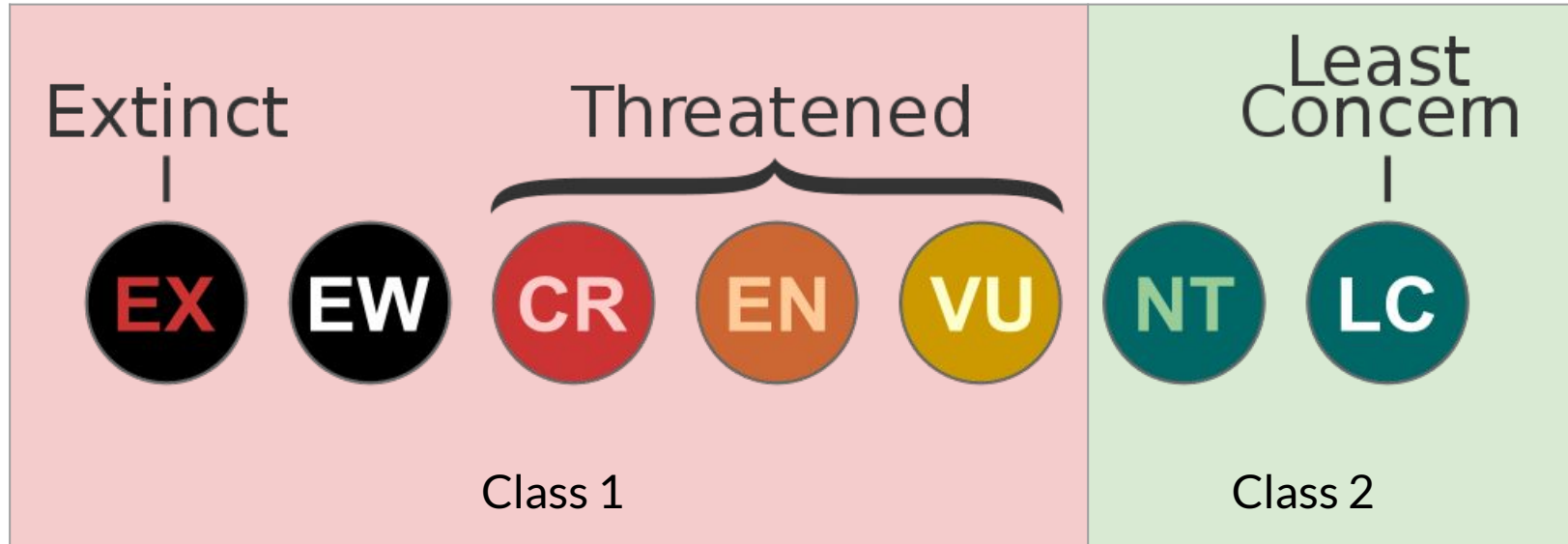




# Predicting Threatened Species

Solomon Klein

## What is the IUCN Redlist?





## Data Collection



Encyclopedia of Life

## Data Processing

- Download via Cypher queries
- Format in SQL

## Analysis

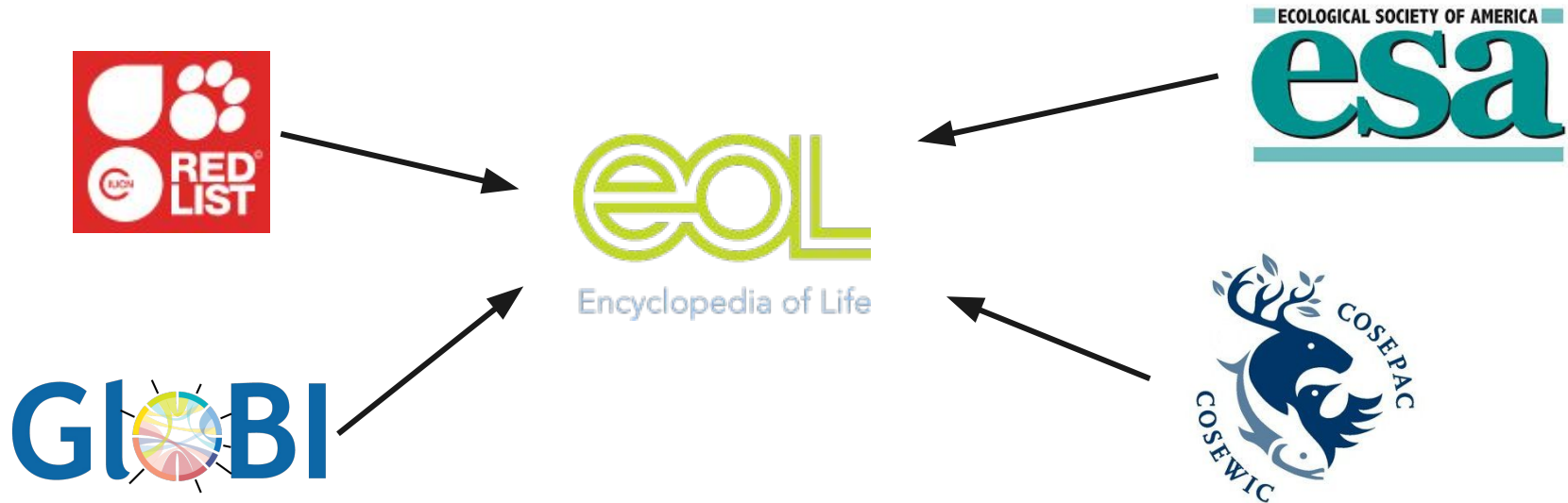
ADASYN

Gradient Boosted  
Classifier

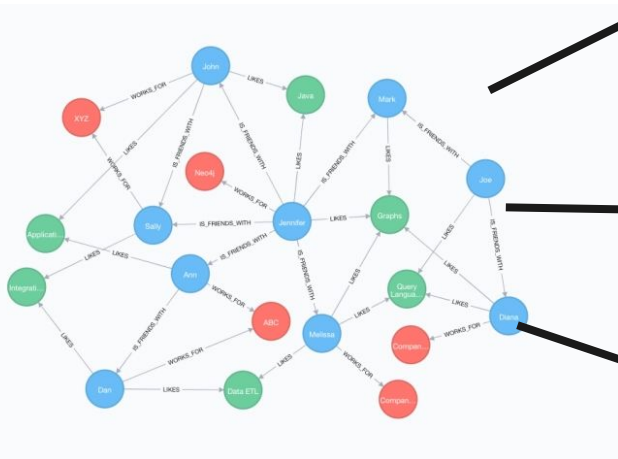
## Dashboard



# Encyclopedia of Life



## Data Collection & Cleaning



trait1 | animal\_id

-----+-----  
50.0 | 11728

trait2 | animal\_id

-----+-----  
3.2e5 | 11728

trait3 | animal\_id

-----+-----  
0.05 | 11728

animal\_id | trait1 | trait2 | trait3

|       |      |       |      |
|-------|------|-------|------|
| 11728 | 50.0 | 3.2e5 | 0.05 |
|-------|------|-------|------|

