Wildfire Prediction Based on Environmental Factors

Member 1	Guangkai Chen	guangkai22@vt.edu
Member 2	Kaiyi Chen	kennychen@vt.edu
Member 3	Chujia Chen	cjchen@vt.edu

Introduction

In recent years, the world has witnessed a surge in the frequency and intensity of wildfires. These catastrophic events, driven by a myriad of environmental factors, have wreaked havoc on ecosystems, displaced communities, and caused billions in damages. The unpredictability and rapid onset of these fires have made them especially challenging to manage. However, with the advancements in data science and machine learning, there lies an opportunity to harness vast datasets to predict these occurrences, facilitating early preventive measures. Our team is committed to developing a predictive model that factors in key environmental variables to anticipate the likelihood of wildfire events. By doing so, we aim to provide a valuable tool for forest departments, local communities, and policymakers, enabling them to take proactive steps and mitigate potential damages.

Project Problem Statement

Problem: Predicting the likelihood of wildfires based on environmental factors such as temperature, rainfall, humidity, and sunlight.

Stakeholders: Local communities, forest departments, insurance companies, government agencies, and environmentalists.

Data-Driven Approach Justification: Wildfires can cause massive destruction. Predicting them using data analytics can help authorities take preventative measures, potentially saving lives and reducing property damage. A data-driven approach would provide a systematic and reliable method to anticipate such events based on historical data and patterns.

Data Set

We will be leveraging multiple datasets, primarily sourced from renowned institutions like NREL and NOAA:

Solar Resource Maps from NREL
Solar datasets from NREL
Comparative Climatic Data from NOAA NCEI
Daily Climate Data from NOAA NCDC
Precipitation data from Weather.gov
Climate Data Form NOAA
Further datasets from NOAA and USGS as needed.

Features: These datasets will provide a comprehensive set of features including temperature, humidity, rainfall, sunlight, and other environmental factors. The exact feature count will be determined after collating and analyzing the data.

Harnessing the power of big data is crucial for the precision and reliability of our predictive model. We have identified several datasets, primarily from renowned institutions such as NREL and NOAA. Our preliminary estimates suggest that our consolidated dataset, after merging and processing, will comprise several hundred thousand records. Each record will represent specific geographical locations and timeframes, providing a comprehensive snapshot of environmental attributes juxtaposed with wildfire occurrences. The richness of our data lies in the multitude of features we plan to incorporate. By combining data from various sources, we anticipate a feature set spanning around 50-100 variables. These features will not only include the obvious suspects, such as temperature, humidity, and rainfall, but also delve deeper into factors like solar radiation, wind patterns, historical fire incidents, and other environmental indicators. Each feature will be meticulously analyzed for its potential predictive power, ensuring a robust model.

Preprocessing steps

- Data Cleaning: Removing inconsistencies, handling missing values, and eliminating outliers.
- Data Merging: Combining data from multiple sources to create a unified dataset.
- Feature Engineering: Generating new features that could be relevant for prediction.
- Data Normalization: Scaling features to ensure they have equal significance in models.
- Data Transformation: Converting non-numeric data into a format suitable for machine learning.
- Methods and Models
- Machine Learning Problem: Supervised learning - Classification
- Target Variable: Whether a wildfire is going to occur or not.

Models to be Used

- Decision Trees
- Logistic Regression
- Support Vector Machines
- Random Forests
- Gradient Boosting Machines
- Other methods such as Principal Component Analysis (PCA) might be used for dimensionality reduction if required.

Conclusion

Our project aims to provide a reliable and systematic approach to predicting wildfires. The insights derived from our models can be used to take proactive measures, thereby reducing the impact of such disasters. By comparing our predictions with traditional methods, if any, we aim to demonstrate the efficacy and importance of a data-driven approach in mitigating natural disasters.