A Model to Predict the Construction Schedule of High-rise Building Projects Using Minimal Project Information

A. P. Sanyal[1] and S. P. Bhattacharya[2]

[1]Assistant Professor, IIT (BHU) Varanasi, Varanasi, India
[2]Associate Professor, IIT Kharagpur, Kharagpur, India
Corresponding author's E-mail: apsanyal.apd@iitbhu.ac.in

Abstract

Construction involves the design and assembly of immovable, site-specific structures, often delivered through temporary teams. The construction industry faces ongoing pressure to provide cost-effective, sustainable solutions. Many construction projects experience significant delays due to prediction errors at the initial phases of the project, and case-based reasoning (CBR) could offer an effective method for estimating and predicting accurate construction schedule to reduce delays. While CBR has been used to forecasting construction delays, research on applying it to construction schedules is limited.
This study focuses on developing a CBR-based methodology to predict construction schedule in residential high-rise construction, using equal weighting across various attributes (a "feature counting" approach). The predicted schedules are compared to observed data to assess model accuracy and reliability. A framework for continuously capturing and applying lessons learned can help future projects better anticipate and control project schedules.
Though the overall prediction accuracy is not exceptionally high, the models demonstrate improved reliability with larger case databases. Further research is still needed to refine and address practical implementation challenges.

Keywords: schedule prediction, high-rise, case-based reasoning, AI

1. INTRODUCTION

The construction industry relies heavily on lessons learned from past projects. Identifying and predicting construction schedule early on is crucial for project success, as failure to do so can lead to issues with budgeting, execution, monitoring, and even litigation (Gondia et al., 2020). Construction timelines involve complex interactions across many project elements, with local factors like productivity, climate, and site conditions often playing a major role in determining the actual duration.
The Indian construction industry is divided into three main segments - infrastructure, real estate, and industrial. The real estate sector is a vital contributor to the industry's success, encompassing housing, retail, hospitality, and commercial sectors. Residential projects, which make up 60-70% of construction activity, typically involve high-rise buildings or complexes of varying heights.
Accurate schedule estimation is critical for assessing a construction project's viability, especially for large-scale investments (Bayram, 2017). Time deviations from the planned schedule are common in practice. Research has highlighted the relationship between construction delays and their potential causes. Predicting construction schedule accurately is a key focus of modern construction management for reducing construction delays, yet few studies have tackled this challenge Sanyal and Bhattacharya, 2023). A model that can forecast construction schedule would greatly benefit project stakeholders. The present study aims to address this need, with a specific focus on residential high-rise projects, providing construction managers with a tool to predict schedule using preliminary data.

2. LITERATURE REVIEW

2.1. Prediction models for construction

In construction projects, the insights and expertise acquired from previous similar endeavors can serve as essential guidelines for addressing current challenges. Utilizing this accumulated knowledge is a crucial strategy for evaluating progress in future projects. This approach, where historical experiences shape contemporary decisions, exemplifies inductive learning and is frequently employed to develop knowledge repositories. Within the realm of construction management, widely used predictive techniques encompass multiple regression analysis (MRA), genetic algorithms (GA), artificial neural networks (ANN), fuzzy logic methods, case-based reasoning (CBR), and various hybrids of these strategies. Nonetheless, the reliability of these models has been scrutinized by researchers, who contend that they often overlook significant qualitative factors that affect construction schedules. CBR, in particular, presents distinct advantages over alternative methods (Kim et al., 2020). It equips project managers with a framework for addressing current issues by recalling and applying knowledge from analogous past situations. CBR is particularly effective for tackling unstructured problems, as it does not necessitate the cumbersome establishment of rigid, predefined rules, unlike some other methodologies. Furthermore, CBR can be employed successfully even when certain aspects are not fully comprehended by the end-users, making it especially beneficial for less experienced practitioners. While MRA yields outcomes through statistical evaluation, its linear characteristics restrict its applicability as a universal model. In contrast, ANN is regarded as a more precise technique; however, its opaque "black box" nature complicates the understanding of the model's foundational structure. Research has shown that CBR prediction models can sustain high-quality information over extended periods and maintain robust solution-generating capabilities, even when faced with uncertain data inputs (Tayefeh, Ebadati, and Kaur, 2020).

