Women in Data Science Worldwide | Charlotte @ Bank of America

# WiDS 2026 Datathon
# Introduction to Machine Learning

**Tej Gorde**

Analytics & model Development
Consumer & Small Business
12/05/2025

# The WiDS Datathon – What is it?

**What**: Global predictive analytics competition for applying data science skill to solve a critical social impact challenge.

**How**: WiDS provides well-curated, real-world datasets that are not readily available in the public domain.

**Who**: Aspiring data scientists, and those with experience. Both undergraduate and graduate college and university students.

**Why:** Access to unique datasets and challenges allow students to stand out to employers and research opportunities.

**RESOURCES PROVIDED:** Webinars and Office Hours | Community | Global Workshops | Guides and Tutorials

# Global Challenge 2026:
# Infrastructure-Disruption Forecasting

Women in Data Science Worldwide | Charlotte @ Bank of America

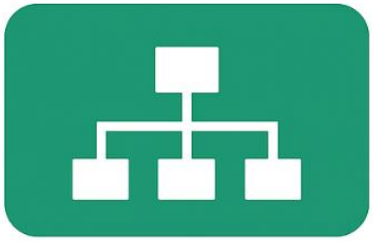
Prediction Challenge Problem Statement

## Infrastructure-Disruption Forecast

**Problem:** Participants in this challenge will build predictive models to estimate the probability of active wildfires intersecting any segment of infrastructure in the affected region.

**Background:** Wildfires now spread faster and burn hotter, putting power lines and other critical assets at daily risk. Early, location-specific forecasts let operators prevent blackouts and protect communities. Existing warning systems rarely deliver location-specific alerts early enough for targeted mitigation. When wildfires intersect such infrastructure, risks escalate significantly, intensifying fire behavior and causing widespread outages and public safety hazards.

# What Is Supervised & Unsupervised Learning?



Machine learning

| Supervised learning | |
|---|---|
| Classification model | Regression model |
| Support-vector machines | Linear regression |
| Discriminant analysis | Logistic regression |
| Naive bayes | SVR, GPR |
| Nearest neighbor | Ensemble methods |
| Neural networks | Decision |

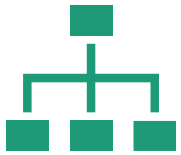
| Unsupervised learning |
|---|
| Clustering |
| Connectivity-based clustering |
| Distribution-based clustering |
| Density-based clustering |
| Centroid-bases clustering |

Machine learning methods can be broadly divided into **Supervised** and **Unsupervised** learning, depending on whether the data includes known outcomes.

# Why Supervised Learning Fits This Challenge

- Predict wildfire–infrastructure intersections (Classification)
- Forecast time until disruption (Regression)
- Identify at-risk communities (Ranking)

Supervised learning enables early warnings and equitable actions.



Classification – Will the fire hit a power line? → 0/1

Regression – How fast will the fire spread? → continuous value

Ranking – Which areas are most at risk? → score (0–1)

# Supervised Models to Try

Simple & Explainable: Logistic Regression, Decision Tree, Random Forest

Tabular: XGBoost, LightGBM, CatBoost

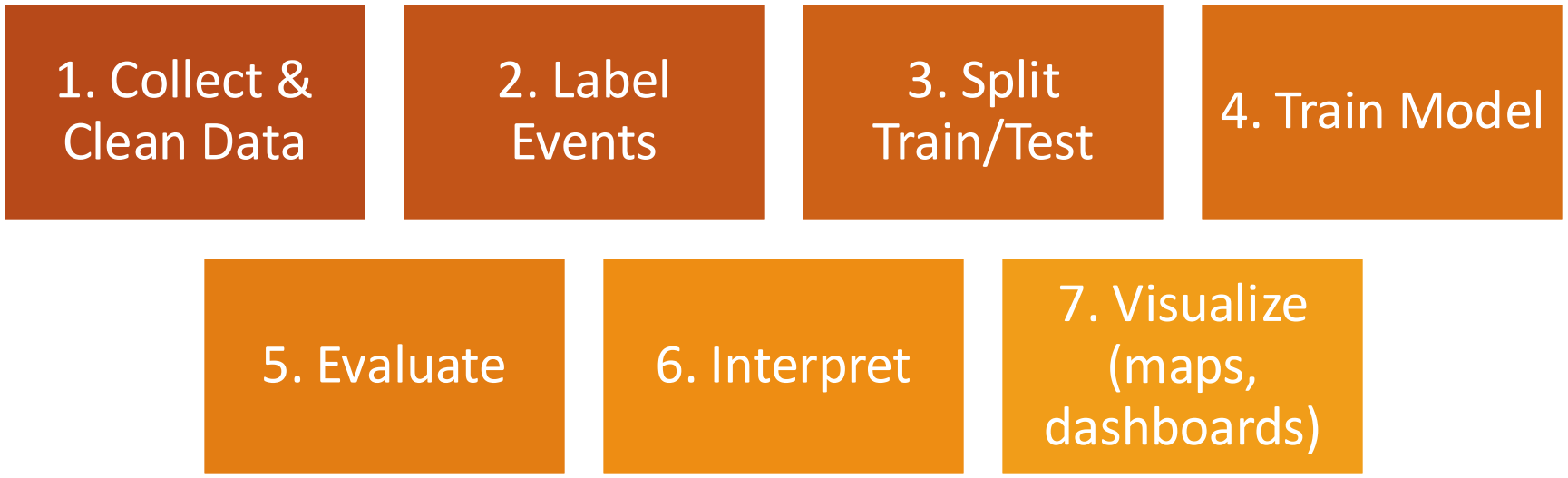
Geospatial/Advanced: CNNs, ConvLSTM, Graph Neural Networks

