Predicting the Frequency of Natural Disasters Due to Climate Change

Astha Rai, Lam Tran

June 15, 2023

Abstract

In our daily lives, we are beginning to experience the unprecedented effects of climate change, primarily attributed to the alarming rise in natural disasters. A notable instance occurred late last year when California suffered significant damage due to a surge in storm occurrences. This study aims to employ time series analysis and feature extraction techniques to quantitatively investigate the impact of climate change-related features on the incidence of floods and wildfires. Specifically, we have selected climate change-related features and collected data across 23 counties in California spanning from 2001 to 2020. Through the time series analysis, we tracked the variations in each feature over the years. Subsequently, utilizing the LASSO technique for feature extraction, we identified the features that exerted the most influence on the frequency of wildfires and flooding. Our findings indicate a positive relationship between the frequency of flooding and the increase in average precipitation. Additionally, average temperature and the loss of tree coverage exhibit a positive association with wildfires, while average precipitation demonstrates a strong negative correlation with wildfires. These results enable us to predict the future frequency of wildfires and flooding caused by climate change and explore potential strategies to mitigate or prepare for such events.

1 Introduction

The effects of climate change are increasing over time, leading to various problems such as food shortages, rising sea levels, and an escalation of natural disasters. California, in particular, faces a higher risk of climate-related hazards like wildfires and flooding. These hazards pose a threat to public health and safety through life-threatening events, damage to public property and infrastructure, and depletion of natural resources. The objective of this paper is to enhance our understanding of the complex interplay and interaction among multiple factors influencing climate change.

To achieve this, we will employ feature extraction techniques coupled with linear regression to effectively identify the most significant features contributing to the causes and effects of climate change. This approach will enable us to make accurate predictions regarding the frequency of flooding and wildfires in California. In Section 2, we will outline the data selected for analysis. Subsequently, in Section 3, we will describe the methods and code implementation, including code validation. Section 4 will be dedicated to analyzing the results, with a specific focus on the key features driving natural disasters. Finally, in Section 5, we will summarize the findings and discuss potential future directions for further experimentation.

2 Model

	County	Date	Tree Coverage Loss	Avg Temp	Avg Precipitation	Ozone	Nitrogen Dioxide	Floods	Wildfire
1	Alameda	2001	60	59.6	22.27	0.087	0.04	0	0
2	Alameda	2002	15	59	18.34	0.106	0.04	2	0
3	Alameda	2003	62	59.7	16.43	0.094	0.035	0	0
4	Alameda	2004	69	59.6	18.35	0.08	0.034	0	0
5	Alameda	2005	36	59.4	24.52	0.09	0.034	1	0
6	Alameda	2006	46	59	21.35	0.101	0.037	1	0
7	Alameda	2007	17	58.9	11.57	0.091	0.035	0	0
8	Alameda	2008	15	59.1	14.18	0.11	0.037	0	0
9	Alameda	2009	18	59	17.32	0.086	0.029	2	0
10	Alameda	2010	16	58.3	23.17	0.097	0.031	1	0
11	Alameda	2011	13	57.9	14.53	0.084	0.032	0	0
12	Alameda	2012	69	59.3	20.78	0.09	0.029	1	0
13	Alameda	2013	6	59.5	5.08	0.077	0.03	0	0
14	Alameda	2014	57	62	19.88	0.08	0.034	4	0
15	Alameda	2015	9	61.1	10.39	0.081	0.024	1	0
16	Alameda	2016	18	60.4	22.55	0.085	0.02	2	0
17	Alameda	2017	22	60.5	24.97	0.086	0.029	20	0
18	Alameda	2018	5	59.8	16.57	0.078	0.036	17	0
19	Alameda	2019	14	59.7	22.29	0.078	0.027	28	9
20	Alameda	2020	63	61.7	8.42	0.077	0.026	12	7
21	Contra Costa	2001	75	60.5	22.8	0.102	0.025	0	0
22	Contra Costa	2002	62	59.9	18.29	0.096	0.026	1	0
23	Contra Costa	2003	32	60.5	16.01	0.082	0.025	0	0
24	Contra Costa	2004	39	60.4	18.97	0.081	0.021	0	0
25	Contra Costa	2005	82	60.1	25.25	0.077	0.022	1	0
26	Contra Costa	2006	66	59.8	20.63	0.09	0.027	1	0
27	Contra Costa	2007	30	59.9	11.26	0.078	0.034	0	0
28	Contra Costa	2008	23	60.2	14.29	0.09	0.019	0	0
29	Contra Costa	2009	127	59.9	17.83	0.094	0.02	1	0
30	Contra Costa	2010	26	59	24.44	0.086	0.02	3	0
31	Contra Costa	2011	16	58.7	15.69	0.078	0.022	0	0
32	Contra Costa	2012	40	60.2	23.25	0.087	0.02	2	0
33	Contra Costa	2013	143	60.7	5.27	0.075	0.019	0	0
34	Contra Costa	2014	125	62.7	21.07	0.071	0.02	0	0
35	Contra Costa	2015	19	62	9.28	0.072	0.017	0	0
36	Contra Costa	2016	15	61.2	22.55	0.08	0.014	0	0
37	Contra Costa	2017	23	61.5	26.84	0.071	0.021	11	0
38	Contra Costa	2018	24	61	17.52	0.078	0.024	8	0
39	Contra Costa	2019	20	60.6	23.29	0.072	0.018	11	9
40	Contra Costa	2020	67	62.8	7.23	0.085	0.016	2	7
41	Fresno	2001	921	56.4	23.21	0.101	0.045	4	5
42	Fresno	2002	562	55.9	15.55	0.113	0.046	5	1
43	Fresno	2003	502	56.5	16.46	0.106	0.044	1	2
44	Fresno	2004	760	55.8	18.64	0.099	0.042	0	11
45	Fresno	2005	1250	55.7	24.86	0.091	0.041	9	11

