George Mason University – SEOR Department

OR/SYST 699 SEOR MS Capstone Project – Spring 2017

Data Analytics to Detect Illegal Fishing

Project Proposal

Jarred Byrnes

Jonathan Gessert

Edward Kerrigan

Jonathan Matteson

Table of Contents

1 INTRODUCTION

Illegal fishing, typically referred to a more inclusive term, IUU (Illegal, Unreported, or Unregulated) fishing, is a global problem. Legal fishing alone depletes the fish populations to unsustainable levels; illegal fishing pushes many fish populations to the brink of extinction. With such a high global demand, there are attractive incentives for fisherman to go outside the law and fish for out of season fish or fish more than they are allowed. Along with very little global regulation black market fish easily spread into the general stock making them almost impossible to track once they are off the boat. To put the problem in perspective Black market fishing is estimated to account for 11 million - 26 million tonnes of fish equal to 14 or 33 per cent respectively of the world's total legal catch (fish and other marine fauna) in 2011. In the same year, legal fishing accounted for 78.9 million tonnes of fish. With such a potentially large subset of the global catch coming from IUU fishing being able to track and prevent it is a top priority (World Ocean View).

With the advent of big data and new tracking system systems becoming available Lockheed Martin sees an opportunity to start becoming proactive on IUU fishing. Previously to find an illegal fishing vessel the Coast Guard or another entity would have to physically patrol the space. With such limited resources and such a wide area to search these tactics have been largely ineffective. Several large companies including Google have formed the Global Fishing Watch (GFW) to utilize this new tracking data to start to solve this problem. With the increased global focus and availability of data a real dent can start to be made in IUU fishing (Cheeseman).

2 PROBLEM STATEMENT

Detecting IUU fishing is often a law enforcement human resource intensive task limited by funding. Newly available sources of data, along with models currently being developed have the potential to make a major difference in detecting illegal fishing activity by reducing and/or refocusing human resources on more productive investigations.

2.1 PROBLEM DESCRIPTION

Currently there are very few models or analytics that exist for detecting illegal fishing without the physical search and seizure by law enforcement human resources. Modeling both fishing behavior and the illegal fishing enterprise will expose the data necessary to model and predict potential illegal fishing activity to focus law enforcement human resources to physically searching areas with a higher probability of detecting illegal activity. Thankfully major companies like Google are taking an interest into the problem. By using shipping vessels on board trackers, a methodology has started to be derived for describing typical fishing behavior using data from legal fishers. Multiple institutions are collecting and refining this data as well as developing algorithms to detect illegal activities. Leveraging their work as a starting point and fusing these and other data sources an effective model can be developed to identify IUU fishing.

Even with the available data trying to solve this problem globally is a very difficult task. It will be important to scope the project to a specific region, potentially targeting only certain fish populations to make the models more meaningful.

2.2 BOUNDS AND ASSUMPTIONS

With so many fishing vessels and fishing areas across the globe it will be important to focus on specific regions to make both the problem and data manageable. The area of focus for identifying IUU fishing will include the West coast fishing areas within the United States Exclusive Economic Zone (EEZ) and Marine Protected Areas. This allows the focus to be on only U.S fishing laws and minimize the area while still making the model meaningful. Only boats large enough to carry transponders and GPS equipment will be tracked as to provide enough freely available data for tracking. The targeted behaviors that will be potentially identified by the model include overfishing as well as fishing without a permit.

IUU fishing presents a potentially large amount of data, so restrictions will need to be put in place so the team can process it. Since the team does not have access to any major computing resources, datasets will have to be small enough that a typical laptop computer can handle the calculations. Spiral 1 training and initial test data will come from archived data and final test datasets will come from near-real time data sources to evaluate the model using data in the West coast US EEZ.

2.3 STAKEHOLDERS

The primary stakeholder for our project is our project sponsor Lockheed Martin. However, any branch of law enforcement, such as the Coast Guard and local law enforcement will benefit from being able to detect criminal shipping vessels. State, federal, and even foreign governments will also benefit from a better posture on being able to detect and predict illegal fishing.

