Topic: Environment

Title

**Wildfires in Southern Europe**

**INTRODUCTION**

**Aims and objectives of the project**

Our project focuses on wildfires, an important environmental issue, strongly connected to climate change. Wildfires cause great damage to the environment and the economy. In this project, we analyze how such fires depend on weather conditions and try to predict the development of the situation in the future using machine learning. Particular emphasis is put on the regions of Portugal and Spain. Time coverage of the datasets is equal to 20 years (2001 - 2021). This project aims to provide useful information based on the analysis of historical data and the use of appropriate machine learning methods to solve the above problems.

**Roadmap of the report**

1. Preparing data:
   1. Downloading data (Wildfires\_get\_data.ipynb)
      1. Downloading MODIS fires locations data for Portugal and Spain, combining into one Iberia\_all.csv file
      2. Downloading burned area data, obtained from MODIS satellite data, monthly values, Global\_Burned\_Area\_2002\_2019.csv
      3. Downloading meteorological data by API Meteo\_2001\_2021.nc, Meteo\_2001\_2021\_monthly.nc (note: data is free, however, login is required. Also data file is quite large)
   2. Merging datasets (merging\_meteo\_and\_fire\_data.ipynb)
2. Exploratory analysis and cleaning data (EDA\_and\_cleaning.ipynb)
3. Applying regression machine learning models, using Fires\_final\_data.csv data file (ML\_models.ipynb)

**BACKGROUND**

A wildfire is an unplanned fire that burns in a natural area such as a forest, grassland, or prairie. Wildfires are often caused by human activity or a natural phenomenon such as lightning, and they can happen at any time or anywhere. The risk of wildfires increases in extremely dry conditions, such as drought, and during high winds. Wildfires can disrupt transportation, communications, power and gas services, and water supply. They also lead to a deterioration of the air quality, and loss of property, crops, resources, animals and people. [[WHO](https://www.who.int/health-topics/wildfires#tab=tab_1)] Historically, fire was an important element for the ecosystem dynamics in the Mediterranean landscapes and represented a major element in many forested ecosystems, particularly in the Southern Europe, where rural communities learned how to use it as a land management tool. However, on the second half of the twentieth century, with the depopulation of rural areas, wildfires increased in area in the Iberian Peninsula and forest fires became one of the main environmental problems and one of the most important natural hazards. [[Nunes2019](https://annforsci.biomedcentral.com/articles/10.1007/s13595-019-0811-5)].

The questions that we tried to asnwer:

1. Has the number of forest fires been growing over the past 20 years?
2. Are wildfires more severe (based on fire radiative power)?
3. What are the annual and seasonal changes in wildfires occurrences?
4. Are changes in wildfires correlated with changes in meteorological conditions?

**STEPS SPECIFICATIONS**

Describe how your team approached each of the key steps of the data analysis: framing questions, data gathering, preprocessing, in-depth analysis etc.

Our work was divided into 4 main parts, that are included in the separate notebooks:

* Downloading data (Wildfires\_get\_data.ipynb)
* Merging datasets (merging\_meteo\_and\_fire\_data.ipynb)
* Exploratory analysis and cleaning data (EDA\_and\_cleaning.ipynb)
* Applying regression machine learning models (ML\_models.ipynb)

Describe your data sources and how you found them.

All data is available for everyone, free and has scientific quality:

* fires locations data derived by MODIS instrument onboard Terra and Aqua NASA satellites ([*Fires*](https://firms.modaps.eosdis.nasa.gov/country/)) [downloaded with urllib.request; info: date, geolocation, brightness temperature, fire radiative power, confidence (and other variables)]
* meteorological data from ERA5 reanalysis provided by Copernicus system (downloaded with API; [*Meteo*](https://cds.climate.copernicus.eu/#!/home), free account needed) [info: temperature, dew point temperature, rain)
* Global dataset burned area data, derived from the MODIS MCD64A1 burned area product. Global Wildfire Information System ([GWIS](https://gwis.jrc.ec.europa.eu/apps/country.profile/downloads)) [downloaded with urllib.request; info: burned area]

**IMPLEMENTATION AND EXECUTION**

Development approach:

● collecting data:

○ fires location (MODIS),

○ meteorology (API)

○ burned area (monthly data)

● merging datasets - to have one, nice .csv :)

● EDA

○ analysis of fires occurrence (annual, seasonal changes,etc).