Figure 1: Data set from 23 counties in California from 2001-2020 include: loss of tree coverage, average temperature, average precipitation, ozone levels, nitrogen dioxide levels, frequency of floods and wildfires throughout the year

We collected our data from 23 counties in California spanning the years 2001 to 2020, assuming that these counties would provide a representative sample of the entire state. The features we analyzed are closely linked to and influenced by climate change. These features include the loss of tree coverage, average temperature, average precipitation, nitrogen dioxide levels, ozone levels, and the frequency of wildfires and floods throughout the year. Our data is sourced from multiple reliable sources, namely NOAA, Global Forest Watch, and CARB. We utilized Global Forest Watch to gather data on the loss of tree coverage, while the NOAA database provided us with data on average temperature and precipitation. CARB supplied us with data on nitrogen dioxide levels, and we focused on these emissions since the data for other gases were either unavailable within our desired time frame or not accessible for specific counties. By relying on NOAA storm events data, we obtain information on the frequency of wildfires and flooding. After collecting and merging the data from these sources, our resulting matrix comprised the county and year as the first two columns, with the mentioned features distributed across the remaining columns.

3 Method

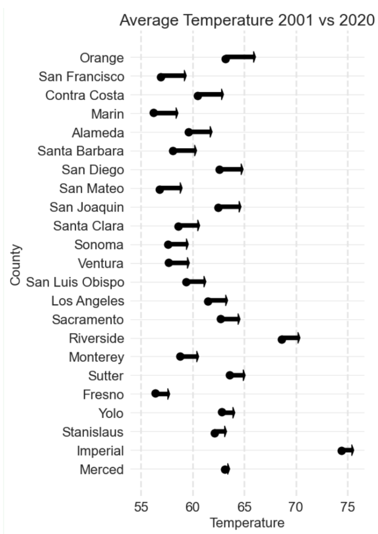
3.1 Code

We began by generating individual time series for each feature, allowing us to analyze their temporal variations and identify any potential similarities in trends across different features. To specifically examine the temperature changes between 2001 and 2020, we employed an arrow graph that visualizes the average temperature fluctuations. In addition, we utilized a heat map representation of the time series data, which provided us with a comprehensive view of the range of values for features such as wildfires, floods, forest loss, average precipitation, and others, spanning from low to high values. For assessing the stability or dynamism of ozone and nitrogen dioxide, we utilized a line plot to visualize their respective patterns.