Agencies involved with environmental and fishery protections such as the Game & Fisheries Commissions, National Marine Fisheries Service, and the NOAA will benefit from a more sustainable level of fishing. Legal fisherman stand to gain from eliminating illegal competition and the illegal fisherman would also be effected from increased detection.

3 SCOPE

This project will develop and refine a model or series of models to progress towards an all-encompassing illegal fishing enterprise model used for predicting illegal fishing. The project will define use cases, identify publically available data sources (prioritizing geospatially referenced, physics-based sensor data) , develop models for analyzing data, exercise models to identify vessels with patterns of interest, and, ultimately, execute the data analytics lifecycle for statistics and machine learning to provide the fishing scoring heuristics and evaluation of their effectiveness, as well as development of supervised machine learning algorithms. This project will strive to characterize what "normal" fishing behavior looks like, then use this to target behaviors indicative of suspect behavior and develop models to fuse geospatially referenced, physics-based sensor data to identify illegal fishing activity.

4 PRELIMINARY REQUIREMENTS

The following is a list of preliminary requirements derived from the sponsor's deliverables and project expectations.

Requirement #1: The team shall determine necessary datasets required for model training and testing.

Requirement #2: The team shall determine and document the set of behaviors and patterns that signal normal fishing activity, which can be evaluated on ships in Marine Protected Areas (MPAs) to predict whether illegal activity is taking place.

Requirement #3: The team shall define required predictors to evaluate models and perform analytics.

Requirement #4: The team shall deliver analytic capabilities design that parameterizes the key signatures and behaviors about the suspect vessels (e.g. size of vessel, speed of vessel, spatial proximity, etc..) ; such that the solution can be flexibly adapted to other similar illegal fishing scenarios (e.g. different species in a different part of the world) or other maritime tracking activities.

Requirement #5: The team shall deliver a defined process of how to conduct statistical analysis of fishing behavior.

Requirement #6: The team shall conduct model creation, validation trials, and model re-evaluation in an iterative process.

Requirement #7: The team shall document all model rules used in analysis.

Requirement #8: The team shall provide descriptive visualization and insights into how the data provides answers as to which vessels are predicted to be conducting illegal fishing.

Requirement #9: The team shall provide guidance on how to utilize data output from the model.

Requirement #10: The team shall use a Model-Based Systems Engineering (MBSE) approach using Systems Modeling Language (SYSML) to document the illegal fishing architecture.

Requirement #11: The team shall identify and document potentially useful data sources (prioritizing geospatially referenced, physics-based sensor data) for future development.

Requirement #12: The team shall generate a report and briefing describing models created, datasets gathered and results of test trials executed.

5 TECHNICAL APPROACH

The overall technical approach is to conduct parallel efforts in systems engineering and data analytics. The system engineering approach will be to document the architecture using model based systems engineering, tailored to describing the potential behaviors and activities of illegal fishing. The data analytics technical approach will use the spiral systems engineering lifecycle with each spiral being guided by the data analytics lifecycle for statistics and machine learning which includes the subtasks shown in Figure 1.

