PREDICTING URBAN FLOOD VULNERABILITY THROUGH MACHINE LEARNING

Lauren Whittier | URP6931 AI in the Built Environment | Fall 2025

INTRODUCTION

Urban flooding poses a major threat to infrastructure, public safety, and economic stability in cities across the United States. This study aims to answer the question: Can machine learning models predict parcel-level flood vulnerability using built environment and environmental data with accuracy?

Goal: Build a **predictive model** that, given the geographical- environmental attributes of a property parcel, outputs the parcel's **Flood Probability.**

METHODS OVERVIEW

Input: numerical + categorical features
(Topography Drainage, Urbanization, climate variables, etc.)

Output: FloodProbability ∈ [0, 1] (continuous)

Baseline: simple Random Forest (100 estimators, default depth) trained on a 70/30 split achieved an R^2 of ≈ 0.60

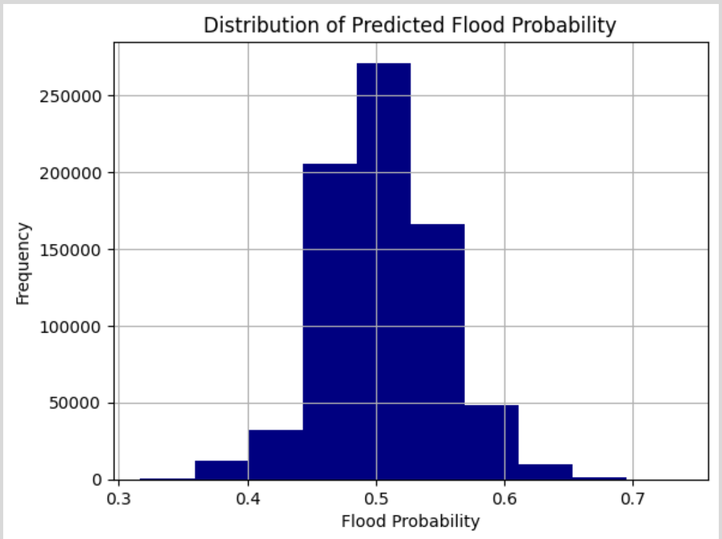
Models: XGBoost, CatBoost, LightGBM

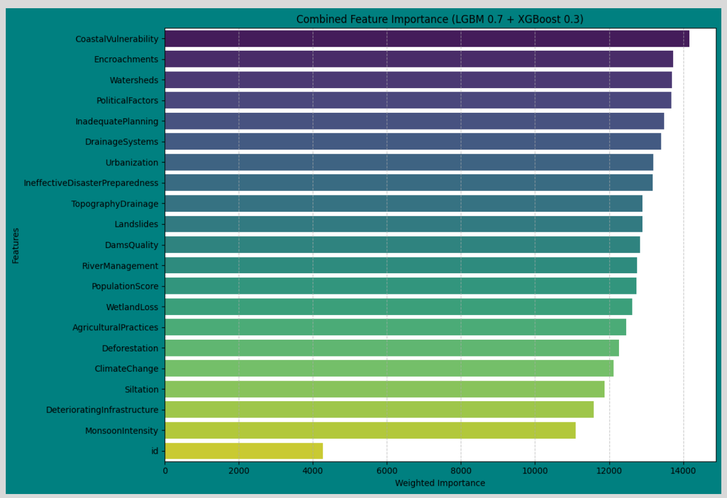
Evaluation Metric: coefficient of determination(R^2)

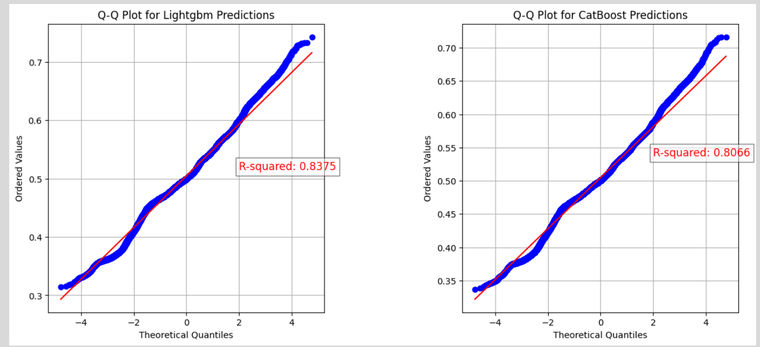
$$R^2 = \frac{\text{RegSS}}{\text{TSS}} = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$$

RESULTS

Model	CV R^2
XGBoost	0.8313
CatBoost	0.8066
LightGBM	0.8375
Ensemble	0.8381







DISCUSSION

An ensemble of LightGBM and XGBoost can predict parcel-level flood probability with an R^2 of 0.83, outperforming a simple Random Forest baseline.

Future steps include:

- **Category-to-number mapping** – assigning each categorical field a numeric code based on its empirical frequency or severity score
- **Spatial regularization** (Gaussian Processes or MRFs) to utilize spatial autocorrelation
- **Temporal validation** to ensure robustness across seasons
- **Risk-mapping implications**
- **Categorical Risk Scale:** probabilities illustrated as low to high risk

These enhancements are low-cost relative to the potential accuracy gains.

Overall, this MLM approach enables planners to allocate green-infrastructure resources at the parcel level rather than relying on coarse FEMA flood zones

REFERENCES

1. Walter Reade and Ashley Chow. Regression with a Flood Prediction Dataset. https://kaggle.com/competitions/playground-series-s4e5, 2024. Kaggle.
2. Satya Prakashshukl. Regression Flood Analysis. Notebook. https://www.kaggle.com/code/satyaprakashshukl/regression-flood-analysis/notebook, 2024. Kaggle.