For feature extraction, we employed LASSO regression to ascertain the features that exerted the most influence on the frequency of natural disasters, namely flooding, and wildfires. The regression gave us predictor coefficients for each of the features. For any feature that LASSO determines to be insignificant, it will set the coefficient to 0. To get the importance of each feature, we took the absolute value of the coefficients returned. Using the identified significant features, we plotted them against their respective natural disasters to gain a deeper understanding of their correlation. To validate our code and the generated coefficients, we ran the regression using Elastic Net which is a combination of LASSO regression and Ridge regression to see that the coefficients generated were close to ones we got from the LASSO regression. We also checked with random forest regression, although how it judges importance is different from LASSO.

4 Results

4.1 Time Series



(a) The average temperature (minimum temperature of 56.4°F, the maximum of 75.4°F) for 23 counties in California in 2001 versus 2020. The dots show 2001 temperatures. The arrow shows 2020 temperatures. The arrow length shows the difference between 2001 versus 2020. The order of the county represents the average temperature difference from highest to lowest. Data source: NOAA

(b) The frequency of wildfires for 23 counties in California from 2001 to 2020. The color from light to dark represents the frequency of wildfires from 0 to 34. Data source: NOAA

Figure 2: The average temperature and the wildfire

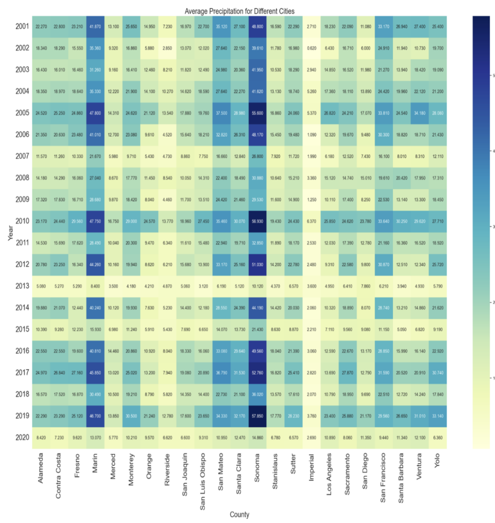
As time progresses, a consistent trend of increasing average temperatures becomes evident across California's counties. Among them, Orange County stands out with the highest recorded change in average temperature. This rise in temperature coincides with the county's unfortunate annual occurrence of wildfires. Similarly, San Diego and Contra Costa counties also grapple with ongoing wildfire incidents. Notably, the average temperatures in these counties rank among the top seven in terms of the highest temperature changes.

It is worth mentioning that Fresno experiences persistent wildfires as well. However, despite these fire events, the average temperature in Fresno remains relatively stable, ranging from 56.4 °F to 57.6 °F. This observation highlights a unique dynamic in Fresno, where wildfires persist while the average temperature exhibits stability.

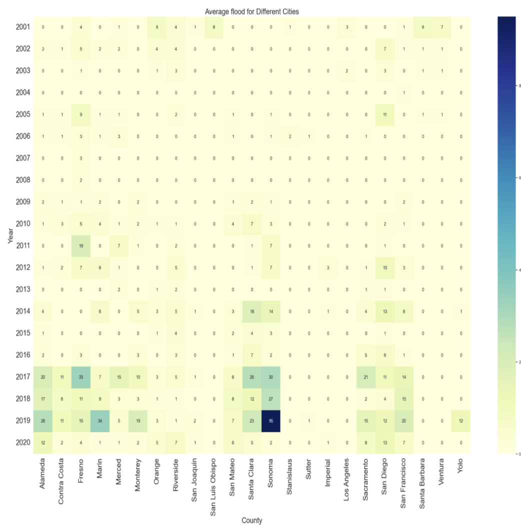
In contrast, several other counties in California are not susceptible to the destructive impacts of wildfires. However, they are not exempt from the overall trend of increasing average temperatures, which signifies the broader impact of climate change across the state.

Flooding stands out as one of the most significant climate hazards faced by California. Periodic extreme precipitation events can lead to destructive floods, posing serious challenges to the affected areas. Figure 3a provides valuable insights into precipitation patterns in Sonoma and Marin counties, which emerge as notable contributors to the state's abundance of precipitation. In 2019, Sonoma experienced its highest recorded precipitation, resulting in a substantial 95 floods. Similarly, Marin received plentiful precipitation during the same year, triggering 34 floods.

