Data Science For Social Good Europe
Summer Fellowship

Developing a fishing risk framework from satellites and ocean data

Iván Higuera-Mendieta (Banco de la República - Central Bank of Colombia), Shubham Tomar (SocialCops, New Delhi), William Grimes (University College London)
Technical mentor: Jane Zanzig (University of Chicago), *Project Manager:* Paul van der Boor (Nova School of Business and Economics)

Problem statement

All indicators on ocean health indicate a biosphere in rapid decline. Demand for seafood is increasing, yet most of the world's fisheries are shrinking due to overfishing. Our partners at the World Economic Forum (WEF) are dedicated to tackling these issues.

Illegal, unregulated, and unreported (IUU) fishing is partly responsible for this decline in ocean health. In partnership with the WEF we are creating an open source tool to help combat IUU fishing. This tool combines multiple data sources to generate components that indicate the risk of a vessel engaging in IUU fishing. This tool can then be used to guide effective enforcement of our oceans.

Project objectives

- Predict the likelihood that a vessel is fishing.
- Create multiple indicators of IUU fishing.
- Generate a prioritized vessel list ordered by their "risk score," defined as the risk of a vessel being involved IUU fishing.
- Integrate multiple data sources into our fishing risk score to augment the ability to identify IUU vessels for example: satellite imagery, automatic identification system (AIS), and synthetic aperture radar (SAR), etc.

Data

- **Automatic Identification System (AIS):** vessel tracking data used for collision avoidance, is a live feed of vessel location, movement, and vessel metadata. Our framework uses historical data data collected from the period of May 2016 till June 2017 globally, by Spire Global, Inc.
- **Satellite Imagery:** visual evidence of vessels for this project was provided by two high resolution satellite imagery companies: Planet Labs, inc, and Digital Globe, inc.

Figure 1: automatic identification system (AIS) vessel trajectory data for all vessels displayed globally from May 2016 till June 2017, aggregate of ~130 million data points.

Risk framework approach

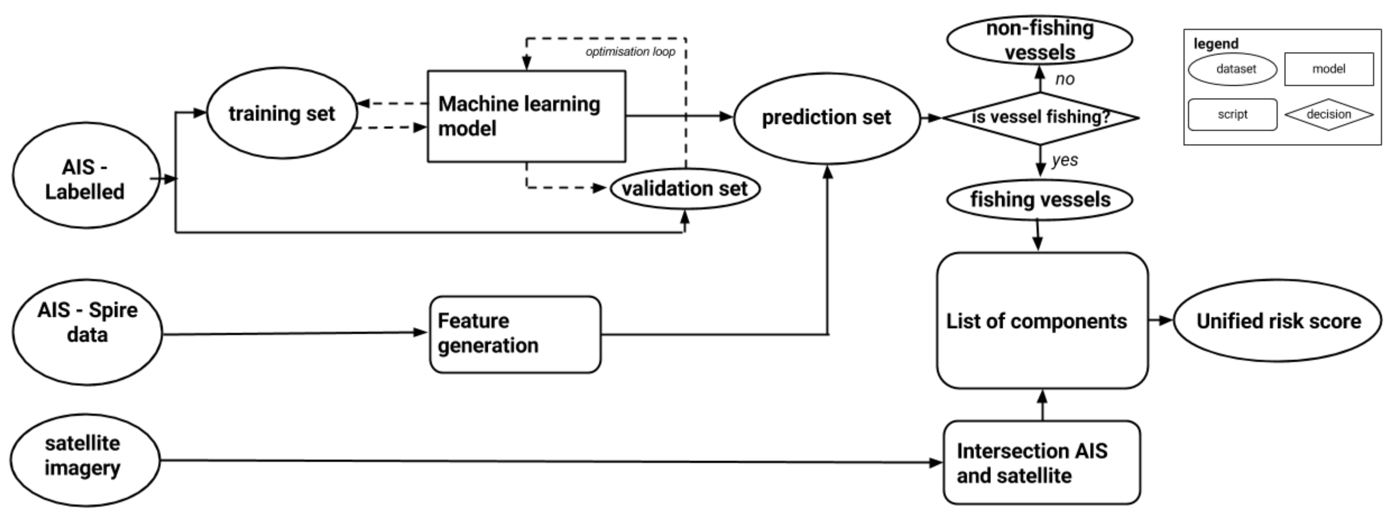

Figure 2: project workflow of fishing risk framework, showing each stage in the modelling, component generation, and integration of satellite imagery.

1. Pre-process data by removing duplicates, null rows, and invalid vessel identifiers.
2. Create features based on positional automatic identification system (AIS) data that are indicative of fishing vessels, for example distance to shore.
3. Train a machine learning model that can identify fishing vessels given their AIS positional data.
4. Create components from AIS data as a proxy for illegal, unregulated, and unreported fishing, for example whether or not a vessel is fishing in marine protected areas.
5. Combine components into a unified risk score and display as a prioritized list of vessels ordered by risk of involvement in IUU activities.
6. Intersect satellite imagery with AIS data, and where available show satellite imagery of vessels.

