



## **Proyecto Big Data I**

# **Instrucciones proyecto: TerraWatt**




Alma Gutiérrez Pesquero

Rafael Borge Mera

Deniz Alcobendas Çakmak

Iñigo Pérez

Profesor: Álvaro Sanchez Pérez



# Índice

<b>Explicación inicial.....</b>	<b>1</b>
Predicción del Consumo Energético.....	1
Predicción del Precio de la Electricidad.....	2
Resultados Esperados.....	2
<b>Cómo ejecutar el proyecto.....</b>	<b>3</b>

# Explicación inicial

Debido a el impulso por parte de la Unión Europea por la eficiencia energética y el alto costo de la electricidad que abarca gran parte de los presupuestos de las familias, nos han animado a poder buscar soluciones que les permitan mejorar y concienciarse de su consumo. Para ello hemos creado la empresa de “**TerraWatt**”, nuestra misión es analizar, calcular y optimizar las facturas de electricidad, con meses de antelación, pudiendo así hacer una previsión de gastos, pudiendo así tomar decisiones informadas, para poder tener un consumo responsable.

Para ello en este proyecto vamos a usar técnicas de análisis de datos, modelado predictivo y la inteligencia artificial, para poder desarrollar los modelos.

Para ello vamos a considerar distintos aspectos como pueden ser la potencia contratada, el número de personas viviendo en el hogar, el tipo, los aspectos meteorológicos... Todo esto quedará explicado posteriormente.

## Predicción del Consumo Energético

El consumo energético está influenciado por muchos factores tanto internos como externos, en este caso nuestro modelo varios de estos factores para saber si son influyentes, en concreto estos incluirán:

- + Características de la vivienda:
  - + Tipo de vivienda
  - + Potencia Contratada
- + Factores demográficos:
  - + Número de residentes
  - + Comportamiento del consumo en función de los horarios y día de la semana.
- + Condiciones meteorológicas:
  - + Temperatura media
  - + Temperatura mínima / máxima
  - + Precipitaciones
  - + Viento
  - + Presión mínima máxima
  - + Horas de luz
- + Días festivos y laborables
  - + Análisis de su influencia

Nuestro proyecto emplea técnicas de regresión avanzada y modelos de aprendizaje automático, como pueden ser el Random Forest, Gradient Boosting y las redes neuronales, aunque deberemos de hacer posteriormente un análisis especializado de cada uno de ellos para saber cuál es el que mejor se adapta a nuestros datos y cuál nos ofrece una mejor predicción.

## Predicción del Precio de la Electricidad

En nuestro caso el precio de la energía es un factor super importante para poder calcular y optimizar la facturación. Buscaremos poder predecir los precios a partir de datos históricos, además de intentar encontrar aquellos factores muy influyentes, aunque sabemos que este precio también es influenciado por muchos otros factores sociales o económicos, como puede ser la última guerra de Ucrania o el precio del oro. Por lo que sabemos que en algunos momentos sociales y económicos se nos hará imposible predecirlo con exactitud, aunque intentaremos tener el menor error posible aumentando su exactitud al máximo. Para ellos hemos decidido analizar los siguiente factores:

- + Condiciones metereológicas
  - + Temperatura media
  - + Temperatura mínima / máxima
  - + Precipitaciones
  - + Viento
  - + Presión mínima máxima
  - + Horas de luz
- + Precio de la electricidad de los últimos 10 años
- + Eventos externos predecibles
  - + Festivos
  - + Días laborables.

Vamos a analizar estos factores, teniendo en cuenta la posibilidad de poder aumentar estos factores en base de la necesidad.

Realizaremos los modelos a través de series temporales como pueden ser ARIMA y LSTM, ya que pueden analizar y capturar esos patrones históricos, que esperamos encontrar, además de prever posibles fluctuaciones. Además analizaremos mediante correlación cruzada todas las variables para poder establecer cuales son las relaciones más fuertes. Todo esto lo realizaremos en la siguiente fase del trabajo.

## Resultados Esperados

Con esto esperamos obtener predicciones del consumo en base a la provincia de manera diario que se podrá ajustar dependiendo de tu hogar proporcionando pequeños datos. Así como el precio de la electricidad pudiendo así hacer previsiones a largo plazo, para que las viviendas tengan un mejor manejo de los datos.

Por otro lado también intentaremos concienciar y mejorar el impacto energético.

# Cómo ejecutar el proyecto

Primeramente se proporcionará una carpeta en la se incluirán diferentes documentos en concreto:

- Datos brutos:
  - Datos brutos generales: esto es una carpeta en la que podremos encontrar 3 csv principales, en ellos encontramos 3 archivos diferentes que constan de:
    - Consumo Energético de Viviendas, en este csv encontramos los datos desde el 2024, del consumo energético diario, dividido por provincia, la media de residentes, la potencia contratada y el tipo de vivienda.
    - Festivos, es una lista con los festivos nacionales y provinciales que aplicaremos a ambos modelos.
    - Precios de energía, este csv se generará posteriormente gracias al código, y se incluirá actualmente en esta carpeta.
  - Datos brutos meteorológicos: dentro de esta carpeta encontraremos los datos de las 947 estaciones meteorológicas, los datos diarios con los factores explicados anteriormente.
- Archivos python de limpieza:
  - Datos limpieza meteorológicos: En este archivo realizaremos el procesamiento y limpieza de los datos, debido a lo que hemos comentado antes, tenemos 927 estaciones meteorológicas, por lo que vamos a juntar aquellas estaciones que pertenezcan a la misma provincia para poder limpiar los datos y que sean más atractivos y simples de comprensión. Por otro lado también tener en cuenta que contamos con los datos desde 1968, como solo tenemos los datos desde el 2010 en el caso del precio y desde 2024 en el caso del consumo, eliminaremos los datos que no nos son necesarios para optimizar también entre otro el tiempo de carga.
  - Extracción precios API: en este archivo nos encontraremos con el proceso de obtener los precios de la electricidad desde el 2010 diariamente.
  - Generación csv modelos: Con este archivo realizaremos la unión necesaria de los datos que hemos limpiado meteorológicos, el precio de electricidad, las festividades... En este caso generamos dos csv en los cuales contendrán las variables explicadas al principio del proyecto.
- Datos limpios: todos los datos que se encuentran en estas carpetas, se han generado gracias a la limpieza de datos que hemos realizado anteriormente en los scripts de python.

