

MIEMBROS:









ENTIDADES FINANCIADORAS









Agenda

- 1. Introducción y presentación convocatoria
- 2. Retos Datathon 2022
 - → Reto A Industrial: Procesamiento de imágenes
 - → Reto B Social: Analítica de datos
- 3. Operativa de trabajo
- 4. Preguntas



















Al4ES es una red de excelencia en tecnologías habilitadoras basadas en el dato

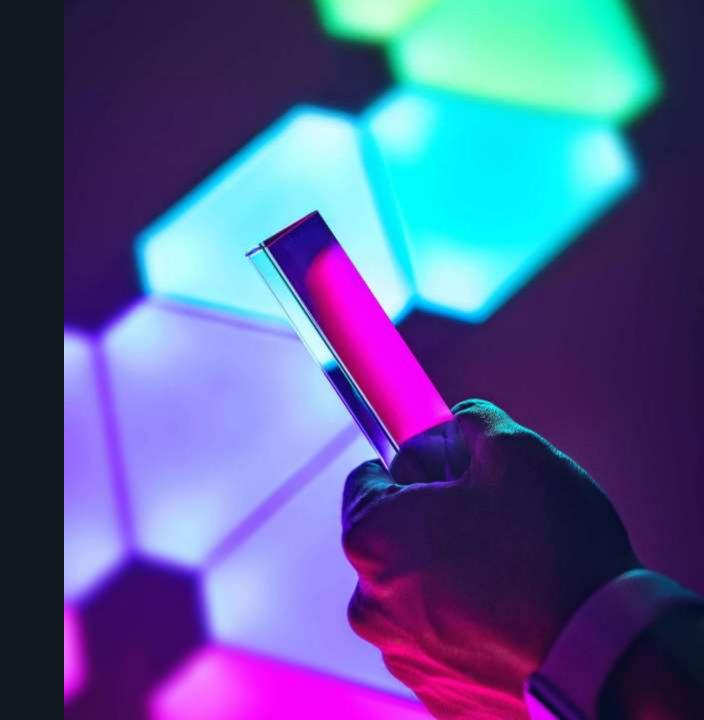
Quienes somos

- → 4 Centros Tecnológicos I+D
- → Capacidades y experiencia en tecnologías Big Data, IA (ML y DL) e infraestructuras computación para manejo datos
- → +280 profesionales (investigadores e ingenieros) en estas tecnologías
- → Amplio posicionamiento en foros y plataformas tecnológicas internacionales
- → +90 proyectos europeos I+D 2018-2020
- → +127 empresas españolas colaboradoras en proyectos I+D europeos



El objetivo de AI4ES es potenciar la economía basada en datos e inteligencia artificial a nivel nacional





Por qué Al4ES?

Credits: The Indian Express



Las tecnologías digitales están cambiando drásticamente la forma en que diseñamos, producimos, comercializamos y consumimos todo tipo de bienes y servicios.

La especialización en estas tecnologías genera efectos significativos y positivos en el rendimiento económico de las regiones, como pueden ser:

- → Impacto sustancial en la **creación de empleo altamente cualificado**
- → Relevancia sistémica en todas las fases de desarrollo de producto, entre diversas cadenas de valor industrial.
- → Capacidad para mejorar la salud, la seguridad, el desarrollo sostenible y la conectividad entre los sistemas y los individuos
- → Habilitador de múltiples **aplicaciones industriales en múltiples sectores**, contribuyendo a la generación de excelencia global, nuevo conocimiento y nuevas formas de participación

Nuestra visión



Ser referente español en I+D+I y transferencia en Tecnologías Habilitadoras Digitales relativas al procesamiento y análisis inteligente de datos

Potenciar la economía basada en datos e inteligencia artificial a nivel nacional

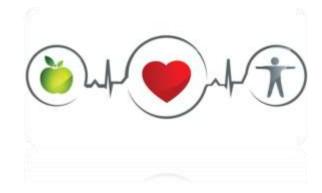


Agregar, de forma armonizada y coordinada, las actividades, capacidades, e infraestructuras de I+D+I en tecnologías de aprendizaje automático y extracción de valor de los datos



