IRJSE
Int. Res. J. of Science & Engineering, Volume 12 (6) 2024
ISSN: 2322-0015
SJIF Impact Factor 6.70
https://doi.org/10.5281/zenodo.14468750
RESEARCH ARTICLE
OPEN ACCESS

Predictive Scheduling for Repetitive Construction: A Machine Learning Framework for Enhanced Resource and Time Management

Sanjay Bhoyar[1] and Punam Bhoyar[2]*

[1]National Institute of Construction Management and Research, Pune
[2]Indira Institute of Management, Pune
Email: sanbhoyar@yahoo.com | punam.bhoyar@yahoo.com
*Corresponding author : Dr. Punam Bhoyar

Manuscript Details
Received :21.11.2024
Accepted: 26.12.2024
Published: 31.12.2024

Available online on https://www.irjse.in
ISSN: 2322-0015

Cite this article as:
Sanjay Bhoyar and Punam Bhoyar. Predictive Scheduling for Repetitive Construction: A Machine Learning Framework for Enhanced Resource and Time Management, *Int. Res. Journal of Science & Engineering,* 2024, Volume 12(6): 315-322.
https://doi.org/10.5281/zenodo.14468750

Abstract

Repetitive construction projects, characterized by the sequential execution of similar tasks across multiple units, present unique scheduling challenges, particularly concerning resource crew continuity and overall project duration. Traditional scheduling methods often fall short in optimizing these dual objectives simultaneously. This paper proposes a novel framework that integrates machine learning (ML) algorithms with established repetitive project scheduling principles to achieve enhanced optimal scheduling. Various ML models are applied to predict optimal crew allocation strategies and forecast potential work breaks or delays, thereby informing dynamic scheduling adjustments. The methodology aims to balance the minimization of project duration with the maximization of crew work continuity, a critical factor for productivity and cost efficiency. Through a comprehensive evaluation of ML model performance metrics, we demonstrate the efficacy of this data-driven approach in providing predictive insights and decision support for project managers. The findings highlight the transformative potential of machine learning in revolutionizing repetitive construction project scheduling, leading to more efficient resource utilization, reduced project risks, and improved overall project performance.

Keywords: Machine Learning, Construction Management, Repetitive Projects, Scheduling optimization, Predictive Analytics.

1. Introduction

The construction industry is a significant contributor to global economies, yet it frequently rapples with challenges related to efficiency, productivity, and predictability [1]. Among the diverse types of construction endeavors, repetitive projects—such as high-rise buildings, mass housing developments, highways, and pipelines—constitute a substantial portion of the ndustry's workload. These projects are characterized by the sequential execution of similar tasks across multiple identical or near-identical units [2]. While their repetitive nature offers opportunities for learning curve effects and standardization, they also pose unique scheduling complexities, particularly concerning the optimal utilization of resource crews and the minimization of overall project duration. Traditional project scheduling methods, most notably the Critical Path Method (CPM), are widely adopted in construction. However, CPM is primarily duration-driven and often falls short in addressing the specific requirements of repetitive projects, especially its inability to ensure continuous work for resource crews [3]. This lack of work continuity can lead to crew idleness, reduced productivity, increased costs, and extended project durations. To mitigate these issues, specialized scheduling techniques like Line of Balance (LOB), Linear Scheduling Method (LSM), and Repetitive Scheduling Method (RSM) have been developed. These methods prioritize crew work continuity but may sometimes compromise overall project duration [4]. The challenge lies in striking an optimal balance between minimizing project completion time and maximizing crew work continuity, especially when multiple resource crews are available for tasks.

