George Mason University – SEOR Department

OR/SYST 699 SEOR MS Capstone Project – Spring 2017

IUU Fishing Detection Final Report

Jarred Byrnes

Jonathan Gessert

Edward Kerrigan

Jonathan Matteson

Table of Contents

1.0 EXECUTIVE SUMMARY

Illegal, Unreported, or Unregulated (IUU) fishing was estimated to account for nearly 30% of the total catch of global fishing in 2010, costing coastal economies billions and depleting fish populations at unsustainable rates. Few countermeasures exist to track down IUU fishing vessels besides physical searches of areas of interest which costs significant resources. Leveraging systems modeling and data analytics skills acquired through the George Mason University Systems Engineering and Operations Research graduate program, Lockheed Martin sponsored a capstone project which resulted in the foundational research and development of an extensible IUU Fishing Detection Architecture, encompassing signal intelligence and modern data analytics techniques to predict vessels performing IUU fishing.

The IUU Fishing Detection Architecture created value through the documentation of use case, activity, and state machine diagrams. Identifying data stores and processes drove the initial development of a feasible IUU fishing detection system that scores a vessel's likelihood of IUU fishing activity by combining (1) a vessels likelihood of fishing using geospatially referenced signal data, (2) the classification of whether or not a vessel is inside an area of interest, and (3) the classification of whether or not a vessel is allowed to be in their inhabited area of interest.

At project completion, portions of the system were prototyped including the process of a highly predictive *fishing or not fishing* classification model for longliner and trawler fishing vessels using logistic regression, and the process of identifying if a vessel is inside an area of interest. Additionally, a large amount fishing vessel registries were identified which define specific vessel fishing rights to fish in a regulated area of interest.

The accuracy the fishing model could predict the probability of a vessel fishing when it has longliner or trawler gear was acceptable, and vessels with purse seine gear could be predicted but will require additional research and analysis. The *In an Area of Interest* classification model should be expanded to score the likelihood that they will enter an area of interest rather than outputting a true/false determination. The data source agnostic and extensible model was designed, so Lockheed Martin can take the IUU framework, models, and associated research and recommendations, along with utilizing additional funding and resources, and further its efforts to predict IUU fishing to enable enforcement and ultimately significantly reduce or prevent IUU fishing.

2.0 INTRODUCTION

2.1 BACKGROUND

The ability to leverage cognitive computing techniques as a way to unburden human analysts is a key problem for several Lockheed Martin (LM) customers including the U.S. military and the Department of Homeland Security. As a result, LM is motivated to sponsor and assist specific areas of academic research that are focused on advancing the field of data analytics, and decision optimization systems. Today's use of satellite and aerial reconnaissance systems to remotely gather geospatial information has proven beneficial to detecting illegal and nefarious activities such as poaching natural resources, human trafficking, and arms smuggling. However, a distinct disadvantage is that the volume and variety of data to be analyzed in order to produce the evidential data needed to support the arrest and prosecution of criminals frequently exceeds human abilities. As a result, LM in collaboration with George Mason University (GMU) identified the detection and identification of Illegal, Unreported, or Unregulated (IUU) fishing activities as an ideal ‘pattern of life' candidate to use as the subject for advanced research in the area of cognitive computing and decisions support systems.

Illegal fishing, typically referred to as the more inclusive term, IUU (Illegal, Unreported, or Unregulated) fishing, is a global problem. With such a high global demand for fish, there are attractive financial incentives for fisherman to perform IUU fishing. These incentives include, but are not limited to, fishing for out-of-season fish, fishing in regions which they are not licensed to fish, or catching more fish than national or regional regulations allow. IUU fishing depletes the fish populations to unsustainable levels and pushes many fish populations to the brink of extinction.

To put the problem in perspective, in 2011, black market fishing is estimated to account for 11 to 26 million metric tons of fish equal to 14 to 33 percent respectively of the world's total legal catch (fish and other marine fauna). In the same year, legal fishing accounted for 78.9 million metric tons 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 [1].

Currently, the U.S. Coast Guard (USCG) and other international entities' current solution to find a vessel conducting IUU fishing is to physically patrol the area of interest. For example, to counteract the threat of foreign encroachment, the USCG patrols the USA's Exclusive Economic Zone (EEZ) with long-range surveillance aircraft, cutters and patrol boats. The foreign fishing activity on the Russian side of the U.S./Russia Maritime Boundary has become of increasing concern. In recent years, the USCG has resorted to near-daily C-130 flights and continuous cutter presence along the boundary line during peak fishing seasons to ensure that the huge foreign fleets, including Russian, Japanese, Polish, Chinese and Taiwanese fishing vessels operating near the line do not violate the U.S. EEZ [2].

With the increased global focus and availability of data, IUU fishing countermeasures can be improved [3]. Several large companies, including Google and SkyTruth, have formed the Global Fishing Watch (GFW) to utilize geospatially referenced, physics-based sensors such as Automatic Identification System (AIS) data to assist with IUU fishing countermeasures. Using geospatial intelligence remote sensing, signals intelligence, and data analytics techniques, Lockheed Martin sees an opportunity to provide solutions to support global IUU fishing detection and enforcement.

2.2 WHY THIS PROJECT IS NEEDED

There is a need for detecting IUU fishing by stakeholders who seek to enforce fishing regulations or prevent the environmentaly unsustainable effects related to illegal fishing and overfishing. Solutions which would more effectively control IUU fishing in addition to reducing expenditures in resources such as time and fuel during these expeditions would be welcomed by law enforcement and the anti-overfishing community. These stakeholders include, but are not limited to, the following:

- Sponsor: Lockheed Martin
- Various Local Law Enforcement (U.S. and International)
- Game & Fisheries Commissions
- Homeland Defense/Maritime Border Security Organizations
- National Marine Fisheries Service
- U.S. National Oceanic and Atmospheric Administration (NOAA)

- Legal Fishing organizations
- Illegal, Unreported, Unregulated Fishing vessels
- Non-Governmental Organizations dedicated to oceanic environmental protection

Predictive data analytic techniques have been recognized as a potential solution to make a major difference in detecting illegal fishing activity. Data analytics models are currently being developed which will provide intelligence for agencies to reduce the amount of resources required to investigate and catch vessels participating in illegal fishing activity.

2.3 PROBLEM STATEMENT

IUU fishing is harming national interests, is environmentally unsustainable, and is expensive to regulate. Furthermore, investigating IUU fishing with only physical patrolling, search, and seizure by law enforcement human resources is an expensive and time-intensive process. Utilizing predictive intelligence on which areas or vessels are more likely to conduct IUU fishing can provide significant benefits to the process. By International convention, vessels travelling internationally over 300 gross tonnage are required to be equipped with AIS systems. AIS data consists of geospatially referenced data used for relaying crucial information such location, heading, speed, and activity that is used as a surrogate for radar to prevent vessel collisions in the open seas. AIS data and predictive analytics can be used to model vessel fishing behavior, and predict the likelihood of fishing for any tracked vessel inside a regulated area of interest. It will also be possible to determine if the vessel is authorized to be fishing inside a regulated area of interest. Improving IUU fishing detection with a predictive model derived from AIS data has the potential to reduce IUU fishing countermeasure expenses while increasing the likelihood of apprehending IUU fishers.

2.4 OBJECTIVES, SCOPE & DELIVERABLES

The scope, objectives and deliverables of this project were iteratively refined during weekly collaboration with the sponsor, Lockheed Martin, but is ultimately described as the following: to develop and deliver an IUU fishing detection architecture, along with identified data sources which could be used for a predictive analytics model to determine IUU fishing behavior. The IUU fishing detection architecture was to be modeled to explore avenues that may be exploited by vessels conducting IUU fishing, and to be extensible for use in post project maritime domain awareness efforts.

A data repository maintained by Global Fishing Watch was identified, which documented success using fishing-gear specific vessel data observed over several time windows and modeled using logistic regression models. However, the repository contained little to no documentation on how to recreate their described model using their maintained data and code. After these highly-relevant repositories were identified, an objective was to explore, understand, and maximize the derived value of Global Fishing Watch's existing repositories. The scope of work was refined to include exploring their repository and document the recreation of their gear-specific, multiple time window, and logistical regression model using the AIS data included in this repository.

The scope of the project was to use geospatially referenced, physics-based sensor data to research, develop, and refine an approach to IUU fishing detection, while simultaneously developing a series of descriptive and predictive IUU fishing detection models. The scope of IUU fishing detection was limited to Marine Protected Areas (MPAs). The project objectives included defining use cases, identifying publically available data sources, developing models for analyzing data, and creating models to identify vessels with patterns of interest, by executing the data analytics lifecycle to provide a fishing scoring model. The expected deliverables at the end of the project were the following:

- IUU fishing detection architecture
    - High-level use case detailing the current fishing industry
    - Use Case diagram of what is considered "Normal" fishing behavior
    - Activity diagrams further dissecting specific illegal activities of interest
    - Data models for each spiral documenting the flow of data from its origin to the model algorithms
- Set of algorithms to analyze the data sets using Bash and Python libraries
    - Descriptive analytics code
    - Predictive analytics code
    - Documentation on how to recreate data analytics environment including

the library of all datasets used during project execution

Through collaborating with the sponsor, the following items were mutually agreed upon to be out of scope and would be deferred for future efforts:

- Ingesting live data into the model
- Modeling all aspects of IUU fishing
- Model implementation

3.0 TECHNICAL APPROACH

The technical approach includes research and generating systems engineering diagrams to explore and document the IUU fishing detection architecture, while supporting a data analytics lifecycle which included data preparation, descriptive analytics, predictive modeling and model validation. The Sponsor and the GMU team had limited background in IUU fishing so the first step of the process involved performing research to become experts. During the research phase, data that could potentially be used for data analytics was identified. Next, the data analytics team focused on a model for predicting fishing while the systems engineering team focused on a hierarchical approach to modelling IUU fishing. Initial IUU modeling focused on the high level maritime vessel operations to model fishing activities, followed by modeling IUU fishing activities that were focused on key areas as guided by the sponsor. The systems engineering team focused their efforts on documenting the architecture using model based systems engineering, tailored to describing the potential behaviors and activities of illegal fishing. Due to readily available data and time constraints, the data analytics was limited to focusing on a subset of the IUU architecture framework. The specific systems engineering and data analytics approaches are detailed in their respective sections below.

3.1 ASSUMPTIONS

- Not all vessels turn off their AIS transmission when conducting IUU fishing
- All vessels fish in similar tracks when fishing based on their gear type, regardless of legality or geographic location
- 2014 MPA shapefile sufficiently defines current MPA boundaries
- Kristina Boerder's data based on her research is reliable; was verified during

project execution

- Python packages used such as SciKit-Learn accurately perform their intended functions to an acceptable standard

3.2 LIMITATIONS

- Limited time to accomplish tasks
- Access to affordable satellite imagery
- Access to IUU subject matter experts
- Global Fishing Watch repositories data was limited to data hand labeled by Kristina Boerder because the integrity of the other data could not be validated
- Understanding of the repository code
- Missing datasets referred to in the repository
- New software libraries/packages which had never been used before by team members

3.3 RESEARCH

Research and architected model areas included:

- Prior predictive analytics projects using AIS data
    - Global Fishing Watch Repositories
    - Kristina Boerder's *Improving Fishing Pattern Detection from Satellite AIS Using Data Mining and Machine Learning* [4]
- Data Analytics Software Stack (in appendix)
- AIS Functionality and existing Data Sources
- Vessel Registries
- Regulations from Licensing Authorities in US Western Coast Pacific: EEZ, MPA, Regional Fisheries Management Organization (RFMO)
- What are visible and observable indicators of IUU fishing activities
- How illegal fishing is determined by law enforcement
- Fishing patterns by gear type: Longliner, Trawler, Purse Seine

4.0 SYSTEM MODELS AND ARCHITECTURE

4.1 IUU FISHING ARCHITECTURE

Developing the IUU fishing architecture focused on the creation of logical processes that the analytical model must follow. To accomplish this, general use cases were created to determine what the scope of the model would be, what actors would be involved, and what processes would need to be identified for each. The use cases would be used to create and eventually to validate the efficacy of the model.