Analyzing the data from 2017 to 2020, we observe an upward trend in flood



(a) The average precipitation measured in inches for 23 counties in California from 2001 to 2020. The color from light to dark represents light to heavy precipitation. Data source: NOAA

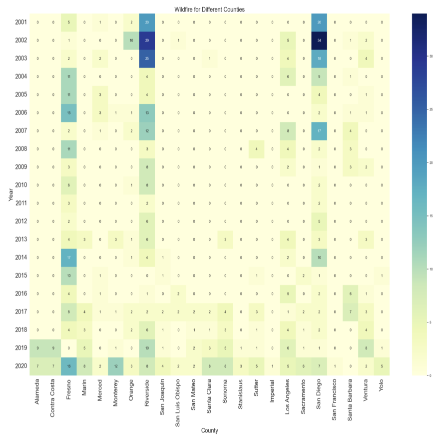


(b) The frequency of floods for 23 counties in California from 2001 to 2020. The color from light to dark represents sporadic to regular floods. Data source: NOAA

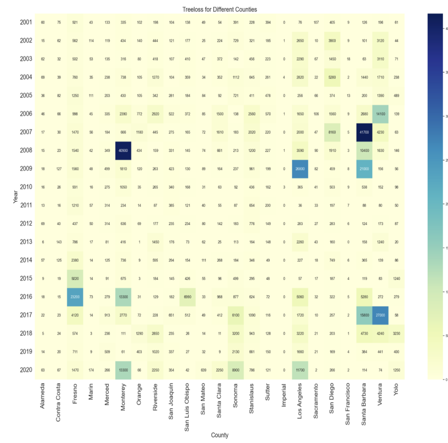
Figure 3: The average precipitation and the flood

frequency for both Marin and Sonoma counties, directly correlated with the increasing precipitation levels. However, this trend witnessed a decline in 2020. In contrast, Imperial County exhibited the lowest occurrences of floods, aligning with its comparatively lower levels of precipitation. By examining the interplay between precipitation and flood frequency, we gain valuable insights into the complex relationship between extreme weather events and their consequences.

The tree loss report in California unveils a disheartening reality, exposing the far-reaching consequences of various factors on the state's forest ecosystems. In recent years, the relentless onslaught of devastating wildfires and prolonged droughts has subjected California's trees to unprecedented challenges. By examining the relationship between wildfires and tree loss, we can delve into the underlying causes of the alarming tree loss crisis gripping the state.



(a) The frequency of wildfires for 23 counties in California from 2001 to 2020. The color from light to dark represents the frequency of wildfires from 0 to 34. Data source: NOAA

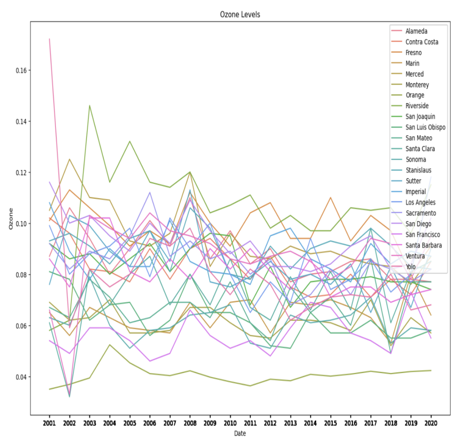


(b) Forest loss measures in ha for 23 counties in California from 2001 to 2020. The color from light to dark represents the frequency of wildfires from 0 to 41700. Data source: Global Forest Watch

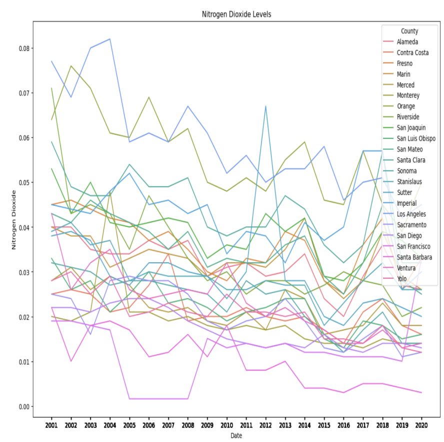
Figure 4: Wildfire and forest loss

Figure 4b presents a compelling picture of Monterey's deforestation, reaching a staggering peak of 40,900 hectares in 2008. Astonishingly, this substantial deforestation surge occurred in the absence of any reported wildfires during the same year. Likewise, Santa Barbara experienced its own peak of 41,700 hectares in 2007, also devoid of documented wildfire occurrences. However, it is worth noting that while 2002 did not witness the highest peak of wildfires in the dataset, it left a significant mark on San Diego due to its notable wildfire activity. A total of 34 wildfires were reported during that period, making 2002 a standout year for wildfire incidents in the region.

