W210 | Summer 2019 | Section 4 | WildFire

- Nick Conidas
- Scott George
- Cassie Seney
- Yulia Zamriy
- Jay Zuniga

Problem Statement

Managing the risk of wildfires has been a critical problem in California for many years. Every fire season, California firefighters battle to contain multiple wildfires threatening lives and property. The 2018 wildfire season was the most destructive on record, resulting in at least 108 deaths and burning 1.6 million acres of land, the largest amount of burned acreage in a fire season. The cost to the California Department of Forestry and Fire Protection in 2018 exceeded 773 million USD[1]. Many of these fires were started by human factors and may be preventable.

The deadliest fire in California wildfire history was the Camp Fire, which started the morning of November 8, 2018, and burned a total of 153,336 acres, destroying 18,804 structures and resulting in 85 civilian fatalities and several firefighter injuries[2]. Fire investigators attributed the cause of this fire to electrical transmission lines owned and operated by Pacific Gas and Electric (PG&E) coming into contact with tinder dry vegetation during conditions of strong winds, low humidity and warm temperatures. This fire could have been prevented by shutting down power to the implicated transmission lines. Currently PG&E faces up to 50 billion USD in liabilities for this incident. Fire investigators have determined that PG&E was also responsible for 18 previous wildfires in 2017 [3].

These events highlight the need for power companies to make better informed decisions with respect to providing power to their customers or shutting down power to areas at high risk of wildfire. Utility companies face massive risks with much of their equipment located in dry, forested areas where the susceptibility to fire is being exacerbated by climate change. Our project will seek to provide a tool for both utility companies and local governments to visualize and evaluate the fire risk associated with transmission lines and substations and make good decisions as to whether the risk outweighs the impact of turning equipment off.

---

[1] "Tracking California's deadly fires", Julie Cart, 10/17/2018, https://calmatters.org/articles/california-wildfires-statistics-tracker/
[2] CAL FIRE News Release, May 15 2019 http://calfire.ca.gov/communications/downloads/newsreleases/2019/CampFire_Cause.pdf
[3] LA Times. https://www.latimes.com/local/lanow/la-me-ln-pge-judge-wildfires-20190109-story.html

The absence of electrical power may have adverse consequences to some public services such as hospitals. Residents without power may suffer heat related complaints during hot summer weather and communication of fire related evacuations may also be impacted. Making the right decision is challenging, which makes our data service more valuable.

"Power shutoffs are more controversial than traditional fire mitigation measures such as vegetation management, which goes a long way toward explaining why utilities have avoided them. But shutoffs are a more surefire way to prevent fires, which is why state officials are preparing for utilities to use them more often." [3]

Proceedings within the California Public Utilities Commission (CPUC) are currently revising and developing rules for utilities to implement safety-related de-energization, called Public Safety Power Shut-offs (PSPS).

Our effort is based on several assumptions:

- Wildfire can be reliably predicted using open data in a couple of days
- Turning off power will have a meaningful impact
- Power can be restored easily after being shut off
- The power grid may compensate to provide power when a transmission line is disabled
- The wildfire problem will remain at least as severe or get worse
- The wildfire prevention technology will not improve drastically in near future
- It will take significant time to implement more effective wildfire prevention methods that are being proposed by research institutions [4]
- Utility companies are reluctant to invest in infrastructure that is more resilient to wildfires
- Utility companies are liable for consequences

Value Proposition

Utility companies must balance the need to provide power to their customers against the risk of wildfire in areas where transmission lines are present. Protecting lives, property and forests while also protecting utility companies from liability can be achieved through better analysis of the data. Our product would combine data from weather, vegetation and historical records of wildfires to make good short-term predictions of wildfire risk and enable utility companies to make better informed decisions whether to shut down power in areas where utility equipment is deployed.

Our product goes beyond traditional fire danger ratings by incorporating geographically specific weather forecasts and information about local utility infrastructure. It will model wildfire probability and provide a greater level of confidence and insight. Incorporating utility data such

[3] See California looks to electricity shutoffs as a faster, cheaper wildfire solution
[4] See California's Fourth Assessment: Climate Change Summary Report

as the voltage of power lines and age of power stations, in addition to allowing utility companies to incorporate proprietary information into the evaluation of risk, helps improve decision making. For making complex and high risk decisions, these product features offer utility companies better predictions and data than has been available to them in the past -- all in one place.

The product will implement a dashboard with a map covering the California geographical area. Layers will display utility infrastructure and wildfire risk with confidence intervals. The dashboard will be interactive allowing customers to use their proprietary knowledge of their equipment to adjust risk, and obtain insight into the factors contributing to risk.

As a stretch goal, we would like to be able to reflect the impact of shutting off the power, by identifying the scope of residential areas affected and priority areas such as hospitals which may be impacted.

Use Case / Customer Segment

While our primary customers are investor-owned utilities, we also anticipate the product being useful to local governments, thereby providing oversight and accountability.

Our data provides transmission line and power utility records for about 50 utility companies including some of the largest operators in California: Pacific Gas and Electric Company (PG&E) (28% market share[5]) and Southern California Edison Company (SCE) (29%).

We will be focusing on utility companies as our primary customer because they were found directly responsible for some of the largest fires in Northern California and their decisions have significant impact in the area.

