



# Zero deforestation mission

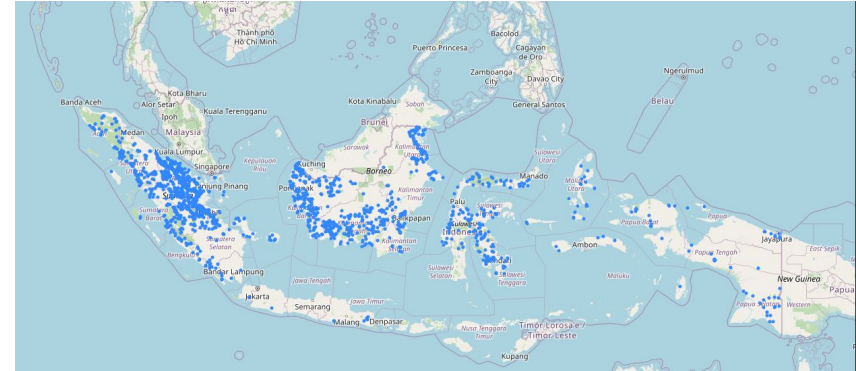
**Data Science  
Pyteam**

Hector Navarro - PhD Environmental Engineer (BSC-CNS)  
Calum Meikle - Data Analyst (BSC-CNS)

BSC-CNS - Barcelona  
Supercomputing Center

# Introduction

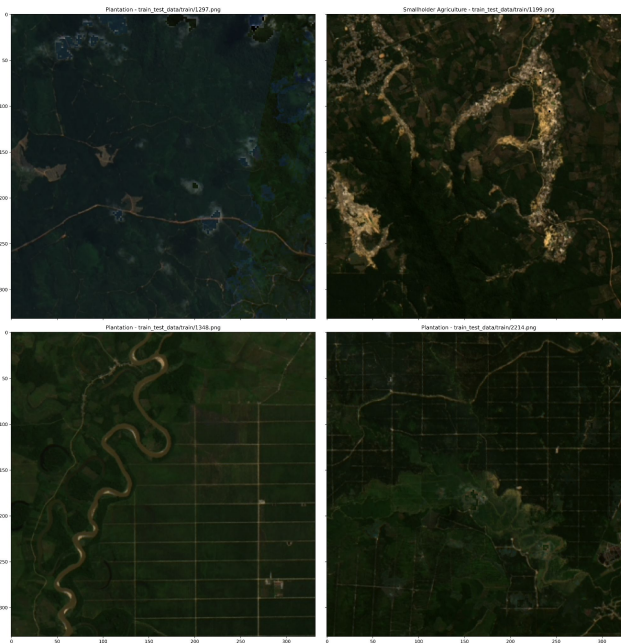
- Data science has gained popularity in the last decades due to the fact that it can be applied in multiples topics, e.g. medical sciences, earth sciences, business, etc.
- Despite this there are still a multitude of fields and topics that could benefit from the use of machine learning.
- One of these is the challenge that we were set today - image classification of different types of land use in South East Asia. Land use that has come about after mass deforestation, which has a devastating impact on local and indigenous peoples but also the world as a whole.



**Fig 1:**  
Places  
where data  
images  
were taken

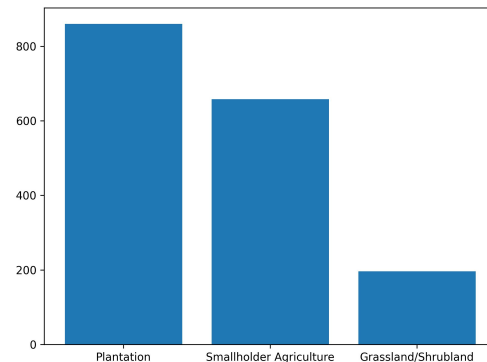
# Dataset exploration

- As described in every medium article you must visualise the data first.
- So we investigated the different types of satellite images and their corresponding labels. (Fig 2)
- We were able to see the daunting task ahead of us as some of the images seemed very similar
- We also investigated the size of the dataset and the value counts for the different classifications. (Fig 3)
- Future research would have been into image augmentation, to be able to run the model over a much larger dataset, but augmenting the data that we already had.



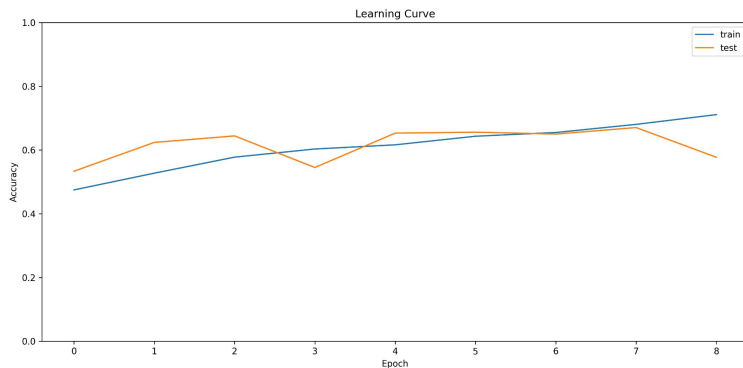
**Fig 2:**  
Satellite  
images with  
different  
types of  
classification

**Fig 3:** Number  
of images per  
classification.



# Model Results

- The model results were a mixed bag for us. Our F-score and accuracy plots (Fig 4) were encouraging, considering the size of the dataset and the similarity between the satellite images we were pleased with this.
- However we were getting some disparaging results when predicting the labels in the test data. Often seeing all data being classified as Plantation or Smallholder Agriculture and very little Grassland/Shrubland.
- This may have been a sign that we were overfitting the data. Another point of future investigation along with the augmentation of the images



**Fig 4:** Accuracy plot for our model.