FIRE GUYS

Goal

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 and relevant for this abstract, if a fire will occur in a given zip code, predict what type of fire will occur based on the day's weather attributes.

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 VissualCrossing 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 information 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, Wind Chill, Precipitation, Snow Depth, Wind Speed, Wind Gust, Visibility, Cloud Cover, Relative Humidity, Weather Type, Conditions).

Model and Evaluation Setup

The three most common fire incident types are Structure (45%), Natural Vegetation (20%), and Vehicle (17%). These make up over 80% of all fires, and the remaining fires are split among another four categories. To remove noise from our prediction and prioritize, we decided to only focus on predicting the three most common types of fires (the remaining fire-types are highly specialized and significantly rarer). After dropping all incomplete rows with incomplete weather data, this left us with around 50,000 records, which we split into training and testing data using an 80/20 train/test split. We would train our models on the training data and record both training and testing performance to evaluate the model. Testing performance was especially useful in detecting overfitting (particularly when tuning our max_depth hyperparameter in our decision tree model to ensure test results were as close to train results as possible).

RESULTS and ANALYSIS

Claim #1: Our models both predict the type of fire based on weather factors considerably better than random chance and only the decision tree performs slightly better than guessing the most common fire type.

Support for Claim #1: Tables 1 and 2 are decision tree (left) and Naive Bayes(right) test metrics

	precision	recall	f1-score	support
Natural_Vegetation	0.49	0.44	0.46	2173
Structure	0.56	0.89	0.69	5583
Vehicle	0.54	0.00	0.00	3026
accuracy			0.55	10782
macro avg	0.53	0.44	0.39	10782
weighted avg	0.54	0.55	0.45	10782

	precision	recall	f1-score	support
Natural_Vegetation	0.37	0.67	0.48	2118
Structure	0.59	0.71	0.64	5565
Vehicle	0.34	0.02	0.05	3099
accuracy			0.50	10782
macro avg	0.43	0.47	0.39	10782
weighted avg	0.47	0.50	0.44	10782

Here, the accuracy represents the proportion of total predictions where our model correctly predicts the type of fire incident based on attributes of the weather. This table shows that the accuracy for our decision tree and naive bayes models are 0.55 and .5 respectively. This is considerably stronger than the 0.33 number that would represent random chance. However, only our decision tree model accuracy was slightly better than .5, while our naive bayes is exactly .5, which is the frequency of the most common fire type in our training data (structure fires). This suggests perhaps only our decision tree model is reasonably useful in a real life setting.
Claim #2: The models are heavily influenced by the relative frequency of the fire types in the data, leading to discrepancies in recall performance, especially on vehicle fires.
Support for Claim #2: The tables supporting Claim #1 also support Claim #2. In both models a similar pattern in the recall values emerged. Focusing on table 1 (decision tree), the more extreme case of the recall discrepancy, vehicle, natural vegetation, and structure fire recall were .00, .44, and .89 respectively. Vehicle, natural vegetation and structure fires respectively represented about 20%, 30%, and 50% of the data we trained/tested on. Thus, it seems that a higher frequency in the dataset leads to higher recall. This was further reflected in the largest depth visualization (in our visualization deliverable) of our decision leaf nodes which were more likely to have labels from the two most frequent fire types. As a result, there were many leaf nodes for structure, less for natural vegetation, and even less for vehicle fires. This made it such that there was a very low false negative rate for the most common fire type (structure), a medium false negative rate for the second most common type (natural vegetation) and an abysmal false negative rate for vehicle fires. This leads to the discrepancies in recall, (and therefore resulting F1 score since precision were all fairly close). Although the relative frequency of fire types was expected to show up in how many predictions were of a given fire type for each model, the severity of underperformance regarding false negatives on vehicles especially was surprising.
Claim #3: The Naive Bayes Model performed better across all fire types, but the Decision tree performed better on the overall data.
The tables supporting Claim #1 also support Claim #3. We can see that the F1 scores for the Naive Bayes outperforms the decision tree F1 scores on Natural Vegetation and Vehicle Fire types. Meaning, based on the F1 metric, Naive Bayes does a better job predicting if natural vegetation or vehicle fires will occur. However, the decision tree outperforms on F1 score for the most common structure fires. Although F1 is not a direct accuracy metric, and is a harmonic mean of recall and precision, the decision tree having a higher F1 for the more common fire type (meaning it performed better on the majority of the data) helped contribute to a higher overall accuracy for decision over Naive Bayes by .05. Thus, our decision tree model was more accurate overall, but perhaps if our dataset was more balanced between fire types, naive bayes might have achieved a higher overall accuracy as it does a better job of having equitable predictive success across attributes.