

RF Earth, Wind and Fire

**Modelo para inferir
el comportamiento
de un incendio**



AI Saturdays

Contenidos

Descripción del proyecto

Framework

Visualización de Datos

Valor Social

Orígenes de Datos

Diseño del modelo

Histórico en Península Ibérica

Principales Atributos

Estructura del Notebook

Problemas que Resolver

Índices y Relaciones

Interpretación de Resultados



Descripción del Proyecto

Proyecto de integración y **ciencia de datos**, que trata de inferir, la **relación entre variables** climatológicas, geo-biológicas y de intensidad de los **potenciales incendios** que se pudieran desarrollar en las áreas de estudio. Con ello, poder **estimar la afectación** de la zona, ante un posible incendio



Valor Social del Proyecto

El proyecto pretende estudiar el comportamiento de los incendios, relacionando la frecuencia de acontecimientos, su intensidad, relaciones con la climatología o aspectos zonales, tratando de inferir su comportamiento.

Nuestro país se enfrenta cada año a la misma emergencia relacionada con incendios forestales. Los montes arden con pequeños incendios, en general decrecientes en número y superficie quemada, pero con un riesgo permanente de sufrir grandes incendios forestales (GIF). Se dan las condiciones para que se produzcan crisis incendiarias como la acontecida en Portugal en junio de 2017, un incendio de sexta generación, vinculado al cambio climático, con más de 156 fuegos simultáneos, que arrasaron unas 50.000 hectáreas, alguno de ellos causaron más de 60 fallecidos y un centenar de heridos.



Península Ibérica On Fired

Portugal

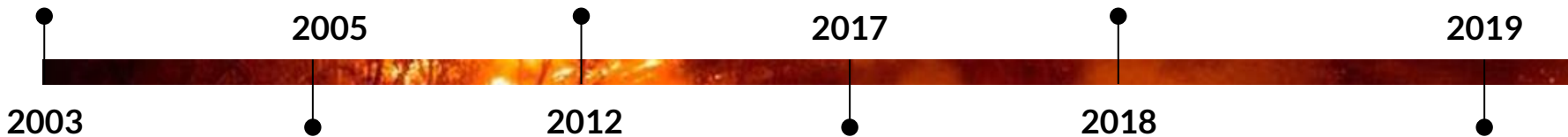
Portugal se vio afectada por una de las olas de calor más largas de su historia: 18 personas perdieron la vida y ardieron un total de 425.839 hectáreas. A comienzos de agosto el gobierno declaró estado de calamidad pública.

España

En España se quemaron cerca de 220.000 hectáreas, el doble de lo habitual

Incendio en Tarifa

Por primera vez en la historia reciente, se registraron intensas y preocupantes oleadas de incendios forestales en Europa Central y del Norte



