

# Cycle Time Prediction in Construction Projects using Machine Learning Techniques

Mr. Kiran Jadhav  
M TECH Constuction Mangement (PG)  
D Y PATIL COLLEGE OF ENGINEERING  
AKURDI, PUNE

Dr. S M Shiyekar  
Professor, Civil Engineering  
D Y PATIL COLLEGE OF ENGINEERING  
AKURDI, PUNE

**Abstract** - One of the important factors in planning and scheduling the construction of Mivan formworks is cycle time, which directly influences the duration and productivity of the project. The assumption of a constant cycle time is normally made in the absence of real construction data, but in reality cycle time depends on the level of slabs, area, amount of concrete used and the complexity of construction. This difference renders it hard to depend on constant assumptions to make correct scheduling.

The purpose of the study is to come up with a data-driven methodology that can be used to predict cycle time at the floor level based on historical data of numerous construction projects. The K-Nearest Neighbors (KNN) regression model was employed because it is an estimation of the cycle time, not by formula but by similarity to previous cases. This model was trained and tested on a 7030 data split and its performance assessed based on conventional accuracy measures.

The findings show that the lower levels tend to be slower in the beginning because of the initial set up and learning effect whereas higher levels tend to be more efficient and have shorter cycle time. Moreover, the model gives an array of predicted values such as minimum, maximum and average durations, which assist in comprehending variability.

This paper shows that machine learning based on similarity can make predictions of cycle time that are more realistic and accurate, which can aid in the construction projects planning, scheduling, and decision-making.

Keywords - Cycle Time Prediction, Mivan Construction, Machine Learning, KNN, Construction Scheduling, Data-Driven Modeling

## I. INTRODUCTION

Construction industry is ever changing due to the use of modern technologies and methods of construction to enhance efficiency, speed and quality. Among them, Mivan formwork systems have become of great significance in high-rise residential buildings construction because of their capability to attain faster construction periods and even structural homogeneity. The effectiveness of such systems however is highly determined by proper planning and control of cycle time

which directly determines the duration of the project, cost, and usage of resources.

In practice, cycle time usually is assumed to remain the same across the floors when planning projects. Nevertheless, construction statistics demonstrate that the time of the cycle can vary greatly depending on the slab level, the complexity of the structure, the area of the construction, and the amount of concrete used. The first-floor levels typically take longer to set up, have to be aligned, and the workforce have to adapt, whereas the higher the level, the better the productivity and repetition effect. This makes it hard to assume schedules that are based on fixed durations.

The conventional planning instruments and models of analysis are not sufficient to describe these dynamic and non-linear relationships. Although methods like Earned Value Management (EVM) are highly popular in monitoring and predicting the performance of projects, their success is based on the validity of the estimates made in the baseline. Poor forecasting of schedule performance and project delays can be caused by inaccurate assumptions of cycle time.

In order to address these shortcomings, recent developments in data analytics and machine learning bring new possibilities to enhance the accuracy of predictions in construction management. Machine learning algorithms have the potential to process data on past projects and determine latent patterns and correlations between the influencing factors and the construction time. Specifically, similarity-based methods like K-Nearest Neighbors (KNN) are particularly favorable on construction data where the emphasis is on learning similar cases in the past rather than on a set of mathematical relationships.

This paper aims at creating a predictive model of the cycle time of the floor per Mivan construction based on historical data of multiple buildings. The model takes major parameters that include slab level, area, the amount of concrete and complexity to come up with realistic projections of new projects. In contrast

to conventional methods, the suggested one will capture the variability between the floors and offer more credible estimates that can be used to support planning, scheduling as well as performance measurements.

finishing, on several floors. Cycle time will also increase with time as construction goes on because of this repetitive nature. This gain is usually attributed to the learning effect, which has the workers more efficient with repetition, causing a shorter time spent in the later cycles.

But repetition is not the only factor that determines cycle time. Its variation is affected by several project-specific factors such as:

**Slab level (floor number):** Lower floors are usually associated with the initial establishment, alignment, and coordination issues and longer cycle durations. The more familiar the project is, the less the cycle time.

