



AI4ES

RED DE EXCELENCIA EN
TECNOLOGÍAS HABILITADORAS
BASADAS EN EL DATO

ANEXO: Retos del Datathon AI4ES 2023

RETO A – Industrial: Procesamiento de Imágenes

*Título del reto: **Detección de defectos superficiales en madera***

Centro promotor del reto: ITI

El "Reto de Detección de defectos superficiales en madera" es una iniciativa que aborda un desafío crucial en la industria, en particular, la identificación de anomalías en superficies de madera. Este reto se enfoca en desarrollar algoritmos que puedan reconocer y distinguir texturas de madera normales de texturas anómalas, como grietas, manchas, porosidad y nudos, que podrían indicar un producto defectuoso o inseguro. La detección de estas anomalías es fundamental para garantizar la calidad y la seguridad de los productos manufacturados que utilizan madera en su fabricación.

El objetivo principal de este reto es desarrollar algoritmos de detección de anomalías que puedan aprender a identificar lo que se considera una textura de madera "normal" y determinar si una textura de madera bajo inspección es anómala.

Los participantes en este reto tendrán acceso a un conjunto de datos que consta de múltiples vistas de objetos con textura de madera de la misma clase. Cada imagen de este conjunto de datos ha sido etiquetada manualmente a nivel de píxel para identificar las áreas anómalas, incluyendo grietas, manchas, porosidad y nudos. Además, cada imagen ha sido etiquetada por dos equipos diferentes, lo que agrega un nivel de confianza a las etiquetas finales. El conjunto de datos se divide en tres particiones: entrenamiento, validación y prueba. Las particiones de entrenamiento y validación contienen solo texturas de madera "normales", mientras que la partición de prueba incluye tanto texturas "normales" como texturas anómalas.



El algoritmo a desarrollar debe tomar una imagen como entrada y generar un mapa de errores con una puntuación de anomalía para cada píxel. Los aspectos a tener en cuenta son:

1. Exactitud de detección: Los algoritmos deben ser capaces de detectar y localizar correctamente las áreas anómalas en las texturas de madera, minimizando los falsos positivos y falsos negativos.
2. Generalización: Los algoritmos deben ser capaces de generalizar su capacidad de detección a nuevas texturas de madera no vistas durante el entrenamiento. Esto implica la capacidad de adaptarse a la variabilidad natural en la textura y el color de la madera.
3. Eficiencia: Los algoritmos deben ser eficientes en términos de tiempo de procesamiento, lo que garantiza su aplicabilidad en entornos industriales donde la velocidad es esencial.
4. Robustez: Los algoritmos deben ser capaces de lidiar con incertidumbres y variabilidades en las texturas de madera, proporcionando resultados consistentes y confiables.

El rendimiento de los algoritmos se evaluará utilizando el conjunto de datos de prueba, y los criterios de éxito se basarán en métricas que cuantifiquen la precisión de la detección de anomalías. Los equipos participantes serán clasificados según su capacidad para cumplir con estos criterios.

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

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.