Predicting the Impact and Severity of Wildfires in Alberta using AI
CxC Data Science Hackathon: EY Challange

Source:
https://www.alberta.ca/system/files/styles/responsive_1040/private/custom_downloaded_images/fpt-wildfire-forest-fire.jpg?itok=aKHZkrLA

Prepared For
CxC Data Science Hackathon

Prepared By
Zia Allahverdiyev, Ajlal Paracha, Mykhailo Briazkalo, Anna Kudiakova
Submitted on February 23, 2024

Abstract

Wildfires impact millions of lives and natural resources each year. In 2023, Canada saw record-breaking wildfires. By the end of the 2023 wildfire season, more than 6,132 fires devastated 16.5 million hectares of land (Canada, 2023). These staggering numbers show Canada is more than ever vulnerable to wildfires. Until recent years, Alberta was the hot spot for wildfires. In this report we analyze 2006 to 2021 wildfire data in Alberta and present:

- FSA regions that are most vulnerable to wildfires.
- Main cause of wildfires in these vulnerable regions.
- Damage caused by wildfires to the general and indigenous population.
- Improvements in government measurements in fighting wildfires.

Some FSA regions of Alberta have shown to be particularly vulnerable to wildfires. We will discover why certain regions tend to be more vulnerable than the rest and what are the main causes of fires in these areas. Moreover, we will look at damage caused by wildfires and how it affects general and indigenous populations. Based on these findings we build an AI model that can predict the risks associated with fires. In addition, we will briefly look at how the Canadian government historically tackled wildfires and provide recommendations on how the government can further improve its wildfire management capacity and capability.

Table of Contents

List of Figures

1.0 Introduction

Wildfires have been a persistent feature of the Canadian landscape throughout history. Despite its long history of wildfires, the government of Canada still does not know how to address this issue (Struzik, 2022).

In 1950, for the first time, Canadians started collecting wildfire data. Since then, the government has developed various policies and strategies to fight wildfires. Recently, the government has made attempts to use advancements in AI and ML to harness the power of collected data in predicting wildfires and the risks associated with them. Predicting where a wildfire may ignite before it occurs could save Canadians up to $5 million a year (Pope, 2017). Cons of most of these models is that they treat all fires equally. Our findings show that some types of fires should be treated individually.

This data analysis shows a decline in the frequency of wildfires, but an increase in the area burned and destruction caused. Our data set from the 2006 – 2021 fire seasons in Alberta does in fact, support this claim.

