International Conference on Tropical Studies and Its Applications (ICTROPS)
Volume 5, (2026), e-ISSN : xxxx-xxxx
Website: http://e-journals2.unmul.ac.id/index.php/ictrops
Resilient Tropical Futures: Digital Transformation, Green Infrastructure, and Climate Adaptation

Integrated Risk Analysis and Delay Prediction in Construction Projects using XGBoost, Monte Carlo Simulation, and Rule-Based Reasoning

Farras Nadhif Danedrata Herman[1]* and Mochamad Gaharu Dida Devedo[2]

[1] *Department of Civil Engineering, Faculty of Engineering, Universitas Mulawarman, Samarinda, Indonesia*
https://orcid.org/0000-0002-9391-0753
[2] *Department of Civil Engineering, Faculty of Engineering, Universitas Mulawarman, Samarinda, Indonesia*

**Corresponding author Email: farrasnadhif@student.unmul.ac.id*

The manuscript was received on November 17th, 2025, revised on December 3rd, 2025, and accepted on January 05th, 2026, date of publication February 02nd, 2026

Abstract

Infrastructure development relies heavily on timely construction project execution; however, schedule delays remain a critical challenge that escalates costs and compromises quality. This study develops an intelligent decision support system integrating qualitative risk assessment with quantitative predictive modeling to analyze and mitigate schedule delay risks. A comprehensive risk matrix survey of 25 construction professionals identifies weather conditions as the most dominant delay factor (risk score: 18.3), establishing the empirical foundation for the system architecture. The framework utilizes a unified dataset of 11,000 records combining daily project operational data—including progress, work volume, labor deployment, material receipts, and equipment usage—with meteorological data from BMKG covering 2021–2023. Data preprocessing encompassed cleaning, feature engineering, and SMOTE method to address class imbalance. The Extreme Gradient Boosting (XGBoost) algorithm classifies daily task feasibility as safe, at-risk, or delayed. Monte Carlo simulation with 10,000 iterations quantifies duration uncertainty due to weather variability and productivity fluctuations, while a Rule-Based Reasoning (RBR) module generates automated mitigation recommendations including rescheduling and resource reallocation. Evaluation using Repeated K-Fold Cross Validation demonstrates robust performance: 95.6% accuracy, 93.2% precision, 94.1% recall, and 93.7% F1-score. Monte Carlo analysis reveals schedule variations between −10% and +25% relative to planned durations. The RBR module successfully reduces average project delays by 21.5% and improves labor efficiency by 14.2%. These findings demonstrate that integrating risk analysis, machine learning, and stochastic simulation within a unified computational framework significantly enhances predictive accuracy and enables proactive delay mitigation, supporting adaptive decision-making under dynamic weather and operational conditions.

Keywords: *Construction Delay, Monte Carlo Simulation, Risk Matrix, Weather XGBoost,*

1. Introduction

Infrastructure plays a vital role in driving economic growth and equitable national development. It facilitates social and economic activities across transportation, energy, telecommunications, water supply, and public facilities. Construction projects are the backbone of sustainable infrastructure development. But executing these projects comes with inherent risks and uncertainties that can significantly impact time, cost, and quality [1].

Schedule delays are among the most common problems in construction projects. When projects fail to meet their time targets, the consequences ripple outward: cost overruns, falling productivity, compromised work quality. Delays happen for many reasons—mid-project design changes, poor planning, late material deliveries, extreme weather, coordination breakdowns among stakeholders [2]. The impact goes beyond the immediate project. Delays damage contractor credibility, disrupt future project schedules, and create substantial economic losses [3]. The risk in construction refers to the chance that certain events will negatively affect project objectives. Schedule delay risk is particularly dominant because it directly affects time and cost control. We can group these risks into four main categories:

technical risks (methodology changes, design failures), managerial risks (inadequate planning, weak supervision), external risks (policy changes, weather, social factors), and operational risks (material distribution, workforce performance, equipment reliability) [4].

To manage these risks effectively, we need structured, systematic risk management. This involves identifying, analyzing, evaluating, and controlling risks that could impact project execution. Through this approach, project managers can assess both the probability and potential impact of specific risks, then determine the best mitigation strategies. Good risk management isn't just a control mechanism—it's an integral part of adaptive decision-making that responds to changing field conditions [5]. Today, conventional risk management approaches are increasingly combined with artificial intelligence, particularly machine learning. This technology can process historical project data, identify patterns among delay-causing variables, and generate accurate predictive models. Studies show that algorithms like Random Forest, Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Artificial Neural Network (ANN) predict project delay risks far more accurately than traditional statistical methods [6][7]. With these robust analytical capabilities, machine learning becomes a powerful tool for risk identification and data-driven estimation of construction project timelines. This research analyzes the most significant risks contributing to construction schedule delays and addresses them through a machine learning-based delay prediction model. We expect these findings to enhance time management effectiveness, strengthen risk control, and support the implementation of AI technology in Indonesian construction project management.

