Designing a data science exploration of 125,000 US wildfires to lower cost, save lives, and increase efficiency.

Janaar Harbour, Olivia Samples, Kun Zhao

[1]*College of Computing and Technology, Lipscomb University Nashville, TN, USA*

Abstract— Wildfire size, frequency, severity, and associated fatalities have upsurged at an alarming rate over the past 25 years, resulting in steep budget increases. A model that predicts the optimal number of personnel needed for fire suppression would significantly trim this firefighting budget, allowing for the focus to be redirected to land management and preventative efforts. By utilizing a unique dataset of 125,367 US wildland fire incidents recorded within a five year span, we propose a real-time model that will predict the optimal number of personnel needed to effectively fight fires. Our project is designed to increase efficiency and help lower the cost of the US Department of Agriculture's rapidly growing firefighting budget through accurate forecasting of personnel needs dependent on understanding the impact of regional differences, wet and dry seasons, and preventative methods on fire frequency.

1. INTRODUCTION

1.1 Significance

Wildfire size, frequency, severity, and associated fatalities have increased at an alarming rate over the past 25 years (North, 2015). In turn, the US budget for wildfires tripled from 1995 to 2015 and is expected to exceed $1.8B in cost by 2025 (USDA, 2015). Our project is designed to increase efficiency and help lower the cost of this rapidly growing budget.

This growing budget is negatively influenced by a human factor. Timothy Ingalsbee claims that fire management is accountable for at least half of fire suppression costs. When comparing fires with similar biophysical features but stark differences in expense, it is clear that the final decisions of fire management are the main influences for excess spending (Ingalsbee, 2015). In other words, there is a need for informative decision making tools that limit most subjective bias.

Currently, US fire management departments are lacking strong predictive modeling for firefighting personnel dispatch. As fire activity rises, the US forest service has had to increase firefighting staff and decrease land management staff. From 1998 to 2015, the fire staff adjusted from 5,700 employees to 12,000 employees, accounting for a significant portion of budget adjustments within that time frame (USDA, 2015). A model that predicts the optimal number of personnel needed for fire suppression would significantly trim the firefighting budget, allowing for the focus to be redirected to land management and preventative efforts.

Hence, efficiency through accurate forecasting of personnel needs is vital to U.S. Forest Service operations. This accuracy is dependent on understanding regional differences and the impact of wet and dry seasons on fire frequency. Additionally, in Reform Forest Fire Management, the benefits of preventative measures such as prescribed burning by the US Forest Service have helped initiate a great paradigm shift in forest fire management (North, 2015).

In this paper, we propose a model that will predict the correct number of personnel needed to effectively fight a wildfire without using unnecessary, additional resources, leaving units to be nimble as other fires and preventative needs arise. We believe that our model paired with wildfire forecasting models would aid firefighting agencies by facilitating more efficient resource management and decision support, enabling those fighting the fires to do so in the most effective manner possible.

1.2 Related Work

An abundance of previous work has been attributed to the prediction of forest fires, using tools such as meteorological data, satellite data, and infrared smoke scanners (Cortez, 2007). While they are able to predict fire location and size, they do not provide insight on how to react in real-time. Historically, the USDA Forest service uses tools such as the National Fire Management Analysis System (NFMAS), the National Park service uses a model called FIREPRO, and even the California fire service uses its own system called the CFES-IAM model. Each fire service utilizes these tools as insight towards protocol in order to optimize resource allocation per annual fire season, incorporating multiple fire incidents.

Conversely, multiple models have been built in order to provide insight on distinct fire incidents. Donovan's Integer Programming model attempts to do this by predicting features such as time and cost of resource per individual fire (Donovan, 2003). Wei's Chance Constrained programming model optimizes a fire fighting perimeter for a particular fire incident (Wei, 2015). Similarly, Haight's scenario based standard response model determines the correct number of dispatched engines to efficiently suppress individual fires (Haight, 2007). However, all of these models fail to provide insight for deploying personnel, an attribute that all of these prior resources are dependent on.

Susana Martin-Fernandez was successful in incorporating the manual resource in her real-time model. After applying discrete simulation algorithms and Bayesian optimization methods, she marked the importance of providing a real-time model so that fire management teams are able to efficiently apply a model onto distinct fire incidents (Martin-Fernandez, 2002).

