Forest Fire Analysis and Prediction

By Team
“Free Thinkers”

Anusha Gangasani, Ashika Anand Babu, Ramya Mahesh, Aman Shah

Forest Fire Analysis and Prediction

By Anusha Gangasani, Ramya Mahesh, Aman Shah and Ashika Anand Babu

Acknowledgements

We want to show our gratitude and appreciation to everyone who made it possible for us to finish this project. Our professor, Mr. Vijay Eranti, deserves special recognition for his assistance, stimulating suggestions, and encouragement throughout the fabrication process and the drafting of this report. We also appreciate the time he spent proofreading and correcting our errors. We owe a debt of gratitude to all of our classmates, particularly our friends, for volunteering their time to assist and support us in the development of our project.

Table of Contents

Abstract

The Forest Fire is a bush or vegetation fire which is an uncontrolled burning of plants in forest, grassland or tundra that spreads on environmental conditions which can seriously affect the frequency and intensity of fires. The problem of forest fires is increasing in dry areas which have droughts and high winds. These forest fires can destroy transport, power, gas services, water supplies and many more. They also cause fall in air quality, loss of crops, resources, animals and many people’s lives. In the year 2021 there are 7,024 fires that took place globally and the area burned in these fires is 1,510,450 acres. In the last five years 1.2 million acres of land has burned due to forest fires which destroyed many things. The forest fires not only lead to damage, even after that the smoke from forest fires can cause health risks to adults, children and heart/lung diseased people. Therefore, the analysis and prediction of forest fire is a major issue to reduce its effects. For this issue many models are developed to determine the forest fire ahead of time. Most of the models require various input attributes for the prediction of forest fire. To overcome this issue, we developed a system which has Fire Radiative Power (FRP) that defines the sensed information to quantify the burned biomass used for the forest fire analysis and prediction. The developed system takes only location parameters, time parameters as input attributes and generates the results related to confidence level at which the forest fire may occur in near future and the percentage of forest fire that takes place at that specified location and time.

Introduction

Our project “Forest Fire Analysis and Prediction” determines the confidence level at which the forest fire may occur in the near future at a given location.We all are aware about global warming, some of us have witnessed it and somehow have experienced it facing floods and high temperatures. Wildfires release carbon dioxide and other pollutants into the atmosphere, exacerbating global warming, and in severe cases, irreparably damaging forest ecosystems. The resulting smoke and haze can travel miles, creating public health crises as people breathe in unhealthy levels of pollutants. Uncontrolled wildfires cause billions of dollars in economic damage each year as property and natural tourist attractions are destroyed, water supplies are polluted, and economies are crippled by evacuations.

As of September 14, 2021, the National Interagency Fire Center (NIFC) reported that 44,647 wildfires in the United States had burned 5.6 million acres of land. Similarly, the Canadian Interagency Forest Fire Centre (CIFFC) announced that 6,317 wildfires burned 10.34 million acres.

To control these forest fires it’s better if we predict earlier whether a region can have a forest fire or not based on environmental conditions. For this reason, we developed the Forest Fire Analysis and Prediction system. We have gathered the data from NASA Earth for the entire world and performed various regression techniques for achieving high accuracy. The model with highest accuracy was pickled and fed to the front end. With the front-end we provide an interface for location input and prediction based on the given location.

Related work

The forest fire analysis and prediction is basically related to the paper “ A Data Mining Approach to Predict Forest Fires using Meteorological Data ”. The main idea of this paper is to develop a system which is capable of predicting the burned area of small fires, which are more frequent. Such knowledge is particularly useful for improving firefighting resource management. The main factor of this system is forest Fire Weather Index (FWI) and the model used to predict forest fires is Support Vector Machines (SVM).

There is a major difference between the system “ forest fire analysis and prediction ” and “ Predict Forest Fires using Meteorological Data ” are:

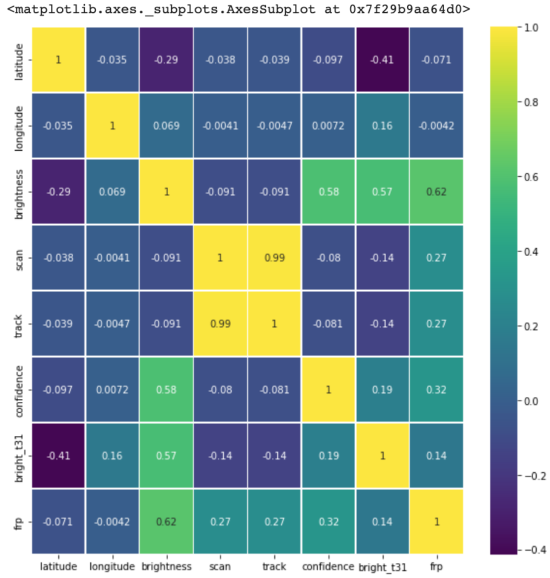
- ➔ Developed a system which works for all the global data
- ➔ Detects and predicts forest fires regardless of how much area burnt
- ➔ Forest Radioactive Power (FRP) is key for the system
- ➔ Used Random Forest Regressor to predict the forest fires
- ➔ Used AWS EC2 instance for the deployment
- ➔ Developed a UI which takes location, time parameters and generates the results