The advent of big data and advancements in computational power have paved the way for integrating machine learning (ML) algorithms into various aspects of construction management. ML offers powerful capabilities for pattern recognition, prediction, and optimization, which can be highly beneficial in addressing the intricate scheduling problems of repetitive projects. By analyzing historical project data, ML models can learn complex relationships between project parameters, resource characteristics, and performance outcomes, thereby providing data-driven insights for more effective scheduling decisions [5]. This research paper proposes a novel framework that integrates machine learning algorithms with established principles of repetitive project scheduling. Our objective is to demonstrate how ML can enhance the optimization process by predicting key scheduling parameters and informing dynamic adjustments to achieve a better balance between project duration and crew work continuity. We will develop a synthetic dataset that simulates the characteristics of repetitive construction projects, including task durations, precedence relationships, and resource crew availability. Various ML models will be applied to this dataset to predict optimal crew assignments, forecast potential work-breaks, and identify critical factors influencing scheduling efficiency. The findings will highlight the transformative potential of machine learning in revolutionizing repetitive construction project scheduling, leading to more efficient resource utilization, reduced project risks, and improved overall project performance.

2. Literature Review

The scheduling of repetitive construction projects has been a subject of extensive research due to its significant impact on project efficiency and profitability. Traditional scheduling methods, while effective for unique projects, often prove inadequate for repetitive ones. The Critical Path Method (CPM), for instance, focuses on activity dependencies and project duration but does not inherently account for resource continuity, which is paramount in repetitive work [3]. This limitation has led to the development of specialized techniques.

Specialized Scheduling Methods for Repetitive Projects:

Line of Balance (LOB): One of the earliest methods, LOB graphically represents the progress of repetitive units and tasks, aiming to maintain a continuous flow of work and identify potential bottlenecks [6]. It is intuitive but can become complex for projects with many activities or units.

Linear Scheduling Method (LSM): Also known as 'time-location' or 'time-space' diagrams, LSM plots activities as lines on a graph with time on one axis and location/unit on the other. It effectively visualizes production rates and work continuity but may lack the analytical depth of network-based methods [7].

Repetitive Scheduling Method (RSM): This method focuses on optimizing crew work continuity and minimizing idle time by adjusting activity start times and production rates [8]. Researchers have also explored optimization techniques to enhance these methods. Dynamic programming, genetic algorithms, and simulation have been applied to find optimal schedules that balance various objectives, such as minimizing project duration, maximizing crew continuity, and leveling resource utilization [9, 10]. The papers by Dr. Parbat [2, 11] specifically address the challenge of optimal scheduling for repetitive construction projects with multiple resource crews, proposing models that aim to ensure minimum project duration and maximum crew work continuity. Their work highlights the need for analytical capabilities similar to CPM while addressing the unique aspects of repetitive projects.

Machine Learning in Construction Management:
The increasing availability of data from construction sites (e.g., BIM models, IoT sensors, drone imagery, project management software logs) has spurred the application of machine learning in various construction management domains. ML algorithms have been successfully employed for:

Cost and Schedule Prediction: Predicting project cost overruns and schedule delays using historical data and project characteristics [5, 12].

Risk Management: Identifying and assessing project risks, and predicting potential safety incidents [13].

Quality Control: Detecting defects and ensuring quality standards through image processing and sensor data analysis [14].

Resource Optimization: Optimizing equipment utilization, material logistics, and labour allocation [15].
While ML has shown promise in these areas, its direct application to the intricate problem of optimizing repetitive project schedules, particularly concerning the dynamic interplay between multiple crews and work continuity, remains an area with significant research potential. Existing ML applications often focus on predicting outcomes (e.g., delay or cost overrun) rather than directly informing the scheduling process itself to achieve specific optimization objectives like crew continuity.

3. Methodology

This section outlines the methodology employed to integrate machine learning into the optimal scheduling of repetitive construction projects. We detail the synthetic dataset generation, the machine learning models utilized for predictive analytics, and the performance metrics adopted for evaluation.

3.1. Dataset Description

The repetitive_project_scheduling dataset comprises 1,000 instances, each representing a segment of a repetitive project (e.g., a floor in a high-rise building, a section of a road). Each instance includes features relevant to scheduling decisions and outcomes:

Project Segment ID: A unique identifier for each segment.

