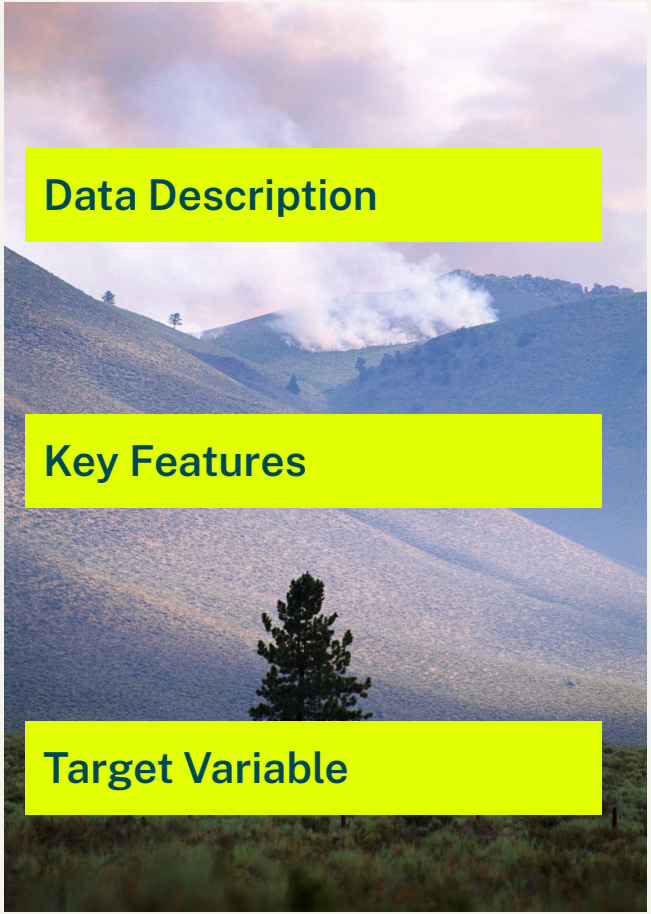
The Environmental Context of Fire Severity: Global Patterns & Insights

Brianna Klabunde / Institute of Data / November 2025

Can fire severity be predicted at the time of ignition?



About the Data

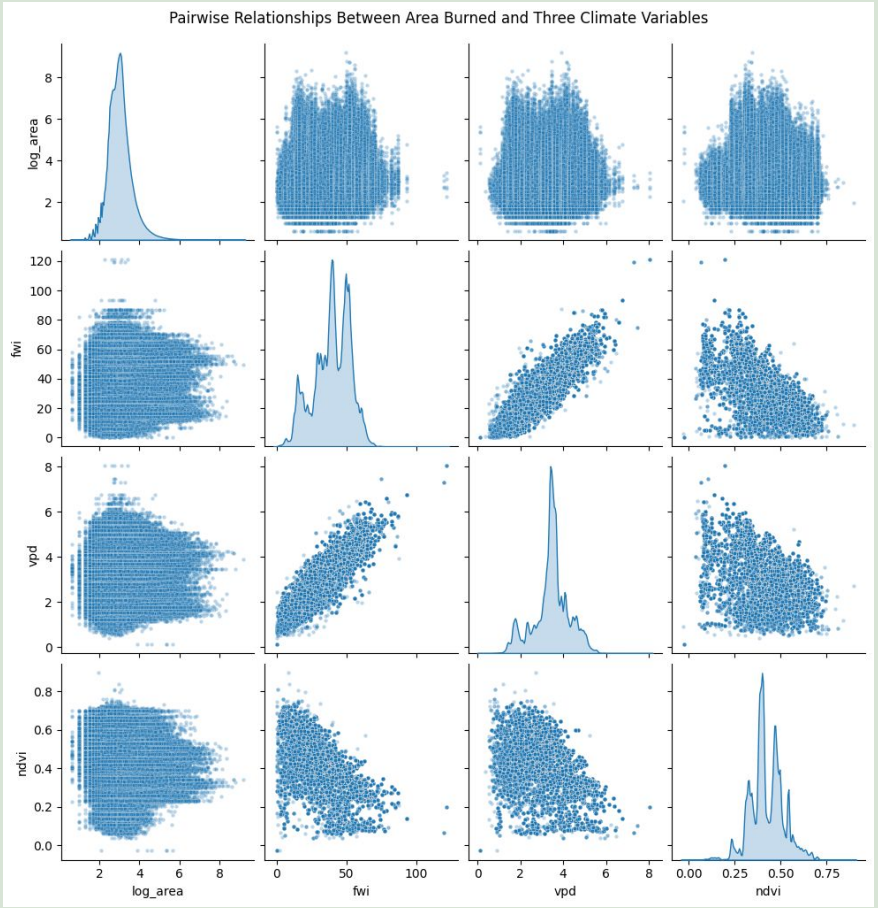
1. Fire clusters - derived from satellite data
2. Exposure - corresponding environmental data aggregated at a country-year level
3. Merged to include ~1.8 million global fire events between 2002-2020

