



# Satellite Data Monitoring Global Deforestation

By James Hoang

# Presentation Structure

1. Overview of Subject Area
2. Dataset & Preprocessing
3. EDA Findings & Visuals
4. Model Comparison & Interpretation
5. Product Proof of Concept

# Overview of The Subject Area



**Problem Statement:** Using machine learning, how might we predict areas at risk for deforestation such that we can determine areas ideal for logging by least environmental impact?

**Solution :** Using carbon and forest density data to help highlight areas which have the highest contribution to deforestation where its suitable for logging

**Potential Impact Estimate :**

- Insights help guide emission reduction policies
- Deforestation risk prediction on future forestry activities



# Overview of Dataset & Preprocessing



## Datasets:

- Carbon emissions & Tree density loss
- Yearly metric 2001-2023
- Forest thresholds to classify forested areas

## Preprocessing Approach:

- Binarized country feature
- Dropped rows with thresholds below 30
- High colinearity



# Important EDA Findings



Most carbon is stored in primary forests which are areas that are at risk for deforestation



Carbon emissions and tree cover loss are very positively correlated



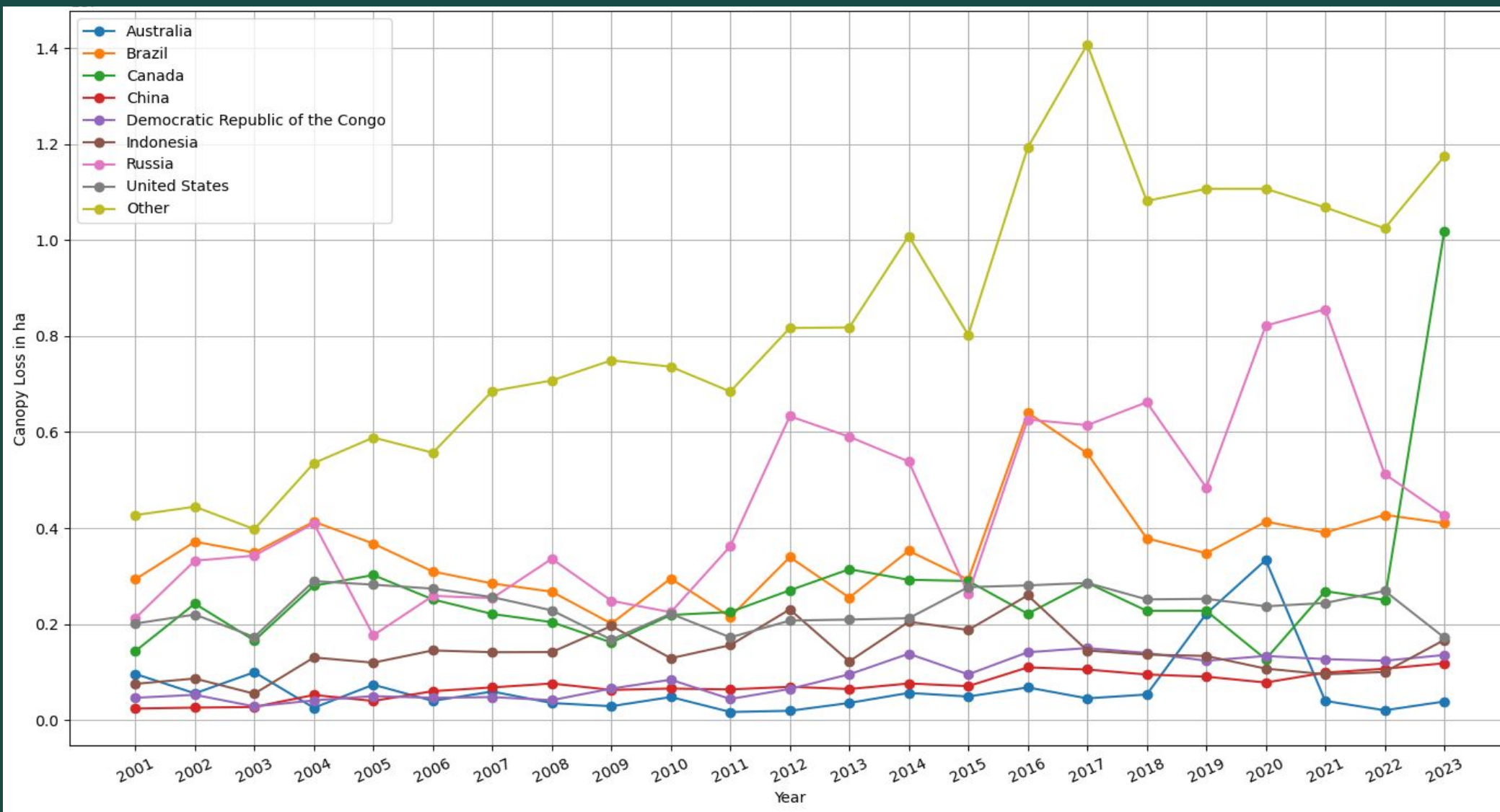
Tree cover loss is becoming more sporadic overtime and more difficult to predict