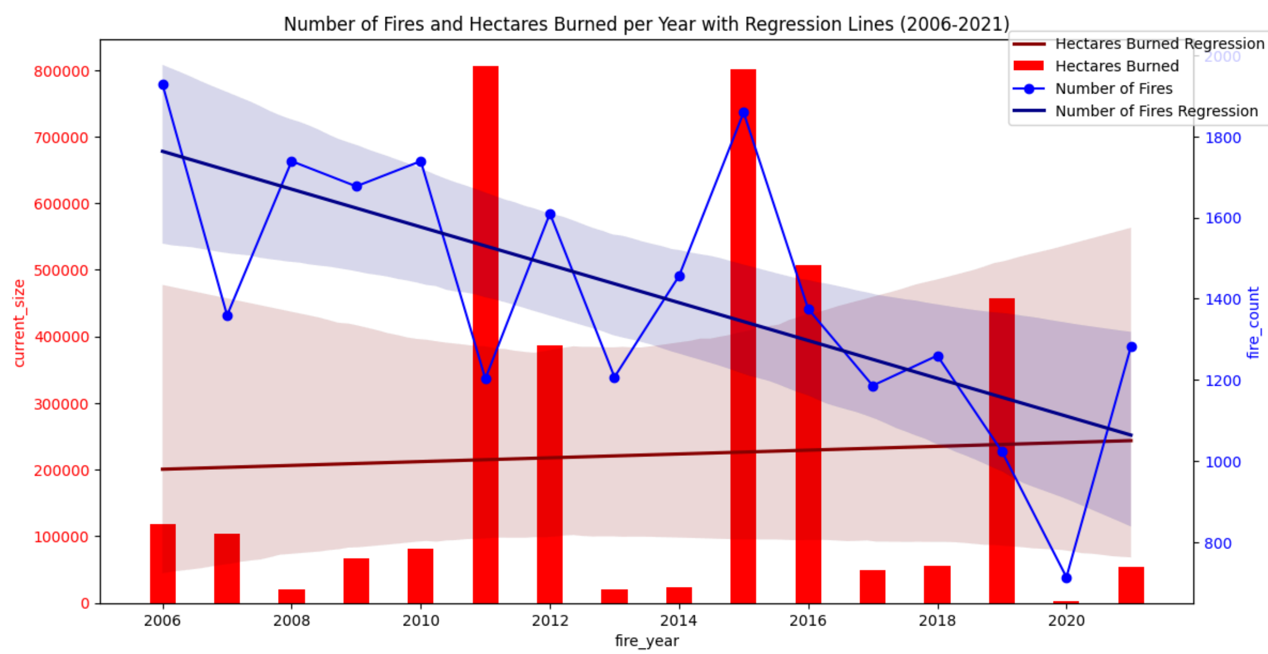

Figure 1.

Furthermore, analysis shows that most of that damage is caused by a relatively small number of fires - about 4% in total. Hence, we will focus on particularly large fires and try to answer the question of why Canadian wildfires have become more severe. We will look at the damage caused by these monstrous fires. In particular, we will model under which conditions these "super wildfires" originate and model their behavior through the burning time.

We define the vulnerability of an FSA region to wildfires as the frequency of wildfires there. It may seem wrong to not have an associated metric for the severity of a fire, say its total size, burn time, etc. However, it is important to realize that all small fires can turn into large fires.

