Research on Engineering Structures & Materials
www.jresm.org
MIM Research Group

Research Article

Profitability-aware machine learning for forecasting fabrication durations in PEB projects

Ringle Raja S [*,1,a], G Hemalatha [1,b], Vincent Sam Jebadurai S [1,c], S. Elizabeth Amudhini Stephen [2,d], Athish N S [3,e]

[1]*Department of Civil Engineering, Karunya Institute of Technology and Sciences, India*
[2]*Department of Mathematics, Karunya Institute of Technology and Sciences, India*
[3]*Department of AI and ML, Karunya Institute of Technology and Sciences, India*

Article Info

Article History:

Received 19 June 2025

Accepted 19 Dec 2025

Keywords:

Fabrication duration forecasting; Pre-engineered buildings; Machine learning; Stage-wise prediction; Random forest regression

Abstract

In Pre-Engineered Building (PEB) production, forecasting production timelines accurately is crucial, especially as tasks like cutting, welding, and painting vary depending on structural complexity and job size. Many conventional scheduling methods fall short when dealing with the dynamic, multi-stage nature of fabrication workflows. This study introduces a profitability-aware machine learning–based approach that predicts both overall and stage level fabrication durations using Random Forest regression. The model is trained on part-level production data, with features including project tonnage, number of parts, and statistical descriptors of task durations. To improve learning performance, projects are grouped into six distinct complexity classes based on fabrication characteristics. A profitability-oriented evaluation is also proposed, which labels prediction outcomes as Profitable, Acceptable, or Excess depending on how closely they align with business targets. The model is tested on real data from 34 completed PEB projects, showing clear improvements over conventional estimation methods. Results demonstrate that stage-level predictions outperform start-date-based forecasts, ensuring profitability-aligned outcomes even in high-complexity projects. The proposed framework helps bridge the gap between technical forecasting accuracy and real-world production goals, offering a practical solution for smarter planning in steel fabrication environments.

1. Introduction

The construction industry continues to adopt industrialized production systems such as Pre-Engineered Buildings (PEBs) to meet rising demands for speed, efficiency, and cost control. Unlike conventional structures, PEBs rely heavily on steel-based components fabricated off-site. These components pass through sequential, tightly coupled stages including cutting, welding, drilling, and surface treatment. While this modularization accelerates installation, it introduces new challenges. Stage-wise fabrication durations are difficult to predict due to component variability, production concurrency, and task interdependencies.

Accurate forecasting of fabrication timelines is vital to ensure just-in-time project delivery, optimize shop-floor operations, and maintain profitability. Traditional estimation methods are often based on historical averages or parametric rules. These methods lack the granularity and adaptability failing to capture the dynamic fabrication environments. In response, researchers have increasingly turned to machine learning to address this complexity [1]. For instance, A demand

*Corresponding author: ringleraja@karunya.edu.in
[a]orcid.org/0009-0002-0304-6605; [b]orcid.org/0000-0001-7067-3786; [c]orcid.org/0000-0002-9820-8103; [d]orcid.org/0000-0002-3794-5214; [e]orcid.org/0009-0002-5648-8984
DOI: http://dx.doi.org/10.17515/resm2026-979ml0619rs

forecasting model for steel manufacturing was introduced using ML techniques [2]. Stage-wise prediction in fabrication was shown to improve accuracy by isolating each task [3]. More recently, an ML-driven approach for material quantity estimation in structural fabrication was proposed [4]. Both highlighted the potential of data-driven models to outperform rule-based systems under uncertainty [3,4].

Beyond material and demand forecasting, several studies have explored ML applications in activity-level duration prediction. Beyond material and demand forecasting, several studies have explored ML applications in activity-level duration prediction. Ensemble-based methods such as Random Forest have been applied to capture non-linear dependencies in production workflows [5,6,7].

Despite these advances, limitations remain. First, they treat fabrication as a monolithic process, forecasting total durations without capturing variability across individual stages. Second, they often fail to incorporate structural and project-level complexity factors which can significantly impact process performance. Recent efforts to overcome these limitations include the use of classification-aware preprocessing [8,9], feature engineering using statistical descriptors [10], and profitability-aligned evaluation frameworks [11,12]. These approaches shape the methodology of the present study.

This research proposes a machine learning based predictive framework that forecasts both overall and stage-specific fabrication durations in PEB. Using Random Forest regression and statistical features such as mean, median, and mode from part-level fabrication records, the model is trained on data from 34 real-world projects. Projects are stratified into six mutually exclusive classes (P1C1–P2C3) based on their structural complexity and production scale, improving model specificity and interpretability. Forecast outputs are further evaluated using a profitability-aware classification system that categorizes predictions into "Profitable," "Acceptable," or "Excess" based on deviation thresholds tied to business performance. The key contributions of this study are as follows:

- Profitability-aware forecasting framework: We propose a machine learning framework tailored to Pre-Engineered Building (PEB) fabrication, which integrates profitability considerations into predictive modeling — an aspect rarely addressed in prior work.
- Activity-level prediction strategy: Unlike conventional project-level or start-date-based forecasting, this study introduces and validates an activity-specific modeling approach that reduces error propagation and enhances stability across fabrication stages.
- Empirical validation across complexity tiers: Using a dataset of 34 projects stratified into low, medium, and high complexity classes, the framework is empirically tested, demonstrating robust accuracy (low RMSE, narrow confidence intervals) and superior profitability classification compared to baseline models.
- Decision-making support in industrial environments: The framework provides actionable insights for PEB managers by linking predictive accuracy to profitability outcomes, enabling informed scheduling, resource allocation, and financial planning.

Together, these contributions advance both the academic understanding of machine learning in construction manufacturing and the practical capabilities of PEB firms to achieve profitability-aware project control.

2. Literature Review

2.1. Machine Learning Applications in Construction and Fabrication Forecasting

Machine learning (ML) has emerged as a transformative approach in construction and industrial fabrication, offering data-driven alternatives to conventional scheduling and forecasting techniques [13]. Early studies in this domain developed ML-based productivity models for construction crews, demonstrating adaptability under site variability [14]. Computational intelligence was also applied to inventory management, where improvements in material flow were achieved within construction supply chains [15].

In steel and modular fabrication, demand forecasting has become a crucial use case for ML. A tailored model was designed to forecast demand in steel production environments [2]. Regression-based learning was later applied to estimate material quantities in structural fabrication tasks [4]. Neural network approaches were also validated for raw material inventory control in steel plants, confirming their usefulness for shop-floor planning [16].

Recent studies in materials and construction domains further underscore the versatility of machine learning in predictive tasks. Advanced neural architectures and optimization techniques have been successfully employed beyond scheduling and fabrication. For instance, an SRS-optimized long short-term memory (LSTME) model achieved high accuracy in predicting the mechanical properties of friction stir processed Al-6061–Alumina composites, with R^2 values exceeding 0.90 across multiple material performance indicators [13]. Similarly, an artificial neural network (ANN)-based soft-computing model was developed to forecast the compressive strength of eco-friendly recycled aggregate concrete, demonstrating strong agreement between experimental and predicted outcomes [1]. These contributions highlight the broader applicability of ML in handling complex, nonlinear behaviors across engineering domains, thereby reinforcing the rationale for its adoption in forecasting multi-stage fabrication durations in PEB projects.

2.2. Activity-Level Forecasting and Multi-Stage Modeling

Traditional duration prediction models often struggle to adapt to the variability present in complex fabrication workflows. To address this, recent studies have proposed breaking projects into discrete stages or tasks. A predictive model was developed that independently forecasted fabrication flow by stage, improving overall forecast fidelity [3]. ML techniques were also applied to estimate production costs at the activity level, showing that disaggregation helps isolate uncertainty [17].

Data-driven timeline forecasting in modular housing was explored through ML applications that estimated process-specific durations across prefabricated units [18]. Multi-objective ML models for scheduling optimization were also introduced, reinforcing the effectiveness of activity-aware modeling [19]. A comprehensive review of hybrid predictive scheduling models in industrial construction further emphasized the value of integrating multiple algorithms across project stages [20]. Ensemble-based models for activity-level forecasting have shown superior accuracy in dynamic workflows. Support vector machines were applied for scheduling in engineering-procurement-construction (EPC) projects [21].

2.3 Feature Engineering and Model Architecture in Predictive Scheduling

Feature engineering plays a pivotal role in the success of ML-based forecasting models. Classification-aware preprocessing was incorporated to improve generalization in scheduling models [8]. Hybrid models that combined statistical features such as mean, median, and mode with learning algorithms were shown to enhance prediction accuracy [22]. The integration of IoT-based sensor data for real-time forecasting in fabrication environments was also explored, marking a shift toward dynamic model updating [23].

The comparison of ML architectures has been a subject of interest in the construction domain. Decision trees, support vector machines, and Random Forest models were evaluated for forecasting applications. Random Forest, among ensemble models, has proven effective in activity-level forecasting by capturing non-linear interactions and handling heterogeneous features, making it suitable for fabrication duration prediction [24,6,7]. Statistical descriptor–based feature sets, similar to those employed in the present study, were also tested and found to improve duration prediction accuracy [10].

2.4 Complexity-Aware Stratification and Classification-Driven Modeling

Given the heterogeneity in project types and part geometries in fabrication workflows, classification-aware approaches have gained traction. Component-type clustering was applied to isolate variance in project complexity, which improved learning performance [9]. A task-classification model was implemented that adjusted its prediction pipeline based on structural and geometric load [22]. These strategies directly inform the use of class-based segmentation in your work (e.g., P1C1 to P2C3).

Profitability-aware forecasting framework further advanced this concept by classifying predictions into zones based on expected value-added versus deviation risk [11]. This aligns with your study's implementation of three forecast-based profitability classes Profitable, Acceptable, and Excess which convert RMSE accuracy into financial insight.

2.5 Profitability-Driven Forecasting

Several studies have emphasized that forecasting precision must translate into operational or financial gains. Forecast deviation was linked to profitability erosion, with recommendations that error tolerance thresholds be embedded into scheduling systems [25]. Cost-sensitive ML models were developed to penalize over- or underestimation differently, introducing a concept later integrated into classification-based evaluation frameworks [12]. Random Forest models were also tested in high-variability scheduling environments, demonstrating that profitability-aware learning produced more practical predictions [26].

