**Project Spending Start Date Prediction: A Machine Learning Approach**

## 🎯 Project Goal

We explored whether we could predict how long it takes for a project to begin spending after it officially starts. Accurately estimating this delay would help departments with financial planning, hiring, and operational readiness.

## 📊 Data & Method

We analyzed 360 past projects, using details such as funding amount, project duration, award type, and internal indicators like approval requirements.  
  
Four machine learning models were tested:  
- Linear Regression  
- Decision Tree  
- Random Forest  
- XGBoost  
  
Each model was evaluated for accuracy using standard measures (RMSE and MAE), and we applied careful data preparation and validation to ensure fairness.

## 📉 Results

- **None** of the models provided consistently **reliable prediction**s.  
- **Random Forest** performed the best but still had a typical error of about 2 months.  
- Most project characteristics did not strongly influence delay predictions.  
- The wide variation in spending delays seems to depend on **missing or untracked factors**, such as:  
 • Internal routing time  
 • Sponsor response delays  
 • Contract processing or compliance steps

## ⚠️ Key Takeaways

- Current **data is not sufficient** for accurate forecasting.  
- Predictions have **high error margins** and should not be used for planning at this time.  
- Delay drivers likely include **process-level variables that were not captured**.

## ✅ Recommendations

1. **Enhance data** collection by including internal workflow indicators (e.g., approvals, sponsor negotiation time).  
2. Consider modeling delay risk categories instead of exact days.  
3. **Revisit** modeling once a **richer dataset is available** — even a few hundred more records with better features can improve performance.