Data

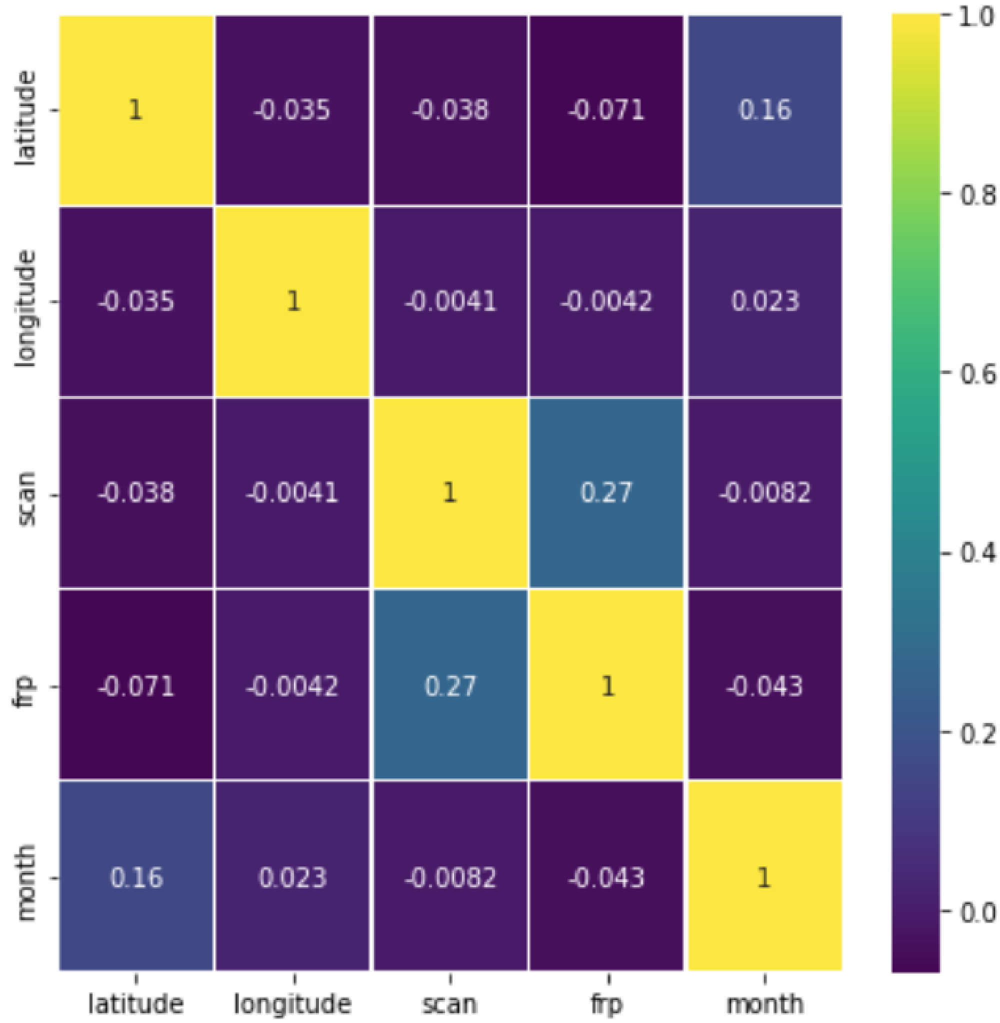
The data used for the forest fire analysis and prediction is gathered from NASA earth(https://nrt3.modaps.eosdis.nasa.gov/archive/FIRMS/modis-c6.1/Global). This contains 61 files of recent data of forest fires that occurred globally. All the 61 files are merged into a single file for forest fire analysis and prediction. The size of the data after combining all the files is around 46MB which has 547910 entries. Some of the main attributes of this data are latitude, longitude, frp, date, time and satellite information.

In the data cleaning and preprocessing section, data is already clean but had many fields that were irrelevant to our project which we removed. The data is feature engineered by adding 'daynight_map', 'satellite_map' and 'month' from the existing features.

