DIEDRO'S SQUAD

ZERO DEFORESTATION MISSION

Ander Aquerreta
Data Science

http://www.linkedin.com/in/ander-aquerreta-montoro

Unai Gurbindo
Data Science

lwww.linkedin.com/in/unai-gurbindo-gonzalez

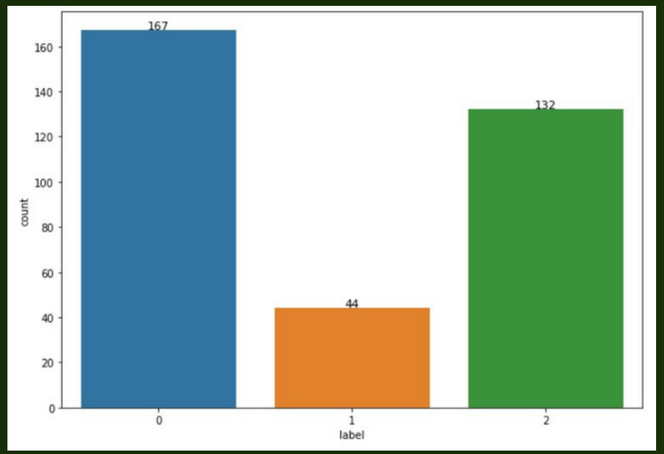
Mikel Aldea
Data Science

https://www.linkedin.com/in/mikel-aldea-42b46b234/

FIRST STEPS

We start preparing the data and loading the libraries. We observe as a first approximation that the data are not balanced so for the image models something that we will take into account when making the models.

Then, we will create some models in which we will take the images and we will try to classify it with the information provided by the pixels. In addition, we take into account the longitude and latitude of the images in another SVM model, since we consider that the distribution of the data is relevant when it comes to see what kind of deforestation we are facing.



OUR MODEL

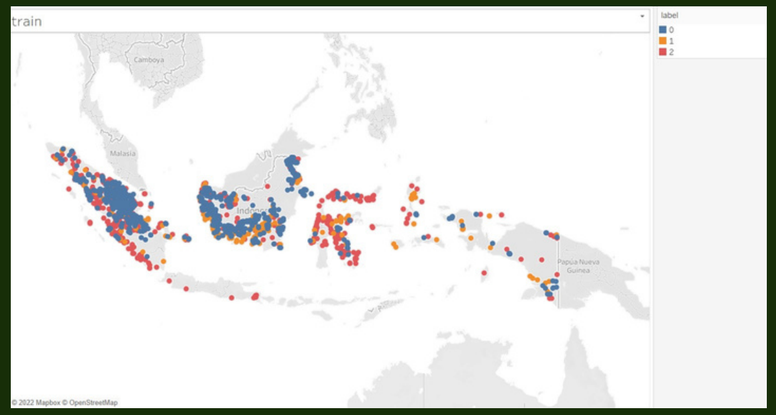
Therefore, we create 4 models through "fastai" with which we take the images as data and create in turn an SVM model with latitude and longitude. We train the models with the train data and classify with these all test and train images.

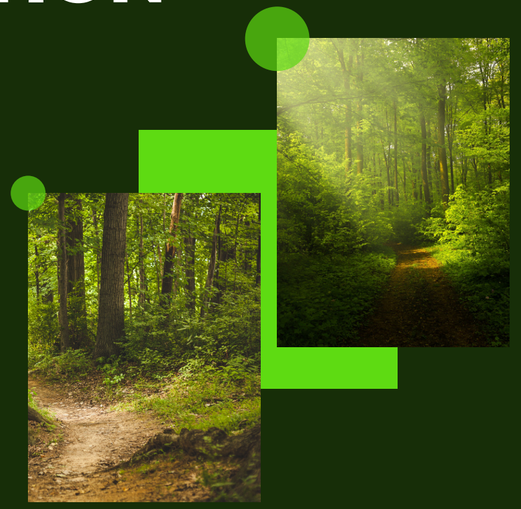
Then, we take the outputs of the models and create a DataFrame for train and another for test, each with the respective outputs of each model for each dataset. We take the DataFrame with the outputs corresponding to train and using the KNN algorithm we train an "ensemble" of the outputs, in this way we manage to make an ensemble of the models obtained with which we can enrich our predictions in test.

Finally, from the ensemble created, we take the DataFrame of the outputs of the 5 models for the test data and calculate the final predictions thanks to this ensemble.

DIEDRO'S SQUAD

ILLUSTRATION OF MODEL CREATION





Taking the special characteristics of the train data, we create the SVM model.

Taking the images from the train dataset we create the following image classification models: Resnet34, Resnet50, Densenet121 and Vgg_16_bn.

K N N

We make the assembly of the outputs of the moledos through the KNN algorithm, with which we manage to correct errors in the outputs of a model with the rest.

OUR RESULTS

DIEDRO'S SQUAD

SHORT COMMENTS

Score of the models

For the creation of all the models we based ourselves on the Macro F1 Score since the objective was to achieve the best result in the competition.

Main execution

To execute the main it is only necessary to download the models that take the images as data so that the main can load them (model_resnet34.plk,...).

Once we have these models we can execute all the main in a simple way.

Fastai image models

The image models created through the fastai library can be seen as they have been built in the notebooks included in the GitHub (model_resnet34.ipynb, ...).

Model quality

We consider that we achieved a high quality model, since we were able to correct the errors thanks to the ensemble, each of the models provides an acceptable f1 score separately and we did not overtrain the models when building them.