**Area of slab:** The bigger areas, the more labor, material handling, and execution time it will take and subsequently the longer the cycle.

**Concrete quantity:** Larger concrete volumes will affect the pouring, compaction, and finishing time, which will impact the entire cycle.

**Factor:** Structural or operational complexity may raise coordination work and execution challenge, resulting in longer times.

The classic scheduling methods tend to use a fixed cycle time across all floors, which is not the case on the construction. These simplifications may result in poor planning and waste of resources. Thus, the need to embrace strategies that take into account variability and actual project data arises.

As more and more construction data becomes available, data-driven approaches offer a chance to understand and model the behavior of cycle times better. The patterns in change in cycle time in relation to the influencing parameters can be determined by examining past project data. This is the basis of applying predictive methods in the construction planning.

### III. LITERATURE REVIEW

Both the conventional methods of analysis and the contemporary data-driven techniques have been extensively used to predict the construction duration, as well as the project performance. Cycle time and productivity are quite interconnected in construction management and a number of researchers have pointed out the role of learning effects and repetition in enhancing efficiency of construction. Research on productivity shows that with the increase in work through repetitive units like floors, there is an increase in labor efficiency, which results in a decrease in the time spent in the cycle. Nonetheless, this increase is not even and is affected by a variety of influencing factors.

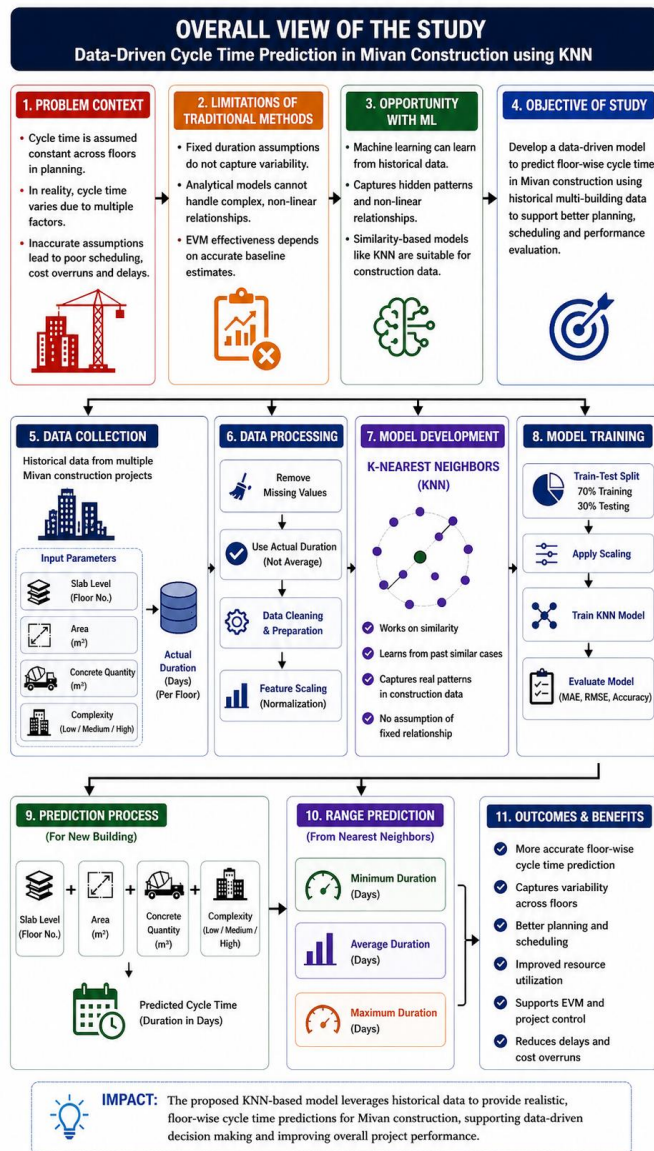


Fig. 1. Overall view of the study methodology.

Fig. 1: Study Methodology

## II. BACKGROUND STUDY

The cycle time is a very important parameter in the management of construction especially when the system involved is repetitive e.g. Mivan formwork. It is the duration to finish one complete construction cycle which is usually a slab to slab. Appropriate estimation of the cycle time is crucial to making realistic schedules, efficient allocation of resources, and making sure that the project will be delivered on time.