In 2019, John Carr and Matt Lewis used a unique data set received from the US Department of Interior to answer the following question: What is the correct number of personnel needed to effectively fight a wildland fire in the U.S. without using unnecessary resources? After acquiring fire incident and fire resource data from the GeoPlatform ArcGIS Online Organization, Carr and Lewis cleaned data selecting 35 features based on the following criterion: relevance, non-correlated or duplicated features, and decently populated. Of the final 35 features, the geographical coordinates (longitude and latitude) of the wildfire locations and the incident ID numbers proved key, prompting the removal of additional records lacking either a geographical entry or a valid incident ID number. Since the ID number was a mandatory field, it was used to join to the resources table with the incidents, allowing the aggregation of personnel, equipment, and other resources associated with each fire. Their exploratory data analysis revealed that most of the incidents were not destructive and were easily contained. Roughly 2.3% of the fire incidents (about 3,000 out of 130,000 fires) burned more than 1 acre, requiring extensive resources and personnel ranging from a few to over 1,000.

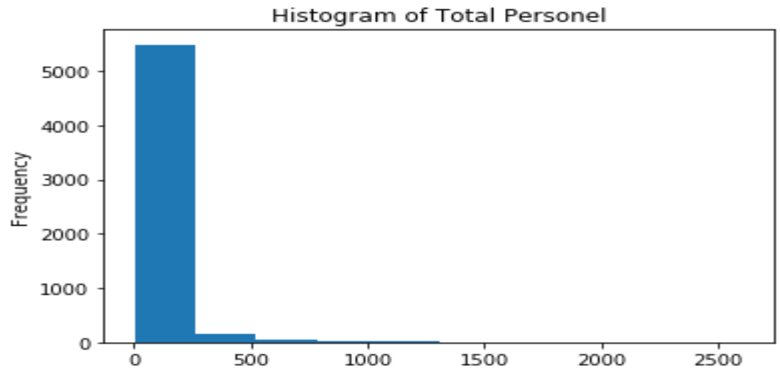


Figure 1: **Histogram of Total Personnel per incident, showing few fires have had more than <u>250</u> personnel deployed.**

Their analysis applied tools such as Rank Order Label Encoding and the model of XGBoost and Shap on cleaned data for optimal predictions and explanations. John Carr and Matt Lewis were able to devise a model capable of explaining 61% of the data variation with a Mean Absolute Error (MAE) of 1.39. Meaning, the error from record to record was off by about +/- 1.4 personnel per record making their predictions off by less than two people as shown in the figure below. Their data set is the basis for our project (Carr, 2019).

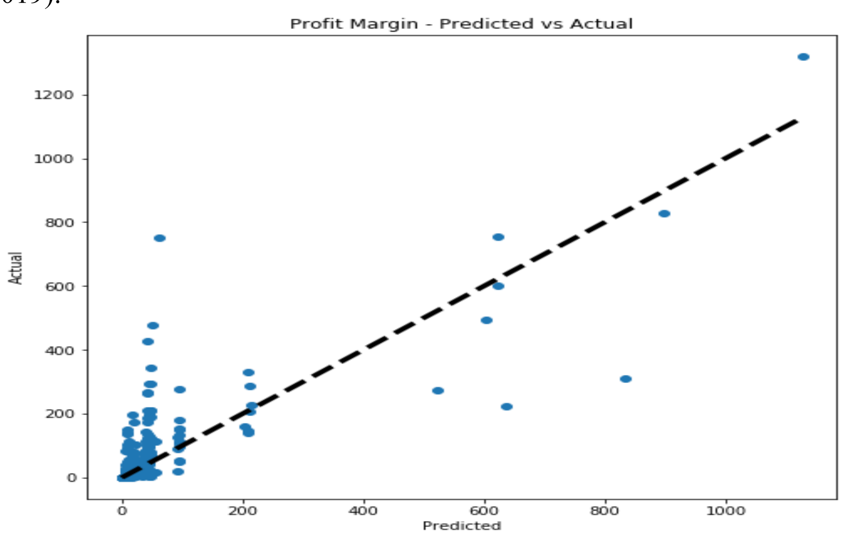


Figure 2: **Scatter Plot of predicted values vs. the actual. Showing fairly-good accuracy for the very few incidents in which a lot of personnel is needed.**

1.3 Research Question

Our research question on this project is the following:

What is the optimal number of personnel to be deployed in order to efficiently and effectively suppress wildland fires?

We will peripherally explore the impact of prescribed fires in regard to decreasing the number of personnel needed for future wildland fires, how the regional location of a fire affects personnel deployment, and if preceding wet and dry weather seasons influence the need for personnel differently? Exploration of this question and its subtopics can ultimately save precious resources in both human and monetary capital.