2.6 Practical ML Integration

Industrial implementation studies demonstrated the practical integration of ML in fabrication environments. Real-time dashboards were developed for monitoring fabrication progress [27]. ML was also embedded into production feedback loops to support adaptive decision-making [28]. Cost and lead-time improvements in a PEB production facility were further documented after the adoption of stage-level ML predictions, affirming the direct applicability of such methods in real-world steel fabrication environments [29].

3. Data Acquisition and Preparation

3.1 Data Source and Collection

The foundation of this study was based on real-time fabrication data collected from a Pre-Engineered Building (PEB) production facility. The structured data pipeline was designed to preserve the fidelity of production workflows while ensuring compatibility with machine learning (ML) algorithms. The primary data source was a centralized Google Sheets interface referred to as the PRED Sheet (Production Raw Entry Database). This live document was managed by the factory manager and is continuously updated to reflect real-time progress. It contains both static and dynamic fields. Static entries, such as job code (JC), member part mark, quantity (Q), and weight in Kgs (TW), are recorded once upon project initiation. Dynamic entries, including date-wise progress for each activity stage, are updated daily for every part mark involved in the manufacturing process.

Table 1. PRED Sheet – Production Raw Entry Database

JC	Part	Q	TW	C	S	BS	SW	......	FP
.	.	.	.	.	.	.	.		.
.	.	.	.	.	.	.	.	......	.
.	.	.	.	.	.	.	.		.
.	.	.	.	.	.	.	.		.
18	Rafter/Portal	1	255.41	17/2/22	17/2/22	20/2/22	22/2/22	......	10/5/22
18	Rafter/Portal	1	219.37	17/2/22	17/2/22	20/2/22	21/2/22	......	10/5/22
18	Rafter/Portal	1	309.63	20/2/22	20/2/22	21/2/22	22/2/22	......	10/5/22
18	Rafter/Portal	1	309.63	20/2/22	20/2/22	21/2/22	23/2/22	......	10/5/22
19	Plates	1	0.3	24/1/22	NA	NA	NA	......	27/1/22
19	Plates	1	0.3	24/1/22	NA	NA	NA	......	27/1/22
19	Plates	1	0.3	24/1/22	NA	NA	NA	......	27/1/22
19	Plates	1	0.3	24/1/22	NA	NA	NA	......	27/1/22
.	.	.	.	.	.	.	.		.
.	.	.	.	.	.	.	.	......	.
.	.	.	.	.	.	.	.		.
.	.	.	.	.	.	.	.		.

Each fabricated component undergoes ten sequential production stages: Cutting (C), Splicing (S), Built-up Section (BS), SAW Welding (SW), Fit-up (FW), Welding (W), Drilling (D), Cleaning & Grinding (CG), Primer Application (P), and Finish Painting (FP), followed by Dispatch. Table 1 presents a sample excerpt from the PRED Sheet.

3.2 Feature Engineering

To support the modeling pipeline, the PRED Sheet was supplemented with calculated fields using spreadsheet automation. These derived metrics fall into two categories: (i) Start Date-based durations, which record the number of working days taken to complete each task from the official project start date (SD), and (ii) individual activity durations (AD), which isolate the duration of each production stage from its own start to end date. Table 2 shows a representative sample of these calculated fields.

Table 2. Automated activity duration metrics from PRED sheet

JC	Part	C		S		BS		SW		FU		W		....	FP	
		SD	AD	SD	AD	SD	AD	SD	AD	SD	AD	SD	AD	....	SD	AD
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
18	Rafter/Portal	1	1	1	1	4	4	6	3	9	4	10	2	....	83	5
18	Rafter/Portal	1	1	1	1	4	4	6	3	10	5	10	1	....	83	5
18	Rafter/Portal	1	1	1	1	4	4	5	2	14	10	16	3	....	83	5
18	Rafter/Portal	1	1	1	1	4	4	5	2	14	10	16	3	....	83	5
18	Rafter/Portal	1	1	1	1	2	2	3	2	7	5	13	7	....	80	5
19	Plates	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	....	4	1
19	Plates	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	....	4	1
19	Plates	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	....	4	1
19	Plates	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	....	4	1
19	Plates	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	....	4	1
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

These engineered features, along with the raw material attributes, were preprocessed for machine learning input. Missing values in activity durations occurred primarily when certain fabrication stages were not applicable (e.g., splicing absent for plate-type members). Categorical variables such as part marks were encoded using label encoding, and missing values were imputed using the median duration for each respective activity within the same project.

3.3 Project Complexity Stratification

Statistical summaries were generated automatically from the PRED Sheet to reveal trends in fabrication timelines. These summaries were pivotal for understanding the variability and central tendency of production activities across projects. Specifically, for each project and each fabrication activity, descriptive statistics including minimum, maximum, mean, median, and mode were calculated. This statistical analysis enabled the identification of outliers, bottlenecks, and typical process durations—insights that are critical for enhancing model accuracy and interpretability.

Two key types of aggregations were performed: the first computed statistics relative to the Start Date (SD) to assess cumulative project timelines, and the second isolated durations for individual activities, independent of their relationship to project commencement. The former allows understanding of the overall project flow, while the latter isolates efficiency metrics for each task.

Tables 3 and 4 provide these aggregated statistics. Table 3 summarizes the project-level fabrication durations from SD, offering insight into end-to-end timeline distributions. Table 4 breaks down the statistical behavior of each fabrication activity across projects, illuminating activity-specific variabilities.

Table 3. Project-Wise Summary of Fabrication Durations (From Start Date)

Code	Total Fabrication Parts	Cutting					Splicing					
		Min	Max	Mean	Median	Mode	Min	Max	Mean	Median	Mode	....
.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	....
.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	
19	50	1	1	1	1	1	0	0	0	0	0	....
21	18	1	1	1	1	1	2	2	2	2	2	....
22	399	1	1	1	1	1	1	2	1	1	1	....
24	550	1	1	1	1	1	1	2	1	1	1	....
27	24	1	1	1	1	1	1	1	1	1	1	....
29	288	1	1	1	1	1	1	2	1	1	1	....
30	339	1	1	1	1	1	1	1	1	1	1	....
31	901	1	1	1	1	1	1	2	1	1	1	....
.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	....
.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	

Table 4. Project-Wise Activity Duration Summary (Independent of SD)

Code	Total Fabrication Parts	Cutting					Splicing					
		Min	Max	Mean	Median	Mode	Min	Max	Mean	Median	Mode	....
.	.	.	.	.	.	.	.	.	.	.	.	....
.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	
19	50	1	1	1	1	1	0	0	0	0	0	....
21	18	1	1	1	1	1	2	2	2	2	2	....
22	399	1	1	1	1	1	1	2	1	1	1	....
24	550	1	1	1	1	1	1	2	1	1	1	....
27	24	1	1	1	1	1	1	1	1	1	1	....
29	288	1	1	1	1	1	1	2	1	1	1	....
30	339	1	1	1	1	1	1	1	1	1	1	....
31	901	1	1	1	1	1	1	2	1	1	1	....
.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	....
.	.	.	.	.	.	.	.	.	.	.	.	
.	.	.	.	.	.	.	.	.	.	.	.	

This level of statistical analysis formed the basis for selecting input features and tuning model performance in subsequent stages. Moreover, it highlights how specific activities vary in duration, guiding the identification of process stages with the greatest predictive uncertainty.

3.4 Classification Based Preprocessing of Data

To ensure that the machine learning model could learn from structurally homogeneous and operationally consistent data, a robust classification scheme was implemented to stratify the dataset based on fabrication characteristics. This multi-step preprocessing and grouping strategy focused on both the type of structural elements fabricated and the scale of production, thereby improving model specificity and interpretability.

The dataset initially included 71 PEB projects. Incomplete, subcontracted, or conventionally fabricated projects were excluded, leaving 34 with verified fabrication records. A further filter was applied based on the presence of critical load bearing components specifically columns and rafters which are essential to PEB structure. Projects containing both components were designated as type P1, while those containing either component were categorized as type P2. This binary classification ensured the inclusion of structurally significant and fabrication-relevant projects.

The filtered projects were then grouped according to fabrication quantity thresholds measured in metric tons (MT). Projects were categorized into three scales: small (0–30 MT), medium (30–80 MT), and large (>80 MT). Only medium and large-scale projects were retained to mitigate skewness and reduce noise from under-representative records, resulting in a final dataset of 10 P1 and 12 P2 projects.

To further delineate complexity, projects were classified into three component-based categories: C1, C2, and C3. These categories represent increasing structural and fabrication complexity, with C1 comprising primary members such as columns and rafters, C2 incorporating additional load-distribution elements like joist and portal beams, and C3 encompassing specialized components such as crane beams and jack beams. Each project was assessed based on the proportion of its total fabricated components that belonged to each combination. Minimum thresholds were imposed to ensure dominant representation: 75% for C1, 80% for C2, and 85% for C3. These values were determined in consultation with domain experts from the fabrication facility to align with operational practice, where a component group is considered representative only when the majority of fabricated parts belong to it. Projects meeting these thresholds were respectively categorized into P1C1–P2C1, P1C2–P2C2, and P1C3–P2C3 classification groups.

This dual-layered classification based on fabrication type and component complexity yielded six mutually exclusive and structurally coherent project groups. Each group reflects a distinct fabrication profile, allowing for stratified training and evaluation of the predictive model. Notably, among the 29 active fabrication components observed across all projects, 15 were common to all classification tiers, while 11 appeared in at least two groups, and three were exclusive to the C3 category. This distribution reinforces the validity of the classification framework and its alignment with real-world fabrication heterogeneity.

After preprocessing and removal of incomplete records, the final modelling dataset comprised 34 completed PEB projects. These instances were distributed across the three complexity classes as follows: P1C1–P2C1: [20] records, P1C2–P2C2: [25], P1C3–P2C3: [28]. This breakdown ensured that each class contained a sufficient number of observations for training and evaluation, while preserving the intended separation between project types and complexity tiers.

The final result of this preprocessing is a dataset that captures the diverse operational realities of PEB fabrication while maintaining sufficient uniformity within each group to enable accurate and interpretable machine learning predictions.

4. Methodology

This section outlined the predictive modeling approach adopted for forecasting fabrication durations in PEB projects using real-time production data. The methodology incorporated ensemble machine learning, complexity-based data segmentation, and a profitability-aware evaluation framework.

