WILDFIRE PREDICTION

CAPSTONE THESIS

ITHACA COLLEGE

BACHELOR OF SCIENCE

DEPARTMENT OF MATHEMATICS

BY

EVAN TURKON

FACULTY MENTOR: TOM PFAFF

ITHACA, NY

FALL SEMESTER 2022

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

A wildfire is classified as an unplanned fire that burns in a natural area such as a forest, grassland, or prairie [2], and for hundreds of millions of years wildfires have posed a threat to various ecosystems on earth. While there may be some minor beneficial environmental impact, wildfires do a disproportionate amount of damage. Which poses the question; How can we reduce damages caused by and prevent wildfires? This paper will explore that question by using exploratory data analysis and machine learning techniques to accurately predict ignition cause probabilities using records of past wildfires. Knowing the likely cause of a wildfire on any given day will provide a great resource as it allows for preventative action, as opposed to fighting an existing wildfire.

CHAPTER 1

INTRODUCTION/ MOTIVATION

This question regarding cutting down on wildfire induced damages and hot to overall prevent wildfires, has become increasingly more prevalent over the years, in the 1990s an average of 78,600 wildfires occurred annually burning 3.3 million acres. Since 2000 an average of 70,072 wildfires occur annually burning 7.0 million acres [1]. While fewer wildfires occur, the more important statistic is how many acres burn, better reflecting potential damages caused by wildfires, a number that has more than doubled since the 1990s, there are many potential wildfire-induced damages, some of the biggest being environmental destruction and air pollutants. Environmental destruction affects ecosystems in nature as well as human dwellings and air pollution greatly affects plant and animal health. On top of these damages, wildfires produce a great deal of carbon emissions, in 2021 wildfires emitted 1.76 billion tonnes of carbon[4] which equates to 4.8% of total emissions on earth.

As these problems do grow more prevalent, the increase in burn area has led to a rapid increase in suppression costs. The United States saw an increase from 2020 to 2021 of over 2.1 billion dollars, to a total of 4.4 billion dollars [3], total costs can be seen in figure 1.1. Getting better at fighting wildfires is a potential way to cut down on burn areas but would assuredly lead to an even further increase in suppression costs. The most worthwhile and cost-effective way to cut down on burn areas would be to put forth efforts for wildfire prevention. This can entail many things including ongoing efforts like wildfire prevention add campaigns, but the best building block would be gathering a better understanding of wildfires.

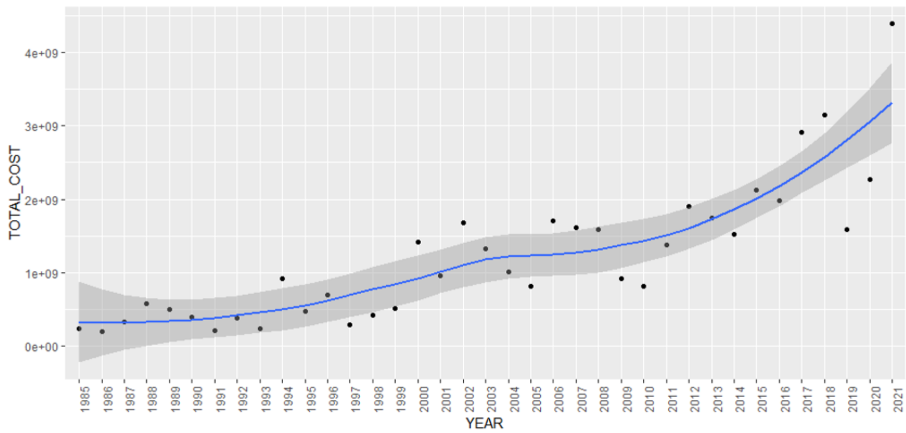

Figure 1.1: Wildfire Cost(USD)

The World Health Organization says the cause is unknown in 50% of wildfires recorded [2], cause has the potential to reveal a great deal about if, when, and where a wildfire will occur along with descriptive attributes like burn area and time.

This paper will explore creating machine learning models with the goal of being able to accurately predict the cause of wildfires given the cause is unknown, along with a model able to predict the cause of a potential wildfire given climate forecasts for the day, with the motivation of finding significant predictors wildfire which will hope to reveal more about what causes a wildfire to occur.

