

# **DelaySense: Project Delay/Success Prediction**

Avish Kaushik (G49697765)

Yashika LNU (G21358006)

Nikhil Shinde (G41194064)

Ruthvik Tarang Goggi (G46248997)

CSCI 4364/6364 Machine Learning - Fall 2025

# The Challenge of Project Delays

## The Problem:

- Development projects often run months or years behind schedule
- Delays disrupt planning, staffing, budgeting, and mission timelines
- Teams rely on reactive monitoring — risks are identified only after delays occur
- Stakeholders need **predictive insight**, not post-mortems

## Traditional Approach Limitations:

- Manual assessments are subjective and inconsistent
- Simple “delayed / on time” labels are not actionable
- Hard to detect patterns across **1.5M+ historical projects**
- No ability to forecast **exact completion dates**

## Our Solution

- Machine learning models that predict **actual completion dates**, not just risk labels
- Early-warning signals for potential delays
- Ability to experiment with scenarios and understand what improves timelines

## Success Criteria

- Accurate predictions
- Clear risk assessment
- Interactive scenario testing
- Fast

# Datasets

**Two independent ML models: AidData & IEG**

## AidData Dataset

- ~630,000 development projects
- Operational and financial attributes
- Donor, sector, region, project cost, start/end dates
- Suitable for predicting **exact completion dates**

## IEG Dataset

- Evaluated World Bank projects
- Institutional and performance attributes
- Supervision quality, M&E ratings, lending instrument, approval FY
- Suitable for predicting **completion year**

## Why Two Models?

- Each dataset requires **its own features, its own modeling pipeline, and its own predictor**
- A single model would not generalize across both datasets
- Therefore, we trained **two independent ML models**

# Dataset 1- AidData

**Goal:** Predict *exact* project completion **date** and flag potential delays

## Key features:

- Donor, recipient country, region
- Sector / purpose classification
- Commitment amount & total project cost
- Start date, proposed end date, commitment date
- Implementing & financing agencies

## Outputs:

- Predicted completion **date**
- Schedule residual (proposed – model prediction)
- Delay-risk categories: **early\_closure**, **on\_schedule**, **delayed\_closure**, **significantly\_delayed**

## Link for the dataset:

<https://www.aiddata.org/data/aiddata-core-research-release-level-1-3-1>

# Dataset 2- IEG World Bank Project Data

**Goal:** Predict *exact* project completion **year** and flag potential delays

## Key features:

- World Bank region
- Practice group and global practice
- Agreement type
- Lending instrument type
- Country lending group
- Implementing agency
- Project name and approval fiscal year
- Quality ratings: quality at entry, quality of supervision, bank performance, M&E quality

## Outputs:

- Predicted completion **year**
- Deviation from expected timeline (delay signal)
- Model supports simplified delay interpretation due to year-only granularity
- Delay-risk categories: **early\_closure**, **on\_schedule**, **delayed\_closure**, **significantly\_delayed**

## Link for the Dataset:

<https://financesone.worldbank.org/ieg-world-bank-project-performance-ratings/DS00053>

# Model Training Pipeline

## Data Preparation

We preprocess each dataset separately — cleaning fields, engineering duration targets, encoding categorical variables, and standardizing date information.

## Model Training

Each dataset (AidData and IEG) is trained on **four models**:

- Linear Regression
- Random Forest
- XGBoost
- Neural Network

## Model Selection

For each dataset, we evaluate all four models and **choose the one with the lowest MAE**. This best-performing model becomes the final predictor.

## Result

Two optimized models — one for AidData, one for IEG — each selected objectively based on error performance.

# Risk Assessment with Statistical Methods

Once we identify the best-performing model and generate predicted completion dates (or years), we compute the schedule residuals and use them to assess delay risk.

These residuals feed into three complementary statistical methods.

## Method 1: Z-Score Analysis

- $z = (\text{residual} - \text{mean}) / \text{std}$
- Early:  $z < -1$  | On Schedule:  $|z| \leq 1$  | Delayed:  $1 < z \leq 2$  | Sig. Delayed:  $z > 2$

## Method 2: Percentile-Based

- Uses 20th, 80th, 95th percentiles from training data
- Distribution-free approach, captures actual patterns

## Method 3: IQR (Interquartile Range)

- $\text{IQR} = Q3 - Q1$ , uses  $1.5 \times \text{IQR}$  fences
- Robust to outliers, emphasizes extreme deviations.

