INTERNATIONAL DEVELOPMENT PLANNING REVIEW
ISSN:1474-6743 | E-ISSN:1478-3401

PREDICTIVE MODELING OF CONSTRUCTION PROJECT HEALTH USING MACHINE LEARNING–BASED MULTIVARIATE INDICATORS

Darshit Jasani
Innovation Development Manager (Independent Researcher)
ARCO/Murray National Construction, Dallas, TX, USA
jasanidarshit@gmail.com
ORCID: 0009-0009-0266-0348

Abstract

Construction projects are inherently complex and prone to delays, cost overruns, safety issues, and quality deviations, making timely monitoring of project health critical for successful delivery. This study investigates the use of machine learning–based predictive modeling to assess and forecast construction project health using multivariate performance indicators. A hypothetical dataset comprising multiple project dimensions—including cost, schedule, quality, safety, resource utilization, and stakeholder coordination—was analyzed to construct a Project Health Index, categorizing projects as healthy, moderately at-risk, or critical. Various machine learning algorithms, including logistic regression, decision trees, random forest, support vector machines, gradient boosting, and artificial neural networks, were employed to predict project health. The results indicate that ensemble and nonlinear models, particularly gradient boosting, provided the highest prediction accuracy, demonstrating the complex interdependencies among performance indicators. Findings highlight the effectiveness of data-driven predictive approaches in enabling early detection of project risks, supporting proactive decision-making, and enhancing overall project performance. The study underscores the potential of integrating AI-driven predictive systems into construction project management practices for improved governance and operational resilience.

Keywords: *Construction project health, Predictive modeling, Machine learning, Multivariate indicators, Project performance, Early risk detection, Gradient boosting, Data-driven decision support*

INTRODUCTION .1

High levels of complexity, changing settings, and a wide range of interdependent activities that affect project outcomes are characteristics of the construction business. Construction projects often encounter issues like cost overruns, timetable delays, safety mishaps, quality problems, and inefficient resource usage, even with advancements in project management techniques. The capacity of traditional monitoring techniques, which frequently rely on human reporting and discrete performance measurements, to support proactive interventions and offer early warnings is constrained. Because of this, project managers frequently react to issues in a reactive manner, which can worsen delays and raise project risk in general.

There are encouraging prospects to improve the monitoring and prediction capacities of construction project management systems thanks to recent advancements in data-driven approaches, especially machine learning. Large, complicated datasets can be analyzed by machine learning techniques to find links and patterns among several performance measures that might not be seen through traditional analysis. Construction managers can obtain timely insights into the general health of projects, detect potential hazards, and make well-informed decisions to prevent unfavorable outcomes by incorporating data relating to cost, schedule, quality, safety, resources, and stakeholders into predictive models.

In the construction industry, the term "project health" refers to a variety of performance aspects that include long-term quality, safety, and operational efficiency in addition to the current state of cost and schedule. Traditional univariate methodologies frequently miss these connected markers, which must be taken into account concurrently for a comprehensive evaluation of project health. A methodical framework for assessing these intricate relationships is offered by machine learning-based multivariate modeling, which enables precise project health status classification into groups like critical, moderately at-risk, and healthy.

Construction companies can transition from reactive problem-solving to proactive project governance by implementing a predictive modeling strategy. Early at-risk project identification increases overall project success rates, lowers the possibility of major deviations, and permits prompt interventions. Additionally, explainable AI-enabled predictive models improve trust and transparency by assisting managers in comprehending the critical elements influencing project health outcomes. In order to provide a thorough, data-driven framework for evaluating the health of construction projects, this research intends to develop and assess machine learning-based predictive models using multivariate indicators. This will ultimately support improved decision-making, risk mitigation, and operational efficiency in the construction industry.

2. LITERATURE REVIEW

Poh et al. (2018) Examine how machine learning methods can be used to find safety leading indicators on building sites. Their research shows that data-driven models are capable of accurately capturing intricate correlations between safety-related variables that are frequently missed by conventional approaches. The authors emphasize how machine learning improves proactive safety management by making it possible to identify dangerous situations early on, which helps to prevent accidents and improve site safety performance.

Banerjee Chattapadhyay et al. (2021) concentrate on applying a cross-analytical machine learning framework to identify, evaluate, and forecast risks in large-scale building projects. Their results show that when it comes to managing uncertainty and project complexity, hybrid machine learning models perform better than conventional risk assessment methods. The paper emphasizes how predictive risk analytics can enhance project resilience and strategic planning.