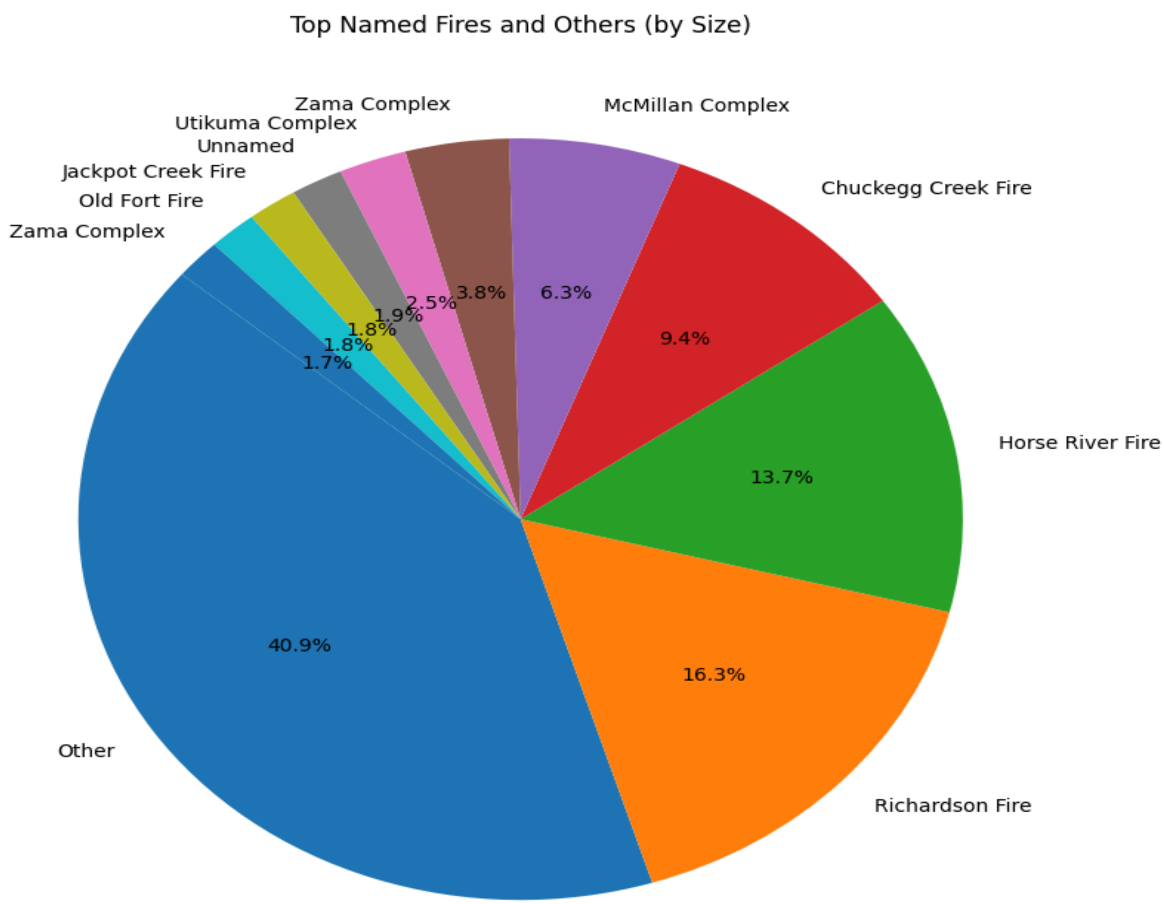

Figure 2.

As this graph demonstrates, huge fires, like 2011's Richardson Fire, account for great amounts of total burned area over the past 15 years. Indeed, of over 22000 wildfires, Richardson accounts for over a sixth of all burned areas, thus stressing the importance of early detection and intervention in mitigating wildfire damage.

2.0 Methodology

Our methodology includes the following approaches:

- Create visualizations and find patterns.
- Cross-reference patterns with 3rd-party research.
- Use data storytelling and logic to explain patterns.

The dataset provided was meticulously preprocessed prior to visualization and model training. We highlight key steps below:

1. Data Imputation: Categorical features were title-cased and missing values were set to a relevant value (“Unknown”, “Inapplicable”, etc). Missing values, both categorical and numeric, were imputed carefully; examples include setting to the nearest neighbor, mean, median, and mode, each with filters (ex. median response time within a given size class).

2. Feature Engineering: Some relevant features created were:
    - Fire Region extracted from the first letter of the fire number.
    - Ordinal encoded the size class.
    - One-Hot encoded the fire origin.
    - One-Hot encoded the fire cause (general cause description) after grouping together small categories to reduce dimensionality.
    - Grouped together activity class and true cause categories to more general ones to group together smaller categories.
    - Create time interval features as days between (ex. days to extinguish, days to report)

We define vulnerability to wildfires first as the frequency of wildfires. At the highest level, we can plot the fires throughout Alberta.

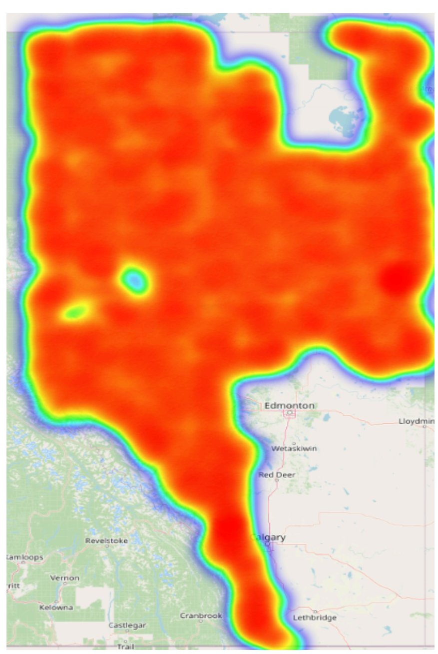

Figure 3. Parts of Alberta affected by wildfires

Clearly, there are large non-fire gaps, namely the bottom right and top middle. It follows that the lack of forestry, and thus fire fuel, is the reason for this. We can immediately conclude these FSA regions are the least vulnerable to wildfires.