CHAPTER 2

LITERATURE REVIEW

In 2013 the U.S. Department of Agriculture(USDA) published an article titled: *Wildfire Ignitions: A Review of the Science and Recommendations for Empirical Modeling* which emphasizes the importance of better understanding the different behaviors of wildfires ignited by various causes. The article breaks down variables to predict ignitions into 3 groups; biophysical, societal, and management variables. In the conclusions of the article, it again emphasized the importance in finding significant predictors of wildfire ignition to find the underlying causes but discusses the societal implications as so many wildfires are human-caused. Therefore the article suggests focusing on finding the "cause behind the cause" [6] because understanding what causes a wildfire can reveal a great deal about wildfire behavior.

CSIRO publishing in Spain pubished an article titled: *A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain*, which went along the lines of the recommendations of the USDA developed a prediction model for specifically human-caused wildfires. The authors focused on a specific region in Spain where 90% of wildfires are caused by humans, broke the human causes into a number of classes and attempted to predict on some past wildfires and found some very significant results. [**?**]

CHAPTER 3

EXPLORING THE DATA SET

From 1992 to 2018 the US government published the Fire Program Analysis(FPA) program which included 2.17 million geo-referenced wildfire records from federal, state and local fire organizations representing a total of 165 million acres burned on US land[5].

Variable	*Description*	*Source*
Cause	Specific Source of Ignition	FPA
CauseClass	Human or Naturally Caused	FPA
StartDOY	Ignition Day of the Year	FPA
LAT	Latitude of Ignition	FPA
LON	Longitude of Ignition	FPA
Precip	Precipitation Amount (tenths of millimeters)	rnoaa
TempMax	Maximum Temperature (degrees Celsius)	rnoaa
DSCI	Drought Severity and Coverage Index	NDMC

Figure 3.1: Variables

Figure 3.1 reveals all the variables that will be tested in hope gain insight in wildfire behavior. As suggested by the USDA in this paper, the independent variable that will undergo investigation is cause, which reveals the cause of the majority of wildfires, and is given from the FDA data set. In the FPA program, cause falls into 13 factors, seen in figure 3.1. Developing an accurate model to predict the cause of a wildfire would not only allow to see with high probability cause of past wildfires, but would also be a great component for wildfire prevention. This can be achieved given conditions for the day, like weather prediction, into the model. Which then, if a wildfire were to occur, the probability of it being started by each particular general cause.

Figures 3.2 and 3.3 explore the distribution of cause through the data set. Figure 3.3 provides a data point for each wildfire given cause on its own US map, omitting the wildfires with undetermined causes assuming there is no significant visual trend in the distribution. These maps show a great deal including, most wildfires caused by Firearms and Explosives use occur in the western half of the US's lower 48, wildfires caused by railroad are much more concentrated in the southeast and majority of the wildfires in Alaska ignite naturally. It is possible for these trends to be caused by local laws that govern the area and not the specific location of the wildfires, for instance as stated above the prevalence of wildfires caused from firearms and explosives use in the western half of the lower 48 could be due to gun control laws. Figure **??** represents the distribution of wildfires in each case by their size. This shows wildfires ignited naturally, by fireworks and by recreation and ceremony can grow larger than the other causes but also that it is more typical for a wildfire caused by recreation and ceremony to burn smaller than 0.2 acres. Figure 7.1 confirms that the cause of over 25% of these fires is unknown and that campfires are the largest cause of human-caused wildfires.