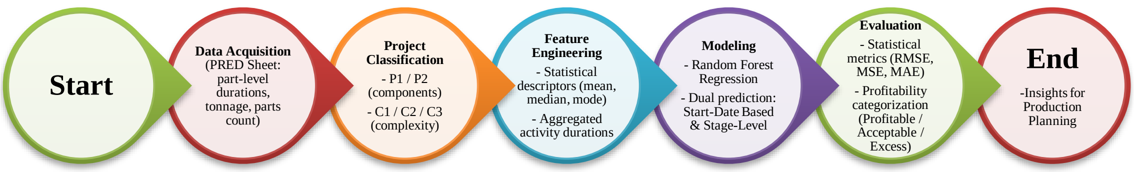

Fig. 1. Methodology flowchart

Fig. 1. Methodology flowchart showing the sequence from classification-based project grouping to feature engineering, Random Forest regression modeling, and profitability-oriented evaluation.

4.1 Model Selection

A Random Forest Regressor was selected for its proven effectiveness in capturing non-linear dependencies and managing overfitting. Unlike linear models that assume additive effects, Random Forests can capture the complex interactions that arise from fabrication sequences, variable workloads, and project heterogeneity. Its robustness to outliers and capability to operate with minimal preprocessing make it suitable for real-world industrial datasets. Implementation was carried out using the Scikit-learn Python package.

Random Forest Regression was chosen not only based on evidence from the literature but also due to its methodological suitability for the nature of fabrication data. The dataset was high-dimensional, with a mixture of categorical and continuous engineered features. Random Forest can handle such datasets effectively without extensive preprocessing, was not sensitive to feature scaling, and captures non-linear dependencies that linear models cannot address. Furthermore, it mitigates overfitting through ensemble averaging, making it robust for real-world production data.

The predictive performance observed in this study (Sections 4.6.1–4.6.5) further confirms that Random Forest aligns with the requirements of multi-stage fabrication forecasting, producing low error rates and profitability-consistent outputs across complexity tiers. These results validate Random Forest as a methodologically appropriate and empirically effective choice for this problem setting.

4.2 Feature Inputs

The input feature set was derived from the PRED Sheet and included a combination of static, dynamic, and engineered features. Static fields such as job code, number of parts, member weights, and total tonnage formed the base attributes. Dynamic activity durations and completion percentages extracted from automated spreadsheet formulas enhanced the temporal granularity. Features were grouped into three primary types:

- Project attributes: Total parts, tonnage, dominant member types (e.g., columns, rafters).
- Activity durations: Mean, median, mode for each of the ten production stages.
- Completion indicators: Real-time percentages derived from daily updates.

Categorical variables were label encoded. Due to the decision-tree architecture of the Random Forest model, feature normalization or scaling was not required.

4.3 Training and Validation Strategy

To preserve stratified project characteristics, model training was performed separately for each grouping category (P1C1–P2C1, P1C2–P2C2, P1C3–P2C3). Within each class, data was split into 80% training and 20% testing sets. A 5-fold cross-validation scheme was employed during hyperparameter tuning. Hyperparameters tuned include:

- Number of trees (n_estimators)
- Maximum depth of trees (max_depth)
- Minimum samples required to split a node (min_samples_split)

Grid search optimization was used, with RMSE as the scoring metric to ensure penalty on large deviations.

4.4 Prediction Targets

Two target variables were defined to satisfy both managerial and operational forecasting needs:

- Overall Project Duration: Measured in working days from the project's start date to final dispatch.
- Activity-Level Durations: Time required for specific fabrication stages such as Cutting, Welding, and Painting.

This dual target approach enabled both macro-level (project-wide) and micro-level (activity-specific) predictive insight.

4.5 Evaluation Metrics

The model was evaluated using traditional regression metrics:

- Root Mean Square Error (RMSE) – Highlights model sensitivity to large prediction errors.
- Mean Squared Error (MSE) – Reflects the average squared deviation across all samples.

Additionally, a custom classification scheme was developed to assess predictions based on practical thresholds for profitability and schedule tolerance. Predictions were categorized as:

- Profitable (≤ −5% deviation from actual)
- Acceptable (±5% deviation)
- Excess (> +5% overestimation)
- Shortfall (> +5% underestimation)

After obtaining predicted and actual fabrication durations for a given part, the percentage deviation is calculated using the formula: ("Predicted"-"Actual")/"Actual"×100. For instance, if the model predicts 12 days and the actual duration is 10 days, the percentage deviation is (12-10)/10×100=20%. According to the study's classification criteria, predictions with a deviation between 0% and +10% are considered Profitable; those exceeding +10% are classified as Excess; and deviations between 0% and -10% indicate More Profitable scheduling. In this example, the 20% deviation places the prediction in the Excess category, signaling an overestimation of time that could lead to inefficient resource allocation in project planning.

4.6 Model Performance Summary

4.6.1 Statistical Error Analysis

This subsection presents a detailed analysis of the model's prediction performance using Root Mean Square Error (RMSE) and Mean Squared Error (MSE), computed for total project durations and activity-level forecasts across the three project complexity groups—P1C1–P2C1 (low complexity), P1C2–P2C2 (medium complexity), and P1C3–P2C3 (high complexity).

4.6.1.1 RMSE-Based Evaluation by Classification

Low Complexity (P1C1–P2C1): The model demonstrated excellent predictive performance. The RMSE for total duration from Start Date (TSD) was 2.12 days, and activity-level RMSEs for deterministic stages like Cutting (SCA, PCA) and Splicing (SSA, PSA) were 0.00, indicating high confidence and standardization. Mid-stream processes such as SAW Welding (SSWA: 2.27), Fit-up (SFA: 2.27), and Welding (SWA: 0.87) had relatively low RMSE values. Finishing activities showed higher variability: Painting average from SD (SPA: 7.69) and Finish Paint average from SD (SFPA: 7.67) highlighting their dependency on external variables. Median RMSEs were closely aligned, reinforcing stable model performance, as shown in Fig. 2 (a) which depicts RMSE for Start-Date-Based and Activity-Level Predictions in Low-Complexity Projects (P1C1–P2C1).

Medium Complexity (P1C2–P2C2): With increased fabrication diversity, RMSE values moderately increased. The RMSE for TSD rose to 2.50 days. SAW Welding (SSWA: 3.06) and Drilling (SDA: 3.55) became more variable. Painting (SPA: 8.93), Finish Paint (SFPA: 8.97), and Grinding average from SD (SGA: 2.78) continued to exhibit significant dispersion. However, base activities remained stable (Cutting and Splicing: 0.00). Median RMSEs again mirrored average values, confirming that errors were not driven by extreme outliers, as detailed in Fig. 2 (b) which is RMSE for Start-Date-Based and Activity-Level Predictions in Medium-Complexity Projects (P1C2–P2C2).

High Complexity (P1C3–P2C3): Although the TSD RMSE was comparable (2.40 days), increased component variability led to sharper divergence in certain activities. SAW Welding average from SD (SSWA: 2.62), Fit-up average form SD (SFA: 3.94), Welding average from SD (SWA: 4.35), and Grinding (SGA: 6.27) saw higher RMSEs. Notably, Primer (SPA: 9.81) and Finish Paint (SFPA: 9.97) recorded the highest variability across all activities. In contrast, Cutting and Splicing remained predictable with 0.00 RMSE. The median RMSEs, such as SPMe and SFPMe, were significantly lower (~1.00), demonstrating that while outliers increased the mean error, core predictions remained centered, as shown in Fig. 2 (c) RMSE for Start-Date-Based and Activity-Level Predictions in High-Complexity Projects (P1C3–P2C3).

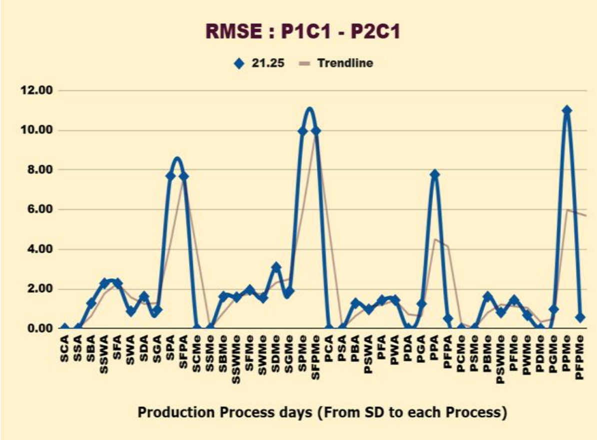

Fig. 2 (a). RMSE for Start-Date-Based and Activity-Level Predictions in Low-Complexity Projects (P1C1–P2C1)

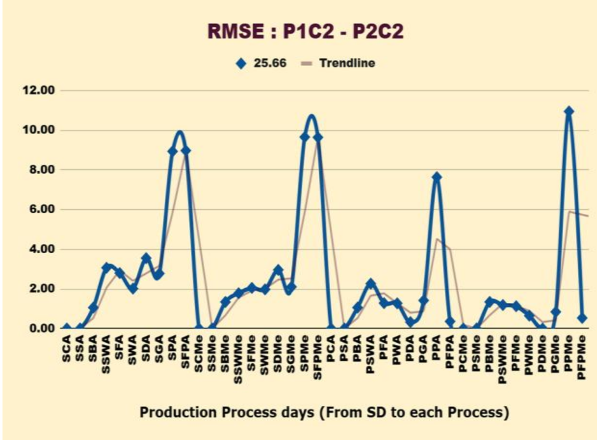

Fig. 2 (b). RMSE for Start-Date-Based and Activity-Level Predictions in Medium-Complexity Projects (P1C2–P2C2)

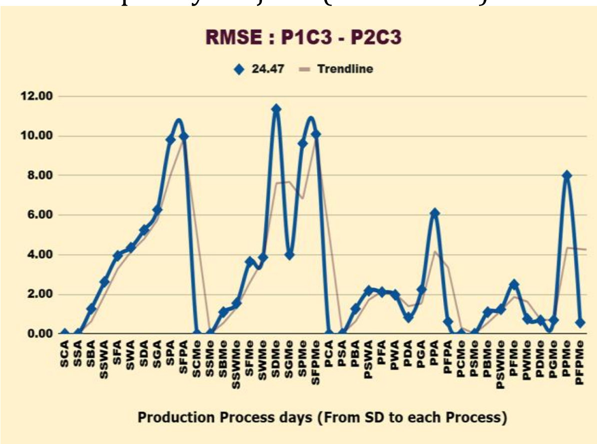

Fig. 2 (c): RMSE for Start-Date-Based and Activity-Level Predictions in High-Complexity Projects (P1C3–P2C3)