These observations shed light on the distinctive nature of deforestation events, which, in some cases, do not align with the typical association with wildfires. Such findings emphasize the need to delve deeper into additional factors that contribute to forest loss in these specific regions. Within this section, we explore the factor of wildfire that has been identified as a significant contributor to forest loss. In the subsequent section, we delve into an examination of the factors that contribute to forest loss in California.



(a) Ozone measured in parts per million (ppm) for 23 counties in California. Data source: CARB



(b) Nitrogen dioxide measure in parts per million (ppm) for 23 counties in California. Data source: CARB

Figure 5: Ozone and Nitrogen Dioxide

This part focused on the relationship between nitrogen dioxide (NO2) and tropospheric ozone (O3) levels, both of which play a role in climate change. It is well-established that NO2 has direct and indirect effects on climate change, whereas tropospheric ozone directly influences climate patterns.

The analysis revealed a notable increase in nitrogen dioxide levels, displaying a chaotic pattern with significant variations observed between different counties. Strikingly, Los Angeles exhibited the highest concentrations of NO2, while its ozone levels remained relatively stable. Similarly, Orange County had the second-highest nitrogen dioxide levels, but paradoxically, it showed the lowest ozone concentrations among the counties studied. These observations indicate a lack of correlation between nitrogen dioxide and ozone levels.

These findings highlight the intricate nature of the relationship between nitrogen dioxide and tropospheric ozone and their respective impacts on climate change. Although nitrogen dioxide is known to contribute to the formation of tropospheric ozone, the stability of ozone levels in the face of varying nitrogen dioxide concentrations suggests the involvement of additional factors.

4.2 Feature Extraction

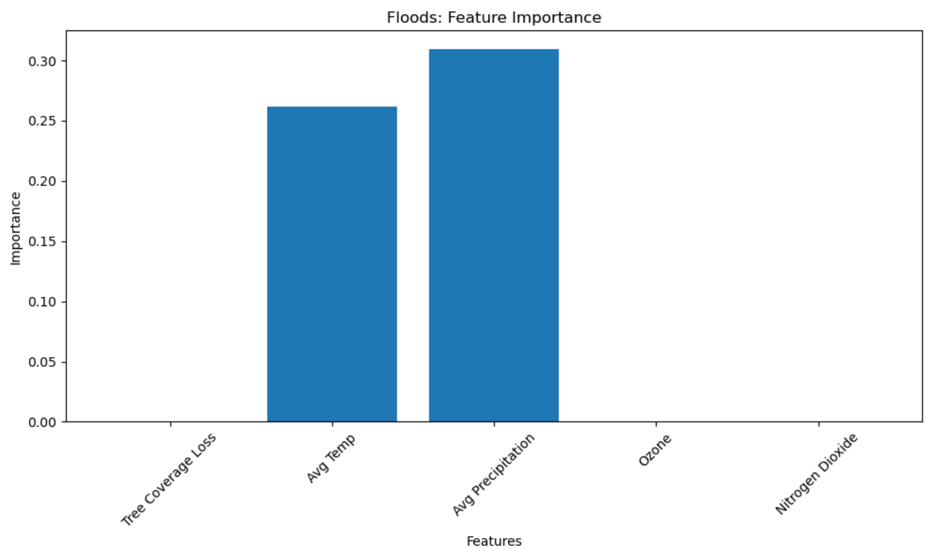


Figure 6: On the bottom is a table containing predictor coefficients found through LASSO, while the top is a bar chart showing the importance of the features for floods. The measure of importance is along the y-axis, while the features are on the x-axis. In this case, average temperature and average precipitation are the most important features.