Modelling

A random forest model to predict at each time point whether a vessel is fishing or not.

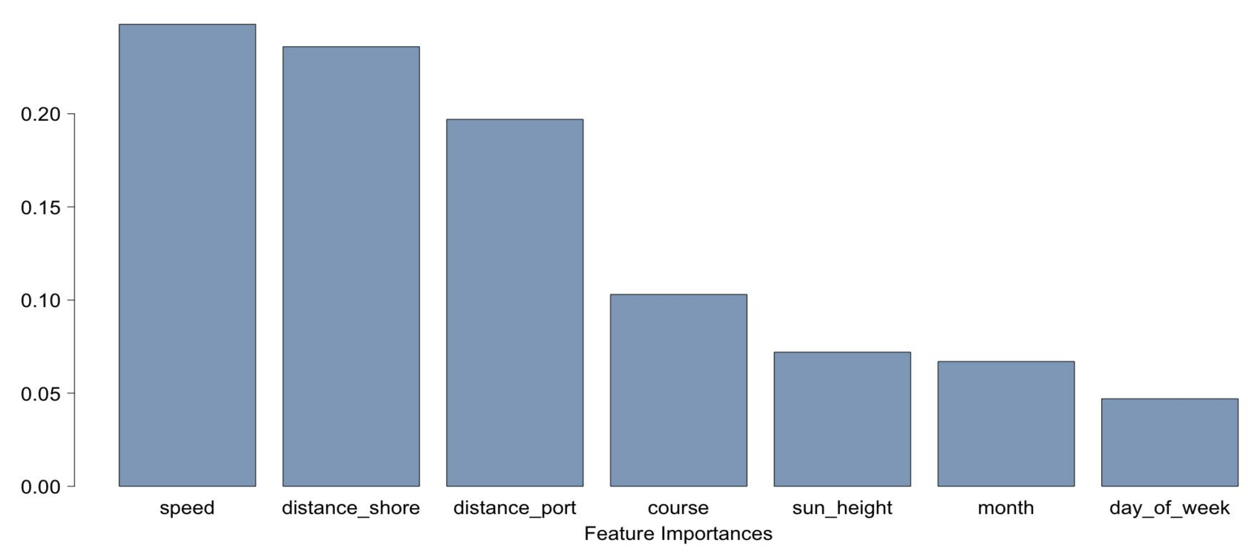

Figure 3: feature importance shown for features in a random forest model used to predict at each time point in automatic identification system (AIS) positional vessel data whether the vessel is fishing or not.