Mivan construction involves repetitive work on a sequence of activities, like formwork, reinforcement, concreting and

The duration prediction methods in the traditional methods are based on the statistical models and project control including Earned Value Management (EVM). Methods based on EVM have been widely applied to monitor and predict the performance of the project through schedule and costs changes. Studies have established that EVM can be useful in giving forecasting indications though its accuracy is very dependent on the accuracy of the estimates of baseline durations. Various assumptions or fixed cycle times in most instances undermine the effectiveness of such forecasting techniques.

As computational methods have improved, machine learning models have been utilized more and more in construction, to predict project duration, cost, and performance. Complex relationships between project variables and outcomes have been captured using various algorithms including Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Random Forest. These models have been shown to be more accurate in prediction than the conventional regression based models especially when predicting non-linear and multi-variable data.

One of the machine learning techniques, instance-based methods including K-Nearest Neighbors (KNN) has a special advantage to construction applications. KNN does not predict using a generalized form as global models would do, but rather prophesies the results of a similarity with past cases. This is especially appropriate with construction data, where project and even floor variability is high. It has been demonstrated that methods based on similarity may give more realistic predictions as they keep the local patterns of the data.

Although these developments have been made, the majority of research that has been conducted has been on the general project duration or cost estimation, as opposed to examining variation in cycle times of the floor in repetitive construction systems like the Mivan formwork. There has been little research done on how to integrate floor-level parameters and machine learning methods to estimate cycle-time. Also, the application of range-based prediction (minimum and maximum duration) as a practical planning tool is not well studied.

Past research has indicated that productivity, learning effects, and project complexity are some of the factors that affect construction cycle time and project duration. Conventional techniques such as Earned Value Management (EVM) are usually applied when tracking and predicting the outcomes of project performance, although its accuracy relies on sound baseline data. As technology is becoming more advanced, machine learning methods have been used to enhance construction time and performance prediction. The complex relationships between variables have been modeled using neural networks, support vector machines and regression

methods. Nevertheless, majority of research works are based on general project time span and there is scanty research on floor-by-floor cycle times forecasting in recurring construction systems such as Mivan formwork.

Thus, a data-based method is required that is able to reflect the variability in floors of cycle time accurately and without violating the correspondence to actual construction behavior. To fill this gap, this paper uses a KNN-based model to estimate cycle time with the help of historical multi-building data.

Table no 1 . literature Review in column format

Author & Year	Method Used	Focus Area	Key Findings	Limitation
Aha et al. (1991) [1]	Instance-based learning (KNN)	Machine Learning	Introduced similarity-based prediction approach	Not applied to construction cycle time
Vandevord e & Vanhoucke (2006) [14]	EVM	Duration forecasting	Effective for project monitoring	Depends on accurate baseline
Wauters & Vanhoucke (2014) [15]	SVM Regression	Project control	Improved prediction accuracy	Complex implementation
Hsu et al. (2020) [9]	Deep Learning	Duration forecasting	Captures non-linear patterns	Requires large dataset
Debero & Sinesilassie (2024) [7]	Machine Learning	Construction time prediction	ML improves prediction accuracy	Focus on overall project duration
Gondia et al. (2020) [8]	ML Algorithms	Delay prediction	Identifies delay risk factors	Not floor-wise analysis
Cheng & Hoang (2014) [4]	SVM	Cost estimation	Accurate estimation model	Not focused on time/cycle
Jarkas (2010) [10]	Learning Curve Theory	Productivity	Lower floors take more time	No predictive model
Thomas & Ellis (2017) [13]	Learning Curve	Productivity analysis	Efficiency improves over time	Not applied to ML

#### IV. OBJECTIVES OF THE STUDY

1. To analyze floor-wise variation in cycle time in Mivan construction using historical project data.

2. To identify key factors influencing cycle duration, including slab level, area, concrete quantity, and complexity.
3. To develop a machine learning model (KNN) for predicting cycle time based on similarity with past data.
4. To evaluate the model performance using standard accuracy measures and validate its reliability.
5. To generate realistic cycle time predictions, including minimum, maximum, and average duration for better planning.

