DFBI 2025 aims to encourage the international exchange of innovative ideas between researchers from academia and industry. In addition to knowledge dissemination, the conference offers a valuable platform for professional networking, particularly benefiting university professors, graduate students, and postdoctoral researchers.

Research Article

Exploring the Use of Machine Learning in Enhancing Bidding Decisions for Construction Projects

Qadri Shaheen[1], Suleiman Ashur[2], Huthaifa Ashqar[3]

Department of Civil and Environmental Engineering, A. James Clark School of Engineering, University of Maryland, College Park, MD 20742, USA
GameAbove College of Engineering and Technology, Eastern Michigan University, Ypsilanti, MI 48197, USA
Data Science Department, Arab American University, Jenin, Palestine. and AI @ Columbia University, New York, NY 10027, USA
Correspondence: shaheen@umd.edu

Abstract

This study addresses the critical need to enhance competitive bidding strategies in construction by revisiting Friedman's 1956 bidding model and incorporating Ioannou's revised equations to improve predictive accuracy. The research fills a significant gap in integrating machine learning techniques into bidding theory, which offers a data-driven approach to optimize bid decision-making. Using synthetic data, logistic regression served as a baseline model, while Random Forest classifiers outperformed with 98% accuracy by addressing class imbalance and effectively capturing the non-linear relationships among key variables, such as reserve price and bid-to-cost ratio. The findings revealed that machine learning models could simplify complex bidding theories and provide contractors with actionable insights, supporting bid or no-bid decisions. However, reliance on synthetic data limits the generalizability of these results. Future work should focus on validating the proposed models using real-world bidding datasets and exploring advanced techniques, such as ensemble methods, to enhance predictive performance. This study underscores the potential of machine learning to transform traditional bidding practices, which offers both theoretical advancements and practical implications for construction management.

Keywords: Bid Decision-Making; Synthetic Bidding Data; Machine Learning Algorithms; Construction Bidding.

Highlights

- Machine learning improves accuracy of bid/no-bid decisions in construction projects.
- Random Forest model achieved 98% accuracy using synthetic bidding data.
- Reserve price and bid-to-cost ratio are key predictors in competitive bidding outcomes.

Cite this article:
Shaheen, Q. H., Ashur, S. A., & Ashqar, H. I. (2025). Exploring the Use of Machine Learning in Enhancing Bidding Decisions for Construction Projects. Title of the Journal, Volume 2025, Page range: 24-32

1 Introduction

The prediction and optimization of competitive bidding strategies in construction management have been extensively studied since Lawrence Friedman's seminal work in 1956. Friedman's competitive bidding model formalized a method for determining optimum bids in closed bidding scenarios, where a contractor competes against an unknown number of opponents. This model utilized historical bidding patterns to estimate winning probabilities, providing a systematic approach to bid decision-making (Friedman, 1956). Despite its foundational impact, Friedman's model has notable limitations, especially in handling scenarios involving zero or unknown numbers of competitors. Photios G. Ioannou's critical assessment highlighted inaccuracies in Friedman's original formulations when competitor numbers approached zero. Ioannou's revised equations offered significant corrections, enhancing the reliability of probability estimates under diverse competitive conditions (Ioannou, 2019).

In parallel with these theoretical advancements, the construction industry has witnessed a growing adoption of machine learning (ML) and artificial intelligence (AI) techniques to address the complexities of bid prediction and risk management. Machine learning applications, such as Random Forest, have been extensively tested and applied across various research disciplines since their introduction by Breiman (2001). In recent years, the adoption of machine learning techniques within the construction industry has become increasingly prevalent, with applications ranging from cost prediction (Kim & Jung, 2018) and road construction cost estimation (Jaafari et al., 2021) to the analysis of work zone crashes (Ashqar et al., 2021) and the prediction of construction-stage carbon emissions during the early design phase (Fang et al., 2021).