After the use case creation, activity diagrams were developed. The activity diagrams expanded upon the use cases to show specific actions and the logical flow between them. An activity diagram was created for each scenario that the analytical model would be investigating. Alongside the activity diagrams, state machine diagrams were created to show specific states of the actors.

Once the data sources for the model were determined a high-level systems architecture was developed. This allowed for easy communication between the sponsors and the group as to where data would be coming from and how it would be ingested into the system.

After the data for the model was gathered the system engineering team focused on generating a scoring system. The vessel scoring system was used to identify how likely a ship is or is not doing an activity of interest. By using key identifiers, the weighted attributes were fed into the analytical model to determine if investigation is warranted. Section 4.5 will provide more details on the incorporated vessel scoring system.

4.2 USE CASE DIAGRAMS

The use case diagrams were used to scope the model and define clear processes that would be investigated. The use case diagrams display how each actor is involved in the given scenarios and what general actions are taken. Figure 1 displays how a typical legal fishing vessel captures fish and sells them to a buyer. Figure 2 displays the high level use case of conducting illegal fishing, and how the fisherman interacts with the other two actors, an illegal buyer and a potential third party transfer. Figure 3 further decomposes the "Evade Fishing Detection" activity housed in the Illegal Fishing Use Case. In this diagram the actor "Fisher" has to make a choice as to whether they stop broadcasting AIS data to avoid detection, or they spoof AIS by broadcasting a fake AIS signal, thus pretending to be a vessel not conducting commercial fishing. Following that, the fisher conducts the illegal fishing activity in a number of different ways, and at some point goes back to broadcasting a true AIS signal, and returns to port to sell their catch. The follow-on activity and state machine diagrams provide a more in-depth look at those processes of evading fishing detection outlined in Figure 3.

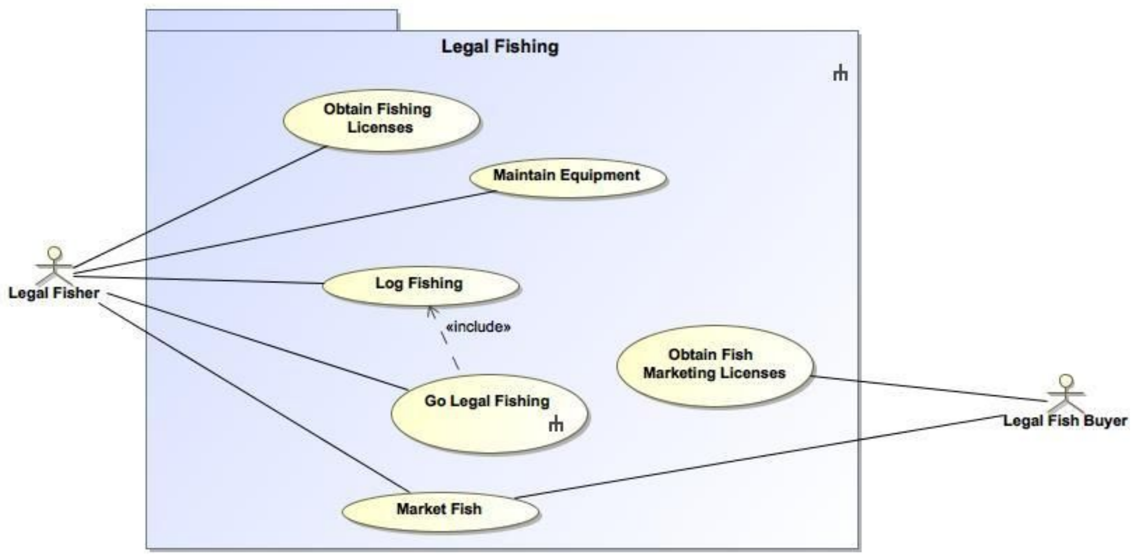


Figure 1 - Legal Fishing Use Case Diagram

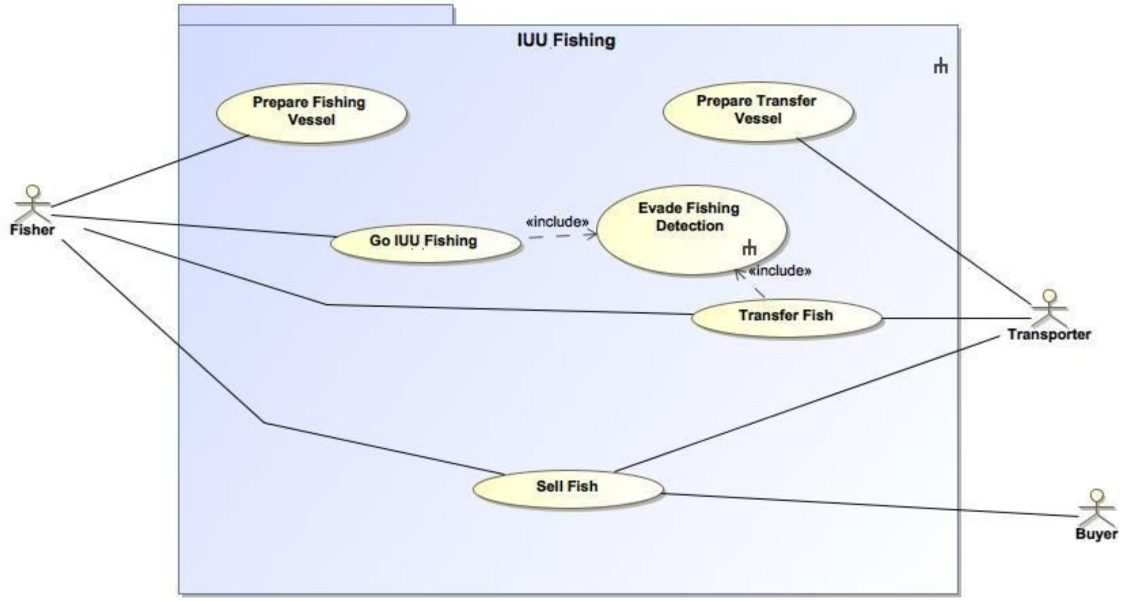


Figure 2 - Illegal Fishing Use Case Diagram

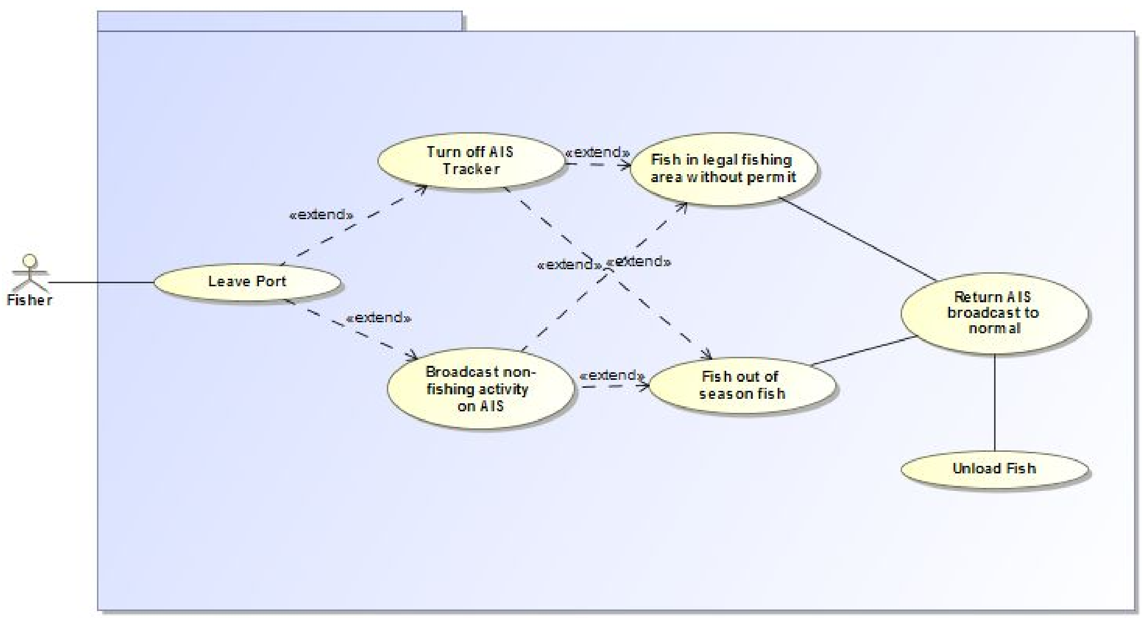


Figure 3 - Evade Fishing Detection Decomposed

4.3 SYSTEM ARCHITECTURE DIAGRAM

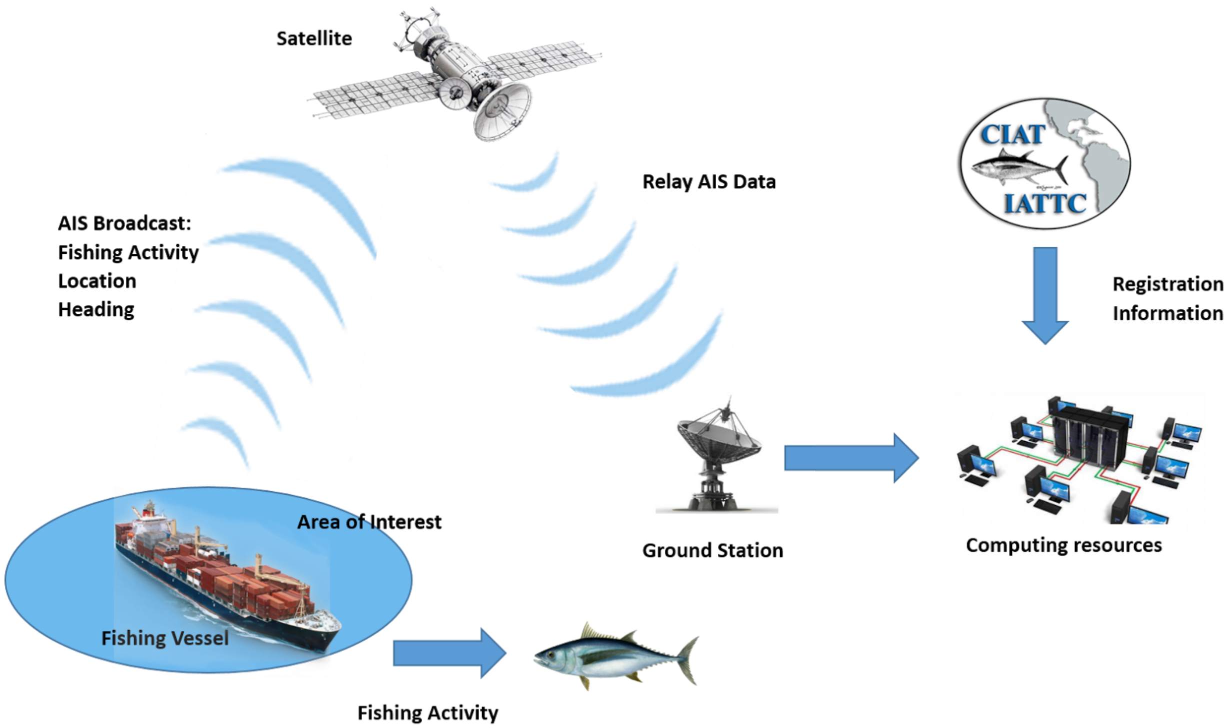


Figure 4 High-level architecture diagram of the Data Analytics to Detect IUU Fishing System

4.4 STATE MACHINE / ACTIVITY DIAGRAMS

The state machine and activity diagrams are used to provide a more detailed look at the steps taken to perform an illegal fishing activity. These diagrams were used to help define what the analytical model would look at to determine behavior that needed to be investigated.