<matplotlib.axes._subplots.AxesSubplot at 0x7f29b9aa64d0>



Heatmap of data before cleaning and preprocessing



Heatmap of data after cleaning and preprocessing

Technical identifier	Variable	Description
Latitude	Latitude	Center of 1km fire pixel but not necessarily the actual location of the fire as one or more fires can be detected within the 1km pixel.
Longitude	Longitude	Center of 1km fire pixel but not necessarily the actual location of the fire as one or more fires can be detected within the 1km pixel.
Brightness	Brightness temperature 21 (Kelvin)	Channel 21/22 brightness temperature of the fire pixel measured in Kelvin.
Scan	Along Scan pixel size	The algorithm produces 1km fire pixels but MODIS pixels get bigger toward the edge of scan. Scan and track reflect actual pixel size.
Track	Along Track pixel size	The algorithm produces 1km fire pixels but MODIS pixels get bigger toward the edge of scan. Scan and track reflect actual pixel size.
Acq_Date	Acquisition Date	Date of MODIS acquisition.
Acq_Time	Acquisition Time	Time of acquisition/overpass of the satellite (in UTC).
Satellite	Satellite	A = Aqua and T = Terra.
Confidence	Confidence (0-100%)	This value is based on a collection of intermediate algorithm quantities used in the detection process. It is intended to help users gauge the quality of individual hotspot/fire pixels. Confidence estimates range between 0 and 100% and are assigned one of the three fire classes (low-confidence fire, nominal-confidence fire, or high-confidence fire).
Version	Version (Collection and source)	Version identifies the collection (e.g. MODIS Collection 6) and source of data processing: Near Real-Time (NRT suffix added to collection) or Standard Processing (collection only). "6.0NRT" - Collection 6 NRT processing. "6.0" - Collection 6 Standard processing. Find out more on collections and on the differences between FIRMS data sourced from LANCE FIRMS and University of Maryland.
Bright_T31	Brightness temperature 31 (Kelvin)	Channel 31 brightness temperature of the fire pixel measured in Kelvin.
FRP	Fire Radiative Power	Depicts the pixel-integrated fire radiative power in MW (megawatts).
DayNight	Day / Night	D = Daytime, N = Nighttime

Attributes that we used in the Forest Fire Analysis and Prediction

Methods

With the obtained cleaned data after preprocessing built three different models for predicting the accuracy of forest fires occurrence. The data is split into training and test data sets for building the model. The train data set is used to train the model. The test data set(which the model has never encountered before) is used to test the model. The three models implemented in this project are Random forest regressor, logistic regression and decision tree regressor.

Random Forest Regression is a supervised learning algorithm that uses ensemble learning methods for regression. It provides higher accuracy through cross validation. For building this random forest model the trained and test data sets are passed through the models with 300 estimators and predicted the accuracy of the model.

```
random_model = RandomForestRegressor(n_estimators=300, random_state = 42, n_jobs = -1)
```

```
#Fit
random_model.fit(Xtrain, ytrain)
```

```
RandomForestRegressor(n_estimators=300, n_jobs=-1, random_state=42)
```

ict the model

```
y_result = random_model.predict(Xtest)
print(y_result)
```

```
[65.62       62.75333333 99.98       ... 64.01       76.1
 91.42666667]
```

Logistic regression is used to understand the relationship between the dependent and independent variables by estimating the probabilities. Decision tree regressor breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed.

Among these three models the Random Forest Regressor stands out with an accuracy of 94.35% which is used for further process of predicting and analysing the forest fires.

Check the accuracy

```
[86] #Checking the accuracy
     random_model_accuracy = round(random_model.score(Xtrain, ytrain)*100,2)
     print(round(random_model_accuracy, 2), '%')

     94.35 %
```

Experiments and Results

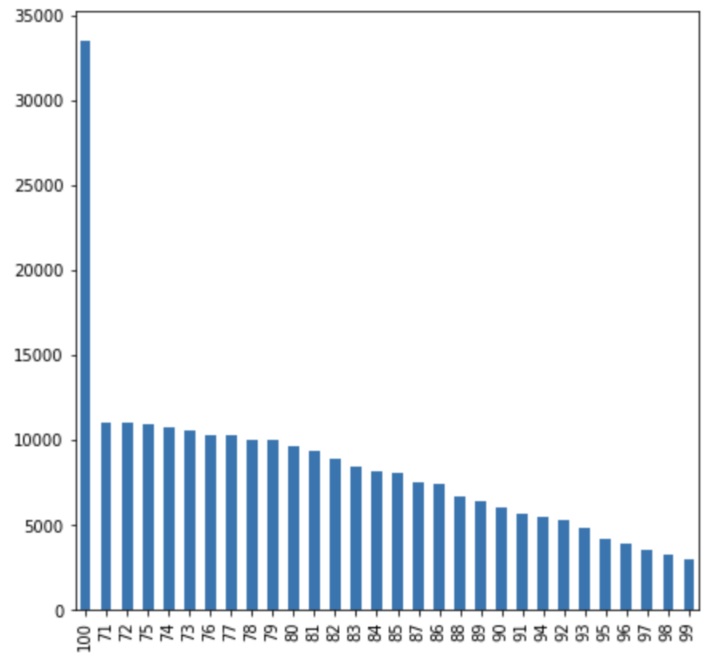
The data is obtained from NASA earth and merged into a single file which doesn't have any noisy data. After obtaining the data the feature extraction is performed to derive some of the attributes to build a model for predicting and analysing the forest fires. The new data set is split into training and test data sets and passed through three different models. The trained data set is passed with 300 estimators to fit in the model and the test data is used for predicting the accuracy of the model.