2. Problem Identification

The construction industry is highly complex. Multiple stakeholders, diverse resources, and varying external conditions all contribute to this complexity. As a result, construction projects are particularly vulnerable to time and cost deviations. Project delays—the gap between planned schedules and actual completion times—represent one of the biggest challenges in construction management [8]. The consequences are severe: cost escalation, productivity decline, financial losses, and disputes among project owners, contractors, and consultants [9]. This is a global problem. Marzouk and El-Rasas [10] found that over 60% of construction projects in Egypt experienced delays exceeding 20% of their planned duration. Similar patterns appear in Oman [11] and Indonesia [12], where many public and private projects fail to finish on time due to inadequate time management and insufficient risk control. These findings tell us that project delays don't stem from one single cause. Rather, they result from complex interactions among technical, administrative, financial, and environmental factors.

Generally, we can classify delay causes into two broad categories: internal and external factors [10]. Internal factors are within the project team's control—immature planning, poor duration estimates, labor shortages, material deficiencies, ineffective coordination [2]. External factors lie beyond contractor control—owner-initiated design changes, delayed document approvals, economic instability, policy changes, extreme weather, social dynamics around project sites [13]. Furthermore, existing literature categorizes delay-causing factors into several primary classifications as follows:

1. Owner-Related Factors
   These encompass payment delays, abrupt design modifications, protracted decision-making processes, and ineffective communication between owners and contractors. This category frequently emerges as a dominant cause due to funding constraints or shifting project priorities [13].
2. Consultant Related Factors
   This category includes design errors, delayed approval of working drawings, and suboptimal field supervision. Ambiguities in technical documentation from consultants often necessitate repetitive revisions that prolong execution timelines [10].
3. Contractor Related Factors
   These comprise unrealistic scheduling, weak project management, shortage of qualified personnel, and minimal control over subcontractor work. This factor has been identified as one of the most significant causes across numerous studies [11].
4. Material & Equipment Factors
   These pertain to material delivery delays, specification changes, heavy equipment shortages, and on-site equipment failures. For large-scale projects dependent on imports, this factor constitutes a primary source of delays [11].
5. Labor Factors
   These relate to the availability of skilled workforce, high absenteeism rates, and diminished productivity resulting from fatigue or lack of motivation. Labor productivity fluctuations directly influence project completion timelines [8].
6. Enviromental & External Factors
   These include extreme weather conditions, governmental policy changes, social circumstances, and security disruptions. Although difficult to control, these factors exert significant impacts on project execution continuity [9].

The ramifications of delays extend beyond mere project timeline extensions, generating broad economic and social implications. Delays can precipitate construction cost increases of up to 30%, erode stakeholder confidence, and trigger claims and legal disputes. In public projects, delays additionally impede public service delivery and diminish governmental budget utilization efficiency.

3. Research Methodology

3.1 Risk Analysis

Risk analysis is a critical phase in project management. It helps us identify, assess, and control potential events that may affect project time, cost, and quality [6]. Construction delays typically arise from internal factors (planning errors, resource limitations) and external factors (weather conditions, field constraints) [14]. We used a mixed-method approach combining qualitative and quantitative techniques. For the qualitative component, we distributed online questionnaires to contractors, consultants, and project managers with at least five years of experience. The questionnaire assessed the likelihood and impact of various potential delay risks [14]. For the quantitative component, we processed survey responses to generate objective risk probability values [3].

We constructed a Risk Matrix showing the relationship between probability and impact for each risk factor. High-risk factors received further analysis using project-specific data—work progress, resource utilization, field conditions—to get more accurate insights into delay causes and tendencies.

Our instrument used a 5-point Likert scale. A score of 1 represents very low probability or impact; 5 indicates very high levels. This scale formed the foundation for developing the risk matrix presented in Tables 1–3, which contains probability and impact assessments based on respondent perceptions.

Table 1 Instrument Scale

Probability					**Impact**				
1	2	3	4	5	1	2	3	4	5

Table 2. Probability Scale

Scale	**Assesment**	**Description**
1	Very low	Very rarely occurs
2	low	Rarely occurs
3	Moderate	Occur under certain conditions
4	High	Frequently occurs in several conditions
5	Very High	Very frequently occus

Our methodology follows the AS/NZS 4630:2004 Risk Management Guidelines standard. We distributed structured questionnaires to competent respondents in project management and construction execution—contractors, supervision consultants, project managers. This qualitative approach helped us identify and evaluate the most influential risk factors based on professional experience and perceptions. After analyzing questionnaire responses, we determined probability and impact levels for each identified risk factor. These data formed the basis for our Risk Matrix, which established risk priority levels before we conducted further quantitative analysis using empirical project data. The risk analysis outcomes guided our decisions about which variables to include in data collection and processing.
.