○ analysis of the meteorological parameters

○ preparing correlation matrix

○ choosing variables for the ML model

● cleaning data (handling nulls, outliers, duplicates, skewness)

● creating new variables from existing ones

● preparing Machine Learning regression models

○ linear regression

○ decision tree regression

○ random forest regression

○ extra trees regression

● Prepare presentation

● Prepare report

● Upload project to Github

**Team member roles:**

Dora: merging data, analysis of the meteorological parameters, cleaning data, prepare presentation

Iuliia: analysis of fires occurrence, creating new variables, prepare report

Olga: collecting data, EDA&cleaning, choosing variables for the ML model, preparing Machine Learning models, , prepare report

**Tools and libraries:**

Main tools and libraries:

pandas, numpy, seaborn, matplotlib, scipy, datetime, xarray, sklearn (metrics, DecisionTreeRegressor, RandomForestRegressor, ExtraTreesRegressor), urllib.request, os, zipfile, cdsapi, ImageMagic.

**Implementation process** (achievements, challenges, decision to change something)

Main challenge was to combine the data with different time and spatial resolution into one table. The geographic latitude and longitude did not match. We decided to take approximate values.

We did not manage to use data on burned area (monthly values), even though we downloaded it. Although very interesting it would be time consuming, and we did not have much time left for project preparation.

**RESULT REPORTING**

What are the key findings of your project?

* Highest mean temperature was reported for July (Fig. 1)
* Most fires were detected in August which is consequence of high temperatures during whole summer. (Fig. 2)
* During last 20 years number of fires varies, some cycles can be seen (Fig. 3). However, generaly number of fires was highest in years 2003, 2005.
* When fires sevirity is taken into account (fire radiative power) maximum values were noted in year 2021 (Fig. 4).
* High values of fire radiative power are connected to the high temperatures (Fig. 5)

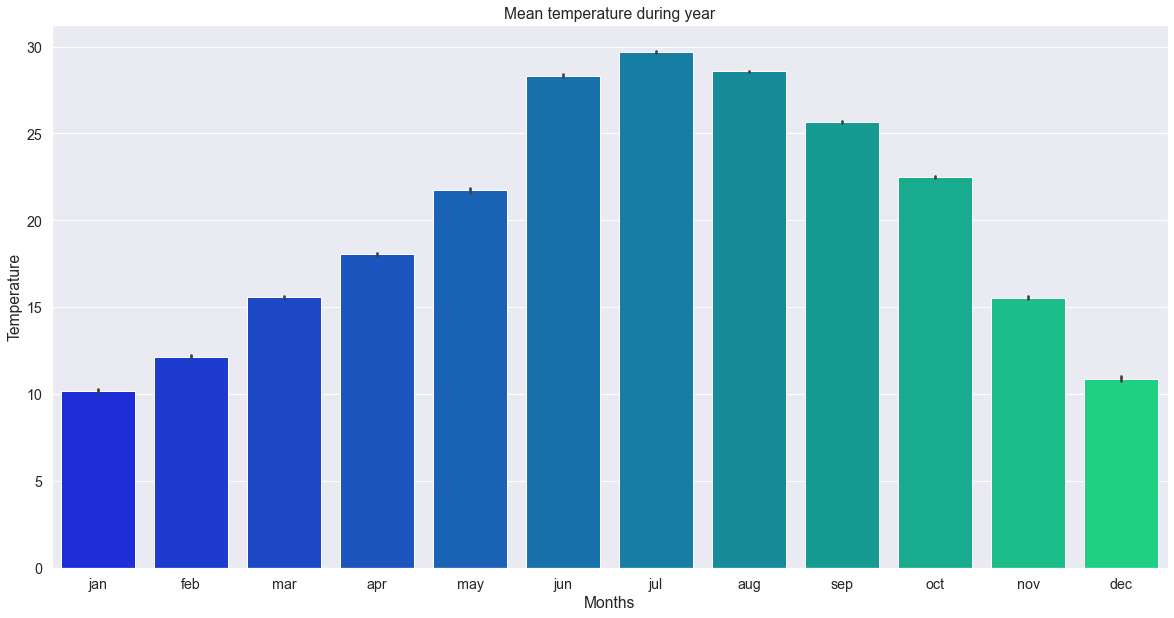


Fig. 1. Mean temperature during year.

