

Weather Patterns and Wildfire Analysis

Matthew Presti
Holly Schwecke

Description

We will be analyzing the relationship between weather patterns and wildfire activity in the western United States using weather data from NOAA and NASA MODIS satellite wildfire data. We seek to identify causation factors in the wildfire activity and perhaps develop predictive insights for fire season severity as well as search for predictive trends related to periods of extreme winds both in the presence and absence of wildfire to ideally reveal novel insights for wildfire prediction and mitigation.

Research Questions

1. Can we identify causation factors in years of elevated wildfire activity based on weather data trends?
2. Can we predict periods of extreme winds that pose elevated fire risks?
3. What trends could be valuable for wildfire prediction or forecasting fire season severity?

Prior Work

NOAA Daily Summaries

- Source: National Oceanic and Atmospheric Administration
- <https://www.ncdc.noaa.gov/cdo-web/datasets/GHCND/locations/CLIM:109/detail>

USA Wildfire Dataset

- Source: NASA MODIS Satellite data
- <https://www.kaggle.com/datasets/avkashchauhan/california-wildfire-dataset-from-2000-2021>

Proposed Work

Preprocessing

1. Weather Data Cleaning:
 - a. Handle missing values
 - b. Standardize measurements
 - c. Filter relevant geographical regions
2. Wildfire Data Preprocessing
 - a. Normalize coordinates
 - b. Validate fire detection data
 - c. Filter for relevant time periods
3. Data Integration
 - a. Align temporal resolutions
 - b. Match geographical coordinates
 - c. Create unified dataset structure

Dataset 1:

NOAA Global Surface Summary of Day (GSOD)

Dataset Overview

- **Source:** National Climatic Data Center (NCDC) & USAF Climatology Center
- **Coverage:** Over 9,000 weather stations
- **Time Range:** 1929-present, updates daily
- **Size:** Multiple millions of records (3.3GB file)

Weather Data Elements

- Temperature (mean, max, min)
- Wind (speed, gusts, sustained)
- Precipitation & Snow depth
- Pressure
- Visibility & Dew point
- Weather events (fog, rain, snow, hail, thunder)

Format

- CSV structure
- Machine-readable format
- Tar Archive

Dataset 2:

NASA MODIS satellite

Dataset Overview

- **Source:** NASA MODIS satellite data (via Kaggle)
- **Coverage:** 2000 - March 25th, 2022
- **Geographical Range:**
 - Latitude: 32° 30' N to 42° N
 - Longitude: 114° 8' W to 124° 24' W
- **Size:** 7.8 million data points

Comprehensive geographic wildfire data

Format

- CSV structure
- Machine-readable format

List of Tools

- Python
 - Pandas, numpy, scikit-learn
- Jupyter Notebooks and/or VS Code
- GIS tools
- Data visualization libraries
 - Matplotlib, seaborn
- Version control
 - GitHub
- Project management
 - Trello

Evaluation

Techniques:

- Accuracy
 - Overall proportion of correct predictions made by a model
- Precision
 - Proportion of positive predictions that are actually correct/relevant
- Recall
 - Proportion of actual positive cases that the model correctly identifies

Method:

- Cross-Validation
 - Divide the data into multiple folds
 - Train the model on each fold while using the remaining folds for validation
 - Average the results to get a more robust evaluation
 - Provides a more robust estimate of model performance by training and evaluating one multiple data splits, compared to the holdout method