As with other construction projects, high-rise projects can be defined by various parameters or attributes that characterize them. Previous research has identified several of these attributes that influence construction duration, with the selection of attributes often based on their availability during the preliminary stages of the project.

2.2. Case-based Reasoning (CBR) – An overview

Knowledge acquisition through the use of experiences is considered a specialized form of inductive learning. This method is commonly employed in the creation of knowledge databases. Case-based reasoning (CBR) has emerged as an efficient alternative to using past schedules, applying reasoning capabilities to derive solutions for new situations from prior experiences. CBR can be defined as a problem-solving approach that solves new cases by adapting the solutions used for similar past cases. The CBR system maintains a case-base that stores previous cases and their corresponding solutions. CBR is a rapidly growing area of research in cognitive science and artificial intelligence (Hammad et al., 2020).

The general CBR framework consists of four key stages: retrieve, reuse, revise, and retain. The retrieve stage aims to identify the case from the case-base that is most similar to the current problem. This retrieval process involves three sub-steps:

1. Determining the attribute similarity (AS),
2. Deriving the attribute weights (AW) and,
3. Calculating the case similarities (CS).

3. ANALYSIS METHODOLOGY

3.1. Data collection

A preliminary survey was conducted to identify the factors contributing to construction delays in high-rise residential projects. This survey involved nine experts from Kolkata's real estate construction industry, each possessing over 15 years of experience and holding positions of Associate Vice-President or higher. The experts were tasked with pinpointing attributes present at the initial stages of high-rise projects that significantly influence construction delays. Attributes that received less than 50% positive feedback were excluded from further analysis. Out of the 28 attributes initially identified, 14 were selected for inclusion in the main questionnaire, with 'Project duration' designated as the dependent variable. Questionnaires were then distributed to project managers overseeing 50 residential high-rise projects nearing completion in Kolkata. Data was collected based on the actual conditions observed at each project site, yielding responses from approximately 32 projects. All projects included in the study had a minimum of 11 stories, with the tallest building exceeding 30 floors.

3.2. Case Base Formulation and Validation

The case base is composed of distinct cases, each assigned a unique case number and encompassing specific problem features along with their corresponding solutions. Each case is characterized by two primary elements: the problem description, which delineates the characteristics of the problem, and the solution description, which elaborates on the resolutions provided. The cases are represented through various attributes, with their respective values referred to as attribute values (AVs). The cases are labeled from ‘Pr01’ to ‘Pr32’, and they are further classified according to the number of floors, which includes four distinct categories: 'Flr 1-15', 'Flr 16-20', 'Flr 21-25', and 'Flr >26' for those exceeding 25 floors. To assess the accuracy of predictions made by different methods, approximately 10% of the cases from each category were chosen for validation purposes. In total, the case base comprises 27 cases, including 5 designated for validation. A summary of these details is presented in Table 1.

Table 1 Details of case bases

Maximum Number of Floors in the Project	Number of Projects	Validation Cases
Flr 1-15	16	2
Flr 16-20	6	1
Flr 21-25	5	1
Flr >26	5	1
Total	**32**	**5**

Researchers have employed various methods to determine attribute weights (AW) for similarity-based models, including Feature Counting (FC), importance weights from Genetic Algorithms (GA), Artificial Neural Networks (ANN) and multivariate analysis. However, there is no consensus on which weight assignment method is superior. The weights of features reflect their influence on similarity calculations. Features with greater influence should be assigned larger weights. The Feature Counting method does not eliminate any attributes.

