Aditya

Data Science

Shrujal Ambati
Aditya Dixit
Ioana-Andreea Cristescu
Farzad Siraj
Islina Shan

Project objective

Islina

Working with Axians' Location Intelligence Team, we developed algorithms using **satellite images** to determine **whether vegetation growth will interfere with power line construction**. This will involve using machine learning to identify and monitor power lines, geographical features, obstructions and other changes in the landscape. This project will involve a lot of **Random Forest** based ensemble methods and segmentation to do **pixel-based classification**.

Geographical Location

Ioana

- The Portuguese Archipelago of Azores
- São Miguel Island
- The surface of the island is 760 km^2
- Area of study - denser vegetation that can cause power outages

Steps

Farzad

1. Dataset - Sentinel-2 SR- 2A
2. Cloud Masking
3. Time Series Analysis
4. Random Forest Image Classification

Cloud Cover

Aditya

- Clouds covered a significant portion of our area of interest
- Need to remove before land analysis
- Many problems with existing cloud masking algorithms
- Solution: sentinel-2 cloudless algorithm

Vegetation Management

Aditya

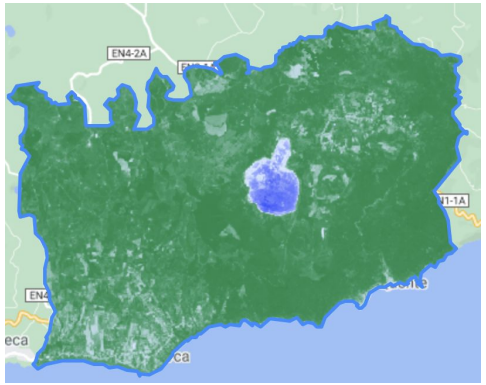
- The detection of vegetation cover changes through the creation of vegetation indices has become an important application of remotely sensed imagery.
- Remotely sensed information from Landsat, SPOT, MODIS and NOAA satellites are a valuable tool for graphically understanding changes in vegetation and land use at a range of temporal and spatial scales.

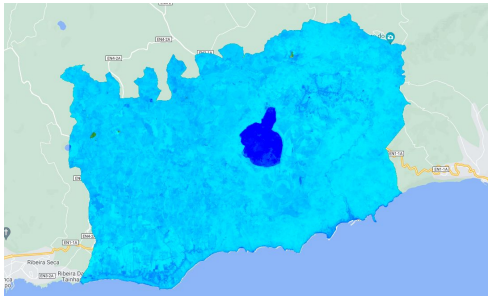
NDVI and other indices

Ioana

- NDVI is the most commonly applied index
- NDVI is a proxy of primary productivity and is directly related to plant growth and density
- Vegetation indices produce a single image indicating the amount of green vegetation production and biomass
- Low index values usually indicate less vegetation while high values indicate higher photosynthetic activity.
- Other spectral indices: NDWI, EVI, ARVI, SAVI

$$NDVI = \frac{(\text{Near Infrared} - \text{Red})}{(\text{Near Infrared} + \text{Red})}$$





Monthly composites

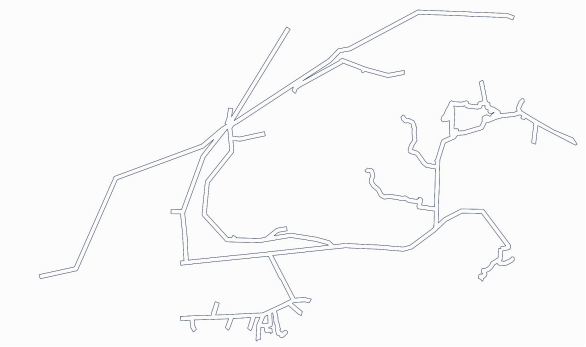
Farzad

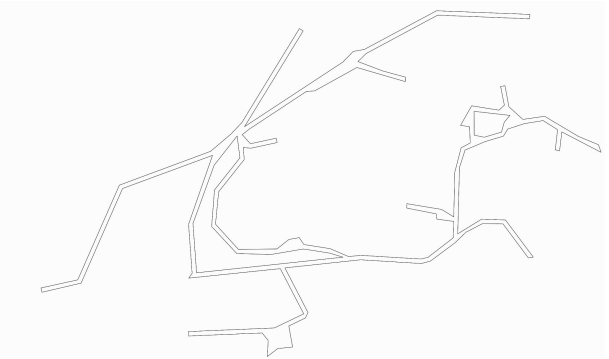
- Sentinel-2 satellite has a 10 days temporal resolution
- Calculated median images
- This resulted in diminished cloud cover and better quality images
- Compared the results with daily images, as well as biweekly and quarterly composites

Working with Geospatial data

Shrujal

- The vector format indicates the location and shape of features and boundaries accurately.
- Point, line, shape, polygon or area
- E.g. shapefile (.cpg, .dbf, .prj, .shp, .shx), kml, or geoJson format
- Using the Earth Engine API we loaded our powerline corridor in all these different formats, however the error was with collabs capacity
    - Possible solution using image reduction (reduce complexity of the multipolygon)