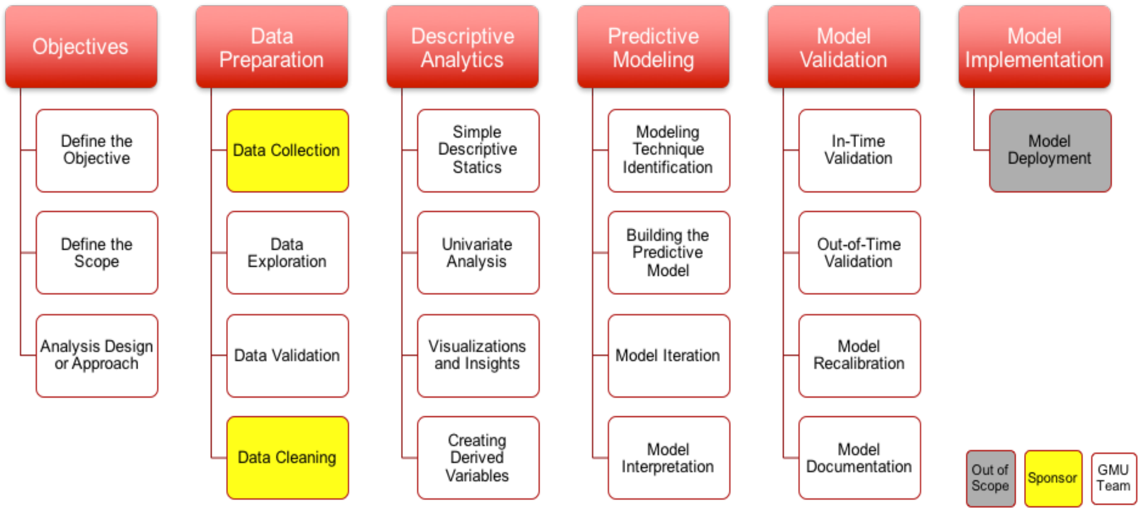

Figure 1: Data Analytics Process

5.1 SYSTEMS ENGINEERING

The systems engineering approach will start by developing top-level legal and illegal fishing use case diagrams. As it is expected that illegal fishers must perform a subset of the legal fishing tasks, a legal fishing use case will be developed to support the development of the illegal fishing use case. A top level illegal fishing use case will be developed and refined with input from the Sponsor. The prioritization of which use cases to decompose and level of decomposition will be determined during project execution with input from the Sponsor. This further decomposition into lower-level, finer detail behavior diagrams (activity, sequence, and state) should expose behaviors (potential activities) and requirements (required activities) for an illegal fishing activity from start to finish. IDEF-1X and Entity Relationship data models will be generated to support the setting of the objectives for Data Analytics Spiral 2 and to document data aggregation and models developed in Data Analytics Spiral 1 and Spiral 2.

5.2 DATA ANALYTICS

The data analytics approach will use a spiral development model with two planned spirals during project execution. Data Analytics Spiral 1 will develop a model to detect fishing vessel behavior in AIS data to support the generation of simulated illegal fishing data or identification of illegal fishing from non-AIS data sources. For Spiral 1, the team plans to evaluate AIS data hand labeled with fishing behavior provided by Kristina Boerder and explore four existing models authored by Timothy Hochberg and Egil Möller described as their Heuristic Model, Generic Model, Multi-Window, and Multi-Window Gear-Types-Specific Model from existing GitHub repositories relating to the Global Fishing Watch project. The types of fishing in the data set include: purse seine, trawling, longlining, and longlining crowd. The top risks associated with Spiral 1 are shown in Table 1.

Name	Description	Mitigation	Response
Computational Time	If computations cannot be performed using student laptops, then additional computing sources will be required	Evaluate computation time as soon as possible to allow for obtaining additional computing resources.	Seek additional computing resources from the sponsor or GMU. Reduce the number of models being explored.
Model Recreation	If the documentation for the four models are not able to be recreated and understood, then the model may not be able to be recreated.	Plan time for researching models and documentation.	Reduce the number of models to be completed during Spiral 1. Extend time for Spiral 1 and reduce or eliminate Spiral 2
Technical Gap	If the use of tools, data types, and Python packages unfamiliar to the team takes more time to master, then Spiral 1 may not be able to be completed as scheduled.	Allow time for tool familiarization.	Extend time for Spiral 1 and reduce or eliminate Spiral 2

Table 1: Spiral 1 Risks

To allow for exploration of the illegal fishing problem and Sponsor flexibility in the objectives of Spiral 2, the objectives for Spiral 2 will be determined during project execution with agreement from the Sponsor.

Some potential model expansion ideas are as follows:

1. Use normal fishing vessel behavior to identify illegal fishing on vessels not indicating fishing in their AIS data feed.
2. Simulate a fishing vessel voyage that is spoofed, and see if fishing model can detect any discrepancy.