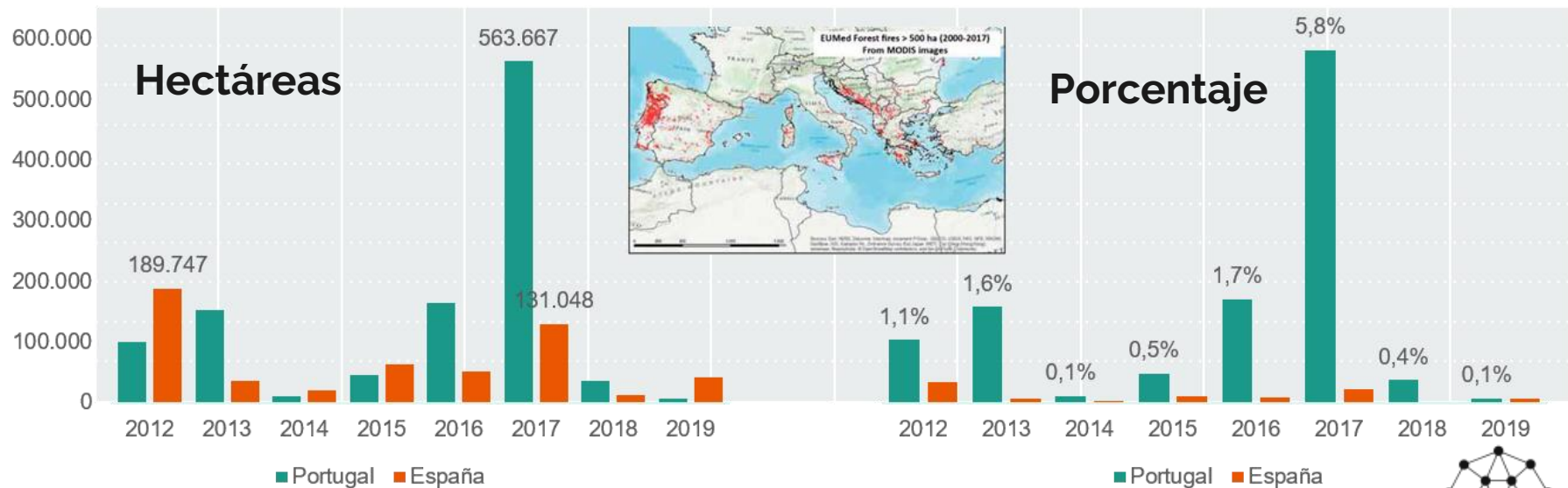
Problema Medioambiental Grave

En España ardieron 189.000 hectáreas, un 80% más de lo que se quema de media al año. En Portugal se quemaron 340.000 hectáreas, el doble de que lo se quema un año normal.

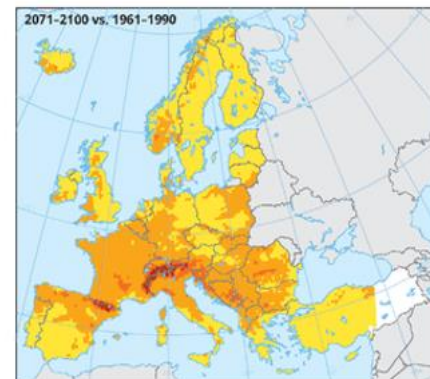
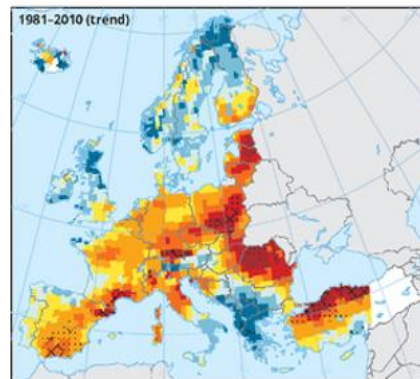
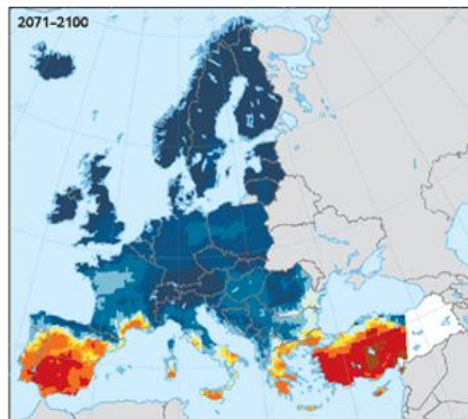
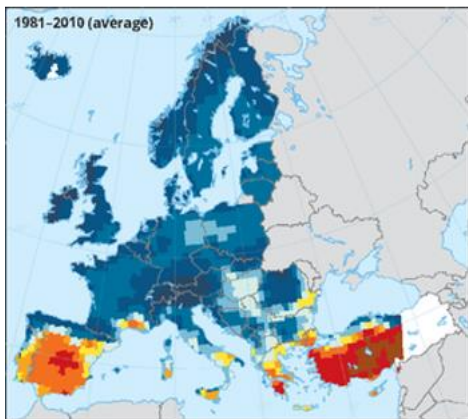
Catástrofe humanitaria y medioambiental

En Portugal 540.000 hectáreas fueron pasto de las llamas, un 250% más de lo que se quema de media al año. En España ardieron unas 180.000 hectáreas, un 70% más de lo que arde un año normal. Además, en Portugal perdieron la vida 119 personas y 4 en España.

