Development of Google Earth Engine-Based App for National Forest Inventory Plot Data Exploration

PURDUE UNIVERSITY | The Data Mine

SCOTT CLAYTON | ANGELINA DE LA TORRE | GABRIEL FERRER | DUWAYNE GRAY | MASHELLE GREGORY
BRUCE NGUYEN | ADAM ORNELAS | PAPA QUAINOO | JACQUELINE SANCHEZ | CARSON SCHUMM | EYAAN SYED

USDA Forest Service

Introduction

Efficient land use starts with understanding the available resources. Our team is focused on developing a comprehensive approach to land resource monitoring, leveraging both in-house and geostationary data to create an algorithm that displays the land's different use cases. Our objective is to provide land resource information and usage trends to those who depend on the forests. In this poster, Costa Rican land data is used.

1. Nature of the Data

- Nine land use classifications data points are determined and grouped per hectare. (Fig 1)
- These groups are called plots and are obtained for every hectare across the country (Fig 2)
- This totals to 101160 individual data points from 11240 plots, with additional information for vegetation, water, and ground types.

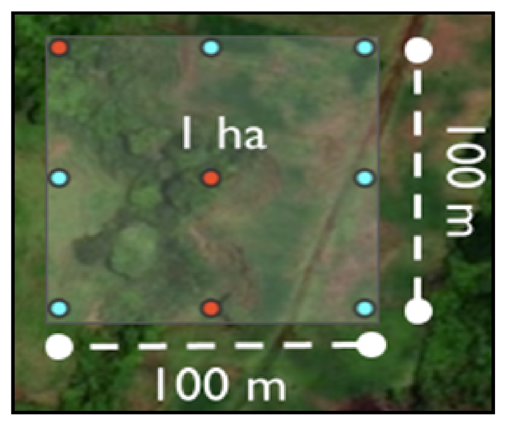


Fig 1. Land Use Data Points per Hectare

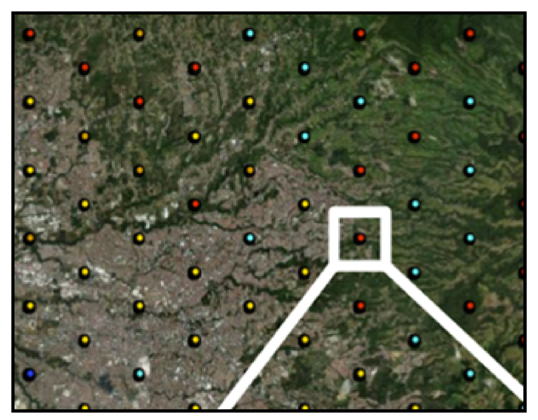


Fig 2. Land Use Plots

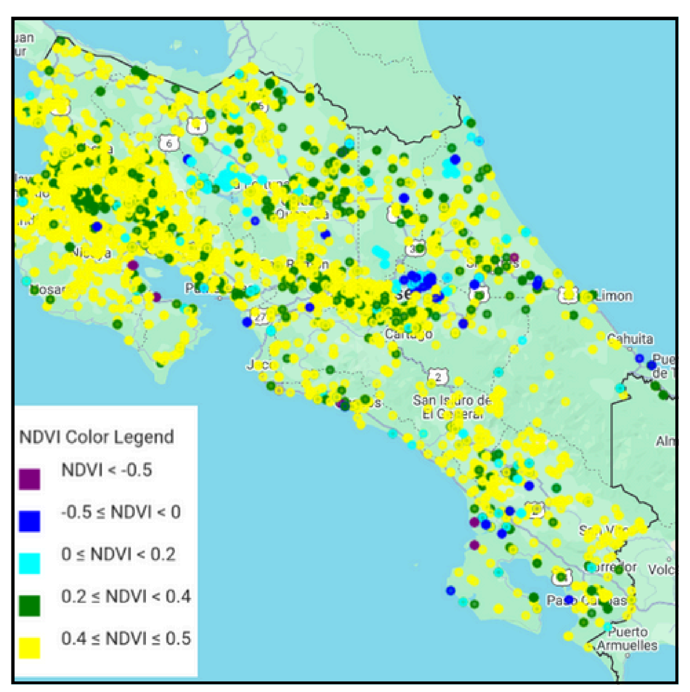
2. Exploratory Data Analysis (EDA)