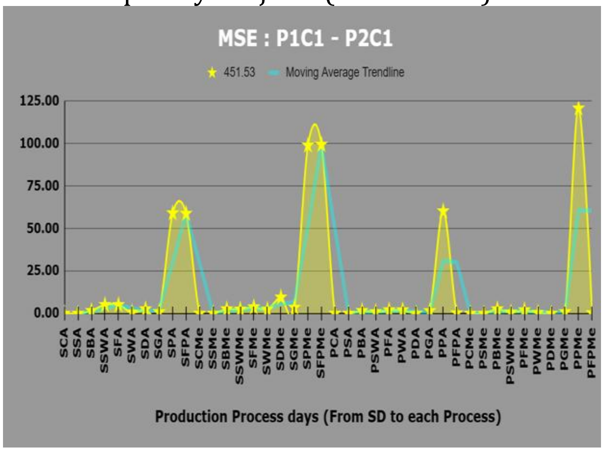

Fig. 3 (a). MSE for Start-Date-Based and Activity-Level Predictions in Low-Complexity Projects (P1C1–P2C1)

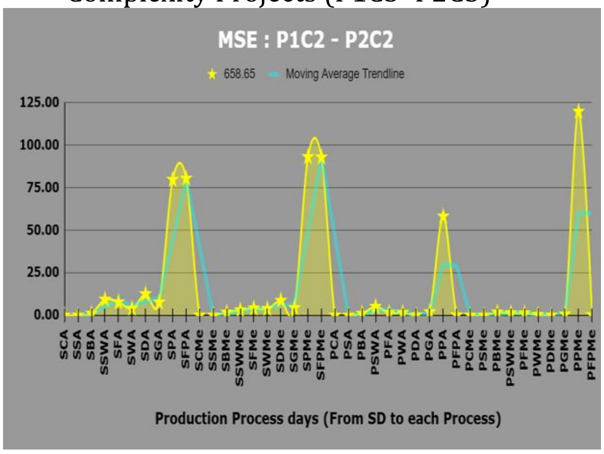

Fig. 3 (b). MSE for Start-Date-Based and Activity-Level Predictions in Medium-Complexity Projects (P1C2–P2C2)

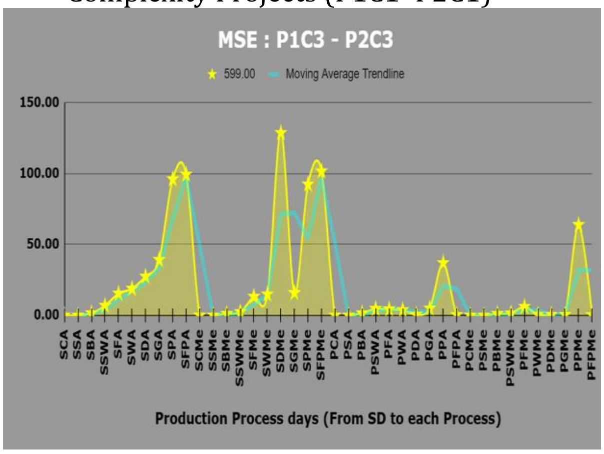

Fig. 3 (c). MSE for Start-Date-Based and Activity-Level Predictions in High-Complexity Projects (P1C3–P2C3)

4.6.1.2 MSE-Based Evaluation by Classification

Low Complexity (P1C1–P2C1): The MSE for (TSD) was 4.52 days^2. SAW Welding (SSWA: 5.17), Grinding (SGA: 0.90), and Painting activities like SPA (5.91) and SFPA (5.88) showed notable squared errors. Other stages remained within 2.5 days^2, indicating the model's strong resistance to large deviations in simpler projects, from Fig. 3 (a) which shows MSE for Start-Date-Based and Activity-Level Predictions in Low-Complexity Projects (P1C1–P2C1).

Medium Complexity (P1C2–P2C2): The MSE for TSD increased to 6.50 days^2, with elevated squared errors in SAW Welding (SSWA: 9.38), Grinding (SGA: 7.70), and Painting stages (SPA: 7.90, SFPA: 8.00). Fit-up and Welding also crossed the 4.0 mark, indicating increased forecast instability as task complexity and variability grew, as presented in Fig. 3 (b) - MSE for Start-Date-Based and Activity-Level Predictions in Medium-Complexity Projects (P1C2–P2C2).

High Complexity (P1C3–P2C3): The MSE for TSD dropped slightly to 5.00 days^2, potentially due to overcorrected predictions. Activity-level MSEs for Saw Welding mean average (SSWA: 6.89) and Grinding average (SGA: 3.90) remained significant. Interestingly, Painting stages had lower MSE values (SPA: 0.90, SFPA: 0.90) possibly due to conservative adjustments in response to previous overestimations. The pattern supports that model variance was concentrated in a limited set of outlier tasks, as depicted in Fig. 3 (c) which represents MSE for Start-Date-Based and Activity-Level Predictions in High-Complexity Projects (P1C3–P2C3).

As shown in Figs. 2 (a), 2 (b) and 2 (c), the RMSE values for both start-date-based and activity-level predictions differ across low, medium, and high-complexity projects, respectively. Similarly, MSE results for start-date-based and activity-level predictions are shown in Fig. 3 (a) for low-complexity, Fig. 3 (b) for medium-complexity, and Fig. 3 (c) for high-complexity projects. This explains that the Random Forest model consistently delivered high predictive accuracy in low-complexity projects, with moderate performance in medium and high-complexity settings. Errors increased notably in stages susceptible to variability such as SAW Welding, Grinding, and Painting. Median RMSE values were generally lower than their means, reflecting resilience to outliers. This group-based insight confirms that while deterministic activities maintain high fidelity, finishing operations remain sensitive to contextual fluctuations.

Overall, both RMSE and MSE metrics affirm the model's robustness and applicability in operational forecasting. Its performance supports deployment in industrial PEB scheduling environments with potential future improvements through integration of real-time production signals and environmental parameters.

4.6.2 Classification-Wise Prediction Performance

To assess the consistency and adaptability of the Random Forest model across different fabrication project types, three set of pairs were analyzed: P1C1 vs P2C1, P1C2 vs P2C2, and P1C3 vs P2C3. Each pair represented a balance of structural complexity and production scale. The goal was to evaluate model robustness in predicting both total project duration (from Start Date to dispatch) and individual activity-wise durations. The predictions were evaluated using a structured tabular framework, the sample table template as shown in Figs. 4 (a–c). For each classification, four sub-tables were generated:

- Original values (actual durations),
- Predicted values from the Random Forest model,
- Absolute error (difference between actual and predicted),
- Cumulative interpretation: percentage-based summary and qualitative categorization.

The comparative analysis of original data, ML predictions, error heatmaps, and profitability classification are presented in Fig. 4 (a) for low-complexity. The model demonstrated excellent agreement with actual values. Most predictions remained within an absolute deviation of ±5%, and the cumulative summary showed that 95% of the outputs fell under profitable, more profitable, or acceptable categories. Only one instance approached a shortfall boundary, suggesting conservative tendencies, which are often favorable in production planning.

The comparative analysis of original data, ML predictions, error heatmaps, and profitability classification are presented in Fig. 4 (b) for medium-complexity. The projects performance remained consistent despite increased complexity. The average absolute error was 2.03%, and all predictions stayed within tolerable thresholds. The model exhibited stability in forecasting mid-range projects, balancing precision with production-safe margins. The comparative analysis of original data, ML predictions, error heatmaps, and profitability classification are presented in Fig. 4 (c) for High-complexity.

Sl.No.	Code	Total MT	TSD	SCA	SSA	SBA	SSWA	.....	PWMe	PDMe	PGMe	PPMe	PFPMe		
1	P29	45.20	46	1	1	4	8	.....	2	2	1	4	2	Original Data	
2	P36	42.78	33	1	1	2	3	.....	3	2	3	3	2		
3	P43	99.08	70	1	1	3	5	.....	2	2	3	28	2		
4	P46	90.13	40	1	1	4	6	.....	2	2	2	5	2		
5	P51	42.97	20	1	1	2	5	.....	3	2	2	6	1		
Sl.No.	Code	Total MT	TSD	SCA	SSA	SBA	SSWA	.....	PWMe	PDMe	PGMe	PPMe	PFPMe	ML Predictions	
1	P29	45.2	42	1	1	4	7	.....	2	2	1	5	2		
2	P36	42.78	36	1	1	2	4	.....	3	2	3	5	2		
3	P43	99.08	59	1	1	3	5	.....	2	2	3	21	2		
4	P46	90.13	40	1	1	4	6	.....	2	2	2	7	2		
5	P51	42.97	28	1	1	3	6	.....	3	2	2	6	1		
Sl.No.	Code	Total MT	TSD	SCA	SSA	SBA	SSWA	.....	PWMe	PDMe	PGMe	PPMe	PFPMe	Difference %	
1	P29	45.20	-8.63%	0.00%	-0.03%	-1.13%	-1.00%	.....	0.42%	0.00%	0.86%	1.55%	-0.24%		
2	P36	42.78	7.22%	0.00%	0.00%	1.68%	2.81%	.....	-0.56%	0.00%	-0.94%	4.68%	-0.21%		
3	P43	99.08	-11.61%	0.00%	-0.44%	0.41%	0.51%	.....	0.13%	0.00%	-0.31%	-7.45%	-0.09%		
4	P46	90.13	-0.52%	0.00%	0.00%	-0.71%	-0.33%	.....	0.22%	0.00%	0.06%	2.01%	-0.17%		
5	P51	42.97	19.50%	0.00%	0.00%	1.59%	1.53%	.....	-0.70%	0.00%	-0.21%	0.26%	0.79%		
Excess Prediction (above 5%)	min		7.22%	0.00%	0.00%	0.00%	0.00%	.....	0.00%	0.00%	0.00%	0.00%	0.00%	Cumulative Average to see overall trends	0.18%
	max		19.50%	0.00%	0.00%	0.00%	0.00%	.....	0.00%	0.00%	0.00%	0.00%	0.00%		0.48%
	Total %		40.00%	0.00%	0.00%	0.00%	0.00%	.....	0.00%	0.00%	0.00%	0.00%	0.00%		0.98%
Shortfall Prediction (below -5%)	min		-11.61%	0.00%	0.00%	0.00%	0.00%	.....	0.00%	0.00%	0.00%	-7.45%	0.00%		-1.17%
	max		-8.63%	0.00%	0.00%	0.00%	0.00%	.....	0.00%	0.00%	0.00%	-7.45%	0.00%		-1.10%
	Total %		40.00%	0.00%	0.00%	0.00%	0.00%	.....	0.00%	0.00%	0.00%	20.00%	0.00%		3.90%
Profit (P) : (0 to 5%)			0.00%	0.00%	0.00%	60.00%	60.00%	.....	60.00%	0.00%	40.00%	80.00%	20.00%		43.90%
More Profit : (0 to -5%)			20.00%	100.00%	100.00%	40.00%	40.00%	.....	40.00%	100.00%	60.00%	0.00%	80.00%		51.22%
Acceptable P : (-5 to 5%			20.00%	100.00%	100.00%	100.00%	100.00%	.....	100.00%	100.00%	100.00%	80.00%	100.00%		95.12%