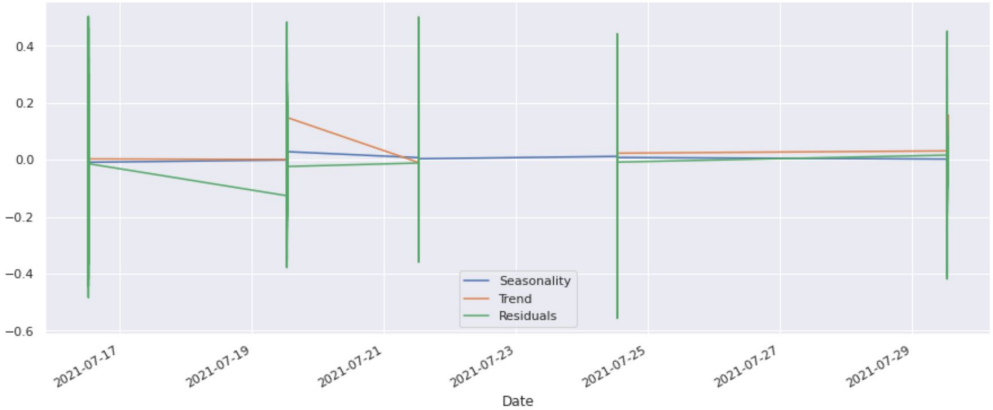




Time Series visualizations

Shrujal

- Started by using the sentinel-2 cloudless collection using a function in python applied to all the images
- Tried using geemap, geopandas, and functions from the earth engine api to convert a shapefile of the power line corridors to a geojson format or just a plain polygon
    - Issue was size and complexity of the corridors
- To see if it would work tried using a point in the study area with a large buffer and converted that region to information I could use to
- calculate NDVI from the bands and then put the NDVI information into a pandas dataframe which was successful, and we saw a plot of NDVI over time which also looked reasonable

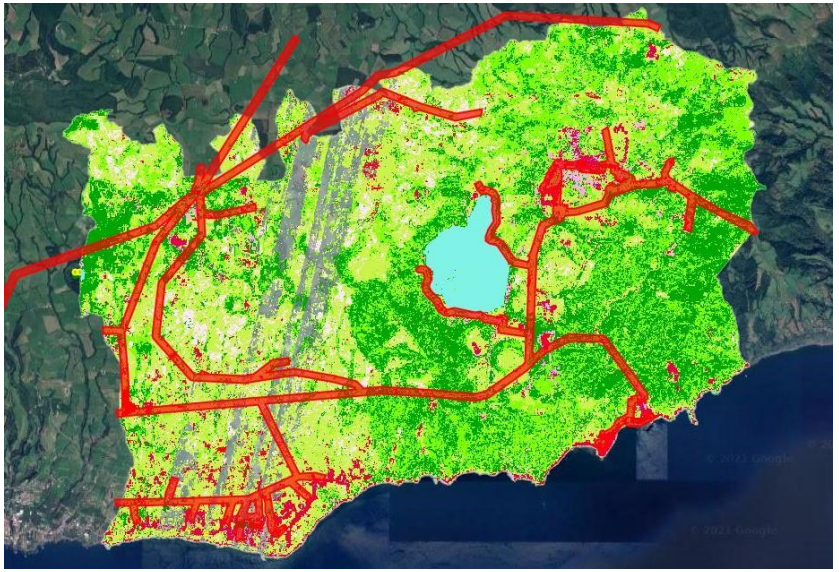


Random Forest Classification

Shrujal

1. Importing and Visualizing Data in GEE
    a. Import Study Area
    b. Import training points and validation points
    c. Cloud masking with sentinel 2

2. Setting up and implementing classification
    a. select the spectral bands that are relevant for the classification
    b. The training points will be used to extract the reflectance values of the pixels from all spectral bands and this will be passed to the classifier algorithms.
    c. Used Copernicus CORINE Land Cover catalog to catalog pixels classified by color
    d. CART and RF



overall accuracy: CART **1.000**, RF **0.968** for Date Range: **2021-01-01 → 2021-07-01**

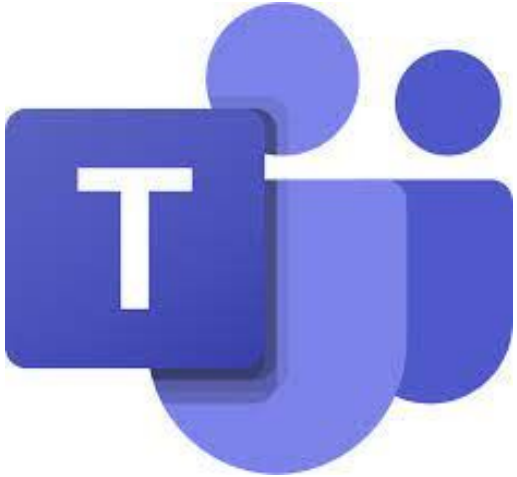
Future Steps

- Use our different composites as testing data to test accuracy of random forest model
- Apply prediction algorithm to predict vegetation encroachment on power lines, in order to advise current and future construction
- Organize all of our past work in a way that fits industrial standards

What we learned

Farzad

- Geospatial Data Files and Formats
- Machine Learning Algorithms
- Google Earth Engine
- Image Classification
- Cloud Masking
- Teamwork, Organization and working across timezone



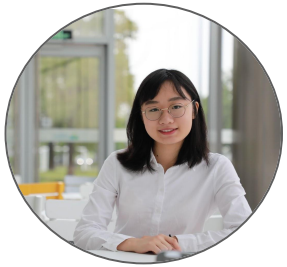
Meet The Team

Shrujal

Aditya

Ioana

Farzad



Islina

Nuno

Orquidea

Daniela

Thank you.