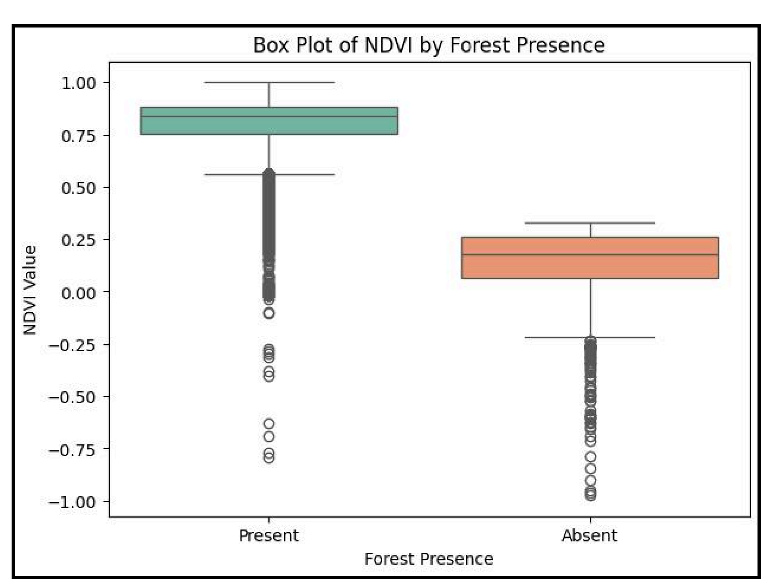
Cleaning the Data

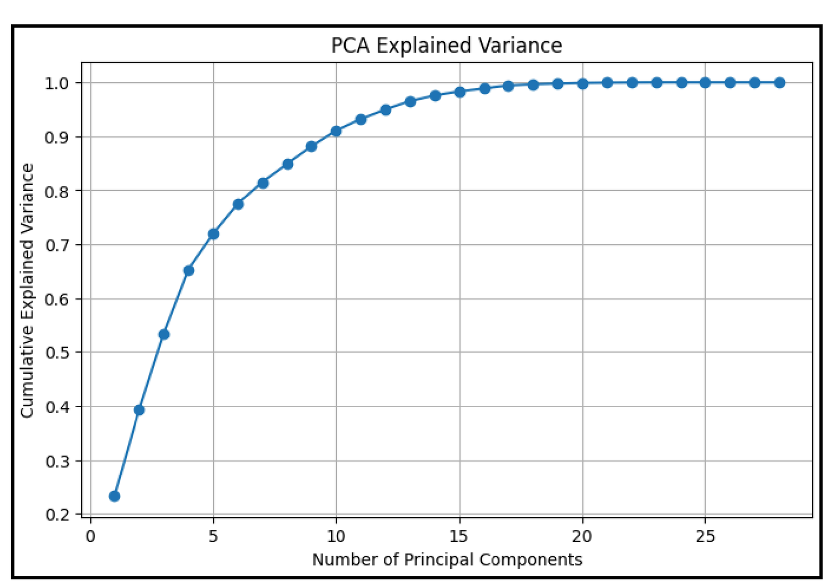
- Raw data had to be preprocessed and cleaned due to duplicate instances of data:
  - Concluded the duplicates were for quality assurance reasons
  - Compared QA and initial entries and removed extras