Fig. 4 (a). Comparison of Original Data, ML Predictions, Differences Heatmap and Profitability Classification (Acceptable, Excess, Shortfall) for Low-Complexity Projects (P1C1 vs. P2C1)

Sl.No.	Code	Total MT	TSD	SCA	SSA	SBA	SSWA	....	PWMe	PDMe	PGMe	PPMe	PFPMe		
1	P29	45.20	46	1	1	4	8	....	2	2	1	4	2	Original Data	
2	P31	73.82	77	1	1	2	11	....	2	2	2	7	2		
3	P36	42.78	33	1	1	2	3	....	3	2	3	3	2		
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮		
Sl.No.	Code	Total MT	TSD	SCA	SSA	SBA	SSWA	....	PWMe	PDMe	PGMe	PPMe	PFPMe	ML Predictions	
1	P29	45.20	42	1	1	3	7	....	2	2	1	5	2		
2	P31	73.82	65	1	1	2	9	....	2	2	2	6	2		
3	P36	42.78	40	1	1	2	4	....	3	2	3	6	2		
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮		
Sl.No.	Code	Total MT	TSD	SCA	SSA	SBA	SSWA	....	PWMe	PDMe	PGMe	PPMe	PFPMe	Difference %	
1	P29	45.20	-8.01%	0.00%	-0.03%	-1.42%	-0.95%	....	0.51%	0.00%	0.95%	1.11%	-0.27%		
2	P31	73.82	-16.36%	0.00%	-0.02%	-0.03%	-1.91%	....	0.15%	0.00%	0.07%	-0.89%	-0.01%		
3	P36	42.78	15.40%	0.00%	0.00%	1.32%	3.23%	....	-0.61%	0.00%	-0.65%	8.09%	-0.23%		
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮		
Excess Prediction (above 5%)	min		15.40%	0.00%	0.00%	0.00%	0.00%	....	0.00%	0.00%	0.00%	8.09%	0.00%	Cumulative Average to see overall trends	0.83%
	max		18.15%	0.00%	0.00%	0.00%	0.00%	....	0.00%	0.00%	0.00%	8.09%	0.00%		0.90%
	Total %		33.33%	0.00%	0.00%	0.00%	0.00%	....	0.00%	0.00%	0.00%	16.67%	0.00%		2.03%
Shortfall Prediction (below -5%)	min		-16.36%	0.00%	0.00%	0.00%	0.00%	....	0.00%	0.00%	0.00%	-8.93%	0.00%		-1.45%
	max		-8.01%	0.00%	0.00%	0.00%	0.00%	....	0.00%	0.00%	0.00%	-8.93%	0.00%		-1.19%
	Total %		50.00%	0.00%	0.00%	0.00%	0.00%	....	0.00%	0.00%	0.00%	16.67%	0.00%		4.47%
Profit (P) : (0 to 5%)			16.67%	0.00%	0.00%	50.00%	66.67%	....	66.67%	0.00%	50.00%	33.33%	16.67%		41.87%
More Profit : (0 to -5%)			0.00%	100.00%	100.00%	50.00%	33.33%	....	33.33%	100.00%	50.00%	33.33%	83.33%		51.63%
Acceptable P : (-5 to 5%)			16.67%	100.00%	100.00%	100.00%	100.00%	....	100.00%	100.00%	100.00%	66.67%	100.00%		93.50%

Fig. 4 (b). Comparison of Original Data, ML Predictions, Differences Heatmap and Profitability Classification (Acceptable, Excess, Shortfall) for Medium-Complexity Projects (P1C2 vs. P2C2)

Sl.No.	Code	Total MT	TSD	SCA	SSA	SBA	SSWA	....	PFMe	PWMe	PDMe	PGMe	PPMe	PFPMe	
1	P24	62.58	28	1	1	2	4	....	2	1	1	2	8	2	Original Data
2	P31	73.82	77	1	1	2	11	....	2	2	2	2	7	2	
3	P32	100.99	73	1	1	2	5	....	8	3	3	3	11	2	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
Sl.No.	Code	Total MT	TSD	SCA	SSA	SBA	SSWA	....	PFMe	PWMe	PDMe	PGMe	PPMe	PFPMe	ML Predictions
1	P24	62.58	34	1	1	2	5	....	3	2	1	2	7	2	
2	P31	73.82	63	1	1	2	9	....	2	2	2	2	6	2	
3	P32	100.99	59	1	1	2	5	....	6	3	3	3	9	2	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
Sl.No.	Code	Total MT	TSD	SCA	SSA	SBA	SSWA	....	PFMe	PWMe	PDMe	PGMe	PPMe	PFPMe	Difference %
1	P24	62.58	10.21%	0.00%	-0.08%	0.38%	0.66%	....	1.15%	0.80%	0.54%	0.22%	-0.91%	-0.16%	
2	P31	73.82	-18.38%	0.00%	-0.02%	-0.31%	-2.33%	....	0.54%	0.22%	-0.07%	0.20%	-1.04%	-0.11%	
3	P32	100.99	-13.86%	0.00%	-0.04%	0.48%	0.24%	....	-1.79%	-0.31%	-0.50%	-0.27%	-1.73%	-0.12%	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

		TSD	SCA	SSA	SBA	SSWA	....	PFMe	PWMe	PDMe	PGMe	PPMe	PFPMe	Cumulative Average to see overall trends
Excess Prediction (above 5%)	min	5.14%	0.00%	0.00%	0.00%	0.00%	....	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.84%
	max	14.94%	0.00%	0.00%	0.00%	0.00%	....	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.54%
	Total %	50.00%	0.00%	0.00%	0.00%	0.00%	....	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	8.54%
Shortfall Prediction (below -5%)	min	-18.38%	0.00%	0.00%	0.00%	0.00%	....	0.00%	0.00%	0.00%	0.00%	-7.35%	0.00%	-3.52%
	max	-11.28%	0.00%	0.00%	0.00%	0.00%	....	0.00%	0.00%	0.00%	0.00%	-7.35%	0.00%	-2.68%
	Total %	37.50%	0.00%	0.00%	0.00%	0.00%	....	0.00%	0.00%	0.00%	0.00%	12.50%	0.00%	6.71%
Profit (P) : (0 to 5%)		12.50%	0.00%	0.00%	62.50%	62.50%	....	62.50%	62.50%	50.00%	62.50%	50.00%	25.00%	42.07%
More Profit : (0 to -5%)		0.00%	100.00%	100.00%	37.50%	37.50%	....	37.50%	37.50%	50.00%	37.50%	37.50%	75.00%	42.68%
Acceptable P : (-5 to 5%)		12.50%	100.00%	100.00%	100.00%	100.00%	....	100.00%	100.00%	100.00%	100.00%	87.50%	100.00%	84.76%

Fig. 4 (c). Comparison of Original Data, ML Predictions, Differences Heatmap, and Profitability Classification (Acceptable, Excess, Shortfall) for High-Complexity Projects (P1C3 vs. P2C3)

The projects involved higher variability in structure, sequencing, and task durations. Yet, the model maintained reliable predictive behavior. The average cumulative deviation stood at 1.84%, and most outputs remained within acceptable or better classifications.). Only one project approached a mild shortfall, reinforcing the model's generalization capability across fabrication complexities. Overall, the structured comparison across these classification pairs validates the model's ability to produce accurate and consistent predictions across diverse project types. This forms the baseline for deeper trend analysis (Section 4.6.3) and classification-based profitability assessments (Section 4.6.4

4.6.3 Profitability Classification Summary

To interpret the real-world scheduling implications of the model's predictions, a profitability-based evaluation framework was applied. Each predicted duration was compared to its actual counterpart and categorized as Excess (prediction >5% longer), Shortfall (prediction >5% shorter), or Acceptable (within ±5%). Within the Acceptable range, further refinement distinguished between Profitable (0 to −5%) and More Profitable (<−5%) predictions. Figs. 5 (a-c) and Figs. 6 (a-c) present these classifications for start-date-based and activity-specific predictions, respectively, across the three complexity tiers.

4.6.3.1 Start Date Based Predictions

For start-date-based predictions shown in Fig. 5 (a), low-complexity projects (P1C1–P2C1) revealed a perfectly balanced outcome: 40% of predictions were Excess and another 40% were Shortfall, with only 20% falling in the Acceptable range. While this highlights limitations in predicting total project timelines from the start date alone, the fact that prediction errors were evenly distributed suggests there was no systemic bias toward over- or underestimation in these simpler projects.

In Fig. 5 (b), for medium-complexity projects (P1C2–P2C2), start-date-based predictions became less reliable. A total of 50% of predictions fell under the Shortfall category, and 33.3% were Excess, leaving only 16.7% as Acceptable. This trend indicates a notable increase in prediction dispersion, which may be attributed to variability in part sequencing or minor overlaps in mid-stage activities.

The pattern continued in Fig. 5 (c) for high-complexity projects (P1C3–P2C3), where the model's prediction quality further degraded. Half the predictions were Excess, and 37.5% were Shortfall, with a mere 12.5% falling within the acceptable margin. This steep drop in predictive reliability highlights the challenge of accurately estimating total project timelines in projects characterized by component diversity, nonlinear production flows, and extended finishing operations.

