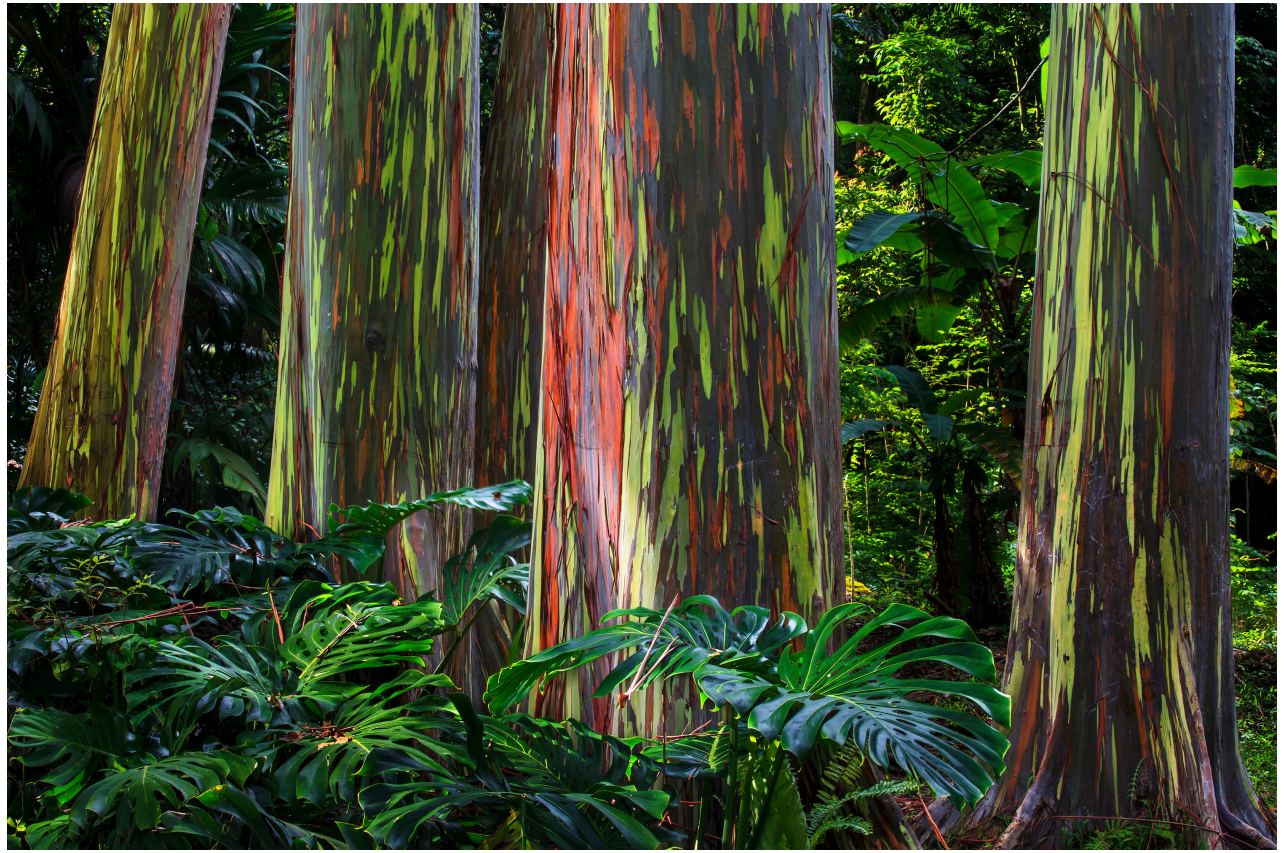


# Schneiden Electric European Hackathon

## Data Science Zero deforestation mission 19 Nov 2022

**AHMED BEGGA**
**TEAM: HAL9000**

# DESCRIPTION

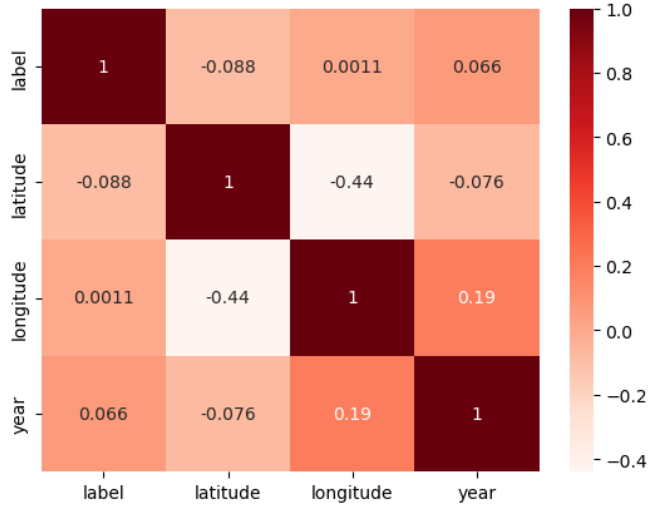
Today's challenge consisted of classifying the data provided concerning satellite images and coordinates. These images are designed to help predict which type of deforestation is being faced. Thus, in the first data analysis, we tried to find a correlation between the coordinates given and the tags. After observing such a correlation did not exist, we started classifying the images, trying to attain the maximum score possible. To achieve our goal, we focused on data augmentation and transfer learning. The latter strategies did not provide a solution to the issue because classes are very similar to each other.

Finally, we opted for the development of a much more simple model, that had the ability to differentiate the former classes.

# ANALYSIS

For this part of the assignment, we have especially focused on data analysing and label distribution. For this sole purpose, we have  represented the given data with different functions . For instance, we have displayed different histograms showing the distribution label of the training data and validation, the correlation between the features and the *TSNE* of all images to see if there was any pattern.