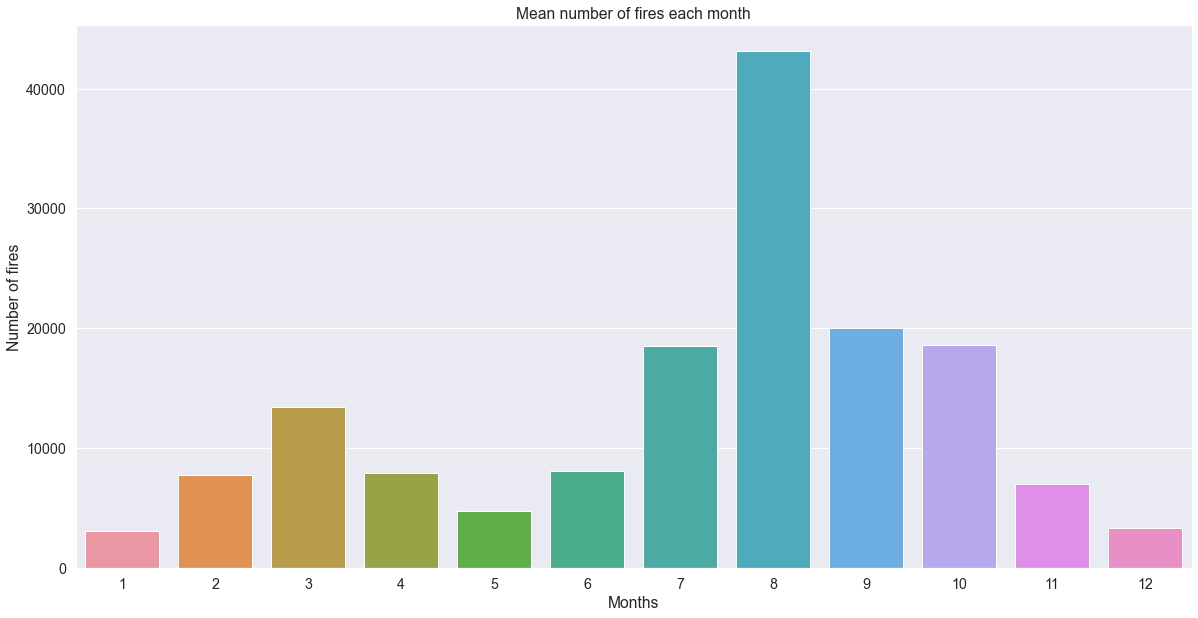


Fig. 2. Mean number of fires each month.

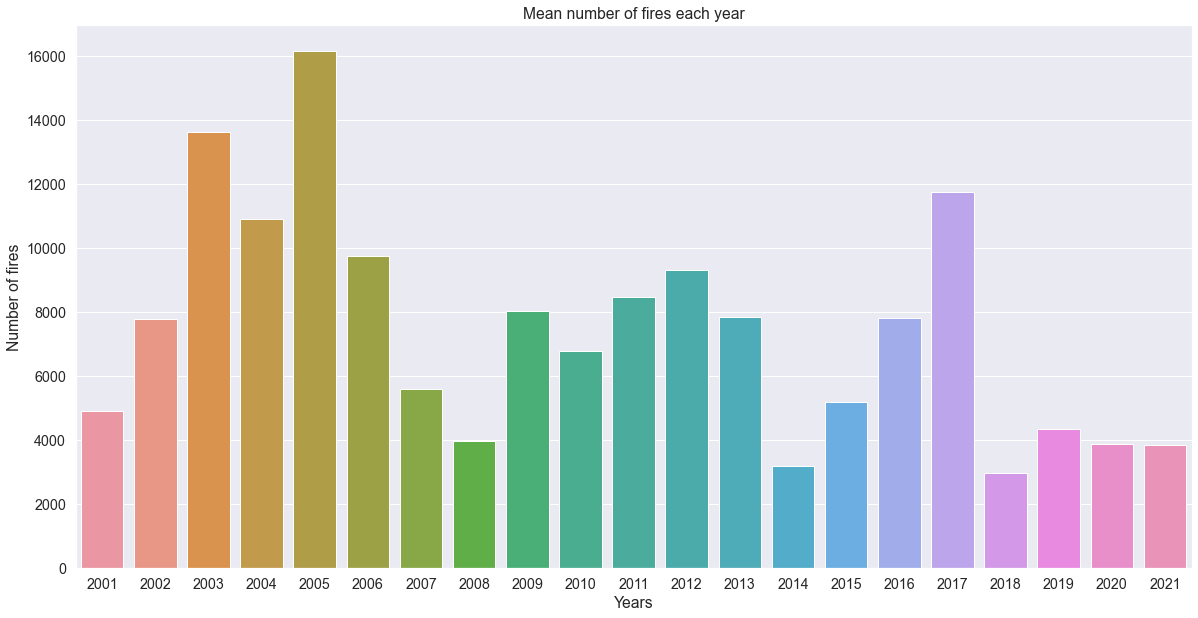


Fig. 3. Mean nuber of fires for years 2001-2021.