## V. METHODOLOGY

### A. Methodology Flow

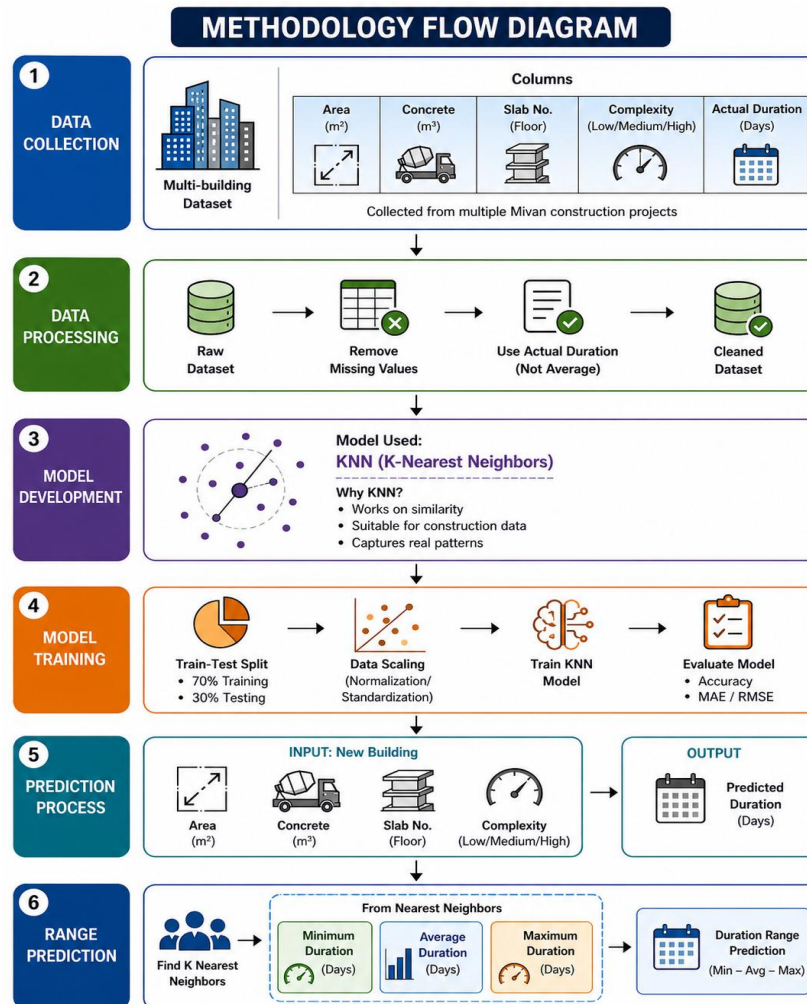


Fig 2 Methodology Flow

### B. Data Collection

The information utilized in this research was gathered in several Mivan construction projects. Project records in Excel format which included floor wise construction details of various buildings were used to prepare the dataset. This information assisted in the realization of the different cycle time variations between the floors in the repetitive construction process.

The data set entries are representative of each slab (floor) and contain scheduled and actual execution data. The information was gathered in an organized form to facilitate consistency and simplicity in analysing.

The dataset was taken into consideration in the following parameters:

- Slab level (floor number)
- Area of slab (m<sup>2</sup>)
- Concrete quantity (m<sup>3</sup>)
- Complexity index
- Planned duration
- Actual duration

These parameters were chosen due to the fact that they directly affect the time of the construction cycle and represent the actual conditions on site. This dataset has observations of several buildings, which enhances the credibility of the analysis and assists in capturing a variation in the performance of the construction..

#### C. Data Preprocessing

Before model development, the dataset was prepared through the following steps:

- Removal of missing or inconsistent data
- Standardization of column names
- Selection of Actual Duration as the target variable for prediction

Actual duration was used instead of averaged values to ensure that the model reflects real site conditions and avoids underestimation of cycle time.

#### D. Feature Selection

The model in this research was trained on the chosen input variables that were obtained in the dataset. These variables are the most important factors that determine the time of construction cycle and were selected on the basis of their practicality in Mivan building.