Áreas Quemadas (2012-2019)

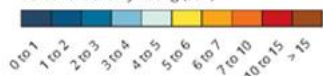


Escalada del Índice de Severidad (SSR)

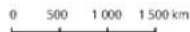


Average forest fire danger (1981-2010)

Seasonal Severity Rating (SSR)



Projected forest fire danger (2071-2100)

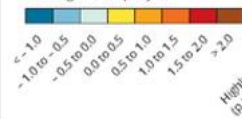


© European Union 2016,
source: Joint Research Centre

No data
Outside coverage

Observed trend in forest fire danger (1981-2010)

% change in SSR per year

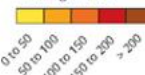


Statistical significance
of fire danger trend



Projected change in forest fire danger (2071-2100 vs. 1961-1990)

% change in SSR



© European Union 2016,
source: Joint Research Centre

Problemas que resolver

1

Integración de múltiples fuentes de datos de distintos proveedores

3

Diseño del modelo predictivo basado en algoritmos de machine learning

2

Limpieza y normalización de los datos

4

Interpretación y extrapolación de los resultados en el ámbito de estudio



Framework & Stack Tecnológico

1

API Services

- AEMET
- Google Earth Engine (ESA y NASA)

3

Notebook en Google Colab

- GPU Runtime
- Python 3.7

2

Life Cycle Development

- GitHub
- Google Drive

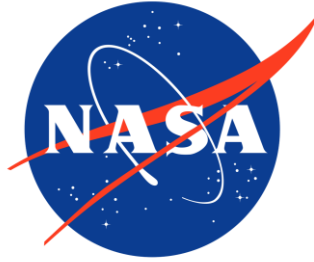
4

Machine Learning Libraries

- Fast.AI
- Pandas



Orígenes de Datos

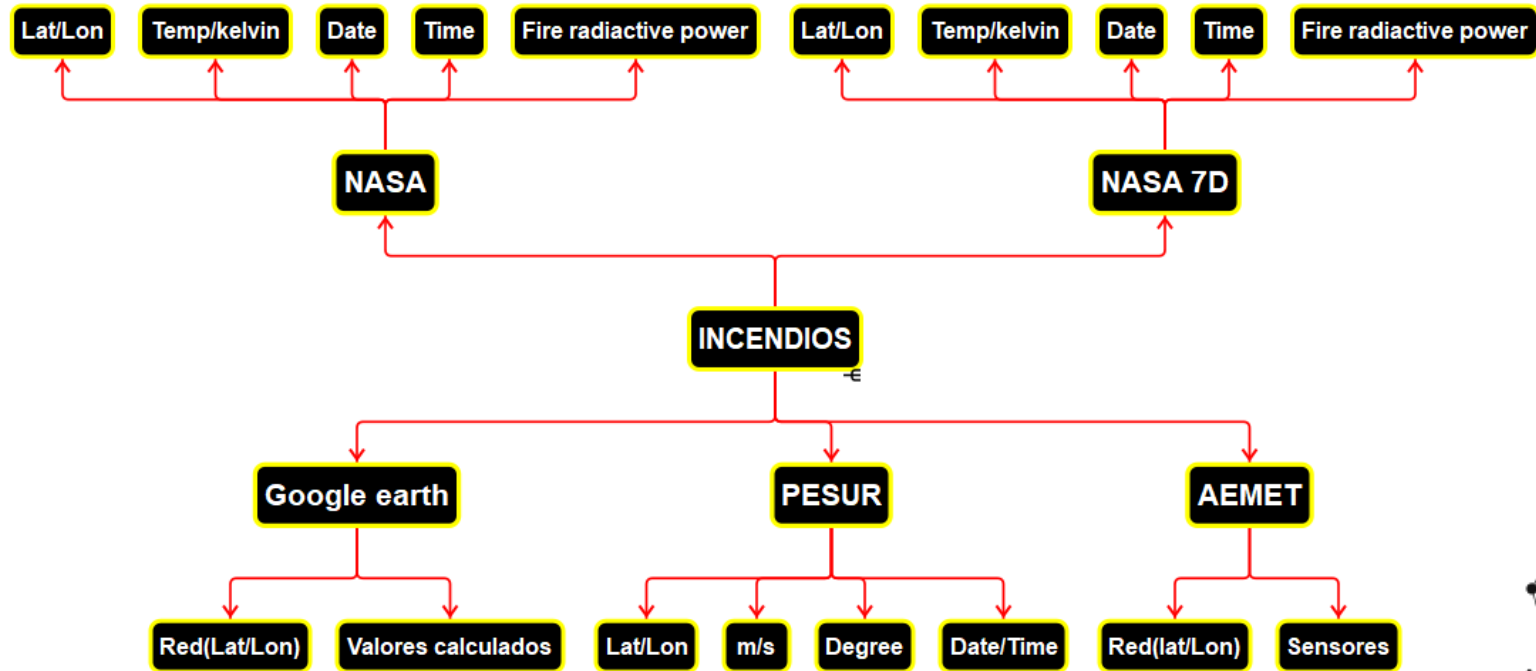


Principales Atributos

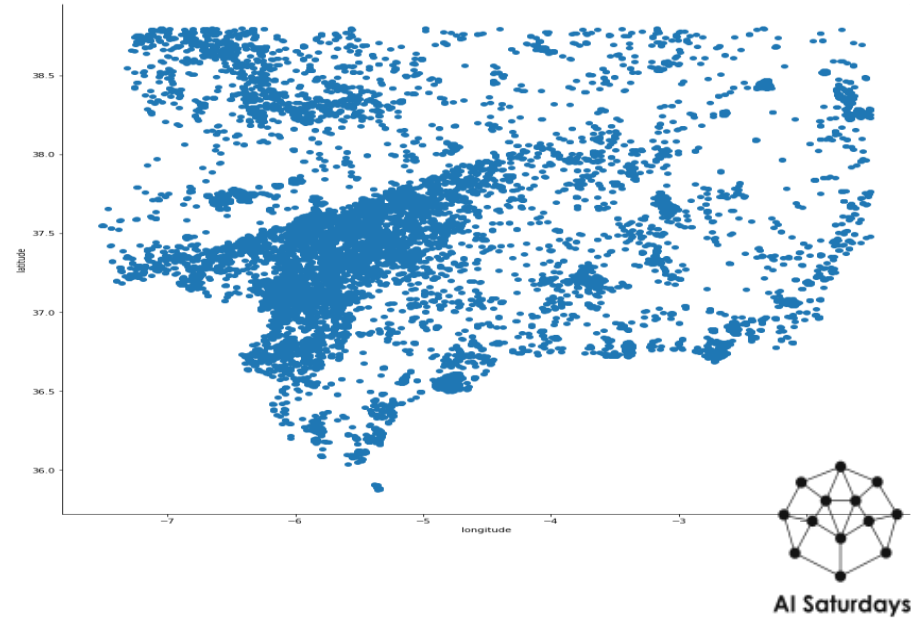
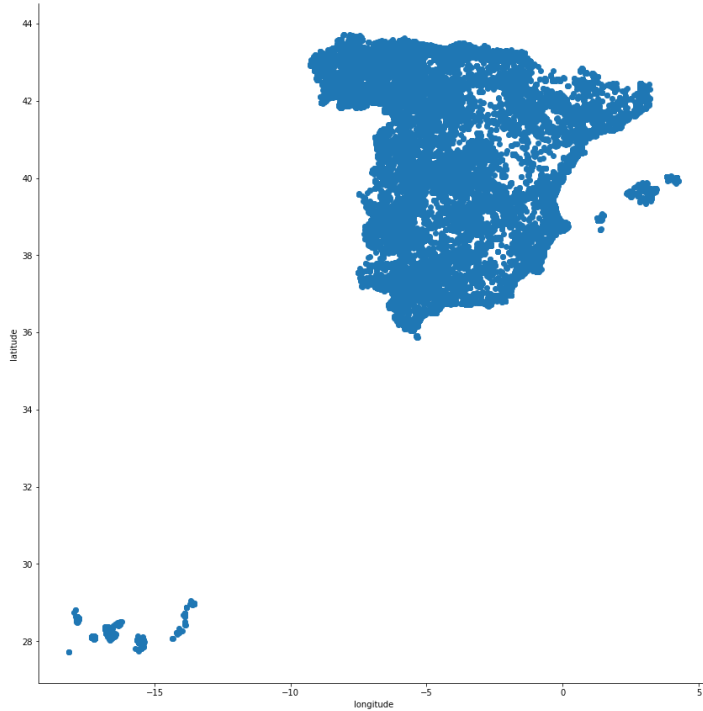
- **FRP** (FIRE RADIACTIVE POWER) : by FIRE INFORMATION FOR RESOURCE MANAGEMENT SYSTEM (FIRMS) by NASA
- **NDVI** (NORMALISED VEGETATION INDEX): by GEE (MODIS)
- **LST** (LAND SURFACE TEMPERATURE): by GEE API (MODIS) y AEMET API
- **DIRECCIÓN DE VIENTO E INTENSIDAD:** by PESUR PARQUE EÓLICO DEL SUR (TARIFA)
- **DATOS CLIMATOLÓGICOS:** by RED DE ESTACIONES CLIMATOLÓGICAS DE ANDALUCÍA (AEMET)



