URBAN TREEHUGGERS

Team 4: Sean Liu, Chris Liding, Krithika Chockalingam, Ananya Kumar, Nikita Gupta, Tejaswini Ashok

Motivation & Introduction

Problem: Urban forests are vital for environmental quality and community well-being. However, there is limited research that integrates tree condition, environmental quality, and public health.
Importance: Our project analyzes the health of urban forests across US cities and evaluates how environmental conditions affect both tree health and human health, specifically in **Los Angeles**. Healthy trees reduce city temperatures, filter pollutants, and serve as gathering centers. By identifying current strengths and weaknesses, we will guide smart city planning and urban forestry policies, improving both environmental and human health outcomes.

Approaches

- **Random forest classifier** & **XGBoost classifier** to predict the **health of trees** across major US cities
  - Inputs: species, city, nativity, height, diameter, age
  - Outputs: tree condition (excellent, good, fair, poor, dying, dead)
  - Preprocessing: city averages for missing values; encoded categorical features
- **Random forest regressor**, **gradient boosting regressor**, & **deep neural network** to predict **public health risk** in Los Angeles
  - Input variables: weather & pollution data (temperature, humidity, AQI, wind, speed, UV index)
  - Preprocessing: RFECV & linear regression → top 8 features for inputs
- **Interactive Tableau visualization** to help users explore the data and models
  - State & city level density maps; drill down dashboards
  - Geospatial exploration of tree health & environmental metrics
- **Why they are effective:**
  - Connects tree health, environmental quality, & public health
  - Offers holistic insights for urban planning
- **Novelty of approaches:**
  - Leverages machine learning to study cross-domain relations between environment and health
  - Enables data-driven comparison of urban forestry and public health

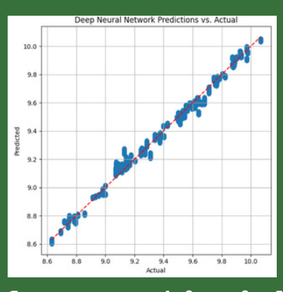
Data

- **Sources:**
  - **Downloaded** as **.csv** files from **Kaggle** (5M Trees & Urban Air Quality and Health Impact Analysis datasets)
- **Characteristics:**
  - **5M Trees:**
    - Total 1.24 GB across 65 .csv files (1 per city)
    - Features include city, observation date, species, longitude & latitude, dimensions, nativity, condition
  - **Urban Air Quality and Health Impact Analysis:**
    - Total 635 KB from 9/6/2024 to 9/20/2024
    - Features include date and time, max and min temperatures, humidity, precipitation, wind speed, visibility, air pressure, sunrise and sunset, heat index, UV index, and health risk score

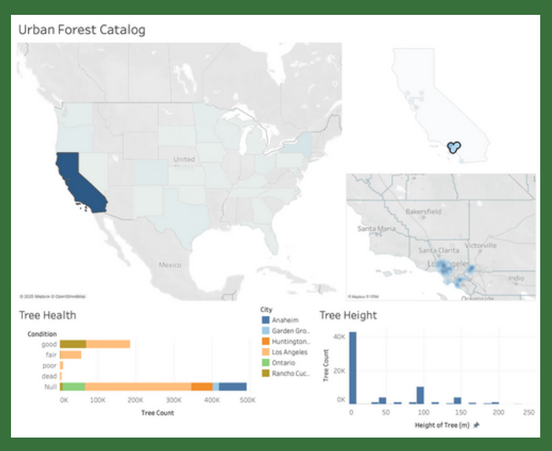
Experiments & Results

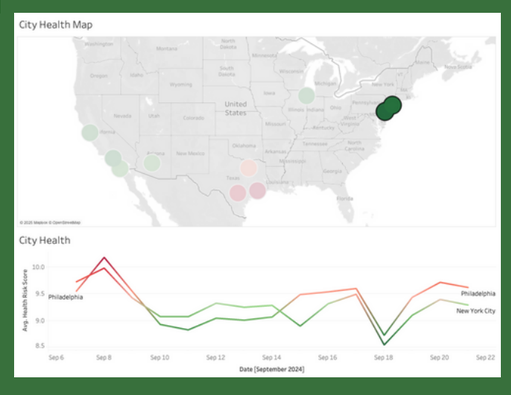
- **Tree Health Classification:**
  - Evaluated by **accuracy** & **feature importance**
  - Best model accuracy: 76%
  - Performs well on good trees; struggles with dead & dying trees
  - Top features: diameter, species, city, age
  - Los Angeles should remove dead trees and plant new trees targeting excellent condition.

- **Health Risk Prediction:**
  - Evaluated by **R^2** & **MSE**
  - R^2 = 0.99; MSE = 0.0012
  - Top predictors: temperature, humidity, UV index



- **Visualizations:**
  - Evaluated by **feedback** on **ease of use** and **insights derived**
  - Average time to assess health of trees near Georgia Tech: 76.5 seconds
  - Users found maps informative but slightly difficult to navigate





- **Comparison to other methods:**
  - Higher accuracy & lower error
  - Stronger generalizability & feature interpretability
  - More interactive & user-friendly