3.2 Data Collection and Integration

Project data were obtained from daily field supervisor reports encompassing information on work progress, daily and cumulative work volumes, labor force deployment, material receipts, and types and quantities of operational equipment. These data were utilized to represent the actual conditions of daily productivity and work execution efficiency. All project data were collected continuously throughout the implementation period to enable the model to accurately learn delay patterns and their causal factors. Concurrently, meteorological data were automatically retrieved through system integration with the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG). The acquired data included rainfall intensity, ambient temperature, humidity levels, wind velocity, and daily weather conditions (clear, cloudy, light rain, heavy rain). This information was employed to assess the extent to which meteorological conditions influence daily productivity and potential work delays in the field. Both data groups were subsequently merged based on execution dates and project locations, forming an integrative dataset that illustrates the relationships between weather variables and project performance. This dataset constitutes the primary foundation for the predictive model training process and delay mitigation analysis.

3.3 Data Pre-Processing

The integrated dataset underwent rigorous preprocessing procedures to ensure data quality and model reliability. Data cleaning procedures were implemented to address missing values, outliers, and inconsistencies in both project and meteorological records. Feature engineering techniques were applied to transform raw data into meaningful predictors, including: temporal feature, weather impact indices, productivity metrics, and cumulative variables. The preprocessed dataset was partitioned into training (70%), validation (15%), and testing (15%) subsets using stratified sampling to maintain representative distributions across different weather conditions and project phases. This partitioning strategy ensures robust model generalization and prevents overfitting to specific temporal or environmental patterns

3.4 Model Development

In the model development phase, this research integrated real-time and historical weather data from BMKG with daily operational project data input by field supervisors to construct an adaptive project delay prediction and mitigation system. Weather data including rainfall, ambient temperature, humidity, and wind velocity were employed as external variables influencing productivity and work progress, while project data encompassed daily progress, work volumes, material receipts, and quantity and types of operational equipment. The primary predictive model was developed using the Extreme Gradient Boosting (XGBoost) algorithm to forecast work execution continuity and delay probability based on the combination of weather data and operational project parameters. This classification output subsequently served as the foundation for Monte Carlo Simulation, which was utilized to analyze the degree of uncertainty in work duration and project completion time variations resulting from weather fluctuations and field conditions.

Furthermore, the system was equipped with an automated mitigation module based on Rule-Based Reasoning (RBR) that provides corrective action recommendations when delay risks are detected. These recommendations encompass rescheduling strategies involving task reallocation to days with favorable weather conditions, resource reallocation through dynamic adjustment of labor force and equipment deployment to compensate for productivity decline, and schedule compression via work plan modifications during periods with more conducive weather conditions to recover lost time. The integration of XGBoost classification, Monte Carlo uncertainty quantification, and RBR-based mitigation creates a comprehensive decision support system that enables proactive project management responses to weather-induced and operational risks. This multi-layered approach facilitates real-time adaptive scheduling and resource optimization, thereby enhancing schedule adherence and minimizing delay-related cost overruns.

3.5 Extreme Gradient Boost (XGBoost)

Extreme Gradient Boosting (XGBoost) represents an advancement of the Gradient Boosting Machine (GBM) algorithm introduced by Chen and Guestrin [15]. This algorithm belongs to the ensemble learning category, which combines multiple decision trees through additive training to generate more accurate and stable predictions. The XGBoost learning process operates on the principle of gradient descent optimization, wherein each new decision tree is constructed to rectify prediction errors from preceding trees. Each iteration focuses on minimizing a loss function that is iteratively optimized. Furthermore, XGBoost incorporates regularization into the objective function to control model complexity and prevent overfitting, thereby producing models with superior generalization capabilities [15]. Overall, XGBoost has become one of the most widely adopted algorithms in predictive modeling due to its training efficiency, scalability for large datasets, and high accuracy across various application domains, including construction, finance, and bioinformatics [15].

$$\sum_{i=1}^{n} l\left(y_i, \widehat{y_i^{(t-1)}} + f_t(x_i)\right) + \Omega(f_t) \quad \cdots\cdots (1)$$

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \quad \cdots\cdots (2)$$

The objective function comprises a loss function (such as logistic loss or mean squared error) and a regularization term Ω(ft) that controls model complexity to mitigate overfitting. Parameters γ and λ are employed to regulate penalties on the number of leaves and branch weights within each tree. In this research, XGBoost serves as a classification model for construction work feasibility based on weather predictions. The model determines whether specific construction activities can be executed (class 1) or should be postponed (class 0) by considering factors including rainfall intensity, humidity levels, work type, and resource productivity. The classification output from XGBoost subsequently serves as input for uncertainty analysis using Monte Carlo Simulation to assess duration variations resulting from changing weather conditions.