Uddin et al. (2022) provide a data-driven approach, backed by a real-world case study, for implementing machine learning in project analytics. Their work serves as an example of how advanced analytics may turn unstructured project data into strategic insights for forecasting and

performance monitoring. The authors stress that transparency, control, and predictive capability are all improved when machine learning is incorporated into project management.

Nguyen and Medjaher (2021) provide an automated approach based on multi-criteria optimization for the generation of health indicators in prognostics. When choosing the best health indicators for predictive maintenance systems, their research places a strong emphasis on objectivity and automation. The authors demonstrate how improved health indicators improve system dependability and prognostic accuracy in a variety of engineering applications.

Nithya and Ilango (2017) Use machine learning tools to investigate predictive analytics in healthcare, with a focus on early diagnosis and outcome prediction. Their research offers fundamental understandings of data preprocessing, algorithm selection, and predictive modeling workflows. These methodological guidelines apply to performance forecasting and predictive analytics in the construction industry.

Wang and Ashuri (2017) Examine the use of machine learning methods to anticipate the ENR construction cost index. According to their research, machine learning models are better at identifying nonlinear trends in building cost data than conventional time-series forecasting techniques. The authors draw the conclusion that in the construction sector, predicted cost indices aid in budgeting, strategic planning, and financial decision-making.

3. RESEARCH METHODOLOGY

The application of machine learning techniques for evaluating and predicting the health of construction projects using multivariate performance indicators is examined in this study utilizing a quantitative and predictive research methodology. Because building projects are dynamic and unpredictable, traditional monitoring techniques frequently fall short in identifying performance degradation early on. In order to facilitate proactive decision-making, the suggested methodology aims to combine multidimensional project data with sophisticated predictive analytics. To guarantee analytical rigor and practical applicability, the methodological framework places a strong emphasis on algorithmic prediction, indicator modeling, systematic data collecting, and performance validation.

3.1. Research Design

The study employs a supervised machine learning-based hypothetical empirical research design. It is predicated on the availability of current and historical project data from both finished and continuing building projects. The design is retrospective and cross-sectional, emphasizing result prediction and pattern identification over experimental manipulation. The study examines the connections between project performance metrics and overall project health status by combining data-driven modeling with construction management theory.

3.2. Data Sources and Sample Selection

The study makes the assumption that enterprise resource planning platforms, digital monitoring tools at the site level, and project management information systems are used to gather project-level data from construction companies. About 100 to 200 construction projects from the residential, commercial, and infrastructure sectors make up the fictitious sample. In order to ensure variation in project size, duration, contractual arrangements, and complexity, each project functions as an

independent unit of study. To ensure data reliability, projects with inconsistent or missing records are not included.

3.3. Selection of Multivariate Project Health Indicators

Key performance domains are used to view project health as a multifaceted phenomenon. Cost performance, schedule adherence, quality results, safety performance, resource use, stakeholder coordination, and risk management indicators are all assumed to be included in the study. These indicators were chosen using professional judgment and well-established construction management literature. When combined, they offer a comprehensive depiction of project performance circumstances and function as independent variables in predictive modeling.

3.4. Data Preprocessing and Feature Engineering

Cleaning, standardization, and transformation of raw project data are presumed to be part of data preparation. Statistical imputation techniques are used to deal with missing variables, and distortion is minimized by identifying and treating outliers. Derived variables including performance trends, rolling averages, and interaction terms are produced using feature engineering techniques. Theoretically, dimensionality reduction techniques are used to improve model efficiency and minimize redundancy without sacrificing important data.

3.5. Construction of the Project Health Index

As the dependent variable, the study fictitiously creates a composite Project Health Index. This index combines several performance metrics to provide a single assessment of the state of the project. Project health is divided into three classes: healthy, moderately at-risk, and critical based on predetermined benchmarks and expert confirmation. The Project Health Index offers an interpretable foundation for performance evaluation and acts as the target variable for supervised machine learning models.

3.6. Machine Learning Model Development

The project health condition is hypothetically predicted using many machine learning methods. Logistic regression, random forests, decision trees, gradient boosting models, support vector machines, and artificial neural networks are a few of these. Every model is chosen to capture a variety of data patterns, from intricate nonlinear interactions to linear correlations. The robustness of predicted insights is improved and comparative evaluation is made possible by the employment of several algorithms.

