# Predicting Wildfire Ignition Using Weather and Ecological Data


**Zoe A. Bean**
Class of 2021
Mount Holyoke College
bean22z@mtholyoke.edu

**Anna Pickett**
Class of 2022
Mount Holyoke College
picke23a@mtholyoke.edu

**Fahmida Rafa**
Class of 2021
Mount Holyoke College
rafa22f@mtholyoke.edu



## Abstract

Our project used machine learning methods to predict the likelihood of a fire occurring in the near future based on local weather and geographic data. We used multiclass classification for our predictions, and our model had 76.6 percent accuracy on our test set. While our model was unable to predict if a fire was likely to start, it had some accuracy in predicting whether or not a fire was currently burning.


## 1 Introduction

There has been an alarming increase in the number of natural disasters the United States faces each year. Tornadoes, hurricanes, even wildfires are rapidly becoming more common as climate change worsens. While tornadoes and hurricanes tend to be more readily predictable, it is still difficult to predict the behavior of wildfires. In our project, we sought to solve that problem by creating a model that would predict when and where these wildfires will start, based on weather and geographical data. We have decided that since California faces a significant number of the wildfires in the US, we would investigate the Angeles National Forest, an area of high fire risk in California that also has a large population nearby.

## 2 Related work

In our research, we saw widespread efforts to use machine learning to make more accurate fire predictions, especially to predict how large a small fire, once set, will grow. As fire seasons in fire-prone areas grow worse as a result of global warming, accurate fire modeling allows fire protection agencies to allocate their resources more efficiently.

### 2.1 Alaskan Investigation

Previous research into fire prevention in Alaska used spectroradiometer data to find data on vapor pressure deficit and the fraction of spruce cover near the ignition point in order to predict the final size of a fire.[1] They compared a variety of possible data sets to find the input data that most accurately predicted final fire size and found that both of their most effective pieces of data could be gathered relatively easily from satellite data, which is important because information such as fuel moisture, which is also commonly used to predict favorable fire conditions, is much more labor intensive to gather.

[1] Coffield, Shane R., Casey A. Graff, and Yang Chen. "Machine learning to predict final fire size at the time of ignition."

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

### 2.2 Moroccan Investigation

Further research on fire prediction done in Morocco used Big Data, Remote Sensing and Data Mining algorithms to take data from satellite images in order to predict fire risk. They successfully used easy to collect satellite data to create an extremely accurate fire prediction model. Their work modeling fire risk from satellite data allowed them to train their model over large swaths of land.

### 2.3 Australian Investigation

Other research has looked into predicting where fire will not spread in Australia which can impact the behavior of local wildlife.[2] The researchers used machine learning methods to determine which variables were most important in determining the areas that did not burn. As with previous research discussed, machine learning methods have frequently been used not just to create general predictive models, but also to determine what data is the most helpful to collect in order to predict fire behavior.

## 3 Datasets

We worked with three different types of datasets- weather, ecological geographic, and fire records.

### 3.1 Weather Data

The weather dataset was downloaded from MesoWest Data[3] , where we downloaded the data from three different weather stations. These stations were: the Burbank-Bob Hope Airport(KBUR), the El Monte weather station(KEMT), and the Brackett Field Airport(KPOC). For each station, we downloaded three years: 2018, 2019, and 2020. Each station was assigned an area within our main area of interest to cover.