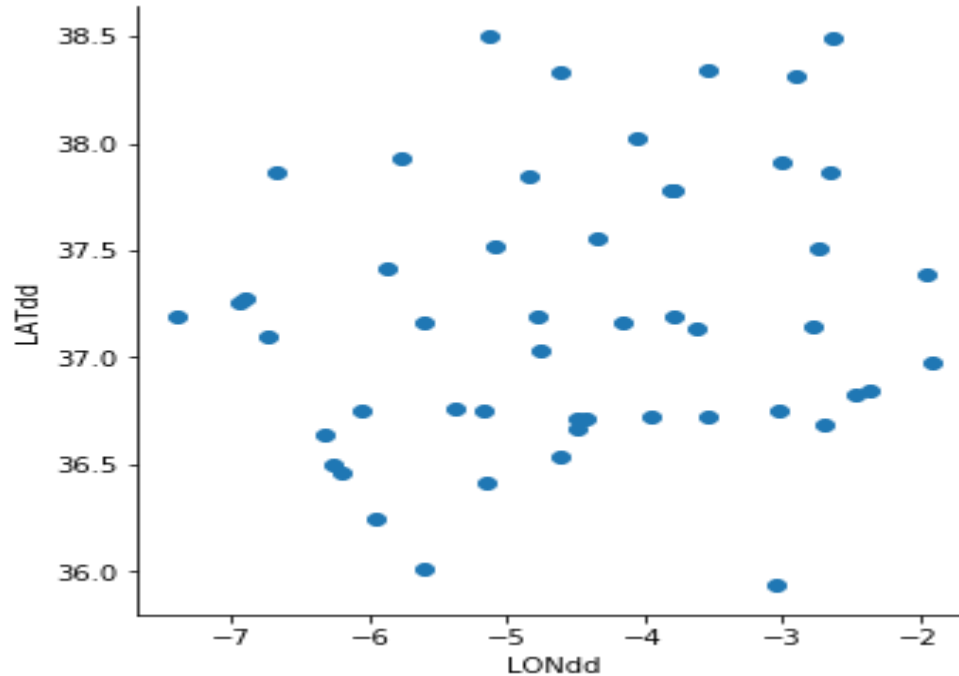
Main Features Map Mind



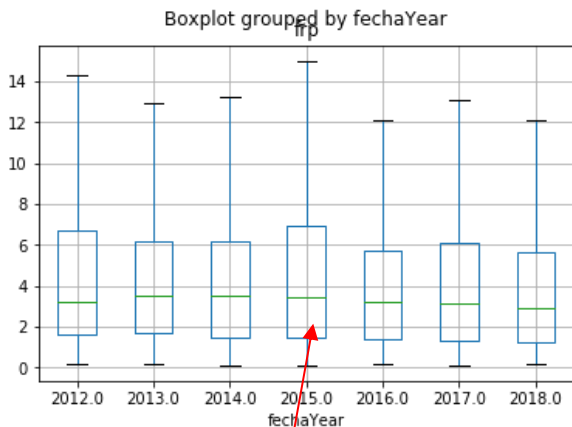
Serie Histórica de Incendios (2014-2018)



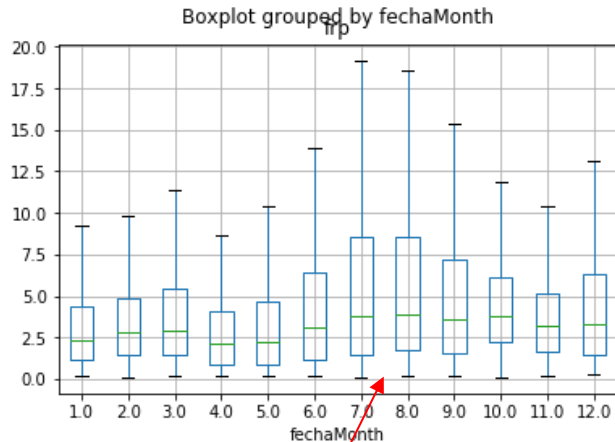
Estaciones Meteorológicas de Andalucía (AEMET)