The model is then dumped into a pickle file to load it into the AWS EC2 instance. The data obtained from the AWS EC2 instance is then passed to the UI to generate the results for forest fire prediction and analysis. Therefore, the forest fire analysis and prediction takes the location and time attributes and generates the confidence level at which the forest fire may occur in near future and the percentage of forest fire that takes place at that specified location and time.

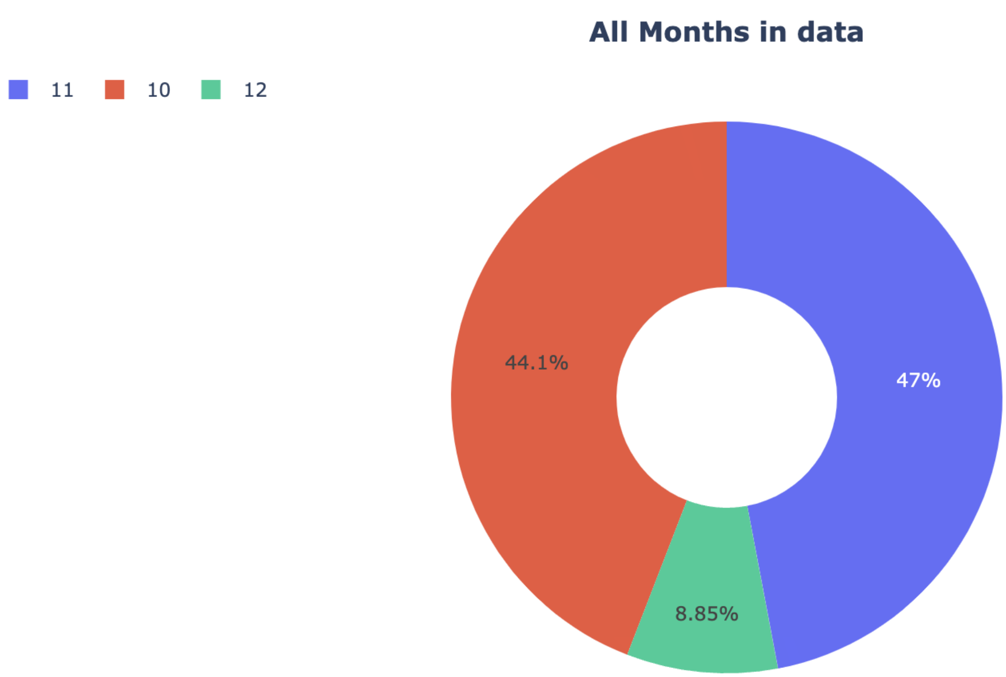
We have used five visualization techniques to represent the data:

- Bar plot
- Pie-chart
- Scatter-mapbox
- Histogram
- Graph plot

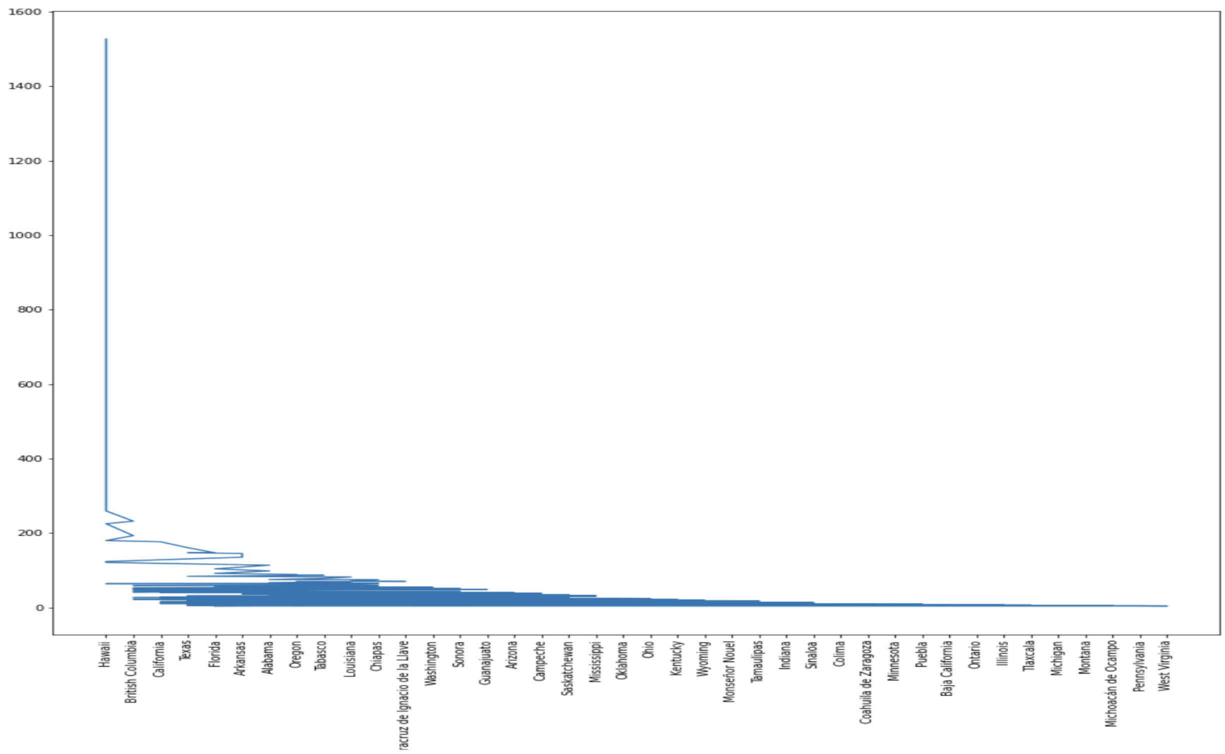
- The below bar graph shows the data records which have a confidence level greater than a threshold value (80).



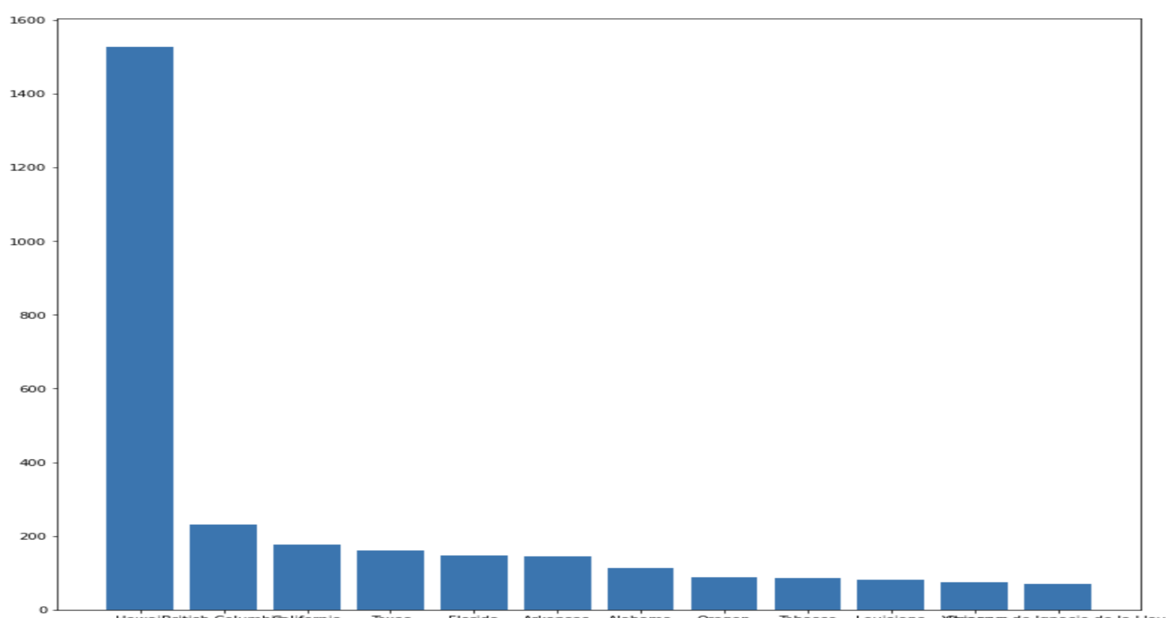
- The below pie chart shows the data regarding the forest fires that took place in the past three months.