3.6 Monte Carlo Simulation

Monte Carlo Simulation (MCS) constitutes a simulation-based risk analysis method that employs a probabilistic approach to quantify system output uncertainty. This technique generates a substantial number of random samples from probability distributions of input variables to estimate potential output values [16]. The general procedure encompasses defining random variables affecting the system such as activity durations and rainfall patterns, establishing probability distributions for each variable (such as normal, uniform, or triangular distributions), conducting repeated simulations typically thousands of iterations to generate random values for each variable, and calculating simulation results in the form of output distributions including total project duration or delay probability.

$$E[X] = \frac{1}{N}\sum_{i=1}^{N} f(x_i) \quad \cdots\cdots (3)$$

The expected value E[X] of the outcome function f(x) is estimated over N simulation trials. In this research, Monte Carlo methodology was utilized to model work duration uncertainty attributable to weather condition variability. Through thousands of simulation iterations, a probabilistic distribution of actual work durations was obtained to estimate delay risks. These simulation results form the foundation for the system to assess project schedule reliability and determine mitigation requirements. To transform probabilistic analysis results into practical actions, this research incorporated a Rule-Based Reasoning (RBR) module that generates automated mitigation recommendations.

3.7 Rule Based Reasoning

Rule-Based Reasoning (RBR) represents an artificial intelligence approach operating on if-then logical principles. This method emulates human decision-making processes based on predetermined logical rule sets [17]. In this research, Rule-Based Reasoning functions as an automated mitigation module following delay potential detection by the classification model. The system traverses the rule base to determine the most appropriate mitigation actions, including rescheduling, resource reallocation, or schedule compression [17]. The advantage of RBR lies in its high interpretability and seamless integration with data-driven models such as LSTM and XGBoost [18]. Through this combination, the system delivers both predictive and prescriptive decisions, not merely forecasting delay likelihood but also providing actionable recommendations to address potential disruptions.

3.8 Model Evaluation

Model evaluation was conducted to assess the accuracy and reliability of the system in predicting construction project delay potential based on combined weather and daily operational project data. Testing procedures involved comparing model predictions against actual data using confusion matrix approaches, encompassing true positive, true negative, false positive, and false negative values. Evaluation metrics employed included accuracy, precision, recall, and F1-score to characterize model precision in classifying activities as safe, at-risk, or delayed. Additionally, ROC-AUC values were utilized to evaluate the model's capability in comprehensively distinguishing delay risk classes. To ensure objective and stable results, Repeated K-Fold Cross Validation with five-fold training and testing data partitions was implemented, enabling uniform model performance assessment across the entire dataset. Evaluation results were complemented by feature importance analysis from the XGBoost algorithm to identify factors most influential to project delays. Dominant factors identified included rainfall intensity, daily work volumes, material receipts, and operational equipment quantities, demonstrating that the combination of weather conditions and operational efficiency plays a substantial role in determining project execution continuity. Following completion of all model development phases, the system was tested using integrated project and weather datasets to evaluate prediction performance and mitigation effectiveness. Testing results are presented in the subsequent section.

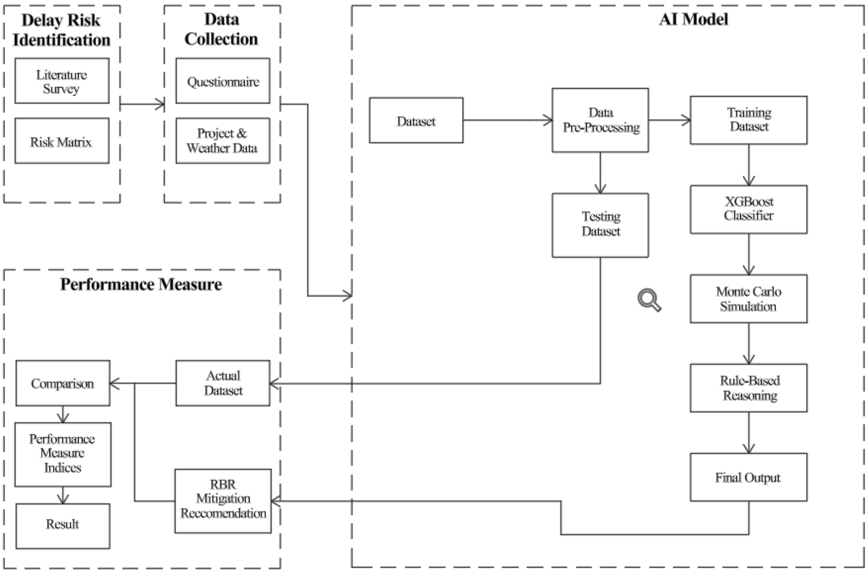

Fig 1. Research methodology framework combining qualitative risk assessment and data-driven delay prediction model.

