Fire Guys

An Analysis of when, where, and why certain fires occur

Guys: Toly Brevnov, Kolya Illarionov, Maximo Moyer, Chotoo Amin

Project Goals

Inspired by a volunteer firefighter in our group, our goal is to analyze data to help fire departments better allocate resources and prepare for the most common types of fires. Our project thus aims to give a breakdown of where the most fires occur, eyeball if seasons might influence the frequency of different types of fires, and most importantly <u>if a fire will occur in a given zip code, predict what type of fire will occur based on the day's weather attributes.</u>

Data

Our data is the result of the combination of two datasets. The first comes from FEMA's National Fire Incident Reporting System and originally contained over 5.5 million incidents reported through calls to 911. However, we narrowed this dataset down to about 600,000 records that were specifically fire-related incidents in the year 2018. The second comes from the VisualCrossing weather API which we queried based on location (zip code) of the fire incidents using a python script. Finally, we joined these two datasets on zip code so that we were left with a single table where each row contained data on one fire incident (date, location, incident type, etc.) as well as data on the weather in the specified zip code where the fire occurred (temperature, precipitation, wind speed, etc.).

Methodology

Why: Trained and tested a decision tree with depth 6 (tweaked depth to optimize testing results), and a Bernoulli Naive Bayes prediction model using 11 weather attributes. Data for prediction model only included fires that fell into the 3 most common fire types that made up 90% of all fires: vehicle, structure, and natural vegetation. Used a 80/20 train test split to ensure unbiased training.

When: Grouped data by # of fires per season, for each of 3 most common fire types. Displayed Bar Chart showing showing seasonal counts of fires by type.

Where: Heat Map showing fires per capita by (available) zip code in the United States.

Why Fires Occur: Decision Tree Depth-3 Viz, Decision Tree (Depth-6) and Naive Bayes Prediction Results and Discussion

Note: Color and boldness of node determined by relative frequency of most common fire type.

Relative Humidity <= 63.785
43127
[8705, 22192, 12230]
Structure

-Attribute Split on (left node is true, right node false)
-Data points at node
-Data points remaining by fire type (alphabetical)
-Most Common Fire Type at Node

Temperature <= 37.45
11775
[4632, 4653, 2490]
Structure

Cloud Cover <= 67.85
31352
[4073, 17539, 9740]
Structure

Snow Depth <= 0.07
2862
[671, 1556, 635]
Structure

Cloud Cover <= 60.75
8913
[3961, 3097, 1855]
Natural Vegetation

Temperature <= 39.85
14305
[2670, 7368, 4267]
Structure

Temperature <= 41.35
17047
[1403, 10171, 5473]
Structure

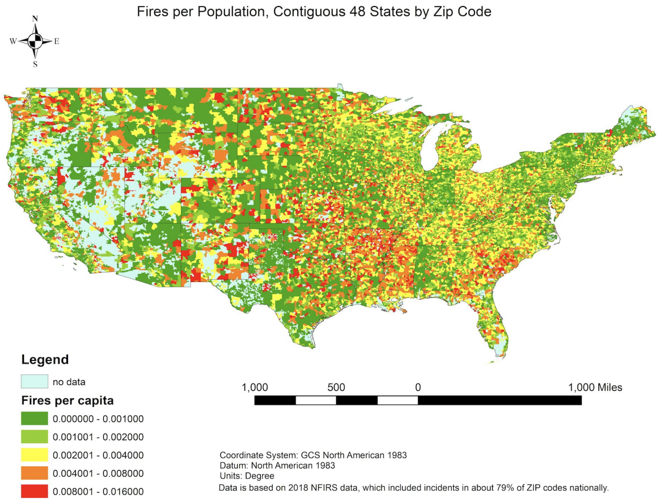
Decision Tree	Precision	Recall	F1	Acc
Nat. Veg.	0.49	0.44	0.46	
Strucutre	0.56	0.89	0.69	
Vehicle	0.54	0	0	
Overall	0.54	0.55	0.45	0.55

Naïve Bayes	Precision	Recall	F1	Acc
Nat. Veg.	0.37	0.67	0.48	
Strucutre	0.59	0.71	0.64	
Vehicle	0.34	0.02	0.05	
Oveall	0.47	0.5	0.44	0.5

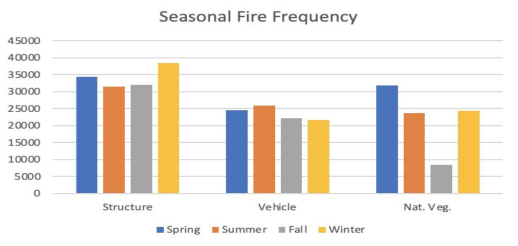
<u>Prediction Results (Naive Bayes and Decision Tree)</u>

- Training Attributes: Temperature, Wind Chill, Precipitation, Snow Depth, Wind Speed, Wind Gust, Visibility, Cloud Cover, Relative Humidity, Weather Type,Conditions
- Vehicle Fire Recall and resulting F1 are very poor
  - Perhaps due to vehicle being least represented fire type at 20%, relatively more false negatives
- Structure recall higher than precision
  - Perhaps due to structure being most represented fire type at 50%, relatively more false positives
- Both models achieved prediction accuracy much higher than chance
  - Both had balanced overall Recall and Precision
  - Naive Bayes better F1 on vehicle and nat. Veg
    - May have underperformed overall... violation of independence assumption?
  - **Decision tree more accurate across all measures in overall results**
  - **Only Decision tree outperformed always guessing most common fire type: Only useful model in practice**
- We showed the first three levels of the decision tree to illustrate the most important attributes in discerning fire type are relative humidity, cloud cover, temperature and snow depth

Where Fires Occur: US Heat Map of Fire Frequency



When Fires Occur: Fire Type Count by Season Bar Chart and Discussion



<u>Seasonal Results</u>

- Structure
  - Fairly equitable dist.
  - More in winter when people have to use indoor heaters?
- Vehicle
  - Fairly equitable dist.
  - Most in summer when engines overheat?
- Natural Vegetation
  - Clearly, far fewer incidents in the fall maybe due to wildfires starting in summer and continuing into fall?
  - Most in spring due to increase in tree, brush, and grass fires (all natural vegetation)

Challenges, Conclusions, and Next Steps

<u>Challenges</u>

- Gathering the data into one table, having to use two different data sources
- Missing weather data points, given different weather data gathering for different zip codes
- Having to discretize and categorize the data for ML models

<u>Conclusions</u>

- Both models better than chance, only decision tree better than guessing most common fire type
- Decision tree viz indicates that low-humidity and high-temperature conditions lead to natural vegetation fires
- Models were heavily influenced by frequency of fire type in the data
- Naive Bayes performed better on more fire types, but decision tree more accurate overall
- Frequency of fires seems to vary by region, but different reporting incentive in states makes results very unreliable
- Does appear to be some relations between season and frequency of fire, especially among natural vegetation

<u>Next Steps</u>

- Seasonal Fire: Chi-Squared tests if there is a relationship between fire type and season, or difference of means t tests to see which seasons have more or less fires of a given type (would need multiple years of data for this)
- ML models : Train on data with equal dist. of fire types (vehicle, structure, veg) to attempt to avoid difference in recall/precision and see if perhaps Naive Bayes would be the better classifier on this data