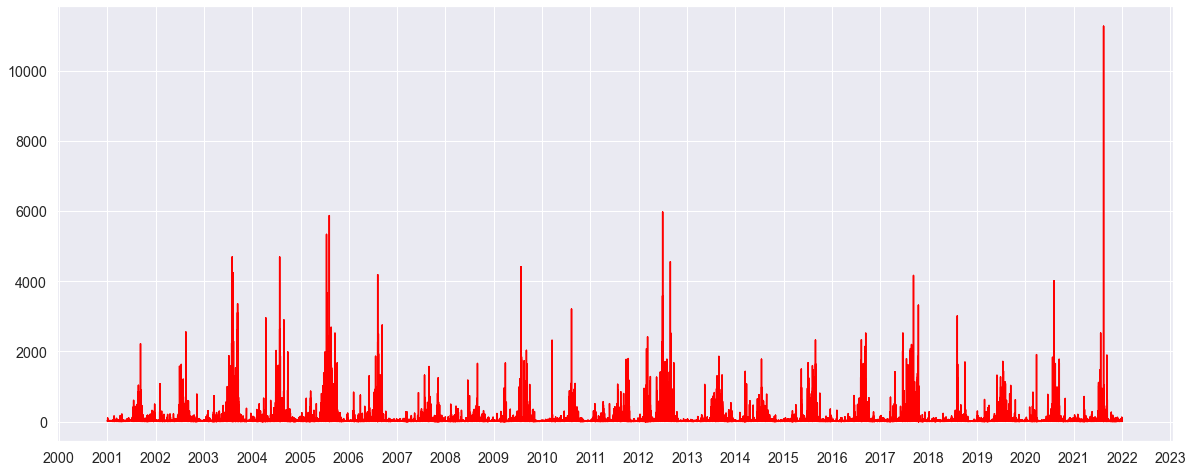


Fig. 4. Fire Radiative Power (MW) for years 2001-2021.

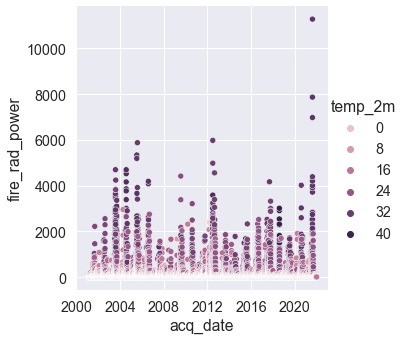


Fig. 5. Fire Radiative Power (MW) vs. temperature for years 2001-2021.

We decided that our problem to be solved will be to **predict intensity of the fire (fire radiative power) based on set of variables.** After trying different approaches (3 versions of ML\_models) we decide to use following setup. Based on correlation matrix we decided to use following fetures: ['brightness','confidence', 'scan\_binned', 'temp\_2m', 'month', 'type\_0\_vegetation', 'daynight', 'year', 'dew\_point\_2m']. We used 4 different regression models: Linear Regression, Decision Trees Regression, Random Forest Regression, Extra Trees Regression. We train the model on 80% of the data and test it on the remaining 20%.