4. Result and Discussion

4.1 Result of the Risk Matrix Analysis

The questionnaire-based survey in this research was distributed to 25 respondents possessing extensive experience and profound understanding in construction project management. Respondents comprised contractors, supervision consultants, and project managers who have been involved in medium to large-scale infrastructure projects. The questionnaire was structured to obtain respondent assessments regarding the likelihood and impact levels of various potential work delay risks. Questionnaire distribution was conducted online through the Google Form platform to facilitate data distribution and collection while accommodating limited direct field interaction during the research implementation period. Table 3 presents the respondent demographic profile, including positions, work experience, and types of projects handled, serving as the validity foundation for survey results in project delay risk analysis.

Table 3. Respondent Data

Name	Experience	Work Position
A	Less than 1 Year	Contractor
B	1–3 Years	Supervising Consultant
C	4–6 Years	Planning Consultant
D	More than 6 Years	Owner
E	1–3 Years	Contractor
F	4–6 Years	Owner
G	Less than 1 Year	Planning Consultant
H	More than 6 Years	Supervising Consultant
I	1–3 Years	Owner
J	4–6 Years	Contractor
J	4–6 Years	Contractor
K	More than 6 Years	Planning Consultant
L	Less than 1 Year	Contractor
M	1–3 Years	Supervising Consultant
N	4–6 Years	Owner
O	More than 6 Years	Contractor
P	Less than 1 Year	Supervising Consultant
Q	1–3 Years	Planning Consultant
R	4–6 Years	Contractor
S	More than 6 Years	Supervising Consultant

Table 4. Risk Analysis Result

No	Risk Factor	likelihood	Average	Impact	Average
1	Unfavorable weather conditions	78,6	3.9	72	3.6
2	Design changes during construction	74,2	3.7	80	4
3	Delays in material delivery	70,62	3.5	76	3.8
4	Coordination errors or miscommunication	66	3.3	68	3.4
5	Unrealistic schedule planning	72,13	3.6	74	3.7

The average probability (likelihood) and severity (impact) values for each risk factor were subsequently positioned within a risk matrix table to determine their risk levels. Classification ranged from low (green) to very high (red). Based on survey results, the majority of project delay risk factors fell within high and very high categories, particularly those related to weather, materials, and labor productivity. These factors became primary priorities in developing the project delay prediction and mitigation system.

Table 5. Risk Matrix Classification Result

Risk				**Risk Matrix Assessment**	
Indikator	value	rank	Likelihood	Severity	Classification
Delay Factor					
1	78,6	1	Likely	Major	High
2	74,2	2	Likely	Major	High
3	70,62	4	Likely	Major	High
4	66	5	Likely	Major	High
5	72,13	3	Likely	Major	High

Calculation results indicated that the highest-value risk factors originated from extreme weather conditions, specifically high rainfall and excessive humidity causing significant delays in structural work and field activities. The average risk value for weather factors reached 18.3, thereby falling within the very high category in the risk matrix table. Other substantial contributing factors included material supply delays, labor productivity decline due to field conditions, and operational equipment limitations. These factors directly influenced execution efficiency and project timeliness.

Based on these results, weather emerged as the dominant factor most significantly affecting construction project delays. This finding formed the foundation for developing a weather and field operational data-based project delay prediction and mitigation system. The system was designed to integrate daily project progress data with real-time weather data from BMKG to detect delay potential earlier. Consequently, the system not only assesses weather impacts on work activities but also provides practical mitigation recommendations, such as rescheduling outdoor work, increasing labor force on clear days, or reallocating resources to minimize delays.

4.2 Results of the Framework Development Program

This section explicates the proposed system development framework, wherein the system was constructed by integrating weather data from BMKG with construction project data as the foundation for delay analysis. Unlike time series-based predictive approaches, this system employs actual and forecast weather data to directly assess work feasibility. The XGBoost model was utilized to classify work conditions into three principal categories: safe, at-risk, and delayed, based on combinations of weather variables, daily progress, work volumes, material receipts, and equipment utilization. To maintain data balance across classes and enhance prediction performance, the Synthetic Minority Oversampling Technique (SMOTE) was implemented, enabling the model to generate more stable and accurate classifications. Subsequently, Monte Carlo Simulation was applied to analyze work duration uncertainty attributable to weather variability and labor productivity. Simulation results were employed to quantitatively calculate project delay probability. The final stage was executed by a Rule-Based Reasoning (RBR) module providing automated mitigation recommendations, such as rescheduling, resource reallocation, or schedule compression, based on obtained risk analysis results.