Then, we seek to find greater granularity within regions with wildfires. We start by plotting the number of wildfires per FSA region.

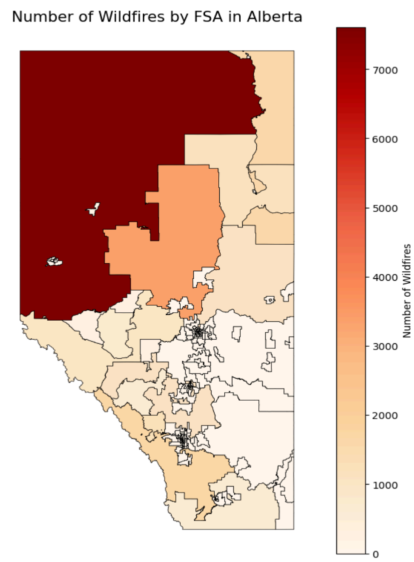

Figure 4.

Plotting the number of wildfires per square kilometer is a good metric taking into account the size of regions. We immediately see the different stories the 2 graphs show, demonstrating that while the topmost FSA regions get the most wildfires, that is largely attributable to their massive size; indeed, even these maps show the top regions as smaller than they are.

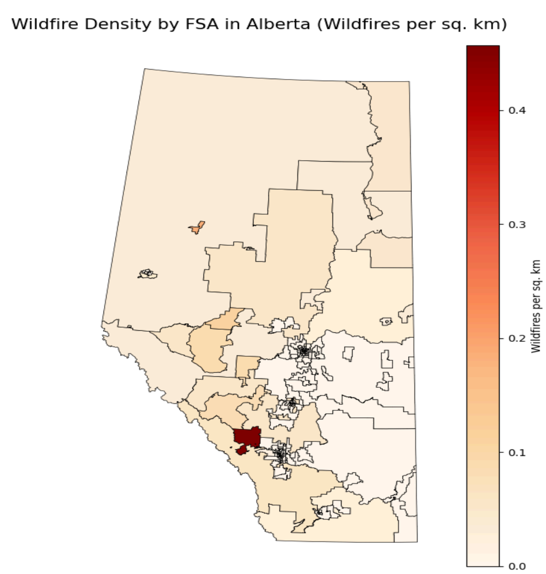

Figure 5.

3.0 Experiments & Results

Our team uses and compares a variety of models for the imbalanced multiclass classification task with their hyperparameters optimized through a grid search. In particular, we predict fire size class based on available data at the point when the fire was discovered and assessed. Our comparison of the models is based on accuracy, precision, recall, f1 score, and macro f1 score as well as the models' tendency to overfit.

Based on the research and our understanding of the data, we excluded data that became available after the fire had been extinguished from the training set. Additionally, we performed feature engineering, encoding, and normalization, which improved the models' performance.

We compared XGBoost, Random Forest, Gradient Boosting, Support Vector Classifier (SVC), and Logistic Regression. The highest accuracy and macro scores are achieved by XGBoost and Gradient Boosting models, while SVC and Logistic Regression show the lowest scores. (Figure 6)