4.4.1 IUU Vessel Turns off AIS Tracker

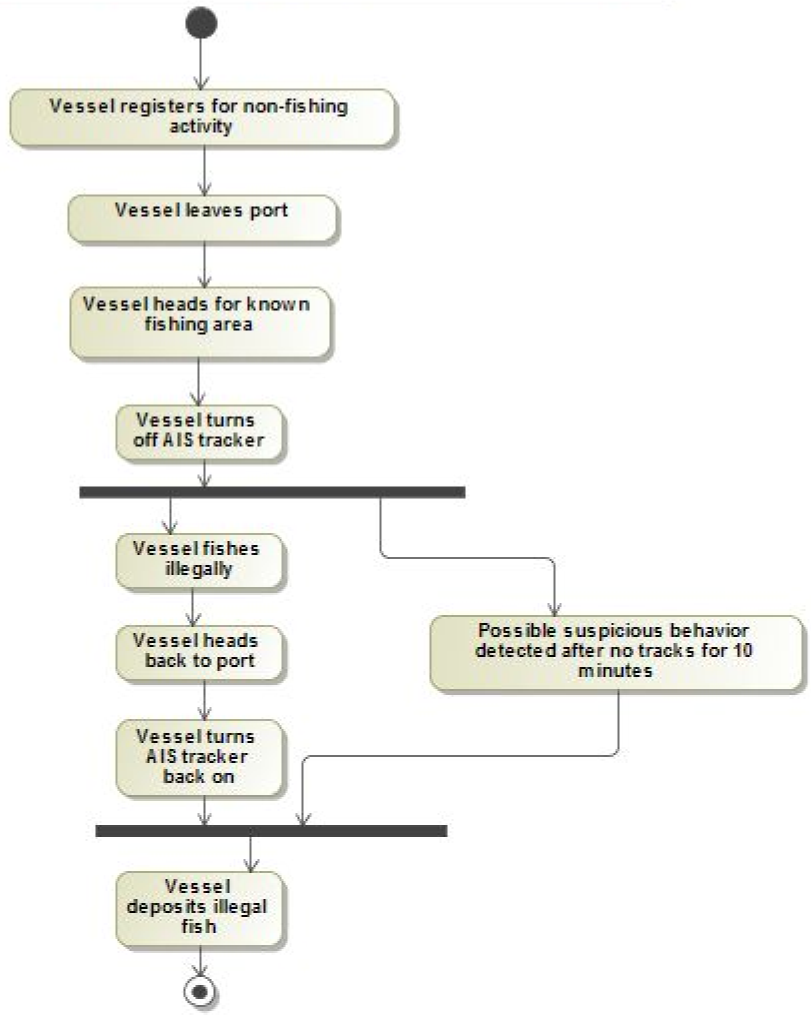


Figure 5 - IUU Vessel Turns Off AIS Tracker

Figure 5 illustrates one of several scenarios that indicates the possibility of illegal behavior. This model focuses on the potential of a fishing vessel turning off their AIS tracker. The scenario begins with a fishing vessel leaving port with no registration or permits for fishing during that voyage. After leaving port, the vessel's heading indicates it is moving in the direction of a known fishing area. Once near the border of the fishing area, the vessel turns off its AIS tracker, signaling suspicious behavior which is determined after not receiving a track for ten minutes. The detection frequency is dictated by AIS equipment, as it is guaranteed to report in at least every ten minutes when operating normally. During the AIS blackout period the vessel performs illegal fishing activity. Once the illegal activity is completed the tracker is turned back on and the fish are off-loaded completing the activity.

4.4.2 IUU Vessel Broadcasts Non-Fishing Behavior

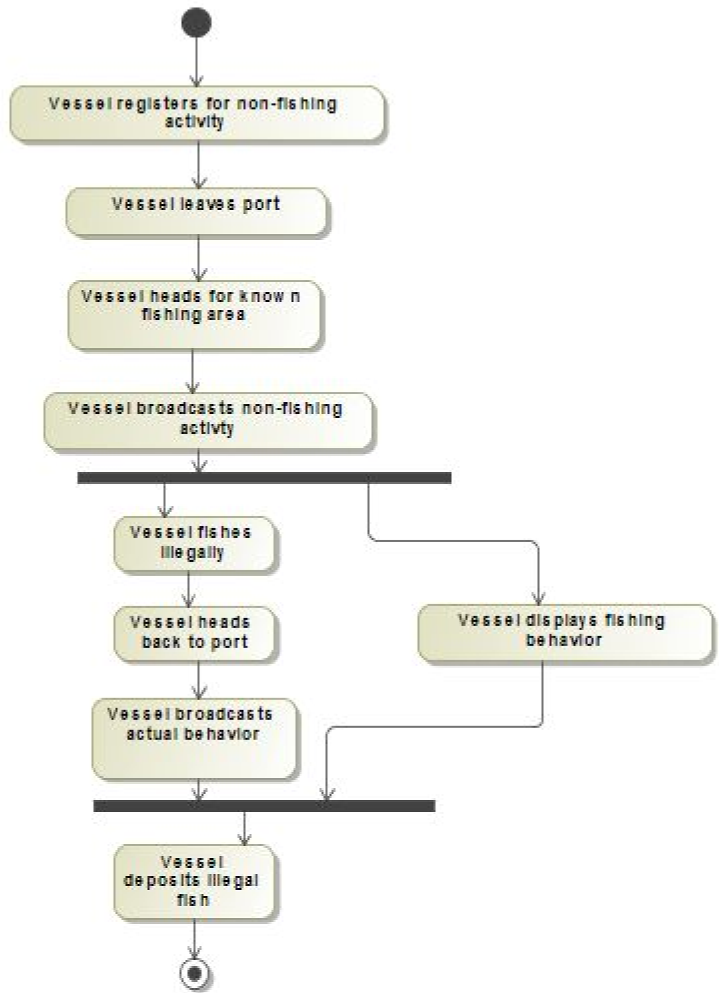


Figure 6 - IUU Vessel Broadcasts Non-Fishing Behavior

Figure 6 depicts a scenario in which a vessel is performing illegal fishing behavior. The focus of the model is a fishing vessel keeping their AIS track on and broadcasting the correct position, but broadcasting as a non-fishing vessel, such as a pleasure craft (ITU-R M.1371-5 code 37 vice 30 for Fishing Vessel), while fishing illegally.

The scenario begins with a vessel leaving port with no registration or permits for fishing during that voyage. After leaving port the vessel's heading indicates it is moving in the direction of a known fishing area. The vessel will begin to fish and show fishing behavior, but their AIS tracker is never switched to broadcast a fishing activity. Once the fishing activity is completed, the vessel deposits the fish to complete the activity.

4.4.3 IUU Vessel is Fishing out-of-season

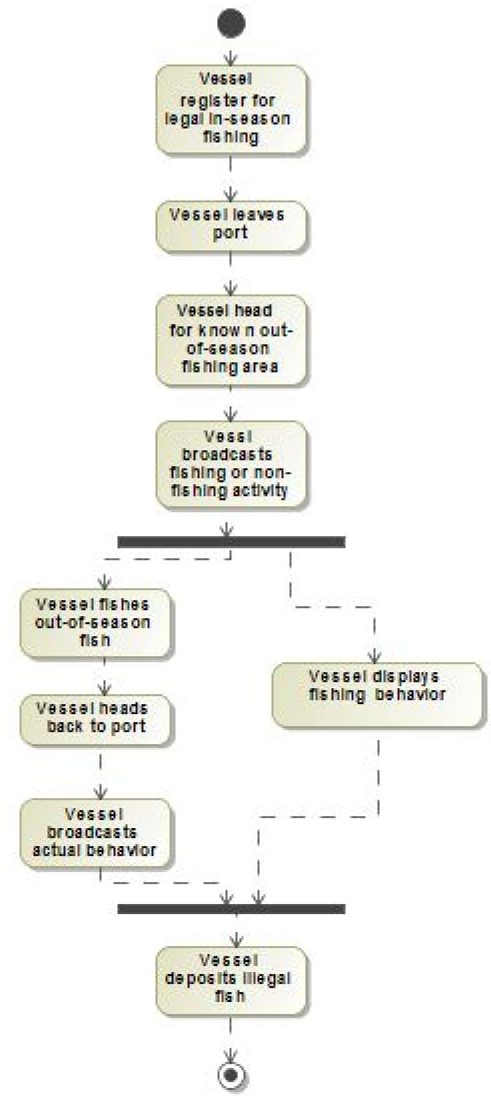


Figure 7 - IUU Vessel is Fishing Out-of-Season

Figure 7 illustrates one of several scenarios that indicates the possibility of illegal behavior. The focus of the model is a fishing vessel keeping their AIS track on and broadcasting the correct position, but broadcast a non-fishing activity while fishing a species of fish that is out-of-season. The vessel leaves port registering to fish for an in-season fish. However, the vessel heads towards an area known to have out-of-season fish. This vessel may or may not broadcast fishing activity at this point. The fishing vessel then illegally deposits the fish, completing the activity.