Cerrar la brecha entre la investigación y la industria, facilitando la transferencia y buscando el impacto, principalmente en las PYMES, proporcionando soluciones y servicios innovadores que requieren de acceso seguro y confiable a datos, procesamiento eficiente, análisis avanzado, aprendizaje automático e inteligencia artificial





Retos Tecnológicos

La actividad de la red se estructura en 4 objetivos científicotecnológicos, que constituyen los 4 pilares de la red



•Modelos de aprendizaje para datos heterogéneos (multimodales, de diversas fuentes), degradados o de baja calidad (datos faltantes, insuficientes, no significativos, erróneos, con anomalías, etc.)

OT 1: Data Analytics

•Modelos de aprendizaje continuo y sistemas que soporten su ciclo de vida, que permitan ir incorporando nuevo conocimiento de forma incremental de forma automática o con una mínima supervisión



OT 2: IA Development



• Nuevos paradigmas de computación especializados para Inteligencia Artificial, que permitan aumentar la eficiencia del cómputo de algoritmos de aprendizaje automático, explorando conexiones con GAIA-X.

OT 3: Cloud y HPC Arquitecturas para la construcción de Espacios de Datos, definiendo modelos de gobernanza apropiados para captar los datos a través de un espacio de confianza.

OT 4: Data Spaces



Retos Tecnológicos y Estratégicos

OE1: Colaboración entre centros + Masa crítica Eragmentación

OE2: Marca Al4ES
+ Reconocimiento
+ Talento y Carrera Investigadora

OE3: Agenda Al4ES 2021-2023 Retos tecnológicos y hoja de ruta para abordarlos

OE4: Infraestructuras cadena de valor del dato Investigar y experimentar con datos

OE5: Impulso economía basada en el dato Proyectos implantación y transferencia tecnologías

OE6: Prospectiva, Diagnóstico y Formación

•Modelos de aprendizaje para datos heterogéneos (multimodales, de diversas fuentes), degradados o de baja calidad (datos faltantes, insuficientes, no significativos, erróneos, con anomalías, etc.)

OT 1: Data Analytics •Modelos de aprendizaje continuo y sistemas que soporten su ciclo de vida, que permitan ir incorporando nuevo conocimiento de forma incremental de forma automática o con una mínima supervisión

OT 2: IA Development

• Nuevos paradigmas de computación especializados para Inteligencia Artificial, que permitan aumentar la eficiencia del cómputo de algoritmos de aprendizaje automático, explorando conexiones con GAIA-X.

OT 3: Cloud y
HPC

 Arquitecturas para la construcción de Espacios de Datos, definiendo modelos de gobernanza apropiados para captar los datos a través de un espacio de confianza.

OT 4: Data Spaces



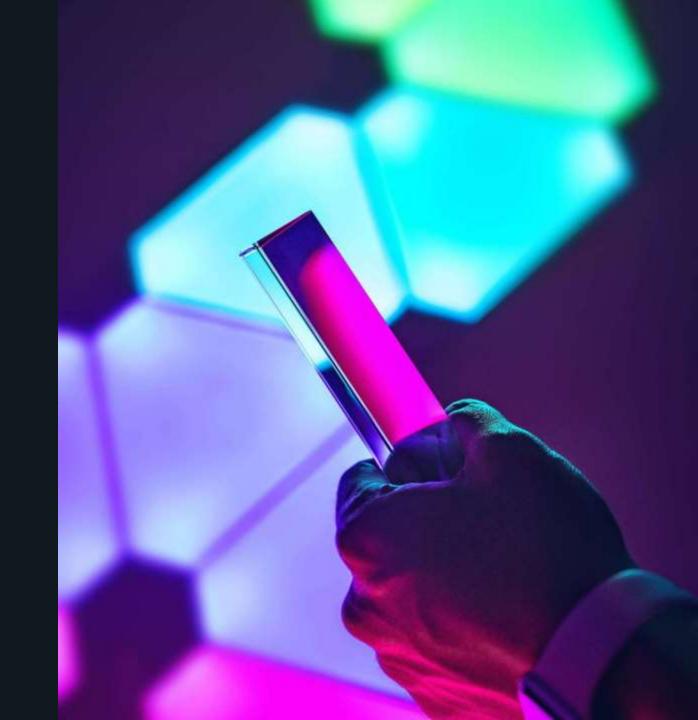
→ Competición online → Demuestra qué eres capaz de hacer con datos → Inscripción hasta el 2 de dicembre