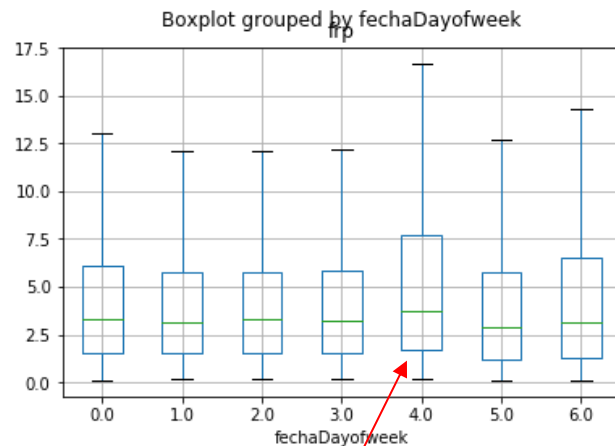
Análisis de las Variables Temporales



2015



Julio-Agosto



Viernes



Diseño de la Solución:

ML Random Forest (Aprendizaje Supervisado)



Estructura del Notebook

1. Descripción del proyecto
2. Fuentes de datos y atributos principales
3. Librerías (Instalación e Imports)
 - Fast.AI, Sklearn, Pandas, Numpy, UTM, Matplotlib, GoogleDriveDownloader, urllib.request
4. Carga de datos
 - Carga de CSV y https requests a APIs (Merge de datasets -> Único Dataframe)
5. Visualización de datos crudos (Matplotlib)
6. Pre-procesamiento de los datos (UTM y Time conversion, normalización, filtrado)
7. Preparación de juegos de datos (entrenamiento y validación)
8. Análisis de relevancia de variables (feature importance)
9. Modelo Predictivo (**Random Forest** -> NEstimators, Iteraciones-> Subsampling, ECM)
10. Optimización de Resultados (Hiperparámetros, Score, Accuracy)
11. Nuestro Modelo





Técnica de Optimización Empleadas

1. Análisis de la importancia de las variables

- Features Importance (Dendrograma)
- Drop de variables correladas
- Categorización

1. Tuning del Modelo

- Análisis de influencia de hiperparámetros
- Búsqueda automática con Grid Search y tuning manual

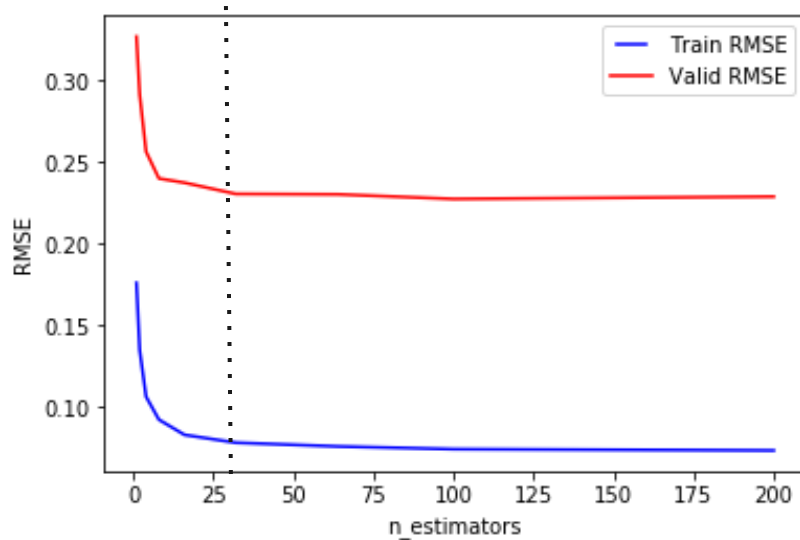
1. Optimización Resultado y fiabilidad

- Balanceado de Score y Accuracy
- Optimización de función de coste (RMSE Mínimos Cuadrados)

