Role of Machine Learning in Predicting project Delays

Mr. Abhishek M Patwardhan[1], Dr. P. M. Alandkar[2], A. B. Khemalapure[3]
[1]*Student, ME construction and management, Department of civil Engineering, RMD Sinhagad Technical Institutes Campus, Warje*
[2]*Professor Head of Department, Department of civil Engineering, RMD Sinhagad Technical Institutes Campus, Warje*
[3]*Assistant Professor, Department of civil Engineering, RMD Sinhagad Technical Institutes Campus, Warje*

***Abstract*—Construction project delays are a pervasive challenge globally, leading to significant cost overruns, contractual disputes, and decreased client satisfaction. Traditional project management and statistical methods often rely on lagging indicators and fail to capture the high-dimensional, non-linear relationships that exist among the delay factors (labor, supply chain, design changes, external factors).**

This project explores the role and effectiveness of Machine Learning (ML) techniques in shifting construction risk management from a reactive approach to a proactive one. The study utilizes a Supervised Learning methodology, employing historical project data as training input, with the goal of predicting a binary outcome: whether a project will experience a schedule delay exceeding 10\%.

The model also successfully identified Design Changes/Revisions, Material Lead-Time Deviation, and Labor Productivity Fluctuation as the most critical predictors of delay risk. This research validates ML as a robust, data-driven tool for delivering timely, high-confidence early warnings, thereby enabling project managers to implement targeted mitigation strategies and enhance overall project predictability and control.

I. INTRODUCTION

Background: Construction Project Delays

The Civil Engineering and Construction (AEC) sector is a cornerstone of global economic development. However, the industry is persistently plagued by project delays, a global issue recognized as the most significant factor undermining project success. Delays are defined as the time overruns beyond the planned schedule date. Industry statistics frequently show that over 70\% of construction projects experience some form of schedule overrun, often leading to cost overruns averaging over 15\% of the initial budget. The impacts extend beyond finance, encompassing contractual disputes, loss of client trust, and reputational damage to the firm.

The complexity of modern construction projects—involving numerous stakeholders, dynamic site conditions, uncertain supply chains, and external regulatory influences—makes accurate forecasting challenging. Identifying potential delays early in the project lifecycle is not merely a benefit; it is a critical necessity for effective project control and risk mitigation.

Overview of Machine Learning (ML)

Machine Learning is a subset of Artificial Intelligence (AI) that focuses on developing algorithms that allow computer systems to learn from and make predictions or decisions based on data, rather than being explicitly programmed to perform the task.

In the context of civil engineering, ML leverages vast amounts of historical and real-time project data (e.g. resource consumption, weather logs, progress reports, contract details) to:

1. Identify intricate patterns and correlations among various project variables.

2. Forecast potential outcomes (e.g. delay or no delay, magnitude of cost overrun).

This data-driven approach allows for the creation of models that can adapt to changing conditions and provide insights far beyond the capability of traditional statistical methods.

II. PROBLEM STATEMENT

Traditional project management tools (like Critical Path Method and Earned Value Analysis) and simple statistical regression models often function as reactive tools, primarily tracking schedule variance after a delay has already occurred. This methodology suffers from two key deficiencies:

1. Inability to Capture Non-linearity: Construction data is high-dimensional and non-linear. The relationship between factors like "number of design changes" and "final delay" is often complex and non-proportional, which linear models cannot accurately represent.
2. Lagging Indicators: Decisions are based on past performance, not future risk.

