking with external data sources such as material price databases and regional cost indices. This enables real-time updates of construction costs based on current market conditions, improving the practical applicability of the model. The framework also supports scalability, allowing the integration of additional datasets and advanced algorithms in future enhancements.

Finally, the overall system performance is analysed by comparing AI-based predictions with traditional estimation methods. The results demonstrate improvements in accuracy, efficiency, and consistency, highlighting the effectiveness of the proposed methodology in reducing human bias and enhancing budget forecasting reliability.

### 3. RESULTS AND DISCUSSIONS

The implementation of the AI-based construction cost estimation framework resulted in significant improvements in prediction accuracy, efficiency, and decision-making compared to traditional estimation methods. The system was evaluated using multiple machine learning models, including Linear Regression, Random Forest, and Gradient Boosting algorithms, across historical construction project datasets. The results demonstrate that AI-driven models effectively capture complex relationships among cost-influencing parameters such as built-up area, location, material quantities, and labour costs.

For the initial objective of accurate cost prediction, the integration of machine learning models significantly enhanced estimation precision. Comparative analysis between traditional manual estimation methods and AI-based predictions revealed a substantial reduction in error metrics

The Random Forest and Gradient Boosting models outperformed Linear Regression by effectively handling nonlinear relationships and feature interactions. The optimized model achieved lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values, along with a higher  $R^2$  score, indicating strong predictive capability and reliability in forecasting construction cost

**Table 3 : AI Construction Cost**

Model	MAE (₹/m <sup>2</sup> )	RMSE (₹/m <sup>2</sup> )	R <sup>2</sup> Score	Accuracy (%)
Linear Regression	1250	1680	0.75	78%
Random Forest	720	950	0.89	88%
Gradient Boosting	480	690	0.92	92%

The Random Forest and Gradient Boosting models outperformed Linear Regression by effectively handling nonlinear relationships and feature interactions. The optimized model achieved lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values, along with a higher R<sup>2</sup> score, indicating strong predictive capability and reliability in forecasting construction costs.

The incorporation of feature engineering techniques, such as cost per unit area and material cost ratios, further improved model performance. These derived features enabled the system to normalize cost variations across projects of different scales and locations. Additionally, the inclusion of regional cost indices and standardized rates (SSR and DSR) allowed the model to reflect real-world market conditions more accurately, reducing discrepancies between predicted and actual costs.

**Table 4: superiority of AI-based estimation over traditional methods**

Parameter	Traditional Method	AI-Based Method
Accuracy	Moderate	High
time Required	High	Low
Human Bias	High	Low
Scalability	Limited	High
Data Handling	Manual	Automated

A key outcome of this research is the development of an AI-powered web-based application that enables real-time cost estimation. The system allows users to input project parameters and instantly receive cost predictions, thereby eliminating delays associated with manual calculations. The user interface simplifies complex machine learning processes and enhances accessibility for non-technical users. The backend model processes inputs dynamically and generates results within seconds, significantly improving time efficiency in project planning.

**Table 5: System Performance Improvements**

Parameter	Before AI	After AI	Improvement
Estimation Accuracy	75%	92%	+17%
Estimation Time	2–3 days	Few seconds	~95% faster
Cost Overruns	20–25%	8–12%	Reduced ~50%

Decision-Making Efficiency	Moderate	High	Improved
Manual Effort	High	Low	Reduced

The system also demonstrates strong capabilities in handling uncertainty and variability in construction data. By leveraging ensemble learning techniques, the model reduces the impact of outliers and noise in the dataset. This leads to more stable and consistent predictions compared to traditional approaches, which are often influenced by subjective judgment and incomplete data.

Beyond prediction accuracy, the AI framework contributes to improved financial planning and risk management. The system enables stakeholders to evaluate multiple project scenarios and understand the impact of variations in material costs, labour rates, and project size. This predictive insight supports better budget allocation, minimizes cost overruns, and enhances overall project feasibility. The ability to integrate dynamic cost data ensures that estimates remain relevant under changing market conditions.

Furthermore, the adoption of AI-based estimation significantly reduces human bias and errors commonly associated with conventional methods. Automated data processing and model-driven predictions ensure consistency and transparency in cost estimation. The system also supports scalability, allowing integration with additional datasets and advanced algorithms for future enhancements.

Despite these advantages, certain limitations were observed. The accuracy of the model is highly dependent on the quality and diversity of the input dataset. Limited availability of standardized and large-scale construction data can affect model generalization across different regions. Additionally, complex models such as Gradient Boosting require careful tuning and computational resources to achieve optimal performance.

Overall, the results confirm that the proposed AI-based construction cost estimation system provides a robust and efficient solution for budget forecasting. The integration of machine learning, data-driven insights, and real-time application significantly improves estimation accuracy, reduces project risks, and enhances decision-making in construction project management.

#### 4. CONCLUSIONS

The present study successfully develops a comprehensive framework for construction cost estimation and budget forecasting using Artificial Intelligence (AI), machine learning algorithms, and data-driven analytical techniques. The research demonstrates that traditional estimation methods based on manual calculations and expert judgment can be significantly

improved through intelligent and automated systems. The proposed framework improves prediction accuracy, reduces human errors, and supports better financial planning and decision-making in construction projects.