4.4.5 Vessels too close

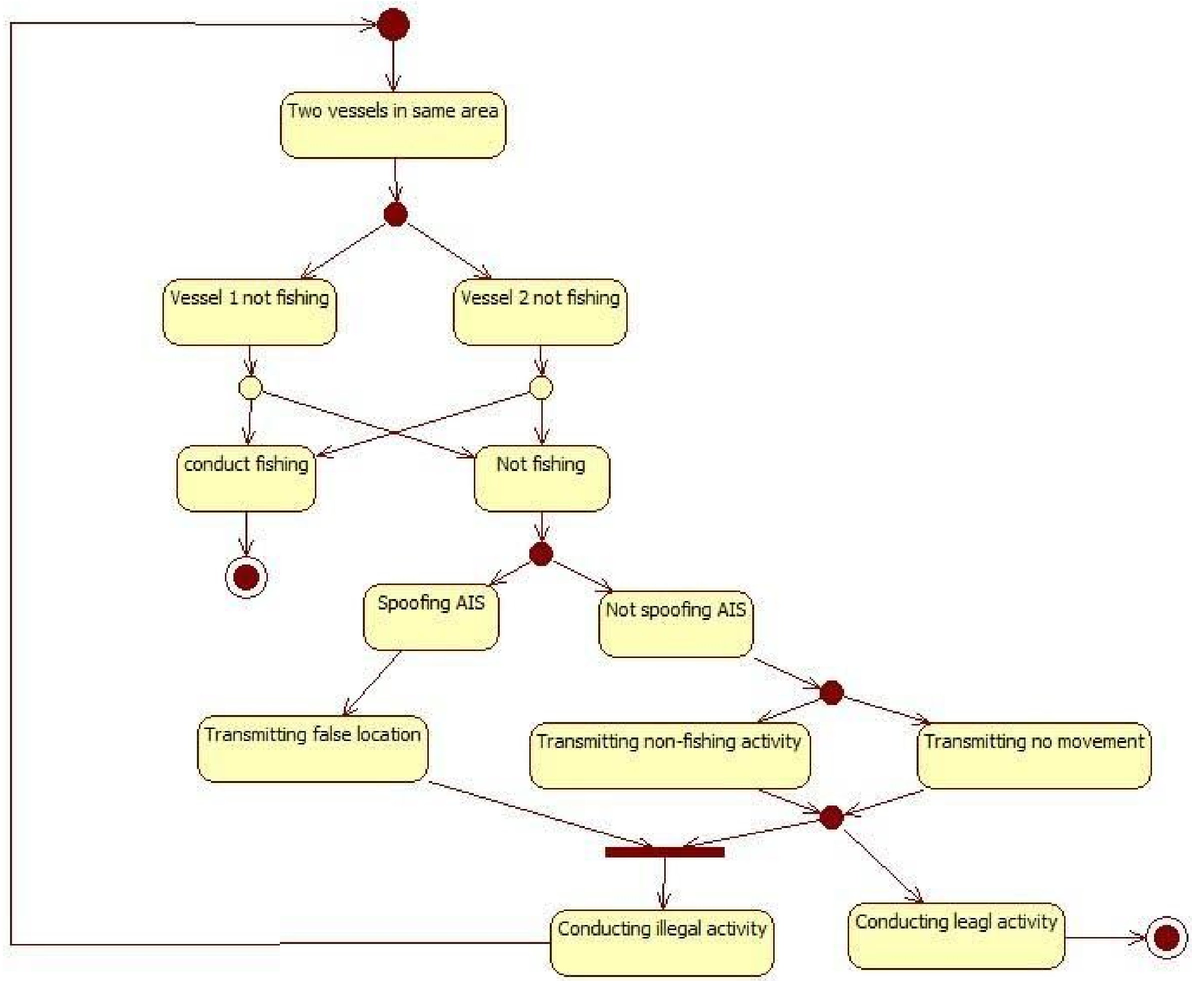


Figure 8 - Vessels Too Close State Diagram

Figure 8 shows a potential illegal activity of interest that was modeled as part of the illegal fishing architecture, it was that of vessels getting too close to one another while at sea. The reason this is of particular interest is because vessels could be transferring fish/cargo from one vessel to another in order to avoid fish catch regulations.

The scenario starts with two fishing vessels in the same area, if both vessels are conducting fishing activity then the scenario is complete, this is perfectly legal, but if AIS data indicates both are not conducting fishing activity then the next step is that they could either be spoofing AIS, or truthfully transmitting that they aren't conducting fishing. If spoofing, they would most likely be transmitting a false location, and because of that we would expect illegal activity. If the vessel(s) is not spoofing AIS, they are either transmitting no fishing activity but movement, or no movement at all. In this scenario, there is the potential for both legal and illegal activity, but at the very least it should be considered "suspicious".

This scenario opens up a few avenues of potential illegal activity, the first being AIS spoofing, which is a something that should be looked at in further detail and incorporated into future illegal fishing models. The second is vessels displaying no movement and a very close proximity, which opens up the possibility for cargo transfer in order to skirt IUU fishing regulations.

4.4.6 Illegal Sale of Fish

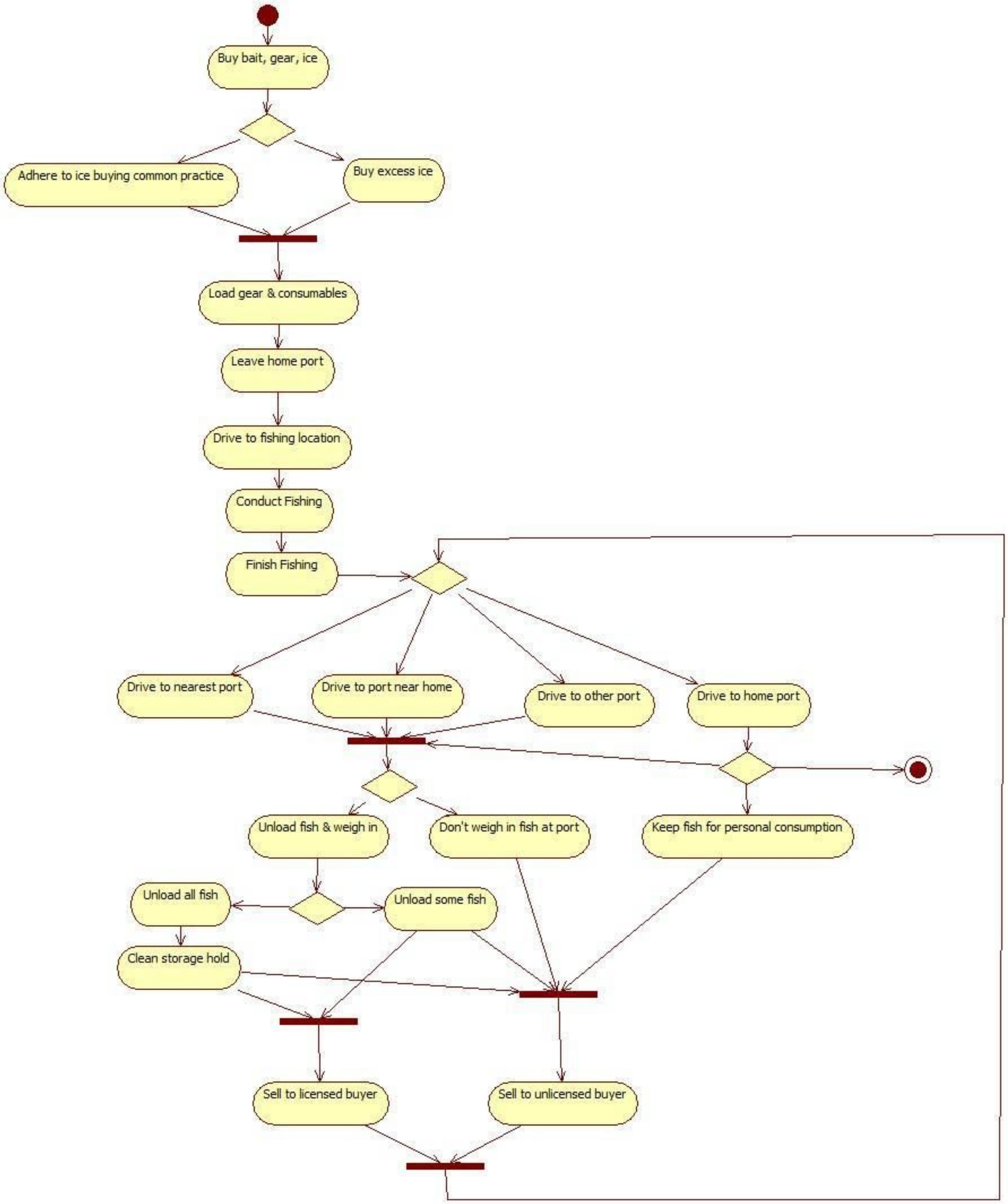


Figure 9 - Illegal Sale of Fish State Machine Diagram

The Illegal sale of fish state machine diagram can be referenced in Figure 9. The illegal fishing use case contains many activities that would be considered IUU fishing, the illegal sale of fish was one of these sub-activities. To more accurately model the process of illegally selling fish, the team decided to break this down into a state machine diagram to analyze the individual steps the fisherman would take.

First the fisherman would need to buy bait, ice, and any gear for the fishing trip. A possible area of further exploration would be whether or not the vessel bought more ice than would reasonably be needed. Next, gear would be loaded, the vessel would leave port, sail to a fishing location, and conduct fishing activities. Once the vessel has concluded fishing, it opens the possibilities for the illegal sale of fish, at this point the vessel has four options we identified: sail to the nearest port, sail to its homeport, sail to a port near home, or sail to a different port. Logically you would think the vessel would either return to home port to unload and sell fish, or it would sail to the nearest port as to minimize the amount of ice they would need, thus maximizing profit. Any activity different would be grounds for flagging the vessel as suspicious.

Next the vessel has the option to unload and check-in fish (all or some), or not unload fish at the port and instead sell to an unlicensed buyer. If only some fish are unloaded and sold to a licensed buyer, this still leaves open the opportunity to sell a portion of the fish to an unlicensed buyer at a higher profit. Once the vessel finishes at this port they have the option of returning to home port, or driving to a different port and once again unloading any remaining fish.

This scenario opens multiple different avenues for the illegal selling of fish that could be explored. This scenario suggests several potential "suspicious" indicators for additional investigation such as a vessel sailing to a port well out of its way, or a vessel is stopping in multiple ports upon the return from a fishing trip. These indicators could be determined through further analysis of AIS/satellite data, using satellite remote sensing imagery as confirmation or, if potentially available, some type of port docking registry.

4.5 VESSEL SCORING SYSTEM

The vessel scoring system was developed as a system engineering tool to assist in detecting the likelihood that a fishing vessel is performing an activity of interest. Several key factors were identified by the team to be the most significant in determining if investigation was warranted. The identifiers were considered for their ability to be incorporated into a multi-factor analytical model. The list of identifiers along with their associated flag and data types are listed in Table 1.

Table 1 Vessel scoring system identifiers

Identifier	Flag	Data Type
In an area of interest (EEZ, MPA, RFMO) [USA-Eastern-Pacific]	1=True 0=False	Boolean
Likelihood of fishing	0-100%	Float [Percentage]
Vessel Not Registered in Area of Interest	1=True 0=False	Boolean

After the fishing model was created and validated, spiral 2 focused on extending the fishing model into an IUU fishing detection model which flagged vessels inside an area of interest, in which they are not allowed to fish, then scores their likelihood that the flagged vessel is fishing, or, in this model, the likelihood that the flagged vessel is illegally fishing. In this phase, the team identified data sources for areas of interest and fishing registries. Due to time constraints, the team decided only to incorporate the Inter-American Tropical Tuna Commission (IATTC) into this iteration of the model. The IATTC is an American agency that is responsible for tracking fishing vessel registration as to determine if a vessel is legally allowed to fish in certain areas for certain species of fish. By combining the GFW data with data from the IATTC, a more refined model can be developed to detect IUU fishing vessels.

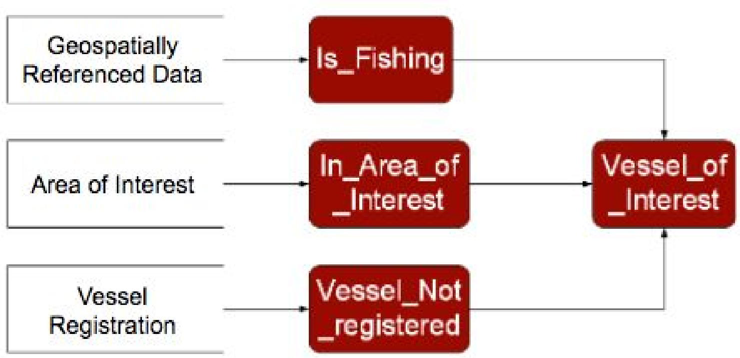


Figure 10: Spiral 2 Data Model

The model cross-references a number of marine protected areas with fishing vessel locations as reported by the AIS data feeds. These Marine Protected Areas (MPAs) are focused in the geographical area of the Eastern Pacific Ocean, in order to make testing of this model reasonable. Areas of interest include the following:

- United States Exclusive Economic Zone located in the Pacific Ocean off the coast of California, Oregon, Washington, and Alaska
- MPAs governed by the United States of America, located in the Eastern Pacific Ocean
- Areas of interest governed by the following Regional Fishery Management Organizations (RFMO):
    - IATTC (Inter-American Tropical Tuna Commission)
    - IPHC (International Pacific Halibut Commission)
    - NPAFC (North Pacific Anadromous Fish Commission)

The second part of the model expansion cross-references the fishing/non-fishing status of vessels, with their physical location. If the vessel is predicted to be both fishing, and is located in or near one of these protected areas, then it is considered as possibly conducting illegal fishing activity and will be flagged as a "Vessel of Interest".