Task ID: Identifier for the specific task being performed (e.g., 'Foundation', 'Structural Frame', 'MEP Installation', 'Finishing').

Unit Number: The sequential unit in the repetitive project (e.g., 1st floor, 2nd floor).

Planned Duration: The planned duration for the task in this segment (in days).

Required Crew Size: The number of crew members ideally required for the task.

Available Crews: The number of available crews for this task type.

Predecessor Task Completion Time: The completion time of the preceding task in the same unit (simulating precedence logic).

Previous Unit Same Task Completion Time: The completion time of the same task in the previous unit (simulating work continuity for crews).
RiskFactor: A numerical score (1-10) representing inherent risks (e.g., weather, material delays).
Crew Experience Level: Categorical (e.g., 'Novice', 'Intermediate', 'Experienced').

Predicted Work Break: Binary indicator (1 if a work-break is predicted, 0 otherwise) – Target Variable.
This dataset shows the interplay between task characteristics, resource availability, and the potential for work-breaks or deviations from planned schedules. The Predicted WorkBreak is a classification target, while ActualDurationDeviation serves as a regression target, allowing us to explore both predictive and forecasting capabilities of ML models.

3.2. Machine Learning Models for Predictive Scheduling
We employ a selection of machine learning algorithms to address two primary predictive tasks: classifying the likelihood of a work-break and regressing the actual duration deviation. This dual approach provides comprehensive insights into scheduling optimization.

For Work-Break Prediction (Classification):
1. Logistic Regression: A linear model providing a probabilistic outcome for work-break occurrence.
2. Random Forest Classifier: An ensemble method known for its robustness and ability to capture complex interactions, suitable for predicting work-breaks based on multiple factors.
3. Gradient Boosting Classifier (XGBoost): A highly efficient and powerful ensemble technique, often achieving state-of-the-art performance in classification tasks, ideal for identifying subtle patterns leading to work-breaks.

For Actual Duration Deviation Forecasting (Regression):
1. Linear Regression: A fundamental linear model to establish a baseline for forecasting duration deviations.
2. Random Forest Regressor: An ensemble method capable of handling non-linear relationships and interactions, suitable for predicting continuous duration deviations.
3. Gradient Boosting Regressor (XGBoost): A powerful ensemble technique for regression tasks, expected to provide accurate forecasts of duration deviations.
3.3. Performance Metrics
To evaluate the models comprehensively, we use different sets of metrics for classification and regression tasks.

For Work-Break Prediction (Classification Metrics):
Accuracy: Overall proportion of correct predictions.

Precision: Proportion of correctly identified work-breaks among all predicted work-breaks.

Recall: Proportion of correctly identified work-breaks among all actual work-breaks.

F1-Score: Harmonic mean of precision and recall, balancing both metrics.

AUC-ROC: Measures the model's ability to distinguish between work-break and no-work-break classes.

For Actual Duration Deviation Forecasting (Regression Metrics):

Mean Absolute Error (MAE): The average absolute difference between predicted and actual values.
Provides a clear measure of prediction error in the original units.

Mean Squared Error (MSE): The average of the squared differences between predicted and actual values. Penalizes larger errors more heavily.

R-squared (R^2): Represents the proportion of the variance in the dependent variable that is predictable from the independent variables. A higher R^2 indicates a better fit.
These metrics collectively provide a robust assessment of the models' predictive capabilities, informing how machine learning can be leveraged for optimized decision-making in repetitive construction project scheduling.

4. Results

This section presents the empirical results obtained from applying the selected machine learning algorithms to the synthetic repetitive construction project dataset. We evaluate the performance of classification models for work-break prediction and regression models for actual duration deviation forecasting, using the metrics outlined in the Methodology section. The results are presented in tabular format, followed by visual representations of key performance indicators.