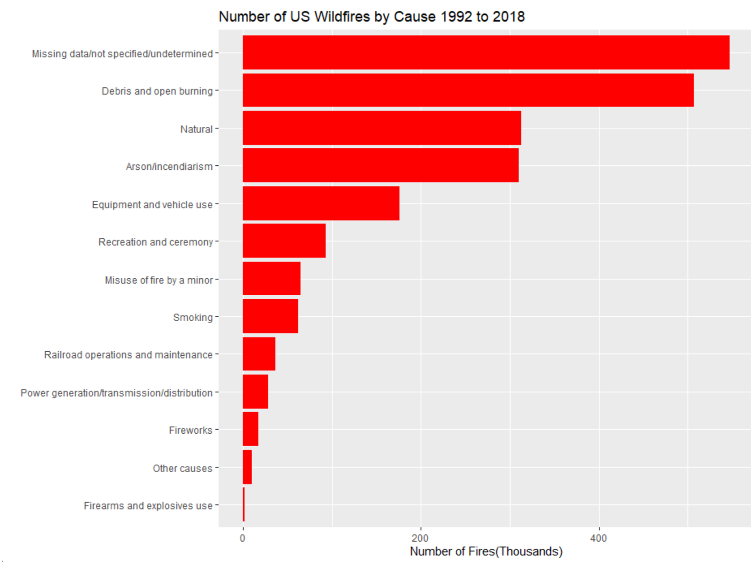

Figure 3.2: Number of Wildfires by Cause

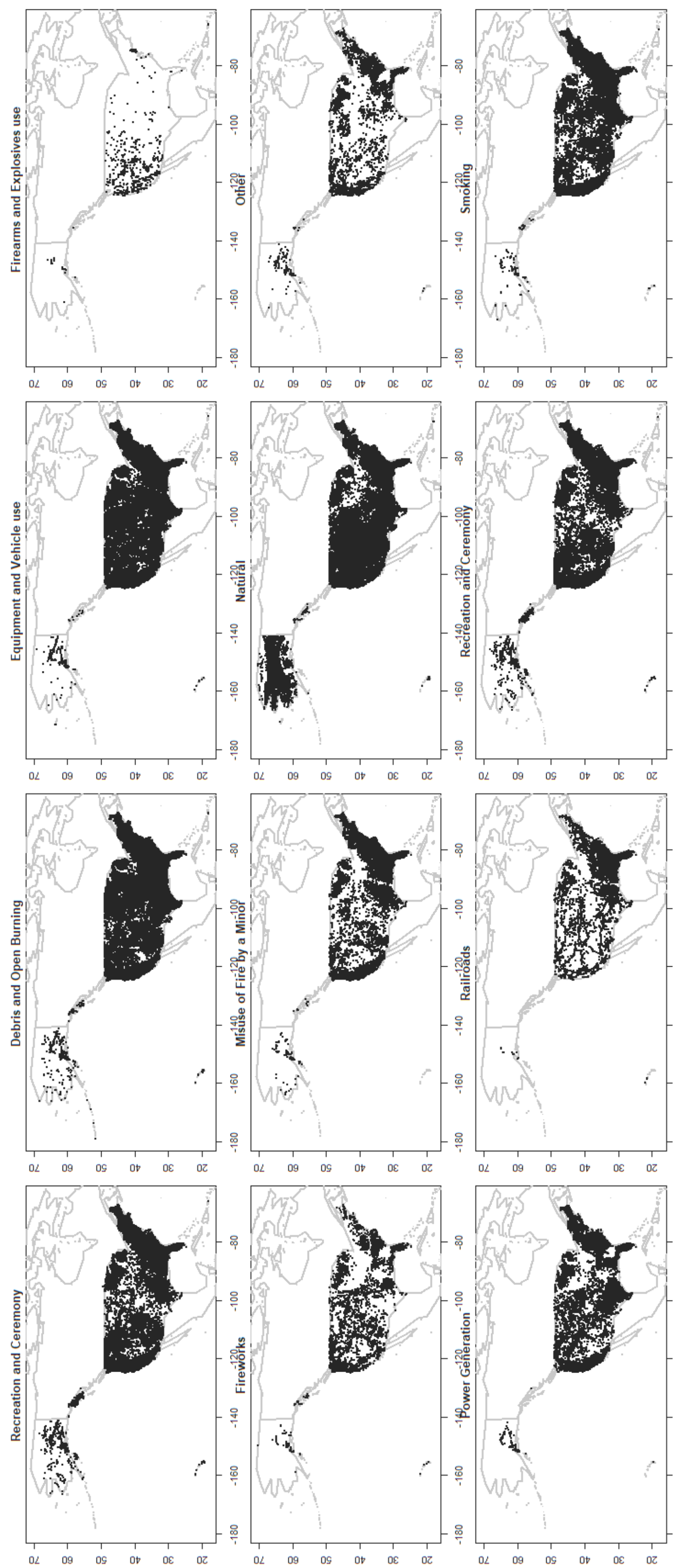

Figure 3.3: Cause Maps

CHAPTER 4

METHODOLOGY

The models will use environmental drivers to predict wildfires, as environmental conditions account significantly toward both human activity and natural occurrences that could account for causing a wildfire. Most environmental drivers are available readily on the internet, many of which already have their own predictions up to a week in advance. There is no problem using these environmental weather predictions in predicting a wildfire, days in the future. To account for these environmental drivers weather data will be implemented into the models. NOAA (National Oceanic and Atmospheric Administration) weather data will be used to achieve a very high level of accuracy as NOAA is a nationwide resource and can therefore be used for the entirety of the wildfire dataset. To achieve a very high level of accuracy when implementing the weather data, an algorithm was created with inputs of; latitude and longitude coordinates, and start date of ignition which results in the maximum temperature and precipitation of the day for the closest possible NOAA weather station. The algorithm achieves this by calculating the euclidean distance from each wildfire to every NOAA weather station that records the maximum temperature of the day and every weather station that records daily precipitation, then selects the closest station for each. Using NOAA's API and the station id for both closest stations (whether or not they are the same station) and the start date of the wildfire the desired information is generated and implemented into the wildfire data set.

Another environmental driver which has potential of being beneficial in this prediction are droughts, which are increasingly growing more prevalent alongside wildfires. It is thought that being in an area that is currently going through a drought, a wildfire is more likely to occur by natural processes. To attempt to account for droughts in the model drought data from the National Drought Mitigation Center was implemented on a state-by-state basis, using their Drought Severity and Coverage Index(DSCI) [7]. DSCI is calculated using a weighted sum of the percentage of the area of each state in a particular drought severity, the higher the DSCI, the more area, and severity of drought in a given state.