5.0 PREDICTIVE MODEL

Due to readily available data and time constraints, the data analytics was limited to focusing on the vessel scoring model. The team was able to complete analytical models for the "Is_Fishing" and "In_an_Area_of_Interest" processes, as well as, data collection for the a subset of Vessel Registration. Unfortunately due to time constraints, the team was not able to complete a prototype of the vessel scoring

5.1 DATA PREPARATION

Data preparation included data collection, data exploration, data validation, and data cleaning. To perform data collection, there were no useable AIS data feeds which were free and labeled as fishing or non-fishing except for the training data which was provided. After exploring the Global Fishing Watch GitHub repository, Training Data was located which included 27 merged datasets. After running the "prepare.sh", described later, script inside the repository which prepared the data to be used for predictive models, the team felt confident to use datasets beginning with "kristina_" because those were the only datasets which were discussed in documentation during research.

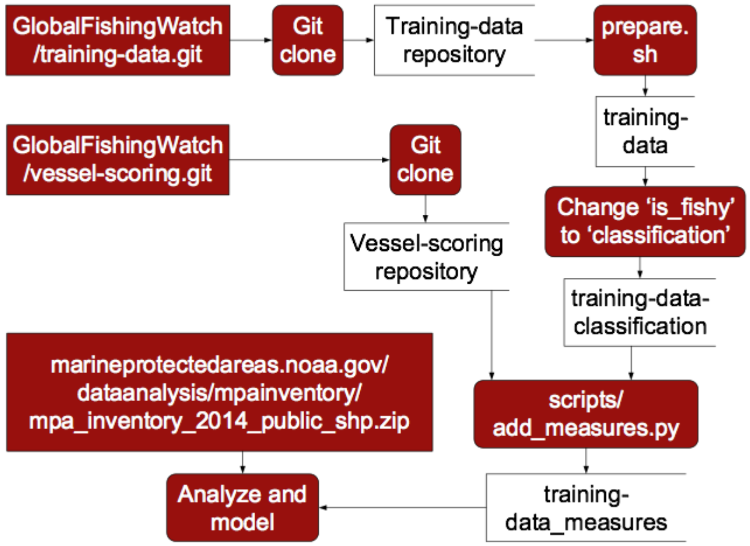


Figure 11: Data preparation for predictive modeling

These datasets were validated through comparison with Kristina Boerder's hand labeled data sets as documented in *Improving Fishing Pattern Detection from Satellite AIS Using Data Mining and Machine Learning* [4]. When explored, these datasets included vessel tracks that were hand labeled as fishing, not fishing, or not labeled status. Based on other supporting documents listed in the appendix, it was noted that gear-types impacted fishing behavior, and that the predictive models would provide better results if analyzed separately. The "prepare.sh" script was used to combine separate vessel track files into a combined gear-specific file; ie one file for each longliner, trawler, and purse seine for a total of three files per source. There were gear-specific data files from several sources but the team could only validate the flagging methodology used by Kristina Boerder; therefore, her datasets were exclusively used to create the model. The team performed data-cleaning by developing code to remove rows with missing or duplicate data. Additionally, rows where fishing was not labeled were removed when preparing the data to create a predictive model.

A qualitative and quantitative description of the data set was contained in Boerder's research paper [4]. Fishing vessels were categorized into three groups of fishing gear type: trawler vessels, longliner vessels, and purse seiner vessels. These fishing vessel groups had distinct descriptions of behavior during a fishing voyage:

A trawler fishing vessel was described to capture fish by dragging a net behind the vessel while moving at a very slow speed. These vessels will typically fish from 3-5 hours at a time travelling approximately between 2.5 and 5.5 knots.

A longliner fishing vessel was described to lay long lines with hooks attached to catch fish. While laying the lines the vessel travels just slightly slower than its cruising speed. However, once the lines are laid the vessel starts to drift, moving at very slow speeds. After several hours of "soaking" the line, the vessel will reverse to haul in the line. The time required to fish in this manner varies, but may take up to a full day.

A purse seiner fishing vessel searches for large schools of fish, once a school is identified large nets attached to floats are deployed. To trap the fish within the nets the vessel must move at quick speeds once the nets are released, roughly 10 knots. After the fish are captured in the net, the vessel drifts to real in the haul.

Boerder used these behaviors and input provided by fisherman and fisheries observers to create training data which can be summarized as vessel track information generated by AIS data and classified at each track point with the classification of fishing, not fishing, or unknown.

5.2 DESCRIPTIVE ANALYTICS

Descriptive analytics were performed by inspecting Boerder's training data as included in the Global Fishing Watch repositories. Analysis was performed using Python 2.7, Jupyter Notebook, Numpy, SciPy, Pandas, and MatPlotLib. Insights into the data were found through simple descriptive statistics, univariate analysis, feature derivation, and visualization creation. A summary of the data is shown in Table 2 below. For each gear type, visualizations were generated, but only the trawling gear type are included in the report.

Table 2 Fishing Classification Data Summary

Vessel Type	Fishing Data Points (unique total)	Not Fishing Data Points (unique total)	Not Flagged Data Points (unique total)	Vessel Count
Longlining	5221 16814	2719 8602	672549 1687630	7
Trawling	16320 38128	21185 48068	290525 648966	5
Purse Seine	297 944	7857 26214	450413 1151857	4

Summary statistics generated by the team using the Boerder dataset are shown in Table 3.

Table 3 Fishing Classification Data Summary

fishing_type	classification		course	distance_from_port	distance_from_shore	lat	lon	speed	timestamp
Longliner	Not Fishing	count	7414	7414	7414	7414	7414	7414	7414
		mean	248.39	816291.22	725804.99	7.45	-55.87	10.35	1373038305
		std	81.19	487247.57	440590.76	24.01	96.93	2.13	9890118
		min	0	14317.47	10049.63	-35.29	-179.98	0	1339297597
		25%	228	461264.44	428353.81	-8.51	-146.76	10	1370201059
		50%	267	714103.62	643770.62	7.99	-40.44	11	1373692204
		75%	303	1187971.66	1068676.25	20.29	7.93	11.6	1379741590
		max	359	2928523.5	2173064.25	74.68	179.93	13.3	1388561206
	Fishing	count	16682	16682	16682	16682	16682	16682	16682
		mean	172.28	1078952.39	955656.68	17.62	-43.09	5.36	1367362376
		std	102.01	565541.64	593926.08	25.9	83.01	3.29	12624892
		min	0	26305.25	10049.63	-17.04	-136.53	0	1338586394
		25%	89	598038.88	368307.41	-6.94	-123.5	2.4	1362186919
		50%	175	1110872.38	1046186.5	8.94	-38.76	5.2	1368336004
		75%	259.1	1568148.38	1457359.88	41.85	-17.41	7.7	1373205889
		max	359.9	2086106.25	1986032.75	75.15	179.34	12.5	1388539774
Trawling	Not Fishing	count	12742	12742	12742	12742	12742	12742	12742
		mean	173.77	761586.64	605976.79	37.01	-28.28	8.6	1402583831
		std	103.14	903418.11	729113.09	39.61	68.85	3.58	29215725
		min	0.2	13038.08	10049.63	-46.25	-179.69	0.2	1327452530
		25%	86.02	122540.86	56043.25	15.67	-52.19	5.5	1383345164
		50%	165.1	355575.28	250581.16	61.22	-13.35	10	1403833590
		75%	270.8	1125814.25	930192.75	64.43	-2.33	11.2	1427274885
		max	359.9	3836963.25	2853625.25	76.24	179.7	16.1	1443685173
	Fishing	count	37574	37574	37574	37574	37574	37574	37574
		mean	179.41	943276.18	840623.26	63.44	-1.58	3.94	1415725649
		std	103.5	599712.61	605220.6	15.45	21.8	2.33	28656390
		min	0	17804.06	10295.38	-44.87	-84.36	0.1	1327628591
		25%	82.3	268038.19	136037.1	61.89	-15.56	2.8	1400820203
		50%	189.5	993167	891140.31	70.38	5.35	3.3	1425170016
		75%	261.77	1554926.25	1470043.12	72.44	7.42	4.2	1440879185
		max	359.9	1933559.5	1612725.12	76.33	157.29	15.1	1443685148
Purse Seine	Not Fishing	count	21442	21442	21442	21442	21442	21442	21442
		mean	187.75	802273.89	575697.67	2.59	-59.8	8.76	1372777717
		std	100.17	496342.53	398673.35	8.75	138.63	5.33	9836071
		min	0	19208.9	10197.79	-10.78	-177.98	0	1338925352
		25%	94.9	420523.88	231996.45	-2.6	-153.14	1.7	1368252014
		50%	211.8	740252.06	501342.88	1.07	-145.89	11.4	1373358386
		75%	271.6	1111535.62	831087.25	4.85	146.14	12.6	1380020856
		max	359.9	2623922.25	1985473.38	43.09	179.94	16.1	1388567609
	Fishing	count	944	944	944	944	944	944	944
		mean	187.61	574714.04	428719.4	-0.6	-15.71	9.08	1375731396
		std	105.53	300097.79	233327.71	5.71	154.95	4.96	8783222
		min	5.4	101254.61	53234.02	-10.77	-174.21	0	1339939781
		25%	89.4	342577.34	194282.3	-4.03	-151.3	2.4	1368085312
		50%	179.6	663004.53	480552.02	-1.06	-145.23	11.75	1375343884
		75%	282.42	781208.75	558309.69	1.56	165.4	12.4	1385902556
		max	356.5	1745820.75	1649774.75	39.92	178.78	15.4	1388495547

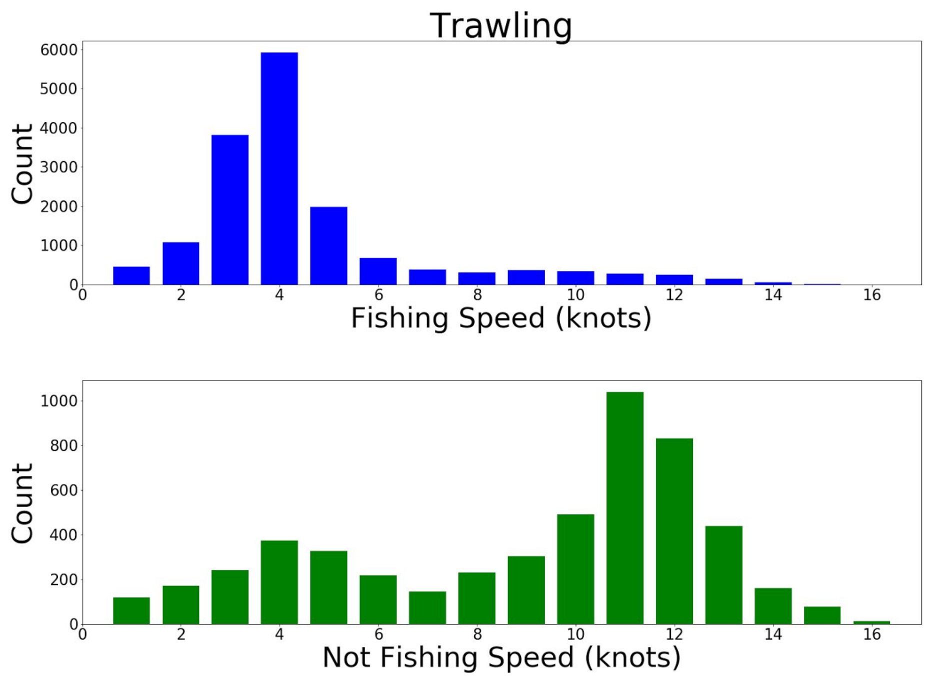


Figure 12 Trawling Fishing Speed Histogram

The histogram in Figure 12 shows that fishing speeds tend to cluster around 3-5 knots while not fishing speeds tend to cluster around 10-13 knots for the trawling gear type.