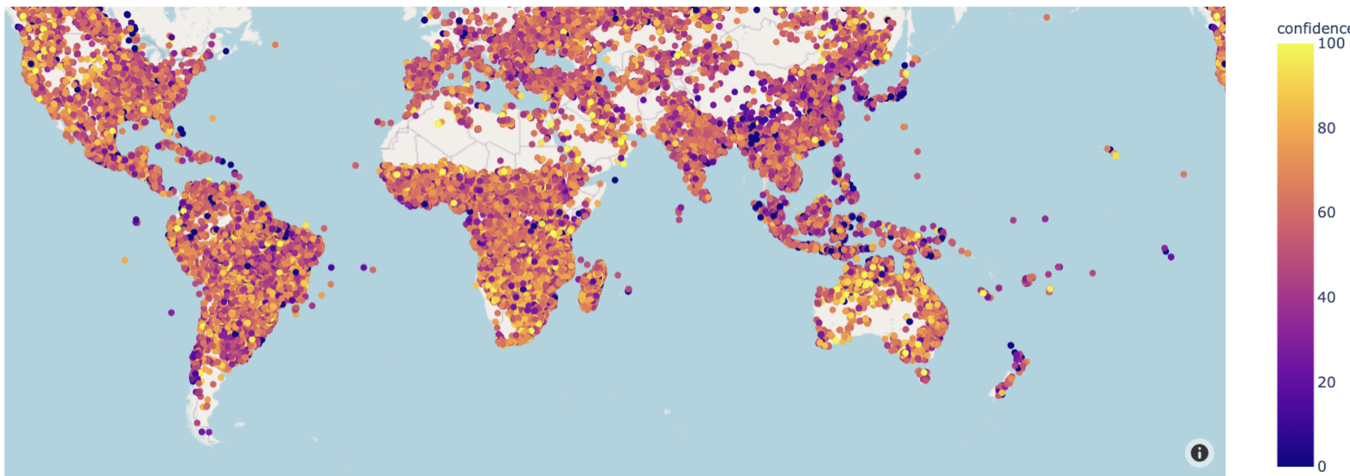
- The below graph plot shows the places which has maximum forest fires and Hawaii struggles the most in the recent past.



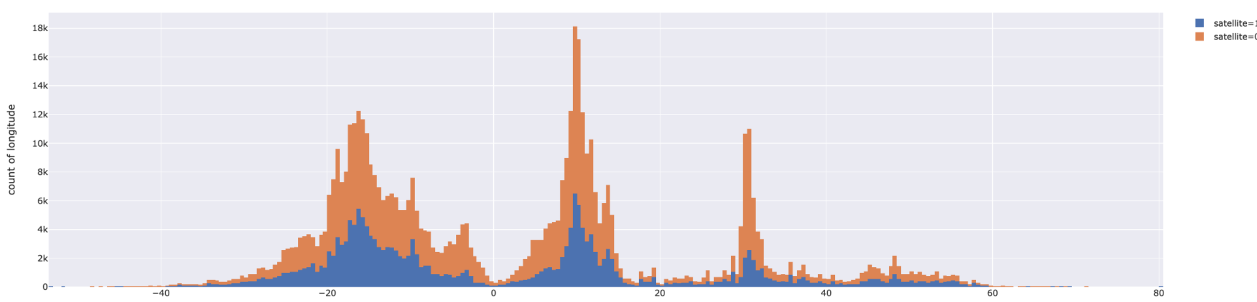
- The below bar graph shows the top 40 states that were affected by the forest fires



- The below mapbox show the confidence and FRP levels of forest fires that took place globally



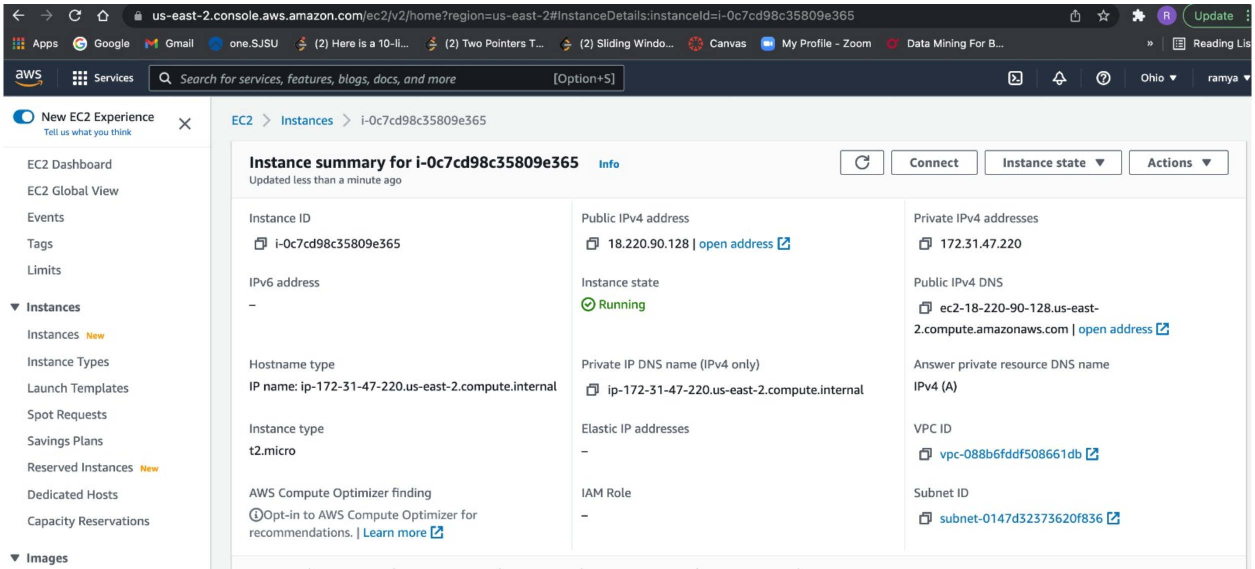
- The below histogram represents the latitude and longitude of all the places captured by two different satellites.