Feature	Coefficients
Tree coverage loss	2.19206301e-05
Average temperature	1.58010466e-01
Average precipitation	2.48556571e-01
Nitrogen dioxide	0
Ozone	0

In Figure 6, we see the features with the most important for flooding obtained from LASSO regression were average precipitation and average temperature. Average precipitation and flooding have a strong positive correlation, which was to be expected. One of the main causes of rivers flooding is heavy rain, therefore this result makes sense. Average temperature and flooding also had a positive relationship. In addition to precipitation, another major reason flooding occurs is melting ice and snow, since these are usually the sources for rivers. This is caused by high temperatures that can melt snow caps and ice on mountains leading to rivers to overflow. Loss of tree coverage and flooding had a very minimal positive relationship, as trees can occasionally help hold back floodwaters, so less trees could cause flooding to occur quicker. The greenhouse gas levels were determined to be relatively unimportant with coefficients of 0, by the LASSO regression for causing flooding, since they mainly influence air quality which does not have much relation with flooding.

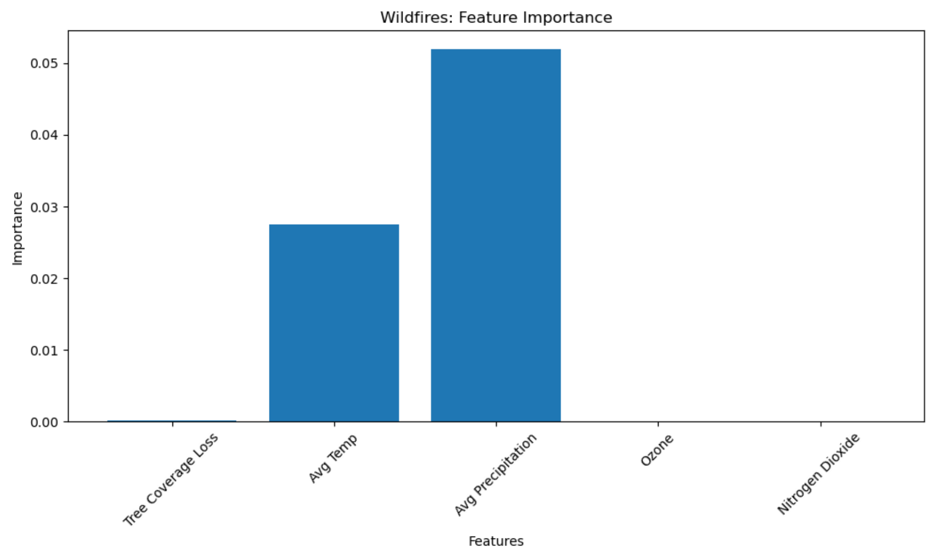


Figure 7: Similar to Figure 6, there is a table for predictor coefficients and a bar chart for feature importance, where average temperature and average precipitation are the most important features for wildfires.

Feature	Coefficients
Tree coverage loss	0.00015721
Average temperature	0.01757961
Average precipitation	-0.06140657
Nitrogen dioxide	0
Ozone	0

Figure 7 depicts the same thing as Figure 6 but for wildfires, although the coefficients are a bit larger. The most important features according to coefficients found through LASSO regression and seen from the bar chart were: tree coverage loss, average precipitation, and average temperature. Average temperature and wildfires have a strong positive relationship. This outcome makes sense since wildfires tend to occur more often due to heat waves. Higher temperatures can also lead to drought and a very dry climate which makes it very easy for wildfires to start and spread. Loss of tree coverage and wildfires also had a positive relationship, although it is difficult to see on the bar chart. This seems a result of the feature extraction since although tree coverage loss may not affect wildfires, wildfires do cause a lot of tree coverage loss. Average precipitation and wildfires have a very strong negative relationship, with it being deemed the most important factor. Rain and snowfall are factors that prevent wildfires from spreading since it leaves the ground moisturized. Once again, the greenhouse gas levels were deemed unimportant by the LASSO regression, but similarly to loss of tree coverage, wildfires may have more of an effect on air quality rather than the other way around.

Average temperature and average precipitation were the two features that had the largest predictor coefficients for both flooding and wildfires. These two features are influenced by changing weather patterns the most out of the features we collected data on, so it makes sense that they would in turn have a heavy importance when it comes to natural disasters. Although worsening air quality is an effect of climate change and is harmful, we found it has little to do with natural disasters. Loss of tree coverage is one of the main factors of climate change as it lessens the consumption of carbon dioxide in the atmosphere, which is the main driver of climate change. As both temperature and precipitation are effects of climate change, it demonstrates how it is important to do our best to slow global warming to reduce drastic weather changes that could lead to more frequent and intense flooding and wildfires.