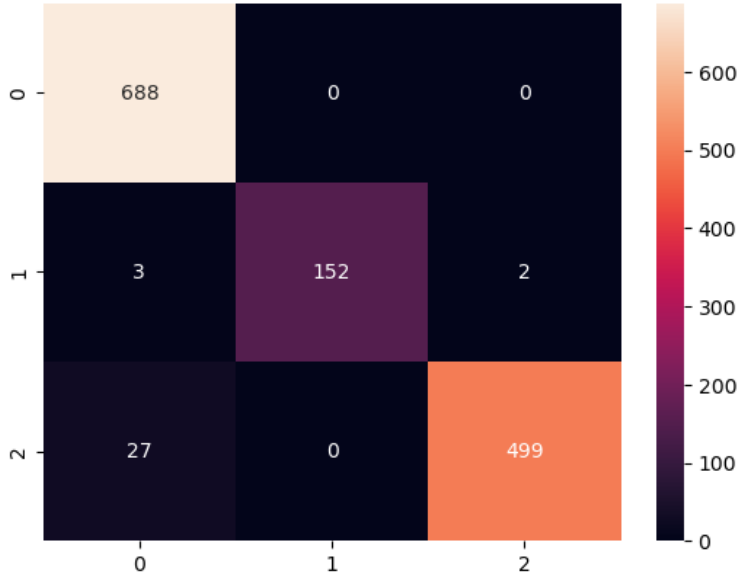
Moreover, we noticed there was a class that had only a few examples, and this difficulted us the task of maximising the *f1_score*. This is the main reason why we employed the data stratification, representing the real distribution of the data during the training.

# TRAINING

For the training of the models we have used different techniques, such as data augmentation or transfer learning. Both options failed due to the fact that the models were too complex for the task they were responsible from. For example, we used *Resnet50* with trained and untrained weights, but we could not achieve to maximise the *f1_score*.

Finally, we opted for using a model created by ourselves with only a few layers. With such action we attained to improve the results.



# RESULTS

During the competition, we have tried to use the most famous techniques in *Deep Learning* to get the best results, either using transfer learning, data augmentation, data stratification, etc.

In the end, we were unable to achieve very competitive results, but at least, we learned a lot about machine learning, most specifically about *Overfitting*.

The final result we obtained on 20% of validation data is 0.40 on *f1_score.*