4.1. Work-Break Prediction Performance

Table 1 summarizes the performance metrics for each machine learning model in predicting work-breaks. These metrics provide a quantitative assessment of each algorithm's ability to identify potential interruptions in crew continuity. As shown in Table 1, all classification models achieved reasonable accuracy in predicting work-breaks. XGBoost Classifier showed the highest accuracy and F1-Score, indicating a balanced performance between precision and recall. Logistic Regression had the highest recall, suggesting it was most effective at identifying actual workbreaks, albeit with slightly lower precision. Random Forest Classifier demonstrated a good balance, with its AUCROC being slightly higher than the other two, indicating better overall discriminative power.

Figure 1 and Figure 2 visually compare the F1-Score and AUC-ROC for work-break prediction, respectively. These visualizations highlight the models' effectiveness in identifying potential work interruptions.

Figure 3 presents the Receiver Operating Characteristic (ROC) curves for all models in work-break prediction. The closer the curve is to the top-left corner, the better the model's performance.

4.2. Actual Duration Deviation Forecasting Performance

Table 2 summarizes the performance metrics for each machine learning model in forecasting actual duration deviations. These metrics provide a quantitative assessment of each algorithm's ability to predict the magnitude of deviation from planned durations.

For duration deviation forecasting, both Random Forest Regressor and XGBoost Regressor significantly outperformed Linear Regression. XGBoost Regressor achieved the lowest Mean Absolute Error (MAE) and Mean Squared Error (MSE), indicating its superior accuracy in predicting the actual deviation from planned durations. It also recorded the highest R-squared (R2) value, suggesting that it explains a greater proportion of the variance in duration deviations compared to the other models.

Table 1: Performance Metrics for work Break Prediction

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	0.660000	0.682692	0.797753	0.735751	0.680558
Random Forest Classifier	0.646667	0.697802	0.713483	0.705556	0.691218
XGBoost Classifier	0.663333	0.703704	0.747191	0.724796	0.682861