4.3 Correlation Matrix

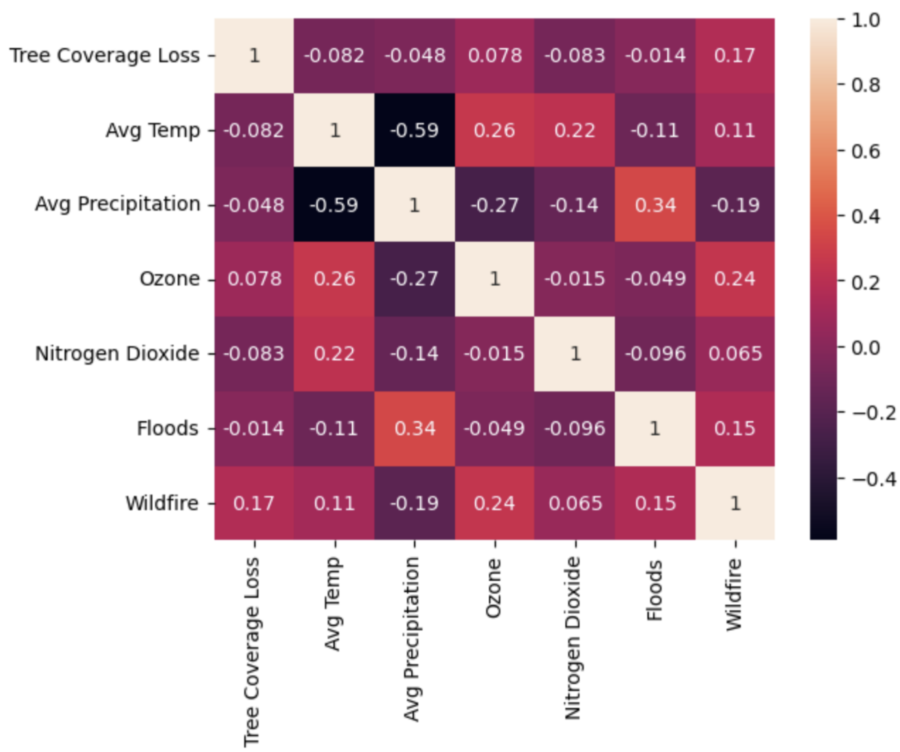


Figure 8: Correlation matrix relating all the features we collected to each other.

In Figure 8, we used a heat map with a correlation matrix to see how all the other features relate to each other outside of feature extraction. The darker squares indicate a strong negative correlation, while the lightest squares show a strong positive correlation. The most obvious result is the strong negative relation between average temperature and average precipitation, which would make sense in most places in California. Although there could always be other factors that affect this, such as location: rain forests and deserts both have high temperatures, but there is a drastic difference in rainfall. However, for the parameters of our research, we can see that as temperature increases, rainfall decreases. As mentioned before, wildfires also had a relatively strong positive correlation with greenhouse gas emissions as the smoke from fires would worsen the air quality. This is also the case for loss of tree coverage and wildfires, since wildfires burn down forests often if they spread. In turn, loss of tree coverage is a major factor for climate change. The majority of the features we have share a relatively strong relationship with each other, which demonstrates how all the features linking to climate change are related to each other and cause a vicious cycle.

5 Conclusion

We collected data on features that contributed to the cyclic nature of climate change in order to observe their relation to and impact on the frequency of natural disasters in California to predict the future frequency of these natural disasters. For flooding, we found a strong positive correlation between average precipitation and flooding, while for wildfires we found a strong positive correlation between average temperature, tree coverage loss, and wildfires. These results indicate that as global warming worsens there is a predicted rise in temperature leading to increasingly volatile weather patterns that may cause an increase in natural disasters.

When collecting data for some features, we were constrained by the lack of continuous and long-term data available online. Therefore our results could be more complete had we obtained some more features that would factor into these natural disasters due to climate change. A few factors may have been: snowfall, water levels, and wind levels. We could also have taken a look at rising sea levels and their effect on another natural disaster: storming, but since sea levels are only recorded for coastal counties we would not have enough data to compare it against the features we obtained.