Photo by Karsten Winegeart on Unsplash

Predicting Wildfires with Weather Forecast Data. - Initial Report

S B S Tarun Reddy – AM.EN.U4CSE19249
Manas P P – AM.EN.U4CSE19233
S S S B Shravanth – AM.EN.U4CSE19250
Satya Sai Rakesh – AM.EN.U4CSE19240

Course: CSE19304 Foundations of Data Science

Problem Statement

Australia experienced one of the worst wildfire events in the years of 2019 and 2020. With an estimated 3 billion animals affected, the wildfires also destroyed homes of many people leading to huge human costs. Although in many ecosystems wildfire is a part of a natural cycle, poor historic forest management and climate change has led to more intense fires each year. Being prepared for these fires is the first step in combating them. However, acting only on the detection of smoke or the starting of a blaze could prove to be too late.

In order to be better prepared for these fires, we have to explore the possibility of forecasting wildfires using the available historic weather forecast, vegetation index and land class data. A number of weather characteristics like precipitation, relative humidity, temperature, wind speed is known to play a role in the spread of forest fires. For example, winds can play a role in blowing the flaming materials to regions of forest that haven't caught fire.

The type of vegetation in a region, like the percentage of area covered with shrubs, cultivated vegetation, closed forests, etc. and the percentage of urban cover is known to lead to varying intensities of wildfires. However, using these as variables to predict the area of possible wildfires in the future can tremendously help in the preparation for these wildfires.

In this project, we attempt to achieve this by carefully studying the data collected from Australia and using tools from data science to establish a relationship between the area of wildfires and the various features observed in our data that broadly fall under the categories of weather characteristics, land coverage and vegetation index.

Data Source

The data used for this project has been taken from IBM's Call for Code Spot Challenge for Wildfire. The data from the challenge is available for open-source use at https://github.com/Call-for-Code/Spot-Challenge-Wildfires.

Data Overview

There are five datasets available to accomplish our task:

- Wildfire.csv : There is data from seven regions in Australia where the fire area was estimated using the pixels of image from satellite data.
- HistoricalWeather.csv: There is data from the seven regions and it gives us information on precipitation, soil water content, relative humididty, etc.
- HistoricalWeatherForecasts.csv: This includes the same features as the dataset above but is the forecasted values.
- VegetationIndex.csv: This is the monthly data for the normalized differential vegetation index in the seven regions.
- LandClass.csv: This includes the percentages of shrubs, water bodies, urban buildup, etc in the seven regions.

Literature Review

1) Climate change had doubled Wetsern U.S. forest fires. https://www.earth.columbia.edu/articles/view/3343

2) Fighting wildfires https://www.bbc.com/news/world-50410481

3) How will California prevent more mega-wildfire disasters? https://www.nationalgeographic.com/science/article/how-will-california-prevent-more-mega-wildfire-disasters

4) An overview of the data. https://github.com/Call-for-Code/Spot-Challenge-Wildfires/blob/main/data/Readme_Docs_Wildfires-Datasets_2020-11.pdf