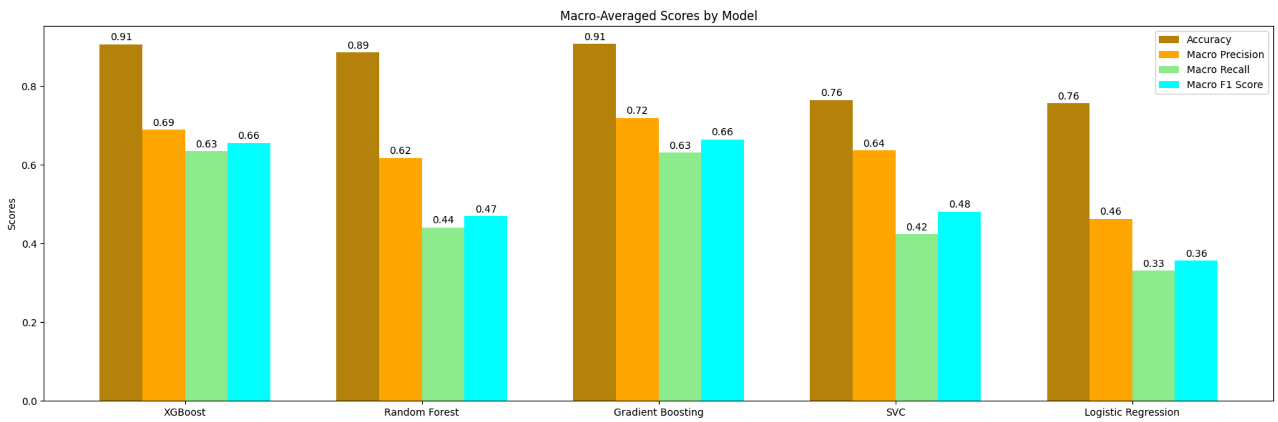

Figure 6.

Among the best two models, Gradient Boosting has slightly better macro scores and overfits less than other models (Figure 7), which makes this algorithm the best choice for the wildfires dataset.

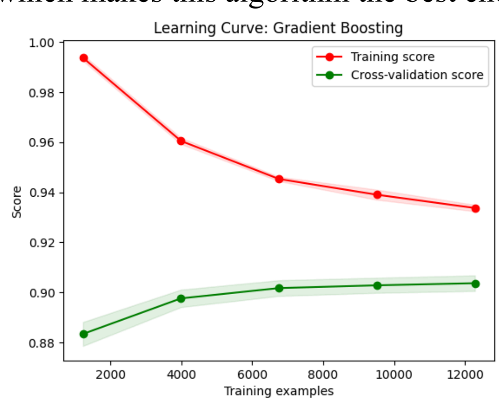

Figure 7.

For future research, we aim to create a more complex model; namely, an ensemble (like soft voting) with various models like XGboost, LightGBM, and Catboost, to name a few. This strategy, we hypothesize, will outperform our current models as the ensemble model will better represent and learn different parts of the dataset, thus yielding a more comprehensive understanding of the data.

We further aim to do more comprehensive testing of the importance of individual features, testing out different encoding strategies, data imputation strategies, etc. Indeed, there are many potential gains in doing the laborious task of best extracting the dataset's value. We were unable to do so due to time constraints for this competition.

4.0 Conclusions and Recommendations

To assess the vulnerability of the population in vulnerable FSA regions we explored various methods incorporating the dataset "Population and dwelling counts: Canada and forward sortation areas" from Statistics Canada with information on the population and dwelling size by FSA regions.

We calculated the number of wildfires per capita for each FSA, which is a vital metric for understanding the relative vulnerability of different regions. This calculation was done by dividing the wildfire count by the population size, offering insights into the frequency of wildfires concerning the number of people potentially affected.