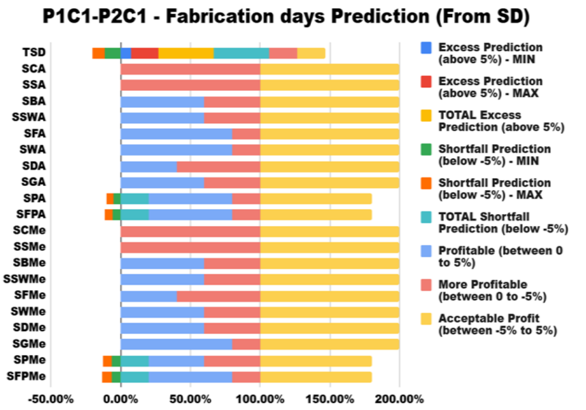

Fig. 5 (a). SD Based Forecasts for Low-Complexity (P1C1 vs. P2C1), including excess, shortfall, and profitability ranges

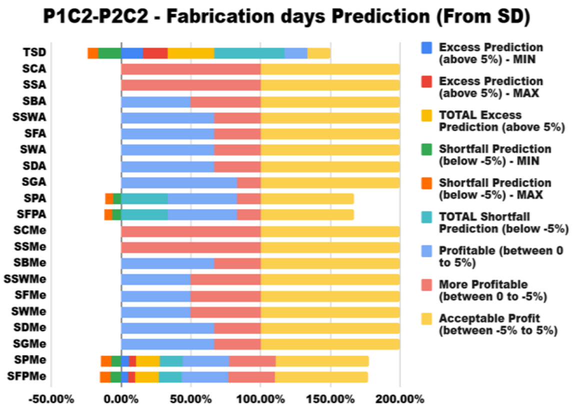

Fig. 5 (b). SD Based Forecasts for Medium-Complexity (P1C2 vs. P2C2), including excess, shortfall, and profitability ranges

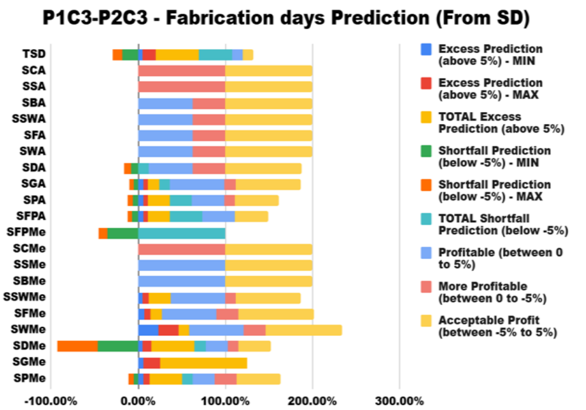

Fig. 5 (c). SD Based Forecasts for High-Complexity (P1C3 vs. P2C3), including excess, shortfall, and profitability ranges

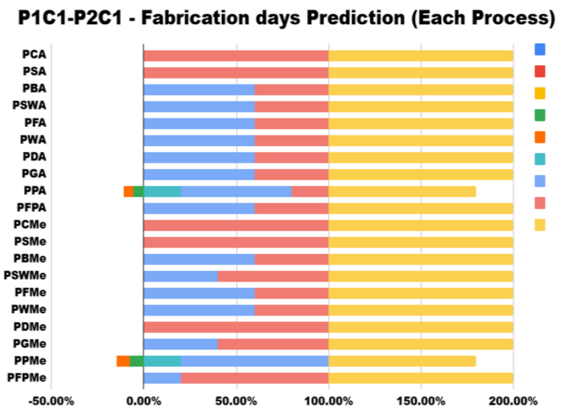

Fig. 6 (a). AD Forecasts for Low-Complexity (P1C1 vs. P2C1), including excess, shortfall, and profitability ranges

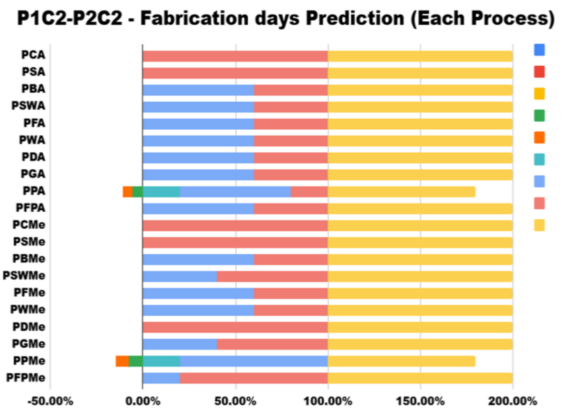

Fig. 6 (b). AD Forecasts for Medium-Complexity (P1C2 vs. P2C2), including excess, shortfall, and profitability ranges

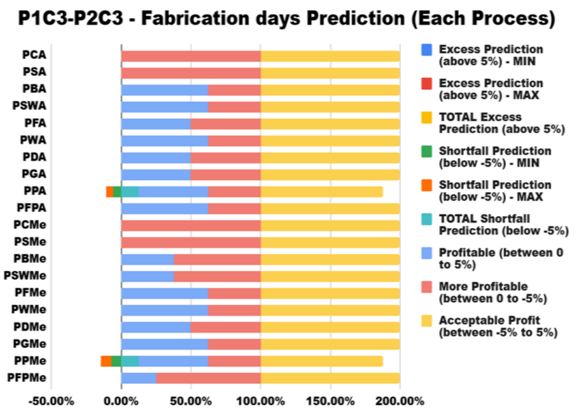

Fig. 6 (c). AD Forecasts for High-Complexity (P1C3 vs. P2C3), including excess, shortfall, and profitability ranges

4.3.6.2 Activity Specific Predictions

By contrast, the activity-specific classification results shown in Fig. 6 (a) for low complexity projects in P1C1–P2C1 projects were markedly better. All fabrication activities except for a few painting stages achieved 100% Acceptable classification, with no instances of Excess or Shortfall. For example, Cutting (PCA) and Splicing (PSA) were fully consistent, while even more variable tasks like Painting (PPA) only registered a 20% Shortfall. This confirms that for low-complexity jobs, activity-wise modeling captures fabrication durations with exceptional accuracy.

In medium-complexity projects (P1C2–P2C2), the results in Fig. 6 (b) reaffirm the advantage of activity-specific forecasting. While Painting (PPA) showed a 16.7% Shortfall, and Painting Median (PPMe) had both Excess and Shortfall at 16.7% each, all other fabrication stages maintained 100% Acceptable classification. Profitable and More Profitable segments were well-balanced, indicating that even under increased part variety and intermediate complexity, the model retains practical usefulness at the activity level.

Finally, in high-complexity projects (P1C3–P2C3), Fig. 6 (c) shows some dispersion but still maintains remarkable control compared to the corresponding start-date-based predictions. Most core fabrication tasks such as Cutting (PCA), Splicing (PSA), SAW Welding (PSWA), and Primer (PFPA) showed no Excess or Shortfall and remained within the Acceptable range. Even in the more variable stages like Painting (PPA) and its median values (PPMe), only 12.5% fell into the Shortfall range, while the rest were well-distributed between Profitable and More Profitable. These comparisons establish that while overall project duration forecasts tend to suffer from growing variance as complexity increases, activity-level prediction classification is consistently stable, accurate, and informative across all complexity tiers. This reinforces the conclusion that micro-level prediction granularity better accommodates structural variation and operational uncertainty inherent in complex Pre-Engineered Building projects.

4.6.4 Profitability Trend Analysis

To gain insight into how predictive performance shifts with increasing project complexity, a trend-based analysis of classification outcomes was conducted. Figs. 7 and 8 summarize the proportion of predictions falling into Excess, Shortfall, and Acceptable Profit categories across the three classification groups: P1C1–P2C1 (low), P1C2–P2C2 (medium), and P1C3–P2C3 (high complexity).

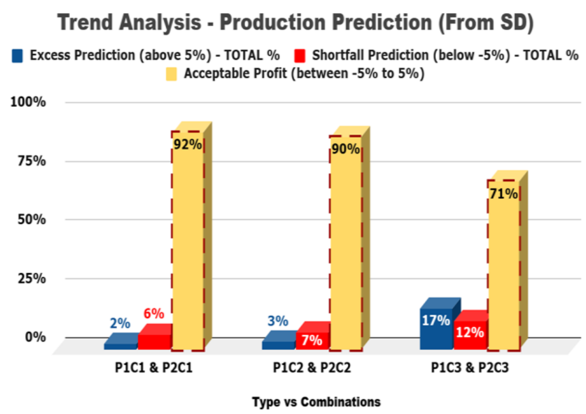

Fig. 7. Profitability Trend analysis of production (From SD)

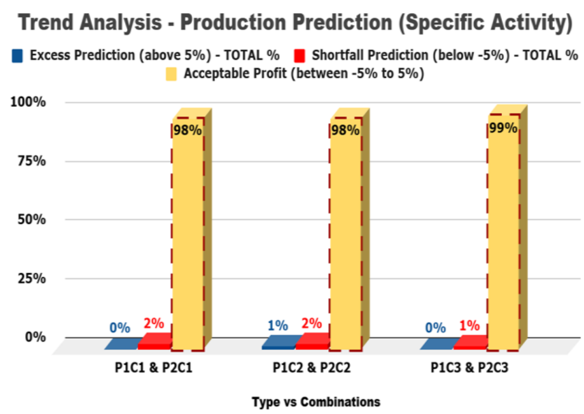

Fig. 8. Profitability trend analysis of production (Activity Based)

In Fig. 7, which visualizes the trends for start-date-based predictions, a clear decline in predictive reliability is observed as complexity increases. In the low-complexity group (P1C1–P2C1), Acceptable predictions accounted for 92%, with only 2% Excess and 6% Shortfall. This demonstrates strong schedule-aligned forecasting in standardized fabrication environments. However, in the medium-complexity group (P1C2–P2C2), Acceptable classification slightly dropped to 90%, while Excess and Shortfall rose to 3% and 7%, respectively. The decline becomes much steeper in the high-complexity group (P1C3–P2C3), where Acceptable predictions fall to just 71%, and Excess and Shortfall rise sharply to 17% and 12%, respectively. This pattern indicates the compounding effects of fabrication variability and inter-task dependencies on the model's cumulative timeline accuracy.