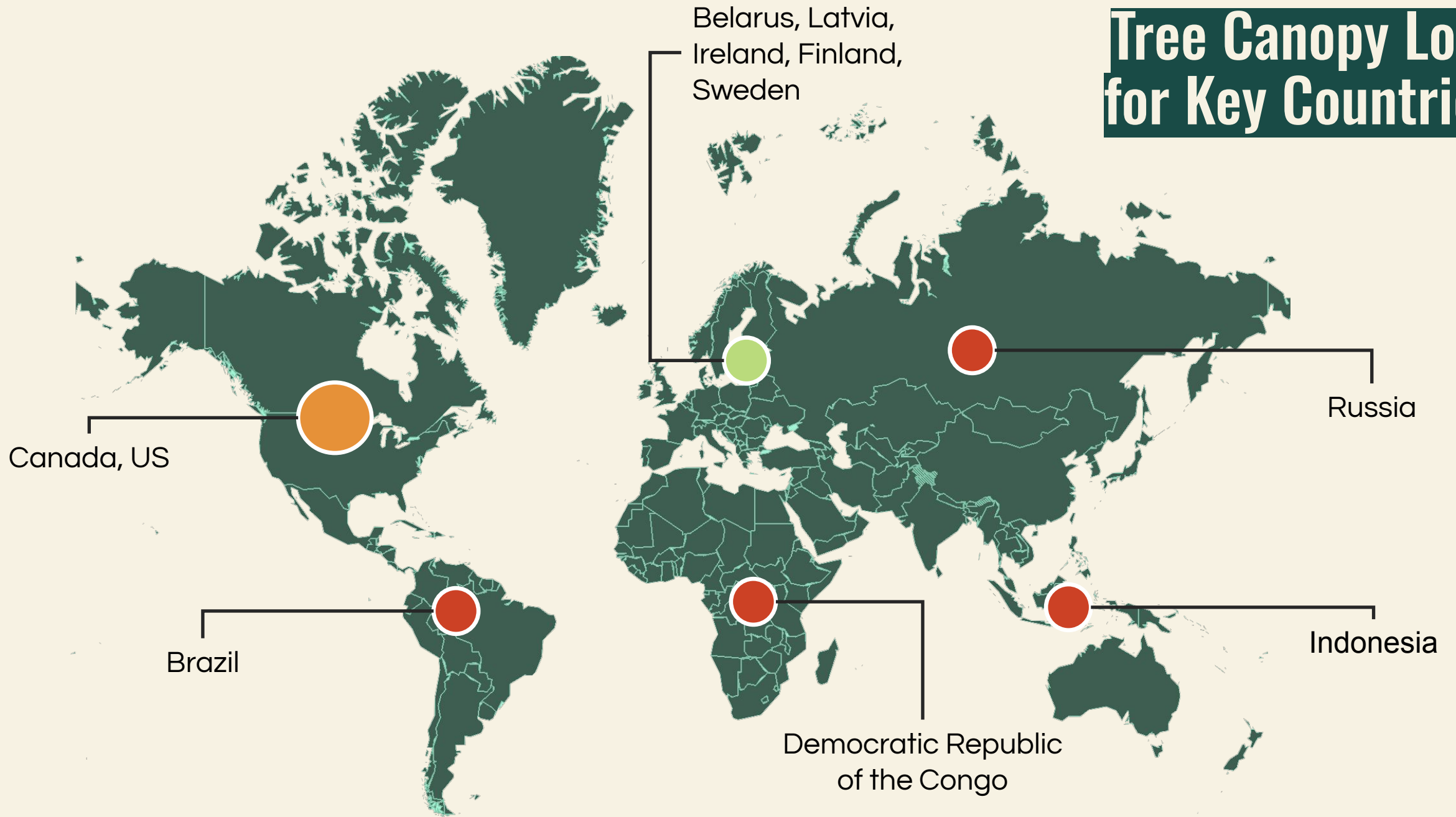
Many large spikes in tree canopy loss largely due to wildfire events