The study shows that AI-based models provide reliable and consistent cost predictions by identifying complex relationships among different project parameters. Machine learning algorithms such as Random Forest and Gradient Boosting effectively analyze historical project data and improve forecasting performance. The use of feature engineering techniques, including cost per unit area and material cost ratios, further enhances the accuracy and adaptability of the model across various construction conditions.

A web-based cost estimation system was also developed to demonstrate the practical application of the proposed framework. The system enables users to enter project details and receive instant budget predictions, reducing the time and effort involved in manual estimation processes. The integration of dynamic data sources such as SSR and DSR helps maintain current market relevance and improves the realism of cost forecasts.

The overall findings confirm that AI-driven cost estimation systems outperform conventional methods in terms of prediction accuracy, efficiency, consistency, and risk management. The framework also offers economic benefits by minimizing cost overruns, improving resource allocation, and supporting effective budget control. Furthermore, the integration of machine learning, data preprocessing, feature engineering, and web deployment creates a unified and scalable cost management system. The developed framework can be further expanded with advanced AI models, larger datasets, and real-time market information to enhance future construction project planning and management.

Future research in AI-based construction cost estimation and budget forecasting should focus on developing intelligent, automated, and real-time decision-support systems. Advanced AI models such as Deep Neural Networks (DNN), Long Short-Term Memory (LSTM), and Transformer-based algorithms can improve prediction accuracy for complex projects. The integration of Artificial Intelligence with Building Information Modelling (BIM), cloud computing, and edge computing can enable real-time monitoring, digital twin environments, and efficient resource management. Furthermore, technologies such as Natural Language Processing (NLP) and computer vision can automate data extraction from BOQs, contracts, and drawings,

reducing manual effort and enhancing estimation accuracy. Multi-output prediction systems that simultaneously estimate project cost, duration, and risks can further improve decision-making and sustainable construction management.

## REFERENCES

- [1] I. Bad El Raze, M., & Awed, A. (2020). Identifying dominant cost factors in highway construction using machine learning techniques.
- [2] II. Abed, A. (2022). Machine learning algorithms for construction cost prediction: A systematic review. *International Journal of Advanced Research in Engineering and Technology (IJARET)*, 13(4), 112–120...
- [3] Abed, Y. G. Has an, T. M., Sahrawi, R. N., & Maser, Z. K. (2022). Machine learning algorithms for construction cost prediction: A systematic review. *Bulletin of Electrical Engineering and Informatics*, 11(3), 112–124.
- [4] Ahab Dayboy, D. D., & Smith, S. D. (2014). Dealing with construction cost overruns using data mining. *Construction Management and Economics*, 32(7–8), 682–694.
- [5] Ahmed, S., Rashid, K., & Omar, H. (2021). Assessment of construction project cost estimating accuracy. *Open Civil Engineering Journal*, 15(1), 290–303
- [6] Eliza, S. F. M. (2025). Explainable machine learning to predict construction cost using Random Forest and SHAP. *Infrastructures*, 10(2), 21.
- [7] Lazuli, A. (2024). Comparison of machine learning models for construction cost estimation. *Journal of Civil Engineering and Construction Technology*, 15(2), 45–58
- [8] Ashok, S., Relish, M., Eudora, M., & Shelton, J. (2025). Smart prediction of construction costs using ML model. *International Journal of Scientific Research and Engineering Development (IJSRED)*, 8(3).
- [9] Bark, G. A. (2019). Factors affecting accuracy of cost estimates at tendering phase. *International Journal of Civil Engineering and Technology (IJCIET)*, 10(1), 1335–1348.
- [10] Carpenter, J., Wu, C. Y., & Sty, N. U. (2024). Leveraging large language models for cost prediction. *Arrive preprint*, arrive: 2409.09617.
- [11] Cassandra, J., Menarche, C., Pancetta, C., & Pagan, A. (2024). Structured cost data integration and validation in AEC/FM industry.
- [12] Chen, G., Zhao, L., & Zhang, Y. (2025). Machine learning-based cost estimation for office buildings. *Buildings*, 15(11), 1802.
- [13] Chen, L. (2023). Construction cost prediction using Random Forest. *AIMS Mathematics*, 8(12), 21900–21920.
- [14] Curt, D., Achebe's, F., González Verona, J. M., & Poza, D. (2024). Impact of uncertainties on project cost reserves. *Arrive preprint*, arrive: 2406.03500.
- [15] Elchaig, T. M. S., & Boussabaine, A. H. (1998). Artificial neural system for cost estimation. In *Proceedings of ARCOM Conference*, 219–226.
- [16] Elmo salami, H., & Ely many, A. (2019). Comparison of AI techniques for cost prediction. *Arrive preprint*, arXiv: 1909.11637.
- [17] Flyvbjerg, B., Hon, C. K., & Fok, W. H. (2016). Reference class forecasting for roadwork projects. *Proceedings of ICE*, 169(6).
- [18] Gupta, R., et al. (2022). ANN analysis for construction cost estimation. In *NCCEI 2022 Proceedings*.
- [19] Hanna, A. S. (2007). Accuracy in cost estimation. *University Malaysia Pahang Repositor*