Conversely, Fig. 8, which illustrates the trend for activity-specific predictions, presents a striking contrast. In low-complexity projects, the model achieved 98% Acceptable predictions, with only 2% Shortfall and no Excess. In medium complexity, Acceptable predictions remained constant at 98%, while Excess and Shortfall slightly increased to 1% and 2%, respectively. Remarkably, in high-complexity projects, Acceptable classification further improved to 99%, with Shortfall dropping to 1% and no Excess predictions recorded. These trends reinforce a key operational insight: as project complexity increases, start-date-based predictions become increasingly volatile and prone to deviation, while activity-level predictions maintain high accuracy and profitability alignment. The ability of stage-focused modeling to isolate task durations independently of project-wide variability makes it a more robust strategy for forecasting in complex Pre-Engineered Building (PEB) fabrication workflows.

4.6.5 Profitability Consolidation Across Prediction Strategies

To consolidate the classification analysis and highlight the consistency of the predictive model across varying complexity tiers, Figs 8 and 9 present grouped bar charts showing the overall profit classification trends for both start-date-based and activity-specific prediction strategies. In Fig. 9, which illustrates results from start-date-based predictions, the overall profitability classification (i.e., combined share of Profitable and More Profitable predictions) shows a declining trend as project complexity increases. In the low-complexity group (P1C1–P2C1), the model achieved a strong 98% profitability classification, which slightly declined to 97% for medium-complexity projects (P1C2–P2C2). However, for the high-complexity group (P1C3–P2C3), profitability dropped notably to 83%, confirming that aggregate predictions become more susceptible to deviation as fabrication workflows become structurally diverse and less deterministic.

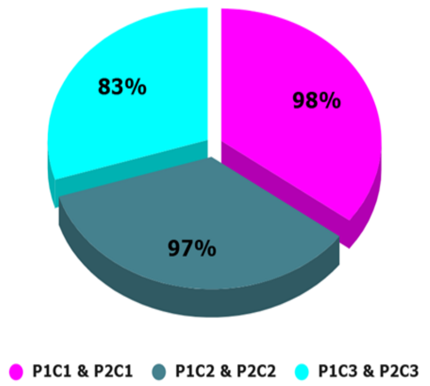

Fig. 9. Overall profit analysis for production (From SD)

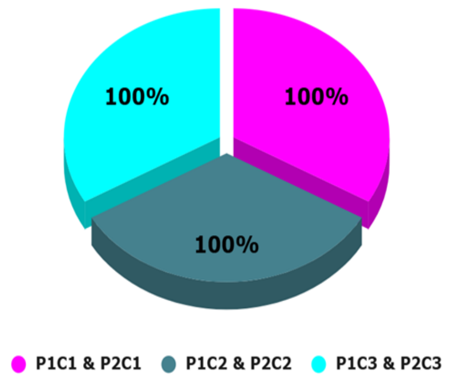

Fig. 10 Overall profit analysis for production (Activity Based)

In contrast, Fig. 10, which summarizes profitability classification for activity-specific predictions, demonstrates remarkable stability and robustness across all complexity levels. The model consistently achieved 100% profitability classification for low, medium, and high-complexity projects alike. This means that every activity-level prediction—regardless of project scale or structural heterogeneity—remained within the Profitable or More Profitable classification zones.

These findings visually reinforce the earlier statistical and trend-based conclusions: activity-specific modeling ensures highly profitable and stable forecasts, even in projects characterized by fabrication variability, concurrent workflows, and extended finishing durations. On the other hand, start-date-based predictions, while initially strong, tend to degrade under project complexity, underscoring the limitations of aggregated forecasting models for detailed production planning.

5. Discussion

The predictive performance analysis of this study provides both theoretical and empirical validation for the structured application of machine learning in real-time fabrication forecasting for PEB. A critical comparison between start-date-based and activity-specific regression models has revealed several important insights into how fabrication workflows, model architecture, and feature behaviors interact under varying degrees of project complexity.

An important methodological consideration in this study was the relatively small sample size within each classification subset. While small data regimes typically risk overfitting, the results showed close agreement between mean and median RMSE values, and profitability classifications were balanced across Excess and Shortfall categories. These outcomes suggest that the model did not memorize isolated instances but generalized patterns effectively. The use of bootstrap aggregation within Random Forest, complexity-aware stratification, and hyperparameter tuning collectively contributed to this resilience.

The stark divergence in error behavior particularly in Root Mean Square Error (RMSE) and Mean Squared Error (MSE) between start-date-based predictions and activity-specific predictions underscores the nonlinear error propagation inherent in cumulative timeline modeling. In high-complexity projects (P1C3–P2C3), aggregated predictions exhibited both amplified residuals and greater variance, suggesting model strain under feature interactions that are not easily decomposable. This is a well-observed behavior in ensemble tree-based models like Random Forests, which excel at isolating decision boundaries for localized predictions (e.g., specific activities), but become less interpretable and less accurate when tasked with predicting long, path-dependent sequences.

In contrast, the task-specific modeling approach constrained the regression task to independent, bounded intervals (e.g., welding duration, painting lead time), which allowed the Random Forest algorithm to operate with higher granularity and lower cumulative variance. The predictive fidelity across even high-complexity projects with significant concurrent tasks and variability in fabrication sequencing reinforces the ensemble model's capacity to effectively partition feature space when provided with stage-isolated data.

From a feature behavior standpoint, the engineered inputs particularly the statistical descriptors such as stage-level mean, median, and mode were key contributors to model generalization. These summary statistics act as stabilizing factors in the presence of noise and outliers. In addition, the classification-based stratification of the dataset into P1C1–P2C3 groups significantly improved model bias-variance trade-off by ensuring homogeneity within training samples. This stratification aligns with complexity theory in project scheduling, where performance is known to degrade in heterogeneous or poorly segmented datasets.

The classification performance summary, visualized through Fig. 4 to Fig. 10, further reinforces the limitations of cumulative prediction models under fabrication variability. The downward trend in profitability and acceptable prediction rates for start-date-based models highlights an exponential sensitivity to component diversity and flow disruptions. Meanwhile, the flat or improving trend in activity-level classifications demonstrates the superior scalability of stage-based models. Notably, no excess classifications were recorded in activity-level predictions for high-complexity projects—a rare behavior in predictive modeling that indicates excellent overfit control and structural resilience.

These findings have implications for both model design and production control. First, they validate the use of ensemble regressors with classification-aware preprocessing for industrial applications involving multistage workflows. Second, they show that disaggregating predictive tasks across operational stages provides not only statistical benefits (e.g., reduced error, variance), but also operational advantages (e.g., task-specific insights, bottleneck detection). Lastly, the introduction of a profitability-aware classification framework represents a methodological advancement, aligning predictive accuracy with actionable performance categories rooted in cost, schedule tolerance, and delivery reliability.

This study has several limitations that should be acknowledged. First, the dataset comprises 34 completed projects, which, while carefully preprocessed and stratified, remains relatively small and may limit the generalizability of the predictive models to other PEB projects with different characteristics. Second, model validation was performed using historical data from past projects, without external prospective validation. These limitations restrict the extent to which the models' predictive accuracy can be assured in new or evolving project contexts. Future research should aim to augment the dataset with additional and more diverse projects and validate the approach prospectively to strengthen robustness and applicability.

5.1 Break-Even Point (BEP)–Based Evaluation of Fabrication Efficiency

To gain a clearer understanding of how individual fabrication stages contribute to overall schedule variability, we evaluated performance using a Break-Even Point (BEP) framework. In this context, the BEP refers to the typical (median) number of days a given process takes across all recorded parts. By design, the median serves as a more robust baseline than the mean, as it reduces sensitivity to extreme outliers and rare delays.

Each fabrication stage was assessed by comparing its mean duration to its BEP. When the mean exceeded the BEP, the stage was flagged for review, with the degree of deviation used to classify the process as efficient, marginally overrun, or significantly delayed. This evaluation helps establish which stages are consistently under control and which ones may require deeper analysis or operational adjustment.

Table. 5 Fabrication Stage Performance Compared to Break-Even Point

Fab Stage	No. of Parts	BEP (Median Days)	Mean Days	Deviation from BEP	Efficiency Status
1. Cutting	31,547	1	1	0	Efficient
2. Splicing	9,473	1	1	0	Efficient
3. Built-Up Section	3,773	2	2	0	Efficient
4. SAW Welding	3,773	2	4	+2	Slight Overrun
5. Fit-up	24,853	2	3	+1	Slight Overrun
6. Welding	22,473	2	2	0	Efficient
7. Drilling	7,670	1	2	+1	Slight Overrun
8. Cleaning & Grinding	31,547	2	4	+2	Moderate Delay
9. Primer	31,547	2	14	+12	Significant Delay
10. Finish Paint	30,840	1	1	0	Efficient

The analysis and the findings are summarized in Table 5 and infers that the majority of fabrication stages particularly Cutting, Splicing, Welding, and Finish Paint operated at or below their BEP, suggesting strong control and consistent execution. A few stages, including Fit-up, SAW Welding, and Drilling, showed mild deviations (+1 to +2 days), which may be acceptable given typical production fluctuations. However, two stages stood out as clear bottlenecks. Primer application averaged 14 days compared to a BEP of 2, representing a substantial delay that could impact downstream scheduling. Cleaning and Grinding also exceeded its BEP by two full days on average, suggesting a need for either process streamlining or capacity realignment. These findings support the motivation for stage-level modeling in this study. By identifying where delays cluster and quantifying how far actual durations deviate from standard expectations, we can target high-variance stages for enhanced machine learning–based forecasting and optimization. Moreover, BEP-based classifications provide a domain-aligned reference point that can be incorporated into profitability-aware ML evaluation metrics. From a theoretical perspective, this study confirms the value of treating time-to-completion not as a monolithic regression target, but as a composition of bounded sub-tasks influenced by domain-specific complexity. Practitioners and researchers in smart manufacturing and construction informatics can extend this approach to integrate uncertainty quantification, real-time sensor feedback, or reinforcement learning-based rescheduling in future work.