Furthermore, the rapid increase in fire management cost is harmful to the overall forest service budget. The US budget for wildfires *tripled* from 1995 to 2015 and is expected to exceed $1.8B in cost by 2025. Our project is designed to increase efficiency and help lower the cost of this rapidly growing budget. Below is a figure taken from the US Forest service and illustrates the percent of the forest service total budget devoted to wildfire costs is expected to increase by more than 400% from 1995 to 2025.

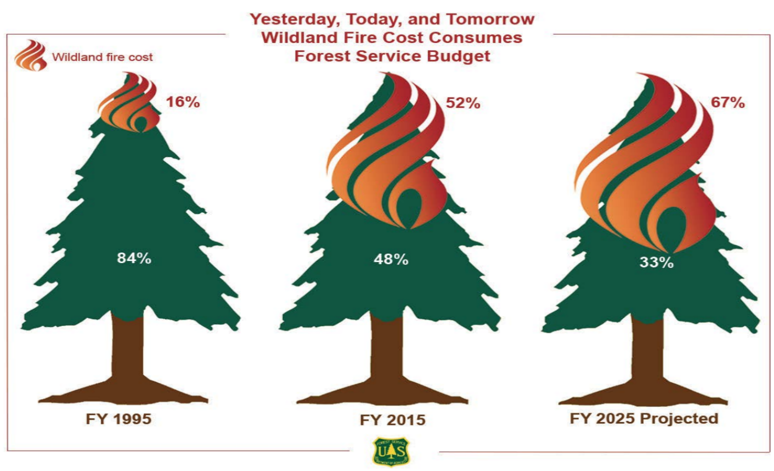


Figure 3. The percent of the forest service total budget devoted to wildfire costs. Taken from (USDA 2015).

1.4 Challenges

The five challenges have been identified that may impact the success of the project are as follows:

1. Time constraint
2. Lack of thorough content knowledge
3. Veracity of the data
4. Implementation of solution may challenge status quo
5. Cost savings in fire management may not directly benefit the forest service

2. MATERIALS AND METHODS

2.1 Data Collection and Preprocessing

2.1.1

IRWIN Observer is a read-only web application designed for viewing data that is being shared through the Integrated Reporting of Wildland-Fire Information (IRWIN) integration services. Access to this application is granted by the Chief Information Officer for the GeoPlatform ArcGIS software and it can be found here: https://irwin.doi.gov/observer. IRWIN Observer provides current and transactional views of incident and resource data being shared by partners within the wildland fire community. This data provides the location of existing fires, size, conditions and several other attributes that help classify fires.

We retrieved fire incident data from the US Department of Interior's IRWIN Observer database which details the Office of Wildland Fire's complete records of wildland fires within the country.

2.1.2

Our data represents all US wildland fires within a five year span between Jan 1st, 2015 and Dec. 31st, 2019. The incident dataset includes 135 features that describe 125,367 fire incidents. Additionally, we pulled a resource dataset detailing 16 resource features associated with these incidents. The resources table includes information such as: number of personnel, trucks, planes, helicopters, boats, and agencies supporting the firefighting efforts. We will combine these datasets together using the fire incident's "Irwin ID".

2.2 Methods

2.2.1

After thorough data cleaning, we will manually examine the dataset to determine and validate the most significant features, detailing our exploratory data analysis. We will use tools such as ANOVA, to statistically analyze the data.

2.2.2

We will utilize Rank Order Label Encoding (ROLE) to convert all categorical variables into ordinal variables. Rank Order Label Encoding (ROLE) provides us with an average of the target variable for every unique value within each categorical feature. In our case, the target variable is the number of personnel. We can use this technique to find the average number of personnel for each POO state or any other unique value in a categorical feature. (Carr, 2019)

2.2.3

The prediction model will be validated by XGBoost, which is a distributed gradient boosting library that provides parallel tree boosting to machine learning algorithms and typically out-performs most machine learning models across a variety of applications. (Carr, 2019)

2.2.4

The prediction model will be adjusted by prescribed fires, region and weather. We will determine the different effects these features have on our model.