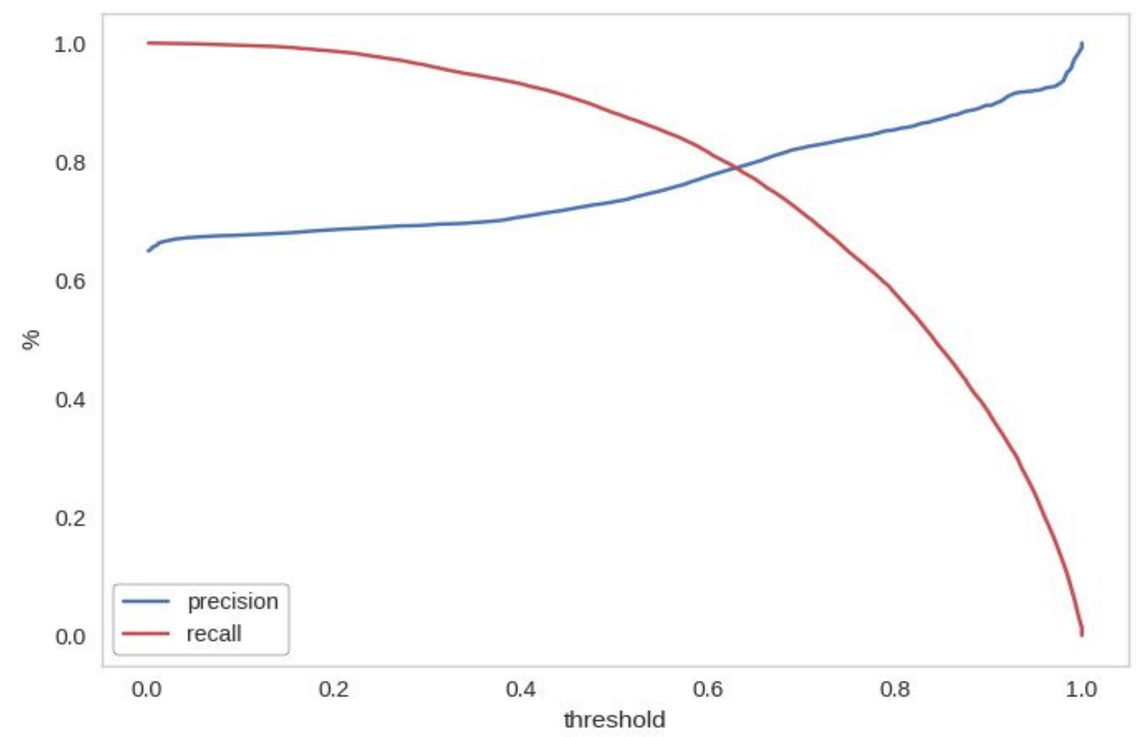

Figure 4: precision and recall plot for random forest model, predicting whether a vessel is fishing or not at a given point in time

Risk components

Component indicators were created based on discussions with domain experts, and combined into a unified vessel risk score. The following are examples of components that were integrated into our risk score:

- Vessels not self-reporting as fishing vessels
- Fishing in an exclusive economic zone (EEZ)
- Fishing in a maritime protected area (MPA)
- Evidence of transshipment
- Visual evidence of vessels involved in IUU activity

Find vessels in satellite images

The overlap between satellite imagery and AIS positional data serves as a primer of a more systematic way to detect unidentified vessels. This novel approach checks for all vessels at all time points in our historical data whether satellite imagery is available. In the future this could be a useful training data set to train a vessel detection classifier.

Figure 5: sample mosaic showing vessels located in satellite imagery based on their automatic identification system (AIS) positional data, source: Planet Labs

Web application

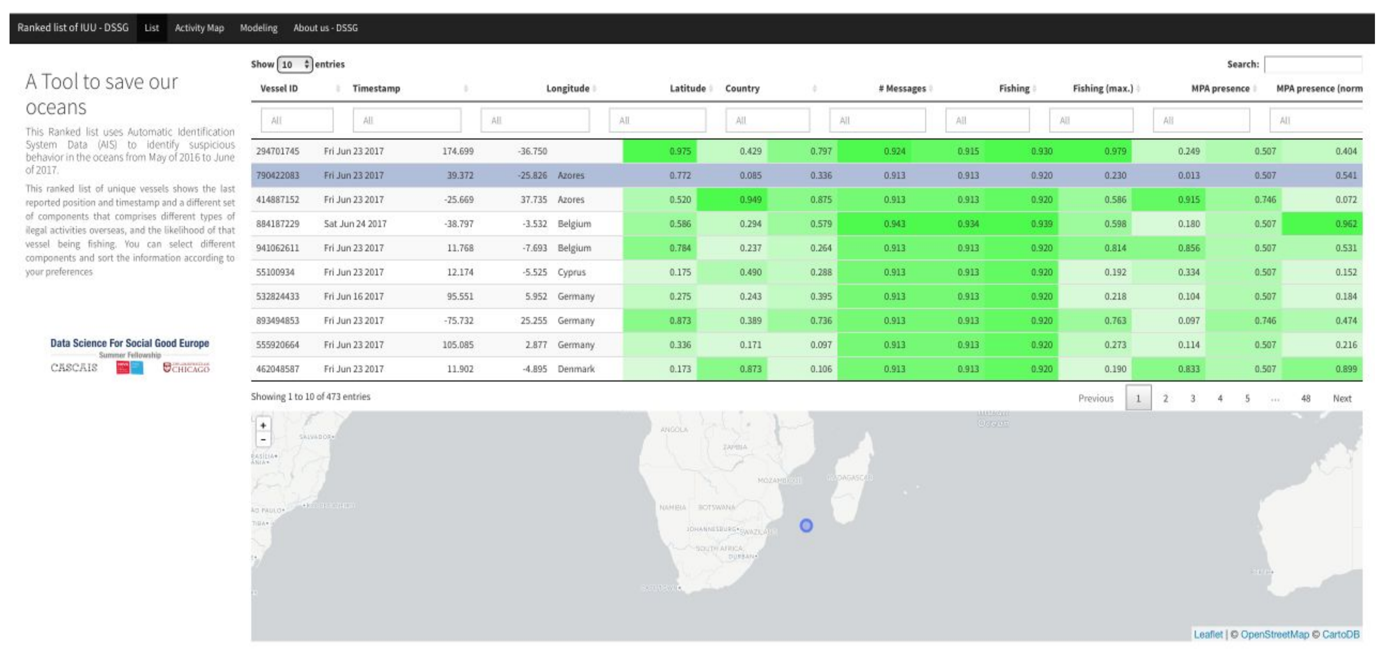

Figure 6: application to display for each vessel in the table the vessel components, and a unified vessel risk index. Each column can be sorted depending on the user's interest. For a selected vessel the most recent location is displayed on the world map upon selection.

This work was done during Cascais Data Science for Social Good Europe Fellowship at Nova School of Business and Economics.