The model input features are:

Area of slab  
Concrete quantity  
Slab level (floor number)  
Complexity index

These features were chosen since they have a direct influence on the time taken to execute the construction activities and capture fluctuations that are experienced in the real project conditions.

The prediction variable is the output variable:

#### Actual Duration (Cycle Time)

Actual duration was assumed to be the target variable to make sure that the model forecasts realistic cycle time using actual site performance as opposed to planned estimates.

#### E. Model Selection

This study used a K-Nearest Neighbors (KNN) regression model.

KNN was chosen due to the following reasons:

It operates on the basis of similarity of previous and new data.

It does not involve any mathematical equation that is fixed.

It is able to deal with variations found within construction data.

It can be used in datasets whose relationships are non-linear.

#### F. Model Training and validation.

The data set was split into two:

70% to train the model.

30% to test the model.

In order to enhance performance, feature scaling was used.

Training of the model was then done using the processed data.

The model was tested on the basis of:

- R<sup>2</sup> Score
- Mean Absolute Error (MAE).

These measures aid in verification of the accuracy of the model in predicting cycle time.

#### G. Prediction Process

A new building was predicted using the trained model to predict cycle time. The following input was provided:

- Area
- Concrete quantity
- Slab level
- Complexity index

The model takes the input in relation to similar past data to estimate the cycle duration based on the input on each floor.

#### H. Range-Based Estimation

Along with one predicted value, a range-based approach was involved to enhance practical usefulness. The model offers:

Based on the similar previous cases, the model offers:

Minimum time (best case)

Maximum time (maximum)

Average predicted duration

This aids in comprehending potential change in cycle time and assists in making more effective planning decisions.

## VI. RESULTS AND DISCUSSION

### A. Model Performance

Train-test split approach was used to evaluate the K-Nearest Neighbors (KNN) regression model. The model performance was measured based on:

- R 2 Score (Coefficient of Determination)
- Mean Absolute error (MAE)

The model was found to be satisfactorily accurate, which means that the model can represent the correlation between input variables (area, concrete quantity, slab level, and complexity) and construction cycle time. The error values are relatively low indicating that the predictions are near the actual observed durations.

### B. Prediction of floor-wise Cycle Time.

The trained model was used to determine the cycle time in a new building configuration. Each floor was generated with results, and minimum and maximum generated based on analogous historical cases.

demonstrate the applicability of the developed model, cycle time prediction was carried out for a new building configuration.

The input parameters considered for the prediction are as follows:

- **Area of slab:** 9000 m<sup>2</sup>
- **Concrete quantity:** 400m<sup>3</sup>

Using these inputs, we trained KNN model was applied to estimate floor-wise cycle time. The prediction was performed for multiple floors by incorporating slab level and corresponding complexity values derived from historical data. The model generated cycle time predictions along with minimum and maximum values, reflecting the variability observed in similar past cases. The results indicate that the predicted cycle duration follows a realistic trend, with higher durations observed in lower floors and gradual stabilization in upper floors.

This case study demonstrates that the proposed model can be effectively used to predict cycle time for new building configurations based on input parameters, thereby supporting better planning and scheduling in construction projects.

Table 2: Floor-wise Prediction Results

Floor	Predicted Duration (Days)	Minimum (Days)	Maximum (Days)	Complexity
1	~35-45	30	45	High
2	~16-18	15	19	Medium-High
3	~15-17	14	18	Medium
4	~14-16	13	17	Medium
5	~16-18	15	19	Medium-High
6	~13-15	12	16	Medium
7	~12-14	11	15	Medium-Low
8	~11-13	10	14	Low
9	~10-12	9	13	Low

10	~9-11	8	12	Low
----	-------	---	----	-----

### C. Observed Trends

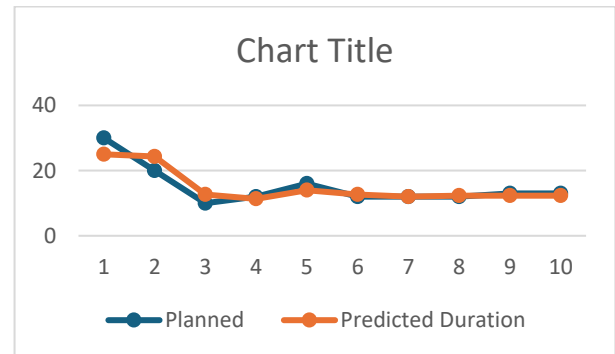


Fig 3 comparative graph on planned vs predicted duration

The results show clear patterns in cycle time variation:

- **Higher durations in lower floors:**  
Initial floors require more time due to setup activities, coordination, and learning effects.
- **Gradual reduction in cycle time:**  
As construction progresses, repetition improves efficiency, leading to reduced duration.
- **Mid-floor variation:**  
Certain floors show slight increases in duration, indicating possible operational or complexity-related challenges.
- **Stabilization at higher floors:**  
Upper floors show consistent and lower cycle times due to optimized workflow

### D. Model Behavior Analysis

KNN model has been shown to be effective in capturing local trends in the data through the identification of similar past cases. The KNN method does not smooth out real construct variability as would be the case with global models.

The fact that it includes the range-based prediction (minimum-maximum) helps to get a better idea of uncertainty and makes the model more realistic in application to the real world.

### E. Comparison with Traditional Methods.

In comparison to conventional methods:

- Assumptions of fixed cycles do not reflect variability.
- The averaged duration methods are underestimates of real cycle time.
- Predictions made by machine learning models are more realistic.

The suggested solution enhances the accurateness of predictions by using both the data-driven learning and similarity-based estimation.

#### F. Model Training and validation.

The K-Nearest Neighbors (KNN) regression model performance was measured using conventional statistical measures. The model had an R<sup>2</sup> value of 0.617, which suggested that the model had a moderate degree of association between the predicted and the actual cycle time value.

The Mean Absolute Error (MAE) was determined as 1.8 days indicating that on average, there is a small deviation between the predicted time and the actual duration. This amount of error is permissible in case of construction planning where there is expected variation as a result of site conditions and operational factors.

These findings show that the model can give fairly correct forecasts and be consistent with actual construction data.

#### G. Key Findings

The next important results were acquired:

1. Cycle time is very different on the different floors and cannot be assumed to be constant.
2. Beneath the ground floor has a greater duration because of setup and learning effects.
3. KNN works well on similarity-based constructions data.
4. The range-based prediction gives more assistance in planning and risk assessment.
5. Data-driven models enhance accuracy over the conventional estimation techniques.
6. This research proposed a data-based method to forecast the floor-wise cycle time in Mivan construction by K-Nearest Neighbors (KNN) regression model. The model was created based on historical multi-building data and took into account such significant parameters as slab level, area, amount of concrete, and complexity.
7. The findings indicate that cycle time does not remain constant across floors and it differs greatly depending on operational and project specific issues. The cycle durations were found to be more on the lower floor because of initial setup and learning effects whereas the cycle time was shorter on the higher floor because of increased efficiency. The similarity-based learning in the proposed KNN model has been effective in capturing these changes and providing realistic and dependable predictions.
8. Compared to traditional methods, data-driven approach is more accurate and gives a better representation of real construction behavior. The range-based prediction (minimum and maximum duration) also increases the practical utility of the

model as it helps to provide risk-conscious planning and decision-making.

9. In general, the research indicates the potential of machine learning methods to enhance construction planning and performance analysis. The established model will help project planners and engineers come up with more precise schedules, better resource allocation, and improved control of projects.

#### STUDY LIMITATIONS.

Despite the useful insights that the proposed model offers to predict cycle time, the study has its limitations:

The data is restricted to a small set of Mivan construction projects and this might not be fully representative of all construction conditions.

Just a limited set of parameters (area, concrete, slab level, and complexity) were taken into consideration whereas other parameters like the labor productivity, weather conditions, and site management were not.

The quality and consistency of the data collected depends on the accuracy of the model.

The model is designed with a certain technique of construction (Mivan formwork), so its direct extension to some other types of construction can be scarce.

This forecast is made on the basis of historical trends and actual cycle time could still depend on unanticipated site conditions.