REFERENCES

Canton-Thompson, J., Thompson, B., Gebert, K., Calkin, D., Donovan, G., & Jones, G. (2006). Factors affecting fire suppression costs as identified by incident management teams. Res. Note RMRS-RN-30. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. 10 p., 30. https://www.fs.fed.us/rm/pubs/rmrs_rn030.pdf Retrieved on Feb. 10th, 2020

Carr, J. & Lewis, M.(2019). Predicting Number of Personnel to Deploy for Wildfire Containment. Research Symposium Lipscomb University

Cortez, P., & Morais, A. D. J. R. (2007). A data mining approach to predict forest fires using meteorological data. http://www3.dsi.uminho.pt/pcortez/fires.pdf. Retrieved on Feb. 10th, 2020

Donovan, G. H., & Rideout, D. B. (2003). An integer programming model to optimize resource allocation for wildfire containment. Forest Science, 49(2), 331-335.

Haight, R. G., & Fried, J. S. (2007). Deploying wildland fire suppression resources with a scenario-based standard response model. INFOR: Information Systems and Operational Research, 45(1), 31-39.

Ingalsbee, T. & Raja, U. (2015) The Rising Costs of Wildfire Suppression and the Case for Ecological Fire Use. The Ecological Importance of Mixed-Severity Fires: Nature's Phoenix. Chapter 12. Elsevier Inc.

Lee, Y., Fried, J. S., Albers, H. J., & Haight, R. G. (2013). Deploying initial attack resources for wildfire suppression: spatial coordination, budget constraints, and capacity constraints. Canadian Journal of Forest Research, 43(1), 56-65.

Martell, D. L. (2015). A review of recent forest and wildland fire management decision support systems research. Current Forestry Reports, 1(2), 128-137.

Martin-fernÁndez, S., Martínez-Falero, E., & Pérez-González, J. M. (2002). Optimization of the resources management in fighting wildfires. Environmental Management, 30(3), 352-364.

National Wildfire Coordinating Group, & Nwcg. (2015, July 29). National Wildfire Coordinating Group (NWCG). Retrieved March 28, 2020, from https://www.nwcg.gov/sites/default/files/stds/standards/geographic-area-coordination-center_v1-0.htm.

NFPA 1. (2018). Retrieved from https://www.nfpa.org/codes-and-standards/all-codes-and-standards/list-of-codes-and-standards/detail?code=1.

North, M. P., Stephens, S. L., Collins, B. M., Agee, J. K., Aplet, G., Franklin, J. F., & Fule, P. Z. (2015). Reform forest fire management. Science, 349(6254), 1280-1281.

Wei, Y., Bevers, M., Belval, E., & Bird, B. (2015). A chance-constrained programming model to allocate wildfire initial attack resources for a fire season. Forest Science, 61(2), 278-288.

United States Department of Agriculture(USDA). Forest Service. (2015). The rising cost of wildfire operations: Effects on the Forest Service's Non-Fire Work..

https://www.fs.usda.gov/sites/default/files/2015-Rising-Cost-Wildfire-Operations.pdf. Retrieved on Feb. 10th, 2020.

APPENDICES

Appendix A: Content Expert

Trent Girard currently serves as a Fire Management officer for Cherokee National Forest. As a strong advocate for prescribed fires, Girard has aided in the optimization of wildland fire control throughout his entire career. Girard provided insight on current US wildland fire protocol and clear explanation of fire management terminology. (2020)

Appendix B: Wildfire Terminology

Terms to note are defined as follows:

Fuel - Describes litter or organic matter, including leaves, moss, smaller trees, etc.

POO - Point of origin, can refer to any location such as state or county that is further detailed.

Containment - Fire progression has stopped.

Suppression - Actions for stopping the fire and its growth have been taken after containment.

Controlled - Fire has officially stopped burning (Girard, 2020).

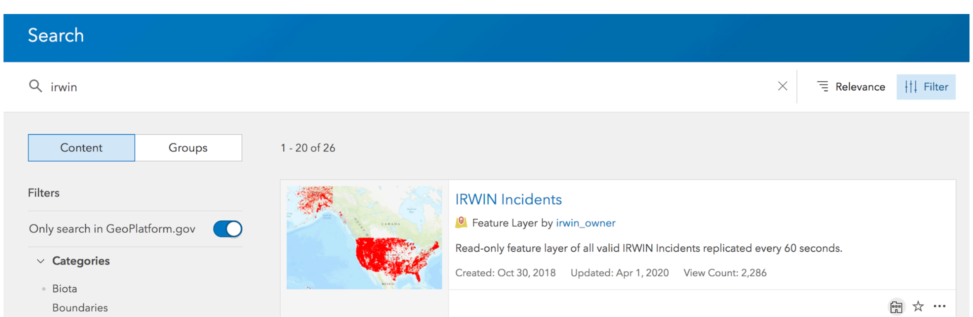
Fire Code - A list of codes that standardizes labeling for fire incidents. This computer assigned code serves as a way of tracking the spending for each incident, providing a legal financial document as well (NFPA 1, 2018).

