

Capstone Progress Report

Last updated: February 13, 2024

**Instructions**: Each capstone team can use this template to capture and summarize progress. This can be shared with the sponsor and mentor. When submitting the report during the course, a PDF file is preferable.

**Stakeholder Names and Roles**

*In the table below, enter information for each team member, mentor, and sponsor. On the sponsor side, there may be several stakeholders including project manager, data contact, etc.*

|  |  |
| --- | --- |
| **Stakeholder** | **Role** |
| *Danielle Katz* | *Team member* |
| *Dana Korotovskikh* | *Team member* |
| *Stephen Kullman* | *Team member* |
| *Samuel Brown* | *Team member* |
| *Tara Valladares* | *Sponsor* |
| *Rebecca Disipio* | *Sponsor* |
| *Heman Shakeri* | *Faculty Mentor* |

**Project Title:** Detection of Illegal, Unreported, and Unregulated Fishing using AIS Data and Unsupervised Learning

**Abstract**

Illegal, unreported, and unregulated (IUU) fishing is a critical issue impacting the health of the marine ecosystem and global security. It is linked to crimes such as human trafficking, and destroys local economies through overfishing, thereby collapsing many local species populations. A major contributor to this is deceptive use of Automatic Identification Systems (AIS) beacons on fishing nets, buoys, or lines, to conceal illegal fishing hauls from local vessels and enforcement authorities. Enhancing the capacity to recognize fraudulent AIS devices, and predict their operational areas of activity, is crucial for regulatory agencies aiming to reform the current state of fishing practices. This paper focuses on analyzing geospatial AIS data, and providing a means to improve the detection of IUU fishing occurrences via AIS devices. We propose a model following machine learning techniques to identify irregular AIS beacons based on their logged activity. This approach processes AIS communication data and explores their relationship with IUU fishing activity. We then apply spatial clustering techniques to discern and highlight common regions of IUU fishing, focusing on fishing nets and buoys. With unlabeled input data, we employ unsupervised machine learning to detect patterns of IUU fishing via AIS characteristics and geospatial elements.

Keywords: IUU Fishing, Geospatial Data, Unsupervised Machine Learning, Clustering

**Outline of the Project**

*This should discuss:*

*- the business purpose of the project*  
*- why the project is important*  
*- who are the stakeholders*  
*- important assumptions*  
*- what is in scope / out of scope (as needed)*

*Example: Legal assistants spend a lot of time reviewing documents related to court cases. This project will use an ML system that will take legal documents as input and output a summary and helpful metadata. This will save legal assistants time, reduce burnout, and allow them to focus on helping their clients. The assumption is that all input documents are legal documents.*

IUU fishing is estimated to cost 10-23 billion dollars per year, which means that this isn’t just an environmental issue but an economic one as well. This project will use a machine learning algorithm to output a probability or classification that an AIS transponder is attached to an illegal fishing net. Having this knowledge will give governments and watchdog organizations a more targeted way to crack down on these illegal activities, aiding the environment and the local economy. For this project, we will focus on just Southeast Asia, specifically two regions near China. The stakeholders are GA-CCRI and the University of Virginia School of Data Science.

**Success Criteria**

*Success criteria define the needs for a successful project. These should be identified with help from the sponsor. It is important to revisit these criteria with the sponsor as the project progresses to ensure alignment.*

|  |  |
| --- | --- |
| *SC1* | *A model that assesses AIS transmissions and discerns true vessels and suspected fraudulent devices based on activity over time* |
| *SC2* | *An analysis of detected areas of heightened illegal AIS activity, suggestive of hotspots* |

**Data Assumptions and Limitations**  
  
*- Discuss aspects which are presenting difficulty: missing data, bad data, etc.*  
*- Discuss and justify data assumptions*  
*- If you are having difficulty accessing data, please discuss*  
*- If you are having difficulty with compute or storage, please discuss*

|  |  |
| --- | --- |
| **Identifier** | **Description** |
| A | *While gaps in AIS transmission signals can suggest a net or buoy (null values), it is important to acknowledge that this may not always hold true. There exist unique AIS transmissions emanating from specific vessels that may falsely appear as buoys in their characteristics.* |
| A | *During the clustering process, it is important to have caution in grouping all the signals into a single category, such as buoy or net. Vessels may be near hotspots and should be recognized distinctly.* |
| L | *There is no accessibility to a comprehensive data dictionary or labelling for this AIS dataset.* |
| L | *The dataset is exclusive to a subset of AIS transmissions from limited East-Asia and Pacific regions. It is important to acknowledge that our dataset, and the derived algorithm, may not be applicable to all global regions.* |
| L | *Difficulty finding methods to ingest and efficiently process the size of the dataset.* |

**Summary of Data Processing, Data Aggregation**

*This should discuss:*  
*- The target variable*  
*- The raw data*  
*- Potential predictor variables  
- Methods and justification for data aggregation*

**Data Visualizations**

*Include a small sample of relevant data visualizations / summaries for illustration*

**Summary of Modeling and Analysis**

*This should discuss:**- Justification and summary of models considered  
- Specialized techniques that may be used, such as NLP  
- Metrics that are important (e.g., F1 score / recall / precision), and explanation of why they are important.*

**Future Work Plan**

*- Discuss your work plan for the remainder of the term. This should include all important tasks.  
- Provide a timeline if possible*

**Potential Concerns [C] and Blockers [B]**

*Concerns* are things to keep in mind but might not yet impede progress.   
*Blockers i*ndicate that you are stuck and need help.  
 *Based on what you know about the project and data, are there things that might present a challenge in moving forward?*

*Are there things you can do to clear these challenges? For example, are there resources you might need? What are they?*

|  |  |
| --- | --- |
| **Identifier** | **Description** |
|  |  |