Outlier Detection

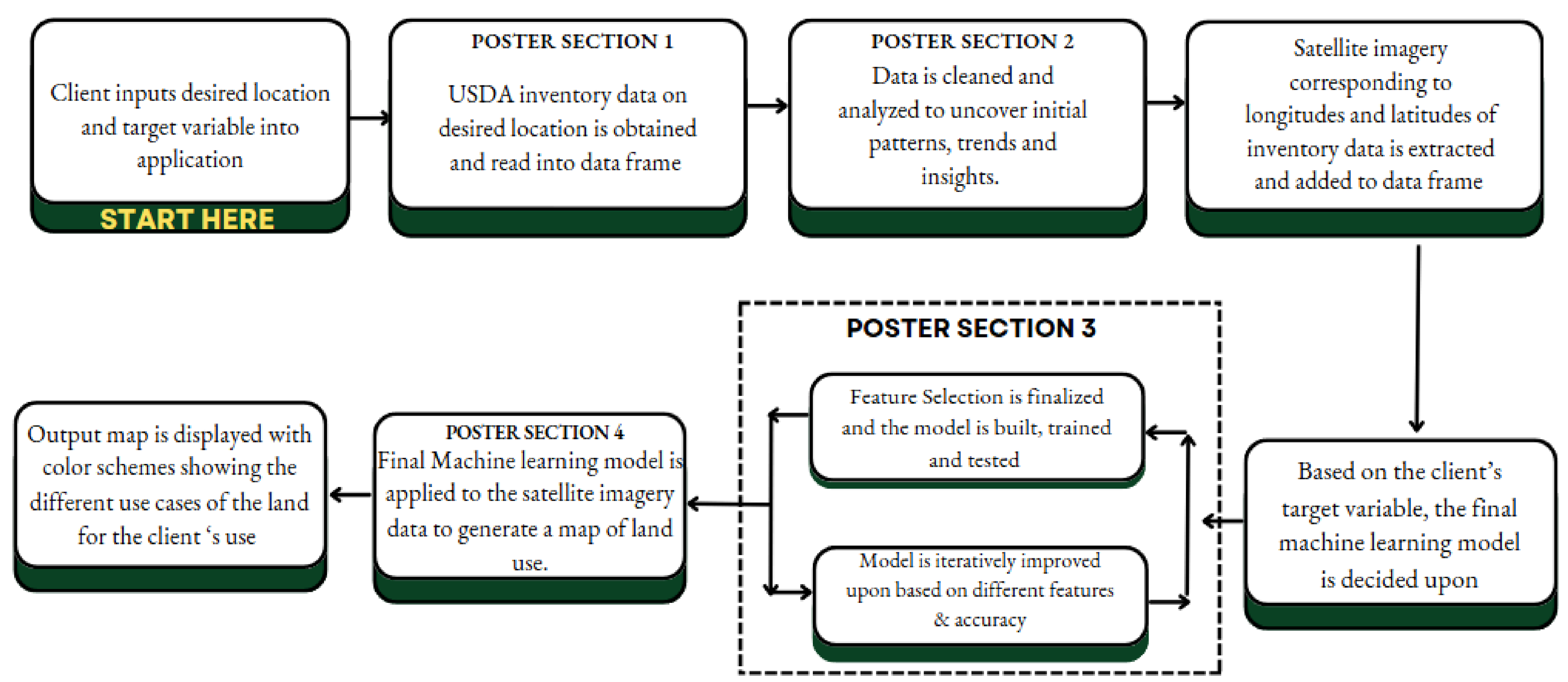
- Using box-and-whisker plots and other graphical analysis tools:
  - Identified and removed outlier points in dataset