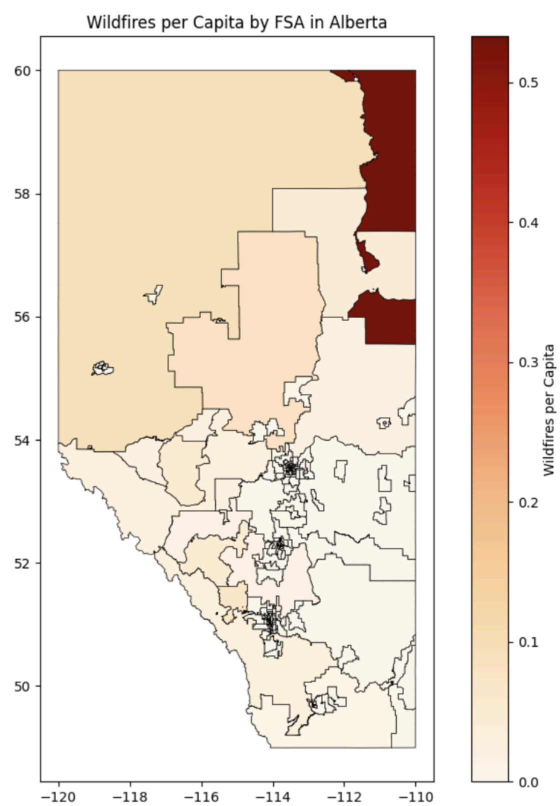

Figure 8.

Bar plots were created to visualize the population size and wildfire count in the top 10 most vulnerable FSAs, allowing for a clear comparison of how population size relates to wildfire count in these regions. Additionally, we produced another bar plot to show the relationship between the number of private dwellings and wildfire count in the top 10 most vulnerable FSAs, adding depth to our understanding of the impact of wildfires on human settlements. A scatter plot was generated to compare the average fire size with the population size in the most vulnerable FSAs, helping us understand if larger fires were occurring in more populated areas. We also calculated and plotted wildfire density and population density for the most vulnerable FSAs, providing insight into how densely populated areas are affected by wildfire density.

To understand the impact on Indigenous populations, we referred to the "Indigenous identity by Registered or Treaty Indian status and residence by Indigenous geography: Canada, provinces and territories" dataset from Statistics Canada. We found that 44,730 out of 4,177,715 Indigenous individuals live in reserve areas in Alberta, which are particularly susceptible to wildfires. These communities often lack the resources and infrastructure to recover quickly from such disasters compared to metropolitan areas.

Our analysis revealed that some areas of Indigenous reserves suffer from high numbers of wildfires, posing a significant threat to these communities. The lack of detailed data on the exact locations of reserves in Alberta was a challenge, but we used available maps to identify regions where Indigenous reserves are affected by wildfires.