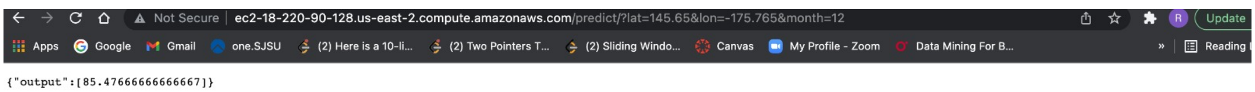
Deployment

The model is dumped into the pickle which is sent to AWS EC2 instance for the deployment phase.

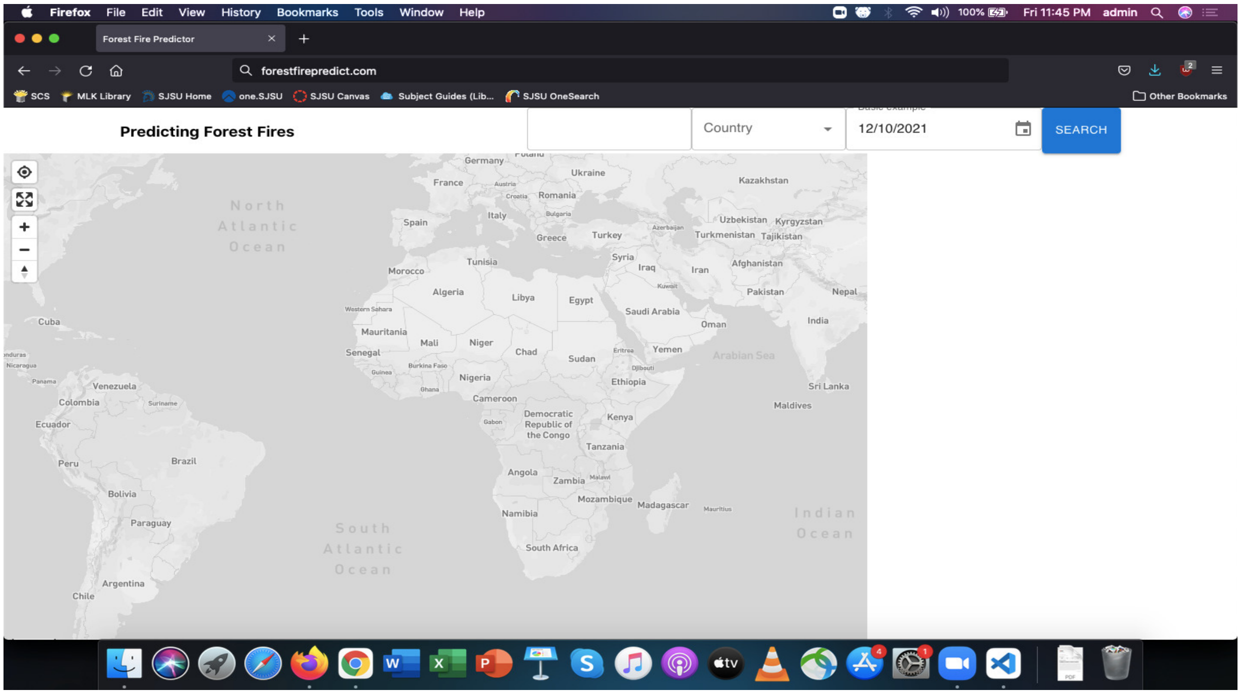
- Below image shows the creation of AWS EC2 instance



