The Environmental Context of Fire Severity: Global Patterns and Insights

Created by Brianna Klabunde
07/11/2025

Problem Statement

Wildfires pose escalating threats to human safety, ecosystems, and infrastructure worldwide. Increasing fire frequency and intensity, driven by climate change and land use patterns, are placing growing pressure on fire management agencies to anticipate and allocate resources effectively.

The problem this project investigates is whether the likely severity of a wildfire can be predicted at the time of ignition using data on fire location, recent fire activity, weather conditions, and vegetation status. Early knowledge of potential fire severity could improve preparedness, support resource prioritization, and reduce response delays.

Currently, fire severity is often assessed reactively, once a fire has already spread. The desired state is a proactive system where data driven tools support early stage decision making. While previous research has explored fire occurrence and spread prediction, relatively few models focus specifically on predicting projected severity using globally aggregated environmental datasets. This project aims to address that gap.

Industry

The project sits within the environmental management and disaster risk reduction sector, particularly in the context of wildfire management.

The industry faces challenges in balancing prevention, suppression, and recovery, especially as fire seasons lengthen and resources remain limited. Predictive analytics offers a potential advantage by allowing agencies to assess risk dynamically and plan ahead of events.

The wildfire management value chain spans data collection (satellite and ground sensors), analysis (modelling and forecasting), and response (allocation of personnel and equipment). Predictive models like this one contribute to the analytical layer, transforming raw data into actionable insights.

Although this project focuses on wildfires, similar methods could apply to other environmental hazards such as floods, droughts, or storms, where early prediction supports better disaster preparedness.

Stakeholders

The primary stakeholders are government fire management authorities, environmental monitoring agencies, and disaster preparedness organizations.

These groups care about the problem because accurate early predictions of wildfire severity could:

- Enable better resource allocation (e.g., prioritizing high risk ignitions for rapid response)
- Reduce costs associated with large scale fires
- Improve safety outcomes for firefighting personnel and affected communities

Stakeholders expect a model that, even with moderate accuracy, provides useful insights into potential fire severity patterns and highlights where improvements in data resolution could further enhance predictive power.

Business Question

Can we predict the likely severity of a wildfire at the time of ignition using aggregated fire history, weather, and vegetation data?

Answering this question could help agencies allocate suppression resources more efficiently and improve situational awareness during early response.

Even a model with modest accuracy can deliver value by highlighting regions or conditions linked to higher severity fires, guiding both immediate and strategic decision making.

For this application, false negatives (predicting a low severity fire that becomes severe) carry more risk than false positives, as they could lead to under preparation or delayed response.

Data Question

What patterns in historical fire activity, weather, and vegetation data are most predictive of potential wildfire severity?

To answer this, the model requires fire level data (area burned) combined with environmental variables that describe recent conditions leading up to ignition.

Data

Description:
The dataset includes global wildfire records combined with environmental indicators from 2002 to 2020. Each observation represents an individual fire event, with associated climate and vegetation data aggregated at the country-year level.

Volume: Approximately 1.8 million records across 9 numerical features.

Attributes:

- Core fire characteristics: severity (target variable built from area data), latitude, longitude
- Engineered features: months since last fire, season progress, fires per year by country, vegetation stress
- Environmental indicators: Fire Weather Index (FWI), Vegetation Index (NDVI), Vapour Pressure Deficit (VPD)

Quality and Reliability:
The data is derived from reputable remote sensing and climate monitoring sources but is limited by spatial aggregation. Climate and vegetation indicators averaged at the country-year level mask regional variability, reducing precision in model predictions.

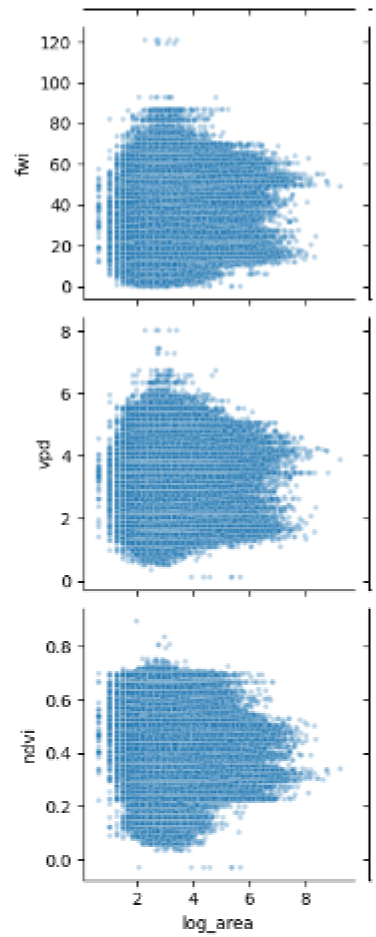
Data Science Process

Data Analysis

Data preprocessing involved cleaning, feature engineering, and scaling. Key steps included:

- Merging two datasets to join corresponding environmental variables with each fire event

- Handling missing values using country and year level seasonal averages
- Engineering temporal and frequency based features such as months since last fire and fire count per year.
- Scaling numerical features for linear regression



Exploratory analysis showed that weather and vegetation indicators showed weak relationships with fire size. Featuring engineering was done to compensate for this weak relationship.