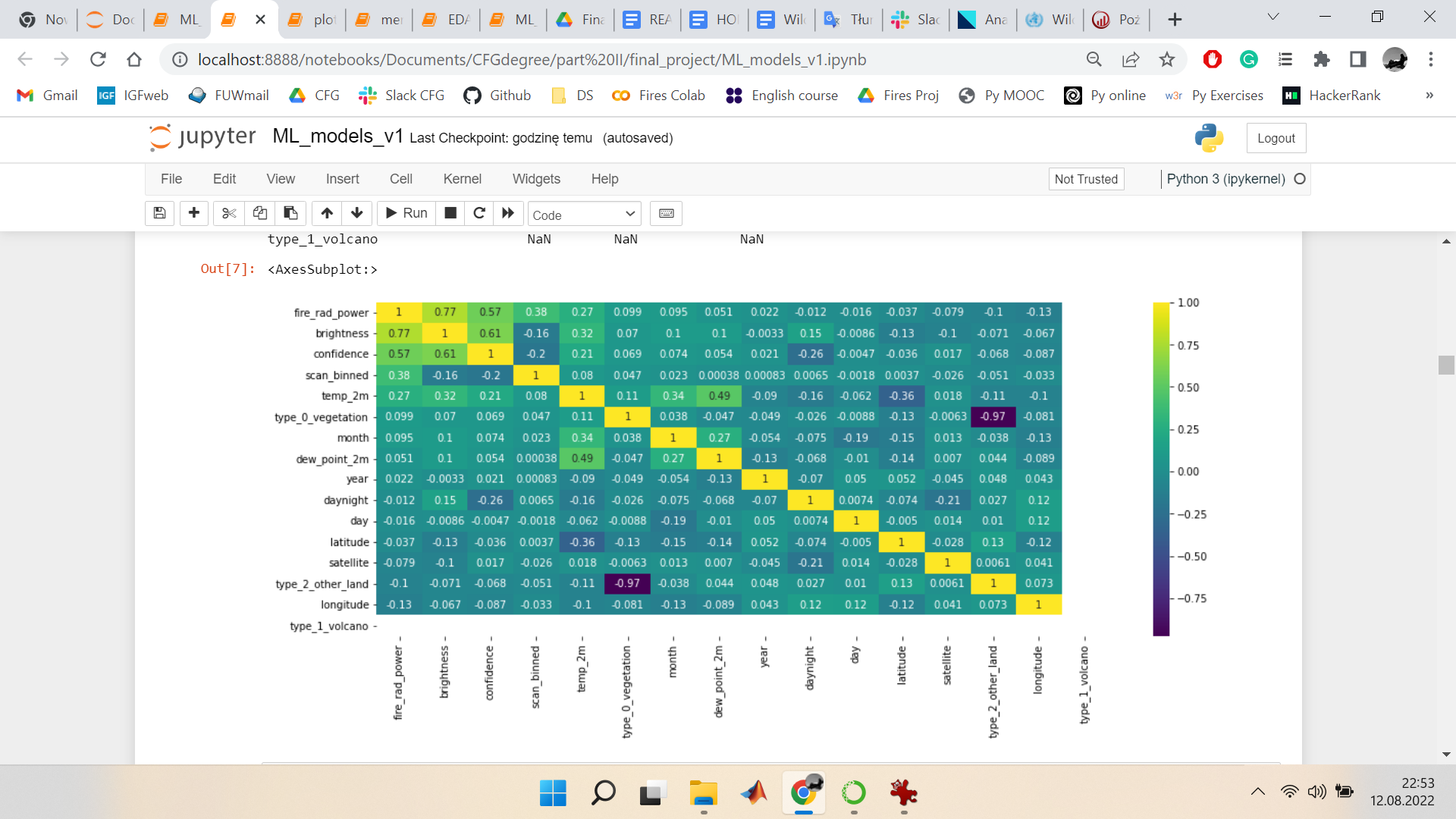
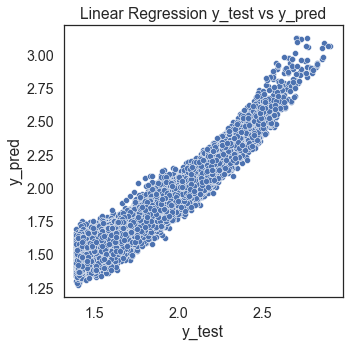


Fig. 6. Correlation matrix.

Below results for different models are presented

**Linear Regression**



The Linear regression model performance for testing set

--------------------------------------

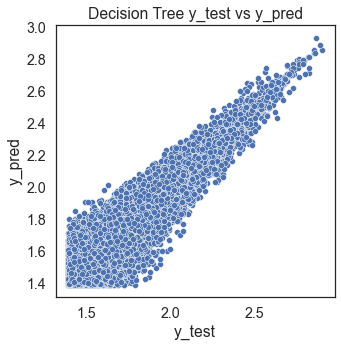
Mean Absolute Error: 0.06909993851811318

Mean Squared Error: 0.007319049751394302

Root Mean Squared Error: 0.08555144505731216

R2 score = 0.9

**Decision Trees Regression**

****

The Decision Trees Regression model performance for testing set

--------------------------------------

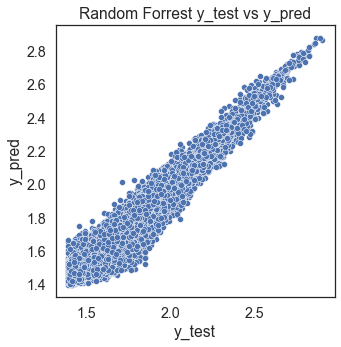
Mean Absolute Error: 0.06679232015315233