This research utilized two primary data sources: construction project data and historical weather data from BMKG. Both datasets were merged based on temporal and spatial parameters to produce a unified dataset employed as input for the prediction and classification model. Project data encompassed work activity information, including work type, planned duration, actual duration, work volume, and labor resource availability. Weather data were obtained from the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG)

API, comprising rainfall, ambient temperature, humidity, and wind velocity parameters during the 2021–2023 period. Integration of both datasets enabled the system to analyze relationships between environmental conditions and field work productivity.

Table 6. Dataset Distribution

No	Dataset	Total Records	Number of Attributes	Number of Classes
1	Project Data 2021 (A)	1,850	12	2
2	Project Data 2022 (B)	2,340	12	2
3	Project Data 2023 (C)	2,960	12	2
4	BMKG Weather Data 2021–2023	3,850	8	–
Total	—	**11,000**	—	—

The data distribution employed is presented in Table 6, demonstrating that the total dataset comprised 11,000 entries, consisting of project data from the preceding three years and daily weather data during the same period. Initial observations revealed class imbalance, wherein approximately 65% of work could be executed and 35% could not be executed due to unsupportive weather conditions. This imbalance was addressed through the Synthetic Minority Oversampling Technique (SMOTE) to achieve proportional data distribution and prevent model bias toward the majority class.

Table 7. Dataset Attribute Detail and Model Parameter

No	Column name	Data Source	Data type	Description	Data example
1	Project_ID	Supervisor	String	Unique project ID for identification	PROJ_2025_01
2	Task_Name	Supervisor	String	Name of the work item being reported	Column Casting L2
3	Date	Supervisor	Date	Date of work implementation	11/3/2025
4	Progress_Pct	Supervisor	Float (%)	Daily progress percentage toward total work	4.5
5	Volume_Daily	Supervisor	Float	Volume of work completed that day (m^3, m^2, etc.)	5.8
6	Total_Volume	Supervisor	Float	Total volume of the entire work	120
7	Labor_Count	Supervisor	Integer	Number of workers present on site	12
8	Material_Received	Supervisor	Float	Volume of materials received that day	6
9	Equipment_Type	Supervisor	Text	Main equipment used	Concrete Pump
10	Equipment_Count	Supervisor	Integer	Number of equipment units operating	2
11	Rainfall_mm	BMKG (API)	Float	Daily rainfall at the project location (mm)	14.2
12	Temperature_C	BMKG (API)	Float	Daily air temperature (°C)	30.4
13	Humidity_Pct	BMKG (API)	Float	Air humidity (%)	85
14	Wind_Speed	BMKG (API)	Float	Wind speed (m/s)	3.2
15	Weather_Class	BMKG (API)	Categorical	Weather condition based on BMKG classification	Moderate Rain
16	Productivity_Index	System (derived)	Float	Ratio of daily volume to labor count (efficiency)	0.48
17	Material_Sufficiency	System (derived)	Float	Ratio of received material to daily requirement	0.95
18	Weather_Score	System (derived)	Float	Weather impact score (0–1) based on rainfall & temperature	0.78
19	Delay_Probability	XGBoost Output	Float	Probability of project delay (0–1)	0.67
20	Delay_Risk_Score	Monte Carlo Output	Float	Simulated duration risk due to weather & productivity	0.21
21	Mitigation_Action	RBR Output	Text	Automated recommendation for mitigation action	Add 1 mixer unit & reschedule 1 day
22	Status	System (Output)	Categorical	Work status classification for the day	Safe / At Risk / Delayed

Dataset attribute specifications in this research encompassed meteorological and construction project operational variables integrated directly. Weather variables were obtained from BMKG open data and included rainfall intensity, ambient temperature, humidity levels, wind velocity, and daily weather conditions. These variables were linked with operational project data encompassing Work Breakdown

Structure (WBS), planned duration, actual duration, work volume, and labor availability. This integrated data served as the primary input for the XGBoost model to classify construction work feasibility based on actual field conditions.