- **Latitude / Longitude** – fire cluster coordinates
- **NDVI** (Normalized Difference Vegetation Index) – vegetation greenness -> fuel availability
- **VPD** (Vapor Pressure Deficit) – measures air dryness
- **FWI** (Fire Weather Index) – composite measure of fire danger (temp, humidity, wind, rain)

Severity - derived from burned area (small, medium, large) to enable classification modeling

Small fire <=20 km^2 - ~57%
Medium fire ~37%
Large fire >=50 ~6%

Weak Correlations



- FWI and VPD are strongly positively correlated (*0.8*) - both reflect hot, dry conditions
- NDVI is negatively correlated with VPD (*–0.31*) and FWI (*–0.61*)
  Together, these patterns show that drier air and lower vegetation health align with higher fire danger
- **NDVI, VPD, and FWI show low/no correlation with area**
- Motivates feature engineering

Feature Engineering

Months Since Last Fire

Calculated using 2° lat/lon grids to track local fire recurrence, avoiding uneven country sizes which miss finer local context

Season Progress

Represents where a fire falls within the year (adjusted by hemisphere)

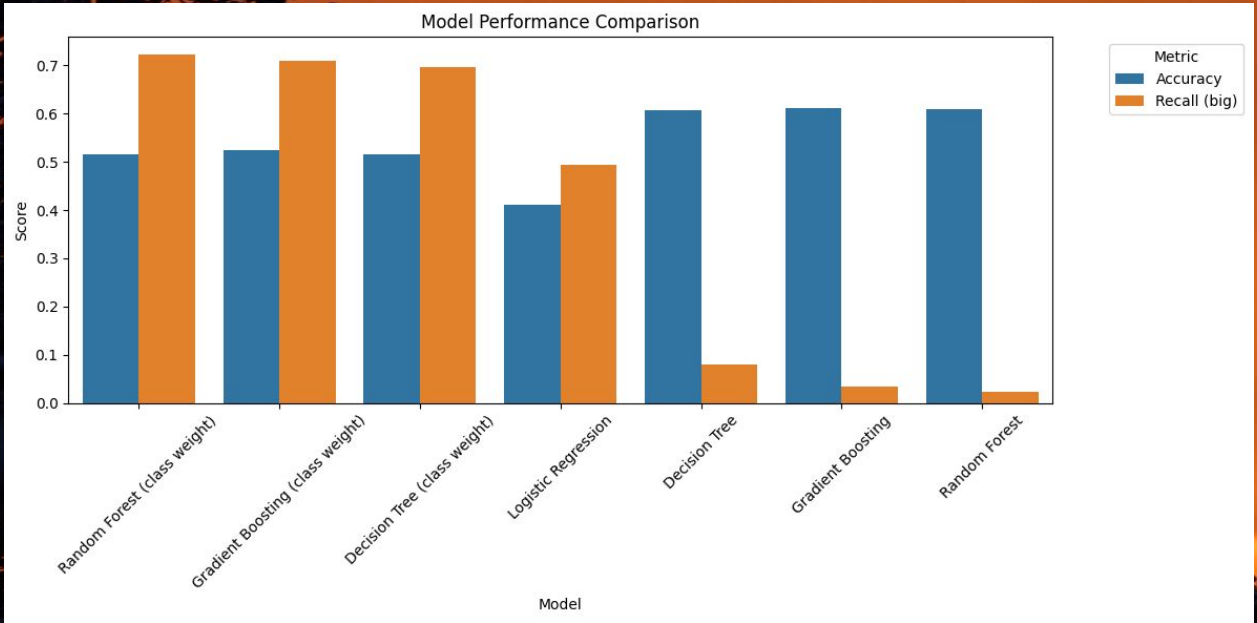
Vegetation Stress

Derived from NDVI and VPD to measure vegetation dryness (flammability)