3. Develop a model based upon AIS data broadcast. Areas of interest include:
    a. Compare new position with position propagated from previous AIS data broadcast raising an alert if data is inconsistent.
    b. Determine the time of the next AIS data broadcast and generate an alert if the next AIS data broadcast is missing or if the timing is out of AIS specification.
    c. Propagate current AIS data to predict time of arrival to EEZ or MPA.
    d. Correlate AIS data with other data sources to determine if vessels spoofing AIS broadcast.
    e. Determine the probability that vessels are exchanging fish that were illegally obtained by generating an alert when two vessels locations are abnormally close together.

5.3 PROJECT TOOLS

To perform project activities, we plan to use MagicDraw for systems engineering, Git for version control, Python 3 and Python packages which will potentially include the Jupyter library for documentation, and common statistical and machine learning libraries such as numpy, scipy, scikit-learn, sklearn, and/or TensorFlow.

6 EXPECTED RESULTS

The goal of this project is to provide the client with an architecture describing the illegal fishing enterprise that can be used to focus further research, as well as validated models which can be extended or used to evaluate data to predict illegal fishing and/or related nefarious vessel behaviors. Specific deliverables are summarized in Table 2: Deliverables.

Deliverable	Description
1. Illegal Fishing Architecture	Top level Legal and Illegal Fishing Use Case Diagrams. Decomposed Illegal Fishing Use Case Diagrams.
2. Data Models	Data models (IDEF-1X, Entity Relationship) documenting Spiral 1 and Spiral 2. Data models identifying potential sources and documenting the fusion necessary for future efforts.
3. Analytic Models	Pattern detection Models, Analytics and Algorithms (models, analytics and algorithms used to associate data to Models) Required predictors and model evaluation.

	Model rules. Model analysis. Descriptive visualization and insights Guidance on model output usage
4. Data	Data used in models and documentation. Actual data with data source identified. Data Feed pre-conditioning mappings (mappings define the cleaning, and correlation of the data used to evaluate models and perform analytics) Potential data sources for future efforts.
5. Data Analytics Environment	Documentation to support the recreation of the environment used for data analytics.
6. Final Report	Final report documenting project assumptions activities, architecture, data analytics, lessons learned, and suggestions for future efforts.
7. Final Presentation	Presentation highlighting the key activities, architecture, data analytics, lessons learned, and suggestions for future efforts.

Table 2: Deliverables

7 PROJECT PLAN

The Work Breakdown Structure (WBS) is described in Table 1: WBS Dictionary. The project schedule showing the key tasks and their relationships is shown in Figure 2: Project Schedule.

8 RESOURCES

World Ocean View. (n.d.). Illegal Fishing. Retrieved February 10, 2017, from http://worldoceanreview.com/en/wor-2/fisheries/illegal-fishing/

Cheeseman, G. (2015, February 07). Impacts of Worldwide Illegal Fishing. Retrieved February 10, 2017, from http://www.triplepundit.com/2013/05/illegal-fishing-big-problem-all-world/

https://www.researchgate.net/publication/304711836_Improving_Fishing_Pattern_Detection_from_Satellite_AIS_Using_Data_Mining_and_Machine_Learning

https://wildtech.mongabay.com/2016/03/watching-for-illegal-fishing-by-keeping-our-eyes-on-the-seas/