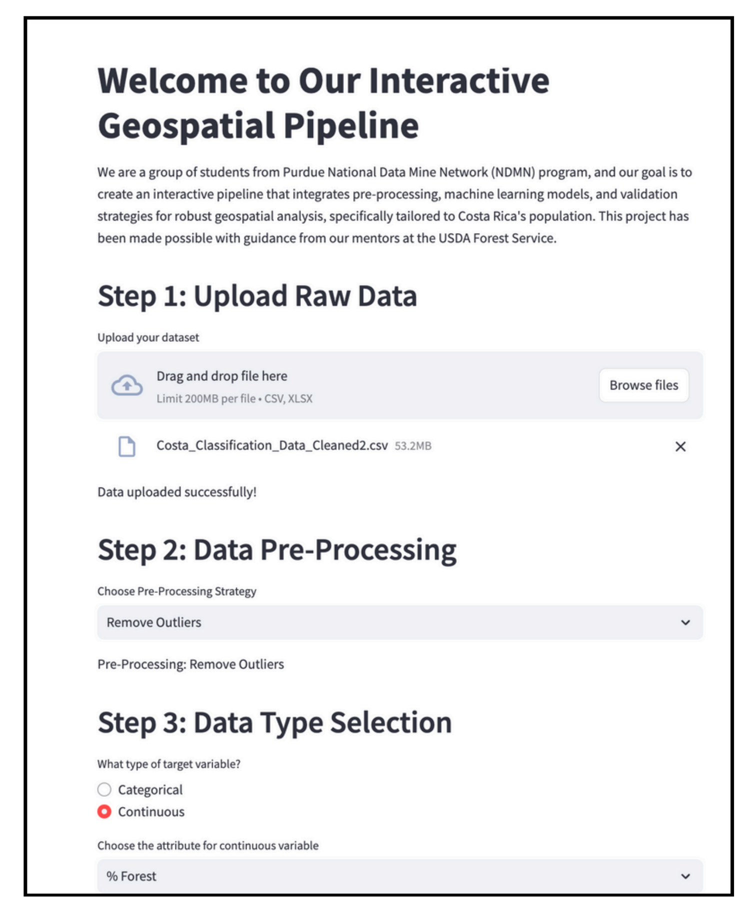
Product Workflow

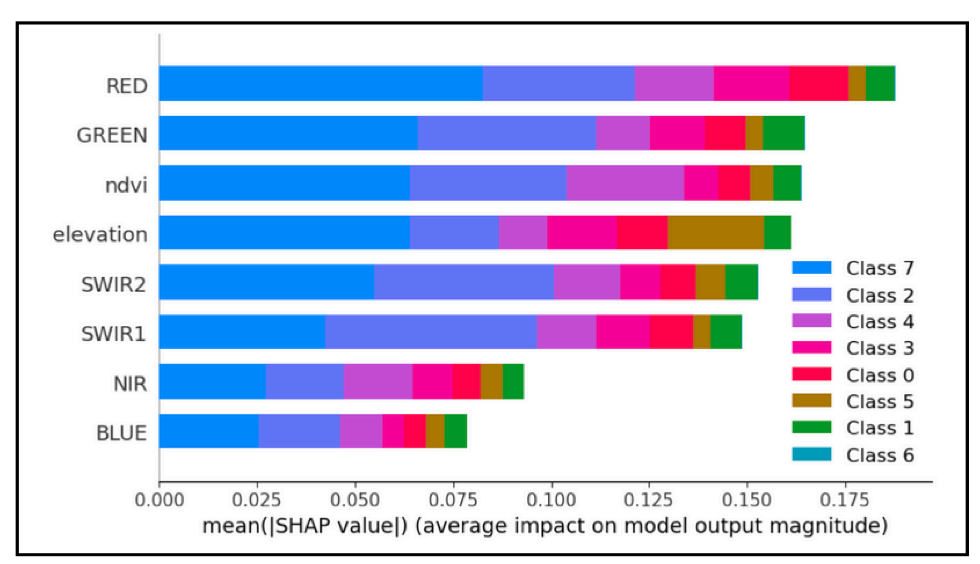


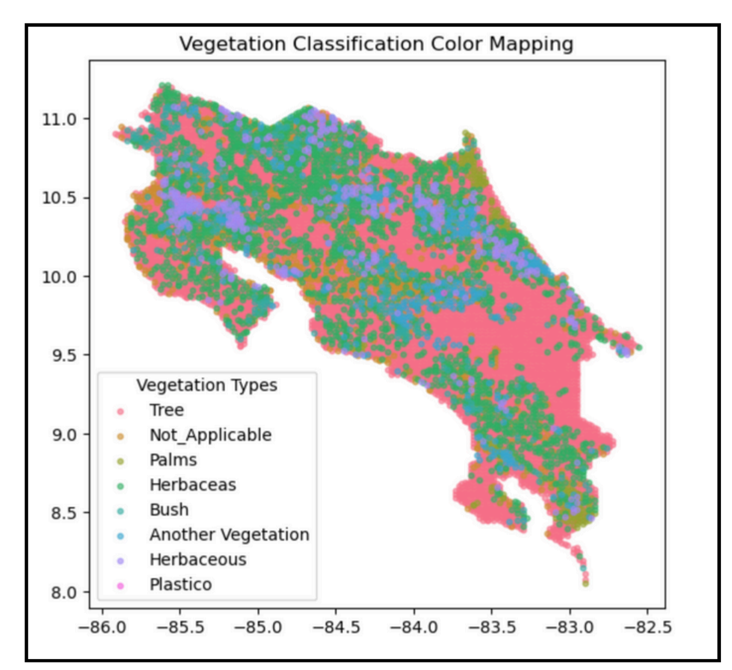
4. App Dashboard

Dashboard Purpose:

- The application integrates remote sensing data with Random Forest models to classify vegetation types and assess forest conditions.
  - Cleans and filters spatial geometries for accuracy
  - Extract key spectral features (e.g. NDVI, NIR, RED)
  - Trains and evaluates classification or regression models
  - Maps predictions using tools like Google Earth Engine and Folium
- The goal is to generate, data driven maps to support land research and
  sustainable land management