Objetivos Datathon

- → Atracción y reconocimiento del talento en tecnologías Al4ES
- → Estimular acciones de formación, capacitación, e I+D, para el avance de las tecnologías AI4ES a nivel nacional
- → Estimular la economía basada en el dato a través del lanzamiento de retos en tecnologías AI4ES

Buscamos a los mejores estudiantes y profesionales en el uso de tecnologías IA para procesamiento de datos







Características

- → 3 categorías de participación, según grado de experiencia
 - → Futuras promesas (menos de 3 años experiencia)
 - → Especialistas (+3 años experiencia)
 - → Investigadores AI4ES

Retos 2022



Premios

- → 3 premios para cada reto, según categoría participación
 - → Futuras promesas: 1000€
 - → Especialistas: 2000€
 - → Miembros AI4ES: 1000€ en especie

Participantes

	Reto A	Reto B	TOTAL
Nº inscritos	36	29	65
Futuras Promesas	21	15	36
Especialistas	12	9	21
Miembros AI4ES	3	5	8









Reto A (Industrial) – Procesamiento de imágenes Detección y cuantificación de enfermedades basadas en imágenes de drones para la protección de cultivos

→Disease detection and quantification based on drone images for crop protection

- →o1 The Challenge
- →02 The Dataset
- →o3 The Code
- →04 Mechanics
- →o₅ Evaluation





The Challenge

.

. . . .





DISEASE DETECTION AND QUANTIFICATION

The Challenge

BASED ON DRONE IMAGING

Context

FOR CROP PROTECTION

BASF's Agricultural Solutions division conducts **hundreds of field trials** each year **to evaluate** the biological **performance** of new **crop protection products**.

Currently, most of the data is generated through visual assessments by experienced field research agronomists.

<u>Use case</u>

Trials for testing wheat protection products: the field agronomist assesses the **percentage of crops infected by each disease** in each plot.

Aim of the Challenge

To train a robust **image regression model**, which **processes images** of individual field plots and **predicts their assessment value** (disease and percentage of infection).



02

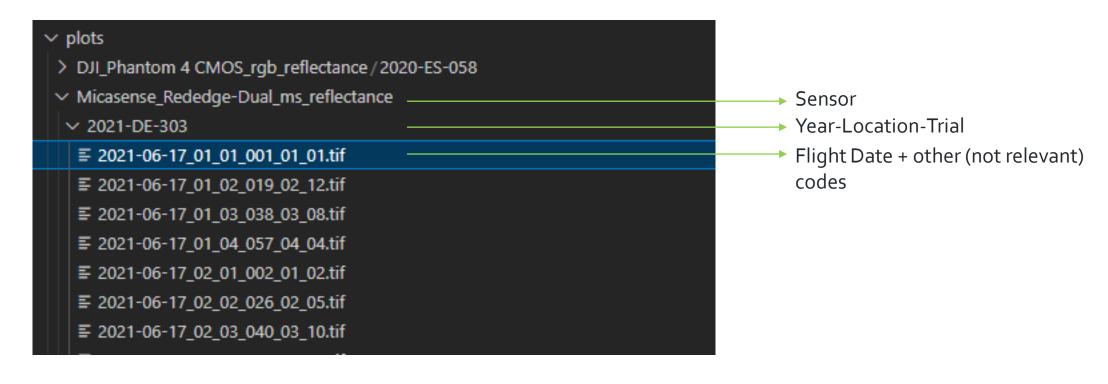
The Dataset

.

. . . .



Multi-spectral images (tif files) captured by **drones** in different years and field trials in wheat crops at various locations.