Delay-risk categories: **early\_closure**, **on\_schedule**, **delayed\_closure**, **significantly\_delayed**

# Introducing the Dashboard

We wanted users to have the freedom to adjust project parameters and immediately see how those changes affect the predicted completion timeline and delay risk.

So, we built an **interactive dashboard** where users can tweak inputs, compare scenarios, and understand what factors help a project stay on schedule.

Dashboard Link - <https://delay-sense-ui.vercel.app/>

The screenshot displays the AidData Project Delay Risk Dashboard interface. At the top, there are two tabs: "AidData model" (selected) and "IEG Data model". Below the tabs, the title "AidData Project Delay Risk Dashboard" is centered, followed by a subtitle: "Powered by your AidData model end-date model + residual-based delay detection."

The dashboard is divided into several sections:

- GLOBAL DELAY RISK:** States that roughly 15–20% of projects are delayed or significantly delayed, with higher risk in small island states and regional programs.
- HIGH-RISK SECTORS:** Lists Rural development, social welfare, biodiversity, WASH, and complex health programs as showing the highest delay rates.
- SHOCK SENSITIVITY:** Notes that COVID-era and cross-country regional projects experienced elevated delays, especially in Africa and Oceania.

**AidData model – Project Risk Calculator:** A form for inputting project details. Fields include:

- Donor: United States
- Recipient country: Kenya
- Recipient region: Africa
- Sector / purpose name: Health
- Flow name: ODA Grants
- Commitment amount (constant USD): 5000000
- Total project cost (optional): 5500000
- Start date: 01/15/2020
- Proposed end date: 06/30/2023
- Commitment date (optional): 12/01/2019
- Implementing agency (optional): Ministry of Health
- Financing agency (optional): USAID
- Project title (optional): Kenya Health Systems Strengthening Project

**Predict end date & risk** button is at the bottom of the calculator section.

**Prediction & context:** This section displays the predicted completion date as **2020-03-07**. It also includes:

- Schedule tightness vs model:** Model expects completion around 2020-03-07. Your proposed end date implies a schedule residual of 1209.531177520752 days (proposed – model).
- Delay risk classifications:**
  - Z-score: significantly\_delayed
  - Percentile: significantly\_delayed
  - IQR: significantly\_delayed
- How to use this for decision-making:**
  - Compare the predicted completion date to the contractual end date to estimate buffer months.
  - For high-risk regions (e.g. small island states, African regional programs) consider additional supervision missions or contingency planning.
  - Combine this prediction with your per-donor / per-sector delay tables to prioritize where portfolio management attention is most needed.

# Model Performance Overview

## IEG Model (Completion Year Prediction)

We evaluated four models: Linear Regression, Random Forest, XGBoost, and a Deep Neural Network.

### Best model: Random Forest

MAE: **1.24 years**

RMSE: **1.70 years**

R<sup>2</sup>: **0.61**

**Error pattern:** Mean residual  $\approx -0.02$  years (almost unbiased)

### Interpretation

The model predicts completion timing within **~1.2 years** for a dataset where project durations vary widely.

## Aid Data Model (Completion Year Prediction)

We trained the same four models, but here the target is exact **end-date** (in days).

### Best model: XGBoost

MAE: **141 days** ( $\approx 4.6$  months)

RMSE: **307 days**

R<sup>2</sup>: **0.87**

**Error pattern:** Mean lateness  $\approx -0.5$  days  
(extremely well calibrated)

### Interpretation:

Because the dataset is much larger and more consistent (~630k projects), the model predicts with **high precision** — average error of only **4–5 months** on multi-year projects.

# Real World Applications

- **Pre-Commitment Risk Assessment**
  - Evaluate characteristics before approval, adjust design to mitigate delays
- **Portfolio Monitoring**
  - Prioritize supervision for high-risk projects, early warning system
- **Evidence-Based Planning**
  - Set realistic timelines based on similar projects
- **Stakeholder Communication**
  - Objective, quantitative risk assessments with transparent methodology

# Future Enhancements

Some of the future enhancements that we have are

- Advanced Models: LightGBM, neural networks for text
- Explainability: SHAP values, counterfactual analysis
- Real-Time Monitoring: Update predictions as projects progress
- Recommendation System: Suggest risk mitigation strategies