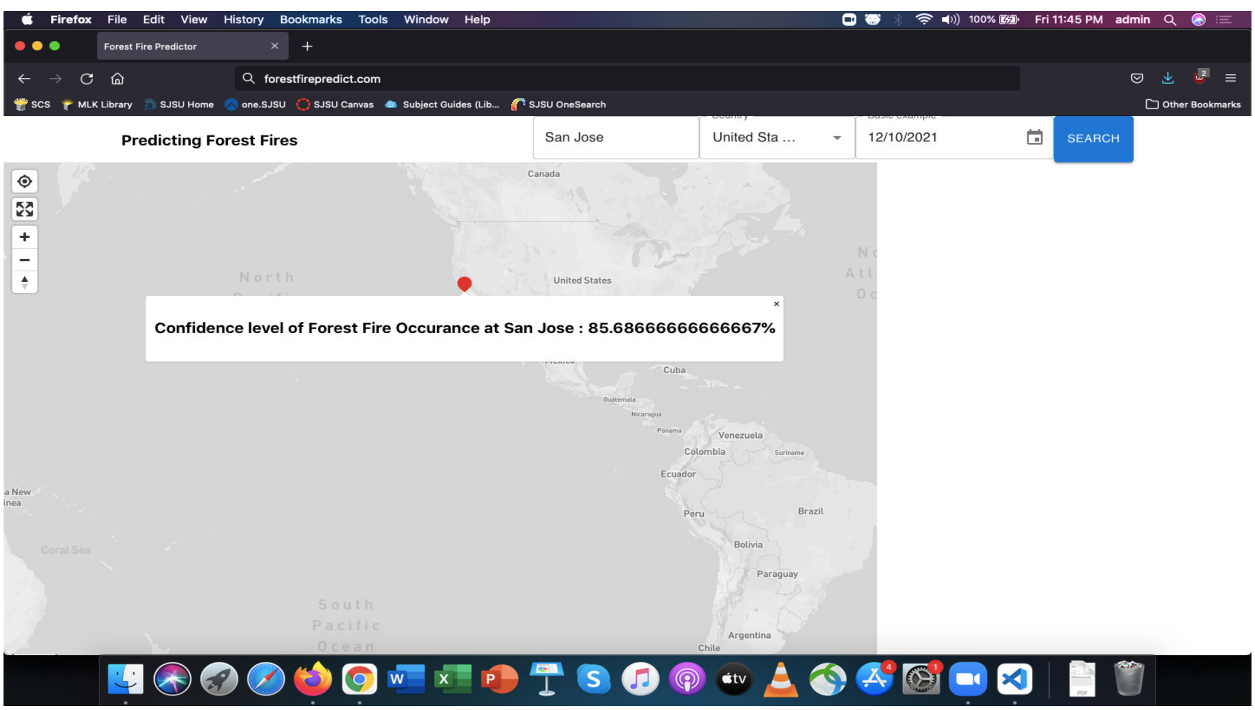
- The result after deployment is shown below



- The data from the deployment is fetched and passed through the UI which takes location and time parameters as input.



- With the input attributes it generates the confidence level of that particular location



Conclusion

The climate and vegetation of the greater Athens area make forest fires a real threat to the environment during the summer. The forest fires will continue to affect source water quality resulting in increasing treatment, maintenance, operating costs and destroying everything.

This system is designed in such a way that it can predict the forest fires at any location by taking the input as latitude and longitude, date and time and generates the results as confidence level at which the forest fire may occur in near future and the percentage of forest fire that takes place at that specified location and time. The Random Forest Regressor model does the internal job of predicting the forest fires which takes trained and test datasets.

Predictive Services provides decision-support information to the U.S. Forest Service and other federal, tribal, state, and local wildland fire management agencies for operational management of and strategic planning.

This system may not help in recovering loss, instead it helps to predict the forest fires in a particular location and time. We say “ Prevention is better than Cure ” which implies the system we developed. The analysis and prediction of forest fires acknowledges everyone when to take preventive measures without having loss of life or resources. By predicting the future, this system will definitely benefit to stop at least a few destructions caused by forest fires.

Google Colab:
https://colab.research.google.com/drive/10g2fgfyI6s3XINKIiMLomKjkuZ4ra0d_?authuser=2#scrollTo=hd2e4BqaNyZ8

Presentation:
https://docs.google.com/presentation/d/1gCoEG9m-zyypJohBpYVs0dRfvOv1Ffns7T11o-_zCbE/edit#slide=id.p

Demo URL:

Github Repository:
https://github.com/AshikaAnand12/ForestFirePrediction

Project Report:
https://docs.google.com/document/d/1AyjehPCYIr95UH9lTKW9UaMHmGEnluV0kL_n57cIPr4/edit#

References:

- https://www.researchgate.net/publication/238767143_A_Data_Mining_Approach_to_Predict_Forest_Fires_using_Meteorological_Data
- https://www.predictiveservices.nifc.gov/outlooks/monthly_seasonal_outlook.pdf
- https://www.sciencedirect.com/science/article/pii/S1877050916307311
- https://fireecology.springeropen.com/articles/10.4996/fireecology.1101106

Anusha Gangasani
Ashika Anand Babu
Ramya Mahesh
Aman Shah