Hackaton Data Science

Zero deforestation mission

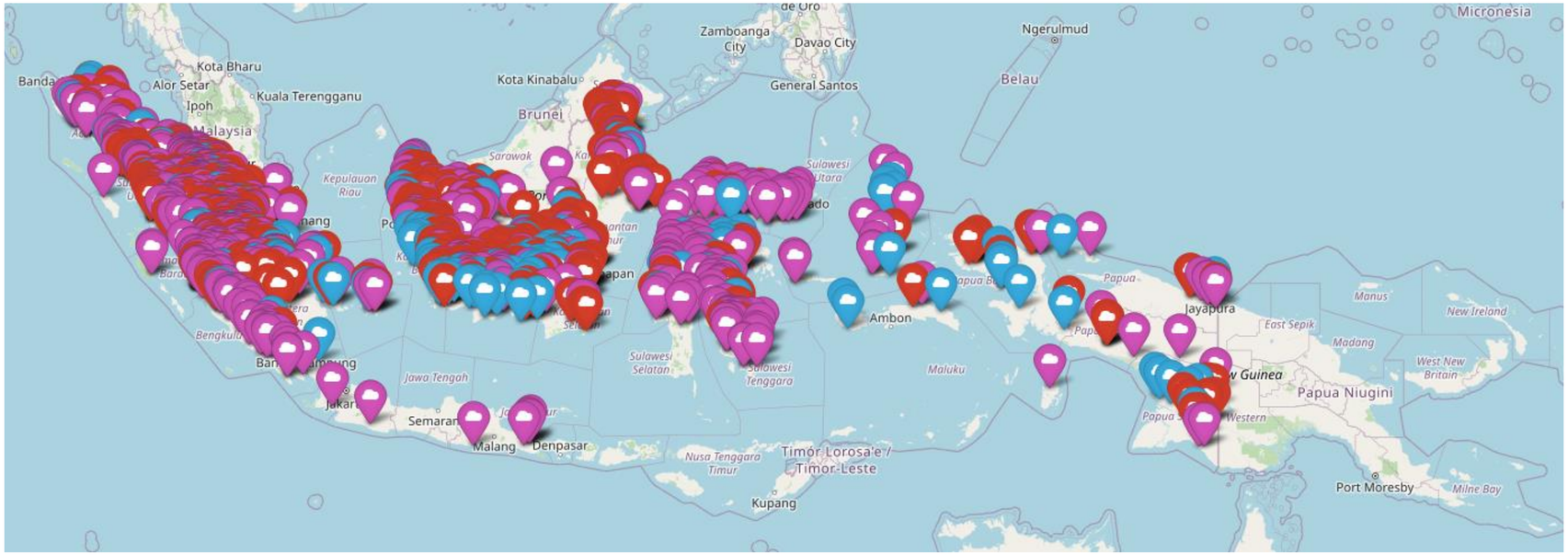
Manolo Boost DT52: Álvaro Beltrán, Sergio Cabezas, Pedro Gallego

ALL

Geospatial data’s distribution

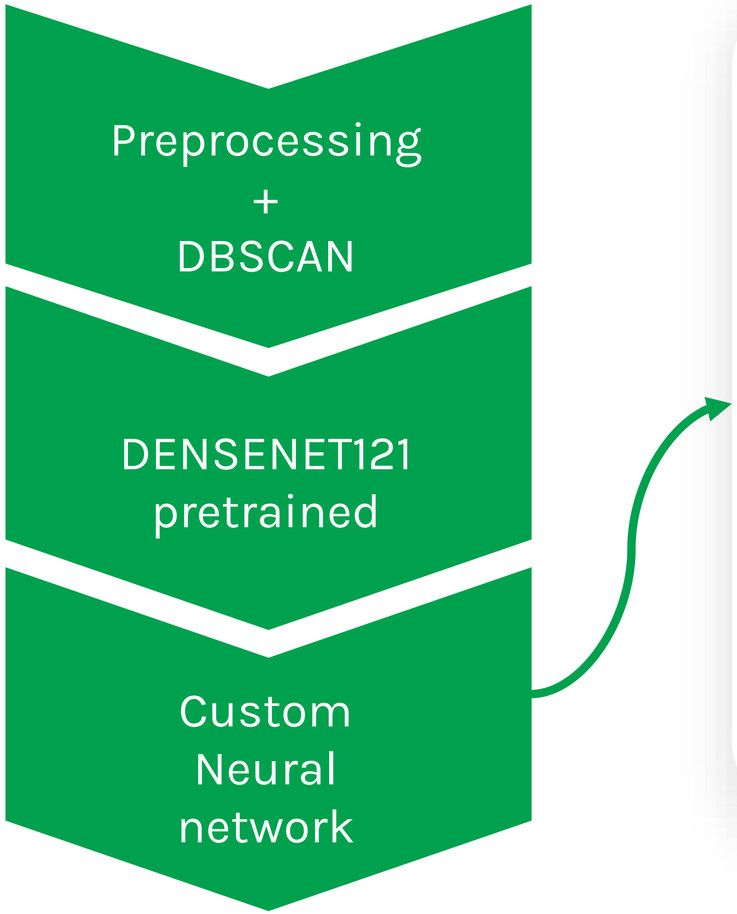
In this representation we can see that geospatial distribution is importat cause of that, we are goint to develop a new feature clustering this points.