Recent advancements in AI and ML have further transformed risk management and bid pricing strategies within Engineering, Procurement, and Construction (EPC) projects. For example, the integration of AI-assisted game theory into bid pricing under uncertainty has demonstrated that combining AI with traditional game-theoretic models can enhance the accuracy of bid estimations and facilitate more effective risk-sharing among stakeholders (Serugga, 2025). This approach addresses the inherent uncertainties in construction bidding environments, providing a robust framework for decision-making. Building upon this, the application of predictive analytics in supply chain management underscores the value of data-driven methodologies, as machine learning models leveraging real supplier data have been shown to optimize procurement strategies and improve supplier evaluation, thereby reducing costs and increasing supply chain reliability.

The evolution of digitalized risk analysis tools for EPC contractors further exemplifies the practical benefits of AI in this domain. Machine learning-based tools for extracting and analysing risks from technical specifications have led to substantial improvements in both the accuracy and efficiency of risk detection processes (Park et al., 2021). These tools automate the traditionally labour-intensive task of risk identification and enable contractors to respond proactively to potential project threats. Complementing these advancements, digital modules such as the Critical Risk Check (CRC) and Term Frequency Analysis (TFA) facilitate the automatic extraction of contractual risks from ITB and contract documents, streamlining the risk assessment process and supporting more informed decision-making (Choi et al., 2021). Collectively, these studies illustrate a clear trajectory toward the integration of AI and ML in EPC project management, where the convergence of advanced analytics, automation, and domain expertise is reshaping traditional practices and setting new standards for risk mitigation and bid optimization.

The integration of machine learning algorithms into construction bidding models has thus gained increasing attention. Studies have shown that these algorithms, such as those based on algorithmic game theory, can significantly improve the bidding decision-making process. For instance, Assaad et al. (2021) developed a simulation framework to assess the impact of learning algorithms in construction bidding, demonstrating that incorporating learning can provide contractors with a competitive advantage. The study found that such integration could either double the chances of winning more projects or reduce the risks of negative profits, often referred to as the "winner's curse." Moreover, the authors highlighted that owners also benefit from the long-term cost reduction associated with these learning-based approaches. This suggests that machine learning can enhance the strategic decision-making of contractors, helping them navigate the complexities and uncertainties inherent in competitive bidding environments.

Additionally, machine learning can address some of the challenges faced by traditional bidding models, such as data incompleteness and dynamic competitor behaviours. Abotaleb and El-Adaway (2017) introduced a Bayesian-based approach to bidding markup estimation, which is particularly useful in situations where competitor data is sparse or continuously evolving. By using a multistage decision theory framework, their model accounts for the stochastic variability of cost estimates and incorporates the latest competitor behaviours, offering a more accurate prediction of bidding outcomes. This approach allows contractors to adapt to changing market conditions and make informed decisions based on incomplete data, ultimately improving both the probability of winning bids and maximizing expected profits. These findings underscore the potential of machine learning to optimize bidding strategies in construction, providing a more robust and flexible framework for contractors.

Building on this, recent research has explored machine learning techniques to further enhance the predictive power of bidding models. This study aims to explore the potential of applying machine learning algorithms—using synthetic data as a test bed—to refine and improve competitive bidding strategies. Rather than assessing real-world prediction accuracy, this research examines the feasibility of using machine learning models like logistic regression and Random Forest classifiers in interpreting bidding scenarios shaped by Ioannou's corrections to Friedman's model.

2 Methodology

The methodology of this study aims to explore the effectiveness of machine learning models in predicting bidding decisions within competitive bidding scenarios. This approach utilizes synthetic data to simulate the key variables influencing a company's decision to bid or not. The study employs both logistic regression and random forest classifiers, with hyperparameter tuning through grid search, to assess model performance and demonstrate the potential of machine learning in bidding decision-making contexts.

