**Hypothesis: Wildfire Duration**

We hypothesize that wildfire duration is primarily driven by the amount of prefire fuel and weather conditions that promote sustained burning. Fires are expected to last longer when more fuel is available, when fuels are drier, and when environmental conditions favor combustion and spread. The following variables from the emissions and weather datasets are expected to be key predictors, with their importance explained below:

**Emission Data:**

1. fuelload, prefire\_fuel – *Amount of fuel available before fire*
   * Importance: More fuel allows the fire to burn longer by supplying sustained energy.
2. fuel\_moisture\_class – *Categorical scale of fuel wetness*
   * Importance: Drier fuel classes (lower moisture) increase burn duration because ignition is easier and combustion is more efficient.
3. cwd\_frac, duff\_frac – *Fraction of coarse woody debris and duff consumed*
   * Importance: High consumption of ground fuels indicates sustained fire behavior and may correlate with longer burning fires.
4. area\_burned – *Total area burned in the event*
   * Importance: Larger fires often take longer to extinguish, though this can be both a cause and effect of long duration.
5. covertype – *Vegetation type*
   * Importance: Different vegetation types burn at different rates; forests may sustain longer fires than grasslands, for example.

**Weather Data:**

1. fm100, fm1000 – *Medium and large fuel moisture content*
   * Importance: Low moisture in larger fuels means deeper burning and longer fire persistence.
2. vpd (Vapor Pressure Deficit)
   * Importance: High VPD means vegetation is drying out, increasing the likelihood of prolonged burning.
3. vs (Wind Speed)
   * Importance: Strong winds spread fire and supply oxygen, helping it burn longer across new fuel sources.
4. tmmx, tmmn (Max/Min Temperature)
   * Importance: High temperatures reduce fuel moisture and preheat vegetation, aiding continuous fire behavior.
5. rmin, rmax (Min/Max Relative Humidity)
   * Importance: Low humidity dries fuels and air, increasing the chances of long-duration fires.
6. srad (Solar Radiation)
   * Importance: Increases surface heating and fuel drying during daylight hours.
7. etr, pet (Evapotranspiration Rates)
   * Importance: High evapotranspiration dries out vegetation, reducing live fuel moisture.
8. pr (Precipitation)
   * Importance: Less rainfall prior to a fire may mean drier conditions and more prolonged fires.
9. bi (Burn Index or Fire Danger Index)
   * Importance: Summarizes fire risk and flammability potential—higher BI typically correlates with longer and more intense fires.

* Add windirection based on professor feedback.

**Consideration of Fire Direction**

Wind direction can influence how wildfire spreads across the landscape. When the wind flows in the same direction as natural slopes or areas with abundant fuel, it can help the fire continue along that path, which may result in longer burn durations.

Because the time available for this project is limited, wind direction will only be included in the final analysis if it is feasible to do so. Incorporating wind direction requires tracking fire movement across space and time, aligning it with environmental factors such as terrain and fuel continuity, and accounting for changes in wind patterns. These steps add considerable complexity to the analysis.

If the data can be used effectively without delaying the main objectives, we will include wind direction in our modeling. Otherwise, it will be documented as a potential area for future research.

**Model Selection**

To predict wildfire duration and related behaviors, we plan to evaluate both traditional machine learning models and advanced deep learning approaches. This hybrid strategy is informed not only by our own dataset structure but also by recent advances in wildfire modeling, such as the work by **Shadrin et al. (2024)**, which successfully applied deep learning to predict fire spread using multimodal geospatial and meteorological data.

**Traditional Models**

* XGBoost and LightGBM will serve as strong baselines. They are highly effective for structured, tabular data such as ours, and they offer good performance with limited tuning.
* k-Nearest Neighbors (k-NN) will be tested for its ability to identify spatial patterns in historical fire behavior using local similarity.

These models are interpretable and computationally efficient, making them ideal for early-stage analysis and feature importance extraction.

**Deep Learning Models**

Inspired by **Shadrin et al. (2024**), we will explore deep learning models that can better capture the spatiotemporal dynamics of wildfire behavior:

* Convolutional Neural Networks (CNNs) are useful if we represent weather, land cover, and fire behavior data as raster images or spatial tensors. The study showed that CNN-based models, such as MA-Net, achieved strong F1-scores (0.64–0.68) in predicting fire spread over a 5-day horizon.
* ConvLSTM and LSTM models may be used if temporal sequences of fire and weather conditions are constructed. These models are well-suited for learning dependencies across time and space, particularly in estimating how fires evolve day by day.

In the article, the authors emphasized that wind direction and land cover were among the most important features. Their CNN-based MA-Net model was able to forecast wildfire spread direction and area with notable accuracy using a grid-based approach. This supports our consideration of CNNs and the integration of high-resolution meteorological and static land data into our modeling pipeline.

**References:**

Shadrin, A., Solovyev, R., Kustov, A., & Krivov, A. (2024). *Wildfire spreading prediction using multimodal data and deep neural network approach*. Scientific Reports, 14, Article 52821. <https://doi.org/10.1038/s41598-024-52821-x>