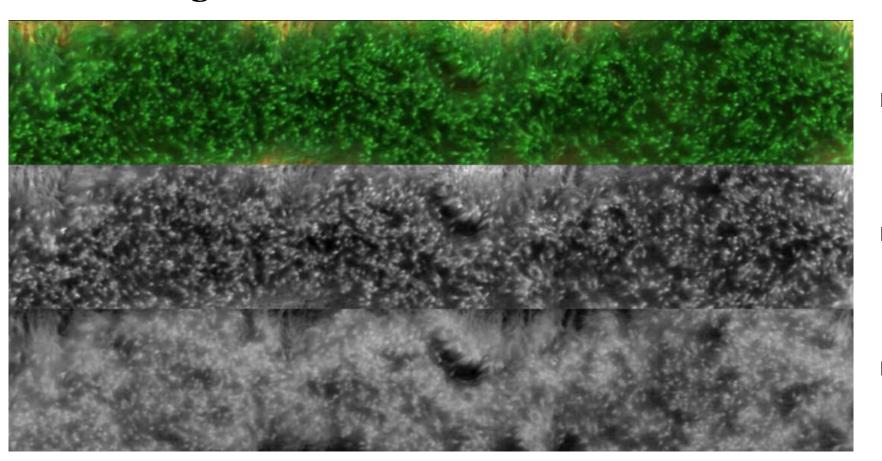
5-band images contain sensor data according to this table:					
Band	Color	Max Intensity (nm)	Band Width (nm)	Sensor	
1	blue	475	32	MX	
2	green	560	27	MX	
3	red	668	14	MX	
4	red edge	717	12	MX	
5	NIR	842	57	MX	



10-band images contain sensor data according to this table:					
Band	Color	Max Intensity (nm)	Band Width (nm)	Sensor	
1	blue	475	32	МХ	
2	coastal blue	444	28	MX Blue (Duo)	
3	green	560	27	МХ	
4	green	531	14	MX Blue (Duo)	
5	red	668	14	МХ	
6	red	650	16	MX Blue (Duo)	
7	red edge	717	12	МХ	
8	red edge	705	10	MX Blue (Duo)	
9	red edge	740	18	MX Blue (Duo)	
10	NIR	842	57	МХ	

Use only these common five bands





RGB

Red edge

NIR



The Dataset: train.csv

GROUND TRUTH

PLOT_FILE	DISEASE1	DISEASE2	DISEASE3
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-320/2021-06-15_04_04_056_04_09.tif	0	0	1
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-018/2021-04-20_13_01_013_13_01.tif	0	0	6
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-999/2021-07-02_06_03_036_03_04.tif	0	0	0
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-321/2021-06-21_01_03_040_03_08.tif	0	0	65
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-429/2021-05-26_02_04_052_04_13.tif	0	0	1
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-030/2021-04-08_18_04_075_08_05.tif	0	0	2
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-P02/2021-02-11_02_02_019_02_05.tif	0	0	98
plots/Micasense_Rededge-M_ms_reflectance/2020-ES-DG4/2020-04-13_08_01_008_01_08.tif	0	0	0
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-322/2021-06-07_04_03_048_03_16.tif	0	0	2
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-428/2021-04-26_01_01_001_01_tif	0	0	10
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-A01/2021-01-11_06_03_010_04_02.tif	0	0	15
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-030/2021-04-28_09_01_009_01_09.tif	0	0	2
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-429/2021-06-07_05_01_005_01_05.tif	0	0	4
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-321/2021-06-21_11_04_060_04_05.tif	0	0	28
plots/Micasense_Rededge-M_ms_reflectance/2020-ES-DG1/2020-01-21_11_02_020_02_08.tif	0	0	92
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-320/2021-06-07_16_02_021_02_12.tif	0	0	2
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-338/2021-08-04_06_02_030_02_01.tif	2	0	0



The Dataset: test.csv

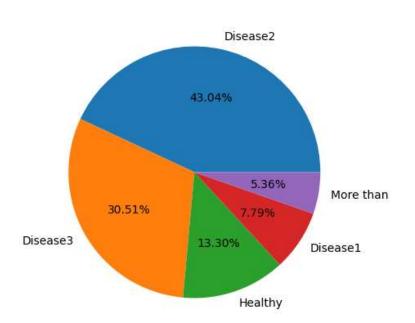
YOUR PREDICTIONS HERE:

PLOT_FILE	DISEASE1	DISEASE2	DISEASE3
plots/Micasense_Rededge-M_ms_reflectance/2020-ES-057/2020-04-13_04_01_004_04_01.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-P08/2021-03-12_11_02_020_02_08.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-P01/2020-12-01_02_03_042_05_02.tif			
plots/Micasense_Rededge-M_ms_reflectance/2020-ES-036/2020-02-25_02_02_031_02_14.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-524/2021-04-26_02_01_002_01_02.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-A02/2021-01-04_05_03_018_03_02.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-P08/2021-01-28_03_01_003_01_03.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-A05/2021-03-03_08_03_009_04_01.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-P08/2021-03-12_09_01_009_01_09.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-P07/2021-03-22_02_03_036_03_12.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-429/2021-06-15_11_01_011_11.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-018/2021-03-23_03_01_003_03_01.tif			
plots/Micasense_Rededge-M_ms_reflectance/2020-ES-001/2020-01-21_04_01_004_01_04.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-322/2021-05-19_03_03_041_03_09.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-303/2021-07-08_05_03_032_03_02.tif			
plots/Micasense_Rededge-M_ms_reflectance/2020-ES-035/2020-03-04_06_03_035_03_09.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-323/2021-06-07_02_01_002_01_02.tif			

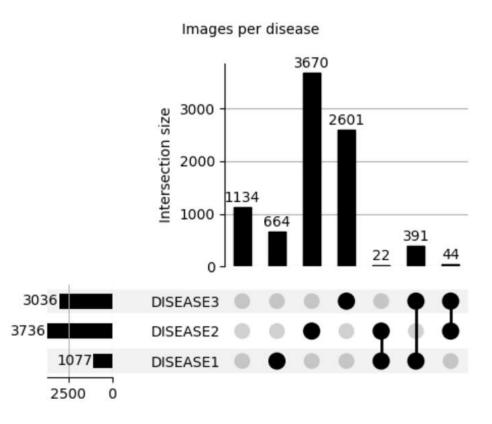


The Dataset: statistics [train: 8526 images / test: 1766 images]

train_healthy_infected.png



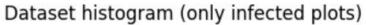
train_diseases_upsetplot.png

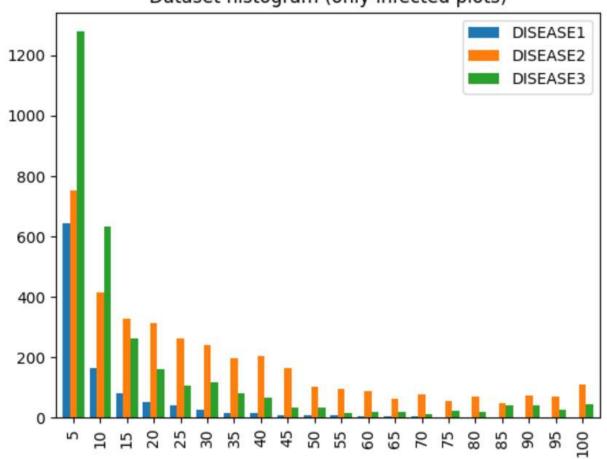




The Dataset: statistics

train_histogram.png







The Dataset: statistics

train_classes_stats.csv

Class	▼	lmages ▼
Healthy		1134
Infected only with DISEASE1:	1 to 5	374
Infected only with DISEASE1:	6 to 10	103
Infected only with DISEASE1:	11 to 15	55
Infected only with DISEASE1:	16 to 20	31
Infected only with DISEASE1:	21 to 25	31
Infected only with DISEASE1:	26 to 30	18
Infected only with DISEASE1:	31 to 35	15
Infected only with DISEASE1:	36 to 40	9
Infected only with DISEASE1:	41 to 45	5
Infected only with DISEASE1:	46 to 50	4
Infected only with DISEASE1:	51 to 55	4
Infected only with DISEASE1:	56 to 60	3
Infected only with DISEASE1:	61 to 65	4
Infected only with DISEASE1:	66 to 70	4
Infected only with DISEASE1:	71 to 75	1
Infected only with DISEASE1:	76 to 80	1
Infected only with DISEASE1:	81 to 85	2
Infected only with DISEASE1:	86 to 90	0
Infected only with DISEASE1:	91 to 95	0
Infected only with DISEASE1:	96 to 100	0
Infected only with DISEASE2:	1 to 5	729
Infected only with DISEASE2:	6 to 10	409
Infected only with DISEASES:	11 to 15	210