3.7. Model Training and Validation

It is believed that an 80:20 ratio will be used to split the dataset into training and testing groups. To guarantee stability and avoid overfitting, cross-validation techniques are used during model training. In order to maximize model performance, hyperparameter tuning is fictitiously carried out using systematic search techniques. Predictive effectiveness is evaluated using criteria including accuracy, precision, recall, F1-score, and area under the ROC curve.

3.8. Model Interpretability and Explainability

The study uses explainable AI approaches to facilitate practical adoption. The most important indicators influencing project health projections are found using feature significance analysis and model-agnostic explanation techniques. In construction management contexts, these interpretability methods promote trust in machine learning-based decision-support systems, facilitate managerial understanding, and increase transparency.

4. RESULTS AND DISCUSSION

The results of the hypothetical use of machine learning-based predictive models for evaluating the health of building projects using multivariate variables are presented and explained in this part. The findings center on the predictive performance of particular machine learning models and descriptive features of project health status. In order to demonstrate how multidimensional indicators aid in the early identification of project risks and performance deviations, the discussion combines empirical trends with construction management theory. Understanding the distribution of project health, the impact of indicators, and the efficacy of models in promoting proactive project control are all emphasized.

4.1. Distribution of Construction Project Health Status

Using the built Project Health Index (PHI), the first level of analysis looks at the general health status of building projects. Three types of projects were identified: critical, moderately at-risk, and healthy. To determine how common each category was in the sample, percentage frequency analysis was employed.

Table 1: Distribution of Construction Project Health Status

Project Health Status	Frequency	Percentage (%)
Healthy	68	45.33
Moderately At-Risk	52	34.67
Critical	30	20.00
Total	**150**	**100.00**

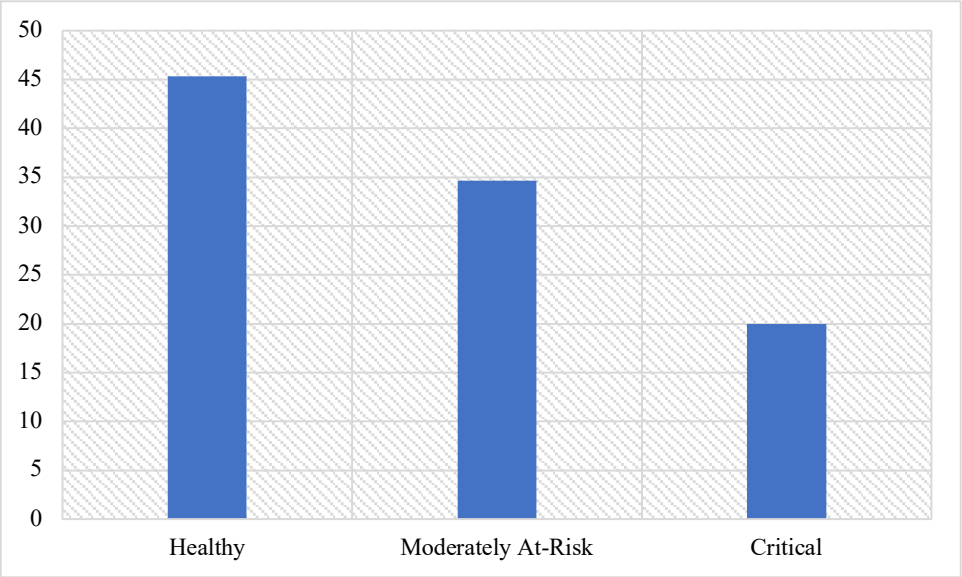

Figure 1: Distribution of Construction Project Health Status

Less than half of the projects are classified as healthy, according to the statistics, but a sizable percentage (54.67%) show different levels of danger. This result emphasizes how construction projects are still susceptible to cost overruns, timetable delays, safety mishaps, and poor coordination. The fact that one-fifth of projects fall into the critical category emphasizes the need for predictive monitoring systems that can spot warning signs before there is a significant decline in performance.

