

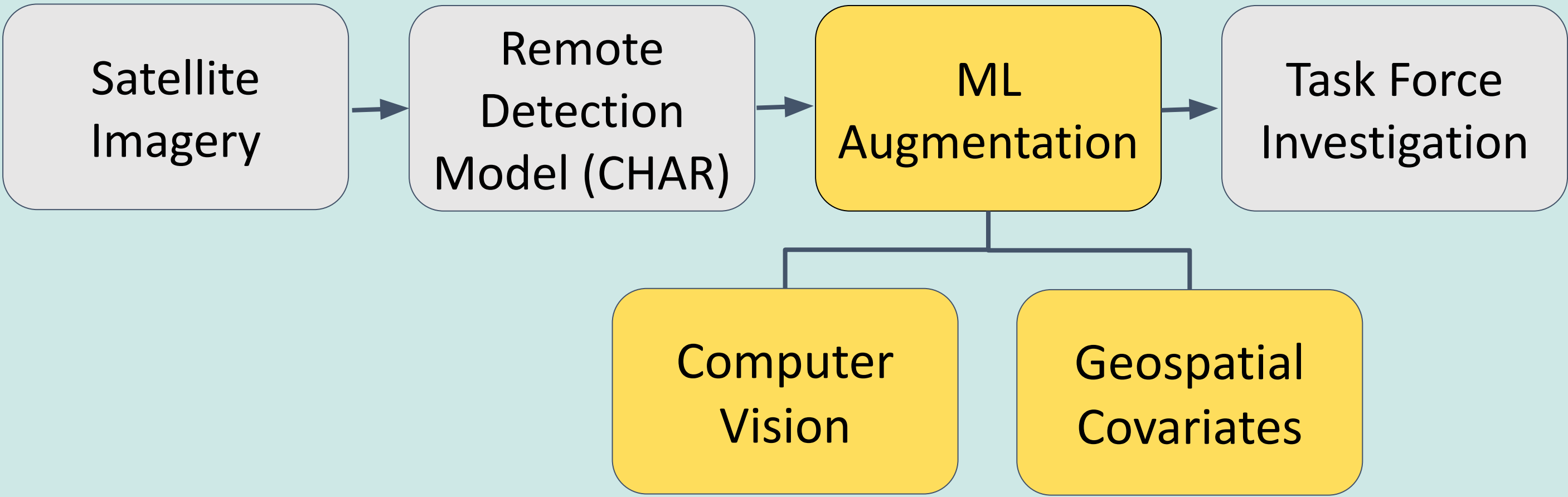


ML Augmented Prediction for Labor Exploitation Detection

Keywords/AMA: Human trafficking, geospatial machine learning, satellite imagery, computer vision

Birds-eye View

We are cracking down on trafficked/slave labor in the charcoal industry using remote satellite detection in Brazil. Our contribution is developing a downstream model to filter the thousands of incoming flagged sites to a manageable, high probability set for periodic human investigations.



About Kyler and the team

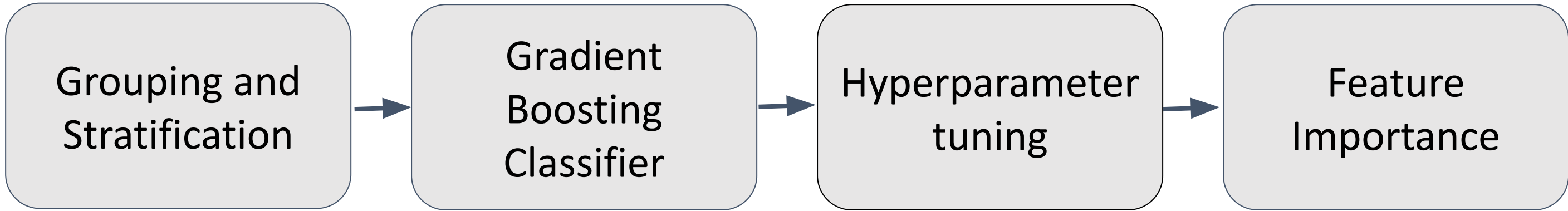
- I’m a rising sophomore interested in data science and its applications in everything from urban planning to the natural sciences
- **Partners-** Leon Reilly (‘27), Enkhjin Munkhbayar (‘25)
- **Technical Mentors (Human Trafficking Data Lab)-** Dr. Benjamin Seiler, Dr. Mike Baiocchi, Dr. Kim Babiarz, Jonas Junnior

Feature Construction

Geospatial feature type	Examples
distance to nearest feature	<i>distance to nearest road, watermass, indigenous land, village/town</i>
feature density	<i>number of deforestation alerts within 10km</i>
land categorization	<i>forest plantation, savannah formation, non-vegetated area</i>
survey/census data by municipality	<i>poverty rate, literacy rate</i>

Image encoding approaches	Examples
Variational Autoencoder with ResNet50V2 backbone	<i>128 dimensional vector embedding</i>
MOSAICS pretrained encodings	<i>4000 features</i>

ML Detection



We implemented grouping by unique location (sites were flagged repeatedly over several months) to prevent knowledge leakage across train/test splits, and stratification to balance positive labels across splits for consistent results. Among the models that we tested, sklearn’s Gradient Boosting Classifier had the best performance. Hyperparameter tuning was carried out using GridSearchCV. We used both TreeSHAP and built in methods from the model to evaluate feature importance, with the results from TreeSHAP displayed below.