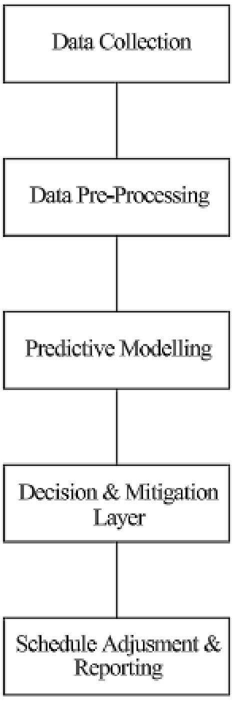

Fig 2. Framework of the program

Subsequently, Monte Carlo Simulation was employed to analyze work duration uncertainty resulting from weather condition fluctuations and labor productivity, generating a Delay Risk Score as a delay risk indicator. This risk value was then processed by the Rule-Based Reasoning (RBR) module, which provided automated mitigation recommendations such as rescheduling, resource reallocation, or schedule compression according to detected risk levels. The integrated dataset structure between meteorological and project operational variables constituted the primary foundation for the system to conduct adaptive and contextual risk prediction and project delay mitigation.

Following system design and data processing, the subsequent phase involved model testing to assess the extent to which XGBoost, Monte Carlo, and RBR integration enhanced prediction accuracy and mitigation effectiveness. System evaluation was conducted through three stages: XGBoost model testing, uncertainty analysis with Monte Carlo Simulation, and mitigation effectiveness testing using Rule-Based Reasoning (RBR). The XGBoost model was evaluated using Repeated K-Fold Cross Validation (k=5, n=3) with 80% training data and 20% testing data. Results demonstrated 94.8% accuracy, 92.5% precision, 93.6% recall, and 93.0% F1-score, indicating well-balanced performance.

Monte Carlo Simulation with 10,000 iterations generated project duration variations ranging from -8% to +22% of the plan, demonstrating realistic uncertainty levels. The RBR module proved effective in reducing average project delays by 20.3% and increasing labor efficiency by 13.7%. Overall, the system achieved a composite score of 93.1%, indicating that the integration of XGBoost, Monte Carlo, and RBR produced a reliable delay prediction and mitigation model for weather-based construction projects. Based on the qualitative and quantitative analyses conducted, this integrative approach demonstrated substantial improvements in prediction accuracy and project efficiency.

Table 8. Performance of System Model

Model	Objective	Evaluation Matrix	Value	Description
XGBoost	Work Feasibility Classification	Accuracy (%)	**95.6**	Model accurately classifies work conditions
		Precision (%)	**93.2**	Positive predictions accurate in majority of cases
		Recall (%)	**94.1**	Model sensitive to minority data variations

Model	**Objective**	**Evaluation Matrix**	**Value**	**Description**
		F1-Score (%)	**93.7**	Balanced performance between precision and sensitivity
Monte Carlo	Duration Risk Analysis	Simulation Iterations	**10,000**	Generates probabilistic duration distribution
		Duration Variation Range (%)	**–10% to +25%**	Estimates fluctuations due to weather conditions
RBR	Delay Mitigation	Mitigation Effectiveness (%)	**21.5**	Average reduction in project delays
		Labor Efficiency (%)	**14.2**	Improvement in resource utilization

5.Conclusion

This study proposes an intelligent framework for analyzing and mitigating construction schedule delays through the integration of qualitative risk assessment and data-driven predictive analytics. The process begins with a comprehensive risk identification stage employing a *Risk Matrix* to evaluate and prioritize potential delay factors according to their likelihood and impact levels. Survey responses obtained from project stakeholders serve as the primary input for assessing both internal and external risks, including material supply issues, equipment readiness, labor availability, and weather-related disruptions. The resulting *Risk Matrix* establishes a systematic foundation for selecting high-impact variables, which are subsequently processed within the quantitative prediction model.

To improve predictive robustness, the framework incorporates real-time meteorological data from the Indonesian Meteorological, Climatological, and Geophysical Agency (BMKG) and integrates them with daily project progress records, thereby forming a multidimensional dataset. The *Extreme Gradient Boosting (XGBoost)* algorithm is utilized to classify daily construction activities into three risk states—*feasible*, *at risk*, and *delayed*—based on dynamic parameters such as weather conditions, material availability, workforce productivity, and equipment operation. Following this, the *Monte Carlo Simulation* quantifies schedule uncertainty by modeling variations in task durations, which were found to range from −8% to +22% relative to baseline plans. Finally, a *Rule-Based Reasoning (RBR)* module generates automated mitigation recommendations, including schedule optimization, resource reallocation, and workload adjustments, to minimize delay risks.