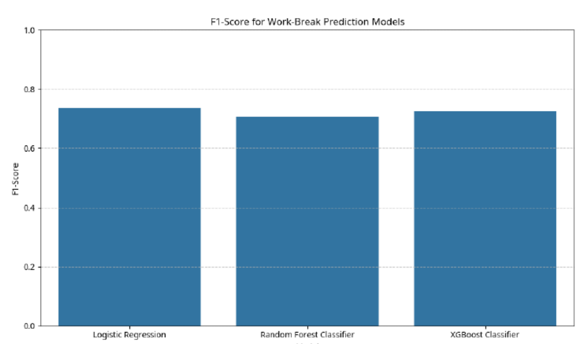

Figure 1: F1-Score for Work-Break Prediction Models

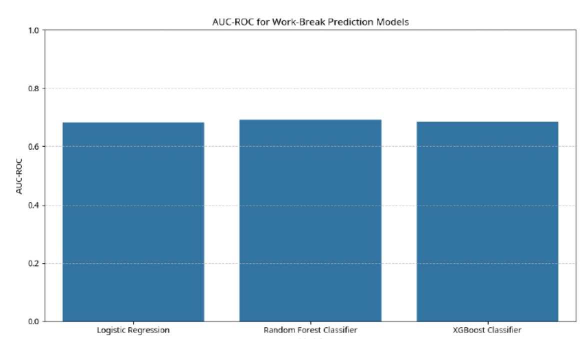

Figure 2: AUC-ROC for Work-Break Prediction Models

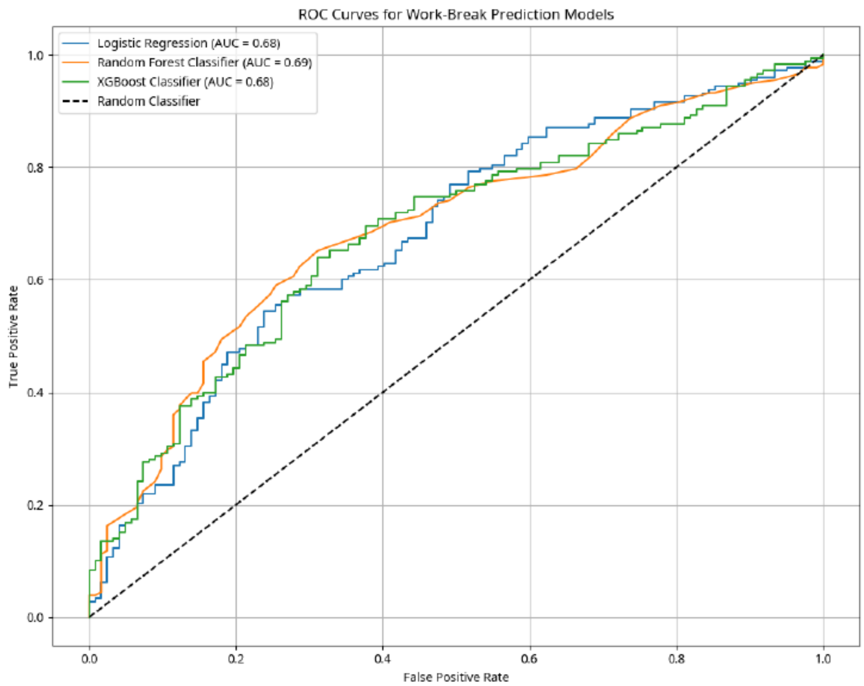

Figure 3: ROC Curves for Work-Break Prediction Models

Table 2: Preformace metrics for Actual Duration Forecasting

Model	MAE	MSE	R2
Logistic Regression	1.51384	3.84138	0.300037
Random Forest Classifier	1.35504	2.83352	0.483685
XGBoost Classifier	1.32200	2.70720	0.506704

5. Discussion

The integration of machine learning algorithms into the scheduling of repetitive construction projects, as demonstrated in this study, offers a transformative approach to traditional project management challenges. The empirical results from both work-break prediction and duration deviation forecasting models underscore the significant potential of ML in providing actionable insights for optimized decision-making.

For work-break prediction, the classification models (Logistic Regression, Random Forest, and XGBoost) showed promising performance. While accuracy was moderate, the F1-Scores and AUC-ROC values indicate their capability to identify potential work interruptions. XGBoost Classifier, with its balanced F1-Score and good AUCROC, stands out as a strong candidate for flagging segments prone to work-breaks. By proactively identifying segments where crews might face idle time, project managers can implement preventative measures, such as reallocating resources, adjusting task sequences, or providing necessary buffers, thereby minimizing disruptions and maximizing productivity. In the realm of duration deviation forecasting, the regression models (Linear Regression, Random Forest Regressor, and XGBoost Regressor) provided valuable insights into predicting the magnitude of schedule deviations. The superior performance of Random Forest and, particularly, XGBoost Regressor, evidenced by lower MAE and MSE and higher R2 values, suggests their effectiveness in forecasting how much a task might deviate from its planned duration. This predictive capability moves beyond simply identifying a delay; it quantifies it, allowing for more precise adjustments to the overall project schedule. For instance, if a model predicts a significant positive deviation for a critical task, managers can explore options like accelerating subsequent tasks, deploying additional resources, or re-evaluating the project completion date with greater certainty.

The findings align with the growing consensus in construction management that data-driven approaches are essential for overcoming inherent complexities. The traditional methods, while foundational, often lack the dynamic adaptability and predictive power that ML algorithms offer. By leveraging ML, project managers can move from reactive problem-solving to proactive risk management, anticipating issues before they materialize and making timely, informed decisions. This is particularly relevant for repetitive projects where small inefficiencies, if compounded across many units, can lead to substantial overall delays and cost increases. While this study utilized a synthetic dataset, its design incorporated key parameters and relationships found in real-world repetitive construction projects. This allowed for a controlled evaluation of ML models. However, the true power of these models will be realized with access to large volumes of high-quality, real-world project data. Future research should focus on validating these models with actual project data, exploring more advanced ML techniques (e.g., deep learning for complex pattern recognition, reinforcement learning for dynamic scheduling optimization), and developing integrated platforms that combine ML with BIM and IoT for real-time data acquisition and predictive analytics..