- Datos limpios generales: En este apartado modificamos y adaptamos los csv de Consumo Energético y de Precios, para que todos nuestros datos tengan el mismo formato a la hora de procesarlos juntos.
- Datos limpios meteorológicos: Dentro de esta carpeta encontraremos un csv por cada provincia con los datos de todas las estaciones meteorológicas de esta, realizando la media para su agrupación.
- Modelos: Gracias al script de Generación csv modelos, se generan los dos archivos con todos los datos para poder aplicarlos posteriormente.

Debemos de tener en cuenta que este proyecto necesita la ejecución con Docker, por lo que también contamos con un archivo llamado Dockerfile en el que encontraremos las instrucciones a ejecutar en docker.

Una vez explicado esto procederemos a explicar cómo se debe proceder a la ejecución.

Todo este proceso lo ejecutaremos a través de la terminal, por lo que primeramente deberemos de abrirla, esto se puede hacer buscando en nuestro ordenador cmd, recomiendo siempre ejecutar el cmd como administrador. Al ejecutarlo como administrador debemos de establecernos en la ruta donde este nuestra carpeta:

```
cd {Ruta_de_la_carpeta}
```

```
cd Desktop\TerraWatt\Limpieza_datos
```

Una vez hecho esto primeramente una network y el contenedor de elasticsearch. Recomendamos primero visualizar qué puertos están disponibles, para no tener problemas.

```
docker network create elasticsearch-network-pdb1
```

creamos el contenedor del elasticsearch

```
docker run -d --name elasticsearch-pdb1 --net elasticsearch-network-pdb1 -p 9202:9200 -p 9203:9300 -e "discovery.type=single-node" -e "xpack.security.enabled=true" -e "ELASTIC_PASSWORD=changeme" elasticsearch:8.10.2
```

Una vez hecho esto debemos de crear la imagen de nuestra image:

```
docker build -t terrawatt_image .
```

A continuación, crearemos el contenedor sin ejecutarlo ya que vamos a proceder a forzar la conexión con la network, ya que ha sido uno de los problemas que hemos tenido y que más nos han retrasado.

```
docker create --name Terrawatt_container terrawatt_image
```

Como hemos comentado procederemos a la conexión forzada a la network

```
docker network connect elasticsearch-network-pdb1 Terrawatt_container
```

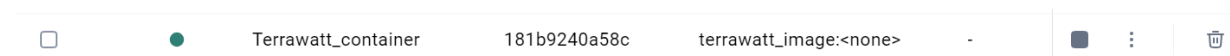
Por lo que una vez hecho esto ya podremos inicializar el contenedor, se nos mostraran los resultados dentro de la app de docker, dentro del contenedor.

```
docker start Terrawatt_container
```

Una vez ejecutado esto, que tardará unos 30 minutos aproximadamente, se recomienda descargar los datos limpios y los modelos para poder visualizar todos los cambios así como para entender mejor los datos y el modelo, luego los subiremos a elasticsearch a partir de ahora realizar las consultas desde nuestra nueva base de datos. Para ellos en la terminal de Visual Studio (o del programa que se haya usado) y deberemos de ejecutar los siguiente comandos:

```
docker cp {id_container}:/app/{archivos_a_descargar} ./{archivos_a_descargar}
```

El id\_container lo encontraréis dentro de la aplicación de docker en la misma sección de Containers en la segunda columna. Como se puede ver aquí, en mi caso el id sería 181b9240a58c



Por otro lado los archivos también podemos obtener estos archivos dentro de la propia app de Docker, dentro de nuestro contenedor, si nos vamos al apartado de “Files”, debemos de entrar en la subcarpeta “App”, dentro de esta carpeta encontraremos todos nuestros datos.

Por otro lado los archivos a descargar que recomendamos serían:

+ Datos_limpios_generales	+ Modelo_Consumo_Met_fest.csv
+ Datos_limpios_meteorologicos	+ Modelo_Precios_Met_Fest.csv

Importante recordar que recomendamos ejecutar este proceso en la terminal como administrador.

Por otro lado tenemos una demo de la página web: en la cual debemos de acceder mediante la terminal, debemos de elegir primero donde esta nuestra carpeta, en nuestro caso de encuentra en:

```
cd C:\Users\Alma\Desktop\TerraWatt\python-website
```

Para poder ejecutar y obtener la url en local para poder visualizar la página web, es necesario tener instalado python en el ordenador, si no se puede descargar en aquí:

<https://www.python.org/downloads/>

Por otro lado es necesario tener instalada la librería de flask, para ello en el propio símbolo de sistema aplicaremos el comando:

```
pip install flask
```

Acto seguido ejecutaremos el archivo py principal

```
python index.py
```

Una vez hecho esto nos aparece el localhost de nuestra página web.

