



**UNIVERSITAT OBERTA DE CATALUNYA**

**PRÁCTICA 1 – WEB SCRAPING**

**M2.851 - Tipología y ciclo de vida de los datos**

**Autores:**

**Jorge Alonso Hernández**

**Inés Vidal Sospedra**



## 1. Contexto

El tiempo forma parte de la vida diaria de las personas ya que, dependiendo de la previsión meteorológica, elegimos qué ropa ponernos y/o elegimos realizar un plan u otro. No obstante, más allá de ese aspecto lúdico, el tiempo es fundamental para el desarrollo de ciertas actividades humanas, desde la generación de electricidad hasta la facilidad o dificultad para transportar bienes y personas. Además, los fenómenos meteorológicos y climáticos extremos, como olas de calor o frío, sequías o lluvias torrenciales, pueden tener enormes impactos sociales y causar daños a la propiedad y las personas.

Por todas estas razones, es crucial invertir en mejorar el pronóstico del tiempo, porque aunque la atmósfera es una entidad desordenada y nunca podemos producir pronósticos del tiempo absolutamente precisos, es especialmente importante mejorar las predicciones meteorológicas para que se puedan tomar las medidas necesarias ante pronósticos meteorológicos adversos, tan pronto como sea posible.

Por ello, en este documento se presenta una descripción de un proyecto que pretende crear una herramienta que facilite el *web Scraping* para recoger, procesar y almacenar en un dataset ordenado, datos relativos al tiempo que permitan realizar un posterior análisis de las predicciones meteorológicas de un país.

## 2. Título

Teniendo en cuenta el contexto anterior, el título del dataset será “**El tiempo en España durante la próxima semana**”.

## 3. Descripción del dataset

El dataset generado como resultado del proyecto de *web Scraping* consta de diferentes registros en los que se recoge información obtenida de la página AEMET tales como temperatura, probabilidad de precipitación, cota de nieve, etc para cada una de las provincias de España. Además, para cada una de ellas se obtienen datos del tiempo para un total de 7 días.

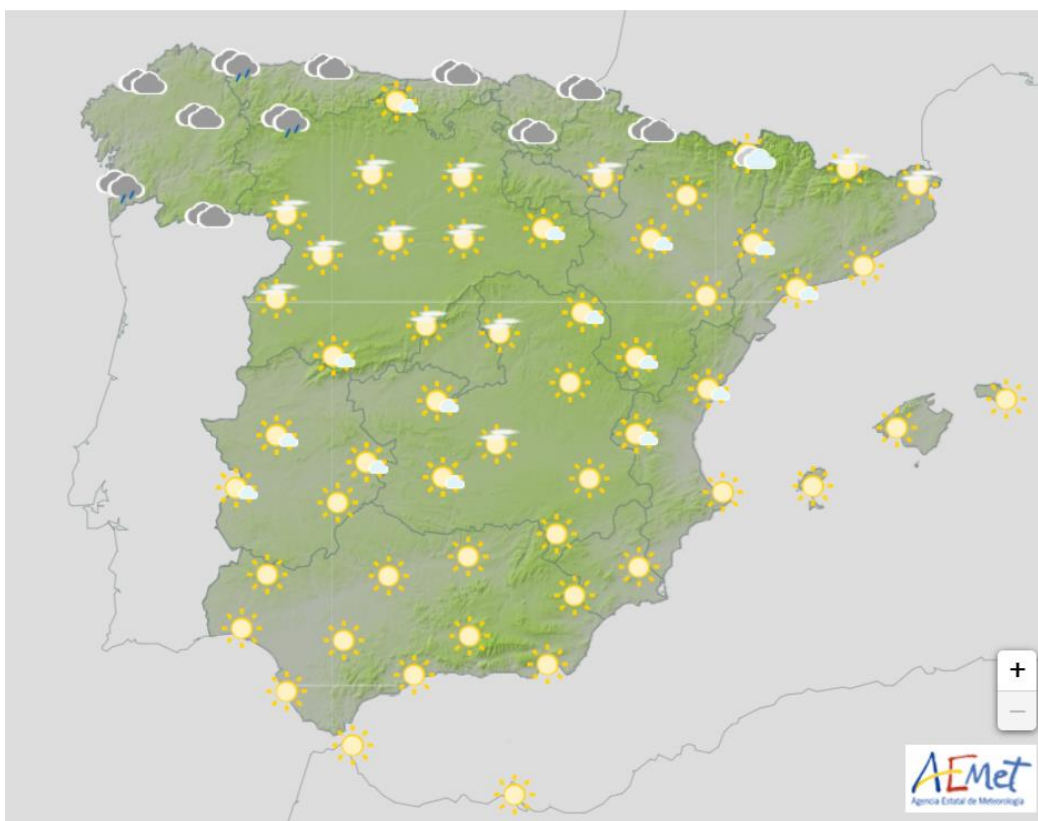
## 4. Contenido

Para cada provincia, se han obtenido varios registros con los siguiente atributos:

