

RETO B – Social: Analítica de datos

Título del reto: **Predicción de consumo energético en edificios**

Centro promotor del reto: **CTIC**

Se plantea como reto la predicción del consumo energético de los dos edificios de oficinas de CTIC, localizadas en Asturias (Gijón y Peón). Dadas las características propias de un contexto laboral con horario de oficina, sus patrones de consumo son relativamente estables a lo largo del tiempo.

Así, se proporcionarán a los participantes los datos reales de consumo energético de **22 meses** con datos recogidos cada hora. Dichos datos serán proporcionados en ficheros estructurados en formato CSV, con la información propia de cada uno de los dos edificios en sendos archivos, donde se incluirán fecha, hora y consumo en kilovatio hora. Se presenta a continuación un ejemplo del formato que se proporciona en cada uno de ellos, con el encabezado y varias entradas asociadas.

```
fecha;hora;consumo_kWh
"2022/01/01";"01:00";"0,168"
"2022/01/01";"02:00";"0,040"
"2022/01/01";"03:00";"0,169"
"2022/01/01";"04:00";"0,042"
"2022/01/01";"05:00";"0,179"
"2022/01/01";"06:00";"0,041"
"2022/01/01";"07:00";"0,201"
"2022/01/01";"08:00";"0,045"
"2022/01/01";"09:00";"0,166"
"2022/01/01";"10:00";"0,040"
"2022/01/01";"11:00";"0,214"
"2022/01/01";"12:00";"0,039"
"2022/01/01";"13:00";"0,163"
"2022/01/01";"14:00";"0,040"
"2022/01/01";"15:00";"0,164"
```

Los participantes deberán utilizar dichos datos para entrenar modelos predictivos del consumo energético. Una vez obtenidos los modelos que consideren que obtengan los mejores resultados, deberán generar las predicciones de ambos edificios para los 2 meses siguientes, con **predicciones horarias**. Dichas predicciones deberán ser enviadas a la organización del datathon, a través de dos ficheros CSV para cada edificio (gijon.csv, peon.csv), donde se disponga de las predicciones en dichos intervalos **siguiendo la misma estructura que los CSV originales** (mismo encabezado, formato de fecha y hora, uso de separador ";", uso de separador decimal ",").

Esta información será la utilizada para validar la bondad de los modelos desarrollados a partir de diferentes métricas que permitan evaluar tanto el nivel de predicción en los instantes temporales iniciales como los más avanzados en el tiempo.

El dataset completo se facilitará al equipo participante una vez inscrito en el reto.