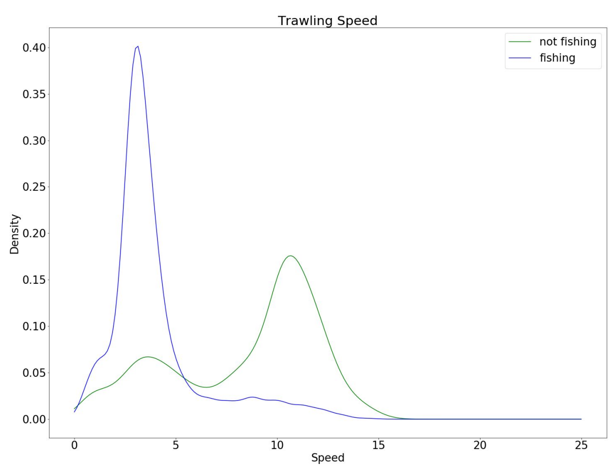


Figure 13 Trawling Fishing Speed Density plot

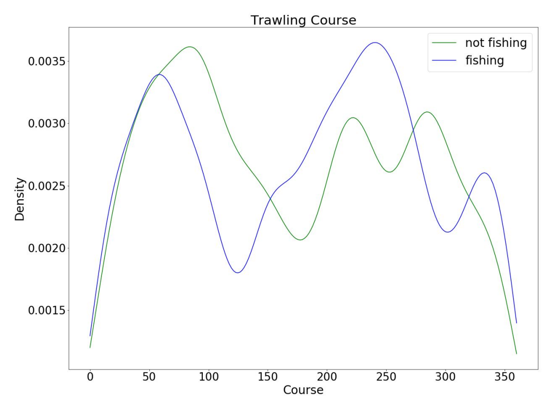


Figure 14 Trawling Fishing Course Density plot

The density plots for the trawling gear-type shows the same clusters for speed, however, there is no visually identifiable difference between fishing and not fishing for course. The course density scale is very low because no specific course, or compass heading, is used for fishing or not fishing activities, therefore the data is relatively evenly distributed among all courses and is not clustered around any particular course.

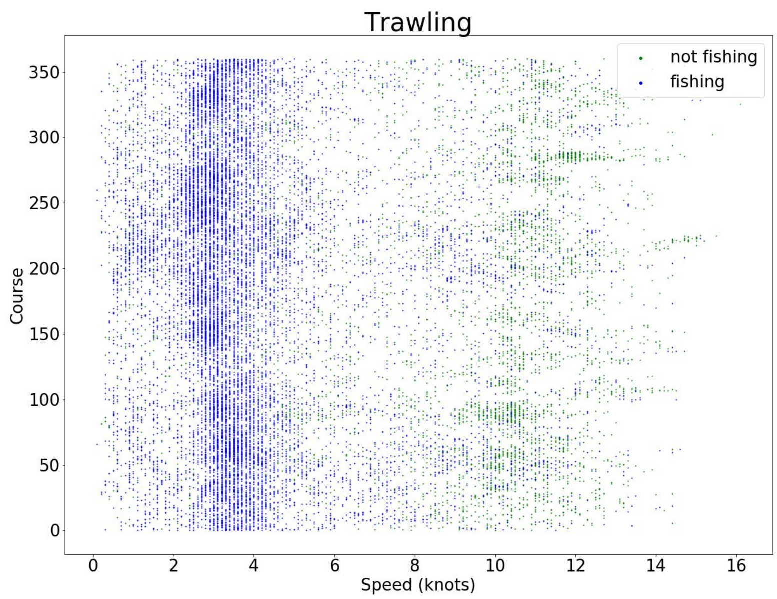


Figure 15 [Trawling Course vs. Speed]

The plot in *Figure 15*, much as the previous ones, only shows the fishing versus not fishing speed clusters for the trawling gear type.

Figure 17 [US MPAs (2014)]

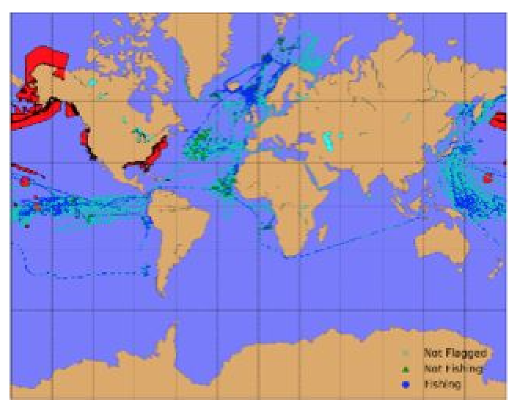
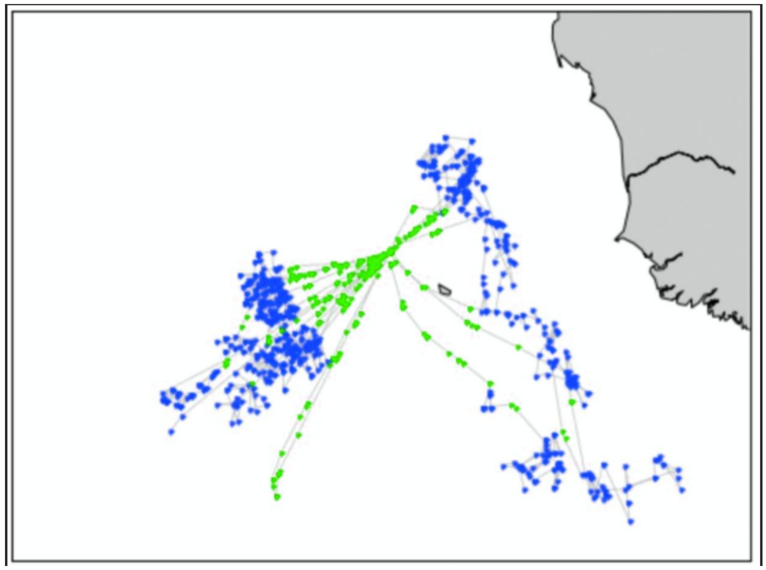


Figure 18 [Dataset Plotted on World Map]

All data points were plotted on a world map that included US MPAs (shown in red). The data is layered from bottom to top; world map, US MPAs, not flagged, not fishing with fishing on the top so that the fishing points could be visually identified. The trawling tracks plotted on a map, see figure below, show that the course deviations for fishing happen much more often than course deviations when not fishing.



Source: https://github.com/GlobalFishingWatch/vessel-scoring/blob/master/docs/ML-Fishing-Score-V1.1.pdf

Figure 19 Trawling Tracks on Map

When comparing all three gear-types data and analytics to that of Boerder, the team was able to validate that data sets in the Global Fishing Watch repositories attributed to Boerder were consistent with her research. It was noted that not all the vessel tracks used in Boerder's research are available in the Global Fishing Watch repositories.

5.2.1 Creating Derived Data

Running the add_measures.py script in the vessel-scoring repository allowed additional derived data (described Appendix D, *Table 9 Added Measures Data Dictionary)* to be visualized. Rolling measures were generated for seven time (seconds) windows; 900, 1800, 3600, 10800, 21600, 43200, and 86400. The more notable derived variables for each of the windows are: normalized course and speed, course standard deviation, and speed standard deviation. These time windows distinguish the difference between fishing and not fishing for the trawling gear-type because that vessel will change its course much more often while fishing than when not fishing.

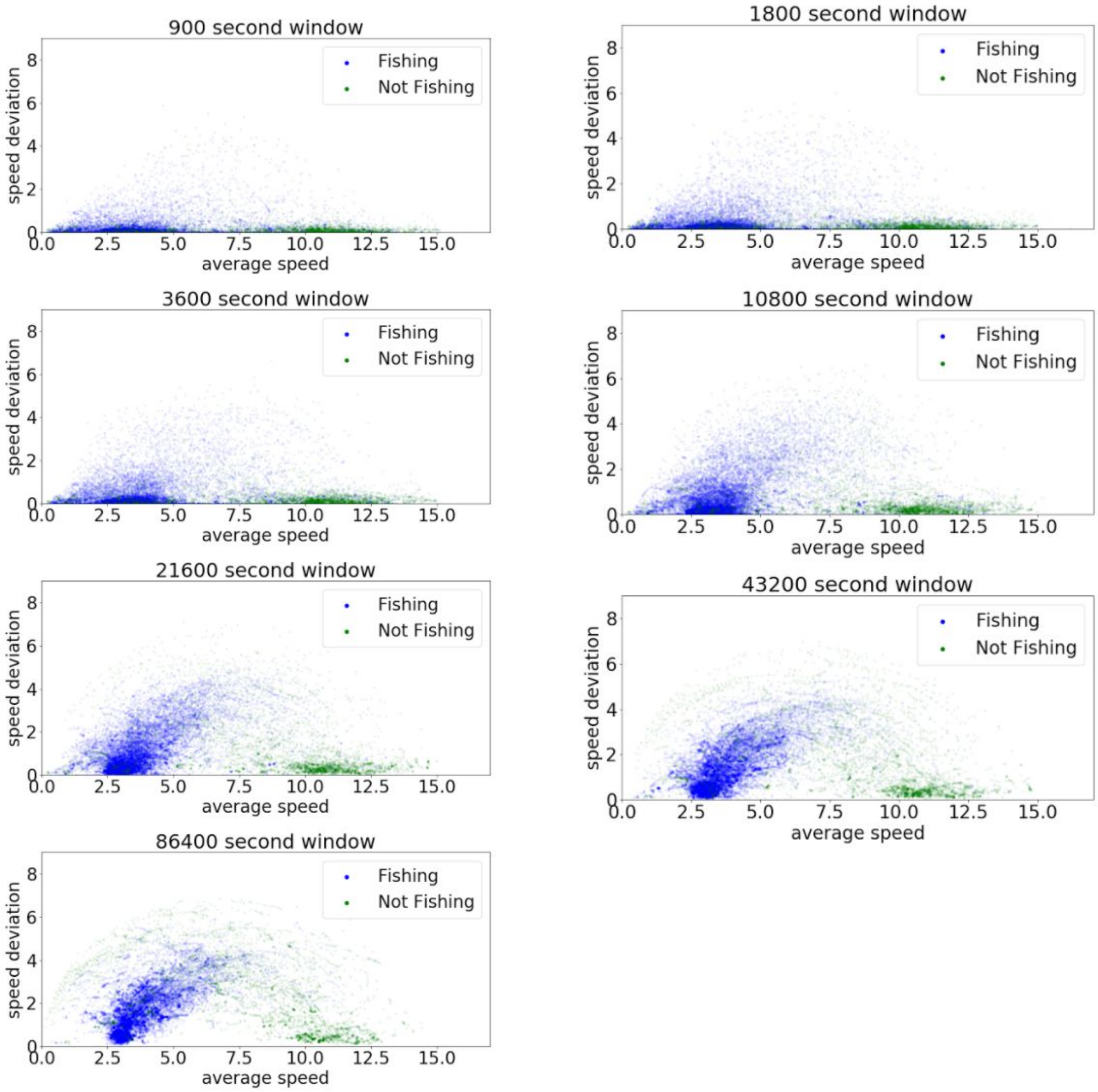


Figure 20 Trawling Speed Deviation vs.Average Speed

The trawling speed deviation vs. average speed plot not only shows the same speed clusters as discussed earlier but also shows there may be a relationship with speed deviation. As the time window is increased there is a difference in speed deviations between fishing and not fishing that becomes more apparent. More importantly, the clustering of fishing from the clustering of not fishing becomes more visually apparent as the time windows are increased.