III. METHODOLOGY

Overall Workflow
Project Creation ↓ Weekly Data Collection ↓ Feature Engineering ↓ Dataset Preparation ↓ Machine Learning Training ↓ Delay Prediction ↓ Risk Classification ↓ Result , Alerts and suggestions

3.1 Dataset Collection
Project data was collected:

Software and Tools Used

Components	Technology
Language	Python
FrontEnd	HTML, CSS
Backend	Python-Flask
Database	MYSQL
ML Library	Scikit-learn

3.2 Step-by-Step Execution Process
STEP 1 – Install Required Software
Install:

- Python 3.11+
- VS Code
- MySQL Server

STEP 2 – Install Python Libraries
Open terminal:
pip install flask pandas scikit-learn joblib mysql-connector-python matplotlib

STEP 3 – Project Folder Structure
construction_delay_prediction/

```
│
├── app.py
├── train_model.py
├── DummyData.csv
├── model/
│   └── delay_model.pkl
│
├── templates/
│   ├── home.html
│   ├── add_project.html
│   ├── weekly_report.html
│   └── result.html
```

STEP 4 – Train Machine Learning Model
Run:
python train_model.py
This will:

- Read dataset
- Train Random Forest model
- Save model as delay_model.pkl

Components	Technology
Language	Python
FrontEnd	HTML, CSS
Backend	Python-Flask
Database	MYSQL
ML Library	Scikit-learn

Parameter	Description
Project Id	Unique Id for each project
Project Name	Associated project name
Project Start Date	Project build starting date
Project End Date	Tentative project end date
Area Covered	Total area under which construction carried
Location	Dropdown for rural, urban or semi-urban type

STEP 5 – Create MySQL Database
Database Name
CREATE DATABASE construction_delay;

STEP 6 – Create Project Table

```
CREATE TABLE projects(
   id INT PRIMARY KEY AUTO_INCREMENT,
   project_id VARCHAR(20) unique,
   project_name VARCHAR(200),
   start_date DATE,
```

```
    end_date DATE,
    area FLOAT,
    location VARCHAR(50)
);

STEP 7 – Create Weekly Report Table
CREATE TABLE weekly_reports(
    id INT PRIMARY KEY AUTO_INCREMENT,
    project_id VARCHAR(20),
    week_date DATE,
    actual_progress FLOAT,
    labor_count FLOAT,
    material_delay FLOAT,
    design_changes FLOAT,

    FOREIGN KEY(project_id)
    REFERENCES projects(project_id)
);

STEP 8 – Configure Database in app.py
Add MySQL connection:
import mysql.connector
conn = mysql.connector.connect(
host="localhost",
user="root",
password="root",
database="construction_delay"
)
STEP 9 – Run Flask Application
python app.py
Open browser:
http://127.0.0.1:5000

STEP 10 – System Workflow
Home Page-Buttons
```

- Add Project
- Weekly Entry

Add Project Page
Store project details in database

Weekly Report Page
Site engineer enters: Weekly report form and saves in Database

Prediction Process
Data sent to ML model.
Model predicts:

- Delay probability
- Risk level

Result Page
Displays:

- Delay percentage
- Risk classification
- Recommendations/suggestions

IV. CONCLUSION

This project successfully demonstrates the application of Machine Learning for predicting construction project delays using weekly monitoring data.
The developed Random Forest based system effectively analyzes construction project parameters and predicts schedule risks with good accuracy.
The web-based implementation provides a practical solution for real-world construction monitoring.
The study proves that machine learning can significantly improve proactive project management and reduce delay-related risks in the construction industry.

REFERENCES

[1] S.O. Abioye et al.Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges Journal of Building Engineering (2021)
[2] AI-enabled organoids: Construction, analysis, and application Bioactive Materials (2024)
[3] W. Duan et al. Be alert to dangers: Collapse and avoidance strategies of platform ecosystems Journal of Business Research (2023)
[4] Applied artificial intelligence for predicting construction projects delay Machine Learning with Applications (2021)
[5] H. Erol et al.
An analytic network process model for risk quantification of mega construction projects Expert Systems with Applications (2022) H. Fan et al.
[6] Novel method for reliability optimization design based on rough set theory and hybrid surrogate model Computer Methods in Applied Mechanics and Engineering (2024)
[7] M.A. Ganaie et al.Ensemble deep learning: A review Engineering Applications of Artificial Intelligence (2022)
[8] Egwim et al 2021 Egwim C.N., et al Extraction of underlying factors causing construction projects delay in Nigeria