- **Municipio:** nombre del municipio de España
- **Provincia:** nombre de la provincia a la que pertenece el municipio
- **Fecha:** día al que se refiere la información obtenida
- **Horario:** rango de horas al que corresponden los datos.
- **Estado del cielo:** estado de la atmósfera en función de la nubosidad y precipitación.
- **Temperatura actual:** temperatura esperada en el horario especificado.
- **Probabilidad de precipitación:** probabilidad de que se produzca cualquier cantidad de precipitación en el período considerado.
- **Cota de nieve:** altitud mínima a partir de la cual las precipitaciones son en forma de nieve
- **Temperatura mínima:** temperatura mínima esperada para el periodo y lugar especificado
- **Temperatura máxima:** temperatura máxima esperada para el periodo y lugar especificado

- **Sensación térmica mínima:** sensación aparente de calor o frío prevista que tienen las personas en función de los parámetros que determinan el ambiente (temperatura, humedad y viento)
- **Sensación térmica máxima:** sensación aparente de calor o frío prevista que tienen las personas en función de los parámetros que determinan el ambiente.
- **Humedad mínima:** valor más bajo de humedad relativa en el día, expresada en porcentaje
- **Humedad máxima:** valor más alto de humedad relativa en el día, expresada en porcentaje
- **Dirección del viento:** dirección media de donde viene el viento en relación a los 8 rumbos principales.
- **Velocidad del viento:** velocidad del viento
- **Racha máxima del viento:** Velocidad máxima del viento en el día indicado, en km/h

## 5. Representación gráfica



En esta imagen podemos observar el estado del cielo en las diferentes provincias de la península. Se trata de una imagen representativa ya que permite visualizar uno de los atributos contenidos en el dataset obtenido.

## 6. Agradecimientos

Los datos han sido extraídos del portal Datos abiertos de AEMET que posibilita la reutilización de la información en el sentido establecido en la Ley 18/2015.

Más específicamente, de la sección en la que se encuentra la predicción del tiempo por municipios (<http://www.aemet.es/es/eltiempo/prediccion/municipios>).

Para su obtención, se ha utilizado el lenguaje de programación Python junto con técnicas de *web Scraping* a partir de las cuales ha sido posible descargar la web y leer los datos de esta.

## 7. Inspiración

Tal como se mencionaba en el apartado Contexto de este documento, el tiempo es fundamental para el desarrollo de ciertas actividades humanas, desde la generación de electricidad hasta la facilidad o dificultad para transportar bienes y personas.

Por ello, el dataset obtenido resulta de gran interés ya que a este se le pueden aplicar técnicas de *machine learning* para realizar predicciones de tiempo futuras y poder organizar las actividades con conocimiento del temporal, mejorando así la eficiencia de ciertas profesiones.

Además, otra opción sería la de aplicar algoritmos de **aprendizaje no supervisado** para agrupar los datos no estructurados según sus similitudes y patrones distintos en el conjunto de datos, lo cual permitiría conocer, entre otras cosas, que provincias tienen climas similares o completamente diferentes.

## 8. Licencia

La licencia elegida para el conjunto de datos ha sido **CC BY-SA 4.0 License**, lo cual implica:

- Se debe mantener el nombre del creador del conjunto de datos y especificar los cambios realizados respecto a la versión original
- Se posibilita el uso de los datos para realizar otros trabajos derivados o basados en los mismos siempre que dichos trabajos se compartan con licencias compatibles a la presente.
- Se permite el uso de estos datos con fines comerciales



## 9. Código

El código, compuesto por 3 archivos .py (DiaHoraMunicipio.py, Scraper.py y main.py) está disponible en el repositorio <https://github.com/jalonsouoc/Practica1-TCVD>

## 10. Dataset

El dataset creado durante el desarrollo del proyecto se encuentra en formato CSV en Zenodo DOI <https://doi.org/10.5281/zenodo.6433623>

## 11. Bibliografia

- [1]. Alternativo, E., Guarda, A. y Guarda, A. (2022). La influencia de la meteorología en nuestras vidas. El Blog Alternativo. Recuperado el 9 de abril de 2022, de <https://www.elblogalternativo.com/2018/02/08/descubre-la-influencia-la-meteorologia-nuestras-vidas-algunas-profesiones/>.
- [2]. La importancia de la previsión - EL ÁGORA DIARIO. EL ÁGORA DIARIO. (2022). Recuperado el 9 de abril de 2022, de <https://www.elagoradiario.com/tribunas/importancia-prediccion-meteorologica-aemet/>.
- [3]. Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.
- [4]. License Selector. (s. f.). Recuperado 9 de abril de 2022, de <https://ufal.github.io/publiclicense-selector/>
- [5]. Subirats Maté, L., & Calvo González, M. (2019). Web Scraping (pp. 1-66). Universitat Oberta de Catalunya.

## ANEXO: Contribuciones

Contribuciones	Firma
Investigación previa	JAH – IVS
Redacción de las respuestas	JAH – IVS
Desarrollo del código	JAH – IVS