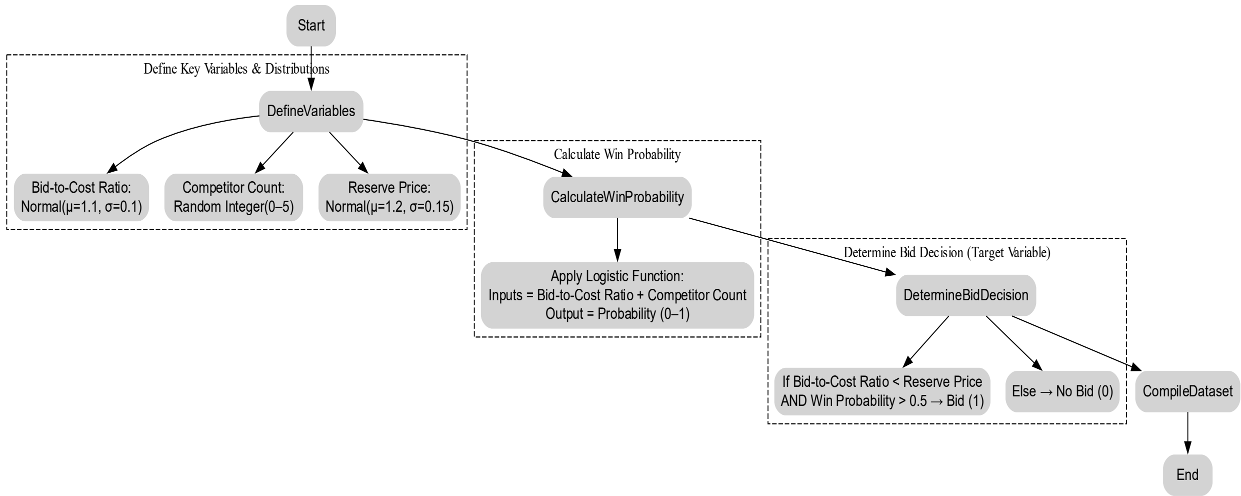
2.1.1 Data Simulation

This subsection defines the A synthetic dataset was generated to represent a competitive bidding environment as illustrated in Figure 1. The key features influencing bid decisions are:

- Bid-to-Cost Ratio (bid_cost_ratio): This feature represents the ratio of the bid price to the actual cost, simulated using a normal distribution with a mean of 1.1 and a standard deviation of 0.1. It indicates how favorable the bid price is relative to the cost.

- Competitor Count (competitor_count): The number of competitors in the bidding process, simulated as a random integer between 0 and 5. This feature reflects the level of competition faced by the bidder.
- Reserve Price (reserve_price): The highest price a client is willing to accept, simulated using a normal distribution with a mean of 1.2 and a standard deviation of 0.15. It represents the threshold price for a successful bid.
- Win Probability (win_probability): This feature calculates the probability of winning the bid, determined by a logistic function that incorporates both the bid_cost_ratio and competitor_count. The win probability serves as a critical factor in the bid decision, influencing the likelihood of success.
- Bid Decision (bid_decision): The target variable, indicating whether the company places a bid (1) or not (0). The decision is based on the condition that the bid_cost_ratio is lower than the reserve_price and the win_probability is greater than 0.5. This simulates a rational decision-making process in competitive bidding.

Figure 1. *Synthetic data generation and application in the model.*



2.2 Model Selection

Two machine learning models were chosen for evaluation: logistic regression and random forest classifiers. These models were selected for their different strengths in handling linear versus non-linear relationships.

Logistic Regression: Logistic regression is a simple and interpretable linear model that estimates the probability of a binary outcome based on one or more independent variables. In this study, it was used to assess how well the bid decision can be predicted based on the features (bid_cost_ratio, competitor_count, reserve_price, and win_probability).

Random Forest Classifier: The random forest model is an ensemble learning method that aggregates the predictions of multiple decision trees. It was chosen for its ability to capture complex, non-linear relationships and interactions between features, which might be difficult for logistic regression to model. This model was expected to perform well given the potential non-linearity of the relationship between the features and the target.

2.3 Model Training and Evaluation

The synthetic dataset was split into training and testing sets, with 80% allocated for training and 20% for testing. This allowed the model to learn from a substantial portion of the data while being evaluated on an unseen test set to assess generalizability. The split was performed using stratified sampling to ensure that the proportions of the target classes (bid decision: 0 or 1) were preserved across both sets. The two models were trained using bid_cost_ratio, competitor_count, reserve_price, and win_probability. The model's performance was evaluated using the classification metrics, including precision, recall, and F1-score. These metrics provide insights into the accuracy and robustness of the model, especially in predicting both classes (bid decision: 0 and 1).