## Selected Features



Red List status

Habitat

Food type

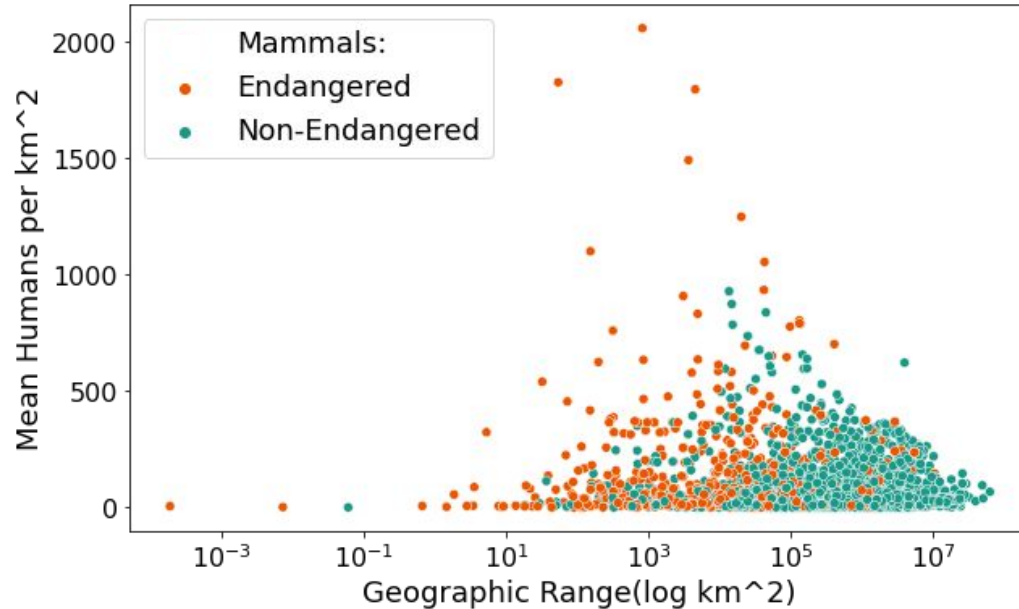


Body mass

Geographic range

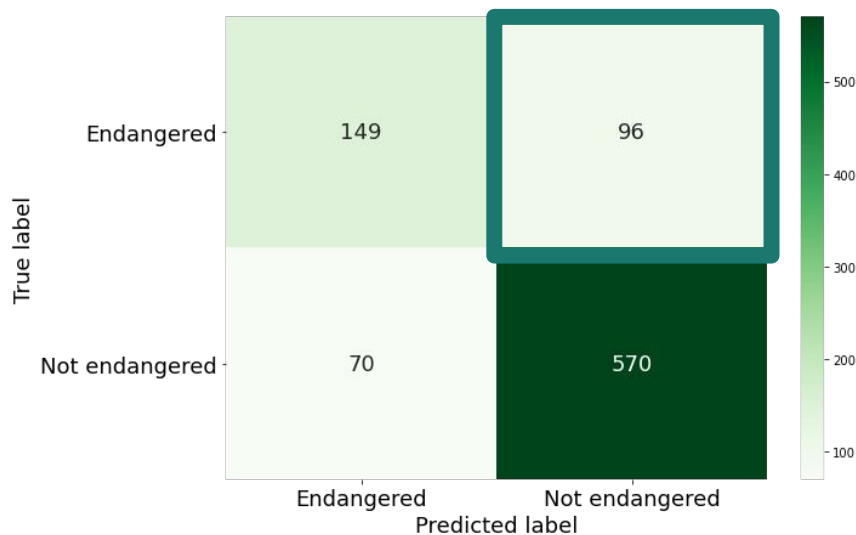
Human population in that range

# Classification Issues: Mixed Data

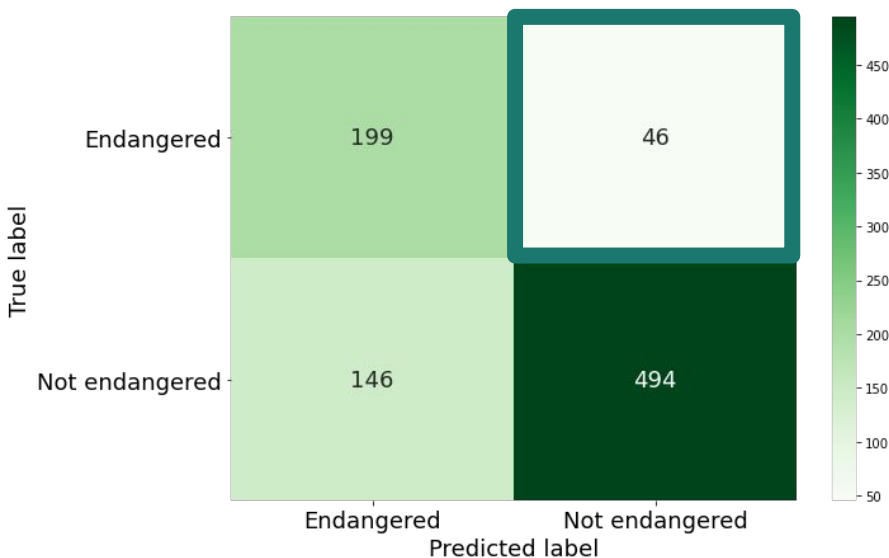


