Deforestation Schneider Electric Hackathon

Omar Aburayyan

Architecture and project setup

- An AlexNet CNN was trained on the dataset from scratch using Pytorch
  - AlexNet is one of the most significant cnn architectures and it’s good to use it as benchmark model
  - I used Pytorch because of its very flexible when it comes to debugging and tweaking models to your liking

- I sat up managed the pipeline using DVC (Data Version Control)
  - dvc is a great tool for managing machine learning pipelines;
  - It allows us to split the pipeline into different stages, easily track, run and do experiments

What would I have added if I had more time?

- The data set is unbalanced, examining the test predictions. It seems that my model is not predicting grassland (class # 1) at all
  - I should have split the training set into a training and validating set to validate the training performance
  - Image augmentation should have been performed on the minority class to help the model generalise on these instances
  - Grassland data is only available after 2012
    - There might be some deterministic behaviour that could be predicted using other attributes than the image (latitude, longitude, year)
    - One idea is to extend the problem solution to a multimodal classification model where the image model is combined together with another model based on other features than the image
- Test other architectures and compare their performance with AlexNet

Thank You for the great hackathon