Predictive Analysis of Wildfire Risk Using Machine Learning

# 1. Problem Definition

**Objective**: Our primary goal is to develop a predictive model for classifying types of wildfires. In the process of achieving this goal, I expect to gain insights into the factors contributing to different types of wildfires. I will also conduct a formal feature importance analysis to specifically identify and quantify the impact of these factors on wildfire classification.

**Importance**: Wildfires have devastating impacts on biodiversity, human lives, property, and air quality. Predicting their likelihood can help in early warning systems, better resource allocation, and mitigating potential damages.

# 2. Data Collection and Refinement

**Data Sources**: The project will utilize satellite imagery, weather data (temperature, humidity, wind speed, etc.), historical wildfire records, and vegetation types from various publicly available datasets, such as the Global Historical Climatology Network (GHCN) and regional environmental monitoring databases.

I found some websites that offers the wildfire data:

* Kaggle: <https://www.kaggle.com/search?q=wildfire>
* Global Historical Climatology Network (GHCN): <https://www.ncdc.noaa.gov/ghcn-daily-description>
* Google Earth Engine: <https://earthengine.google.com/>

Consider my laptop, I cannot run the super large data set so I will focus on the regions with a high frequency of wildfires. Hope, we can use the small dataset to learn more factors and more detailed predictive model.

**Refinement**: Data preprocessing will include cleaning, normalization, and transformation. For some dataset, I will use the date and spatial features to joint two datasets together therefore, I can get more features to fit in my model. Feature engineering will be conducted to identify and create relevant variables for model training.

# 3. Implementation

**Method**: I will use a combination of machine learning techniques, including Random Forests, Gradient Boosting Machines (GBMs), and Convolutional Neural Networks (CNNs) for image data, to develop predictive t model. The approach will integrate historical data trends with real-time data to predict wildfire risks.

**Technology**: Python for data processing and model development, TensorFlow or PyTorch for deep learning models. Most of the code, I will use co lab as the platform.

# 4. Evaluation

**Metrics**: The model's performance will be evaluated using precision, recall, and the F1 score to balance the importance of false positives and false negatives. Additionally, the Area Under the Receiver Operating Characteristic Curve (AUROC) will be used for binary classification tasks. Alongside traditional performance metrics, I will also assess the environmental impact of our machine learning models. This involves calculating the carbon footprint of model training and inference processes. To achieve this, we will integrate the CodeCarbon library, an open-source software that estimates the amount of CO2 produced by the computing resources consumed.

**Validation**: The model will be validated using a split of training and test data, with further validation through cross-validation techniques to ensure its generalizability.

**Outcome**: The expected outcome is a reliable model that can predict wildfire risks with a high degree of accuracy, providing valuable insights for decision-makers in environmental protection agencies, firefighting units, and policy formulation.

**Significance**: Regardless of the result, the project will contribute to the understanding of wildfire and the potential of AI in environmental risk management. For me, my project aims to increase my knowledge about wildfires and explore how artificial intelligence can help manage environmental risks. Even if I don't get the results I hope for, I'll learn a lot along the way. My personal goal is to discover what causes wildfires by studying different factors. Then, using our AI model, we can figure out ways to prevent them. This is not just about building a model; it's about using technology to make a safer environment for us all.