Table: 6 Comparative Summary of Start-Date-Based (SD) and Activity-Based (AD) Forecasting Results across Complexity Tiers, including Error Metrics, Profitability Rates, and Break-Even Status

Complexity Tier	SD Based RMSE (Avg)	AD Based RMSE (Avg)	SD Based MSE (Avg)	AD MSE (Avg)	SD Based Profitability Rate	AD Based Profitability Rate	Break Even Status
P1C1–P2C1 (Low)	2.12	1.85	4.49	3.98	98	100	Achieved
P1C2–P2C2 (Medium)	2.5	2.14	6.25	5.11	97	100	Achieved
P1C3–P2C3 (High)	3.1	2.72	9.61	7.39	83	100	Achieved

To provide a consolidated view of both technical accuracy and business alignment, Table 6 summarizes start-date-based (SD) and activity-based (AD) forecasting results across all complexity tiers. The table integrates RMSE, MSE, profitability rates, and break-even status into a single comparative framework, highlighting the consistent advantage of activity-level predictions. This consolidated comparison is provided in Table 6, shows activity-based (AD) predictions consistently yield lower RMSE and MSE values across all complexity tiers compared to start-date-based (SD) predictions. Notably, while SD-based profitability rates decline with higher project complexity (98% → 97% → 83%), AD-based predictions maintain a 100% profitability rate across all tiers. Furthermore, the break-even status is achieved in every case, underscoring the robustness of the proposed framework.

6. Conclusion

This study developed a machine learning–driven approach for predicting fabrication durations in Pre-Engineered Building (PEB) projects using real-time production data from a structured PRED system. Leveraging an ensemble Regressor (Random Forest), the model was trained on a stratified dataset segmented by project complexity and structural configuration (P1C1–P2C3). Input features included total tonnage, number of parts, and statistical descriptors (mean, median, mode) of task durations, engineered from part-level fabrication records. Two target prediction strategies were evaluated start-date-based and activity-specific. The former suffered increased RMSE and MSE as project complexity rose, while the latter maintained low error rates and exceptional prediction consistency. Specifically, activity-level models achieved over 99% accuracy in profitability-aligned predictions across all complexity tiers, with no Excess in high-complexity projects. These results confirm that granular, task-isolated modeling significantly improves forecast robustness, especially under conditions of sequencing variability and concurrent task execution.

The integration of group-aware preprocessing, ensemble learning, and profitability-based evaluation establishes a scalable and interpretable solution for predictive scheduling in industrial fabrication. By aligning forecast accuracy with profitability classes—Profitable, Acceptable, and Excess—the framework translates technical predictions into direct financial implications. This enables fabrication managers to anticipate schedule risks, detect bottlenecks early, and optimize resource allocation, ultimately strengthening competitiveness in cost- and time-sensitive markets.

Technically, the results highlight the value of statistical feature engineering in stabilizing predictions, as descriptors such as mean and median mitigated the influence of noise and outliers. The classification-based stratification further ensured balanced learning across projects of differing complexity, minimizing bias–variance trade-offs. Together, these design choices explain the framework's superior performance compared to conventional approaches.

The demonstrated robustness of activity-level forecasts provides actionable insights for PEB firms by reducing uncertainty in project timelines and profitability planning. This makes the framework a valuable decision-support tool for managers seeking to balance operational efficiency with financial performance in competitive fabrication environments. Future work may incorporate repeated cross-validation, regularization techniques, or blended ensemble approaches to further reinforce generalization.

While the results are promising, further research may expand on this foundation. Future work could explore repeated cross-validation, regularization techniques, or blended ensemble approaches to strengthen generalization. Integration with IoT-enabled real-time feedback systems may enable adaptive rescheduling under disruptions, while reinforcement learning or probabilistic modeling could improve handling of uncertainty in highly variable projects. Finally, extending the framework to include supply chain and resource allocation considerations would allow for an even more holistic view of PEB project planning.

References

[1] Ramesh V, Pandulu G, Khanam N, Lian O C. Compressive strength prediction of ecofriendly recycled aggregate concrete: A machine learning approach. Res. Eng. Struct. Mater., 2025; 11(6): 2965-2982. http://dx.doi.org/10.17515/resm2025-595me1212rs

[2] Raju SMTU, Sarker A, Das A, Islam MM, Al-Rakhami MS, Al-Amri AM, Mohiuddin T, Albogamy FR. An approach for demand forecasting in steel industries using ensemble learning. Complexity. 2022:9928836. https://doi.org/10.1155/2022/9928836

[3] Zermane H, Madjour H, Bouzghaya MA. Prediction of the amount of raw material in an Algerian cement factory. Eurasia Proc Sci Technol Eng Math. 2022;19:41-46. https://doi.org/10.55549/epstem.1218718

[4] Zermane H, Madjour H, Ziar A, Zermane A. Forecasting material quantity using machine learning and time series techniques. J Electr Eng. 2024;75(3):237-248. https://doi.org/10.2478/jee-2024-0029

[5] Raja Gopal G ST. Analysis of factors affecting labour productivity in construction. Int J Recent Sci Res. 2016;7(6).

[6] Salihu AA, Ibrahim Y, Muhammad S. Impact of labour productivity factors on construction project cost and time. LAUTECH J Civ Environ Stud. 2022;8(1). https://doi.org/10.36108/laujoces/2202.80.0110

[7] Soham M, Rajiv B. Critical factors affecting labour productivity in construction projects: Case study of South Gujarat region of India. Int J Eng Adv Technol. 2013;(2).

[8] Shinde PP, Shah SA. A review of machine learning and deep learning applications. IEEE; 2018. https://doi.org/10.1109/ICCUBEA.2018.8697857

[9] Yusoff MIM. Machine learning: An overview. Open J Model Simul. 2024;12(3):89-99. https://doi.org/10.4236/ojmsi.2024.123006

[10] Tao F, Qi Q, Liu A, Kusiak A. Data-driven smart manufacturing. J Manuf Syst. 2018;48:157-169. https://doi.org/10.1016/j.jmsy.2018.01.006

[11] Abdulfattah BS, Abdelsalam HA, Abdelsalam M, Bolpagni M, Thurairajah N, Perez LF, Butt TE. Predicting implications of design changes in BIM-based construction projects through machine learning. Automation in Construction. 2023;155:105057. https://doi.org/10.1016/j.autcon.2023.105057

[12] Fahle S, Prinz C, Kuhlenkötter B. Systematic review on machine learning (ML) methods for manufacturing processes - Identifying artificial intelligence (AI) methods for field application. Procedia CIRP. 2020;93:413-418. https://doi.org/10.1016/j.procir.2020.04.109

[13] Saini S, Sharma N, Chopra R K, Mehra M, Kumar R, Bhardwaj S, Kumar D. Data-driven prediction of mechanical properties in friction stir processed Al6061-Alumina composite using enhanced machine learning models. Res. Eng. Struct. Mater., 2025; 11(5): 2097-2117. http://dx.doi.org/10.17515/resm2024.438me0907rs

[14] Sadatnya A, Sadeghi N, Sabzekar S, Khanjani M, Tak AN, Taghaddos H. Machine learning for construction crew productivity prediction using daily work reports. Automation in Construction. 2023;152:104891. https://doi.org/10.1016/j.autcon.2023.104891

[15] Villegas-Ch W, Navarro AM, Sanchez-Viteri S. Optimization of inventory management through computer vision and machine learning technologies. Intell Syst Appl. 2024;24:200438. https://doi.org/10.1016/j.iswa.2024.200438

[16] Ali SM, Paul SK, Ahsan K, Azeem A. Forecasting of optimum raw material inventory level using artificial neural network. International J. of Opers. and Quant. Management, 2011;17(4): 333-348.

[17] Duymaz Ş. The application of machine learning algorithms in the estimation of production lead times: A case study of a steel construction manufacturing company. J Adv Manuf Eng. 2024:21-28. https://doi.org/10.14744/ytu.jame.2024.00004

[18] You Z, Wu C. A framework for data-driven informatization of the construction company. Adv Eng Inform. 2019;39:269-277. https://doi.org/10.1016/j.aei.2019.02.002

[19] Elbeltagi E, Ammar M, Sanad H, Kassab M. Overall multiobjective optimization of construction projects scheduling using particle swarm. Eng Constr Archit Manag. 2016;23(3):265-282. https://doi.org/10.1108/ECAM-11-2014-0135

[20] Kang Z, Catal C, Tekinerdogan B. Machine learning applications in production lines: A systematic literature review. Comput Ind Eng. 2020;149:106773. https://doi.org/10.1016/j.cie.2020.106773
[21] Tiwari S. Artificial intelligence implications in engineering and production. ASEC. 2022;31(1):16. https://doi.org/10.3390/ASEC2022-13823
[22] Li Y, Carabelli S, Fadda E, Manerba D, Tadei R, Terzo O. Machine learning and optimization for production rescheduling in Industry 4.0. Int J Adv Manuf Technol. 2020;110(9-10):2445-2463. https://doi.org/10.1007/s00170-020-05850-5
[23] Guo K, Yang Z, Yu CH, Buehler MJ. Artificial intelligence and machine learning in design of mechanical materials. Mater Horiz. 2021;8(4):1153-1172. https://doi.org/10.1039/D0MH01451F
[24] Gao Y, Wang J, Xu X. Machine learning in construction and demolition waste management: Progress, challenges, and future directions. Automation in Construction. 2024;162:105380. https://doi.org/10.1016/j.autcon.2024.105380
[25] Bojer CS. Understanding machine learning-based forecasting methods: A decomposition framework and research opportunities. Int J Forecasting. 2022;38(4):1555-1561. https://doi.org/10.1016/j.ijforecast.2021.11.003
[26] Jin B, Xu X. Machine learning-based scrap steel price forecasting for the Northeast Chinese market. Int J Empir Econ. 2024. https://doi.org/10.1142/S2810943024500112
[27] Hwang SH, Mangalathu S, Shin J, Jeon JS. Estimation of economic seismic loss of steel moment-frame buildings using a machine learning algorithm. Eng Struct. 2022;254:113877. https://doi.org/10.1016/j.engstruct.2022.113877
[28] Guo S, Yu J, Liu X, Wang C, Jiang Q. A predicting model for properties of steel using the industrial big data based on machine learning. Comput Mater Sci. 2019;160:95-104. https://doi.org/10.1016/j.commatsci.2018.12.056
[29] Kawakami K, Kobayashi H, Nakata K. Seasonal inventory management model for raw materials in steel industry. INFORMS J Appl Anal. 2021;51(4):312-324. https://doi.org/10.1287/inte.2021.1073