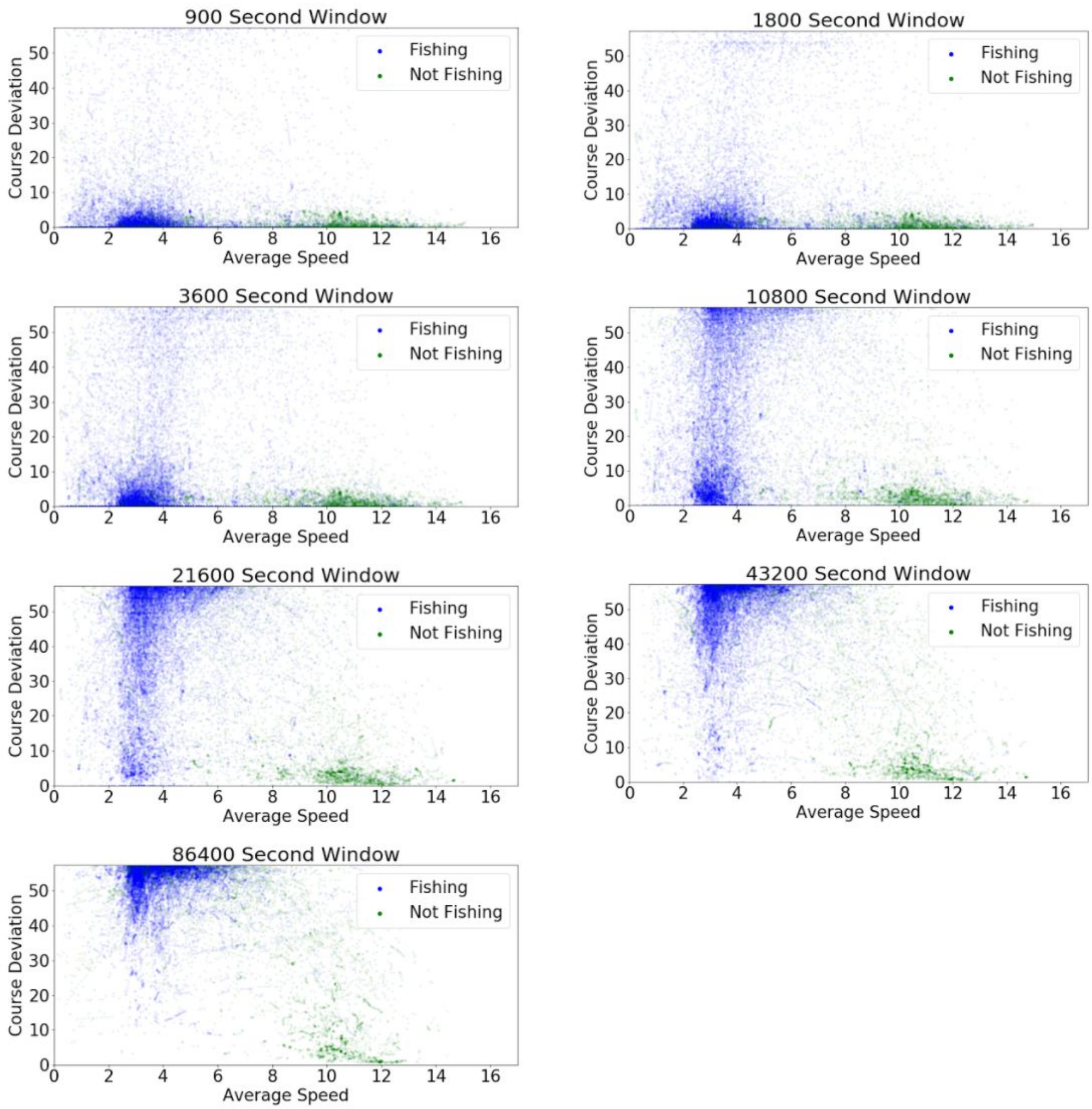


Figure 21 [Trawling Course Deviation vs.Average Speed]

Again, the trawling course deviation vs. average speed plot not only shows the same speed clusters as discussed earlier, but also shows there may be a relationship with course deviation. The maximum course deviations are near 57.5 degrees and are most likely caused by vessel maneuver limitations or AIS data reporting requirements. As the time window is increased there is a difference in speed deviations between fishing and not fishing that becomes more apparent. More importantly, the clustering of fishing from the clustering of not fishing becomes more visually apparent as the time windows are increased up to the 21600 second time window, and then all the fishing data develops a much greater course deviation than the not fishing data that remains relatively constant near lower course deviations.

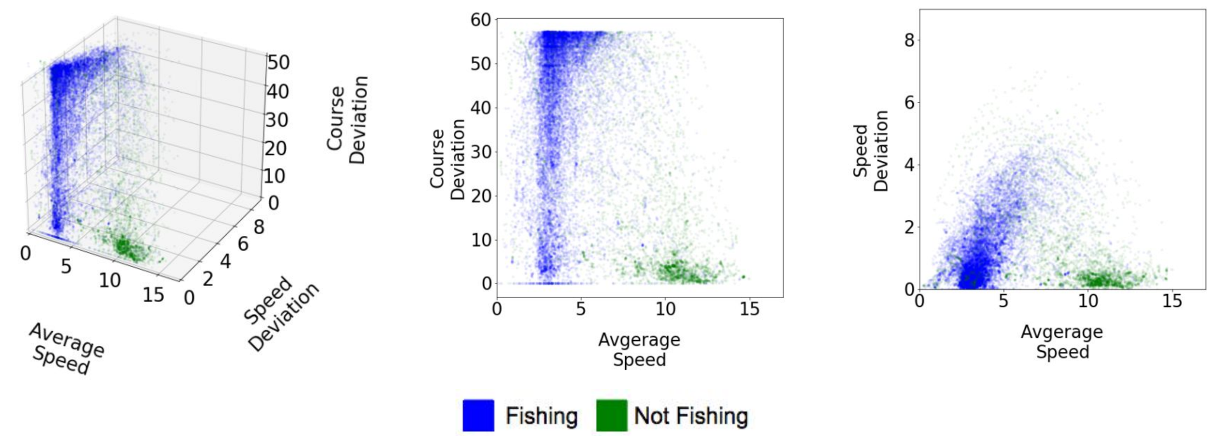


Figure 22 Trawling Course and Speed Deviations vs. Average Speed at 21600 Time Window

Figure 22 shows the course and speed deviation of vessels when fishing and not fishing were compared together. This figure shows the clusters as they relate to course deviation, speed deviation, and average speed.

5.3 PREDICTIVE MODELING AND MODEL VALIDATION

Predictive modeling included modeling technique identification, building the predictive model, model validation, model interpretation and model documentation. For the modeling technique identification, binomial logistic regression was chosen to predict the classification of *fishing* (labeled in the dataset as "1") or *not fishing* (labeled in the dataset as "0") because the variable was discrete with two values. Additionally, logistic regression over multiple time windows and individual gear types was identified in Global Fishing Watch's documentation as the most predictive out of their different predictive models which they trained and discussed in their "vessel-scoring" repository.

To build the predictive model, the *LogisticRegression* function from the Python data analytics package SciKit-Learn was used. The data set was split into training and test data sets (70% and 30% respectively), then a logistic regression model was instantiated, fit to the training data, and evaluated using testing data. Each model’s accuracy was evaluated using

SciKit-Learn's accuracy_score (outputs model accuracy which is calculated by dividing the sum of the correct test data predictions [True Positive + True Negative] by the total all found points), and each model's predictiveness was evaluated using the roc_auc_score (outputs ROC AUC Score) and pr_auc_score (outputs Precision-Recall AUC Score. Following this, stratified 10-fold cross validation was performed to ensure the model was valid across several samples.

5.4 "IN AREA OF INTEREST" MODEL

To model the "in_area_of_interest" code, the team used the Marianas Trench Marine National Monument MPA that includes Guam to define the area of interest. Since the sponsor was interested in algorithms rather than specific software stack implementations, the ray casting method was used to determine if the point fell within the area of interest polygon. The ray casting method draws a horizontal ray from the point of interest and counts the number of times that the ray crosses a line segment that defines the polygon. If the number of crossings is odd, then the point is determined to be in the area of interest whereas if the number of crossings are even the point is said to be outside the polygon. The Python code was compared to Python specific implementations for determining if a point was in a polygon and was found to be consistent. While the ray casting implementation is limited in the information provided (inside or outside the polygon), the team was able to show that the geospatial data contained in an Environmental Systems Research Institute (ESRI) shapefile is suitable for this algorithm and has the potential to be applied to more complex algorithms that can determine the distance from the polygon as well the distance inside the polygon.

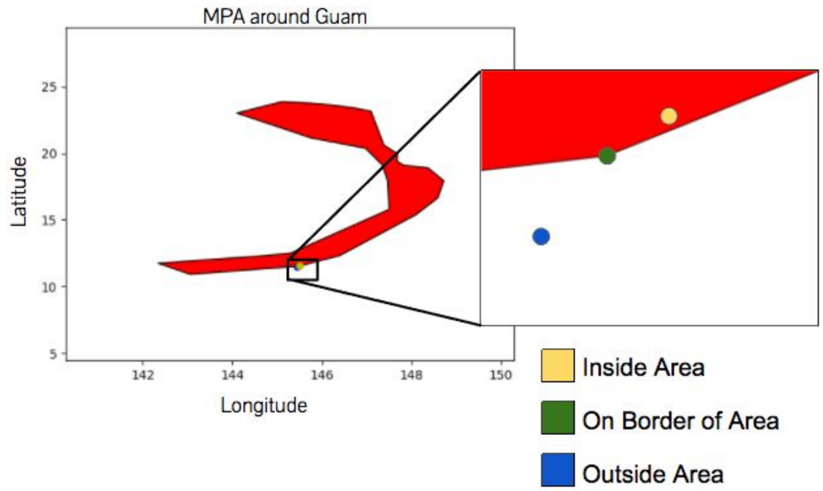


Figure 23 Area of Interest around Guam

6.0 RESULTS, ANALYSIS, AND EVALUATION

For the longliner fishing vessels, the model metrics found are on the following table and plotted curves:

Table 4 Longliner Fishing Model Analysis Data Summary

Metric	Pos. (Fishing) Values	Neg. (Not Fishing) Values	Accuracy	Null Accuracy	True Positive Precision	True Negative Precision	ROC AUC Score	Precision -Recall AUC Score	Stratified 10-Fold Cross Validation Accuracy Scores Mean
Value	9007	3890	91.4%	69.80%	92%	91%	94.7%	96%	90.6%

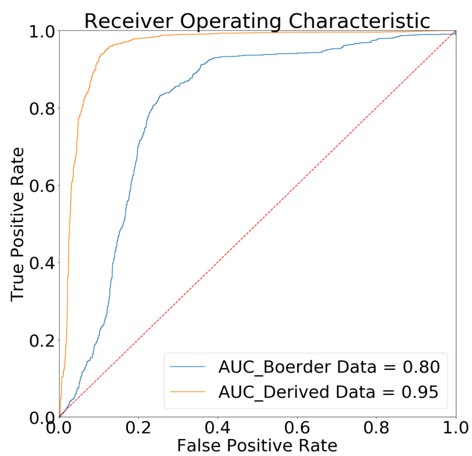


Figure 25 Longliner ROC plot

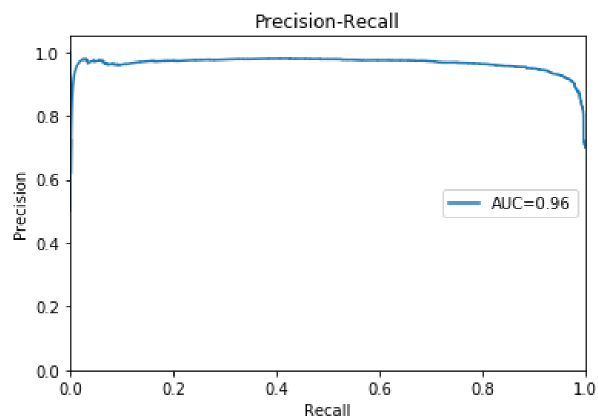


Figure 27 Longliner Model Precision-Recall

The model accuracy was determined to be 91.4% and the model ROC AUC score was determined to be 94.7%. The average accuracy of the cross-validation was found to have an approximate mean of 90.6%. We assessed the model to be highly accurate based on the accuracy score, highly predictive based on the ROC AUC score, and valid based on the stratified 10-fold cross-validation outputs.