train_damage_stats.csv

Damage	₽ ▼	DISEASE1 ▼	DISEASE2 ▼	DISEASE3 ▼
	0	7449	4790	5490
11	to 5	645	754	1279
6 to	o 10	165	416	632
11 to	15	81	326	263
16 to	20	53	315	162
21 to	25	39	263	106
26 to	30	25	241	118
31 to	35	15	196	79
36 to	40	16	206	67
41 to	45	7	163	34
46 to	50	8	102	34
51 to	55	7	97	15
56 to	60	3	87	20
61 to	65	4	61	18
66 to	70	4	77	11
71 to	75	1	56	24
76 to	08 c	2	71	19
81 to	85	2	49	41
86 to	90	0	75	42
91 to	95	0	71	26
96 to	100	0	110	46



The Dataset: some considerations

Images follow a calibration and stitching process that is performed in daily basis:

- Calibration is done each day by using a calibration plate.
- Plot images are stitched and merged together.

On our previous work, we have detected some domain shift and performance reduction due to:

- Different flight dates
- Weather conditions
- Calibration issues

We recomend to analyse and preprocess? the files contents to try to minimize this domain shift:

- Apply normalization (example provided later)
- Remove negative values (example provided later)
- Use trial key numbers and/or flight dates to isolate domain shift caused by calibration



03

The Code

. . . .

. . . .

.

. .

. . . .



The Code

A pyhton script (inference.py) is provided as a help to fill the test.csv file with your predictions:

PLOT FILE	DISEASE1	DISEASE2	DISEASE3
-	DISEASET	DISEASEZ	DISEASES
plots/Micasense_Rededge-M_ms_reflectance/2020-ES-057/2020-04-13_04_01_004_04_01.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-P08/2021-03-12_11_02_020_02_08.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-P01/2020-12-01_02_03_042_05_02.tif			
plots/Micasense_Rededge-M_ms_reflectance/2020-ES-036/2020-02-25_02_02_031_02_14.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-524/2021-04-26_02_01_002_01_02.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-A02/2021-01-04_05_03_018_03_02.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-P08/2021-01-28_03_01_003_01_03.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-A05/2021-03-03_08_03_009_04_01.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-P08/2021-03-12_09_01_009_01_09.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-P07/2021-03-22_02_03_036_03_12.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-DE-429/2021-06-15_11_01_011_01_11.tif			
plots/Micasense_Rededge-Dual_ms_reflectance/2021-ES-018/2021-03-23_03_01_003_03_01.tif			
0000 50 001 (0000 01 01 01 01 01 01 04 01 04 tif			
3_03_041_03_09.tif			
5_03_032_03_02.tif			

```
def predict(image):
   import random
   return [random.randrange(0, 101), random.randrange(0, 101), random.randrange(0, 101)]
if name ==" main ":
   test df = pd.read csv("test.csv")
    for image path in test df["PLOT FILE"]:
       image = load_image(image_path=image_path, remove_negs=True, normalization=True)
       prediction = predict(image=image)
       test df.loc[test df.PLOT FILE == image path, ['DISEASE1', 'DISEASE2', 'DISEASE3']] =
               prediction
   participant = "John Smith"
   version = "v1"
   results filename = '-'.join([participant, 'test results', version])
   test df.to csv(results filename, index=False)
```