We also anticipate the dashboard being useful to government agencies including:
- California Public Utilities Commission (CPUC)
- California Department of Forestry and Fire Protection (Cal Fire)

The Use Case is decision support for power shutoffs. Specifically, it will provide local confidence intervals by area (potentially, zip codes), which update in time based on weather, merged with utility infrastructure information.

To perform customer validation, we will work with people who have experience in the utility industry and who are involved in fire prevention and containment efforts with state/local government. We have identified a couple of contacts with experience working for/with power utilities and we are reaching out to them with questions to validate our approach.

[5] Measured from Total Usage (MM kWh) in California. Source: California Energy Commision

MVP

Our minimal viable product is a dashboard that will show the fire risk in areas where utility equipment is deployed. The initial dashboard can be in tabular form and the fire risk can be described using a numerical rating. The MVP will initially use the minimum subset of the available data, most likely weather data, utility company equipment locations and California terrain information. A simple learning model will also be used, such as logistic regression, with historical wildfire data used in training the model.

On subsequent iterations, we plan to add more data into the model - moisture data and additional risk factors. We will also use interactive visualizations to increase engagement.

We believe our focus on the risks posed by utility company equipment, our assessment of near term conditions and the unique blend of our data will provide a value add over current solutions that are either focused on different wildfire risk vectors or over long term trends. While larger utility companies have tools themselves that they haven't publicly detailed, we believe the way we use data will be beneficial to them. We also believe there will be benefit for smaller utility companies, which might not have dedicated resources.

Data sets

The two different methods for calculating wildfire risk are the Fire Weather Index System and the National Fire Danger Ratings. Both rely on a combination of weather and fuel moisture ratings.

> Synoptic Weather Data API
> Public Web Service to query surface weather stations, free or $5/mth depending on volume of requests. This data set is used by the National Weather Service to produce its own fire weather visualization https://www.weather.gov/fire/. Provides current and historical weather data including precipitation, temperature and wind speed for a set of stations over a time span.
>
> US Forest Services National Fuel Moisture Database
> Public freely downloadable fuel moisture data. Multiple years of historical data are available. The data is retrieved via an interactive form.
>
> Sentinel-2 ESA satellite data on vegetation with update period of 5 days, 10 m spatial resolution
> NAIP Imagery National Agriculture Imagery Program
>
> State of California Wildfire History

Public freely available historical data about wildfires in California. Provides start date, latitude and longitude of wildfires. Text format on web pages would require some web scraping to extract.
As an alternative there is data for just 2018 provided in spreadsheet format here.

Power transmission lines and station data is freely available for visual mapping.

California Electric Transmission Line
Geospatial data for map visualization of California Energy Commission (CEC) Electric Transmission Lines. Includes voltage attributes for the transmission line and can be filtered by owner (PG&E). Provided in formats compatible with ArcGIS and Google Earth. Also available as a downloadable geodatabase file from the State of California Data Catalog.

California Electric Substation
Geospatial data for map visualization of California Energy Commission (CEC) Electric Substations. Can be filtered by owner. Provided in formats compatible with ArcGIS and Google Earth.

California Power Plant
Geospatial data for map visualization of California Energy Commission (CEC) Power Plants. Includes year brought online and can be filtered by owner. Provided in formats compatible with ArcGIS and Google Earth.

Project Management

Jay Zuniga has been appointed to the role of Project Manager. The team will support his efforts to ensure that the project is completed on time and project milestones are met. The role of each member is still being worked out. We expect to have project milestones by next week.

While the team will continuously work together on our slack channel (#w210-su19-wildfire), we have two weekly meetings: our main meeting happens on Sunday 2-3pm and we have a Tuesday check-in that we'll use as needed. Each team member expects to contribute 10-20 hours a week toward the project as agreed in our Team Process Agreement.

We do plan to implement scrum methods including dividing our work into sprints and we will use Asana as our project management tool.

Technical Approach

Exploratory data analysis will be distributed among team members who will have the choice of using R or Python to look through the characteristics of our data.

We will be cleansing and transforming the data using Python pandas. Python will also be used to glue together the various procedures for preparing the data, converting into a form that we will use for learning. For Machine Learning models, we plan to use sci-kit.

We intend to load the data into GCP BigQuery.

This problem can be structured as a supervised binary classification problem. The model can take weather and vegetation moisture data for each geospatial zone as input and return the probability of fire as its output. This structure will allow us to leverage algorithms like: logistic regression, decision trees, tree ensembles and deep learning architectures such as feed forward, convolutional and recurrent neural networks as determined by the data. The output probabilities by geospatial zone will nicely support our goal of creating a map based dashboard displaying fire risk.

A first major technical challenge will be the integration of disparate data sets. Finding the optimal way to combine the various datasets will be difficult. Each dataset will likely be divided into geospatial regions of varying sizes. Furthermore, time series datasets are not guaranteed to be based on the same temporal frequency.

A second major technical challenge will be the fact that fires are relatively rare. This will essentially be an anomaly detection task with heavily skewed class labels (aka the label no fire is far more common than the label fire). In order to combat this label skew, it will be critical to engineer features capable of discerning signal from noise.

For visualization, we will be using Tableau on top of BigQuery and the dashboards themselves will be available through Tableau Online.