Data Science for Localized Climate Prediction Product

Hailee Schuele, Irene Widiaman, Eesha Thaker, Yongzhen Shi James, Farouk Ghandour, Danielle Ma

Abstract

While the heat stress exposure urges the Utilities management to provide extra energy supply to assure AC usage and keep people cool, the wildfire risk points them to induce planned power outages, so fire ignition due to energy grid failure is prevented. This tool will help Energy Utilities to decide how to distribute the energy, and where to apply planned power outages according to vulnerabilities of the communities.

Introduction

In this project, we are developing decision-making tool to help Energy Utilities in deciding how to distribute the energy in situations like extreme weather or fire weather danger. We are focusing on the 6 CalFire locations: San Diego, Mariposa, Mendocino, Santa Barbara, Shasta, and Alameda. We also used public datasets to further analyze the best mitigation measures for areas that are prone to climate danger.

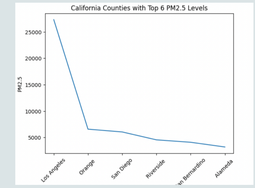
Data

Datasets used to build the forecasting models are historical and future exogenous data. These datasets contained hourly time series data of fire-weather variables, such as temperature and humidity.

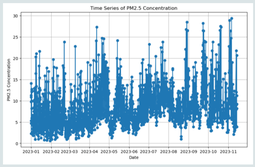
Datasets used to do Exploratory Data Analysis (EDA) to analyze 6 CalFire locations and their pattern with climate danger is from CalEnviroScreen. Time-series datasets for PM 2.5 concentration, wildfire-emitted airborne particulate matter of diameter ≤ 2.5 μm, is from the U.S Environmental Protection Agency (EPA). To measure natural disaster preparedness in each county, we used Social Vulnerability Index (SVI) datasets from Centers for Disease Control and Prevention (CDC).

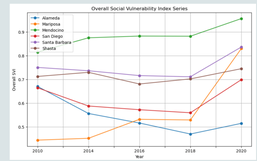
Analysis

The top 6 California counties that have the highest total of PM2.5 level are different with 6 CalFire locations that are used for forecasting and decision making is using. Which shows that PM 2.5 is not the primary significant feature that determines the CalFire locations index



The highest PM 2.5 Concentration in 2023 among 6 CalFire locations based on their average and time series are Mendocino and San Diego. While the most vulnerable county based on its overall social vulnerability index is Mendocino and Santa Barbara. Therefore, mitigation and prevention of fire weather and heat stress has to be focused in Mendocino county.



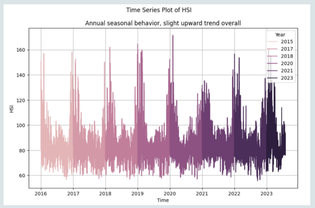


Models

Decision Making Algorithm

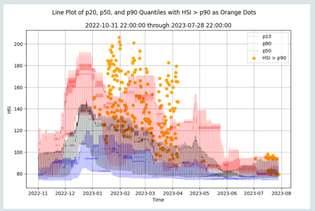
1.) SARIMA Forecast with a Threshold-Based Decision Model

Using a forecasting method that is highly accurate, we can flag points at which the 95% confidence interval crosses over a safe threshold. Our decision-making model would flag these as points of increased AC demand. SARIMA can capture the annual seasonality patterns at fixed intervals, but fitting this model is computationally expensive.



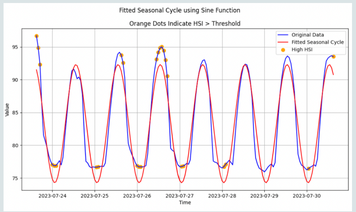
2.) Smoothed Quantile Forecast with an Anomaly Detection-Based Decision Model using NOAA

This forecast method is used by looking back 1 year and taking a 30-day moving average. The smoothing method reduces the effect of outliers and gives an idea of normal behavior. In this case, we can flag points when the NOAA forecast is greater than our normal 95th quantile. The visual shows the smoothed forecast for Mariposa.



3.) Sinusoidal and Climatological Decomposition Forecast with an Anomaly Detection-Based Decision Model using Probability Distributions

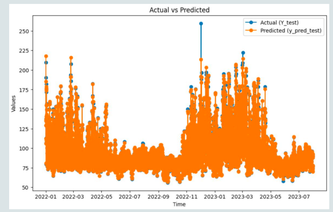
Similar to SARIMA, this model decompose the time series and then look at what's left and use a sine wave to mimic the seasonal cycle and subtract the climatological mean as well. We can detect anomalies that fall outside of a decided threshold and the "normal" sine wave. This approach may work for short-term predictions where the seasonal pattern is relatively simple, but it did not work well for complex pattern.



Forecasting

Heat Stress Forecasting

Trained ML models to produce forecasts for heat stress in 2022-2024. We utilized XGB Regressor as the regression model and MAPE as the primary metric of models accuracy. Mariposa, Redding and Romana received a MAPE of approximately 92% while the other locations received lower. The model of predicted vs actual HSI containing Mariposa, Redding, and Romana



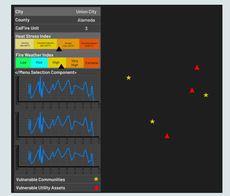
Conclusion

Conclusion

According to the PM 2.5 concentration time series and yearly pattern of Social Vulnerability Index in CalFire counties, Mendocino is the county that we should focus our mitigation measure on. To find the best Decision Making algorithm and Forecasting method, we can use the combinations of methods used. We could build all forecasting methods, compare their flags, and dig into the differences to find the best method.

Final Product

The final product for this project is a dashboard for early alert system which include graphs and multi-layer maps of fire-weather index (FWI), heat stress index (HSI), and other statistics.



Next Step

- Find the most significant features used to determine CalFire locations
- Using hospitalization data due to PM 2.5, predict the number of people that will be hospitalized using the fire-weather forecasting method
- Explore other ways to improve the decision making algorithm, such as : reducing dimensionality, explore other functions to fit model, narrow hyper parameters,etc.
- Further analyze why the heat stress forecasting model for Santa Barbara, Sanel Valley and Union City do not work well. These locations are very humid; thus, the HSI calculation may need to altered to take this into account.