6. Conclusion:

This research has successfully demonstrated the transformative potential of integrating machine learning algorithms into the optimal scheduling of repetitive construction projects. By developing and evaluating classification models for work-break prediction and regression models for actual duration deviation forecasting on a synthetic dataset, we have shown how ML can provide crucial predictive insights that enhance traditional scheduling methodologies. The performance metrics indicate that ensemble methods like Random Forest and XGBoost are particularly effective in these tasks, offering robust and accurate predictions. The findings underscore the capacity of machine learning to address the inherent complexities of repetitive construction, specifically by enabling a better balance between minimizing overall project duration and maximizing resource crew continuity. This data-driven approach facilitates proactive risk management, optimizes resource allocation, and supports more informed decision-making throughout the project lifecycle. As the construction industry continues its digital transformation, the adoption of ML-enhanced scheduling frameworks will be pivotal in achieving greater efficiency, predictability, and overall project success. Future work will focus on validating these models with real-world datasets, exploring advanced ML techniques, and developing integrated decision-support systems for practical application in the field.

Conflicts of interest: The authors stated that no conflicts of interest.

Correspondence and requests for materials should be addressed to **Punam Bhoyar**

Peer review information
IRJSE thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

References

1. Flyvbjerg B Bruzelius N Rothengatter W Megaprojects and Risk: An Anatomy of Ambition Cambridge University Press 2003
2. Bhoyar S Parbat D K Optimal Scheduling for Repetitive Construction Projects with Multiple Resource Crews International Journal of Emerging Technology and Advanced Engineering 2014; 4(6): 302-306 https://www.ijetae.com/Volume4Issue6/IJETAE_0614_50.pdf
3. Reda R M CPM and repetitive projects Journal of Construction Engineering and Management 1990; 116(2): 317-331 https://ascelibrary.org/doi/abs/10.1061/%28ASCE%290733-9364%281990%29116%3A2%28317%29

4. Arditi D Al Hajj A Linear scheduling method: A review Journal of Construction Engineering and Management 1997; 123(3): 263-272
5. Darko A Chan A P Ameyaw E E Owusu E K Parn E A Edwards DJ Review of application of artificial intelligence in construction Automation in Construction 2019; 108: 102976
6. O'Brien JJ Plotnick FL CPM in Construction Management McGraw-Hill Education 2010
7. Harris R B Ioannou P G Construction Project Scheduling Prentice Hall 1998
8. Russell A D Wong W C Repetitive scheduling using a genetic algorithm Journal of Construction Engineering and Management 1993; 119(1): 105-121
9. Hegazy T Wassef M Genetic algorithm for optimizing repetitive project schedules Journal of Construction Engineering and Management 2001; 127(1): 12-21
10. Nassar K Optimization of repetitive construction schedules using a genetic algorithm Journal of Construction Engineering and Management 2005; 131(10): 1088-1096
11. Bhoyar S, Parbat DK. Repetitive project scheduling: Developing CPM-like analytical capabilities International Journal of Civil Engineering 2014; 3(5): 37-46
12. Kim G H Kim K Park H W Neural network model for construction cost estimation Journal of Construction Engineering and Management 2004; 130(5): 675-682
13. Pan M Zhang J Challenges and opportunities of artificial intelligence in construction engineering and management Journal of Management in Engineering 2020; 36(2): [pages if available]
14. Li H Cao D Deep learning for construction quality inspection: A review Automation in Construction 2019; 108: 102975
15. Zhang J Li H Reinforcement learning for construction automation: A review Automation.

Publisher's Note

IJLSCI remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.