Mean Squared Error: 0.008635438729044228

Root Mean Squared Error: 0.09292706133868771

R2 score = 0.88

**Random Forest Regression**

****

The Random Forest Regression model performance for testing set

--------------------------------------

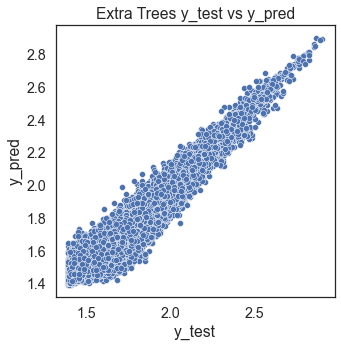
Mean Absolute Error: 0.05022593376801821

Mean Squared Error: 0.0044099999724207155

Root Mean Squared Error: 0.06640783065588513

R2 score = 0.94

**Extra Trees Regression**

****

The Extra Trees Regression model performance for testing set

--------------------------------------

Mean Absolute Error: 0.0508700400006792

Mean Squared Error: 0.004584318702571568

Root Mean Squared Error: 0.06770759708165376

R2 score = 0.94

CONCLUSION

We analyzed fires data from last 20 years. We found out that in August there is highest number of fires. During last 20 years number of fires varies, some cycles can be seen. Number of fires was highest in years 2003, 2005, but maximum values of fire radiative power were noted in year 2021.

Machine learning models trained by use performed quite good taking into account the complexity of the phenomenon. Best results were obtained for Random Forest Regression and Extra Trees Regression. However we belive that this part of project should be further developed and explored.

Also, it would be an advantage to include burned area data into analysis.

During the last 20 years there have been 160 k of fires reported in the countries of Spain and Portugal. Approximately 60% of all wildfires are recorded during 4 months of the year starting from July and ending September while almost 25% of them are recorded in the Month of August. The overall yearly trend tends to decrease on yearly basis with an average of 7.5 k wildfires in a preceding decade into 5.4 k in a current one. 70% of all wildfires were recorded during the day period while the remaining 30% were active during the night. Additionally 57% of the wildfires were active during the week day while 43% of them were recorded during the weekend including Fridays. However, during the 4 months with the most % share of wildfires from July till September the structure is changing slightly, increasing the total % of wildfires recorded during the night up 10 percentage points to 40% of total. We believe that the cause and effect may be tied directly to people’s behavior during the holiday season and their increased tendency to spend more time outdoors enjoying their lives to the fullest while not necessarily focusing on safety itself. This of course is magnified greatly by the hot and dry weather itself during this period and we believe this could be a next step of more in-depth analysis.

Additionally we noticed a great increase of the FRP in 2021 vs the average number for the whole 20 years ( 186 vs 50), however we were not able to find the cause of this occurrence.

When it comes to the correlation matrix in general between the metrics no surprises were discovered between all the relevant metrics.

In general, we do not believe that global temperature warming is a sole reason for the number of total wildfires being increasing or decreasing as based on our data for both Portugal and Spain the temperature looks still while the total number of wildfires is decreasing during the period. We are aware that there is a huge data lack at our end in order to fully understand the wildfires situation. We understand that countries could and probably are deploying a lot of initiatives to stop the situation by investing in technology and social education which helps prevent wildfires. On a side note we have discovered that the total decrease of wildfires during two decades correlates with the decrease of active smokers in both countries which fell since the millennium by about 30% in Spain and Portugal during the period. The bottom line is somewhat tragic, as we, the people, are directly and indirectly causing the most of the wildfires on the globe and there is no significant sign of it being changed anytime soon.

**References**

| **Fires** | <https://firms.modaps.eosdis.nasa.gov/country/> |
| --- | --- |
| **FWIS** | <https://www.nwcg.gov/publications/pms437/cffdrs/fire-weather-index-system> |
| **GWIS** | <https://gwis.jrc.ec.europa.eu/apps/country.profile/downloads> |
| **Meteo** | <https://cds.climate.copernicus.eu/#!/home> |
| **MODIS** | <https://modis-fire.umd.edu/files/MODIS_C6_Fire_User_Guide_B.pdf> |
| **Nunes2019** | <https://annforsci.biomedcentral.com/articles/10.1007/s13595-019-0811-5> |
| **WHO** | <https://www.who.int/health-topics/wildfires#tab=tab_1> |
|  |  |
|  |  |