

# Leon Reilly

B.A.S Mathematics and Philosophy ('27)  
Technical Mentors: Dr. Ben Seiler, Dr. Mike Baiocchi,  
Dr. Kim Babiarz, Jonas Junnior

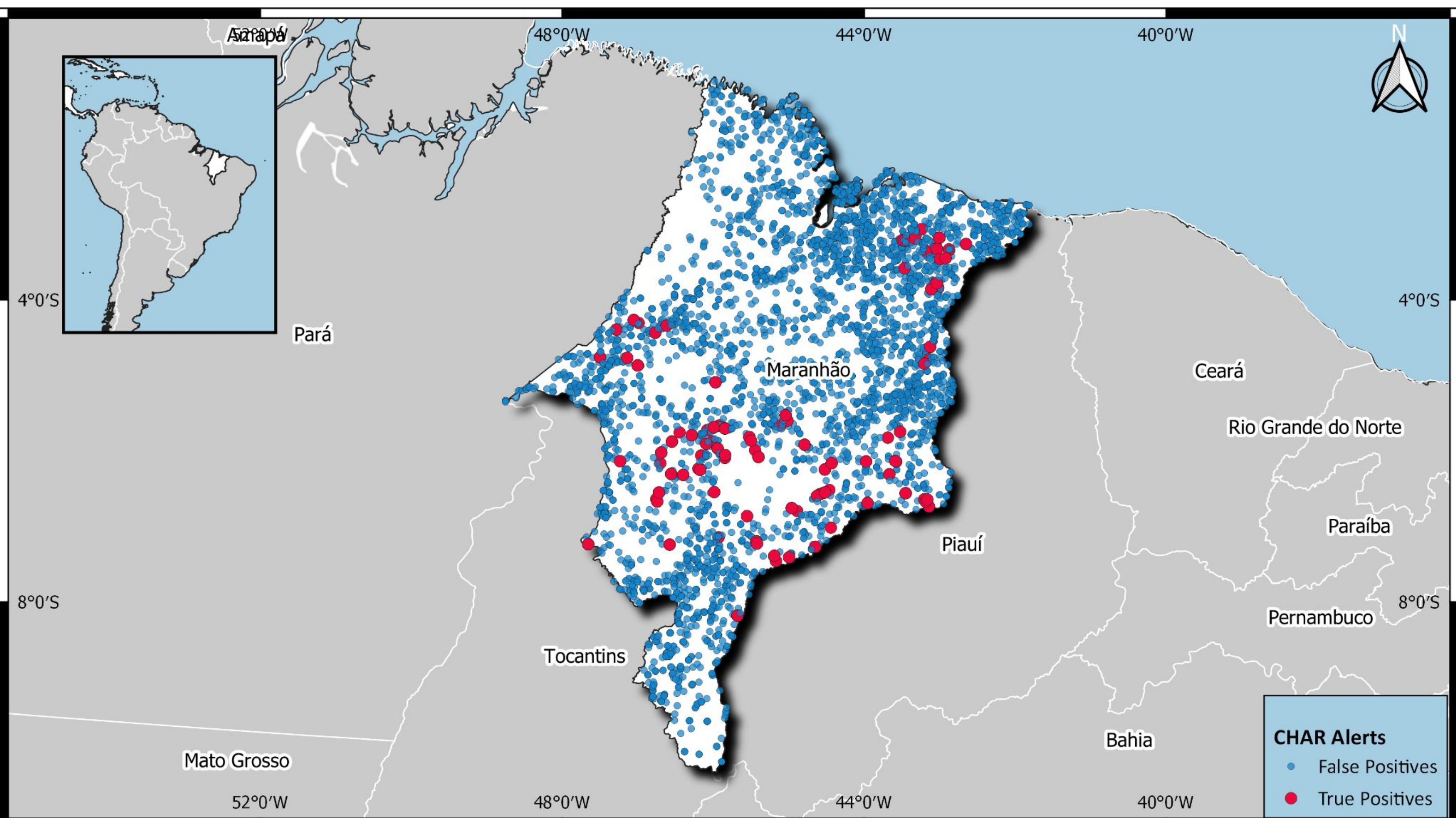


## ML Augmented Prediction for Labor Exploitation Detection

Keywords/AMA: Human Trafficking; Geospatial Machine Learning; Satellite Imagery; Computer Vision