Agencies Supporting - Refers to the Geographic area coordination center associated with the National Wildfire Coordinating Group (National, 2015).

Appendix C: Acquiring Dataset

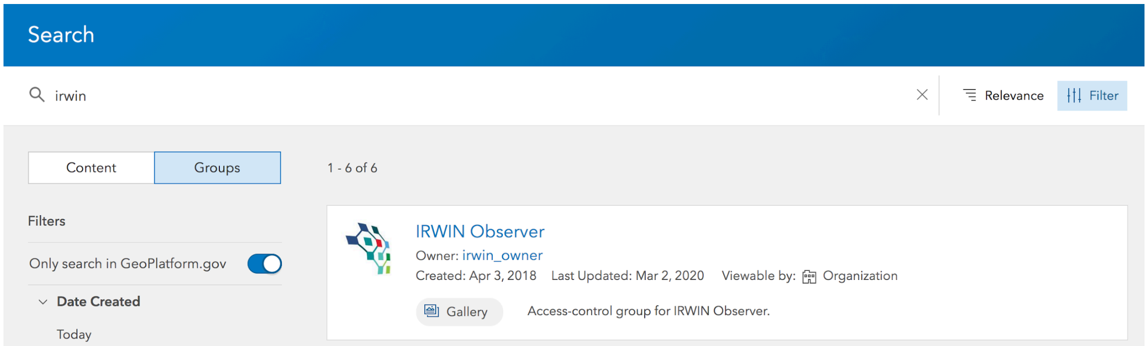
After creating an account on GeoPlatform.gov, we were granted access to GeoPlatform ArcGIS online via the Chief Information Officer for the GeoPlatform ArcGIS software. We then followed this link - https://geoplatform.maps.arcgis.com/home/index.html - which is where we were able to access our data.

We obtained our datasets from "IRWIN incidents" after searching for "Irwin Resources" under "Content".

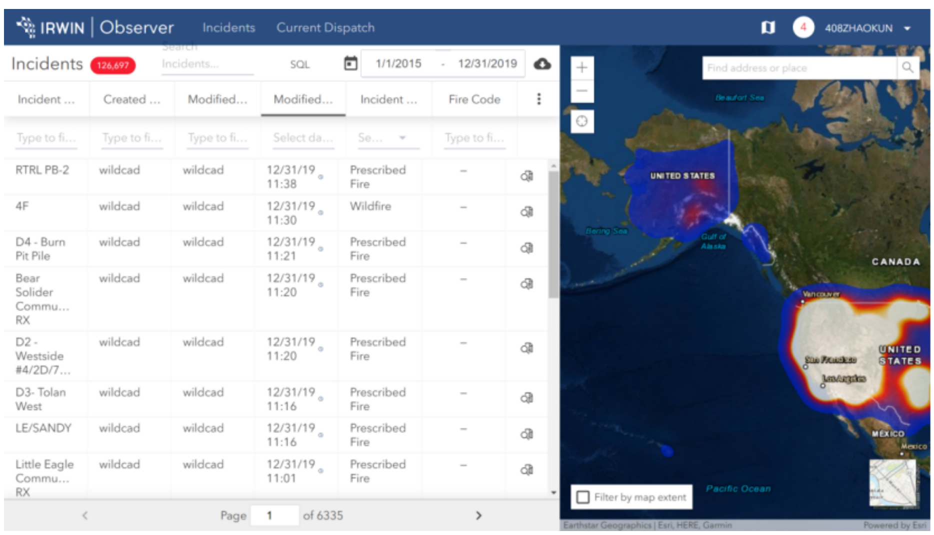


This provided the downloadable "Incident" and "Incident_ResourceSummary" dataset that we utilized in this paper.

The Irwin Observer link - https://irwin.doi.gov/observer/ - was found after searching for “Irwin” under “Groups”.



This showed the real time data as well as a corresponding map that represented the data. While you are able to download data from this link, the prime use for this resource is the visualization of the data.



Appendix D: Data Exploration

Our data exploration and analysis will be stored on Github.

A link to this site will be placed here.