

# Manual de instalación y uso del sistema

## Proyecto 3 - Master IABD

Víctor Angulo | Marcos de Castro | Joshua Winn

### 1. Manual de instalación

Para comenzar hay que descargar los archivos que hay en el repositorio del proyecto, ahí está todo lo necesario para poder ejecutar y empezar a trabajar.

Lo primero sería ejecutar el archivo que contiene el environment con todas las librerías necesarias para poder ejecutar los scripts.

Para poder trabajar también se requiere tener Python, Jupyter, Tableau Desktop y algún IDE para poder abrir los archivos.

Tanto en el archivo README.md en el github como en la Documentación oficial del proyecto hay un resumen de cómo están las carpetas estructuradas y dónde está cada archivo importante.

### 2. Manual de usuario:

Lo primero es obtener datos, si se añadiesen datos nuevos se deberían guardar en la carpeta datos/raw, ahí hay otros datos meteorológicos con los que nosotros trabajamos para crear el proyecto.

En la carpeta src/ETL se encuentra un archivo para homogeneizar los datos usando PySpark. Es un archivo simple de utilizar y para que funcione con tus datos solo habría que cambiar la ruta de donde recibe los datos y algún nombre de variable ya que tus datos no tienen las mismas columnas que los nuestros.

Al terminar de ejecutar este archivo se habrán creado en data/homogenized tus datos particionados.

Ahora ya puedes empezar a pensar qué modelos quieres utilizar en la carpeta src/modeling hay varios ejemplos tanto para datos diarios como horarios. Para que los modelos funcionen bien es recomendable crear nuevas columnas como el desglose de fechas, cambio de fechas a seno y coseno o lags que puedan ayudar a la mejora de la calidad de la predicción del modelo.

Cuando se sepa que datos y como se quieren utilizar en src/ETL hay tambien un archivo dedicado a transformar los datos de homogeneizados a los valores que si los añades a un modelo te devuelve un resultado.

Por último, una vez se tienen los resultados tenemos una visualización hecha con Tableau para que puedas poner los datos o el EDA de una manera más visual como en un mapa o en diferentes graficos.