When specific information is lacking, it is commonly assumed that no attribute is more important than another. The Feature Counting method applies a weight of 1 to each input attribute, implying equal importance.

The evaluation of the accuracy of prediction models has historically posed significant challenges; however, cross-validation provides a nearly unbiased estimate of prediction accuracy applicable to any methodology. This technique is closely associated with the bootstrap method for estimating error rates and is extensively utilized to evaluate the performance of algorithms in both classification and regression tasks. Its primary advantage is the ability to utilize the entire dataset for both training and testing purposes, facilitating the repeated calculation of error measures. To cross-validate the predictive accuracy of various models, the original dataset is partitioned into 20 subsets, with 15% of the cases (5 cases) designated as validation cases and the remaining 85% (27 cases) serving as the base. By employing different combinations of validation and base cases, 10 distinct datasets are generated for the purpose of cross-validating the model.

4. ANALYSIS AND DISCUSSION

With the attribute weights considered using feature counting method, the construction schedule is predicted for each set. The predicted schedules are further analysed for the accuracy of the method used. Figure 1 shows the comparison of comparison between the predicted and observed values for Set 7.

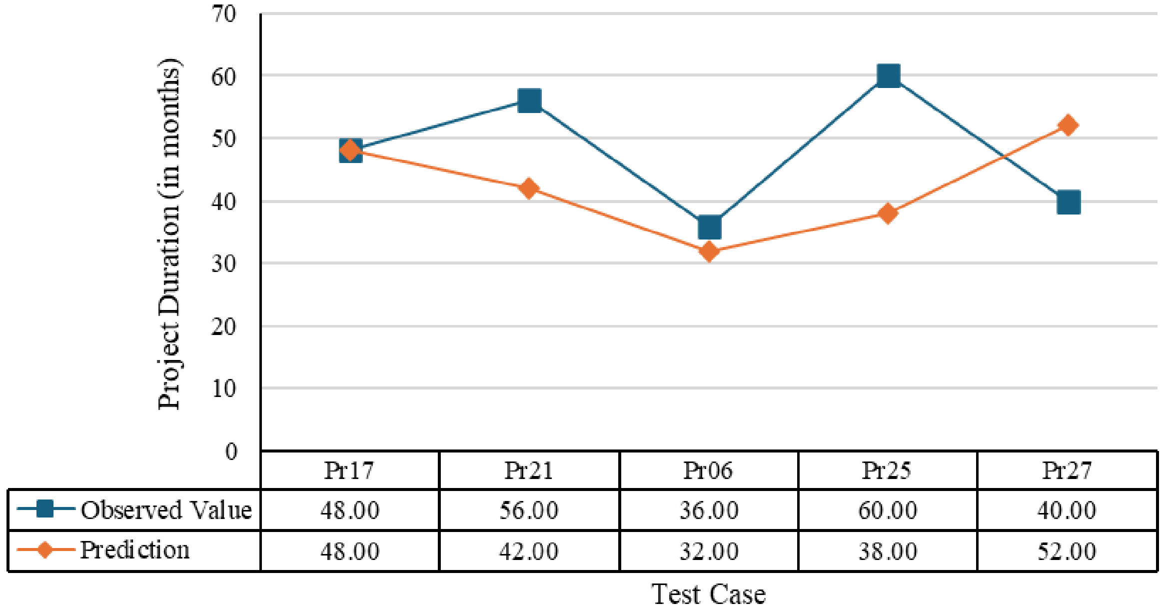

Figure 1 Comparison of Observed and predicted value for Proposed duration in Set 7

Table 2 lists the prediction accuracy for each of the four level categories, namely, 'Flr 1-15', Flr 16-20', 'Flr 21-25', and 'Flr >26'. The average accuracy has been calculated and displayed in Table 2. Results show the average accuracy lies over 70% for all the four level categories.