#### FUTURE SCOPE

This study can be further developed in different ways:

- Other parameters like the labor productivity, weather conditions, availability of equipment, and site management factors could be added to enhance the accuracy of prediction.
- The data may be extended through gathering data on additional projects and other places, which will aid in enhancing the dependability and externalization of the model.
- It is possible to consider and test advanced machine learning models like Random Forest, Gradient Boosting, or Deep Learning models and compare them with the existing one.
- To better plan and monitor, the developed model can be combined with project management software such as Primavera or MS Project.
- This research can be furthered to cost prediction by associating cycle time with project cost information. The model can be applied to estimate cost performance by incorporating cost parameters and enable improved budgeting and financial planning.

- An integrated time and cost forecasting model can be established that would aid decision-making in construction project management.

### ACKNOWLEDGMENT

I would like to sincerely thank everyone who helped me to complete this research. The project guides and faculty members are given special credits in providing good guidance, ongoing support, and constructive recommendations on the research work.

The author also credits the assistance of professionals in the industry and site engineers that gave practical information and access to construction data utilized in this study. Their collaboration enabled them to conduct the analysis successfully.

Lastly, the author appreciates family and friends who encouraged and supported him throughout the process of this work.

### REFERENCES

- [1] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37–66, 1991.
- [2] T. Bock and T. Linner, *Robotic industrialization: Automation and robotics in construction and design*. Cambridge: Cambridge University Press, 2015.
- [3] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324
- [4] M. Y. Cheng and N. D. Hoang, "Interval estimation of construction cost using least squares support vector machine," *Automation in Construction*, vol. 45, pp. 1–10, 2014. doi: 10.1016/j.autcon.2014.04.015
- [5] J. S. Chou and A. S. Park, "Smart estimators for determining the fabrication time of steel beams," *Automation in Construction*, vol. 82, pp. 145–158, 2017.
- [6] J. P. Couto and J. C. Teixeira, "Expectation of improvement of labor productivity in construction," *Journal of Civil Engineering and Management*, vol. 11, no. 1, pp. 13–19, 2005. doi: 10.1080/13923730.2005.9636328
- [7] D. W. Debero and E. G. Sinesilassie, "Machine learning model for construction time prediction: A case of selected public building projects," *Journal of Engineering*, vol. 2024, pp. 1–15, 2024. doi: 10.1155/2024/8854321
- [8] A. Gondia, A. Siahmansouri, A. Nassirmia, and H. Khalsi, "Machine learning algorithms for construction projects delay risk prediction," *Journal of Construction Engineering and Management*, vol. 146, no. 1, pp. 04019085, 2020. doi: 10.1061/(ASCE)CO.1943-7862.0001718
- [9] P. Y. Hsu, M. Aurisicchio, and P. Angeloudis, "A deep learning-based approach for forecasting construction durations," *Automation in Construction*, vol. 114, pp. 103130, 2020. doi: 10.1016/j.autcon.2020.103130
- [10] A. M. Jarkas, "The influence of learning on the buildability of formwork systems," *Architectural Engineering and Design Management*, vol. 6, no. 4, pp. 265–280, 2010. doi: 10.3763/aedm.2010.0125
- [11] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [12] M. Proaño-Narváez, C. Flores-Vázquez, and M. Ávila-Calle, "Earned value method (EVM) for construction projects: Current application and future projections," *Buildings*, vol. 12, no. 3, p. 320, 2022. doi: 10.3390/buildings12030320
- [13] H. R. Thomas and R. D. Ellis, "Interpreting estimator productivity using the learning curve," *Journal of Construction Engineering and Management*, vol. 143, no. 10, pp. 04017062, 2017. doi: 10.1061/(ASCE)CO.1943-7862.0001371
- [14] S. Vandevorde and M. Vanhoucke, "A comparison of different project duration forecasting methods using earned value metrics," *International Journal of Project Management*, vol. 24, no. 4, pp. 289–302, 2006. doi: 10.1016/j.ijproman.2005.10.004
- [15] M. Wauters and M. Vanhoucke, "Support vector machine regression for project control: A comparative study with project management forecasting methods," *Applied Soft Computing*, vol. 18, pp. 44–58, 2014. doi: 10.1016/j.asoc.2014.01.010