CPE-695 Project Proposal (Team 2)

Rahul Gupta, Andrew Bennett, Chinmay Bhagwat

October 18, 2022

1 Background

Machine Learning algorithms have been becoming popular in many research areas due to the high precision for end-to-end prediction. In the wildfire area, there are many types of research that have applied different methods of Machine Learning in predicting fire activities. One of the most recent research [6] used an ensemble model to analyze satellite data, weather data, and historical fire data and achieved 100% precision on wildfire risk prediction and 93% on wildfire risk detection.

2 Research Problem

Wildfire has become a big issue in the United States, especially in the states like California, causing casualties and substantial economic losses every year. In the 2022 wildfire season, as of 21 September, a total of 6,473 fires have been recorded, totaling approximately 365,140 acres (147,770 hectares) across the state and killing 9 people so far in California. Predicting such an environmental issue becomes a critical concern to mitigate this threat. Several factors cause wildfires such as climate change and human activities. In this project, we are going to predict wildfire using machine learning models based on different types of data.

3 Data

The data for training will be focused on California Wildfires from 1950 to 2022. The main source of the training data will be from [1] which will include the following features: Year of occurrence, Fire Name, Alarm Date, Containment Date, Spatial Shape of the fire, Spatial Shape Length, and the Total Acres burned. In addition, to this data source, the following source [2][3] will supplement the previous data set with GeoSpatial information about the fire location and size of the fire, allowing the Latitude and Longitude to be plotted from this data. The GeoSpatial data will be in GBD format for the GIS data. Climate features will also be added to the training data set from the following [4][5] to add the features Wind Speed, Precipitation, Min Temperature, Max Temperature, and Relative Humidity using the mean monthly values. The climate data will be in two different formats netCDF4 and delimited text files.

The first step of using the training data will be to compile the data into aligned data sets. Secondly, cross reference the data sources and use GIS data to ensure that each fire has location data and spread data. If the Longitude and Latitude cannot be found for a particular fire entry, this would be a candidate for removal from the training data set. Next, obtain the climate and weather data for each of the fires within the main data set using monthly intervals. Daily info for climate and weather info may be used as well when looking at the spread and size of fire from the alarm date to the containment data.

4 Methods

Following paper [6], we would like to reproduce the methods that they implemented to check the precision of 2022 data. Their ensemble method includes SVM, XGBoost, Random Forest, CNN, and LSTM. However, there is no published dataset or code from the research as well as we have limited time and resources for this project. Therefore, we would start by deciding which type of data (features) from the data source will be used in our model and which algorithms need to be implemented. The next step is to create the training and testing dataset. After that, we will screen the data to understand each feature distribution as well as missing or abnormal data. Following the screening step, we must clean data to get a trainable set. In parallel with data creation, we will implement algorithms that we agree on at the first step. After having the model and data, we will train the model and check the performance based on the recall, precision, and F1-score metric on the test set.

5 Expected Results

The result should be compatible with the paper. [6]

6 Time Line

- Features and algorithm decision. (Oct 16, 2022)
- Data cloning. (Oct 20, 2022)
- Data screening and cleaning. (Oct 23, 2022)
- Model implementation. (Oct 30, 2022)
- Testing and evaluating model. (Oct 30, 2022)
- Adjusting methods to achieve better results. (Nov 15, 2022)

References

[1] URL: `https://gis.data.ca.gov/datasets/CALFIRE-Forestry::california-fire-perimeters-all-1/about`. (accessed: 10.16.2022).

[2] URL: `https://frap.fire.ca.gov/mapping/gis-data/`. (accessed: 10.16.2022).

[3] URL: `https://www.kaggle.com/rtatman/188-million-us-wildfires`. (accessed: 10.16.2022).

[4] URL: `https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html`. (accessed: 10.16.2022).

[5] URL: `http://ipm.ucanr.edu/WEATHER/wxactstnames.html`. (accessed: 10.16.2022).

[6] Ashima Malik et al. "Wildfire Risk Prediction and Detection using Machine Learning in San Diego, California". In: Oct. 2021. DOI: `10.1109/SWC50871.2021.00092`.