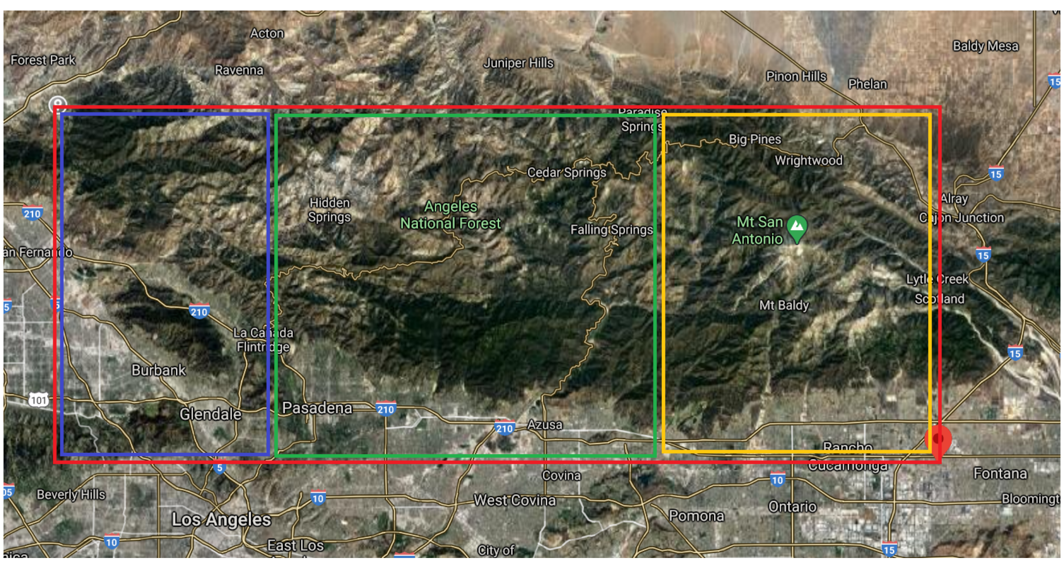


Figure 1: The red outlines our overall area of interest, the blue the area covered by KBUR, the green the area covered by KEMT, and the orange covered by KPOC

[2]Collins, Luke, Andrew F. Bennett, Steve W. J. Leonard, and Trent D. Penman. "Wildfire refugia in forests: Severe fire weather and drought mute the influence of topography and fuel age."

[3]University of Utah Department of Atmospheric Sciences.

### 3.2 Ecological Geographic

For our ecological geographic data we used data from the National Fuel Moisture Database[4]. This dataset contained the location, the plant type used at that location, and the date the data was calculated on. Using the locations given, the data was divided into the station areas and applied to the weather station data using the dates in this dataset.The Fuel type was assigned by area, since we assume that the vegetation remains mostly the same in the same area, even if the fuel moisture percentage changes.

### 3.3 Fire Records and Final Dataset

For the recorded fire data, we used data from Cal Fire[5]in Los Angeles County. This data held starting and ending dates and times for many fires, as well as their geographic location. Those locations were assigned station areas. In the combined weather and geographic data set, the fire status was added by inputting 'START' on the hour the fire started and the four hours preceding it, 'FIRE' while the fire was ongoing, and 'NO _FIRE' when there was no fire. In the end, the overall dataset we created had 9 categories and looked like this:

| Station_ID | Date | Time | air_temp_ | relative_humidity_ | wind_speed_ | Fuel | veg_humidity_percent | fire_state |
|---|---|---|---|---|---|---|---|---|
| KBUR | 9/3/2018 | 13:00 | 66.2 | 93.94 | 4.61 | Chamise | 50 | FIRE |
| KBUR | 9/3/2018 | 14:00 | 66.2 | 93.94 | 4.61 | Chamise | 50 | FIRE |
| KBUR | 9/3/2018 | 15:00 | 66.2 | 88.2 | 4.61 | Chamise | 50 | FIRE |
| KBUR | 9/3/2018 | 16:00 | 68 | 82.89 | 6.91 | Chamise | 50 | FIRE |
| KBUR | 9/3/2018 | 17:00 | 69.8 | 77.94 | 3.44 | Chamise | 50 | FIRE |
| KBUR | 9/3/2018 | 18:00 | 71.6 | 73.32 | 3.44 | Chamise | 50 | FIRE |
| KBUR | 9/3/2018 | 19:00 | 77 | 61.21 | 6.91 | Chamise | 50 | FIRE |
| KBUR | 9/3/2018 | 20:00 | 78.8 | 57.69 | 6.91 | Chamise | 50 | FIRE |
| KBUR | 9/3/2018 | 21:00 | 80.6 | 57.94 | 8.05 | Chamise | 50 | FIRE |
| KBUR | 9/3/2018 | 22:00 | 82.4 | 54.65 | 4.61 | Chamise | 50 | FIRE |
| KBUR | 9/3/2018 | 23:00 | 78.8 | 61.44 | 12.66 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 0:00 | 77 | 65.2 | 12.66 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 1:00 | 75.2 | 69.21 | 10.36 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 2:00 | 71.6 | 73.32 | 11.5 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 3:00 | 68 | 88.29 | 9.22 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 4:00 | 66.2 | 93.94 | 5.75 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 5:00 | 66.2 | 93.94 | 5.75 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 6:00 | 66.2 | 93.94 | 5.75 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 7:00 | 66.2 | 93.94 | 4.61 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 8:00 | 66.2 | 93.94 | 4.61 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 9:00 | 66.2 | 93.94 | 0 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 10:00 | 66.2 | 93.94 | 6.91 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 11:00 | 66.2 | 93.94 | 3.44 | Chamise | 50 | FIRE |
| KBUR | 9/4/2018 | 12:00 | 68 | 88.29 | 3.44 | Chamise | 50 | FIRE |

Figure 2: An excerpt of our data

With veg_humidity_percent and Fuel containing the data extracted from the Fuel Moisture Database, and fire_state being the desired Y output derived from our fire data set that we would like the machine learning algorithm to learn. Station_ID is the category that records what weather station recorded the weather data, and is also the location category for the other two datasets. The Date and Time categories were only useful in order to create the dataset, and were removed from the data input into the model so as to not interfere with the output. The categories of data that were from the weather set were air_temp, which is in Farenheight, relative_humidity, which is a percentage, and wind_speed, which is in miles per hour.

[4]USFS Wildland Fire Assessment System.
[5]CA.gov.

# 4 Methods

We used multiclass classification for our model. We split our data into no fire, start, and fire (0,1 and 2 for our input into the model) where entries where there was no fire at that time were labeled no fire, entries in the hours before a fire started were labeled start and entries during which there was an active fire were labeled FIRE. There was some overlap between the FIRE and START labels as sometimes one fire started while another was still burning and in those instances we recorded those entries as START. We chose multiclass classification because there were a distinct set of possibilities for each entry of data: that there was no fire, that there would be a fire shortly or that there was an active fire. Other methods such as linear regression or binary classification didn't fit what we wanted from our data prediction, so multiclass classification was the best fit.

# 5 Results and evaluation

## 5.1 Accuracy

When One versus All multiclass classification is applied to our data, we get a training accuracy of approximately 74 percent and a similar testing accuracy. This means that our algorithm will be accurate almost 74 percent of the time when categorizing input as no fire, fire, or start. The data does not do well with being normalized, as when normalized the error present increases, as shown:

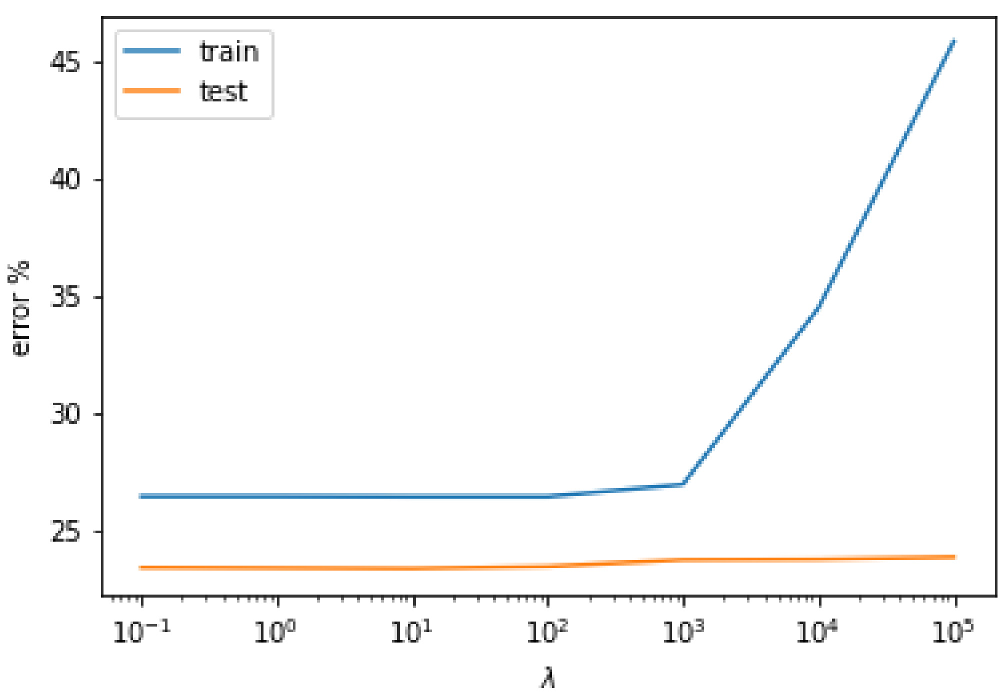


Figure 3: Error increases with normalization

Which means that normalizing the code causes our model to be underfitting. We evaluated our results by calculating the accuracy and plotting a graph of the accuracy versus the iteration of the code, and saw that the accuracy reached a point where it became consistent.

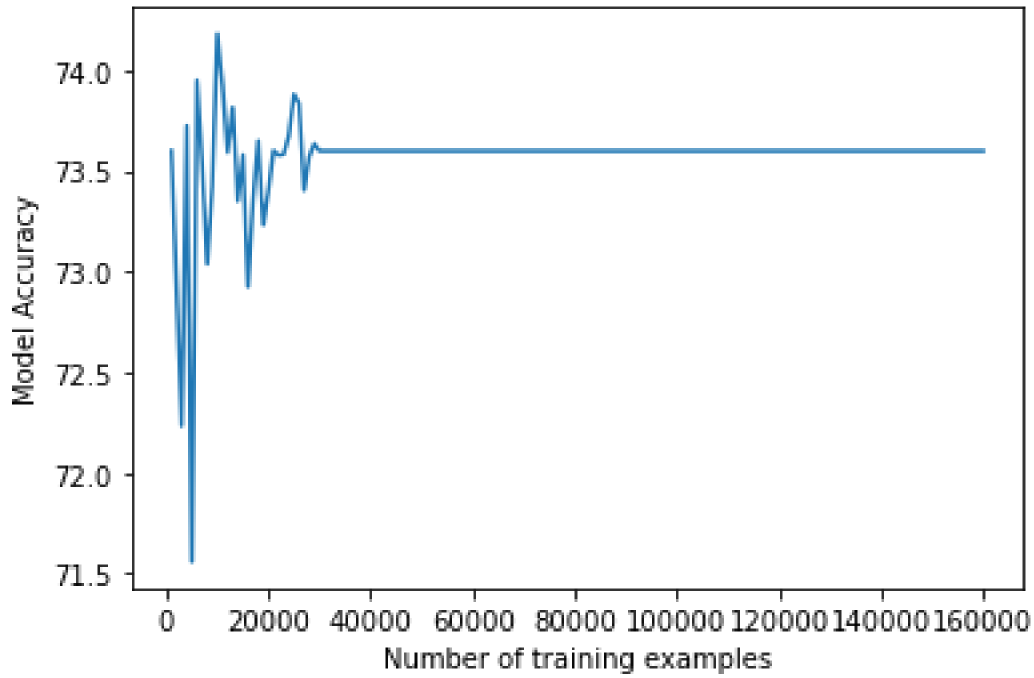


Figure 4: Accuracy becoming consistent

### 5.2 Precision

We also evaluated the code using precision, where the model was proven to be not as accurate as originally thought. The precision for the fire state when the state was 'START' was zero, which is likely due to the fact that we do not have a lot of data points where the fire started. However, our model did predict when fire was actively burning with good accuracy on the training set (70.9 percent). The accuracy was lower on the test set, but we believe this is because the number of fires in the test set did not end up being proportionally similar to the training set.