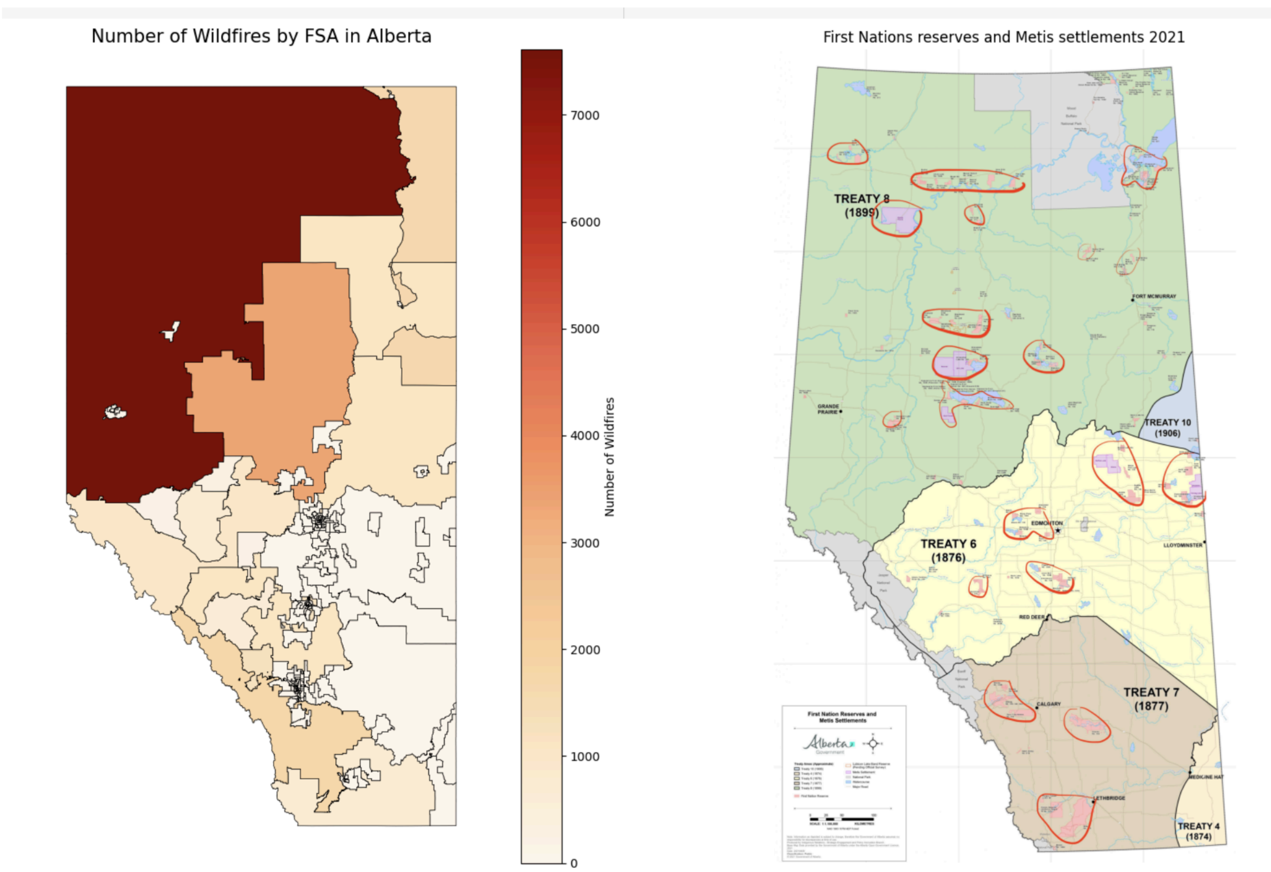

Figure 9.

Indigenous populations in Canada are disproportionately threatened by wildfire smoke and the associated adverse health impacts (Batdorf & McGee, 2023). The effects of wildfires on Indigenous communities are profound, as they not only cause immediate harm but also threaten cultural activities like hunting, fishing, and gathering native plants, which are integral to Indigenous traditions.

References

Coogan, S. C. P., Daniels, L. D., Boychuk, D., Burton, P. J., Flannigan, M. D., Gauthier, S., Kafka, V., Park, J. S., & Wotton, B. M. (2021). Fifty years of wildland fire science in Canada. *Canadian Journal of Forest Research*, *51*(2), 283–302. https://doi.org/10.1139/cjfr-2020-0314

https://natural-resources.canada.ca/simply-science/canadas-record-breaking-wildfires-2023-fiery-wake-call/25303

ALEXANDRA POPE. (2017, September 3). *Alberta to harness artificial intelligence technology to predict forest fires before they begin*. Canadian Geographic. https://canadiangeographic.ca/articles/how-artificial-intelligence-could-help-predict-major-forest-fires/

Struzik, E. (n.d.). *Despite its long history of wildfires, Canada still doesn't know how to live with them*. The Conversation. https://theconversation.com/despite-its-long-history-of-wildfires-canada-still-doesnt-know-how-to-live-with-them-190213

D'Andrea, A. (2023, December 18). Canada 'very vulnerable' to wildfire damage amid record year. How to protect your home. *Global News*. https://globalnews.ca/news/10176775/canada-wildfires-home-protection/

Batdorf, B., & McGee, T. K. (2023). Wildfire Smoke and Protective Actions in Canadian Indigenous Communities. Atmosphere, 14(8), 1204.

https://www.mdpi.com/2073-4433/14/8/1204