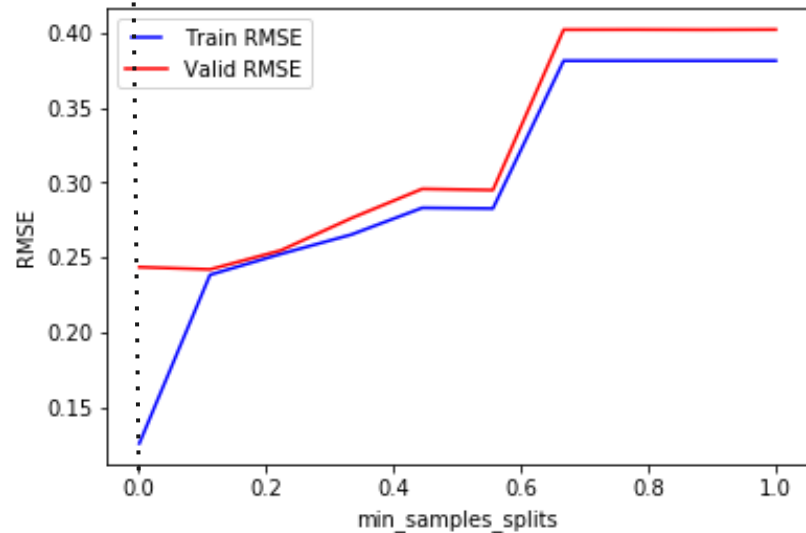


Técnica de Optimización Empleadas

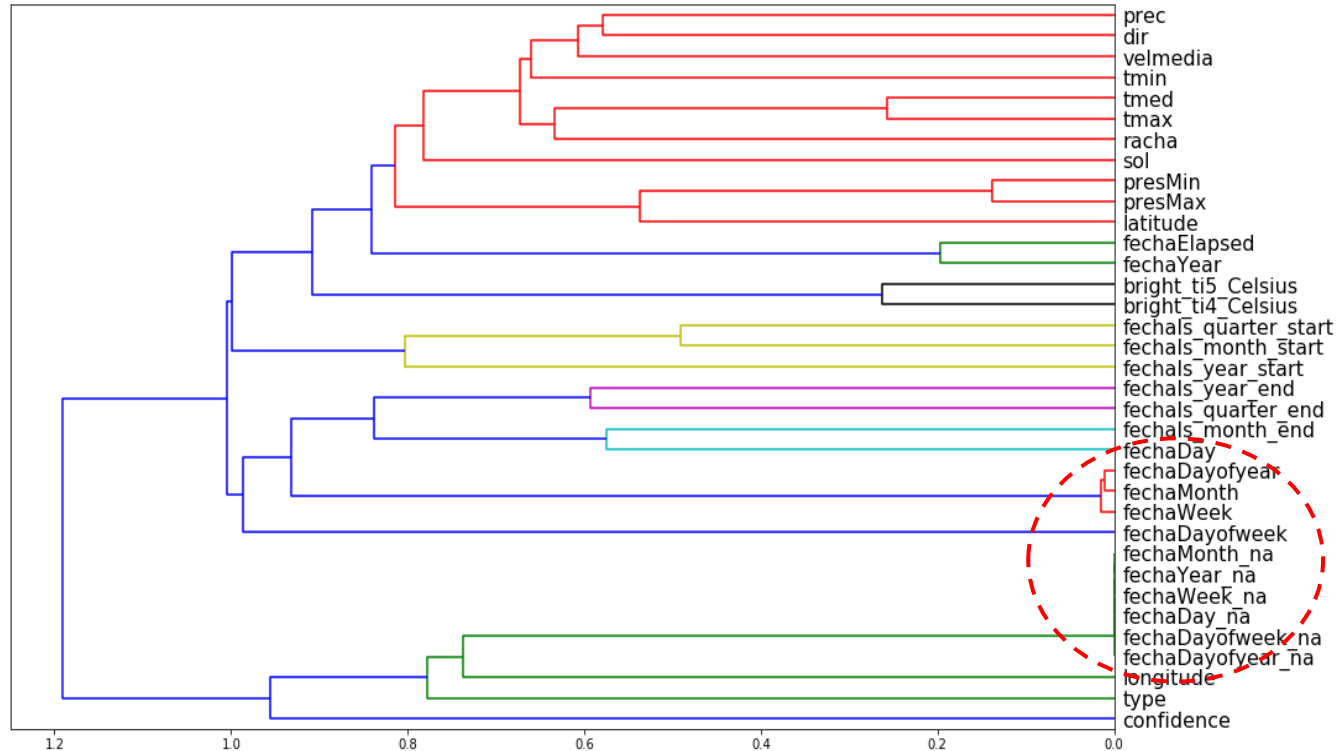
RMSE - N_Estimators



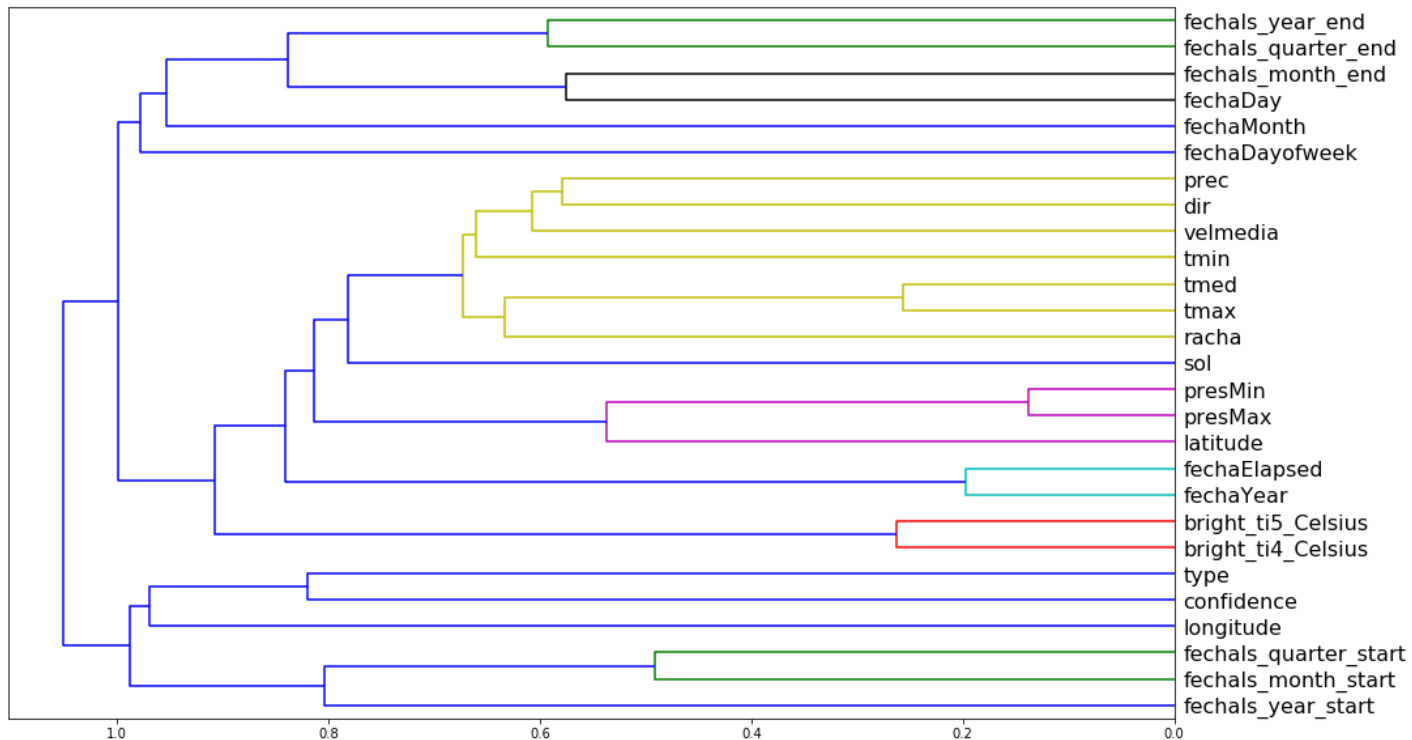
RMSE - Min_Samples_Splits



Dendrograma - Correlación de Variables (Antes)



Dendrograma - Correlación de Variables (Después)



Nuestro Modelo: Características

Random Forest

```
{'bootstrap': True,  
'max_depth': None,  
'max_features': 28,  
'min_samples_leaf': 0.01,  
'min_samples_split': 0.01,  
'n_estimators': 30,  
'oob_score': True}
```

Datasets (50/50)

Training
Validación

Métricas

RMSE

Optimización

Categorización
Eliminación de Nulos
Subsampling
Feature Importance
Redundance Features
Grid Search (hiperparámetros)



Conclusiones del Proyecto

El proyecto basándose en una estructura de ejecución lineal y con **iteraciones de optimización**, en torno a las conclusiones obtenidas de relevancia de parámetros y variables, consigue finalmente, optimizar un modelo basado en el algoritmo **Random Forest**, que es capaz de **estimar el FRP** (intensidad de un incendio) para un potencial futuro incendio.

Project Team



Maribel Luque



Rafael Sánchez



Sergio Gómiz



#AIFellows



AI Saturdays

RF Earth, Wind and Fire

Gracias.



AI Saturdays