# Tree Canopy Loss for 2001 - 2023



# Tree Canopy Loss for Key Countries



# Model Comparison



**Lasso Regression Targets:** Net carbon emissions, Net tree density gain

**Evaluation :** R-squared, MSE, MAE

**Optimizations :** MinMaxScaler

Carbon:

- PCA with 6 components (99.56% variance explained)
- Lasso with alpha 0.01

Tree Loss

- PCA with 8 components (99.59% variance explained)
- Lasso with alpha 0.01

**Best Performance:**

Net Carbon Emissions:  
**83.80%**

Tree Canopy Gain:  
**97.81%**



# Model Interpretations



## Carbon Emissions:

Currently it aims to predict country's net carbon emissions using historical emissions data

- Identify which year emissions or biomass matter most
- Make predictions for countries not included in my dataset or subnational regions

## Tree Density Loss:

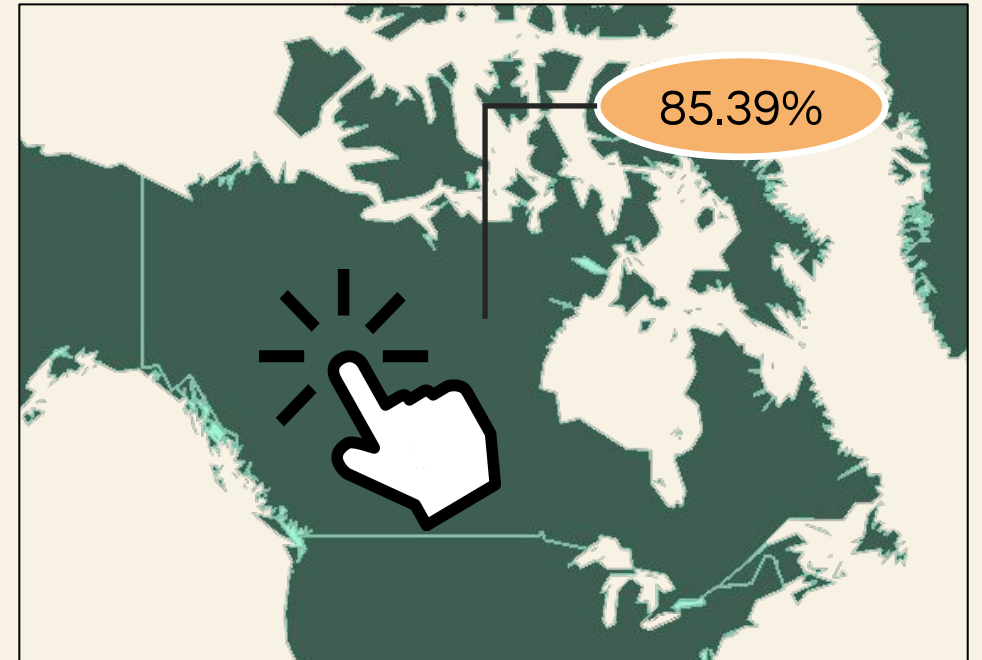
- Make predictions on tree cover gain or loss for regions not in the dataset
- predicting potential reforestation and deforestation outcomes in for designated areas

# Design for the Product / Proof of Concept



## Proof of Concept : Interactive map

- When User clicks a country or region it displays a risk score metric on how at risk of deforestation the specific area is
- Displays the predicted deforestation risk for next year
- Granularity of area can change



# Future Work

## Datasets

- Forest activities like wildfires, forestry, agriculture
- Climate temperature (wildfire direction)
- Granularity can be adjusted from subnational, region, city etc.

## Potential Models:

- Neural Networks
- Random Forests





# Thanks for Listening!

Presentation Template: [SlidesMania](#)