The Code

Help also to open and manage the tif files:

```
def load image(image path, remove negs=True, normalization=False):
    # Load the image as a numpy array
    img = rasterio.open(image path)
    concat list = []
    if img.count == 5: # 1=B, 2=G, 3=R, 4=RE, 5=NIR
        channels list = [1,2,3,4,5]
    elif img.count == 10: # 1=B, 3=G, 5=R, 7=RE, 10=NIR
        channels list = [1,3,5,7,10]
    else:
        raise Exception("Unexpected number of channels in image %s" %image path)
    for i in channels list:
        ch image = img.read(i)
       if remove negs:
           ch image[ch image < 0] = 0.0
        concat list.append(np.expand dims(ch image, -1))
    image = np.concatenate(concat list, axis=-1).astype(float) # float 0-1
    if normalization:
       max pixel value = np.max(image, axis=2)
        max_pixel_value = np.repeat(max_pixel_value[:, :, np.newaxis], 5, axis=2)
        image = np.divide(image, max pixel value)
        image = image[:, :, 0:4]
        image[np.isnan(image)] = 0.0
    return image
```



04

Mechanics

. . . .

. . . .

.

. .

. . . .

∧I4∃S



Mechanics

- Sign agreement to use the data provided by BASF
- Download zip file with data, labels, statistics and code from TEAMS
- Communication (with the organizers and between participants): TEAMS
- Every Friday the participants can send results (v1, v2, etc.) and we will provide evaluation
- January 17th: send final test results and PDF describing the work
- January 18th to 20th: Presentation of the models by the participants
- January 24th: winners will be announced



05

Evaluation

. . . .

. . . .

.

. .

. . . .



Evaluation

Qualitative aspects:

- Novelty and feasibility of the algorithm proposal
- Presentation quality

Quantitative aspects (metrics):

- R2-score
- MAE
- RMSE



Reto B (Social) – Analítica de datos Incidencia y evolución enfermedades víricas

SIVIC

Con los datos abiertos de la evolución de la pandemia COVID-19 en Cataluña, obtenidos a través de la web del SIVIC, el presente reto tiene como objetivo avanzar en la mejora de la creación de modelos predictivos capaces de ajustar/predecir el futuro de la progresión del COVID-19.

Modelos predictivos

- →El objetivo de este reto es por tanto crear modelos capaces de proporcionar 4 tipos de pronósticos para los datos Catalunya:
- →• Incidencia, Ingresos hospitalarios,
 - Ingresos en UCI,
 Defunciones.





Canales comunicación

- → Lista correo para difusión general: inscritos-datathon-2022@ai4es.es
- → Contacto con equipo organizador: <u>datathon@ai4es.es</u>
- →Espacio de trabajo: cada reto contará con un grupo Microsoft Teams
 - → Canal comunicación específico para cada reto: difusión, preguntas, comunicación entre participantes, ...
 - → Datasets accesibles desde repositorio del grupo

Presentación del proyecto

- → Informe resumen del proyecto realizado (5 páginas)
 - → Enfoque adoptado
 - → Características del desarrollo
 - → Principales resultados y conclusiones
- → Presentación oral ante Jurado
 - → 10 mins., slot pre-asignado
 - → En remoto
- → Se facilitarán plantillas para el informe y presentación



Evaluación

- → Criterios evaluación
 - → Impacto e importancia del reto solucionado
 - → Creatividad e innovación en el enfoque seguido
 - → Uso de los datos (grado de resolución del reto)
 - → Grado de acabado conseguido (demo)
 - → Claridad de la presentación oral
- → Jurado Datathon
 - → 5 miembros: un especialista de cada centro de la red + 1 miembro Comité Asesor
 - → Se constituirá un Jurado por cada reto

Fechas importantes

- →02-Diciembre-2022: Inicio Datathon
- →17-Enero-2023 23.59: Fin Datathon (límite envío proyecto)
- → 18 al 20 Enero: Presentacion de soluciones ante el Jurado
- →24 Enero 2023: Fallo del Jurado y comunicación resultados



Preguntas

- → Turno abierto para preguntas
- → Las preguntas y respuestas frecuentes se recopilarán y facilitarán en documento FAQ

→ Preguntas offline: <u>datathon@ai4es.es</u>



Gracias

https://ai4es.com →

MIEMBROS:















