

Project Title: - Identifying the forest through the trees

Group members: Wei Wang, Elle Proust, Sai Ravuru, Nick McCormack

Objectives:

Project Objective:

Use Machine Learning algorithms to accurately label satellite image grid(s) of Amazon forests based on the various classes of land cover and land use

Learning Objective:

1. Apply machine learning techniques with little guidance
2. Develop iterative processes to improve the model
3. Learning Image pre-processing techniques

Data

Example Train Images:

https://drive.google.com/drive/u/1/folders/1m6GXU9ujtiZd14s_NNeomjkXhea7CDv2

(Everyone should be able to view and change this file)

Train Data Labels:

<https://drive.google.com/open?id=1TdhLpojpcR2n0QRQmhEddnjC-DVt2DCq>

Full Data Set: <https://www.kaggle.com/c/planet-understanding-the-amazon-from-space>

Things to do (post June check-in)

1. Complete EDA
 - a. Correlation between labels
 - b. PCA decomposition
 - c. View example images
 - d. Compare TIF/JPG
 - e. Label simplification
2. Imaging Preprocessing:
 - a. Haze Removal
 - b. Rotation
 - c. Edge Identification?, road/water
3. Data Preprocessing
 - a. Min-max scaler
 - b. PCA for faster running
4. Machine Learning
 - a. Can we turn this into simple classification problem?
 - i. Try a two stage approach, classify weather, and then do 2 against the rest

- ii. Single Label approaches
 - 1. Logistic Regression
 - 2. Decision Trees
 - 3. Random Forest
 - b. Multilabel Approaches
 - i. OneVsRestClassifier
 - 1. Decision Trees
 - 2. Random forests
 - ii. Classifier Chains
 - 1. Decision Trees
 - 2. Random forests
- 5. Adapted Algorithms
 - 3. Multi-label nearest Neighbors

June's first checklist

1. Pull up one example of each possible label (Weather, forest, water, other)
2. Cross-validation
3. Pass at all algorithms; focus on one that does the best - we focused on forest models
4. Design precision-recall metric to measure our goal - we ended up focusing on F1, and classification of the rarer tag

June's second checklist

1. Show possible errors with haze identification
2. Run a combined weather/land model