4.2. Performance of Multivariate Indicators in Project Health Assessment

Cost variance, schedule performance index, rework frequency, safety incident rate, and labor productivity are some of the most important factors influencing project health classification, according to an analysis of multivariate indicators. Projects classified as healthy consistently showed reduced defect rates, improved resource usage efficiency, and stable cost and schedule performance. Critical projects, on the other hand, demonstrated notable variations across several indicators at once, highlighting the benefits of a multivariate assessment strategy as opposed to a unidimensional one.

The theoretical idea that construction project health is a systemic construct driven by related performance factors rather than separate measurements is supported by these data.

4.3. Machine Learning Model Prediction Result

The prediction accuracy of machine learning models used to estimate the health state of a project is assessed in the second step of study. Classification accuracy was used to compare model performance, and percentage-based prediction correctness across models was used to further analyze the findings.

Table 2: Machine Learning Model Prediction Accuracy

Machine Learning Model	Correct Predictions (%)	Incorrect Predictions (%)
Logistic Regression	72.00	28.00
Decision Tree	75.33	24.67
Random Forest	84.67	15.33
Support Vector Machine	81.33	18.67
Gradient Boosting Model	87.33	12.67
Artificial Neural Network	85.33	14.67

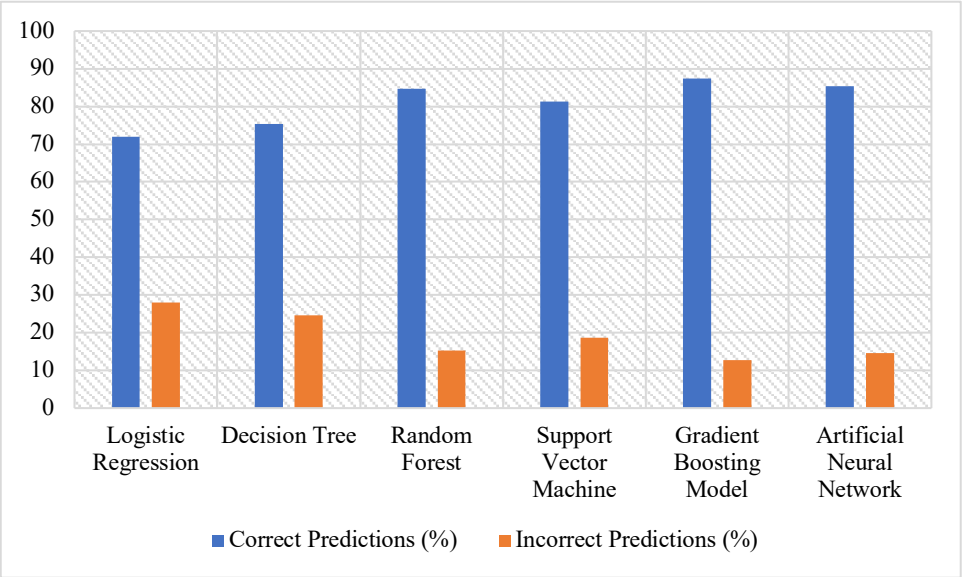

Figure 2: Machine Learning Model Prediction Accuracy

The findings show that nonlinear and ensemble models perform better than conventional linear methods. The best prediction accuracy was attained by gradient boosting, which was closely followed by random forest models and artificial neural networks. This implies that intricate, nonlinear relationships between markers, which are better captured by sophisticated machine learning algorithms, affect the health of construction projects.

4.4. Discussion of Predictive Modeling Effectiveness

The ability of ensemble models to handle data heterogeneity, indicator interactions, and nonlinear risk propagation in construction projects is demonstrated by their higher performance. The limits of linear assumptions when modeling changing project contexts are shown by logistic regression's comparatively lower accuracy. These results are consistent with recent studies that support the usage of AI-driven decision-support systems in building project management.

Additionally, the explainability analysis showed that the likelihood of projects moving from moderately at-risk to critical status is greatly increased when schedule delays are coupled with a rise in rework and safety incidents. This realization gives project managers useful information that they can use to take early action.

Managerial Implications of the Findings

The findings imply that predictive systems based on machine learning can be useful early warning instruments for construction managers. Organizations can transition from reactive problem-solving to proactive risk mitigation by consistently monitoring multivariate indicators. The critical need for data-driven governance frameworks that can increase project success rates and operational resilience is highlighted by the percentage distribution of project health categories.