# Possible Input Features

Integrate geospatial, environmental, and social data:

- Fire behavior: area, spread rate, direction
- Weather: temperature, humidity, wind
- Terrain: slope, elevation, vegetation
- Infrastructure: distance to lines, roads
- Community: population, vulnerability index

# Building a Supervised Learning Model



# Example 1: Predict Infrastructure Disruption

Goal: Predict if wildfire hits a line in 48h.

Inputs: Wind, fire growth, terrain.

Model: Random Forest → Output: Probability (e.g., 0.82 = High Risk).

# Example 2: Equitable Wildfire Response

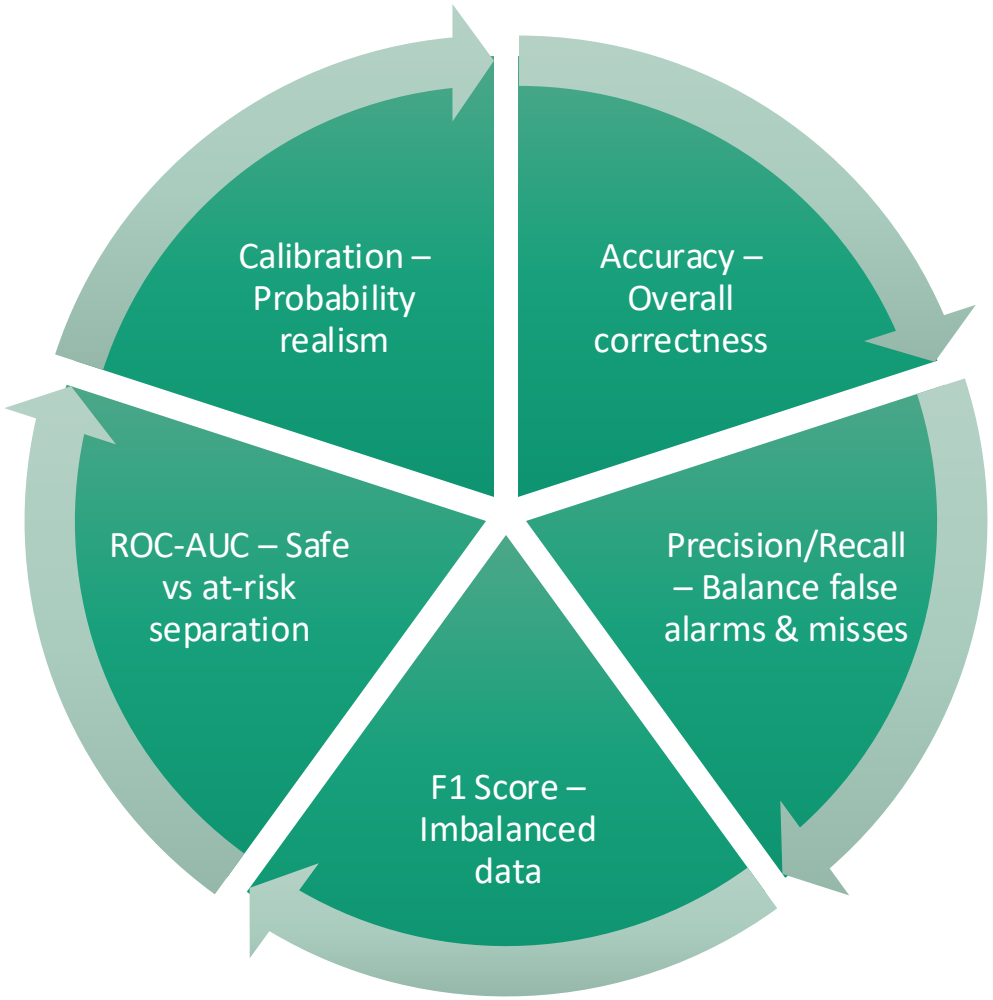
Goal: Predict delayed evacuation communities.

Inputs: Mobility, age, alerts, economy.

Model: Logistic Regression or Gradient Boosting.

Output: Probability of delay → prioritize outreach.

# Evaluating Model Performance



# Making It Human-Centered

Machine learning must serve people:

- Keep models explainable & trustworthy
- Inform real actions (shutoffs, evacuations)
- Present insights via dashboards

"Machine learning saves lives when it informs human decisions."

# Closing Message

"Be Bold. Be Rigorous. Be Human-Centered."

You're not just training models — you're training foresight.

Let data science protect communities and build resilience.