Modelling

Four supervised classification models were tested:

1. Logistic Regression
2. Decision Tree
3. Random Forest

4. Gradient Boosting

The target variable was categorical and the classes were imbalanced:

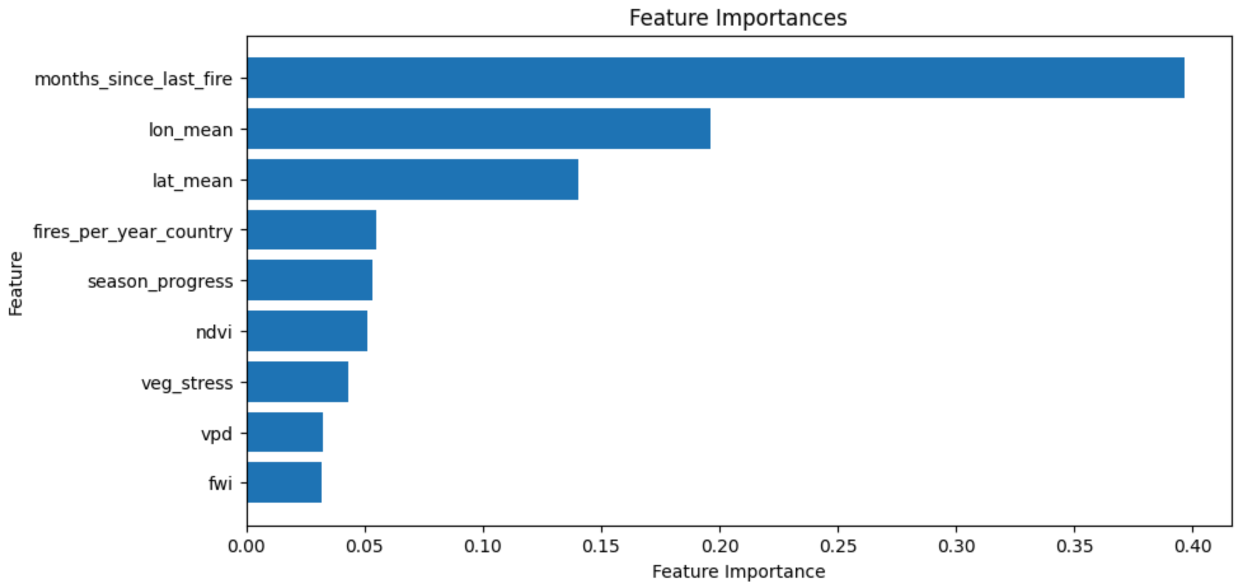
- **Small** (≤ 20 km²) - 57% of the data
- **Medium** (20–50 km²) - 37% of the data
- **Large** (≥ 50 km²) - 6% of the data

Top performing model:

Gradient Boost with a balanced class weight parameter performed reasonably, achieving moderate accuracy (52%) despite the limitations of aggregated data. This model was most successful at capturing the big fire class (71%) therefore minimizing the risk of false negatives. It still struggles to distinguish between small and medium fires as well as medium and large but showed fewer errors when predicting between small and large fires, suggesting potential for improved performance with more detailed weather data.

Feature Importance:
The most influential predictors were:

- Months since last fire
- Longitude and latitude
- Season progress
- Fires per year by country



Weather and vegetation features (FWI, NDVI, VPD, vegetation stress) contributed minimally, reflecting the coarse granularity of environmental data.

Outcomes

The model indicates potential for developing a reliable predictive system for wildfire severity, however, the current dataset, which is aggregated at the country-year level, was not specific enough to achieve strong predictive performance. Results suggest that temporal and locational patterns hold more predictive value than the generalized weather and vegetation indicators currently available.

Implementation

In a production environment, this model could be integrated into an early warning dashboard or decision support system. For operational use, several improvements would be necessary:

- Incorporate higher resolution weather and vegetation data (subnational or gridded)
- Automate real time data ingestion from satellite sources
- Regularly retrain the model as new fire data becomes available

---

Data Answer

The data question was partially answered. The analysis revealed that while historical fire patterns are valuable predictors, aggregated environmental data at the country-year scale limits predictive precision. Confidence in the data driven insights is moderate, constrained primarily by data availability.

Business Answer

The business question was answered satisfactorily within the limits of available data. The models provide early stage insights that could inform preparedness strategies, particularly in identifying areas or times of heightened risk. However, fine grained data would be needed before such a system could be operationally deployed.

Response to Stakeholders

The findings suggest that even with global, aggregated data, there is measurable structure in how fire frequency and seasonality relate to severity. Future improvements should focus on data resolution rather than algorithm complexity.

Key recommendations:

- Prioritize the integration of location specific weather data
- Explore regional model deployment before scaling globally
- Consider using the model as a risk screening tool to flag potentially severe ignitions

---

End-to-End Solution

The overall solution involves:

1. Collecting and cleaning global fire, weather, and vegetation data
2. Engineering temporal and geographic features related to fire recurrence and seasonality

3. Training and evaluating machine learning models (Gradient Boost)

4. Producing a predictive framework capable of estimating fire severity class at ignition

With further refinement, this approach could be integrated into environmental monitoring systems to enhance early warning and response planning.

References

- Pan, T. (2025). *Global wildfire exposure dataset (1.0.0)* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.15254187