Fires per Year by Country

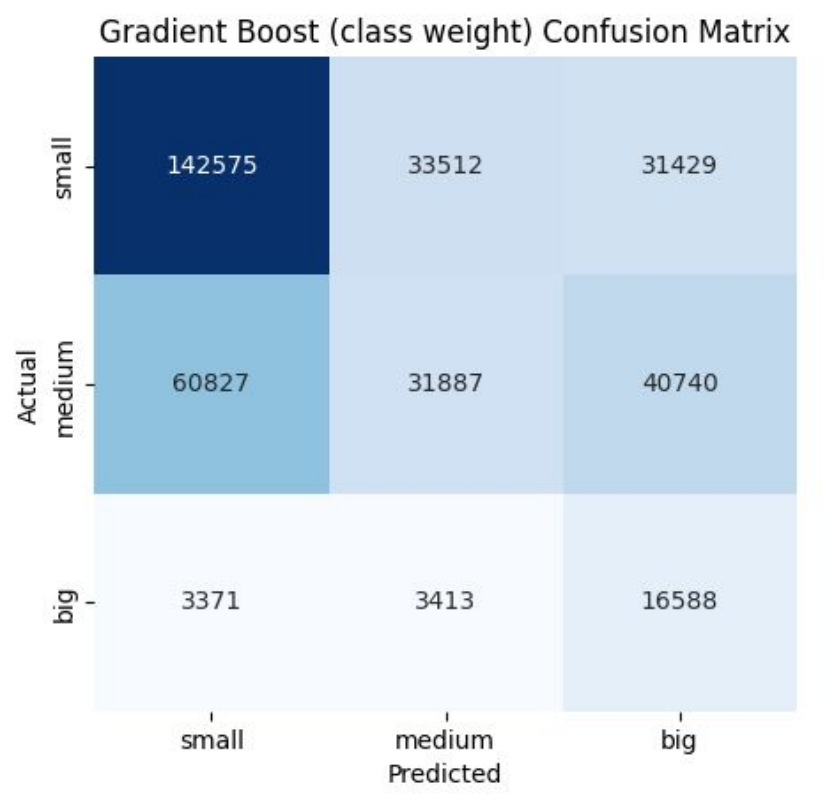
Measures annual fire frequency per country, providing broader contextual information for the model

Model Performance



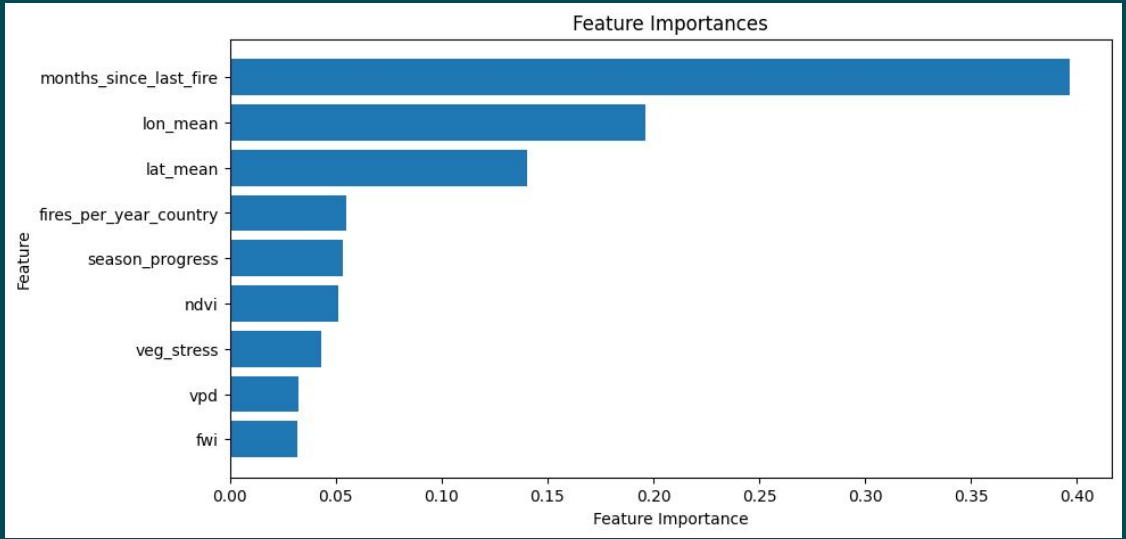
- Initial models tended to overpredict small and medium fires, rarely identifying big fires, while still showing high accuracy due to class imbalance.
- **Recall (big)** measures how well the model detects large fires specifically.
- Misclassifying large fires has greater consequences for human life, making recall more important than overall accuracy.
- After adjusting the models, detection of big fires improved, though this came with a small reduction in overall accuracy.

Gradient Boost (class weight balanced)



- The Gradient Boost model achieved moderate overall accuracy (~52%), similar to the other class balanced models
- Recall for *big* fires was high (71%) - meaning it successfully identified most of the large, high-impact fires
- However, this came with lower precision for big fires (19%) - the model often predicts “big” when the fire is actually medium
- Performance on the medium class remains weak in both directions
- Overall, the model is effective when the priority is catching large, high-consequence fires, even if it means more false alarms

Feature Importances



- Months since last fire was the strongest predictor of severity
- Location (lat/long) was also important, suggesting regional patterns in fire behavior
- Seasonal timing and fire frequency by country contributed moderately
- Weather and vegetation indicators (FWI, NDVI, VPD, veg stress) contributed very little to the model - possibly because they were aggregated by country-year

Conclusion

Can fire severity be predicted?

Recall	Gradient Boost	Better Weather Data	Further Study
For this project, maximizing recall on big fires is extremely important for preparedness and saving lives To improve recall we needed to address class imbalance	Gradient Boost was slightly better at correctly classifying medium sized fires, but more is needed before implementing this model.	The environmental data available was not sufficient enough for accurate prediction	Incorporate regionally specific weather information (wind speed, humidity, temp, drought, rain)

Thank You!

Questions?

References:
Data Source: Pan, T. (2025). Global wildfire exposure dataset (1.0.0) [Data set]. Zenodo.