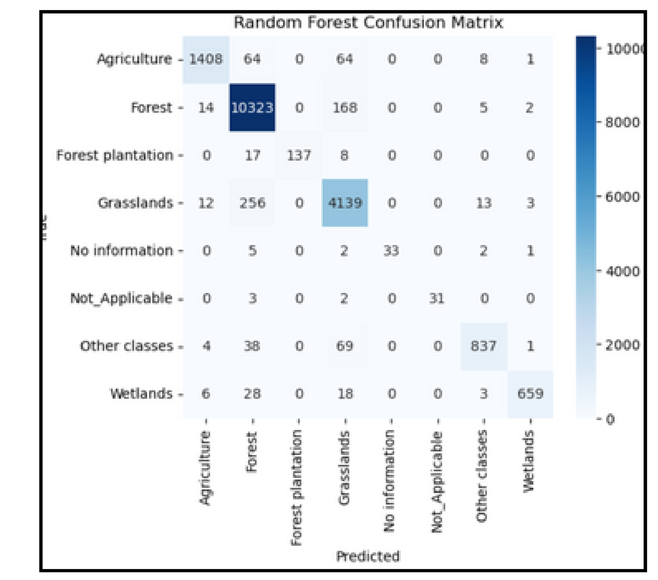


3. Machine Learning & Methodologies

Model & Feature Selection

Multiple Machine Learning (ML) classification models were trained

- Logistic regression, Random Forest, Support Vector Machine
- Predictor variable is land use
- 80-20 split was used to partition the dataset into training and testing sets



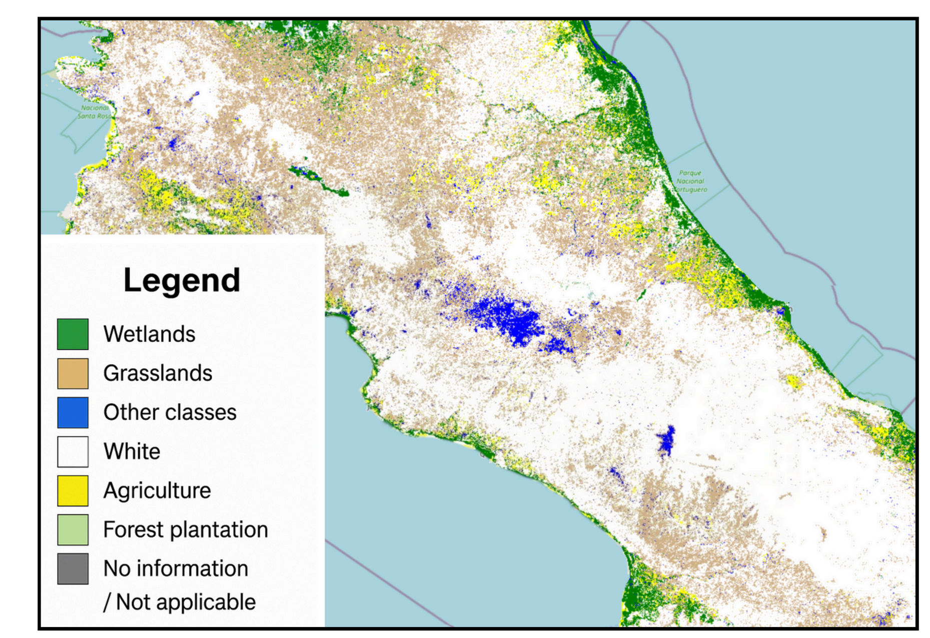
Model Results

- Models were evaluated by ccuracy, precision, recall, F-1 score, and confusion matrix
- Random forest performed the best at 96% accuracy,
  - It may be overfitting, as accuracy scores varied across test samples
- Dettermined that some of the most improtant features across models include elevation, NDVI, and other remote sensing variables such as red, green bands nad shortwave infrared

Your paragraph text

5. Conclusion & Results

- Obtained a map of Costa Rica based on the different land use cases using our Machine Learning Model.
- Built an interactive application dashboard for potential external use



6. Future Work

Adjustments & Future Projection:

- Collect feedback from Costa Rican scientists and authorities on project results and dashboard
- Update the project's pipeline and dashboard as necessary and as requested
- Enable the software and application to be available for external companies and countries

Acknowledgements

Thank you to our Corporate Partners, our professional mentors Andy Lister, John Hogland, and Rachel Riemann, as well as our peer mentor Wendy Jiang.