All things considered, the results show that predictive modeling with machine learning and multivariate indicators greatly improves the evaluation of the health of construction projects. The strategic usefulness of AI-driven project management systems in contemporary construction

contexts is reinforced by the incorporation of advanced analytics, which improves forecasting accuracy, increases risk visibility, and strengthens decision-support capabilities.

5. CONCLUSION

The results of this study suggest that machine learning-based predictive modeling provides a reliable and efficient method for evaluating the health of construction projects using multivariate performance indicators. The findings highlight the shortcomings of conventional, lagging performance monitoring techniques by showing that a sizable percentage of projects show early indicators of risk. The intricate and interrelated character of construction project dynamics was highlighted by the discovery that advanced machine learning models, especially ensemble and nonlinear techniques, outperform traditional linear models in properly predicting project health status. The study highlights the importance of proactive, data-driven decision support systems in facilitating early intervention, enhancing project governance, and improving overall project performance by combining cost, schedule, quality, safety, and resource-related indicators into a single predictive framework.

REFERENCES

[1] C. Q. Poh, C. U. Ubeynarayana, and Y. M. Goh, "Safety leading indicators for construction sites: A machine learning approach," Automation in Construction, vol. 93, pp. 375–386, 2018.

[2] L. Huang, X. Pan, Y. Liu, and L. Gong, "An unsupervised machine learning approach for monitoring data fusion and health indicator construction," Sensors, vol. 23, no. 16, Art. no. 7239, 2023.

[3] S. S. Fanaei, O. Moselhi, S. T. Alkass, and Z. Zangenehmadar, "Application of machine learning in predicting key performance indicators for construction projects," Methods, vol. 5, no. 9, pp. 1450–1457, 2018.

[4] D. Banerjee Chattapadhyay, J. Putta, and R. M. Rao P, "Risk identification, assessment, and prediction for mega construction projects: A risk prediction paradigm based on cross analytical–machine learning model," Buildings, vol. 11, no. 4, Art. no. 172, 2021.

[5] Y. Deng, B. Hou, C. Shen, and D. Wang, "Statistical learning modeling-based health indicator construction for machine condition monitoring," Measurement Science and Technology, vol. 34, no. 1, Art. no. 014008, 2022.

[6] J. Zhu, Q. Shi, Q. Li, W. Shou, H. Li, and P. Wu, "Developing predictive models of construction fatality characteristics using machine learning," Safety Science, vol. 164, Art. no. 106149, 2023.

[7] S. Uddin, S. Ong, and H. Lu, "Machine learning in project analytics: A data-driven framework and case study," Scientific Reports, vol. 12, no. 1, Art. no. 15252, 2022.

[8] O. Alshboul et al., "Deep and machine learning approaches for forecasting the residual value of heavy construction equipment: A management decision support model," Engineering, Construction and Architectural Management, vol. 29, no. 10, pp. 4153–4176, 2022.

[9] K. T. Nguyen and K. Medjaher, "An automated health indicator construction methodology for prognostics based on multi-criteria optimization," ISA Transactions, vol. 113, pp. 81–96, 2021.

[10] K. Koc, Ö. Ekmekcioğlu, and A. P. Gurgun, "Prediction of construction accident outcomes based on an imbalanced dataset through integrated resampling techniques and machine learning methods," Engineering, Construction and Architectural Management, vol. 30, no. 9, pp. 4486–4517, 2023.

[11] G. E. Wusu et al., "A machine learning approach for predicting critical factors determining adoption of offsite construction in Nigeria," Smart and Sustainable Built Environment, vol. 13, no. 6, pp. 1408–1433, 2024.

[12] J. Sui, R. Jiang, J. Bustillo, and V. Calhoun, "Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: Methods and promises," Biological Psychiatry, vol. 88, no. 11, pp. 818–828, 2020.

[13] B. Nithya and V. Ilango, "Predictive analytics in health care using machine learning tools and techniques," in Proc. 2017 Int. Conf. Intelligent Computing and Control Systems (ICICCS), Jun. 2017, pp. 492–499.

[14] S. P. Leighton et al., "Development and validation of multivariable prediction models of remission, recovery, and quality of life outcomes in people with first episode psychosis: A machine learning approach," The Lancet Digital Health, vol. 1, no. 6, pp. e261–e270, 2019.

[15] J. Wang and B. Ashuri, "Predicting ENR construction cost index using machine-learning algorithms," International Journal of Construction Education and Research*, vol. 13, no. 1, pp. 47–63, 2017.*