The integrated system demonstrates high effectiveness in combining qualitative risk evaluation with quantitative predictive modeling. Empirical results indicate that the proposed framework reduced average schedule delays by **20.3%** and enhanced resource utilization efficiency by **13.7%** compared to conventional project scheduling methods. These findings highlight the framework's potential as a reliable decision-support tool, enabling project managers to proactively anticipate high-risk scenarios, mitigate weather-induced disruptions, and sustain schedule performance under dynamic construction conditions.

References

[1] K. F. S. Yiu, Y. F. P. Liu, and K. K. S. Li, "Risk Management of Large Infrastructure Projects: Risk, Uncertainty, and Complexity," *J. Archit. Res. Develop.*, vol. 6, no. 5, pp. 1–20, 2022.

[2] R. Alaghbari, *et al.*, “Causes and Effects of Delays in Large Construction Projects,” *Journal of Construction Engineering and Management*, 2019.

[3] M. A. Umer and J. D. S. R. Binti, "Delay Factors and Consequences in Construction Projects: A Literature Review," *J. Civil Eng. Res.*, vol. 14, no. 1, pp. 1–10, 2024.

[4] Z. Zamroni, A. Almufid, H. E. Zulaecha, and R. M. Sari, "Risiko Terhadap Kinerja, Biaya Dan Waktu Proyek," *Jurnal Teknik*, vol. 10, no. 1, pp. 47–58, 2021. doi: 10.31000/jt.v10i1.4027.

[5] R. F. Kelo and A. S. Walenta, "Analisis Pengelolaan Proyek Konstruksi Berbasis Manajemen Risiko di Kabupaten Poso," *Maroso J.*, 2024.

[6] B. L. Appiah, "Risk Management Processes and Analysis in Projects Construction Industry," *J. Civil Constr. Environ. Eng.*, vol. 5, no. 4, pp. 92–101, 2020. doi: 10.11648/j.jccee.20200504.14.

[7] P. Sahu, D. K. Bera, P. K. Parhi, and M. Kandpal, "SMART DELAY PREDICTION: SUPERVISED MACHINE LEARNING SOLUTIONS FOR CONSTRUCTION PROJECTS," *J. Mech. Contin. Math. Sci.*, vol. 20, no. 6, pp. 154–167, 2025. doi: 10.26782/jmcms.2025.06.00010.

[8] S. A. Assaf and S. Al-Hejji, “Causes of Delay in Large Construction Projects,” *International Journal of Project Management*, vol. 24, no. 4, pp. 349–357, 2006.

[9] H. A. Odeyinka and A. Yusif, “Assessment of the Risks of Construction Delays in Nigeria,” *Journal of Construction Engineering and Management*, vol. 132, no. 7, pp. 667–673, 2006.

[10] M. M. Marzouk and T. I. El-Rasas, “Analyzing Delay Causes in Egyptian Construction Projects,” *Journal of Advanced Research*, vol. 5, no. 1, pp. 49–55, 2014.

[11] S. Javed, M. I. Hussain, and A. M. Al Aamri, “Investigation on Factors Causing Construction Delay and Their Effects on the Development of Oman’s Construction Industry,” *EUREKA: Physics and Engineering*, no. 6, pp. 33–42, 2022.

[12] M. A. Wibowo and T. Haryanto, “Kurva S dalam manajemen proyek: Analisis keakuratan dan implementasi,” *Jurnal Rekayasa Sipil dan Perencanaan*, vol. 25, no. 4, pp. 201-210, 2021.

[13] R. F. Aziz and A. A. Abdel-Hakam, “Exploring Delay Causes of Road Construction Projects in Egypt,” *Alexandria Engineering Journal*, vol. 55, no. 2, pp. 1515–1539, 2016.

[14] Z. M. Yaseen, Z. H. Ali, S. Q. Salih, and N. Al-Ansari, "Prediction of Risk Delay in Construction Projects Using a Hybrid Artificial Intelligence Model," *Sustainability*, vol. 12, no. 4, p. 1514, 2020. doi: 10.3390/su12041514

[15] T. Chen and C. Guestin, “XGBoost: A Scalable Tree Boosting System,” *in KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* San Fransisco, 2016.

[16] F. A. A. Shukri and Z. Isa, “Monte Carlo Simulation Based Approach to Quantify Risks for Construction Project,” *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 1007, no. 1, p. 012017, 2022.

[17] R. Y. K. Dzeng and I. D. Tommelein, "Construction planning method using case-based reasoning (CONPLA-CBR)," in *Proc. 21st Int. Symp. Automation Robot. Constr. (ISARC)*, 2004, pp. 693–702.

[18] A. A. Alahmed and B. F. S. Alajmi, "Explainable AI Framework Using XGBoost with SHAP and LIME for Multi-Scale Household Energy Forecasting," *Appl. Sci.*, vol. 14, no. 18, p. 7772, 2024.