CHAPTER 5

MODELS

Two models with the same overlaying goal will be explored in this section using the variables listed above which can be seen in figure 3.1. Figure 5.1 gives the breakdown of the causes included and how they will be predicted in each of the models. Model 1, seen in figure 5.1 is the most general which attempts to predict whether a wildfire is caused by human or natural processes and only leaves out wildfires where the cause is unknown. Model 2 is the most specific model, its goal is to predict the specific human cause of a wildfire but does not include wildfires with unknown causes or causes generalized with "other", which also leaves out wildfires which are naturally caused. All models leave out wildfires with unknown causes because unknown causes have the potential to include wildfires that were actually ignited with other causes already being accounted for in the model, wildfires ignited from "Other" processes are excluded from model 2 because while "Other" in this data set only includes wildfires caused by humans, "Other" is too general to try to predict. Model 2 leaves out Naturally caused wildfire occurrences because the overwhelming majority of wildfires are Human-caused therefore this model has the potential to achieve a higher accuracy and reveal more about wildfire causes.

Model 1	Model 2
Human Caused	Arson/ Incendiarism
Human Caused	Debris and Open Burning
Human Caused	Equipment and Vehicle Use
Human Caused	Firearms and Explosives Use
Human Caused	Fireworks
Missing/ Not Specified	Missing/ Not Specified
Human Caused	Misuse of Fire by a Minor
Naturally Caused	Natural
Human Caused	Other Causes
Human Caused	Power Generation/ Distribution
Human Caused	Railroad Operations and Maintenance
Human Caused	Recreation and Ceremony
Human Caused	Smoking

Figure 5.1: Cause Broken up by Model

Method: Logistic Regression **n= 13,849**

Significance: (***) 0.001 (**) 0.01 (*) 0.05 (.) 0.1

Variable	Coefficient	Std. Error	p-values
C	10.287 (***)	0.243	2e-16
StartDOY	-0.00315 (***)	0.000385	2.95e-16
LAT	-0.0298 (***)	0.00542	3.81e-8
LON	0.0449 (***)	0.00198	2e-16
Precip	-0.00977 (***)	0.000624	2e-16
TempMax	-0.0931 (***)	0.00394	2e-16
DSCI	0.000453 (.)	0.000239	0.0586