# ADASYN and Gradient Boosted Classification

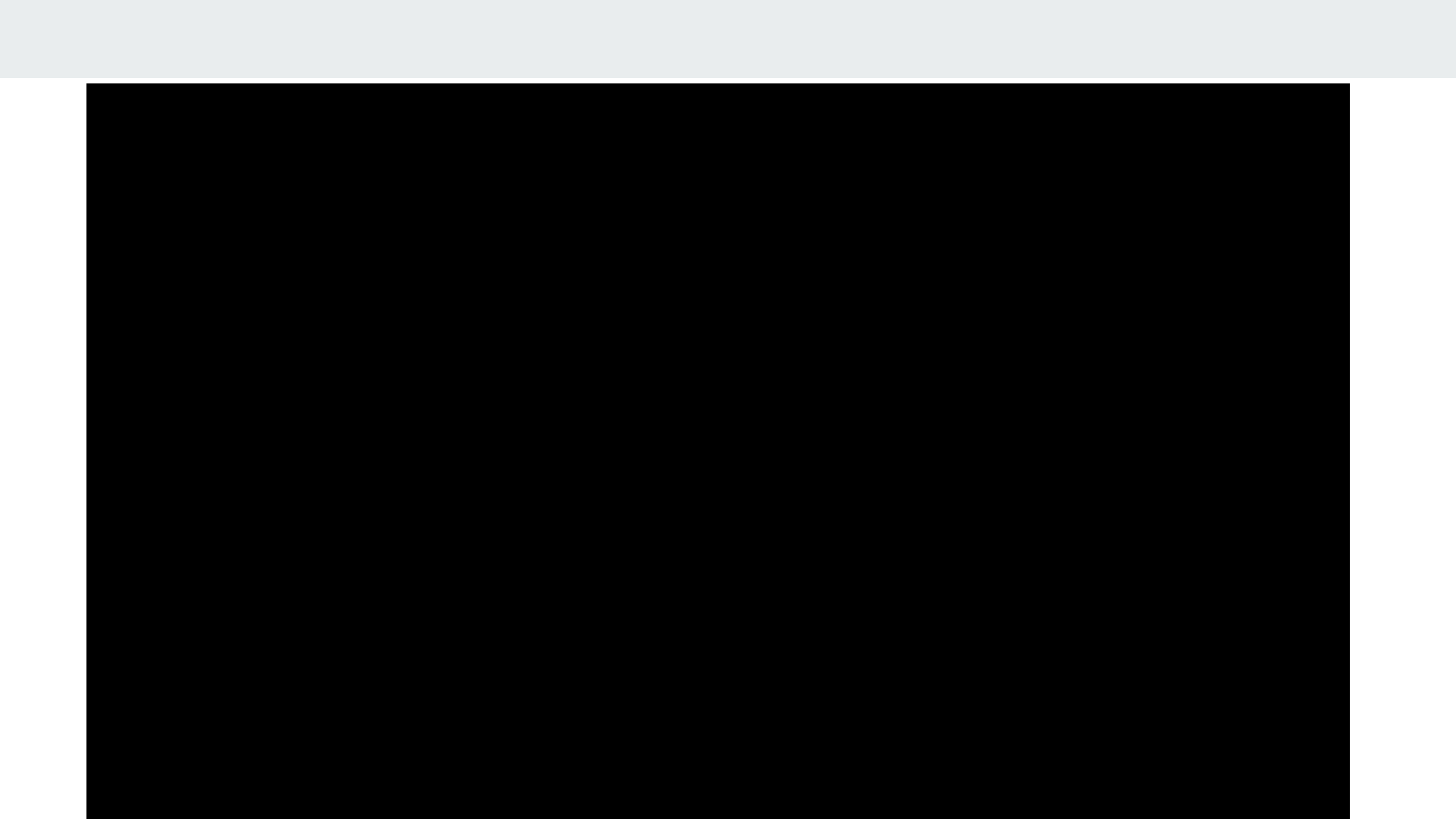
Without Resampling:



With Resampling:









# Questions?

Solomon Klein

[solomon968@gmail.com](mailto:solomon968@gmail.com)

<https://github.com/computerGeologist/>