WBS #	WBS Name	Description
1	Project Management	Project Management Tasks
1.1	Project Initiation	Tasks from Project start to Project Proposal
1.2	Reviews	Required Course Reviews
1.3	Sponsor Synchronization Meeting	Weekly meetings to synchronize project work with Sponsor expectations
2	Research	Research tasks
2.1	Illegal Fishing	Research on Illegal Fishing
2.2	Tools Research	Research on Tools for modeling
3	Model Development	Model development tasks
3.1	Systems Engineering	Systems Engineering Tasks no covered directly in Spiral tasks
3.1.1	Develop Illegal Fishing Use Case	Develop top level Use Case for Illegal Fishing
3.1.2	Develop Illegal Fishing Behavior Diagrams	Decompose top level Use Cases to finer granularity; Sponsor to support prioritization
3.2	Spiral 1 - Fishing Identification	Spiral 1 Tasks – Preliminary objective is to identify fishing activity with AIS data
3.2.1	Spiral 1 - Objectives	Define the objectives for Spiral 1
3.2.2	Spiral 1 - Data Preparation	Prepare data for Spiral 1
3.2.3	Spiral 1 – Develop Data Model	Develop data models related to this Spiral
3.2.4	Spiral 1 - Descriptive Analysis	Perform descriptive Analysis of data for Spiral 1
3.2.5	Spiral 1 - Predictive Modeling	Develop predictive model for Spiral 1
3.2.6	Spiral 1 - Refine Model	Refine Spiral 1
3.2.7	Spiral 1 - Model Validation	Validation and Documentation for Spiral 1
3.3	Spiral 2 - TBD	Spiral 2 Tasks – Objective is TBD with Sponsor during project execution; Tasks consistent with WBS 3.2 Spiral 1
4	Deliverables	Tasks associated with generation of deliverables
4.1	High Level Architecture	Architecture and related diagrams
4.2	Library of Data Sets	Data sets used for data analytics; Required data sets; Training sets; Test sets; Evaluation sets
4.3	Library of Algorithms	Algorithms and tool sets required to execute models
4.4	Final Presentation	Preparation of Final Presentation
4.5	Final Report	Preparation of Final Report
4.6	Web Site	Preparation of Website
4.7	Final Presentations	Delivery of Final Presentation
4.8	Models	Models developed during the project execution

Table 3: WBS Dictionary

WBS	Task Name	Duration	Start	Finish	Predecessors	Resource Names	Add New Column
1	**Project Management**	**94 days**	**Sun 1/29/17**	**Tue 5/2/17**			
1.1	**Project Initiation**	**19 days**	**Sun 1/29/17**	**Thu 2/16/17**			
1.2	**Reviews**	**49 days**	**Fri 3/3/17**	**Thu 4/20/17**			
1.3	**Sponsor Synchronization Meeting**	**92 days**	**Tue 1/31/17**	**Tue 5/2/17**			
2	**Research**	**61 days**	**Sun 1/29/17**	**Thu 3/30/17**			
2.1	Illegal Fishing	61 days	Sun 1/29/17	Thu 3/30/17			
2.2	Tools Research	11 days	Fri 2/17/17	Mon 2/27/17	13		
3	**Model Development**	**57 days**	**Fri 2/17/17**	**Fri 4/14/17**			
3.1	**Systems Engineering**	**47 days**	**Fri 2/17/17**	**Tue 4/4/17**			
3.1.1	Develop Illegal Fishing Use Case	15 days	Fri 2/17/17	Fri 3/3/17	13		
3.1.2	Develop Illegal Fishing Behavior Diagrams	40 days	Fri 2/24/17	Tue 4/4/17	39SS+7 days		
3.2	**Spiral 1 - Fishing Identification**	**38 days**	**Fri 2/17/17**	**Sun 3/26/17**			
3.2.1	Spiral 1 - Objectives	3 days	Fri 2/17/17	Sun 2/19/17	13		
3.2.2	Spiral 1 - Data Preparation	3 days	Fri 2/17/17	Sun 2/19/17	13		
3.2.3	Spiral 1 - Develop Data Model	8 days	Fri 2/17/17	Fri 2/24/17	13		
3.2.4	Spiral 1 - Descriptive Analysis	5 days	Mon 2/20/17	Fri 2/24/17	43		
3.2.5	Spiral 1 - Predictive Modeling Phase I	8 days	Sat 2/25/17	Sat 3/4/17	45		
3.2.6	Spiral 1 - Predictive Modeling Phase II	8 days	Sun 3/12/17	Sun 3/19/17	51FS+2 days		
3.2.7	Spiral 1 - Model Validation	7 days	Mon 3/20/17	Sun 3/26/17	47		
3.2.8	Spiral 1 - Complete	0 days	Sun 3/26/17	Sun 3/26/17	48		
3.3	**Spiral 2 - TBD**	**41 days**	**Sun 3/5/17**	**Fri 4/14/17**			
3.3.1	Spiral 2 - Objectives	5 days	Sun 3/5/17	Thu 3/9/17	46		
3.3.2	Spiral 2 - Data Preparation	5 days	Fri 3/10/17	Tue 3/14/17	51		
3.3.3	Spiral 2 - Develop Data Model	8 days	Fri 3/10/17	Fri 3/17/17	51		
3.3.4	Spiral 2 - Descriptive Analysis	5 days	Mon 3/27/17	Fri 3/31/17	48		
3.3.5	Spiral 2 - Predictive Modeling	7 days	Sat 4/1/17	Fri 4/7/17	54		
3.3.6	Spiral 2 - Model Validation	7 days	Sat 4/8/17	Fri 4/14/17	55		
3.3.7	Spiral 2 - Complete	0 days	Fri 4/14/17	Fri 4/14/17	56		
3.4	Model Development Complete	0 days	Fri 4/14/17	Fri 4/14/17	57		
4	**Deliverables**	**47 days**	**Mon 3/27/17**	**Fri 5/12/17**			
4.1	High Level Architecture	7 days	Sat 4/15/17	Fri 4/21/17	40,49,57		
4.2	Library of Data Sets	7 days	Sat 4/15/17	Fri 4/21/17	49,57		
4.3	Library of Algorithms	7 days	Sat 4/15/17	Fri 4/21/17	49,57		
4.4	**Final Presentation**	**47 days**	**Mon 3/27/17**	**Fri 5/12/17**			
4.5	**Final Report**	**42 days**	**Mon 3/27/17**	**Sun 5/7/17**			
4.6	**Web Site**	**7 days**	**Mon 5/1/17**	**Sun 5/7/17**			
4.7	Final Presentations	0 days	Fri 5/12/17	Fri 5/12/17			
4.8	Models Developed	7 days	Sat 4/15/17	Fri 4/21/17	49,57		
4.9	Deliverables Complete	0 days	Fri 5/12/17	Fri 5/12/17	74,84,86,87		