## 6 Conclusion

In our project, we were able to create a model that predicted when a fire was burning with good accuracy. Although our original intent was to predict whether a fire would start soon, we didn't end up having enough fire data to make accurate predictions for the start of fires. In compiling our data set, we learned why fire prediction can be so complicated. Deciding which factors to include in our model as well as finding complementary datasets to stitch together was surprisingly difficult.

There are a plethora of fire data collections, but they were frequently difficult to navigate and many sets had gaps in the data. Additionally, a large amount of the data we found was map based, not in the csv format we needed. If we could take our project further, we would explore ways to include more specific geographic data and data over a wider area. We would also expand the time span of our data to include more fires. We would also continue searching for additional data sets on other factors that could influence fire risk such as topography and previously recorded fire frequency in an area in order to test which factors gave us the most accurate predictions.

Since our work is in an area that is becoming increasing relevant as climate change advances, our model, with more testing, could be useful to fire prevention agencies in states such as California or Oregon where fires frequently cause severe damage to homes and businesses. In our chosen area of California, the power companies frequently turn off the electricity in order to prevent downed power lines from causing fires so accurate fire prediction modeling could help determine when planned power shut offs would be the most effective. Additionally, if our model was able to incorporate more specific geographic information, geographically specific fire risk predictions could help determine which areas to focus land management resources to reduce the risk of fire.

## References


CA.gov. Active Fires of Interest. Cal Fire. Accessed December 13, 2020. https://www.fire.ca.gov/incidents/.

Coffield, Shane R., Casey A. Graff, and Yang Chen. *"Machine learning to predict final fire size at the time of ignition."* International Journal of Wildland Fire. Accessed December 13, 2020. https://www.publish.csiro.au/wf/Fulltext/WF19023.

Collins, Luke, Andrew F. Bennett, Steve W. J. Leonard, and Trent D. Penman. *"Wildfire refugia in forests: Severe fire weather and drought mute the influence of topography and fuel age."* Global Change Biology 25, no. 11 (November 2019): 3571-994. https://doi.org/10.1111/ Gcb.14735.

Rothermel, Richard C. *How to Predict the Spread and Intensity of Forest and Range Fires.* Technical report no. INT-143. N.p.: United States Department of Agriculture, Forest Service, 1983.

Sayad, Younes Oulad, Hajar Mousannif, and Hassan Al Moatassime. *"Predictive modeling of wildfires: A new dataset and machine learning approach."* Fire Safety Journal 104 (March 2019): 130-46. Accessed December 13, 2020. https://www.sciencedirect.com/science/article/abs/pii/S0379711218303941.

University of Utah Department of Atmospheric Sciences. Page to Download Environmental Data. MesoWest. Accessed December 13, 2020. https://mesowest.utah.edu/cgi-bin/droman/download_api2.cgi?stn=KBUR.

USFS Wildland Fire Assessment System. Fuel Moisture Graphs and Tables. N.p.: Geographic Area Coordination Centers, n.d. https://www.wfas.net/index.php/ national-fuel-moisture-database-moisture-drought-103.

Young, Jesse D., Andrea E. Thode, and Ching-Hsun Huang. *"Strategic application of wildland fire suppression in the southwestern United States."* Journal of Environmental Management 245 (September 2019). Accessed December 13, 2020. https://www.sciencedirect.com/science/article/pii/S0301479719300039bib31.