To further optimize the random forest model, hyperparameter tuning was performed using grid search. The parameter grid included variations in the number of estimators (trees) and the maximum number of features used for splitting each node in the trees. The grid search identified the best combination of parameters of 100 trees and two maximum number of features. The optimized random forest model achieved similar performance to the initial model but with the added benefit of hyperparameter tuning to enhance accuracy and robustness.

3 Analysis and Results

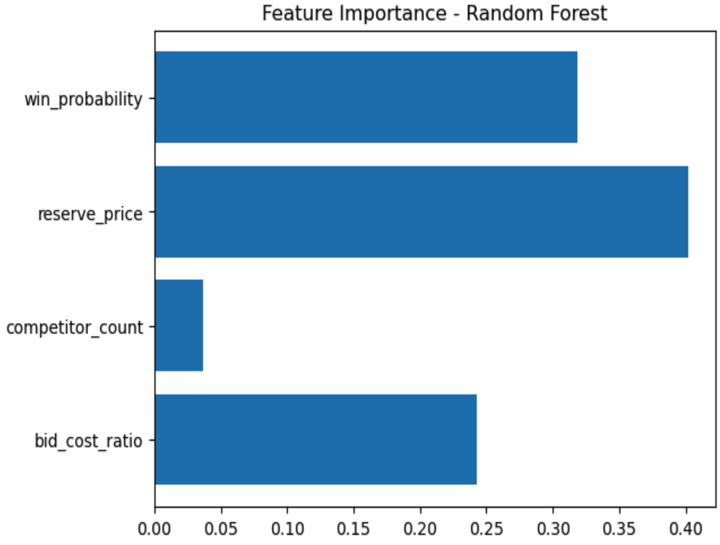
The results presented in Table 1 demonstrate a clear performance advantage of the Random Forest algorithm over Logistic Regression in predicting the desired outcomes. Random Forest achieves an accuracy of 98%, significantly outperforming Logistic Regression of 81% accuracy. This performance is consistent across other performance metrics, including recall, precision, and the F1 score, where Random Forest consistently scores above 0.97, which shows its robustness in classifying instances correctly. The high recall value indicates that Random Forest effectively captures positive instances, making it particularly useful in construction project bidding scenarios where missing a potential winning bid could result in financial losses.

Table I. Accuracy, recall, precision, and F1 Score for Logistic Regression and Random Forest.

Algorithm	Accuracy	Recall	Precision	F1 Score
Logistic Regression	0.81	0.80	0.81	0.79
Random Forest	0.98	0.97	0.97	0.97

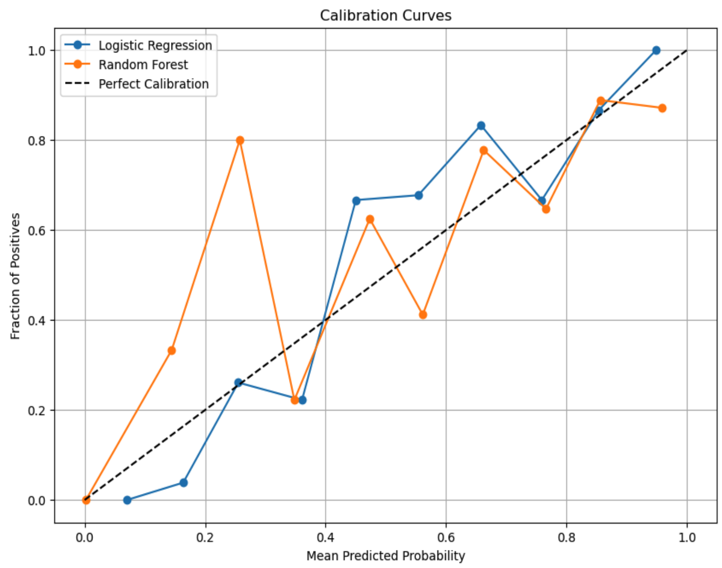
The feature importance analysis in Figure 2 sheds light on the key factors influencing prediction outcomes. The reserve price, win probability, bid-to-cost ratio emerge as the most critical determinants, indicating that financial considerations and historical success rates play a dominant role in bid outcomes. The bid cost ratio also carries a notable weight, reinforcing the practical importance of cost efficiency in competitive bidding. Interestingly, the competitor count appears to have minimal influence, suggesting that bid success is more strongly linked to internal cost structures and strategic pricing rather than the number of competing firms. This insight is particularly valuable for construction managers who need to optimize bid pricing strategies rather than focusing solely on market competition levels.