### Birds-eye View

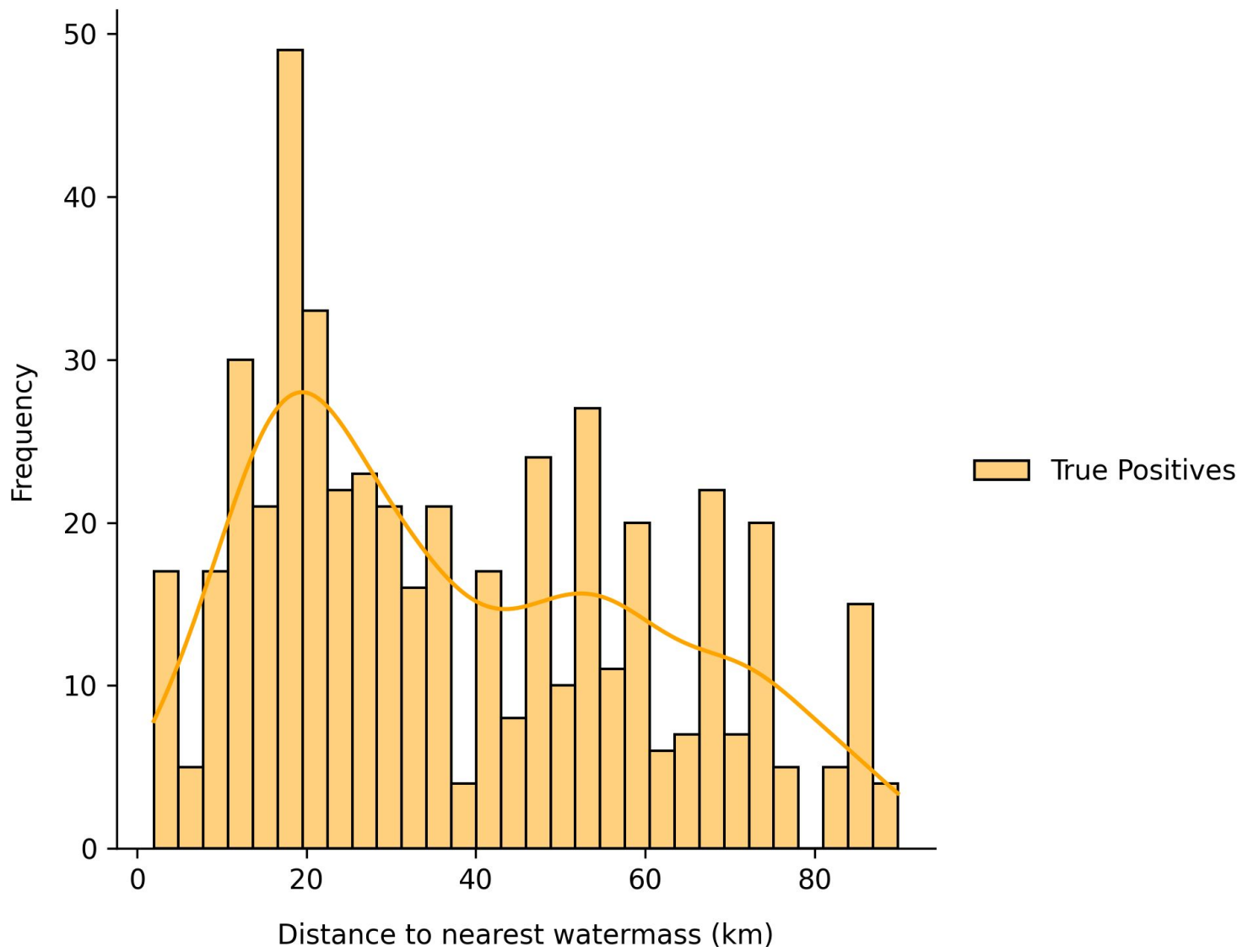
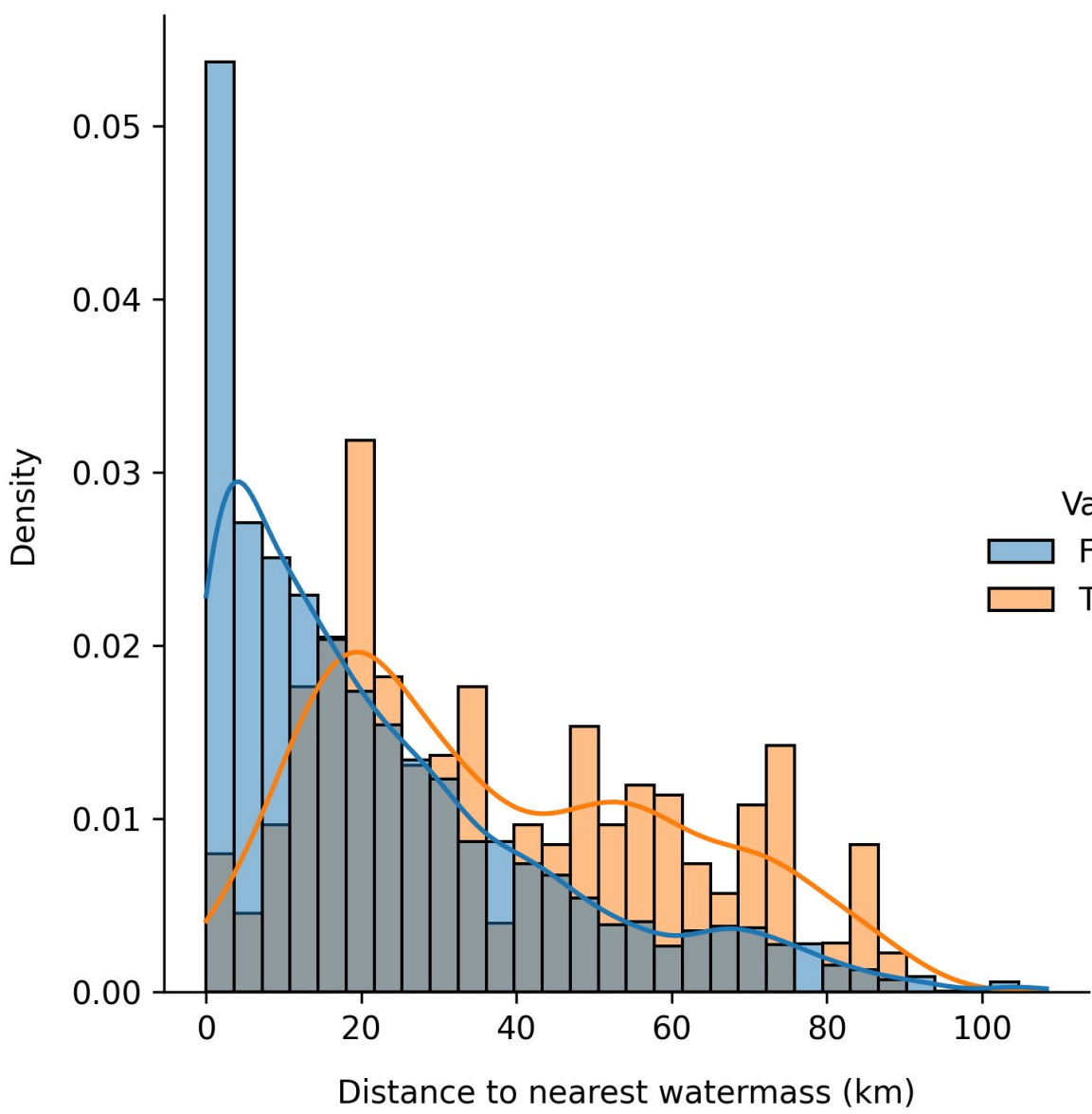
- Charcoal production in Brazil is often linked to exploitative labor and slavery-like conditions, with around 2,830 people rescued since 1995.
- Combating illegal charcoal production is challenging due to its remote locations and temporary operations.
- The Stanford Human Trafficking Data Lab and Brazilian Federal Labor Prosecution Office use a tool called CHAR, which employs computer vision on satellite imagery to detect potential forced labor sites.
- The project aims to improve the accuracy of site identification by developing a downstream Gradient Boosting Classifier model, reducing false positives and enhancing anti-trafficking interventions.



### Feature Engineering

Given the information we have about the charcoal sites, we carefully interrogated the contours of what data might have signal. We used three data sources:

- Geographic Features:** Features like roads, villages, lakes, landcover type etc. may reasonably be expected to be important considerations when determining where to create a charcoal site. For example, proximity to a road is important since charcoal needs to be transported.
- SmartLab:** Contains survey data of every municipality in Brazil. Includes data like literacy rate, poverty rate, number of workers rescued, and so on. This data may be a proxy for likelihood of trafficking in a municipality.
- MapBiomass:** Contains geometries of deforestation alerts that are updated every two weeks by the Brazilian government. Charcoal is made from wood—therefore we might expect charcoal sites to be close to deforestation sites.



### Appropriate Feature Metrics

We constructed shortest-distance-to and density metrics in Python using **GeoPandas**. For example, shortest distance to a watermass and number of charcoal sites are within a given radius. This is because 1) it makes sense to ask how close a site is to a lake—at a minimum, you can't build charcoal kilns on a lake— and 2) it make sense to ask how many lakes are nearby since that information may proxy for land type and how suitable the area is for producing charcoal.

Partners: Kyler Shu ('27)  
Enkhjin Munkhbayer ('25)

