Project Proposal

Project Title: Wildfire Progression Prediction

Group ID: 12

Members: Leonardo Flores Gonzalez and Christina Vo

Project Summary

California wildfires are a recurring issue created due to rising temperatures, dry and hot environments, and high wind speed. In the past few years, wildfire seasons have started earlier and lasted longer than previously. Major wildfires have destroyed many homes, affecting thousands of people and forcing them to evacuate. Because of the increase of wildfires, management and prevention projects are developed and prioritized to protect communities from the public safety issue. For our data science project, we want to develop a machine learning framework designed to forecast the progression of California wildfires. Our primary objective is to build a live geographic visualization that maps current fire boundaries and predicts future spread by synthesizing multi-modal data, which includes CAL FIRE Historical Perimeters, NASA FIRMS (VIIRS) satellite hotspots, and USGS topographic elevation models. By training a PyTorch-based model on dynamic environmental features such as wind speed, humidity, and fuel density, we aim to project real-time movement across the landscapes. While our focus is on the ground spatial visualization, our system would also be architected to perform a core "barebones" evaluation: a regression-based ML model that estimates the total final acreage and burn size of an active fire based on initial ignition conditions. This approach ensures the project provides both insights and high-level impact assessments for disaster response.

Broader Impacts

Our model could be used to predict the size of wildfires which could allow for first responders to make tactical choices to minimize damage. Our goal would be that we can adjust in real time and make predictions based on the current situation. An example of this would be inputting wind speeds and current temperature to determine the location and the spread of a fire. By knowing this information, first responders are able to create a plan to contain the fire, ensuring that residential homes and buildings are not affected. Having a more advanced model to forecast these wildfires and adjust in real time can be beneficial. Preparation can be made early in order to protect people's lives and lessen the money in infrastructure damage.

Data Sources

Google deepmind Forecast model GITHUB: Link
NOAA Severe Storm Data: Link
Gridded Forecast Data: Link
ERA5: Link
HRRR: Link

WeatherBench: [Link](#)
MERRA-2: [Link](#)
CAL FIRE Historical Fire Perimeters: contains GIS boundaries for all major fires dating back to 1878: [Link](#)
Link to CAL FIRE Perimeters DATA 2020-2025 RECENT LARGE Fires: [Link](#)
Link to CAL FIRE Perimeters DATA ALL DATES ALL FIRES: [Link](#)
USGS Combined Wildland Fire Polygons: Massive national dataset that merges 40 different sources into one cleaned file: [Link](#)

Expected Major Findings
- Which environmental factors strongly influence the spread of wildfires the most?
- How would the accuracy change when looking at different time intervals such as 2 hours versus 24 hours?
- How far in advance can the model be able to predict the progression of the wildfire?
- How well will the model perform when analyzing different regions in California?