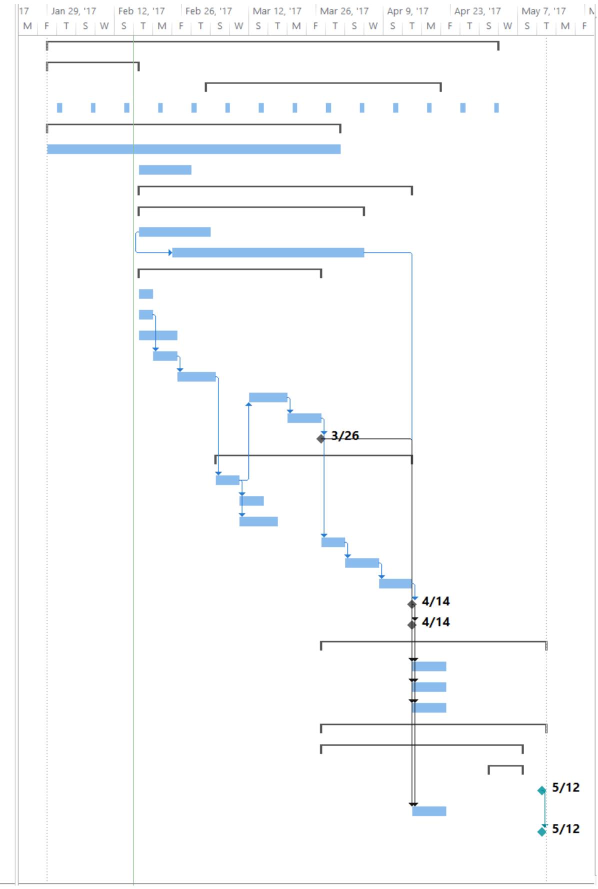

Figure 2: Project Schedule