Smallholder Agriculture Grassland/Shrubland Plantation



ALL

Model



We extracted the images features using a DenseNet121 pretrained.

Then, we created the following model to predict the target for each case, using as input the data from the csv (lat and long) and the image features extracted by DenseNet121.

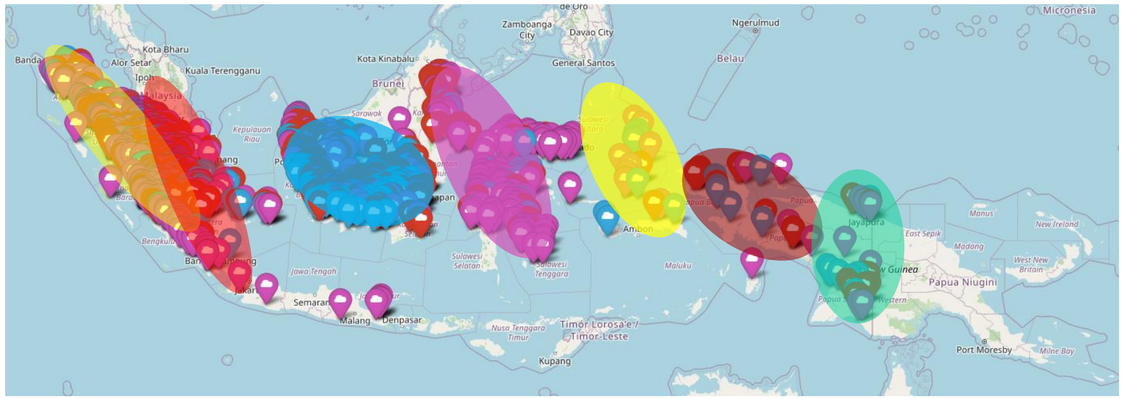
Our preprocessing is based in normalize tabular data, create clusters with DBSCAN and process data in the right way to feed out model.

```
Model: "model"
__________________________________________________________________________________________________
Layer (type) Output Shape Param # Connected to
==================================================================================================
images (InputLayer) [(None, 1024)] 0 []

tabular (InputLayer) [(None, 14)] 0 []

dense_1 (Dense) (None, 512) 524800 ['images[0][0]’]

dense (Dense) (None, 6) 90 ['tabular[0][0]’]

concatenate (Concatenate) (None, 518) 0 ['dense_1[0][0]', 'dense[0][0]’]

priority (Dense) (None, 3) 1557 ['concatenate[0][0]']
==================================================================================================
Total params: 526,447 Trainable params: 526,447 Non-trainable params: 0
__________________________________________________________________________________________________
```

ALL

Clustering: DBSCAN

Example zone clustering



Considering that the prediction is improved by using additional information to the images, we have added to our model during the preprocessing the use of the DBSCAN clustering algorithm. Based on the latitude and longitude data of the images we are going to add the results of the DBSCAN clustering as tabular data since the deforestation situation is usually common in close proximity.

We use DBSCAN as the clustering algorithm because, since we are dealing with images of an archipelago of islands, this will allow us to do the clustering respecting the limits of each island, offering better results than other clustering algorithms such as KMeans.

ALL

Our Results

Experiments with 33% of our data for testing:

0.75 F1 Score

We trained the model with the whole dataset of training for getting our best model. This are our graphics of the traning. Later, we infer the target in the test dataset.