Figure 5.2: Human or Naturally Caused, Logistic Regression

CHAPTER 6

RESULTS/ DISCUSSION

Model 1 performed very well when testing accuracy, achieving 83.5% accuracy and 99.8% sensitivity, the model did extremely well in correctly predicting the occurrence of naturally occurring wildfires. Where the model lacked in performance was in falsely predicting human-caused wildfires, here the model saw an accuracy of 83.5%, meaning in testing the model missed 651 naturally occurring wildfires out of 4000, which is due to so many more wildfires being human-caused. This false positive is represented in the low specificity statistic in figure 4, however, this ends up not having drastic repercussions, as falsely predicting human-caused wildfires would only result in a temporary halt in human activities like a campfire ban. Falsely predicting naturally caused wildfires have the potential for a much worse outcome, taking some drastic preventative measure and the wildfire ends up being caused by humans creates the potential for much larger consequences.

Accuracy:	***0.835***
AIC:	9491
Sensitivity:	0.999
Specificity:	0.0338
AUC:	0.824

Figure 6.1: Model 1 Accuracy

Given Model 2 is predicting 10 factors, the model performed reasonably well, achieving an accuracy of 46.7% and a 0.217 kappa coefficient, significance is seen and shows great potential for future research.

CHAPTER 7

RECOMMENDATIONS

Based on the outcome of model 1 recommendations can be made regarding wildfire prevention. All variables regarding prediction can easily be found with an internet search, which then can then be fed into the models, model 1 will result in probabilities of wildfire ignition being caused by humans or natural causes. Figure 5 gives general recommendations as to when to expect a wildfire to be naturally occurring.

Variable	Recommendation Regarding Naturally Ignited Wildfires
StartDOY	more likely later in the year
LAT	more likely further north
LON	more likely further east
Precip	more likely with more precip
TempMax	more likely with higher temperatures
DSCI	more likely the lower the index

Figure 7.1: Recommendations Based on Model 1

Depending on the outcome of model 1, if the prediction is likely human-caused, then the same variables fed into model 1 can be fed into model 2 which will then result in probabilities of the most likely human ignition causes for the day.

CHAPTER 8

LIMITATIONS

With the significance of the results, there is lots of motivation to continue with these predictions. A large limitation was in implementing the NOAA data into the wildfire data set. The limitation here lay in the speed of the API, for instance, 20 thousand wildfires took the computer 8 hours to load. Another limitation was in accounting for social drivers when predicting human-caused wildfires in Model 2. Knowing, accounting for, and implementing various laws and statistics about an area is likely to reveal a great deal about human-caused wildfires. There is lots of potential for a project building on the results of this paper.

CHAPTER 9

CONCLUSIONS

This major problem causes lots of major damage worldwide and due to climate change is becoming more and more prevalent. More resources being allocated to further research and analysis of predicting and preventing wildfires would do much good and would help with the climate crisis. Developing an algorithm to trust and use proactively has lots of potential and is an option that should be considered when it comes to wildfire prediction. Overall, there is much significance shown in this paper and lots of room to improve. The high accuracy of model 1 given the limitations and time constraints could be greatly improved. Adding more data would assuredly allow for higher accuracy in model 2 and would reveal lots around ignition causes of wildfires.

REFERENCES

[1] Hoover, K., and Hansen, L. A. (2022). Wildfire Statistics. Congressional Research Service. https://sgp.fas.org/crs/misc/IF10244.pdf

[2] World Health Organization(2022). Wildfires. https://www.who.int/health-topics/wildfirestab=tab$_1$

[3] National Interagency Fire Center(2022). Suppression Costs. https://www.nifc.gov/fire-information/statistics/suppression-costs

[4] Abnett, K. (2022). This is how much carbon wildfires have emitted this year. World Economic Forum. https://www.weforum.org/agenda/2021/12/siberia-america-wildfires-emissions-records-2021/: :text=said

[5] U.S. Department of agriculture. https://www.fs.usda.gov/rds/archive/Catalog/RDS-2013-0009.5

[6] U.S. Department of agriculture. https://www.srs.fs.usda.gov/pubs/gtr/gtr$_{s}rs171.pdf$

[7] National Drought Mitigation Center. https://droughtmonitor.unl.edu/DmData/DataTables.aspx

[8] Lara Vilar, Douglas G. Woolford, David L. Martell, M. Pilar Martin, (2009).CSIRO Publishing https://www.publish.csiro.au/wf/wf09030