

Schneider Electric European Hackathon

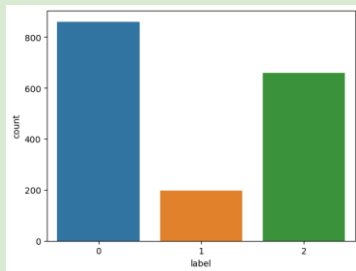
19 de noviembre de 2022



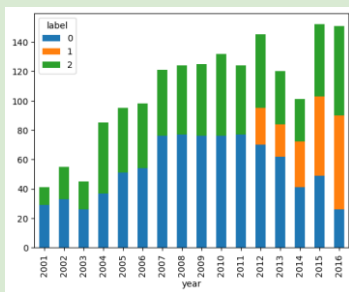
Hackaton Scheider Electric - Nuwe

- Se plantea el problema de clasificar imágenes en tres **tipos distintos de categorías según el tipo de vegetación: *Plantation, Grassland/Shrubland y Smallholder Agriculture***
- Adicionalmente a las imágenes existes los **datos de las coordenadas y año de la foto**
- Inicialmente se han planeado hacer los siguientes desarrollos:
 - **Análisis exploratorio de datos** (✓ completado)
 - **Modelo Deep learning utilizando únicamente imágenes** (✓ completado)
 - **Utilizar coordenadas en el modelo** (✗ NO completado por falta de tiempo)
 - **Utilizas datos de año** (✗ NO completado por falta de tiempo)

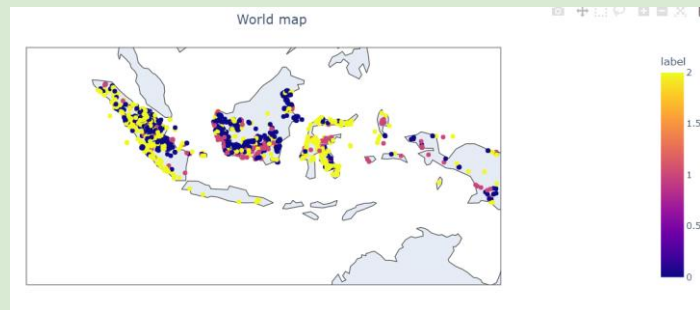
Exploración de datos



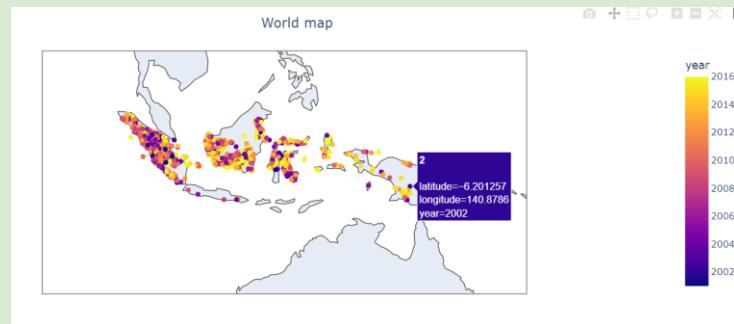
En los datos hay muchas más imágenes del tipo *Plantation* que de las otras clases. **Disponemos de una muestra NO balanceada**



Si nos fijamos de la muestra de *train*, el **tipo de zona Smallholder Agriculture** está creciendo desde el 2013



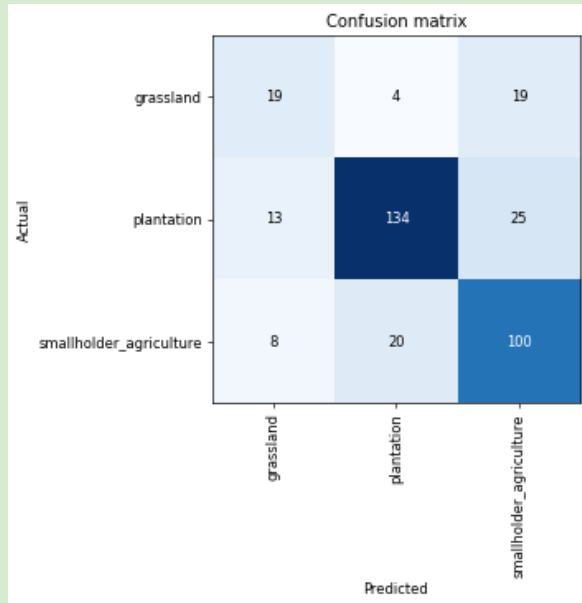
Por la ubicación de las fotos vemos que se trata de la zona de Indonesia y que **sí parece haber un patrón de zonas con más presencia de un tipo u otro de *label***



Las fotos parecen tomadas de forma aleatoria entre las distintas zonas a lo largo del periodo analizado

Resultados

El modelo construido en base solamente a las fotos nos da **un F1-score de 0.67 sobre el test de validación**



	precision	recall	f1-score	support
grassland	0.47	0.45	0.46	42
plantation	0.85	0.78	0.81	172
smallholder_agriculture	0.69	0.78	0.74	128
accuracy			0.74	342
macro avg	0.67	0.67	0.67	342
weighted avg	0.74	0.74	0.74	342

Posibles mejoras / próximos pasos

- Entrenar el modelo FastAI con más epochs
- Balancear la muestra para mejorar los resultados
- Como la muestra está muy localizada en Indonesia, si se quiere **un modelo sólo para Indonesia se pueden hacer los siguientes desarrollos**. Estos desarrollos no se deberían hacer en caso de necesitar un modelo para otras áreas:
 - **Utilizar la variable *year* en el modelo** (la label de tipo 1 empieza a aparecer a partir de 2013)
 - **Utilizar las coordenadas para identificar zonas con más presencia de un tipo de *label***