Table 2 Prediction accuracy achieved using CBR_FC for all 20 sets of data

CBR-FC	Flr 1-15	Flr 16-20	Flr 21-25	Flr 26-30
Set 1	90%	89%	65%	82%
Set 2	42%	69%	63%	70%
Set 3	97%	85%	70%	93%
Set 4	71%	51%	92%	94%
Set 5	49%	77%	60%	53%
Set 6	88%	76%	65%	82%
Set 7	88%	89%	63%	70%
Set 8	93%	69%	70%	93%
Set 9	97%	85%	92%	94%
Set 10	70%	51%	60%	93%
Set 11	61%	77%	65%	82%

CBR-FC	Flr 1-15	Flr 16-20	Flr 21-25	Flr 26-30
Set 12	100%	83%	63%	70%
Set 13	54%	89%	70%	93%
Set 14	100%	47%	92%	94%
Set 15	100%	85%	60%	53%
Set 16	79%	69%	65%	82%
Set 17	76%	53%	63%	95%
Set 18	47%	83%	77%	85%
Set 19	74%	26%	92%	94%
Set 20	84%	92%	60%	53%
Average Accuracy	**78%**	**72%**	**70%**	**81%**

The prediction accuracy for category *'Flr 1-15'* is about for 78% for all sets combined. Similar observations can be seen for *'Flr 26-30'*. However, lower accuracy has been observed for *'Flr 16-20' and 'Flr 21-25'*. Figure 2 depicts the prediction accuracy achieved using CBR_FC for all 20 sets of data.

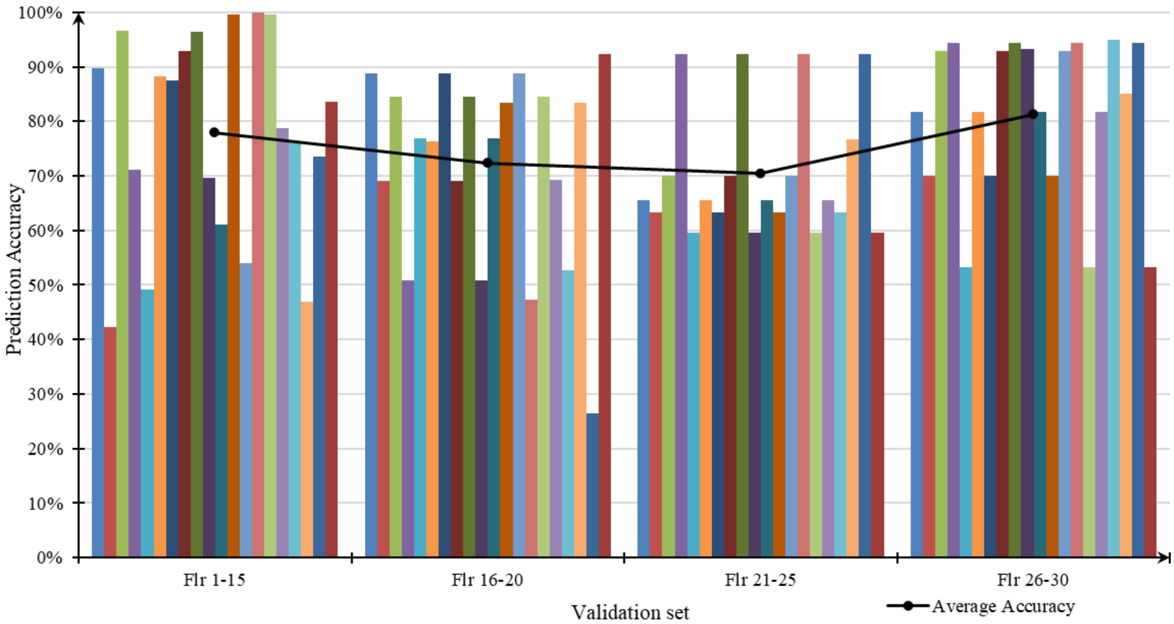

Figure 2 Graph depicting prediction accuracy of achieved using CBR_FC for all 20 sets of data