Figure 2. *Features importance results using Random Forest.*



However, the calibration curves in Figure 3 further compare the predictive reliability of both models. A well-calibrated model should align closely with the dashed black line, which represents perfect calibration. The Random Forest model demonstrates higher fluctuations in lower probability regions, indicating some overconfidence in certain predictions. However, at higher probability levels, both models show improved calibration, with Logistic Regression aligning more closely with the perfect calibration line in some areas. This suggests that while Random Forest offers superior accuracy, it may require additional calibration to improve probability estimation in real-world applications, such as risk assessment in large-scale infrastructure projects.

Figure 3. *Calibration curve to compare the fitness between Random Forest and Logistic Regression.*



From a practical perspective, these findings suggest that Random Forest is a more reliable tool for bid outcome predictions in construction management, particularly when making high-stakes financial decisions. However, Logistic Regression's smoother calibration curve could be advantageous in situations where precise probability estimates are needed, such as risk assessments or contingency planning. In practice, construction managers could benefit from a hybrid approach, leveraging Random Forest's predictive power while incorporating Logistic Regression for fine-tuned probability assessments to guide bidding strategies more effectively.

4 Discussion

This study has revealed important insights into competitive bidding behavior and the potential of machine learning in improving decision-making processes within the construction industry. By revisiting Friedman's 1956 bidding model and integrating Ioannou's corrective equations, the research improves the accuracy of win probability calculations, particularly in low-competition scenarios. These revisions provide a more realistic foundation for modeling competitive bidding, which is crucial for contractors when deciding whether to participate in a bid.

The application of machine learning models, such as logistic regression and Random Forest, has provided deeper insights into how different variables influence bidding decisions. Logistic regression, although useful as a baseline, was unable to capture the complexities of the bidding environment. In contrast, the Random Forest model excelled by handling the non-linearities and interactions between features. The Random Forest model's ability to manage class imbalance is particularly noteworthy, as real-world bidding data often exhibits this challenge.

Feature importance analysis revealed that the reserve price and bid-to-cost ratio were the most influential factors in bid decisions, which aligns with construction industry practices where contractors are focused on profitability and expected returns. Additionally, the ability of the Random Forest model to identify conditions where no competitors are present demonstrates its capacity to model strategic decision-making more effectively than traditional methods.

The development of an automated bidding decision system, powered by the trained machine learning models, represents a significant step toward making data-driven decisions in the bidding process. This system could help contractors assess bid viability in real-time, offering valuable insights that are grounded in data rather than intuition alone. Such a system could streamline the bidding process and improve overall decision-making efficiency.

However, several challenges need to be addressed. The reliance on synthetic data for model training is a significant limitation, as real-world data introduces additional complexities, including economic shifts, contractor strategies, and unpredictable market forces. To ensure the models' robustness and applicability, future research must test them on actual bidding datasets. This will allow for better assessment of their generalizability and refinement of the models to account for the noise and variability inherent in real-world data.

Additionally, although Random Forest demonstrated strong performance in handling class imbalances, further refinement is needed to address these issues more effectively. Techniques such as synthetic oversampling or cost-sensitive learning could be explored to improve the model's ability to predict underrepresented classes (e.g., "bid" decisions in competitive scenarios with fewer bids).

Table II. Comparison of Traditional Bidding Models (e.g., Friedman, Ioannou) vs. Machine Learning (e.g., Random Forest).

Dimension	Traditional Models	Machine Learning Models
Handling non-linearity	Limited	Efficient
Class Imbalance	Not addressed	Handled effectively
Data Requirements	Historical bidding patterns	Historical bidding patterns
Interpretability	High	Moderate (requires feature analysis)
Real-world Applicability	Limited by assumptions	High (with real-world data validation)

5 Conclusions

The methodology demonstrated the feasibility of using machine learning, specifically logistic regression and random forest classifiers. The results suggest that machine learning can be effectively applied to competitive bidding analysis, providing valuable insights into bid decision-making processes. Future work should involve applying these models to real-world data to validate their predictive power and generalizability.

This research explores the integration of machine learning with theoretical bidding models, specifically using Ioannou's corrections to Friedman's 1956 bidding model, to enhance predictive accuracy in construction bidding decisions. The study demonstrates the potential of using machine learning, particularly Random Forest and logistic regression models, to improve the decision-making process in competitive bidding environments. While the models performed exceptionally well on synthetic data, their application to real-world data remains critical to confirm their robustness and practical applicability.

The findings show that the Random Forest model outperforms logistic regression by effectively capturing non-linear relationships and interactions among the bid decision features, such as bid-to-cost ratio, reserve price, and competitor count. The incorporation of Ioannou's revised equations into the data generation process led to more accurate win probability estimations, especially in scenarios with minimal or no competition—an area where Friedman's original model was limited. These advancements lay the foundation for the development of data-driven decision support tools that can aid contractors in optimizing their bidding strategies.

However, as the study relies on synthetic data, its findings are exploratory and not yet conclusive for real-world applications. Future work will focus on testing these models against actual bidding datasets, incorporating the complexities and variations found in real-world bidding environments. Additionally, expanding the feature set to include macroeconomic variables and exploring further model improvements, such as handling class imbalances and refining ensemble methods, will be essential for ensuring the practical utility and generalizability of these models in real-world settings. Lastly, researchers are encouraged to apply LLMs on bidding decisions given the abilities of these models to predict and make decision based on historical data.

Funding

No external funding was provided for this research. The authors conducted the study independently based on their own academic interests.

Data Availability Statement

This research used synthetic data generated through Monte Carlo simulation.

Conflicts of Interest

The authors declare no conflict of interest.

References

Abotaleb, I. S., & El-adaway, I. H. (2017). Construction bidding markup estimation using a multistage decision theory approach. Journal of construction engineering and management, 143(1), 04016079.

Ashqar, H. I., Shaheen, Q. H., Ashur, S. A., & Rakha, H. A. (2021, September). Impact of risk factors on work zone crashes using logistic models and Random Forest. In 2021 IEEE International Intelligent Transportation Systems Conference (ITSC) (pp. 1815-1820). IEEE.

Assaad, R., Ahmed, M. O., El-Adaway, I. H., Elsayegh, A., & Siddhardh Nadendla, V. S. (2021). Comparing the impact of learning in bidding decision-making processes using algorithmic game theory. Journal of management in engineering, 37(1), 04020099.

Breiman, L. (2001). Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Choi, S. J., Choi, S. W., Kim, J. H., & Lee, E. B. (2021). AI and text-mining applications for analyzing contractor's risk in invitation to bid (ITB) and contracts for engineering procurement and construction (EPC) projects. *Energies*, *14*(15), 4632.

Fang, Y., Lu, X., & Li, H. (2021). A random forest-based model for the prediction of construction-stage carbon emissions at the early design stage. Journal of Cleaner Production, 328, 129657.

Friedman, L. (1956). A competitive-bidding strategy. Operations research, 4(1), 104-112.

Ioannou, P. G. (2019). Friedman's bidding model: Errors and corrections. Journal of construction engineering and management, 145(10), 04019058.

Jaafari, A., Pazhouhan, I., & Bettinger, P. (2021). Machine learning modeling of forest road construction costs. Forests, 12(9), 1169.

Kim, J. M., & Jung, H. (2018). Predicting bid prices by using machine learning methods. Applied Economics, 51(19), 2011–2018. https://doi.org/10.1080/00036846.2018.1537477.

Serugga, J. (2025). AI-Assisted Game Theory Approaches to Bid Pricing Under Uncertainty in Construction. *AppliedMath*, *5*(2), 39

Park, M. J., Lee, E. B., Lee, S. Y., & Kim, J. H. (2021). A digitalized design risk analysis tool with machine-learning algorithm for EPC contractor's technical specifications assessment on bidding. *Energies*, *14*(18), 5901.