5. CONCLUSION

In the construction industry, accurately predicting project performance from the outset plays a critical role in project success. Many construction projects in India currently face delays ranging from months to years. As such, finding an effective strategy for forecasting construction schedule is of paramount importance. (Lagos et al., 2024)

This research develops a construction schedule prediction model using a case-based reasoning (CBR) approach, focusing on residential high-rise buildings in Kolkata, a major metropolitan city in eastern India. The CBR model incorporates weights generated through multiple regression analysis (MRA) and artificial neural networks (ANN). The selection of the most similar case in the CBR algorithm can vary based on factors such as attribute selection, weighting methods, and similarity measures – all of which depend on the researcher's approach.

The researchers developed the case-based reasoning (CBR) model using 32 case studies, with 5 cases used for validation. The CBR model assigned equal weights to various attributes, and the prediction accuracy was cross validated. The results confirm the high potential of CBR methods, which

demonstrated over 70% accuracy in predicting construction schedule. Such predictive models provide construction managers with valuable insights by leveraging readily available project data.
To further enhance these predictive capabilities, the researchers recommend expanding the case database with additional construction information. While directly utilizing project data can be tedious, developing a user-friendly interface that draws from a library of similar past projects could facilitate practical implementation. This interface should be regularly updated to incorporate new and more detailed project data (Tariq and Gardezi, 2023).
Overall, the study highlights the value of data-driven approaches like CBR in forecasting construction delays, offering construction firms a strategic advantage in project planning and execution.

6. REFERENCES

Abu Hammad, A. A., Alhaj Ali, S. M., Sweis, G. J., & Bashir, A. (2008). Prediction model for construction cost and duration in Jordan. *Jordan Journal of Civil Engineering*, *2*(3), 250–266.

Alemu, S. K. (2022). Construction time prediction model for public building projects. *Engineering, Construction and Architectural Management*, *29*(5), 2183–2206. https://doi.org/10.1108/ECAM-11-2020-0975

Bayram, S. (2017). Duration prediction models for construction projects: In terms of cost or physical characteristics? *KSCE Journal of Civil Engineering*, *21*(6), 2049–2060. https://doi.org/10.1007/s12205-016-0691-2

Gondia, A., Asce, S. M., Siam, A., El-dakhakhni, W., Asce, F., & Nassar, A. H. (2020). Machine Learning Algorithms for Construction Projects Delay Risk Prediction. *Journal of Construction Engi- Neering and Management*, *146*(1), 1–16. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001736

Kim, Y. J., Yeom, D. J., & Kim, Y. S. (2019). Development of construction duration prediction model for project planning phase of mixed-use buildings. *Journal of Asian Architecture and Building Engineering*, *18*(6), 586–598. https://doi.org/10.1080/13467581.2019.1696207

Lagos, C. I., Herrera, R. F., mac Cawley, A. F., & Alarcón, L. F. (2024). Predicting construction schedule performance with last planner system and machine learning. *Automation in Construction*, *167*. https://doi.org/10.1016/j.autcon.2024.105716

Sanyal, A. P., & Bhattacharya, S. P. (2023). A comparative analysis between CBR based prediction models and MRA models for high-rise construction delay prediction. *International Journal of Construction Management*, 1–13. https://doi.org/10.1080/15623599.2023.2211461

Tariq, J., & Gardezi, S. S. S. (2023). Study the delays and conflicts for construction projects and their mutual relationship : A review. *Ain Shams Engineering Journal*, *14*(1), 101815. https://doi.org/10.1016/j.asej.2022.101815

Tayefeh Hashemi, S., Ebadati, O. M., & Kaur, H. (2020). Cost estimation and prediction in construction projects: a systematic review on machine learning techniques. In *SN Applied Sciences* (Vol. 2, Issue 10). Springer Nature. https://doi.org/10.1007/s42452-020-03497-1