For the trawler fishing vessels, the model metrics found are on the following table and plotted curves:

Table 5 Trawler Fishing Model Analysis Data Summary

Metric	Pos. (Fishing) Values	Neg. (Not Fishing) Values	Accuracy	Null Accuracy	True Positive Precision	True Negative Precision	ROC AUC Score	Precision -Recall AUC Score	Stratified 10-Fold Cross Validation Accuracy Scores Mean
Value	32141	31860	78.7%	50.2%	76%	83%	85.9%	82%	76.9%

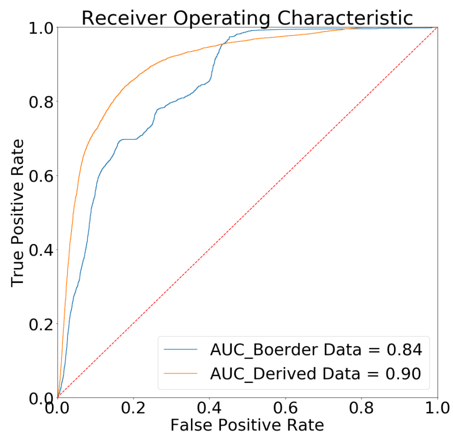


Figure 24 [Trawler ROC plot]

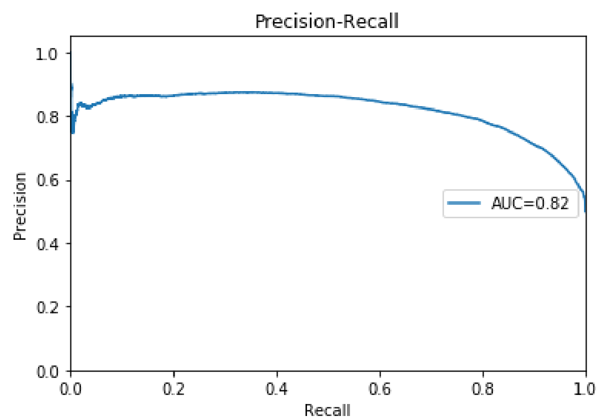


Figure 29 [Trawler Model Precision-Recall]

For the trawler fishing vessels, the model accuracy was determined to be 78.7% and the model ROC AUC score was determined to be 85.9%. The average accuracy of the cross-validation was found to have an approximate mean of 76.9%. We assessed the model to be highly accurate based on the accuracy score, highly predictive based on the ROC AUC score, and valid based on the stratified 10-fold cross-validation outputs.

For the purse seiner fishing vessels, the model metrics found are on the following table and plots:

Table 6 Longliner Fishing Model Analysis Data Summary

Metric	Pos. (Fishing) Values	Neg. (Not Fishing) Values	Accuracy	Null Accuracy	True Positive Precision	True Negative Precision	ROC AUC Score	Precision -Recall AUC Score	Stratified 10-Fold Cross Validation Accuracy Scores Mean
Value	333	12254	97.3%	97.3%	27%	97%	88.3%	16%	97.0%

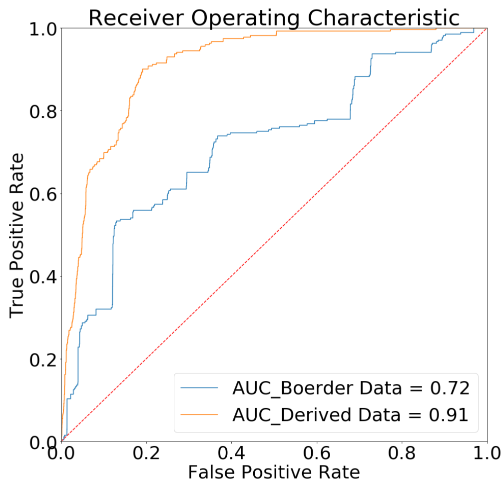


Figure 30 Purse Seine ROC plot

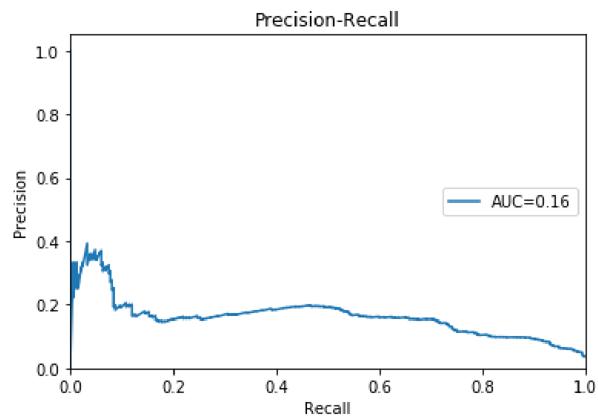


Figure 31 Purse Seine Precision-Recall plot

For the purse seine fishing vessels, the model accuracy was determined to be 97.3% which is considered highly accurate, but it was similar to the null accuracy metric's value of 97.3%. A high null accuracy metric means that incorrectly assessing every positive value where a vessel is fishing would still result in a misleading high accuracy. Due to this class imbalance problem, where the positive values comprise of less than 3% of the evaluated sample, and the desire to obtain a model that predicted positive values where a vessel is fishing, the Precision-Recall AUC Score was a more relevant metric to evaluate model performance. The Precision-Recall AUC score was found to be 16%, so we assessed the model to be highly accurate due to null accuracy, and unpredicted based on the low Precision-Recall AUC score. Further data collection of positive values for the training data, additional data manipulation, and/or a different model is required to obtain a valid predictive model for purse seine fishing vessels.

When comparing the logistic regression model on Boerder's underived training data versus the derived training data, all three ROC plots for the longliner, trawling, and purse seine gear-type models perform significantly better with a difference of .15, .16, and .19 respectively; therefore, the the models created using the derived data are stronger predictive models.

Based on the model assessments, the longliner and trawler fishing vessel models are acceptable for use in IUU Fishing Detection Architecture, and the purse seine fishing vessel model is not acceptable for use. The purse seine fishing vessel model requires additional research and development before it can be used to predict the likelihood of a purse seine vessel fishing and be considered in a model which scores the likelihood of a purse seine vessel IUU fishing.

7.0 FUTURE WORK & RECOMMENDATIONS

The team believes the illegal fishing model produced will serve as an excellent starting point for future expansion of this project. During initial scoping discussion, the team evaluated that developing a full model, with global coverage and the ability to use live data ingest was too large a project to complete in one semester, so the team paid particular attention to how their model could be expanded upon in an effort to work towards an all-encompassing illegal fishing model.

7.1 EXPANSION OF VESSEL SCORING AND FLAGGING

The current vessel scoring system created for this project is just a start, and should be expanded upon to better characterize whether or not a vessel should be marked as "suspicious". Many of the RFMOs maintain databases of both vessels of interest, and vessels cleared to access/fish in restricted areas managed by the RFMO. The team recommends that these databases be incorporated into the model. If a vessel is on an exemption list, it should be marked as "non-suspicious" if it is inside the managed area. Opposite to this, if a vessel in the dataset being analyzed shows up in a suspicious vessel database, housed by one of these RFMOs, its score should be elevated to an appropriate level and indicated as a "suspicious" vessel of interest, and passed along to law enforcement for further investigation.

Expanding the code for determining if a vessel is in an area of interest to calculate the distance from or distance inside the area of interest. This would allow for a weighting system to be developed based upon these distances. These distances could also be incorporated into models that determine that an AIS data transmission has stopped being broadcast, to apply an appropriate weighting to the scoring for those vessels.

The topic of illegal fishing is gaining more and more interest every year, and because of this more data and technological advances for vessel tracking are coming out all the time. Our model considered the AIS data feed limitations as they are in 2017, and what indicators could be gathered from this data. As new and more advanced sensor data becomes available, we recommend incorporating this into the model, and expanding upon the already incorporated illegal fishing indicators.

One very interesting phenomenon the group didn't get a chance to analyze is AIS spoofing. It is critical that vessels exhibiting AIS spoofing behavior be captured as a "suspicious" vessel. There are numerous indicators of spoofing such as:

1. Vessels registered as a non-fishing vessel, but who are exhibiting fishing behavior
2. Vessels whose AIS signal turns ON and OFF
3. Vessels location showing up on land, or who is exhibiting physically impossible route characteristics

AIS Spoofing is a tell-tale sign that some sort of illegal activity is going on, and it should be flagged appropriately and further investigated. Another behavior often indicative of illegal activity are multiple vessels sitting still within close proximity of one another. The group recommends implementing a scoring metric to flag vessels indicating this type of suspicious behavior.

One additional area of interest would be that of incorporating low cost satellite imagery data into the model, once this technology evolves and becomes a cost-effective solution. Satellite data could be used to conduct overhead imagery of a particular MPA or regulated area, cross-reference this data with AIS reporting, and flag any vessels showing up in the satelitte imagery but not transmitting an AIS feed. This would allow the model the ability to target and flag vessels conducting AIS spoofing.

7.2 DATA AGGREGATION OF MMSI NUMBERS & RFMOs

Each registered fishing vessel has a unique Maritime Mobile Service Identity (MMSI) number, and reports this out as part of its AIS data. The team determined that it would be valuable to look at the MMSI number reported in the AIS data, and cross-reference this with RFMO registries to verify whether or not the particular vessels are allowed to be fishing in RFMO regulated zones. AIS data sources are already available, to incorporate this in the model, future efforts would need to identify the RFMO registry databases and write code for comparing the two to look for discrepancies.

7.3 USE OF REAL-TIME AIS DATA

The current model was created using exclusively historic AIS data. It was done this way because data was already broken into testing and training sets, and was already hand labeled as to whether or not the vessel was conducting fishing activity, so it made model training and validation possible. With the hand labeled data, we were able to see how well the model picked up on whether or not a vessel was exhibiting a fishing behavior, regular AIS data wouldn't offer that luxury without further investigation by satellites or law enforcement to determine the nature of the vessel's activity. Satellite imaging data was investigated as part of this project, but at this point isn't deemed a cost-effective method for conducting the model validation.

Now that the model has been created and validated within a reasonable margin of error to detect fishing and non-fishing activity, the team sees it as beneficial to run real-time or at a minimum real-time non-labeled AIS data through it, and conduct additional validation. This step may have to wait until satellite imaging data is reasonably priced, or a small region has the support and buy in of law enforcement to manually check vessels.

7.4 EXPANSION OF MPAS INCLUDED

The current iteration of this illegal fishing model only considers a select group of MPAs. To make data analysis and model validation feasible the team limited the MPAs incorporated to those managed by the United States, in the Eastern Pacific Ocean off the coasts of California, Oregon, Washington, and Alaska. There are many additional MPAs not only governed by the United States, but also ones governed by other countries located in every ocean all over the world. To have an all-encompassing fishing model we recommend the inclusion of all MPAs. This may need to be done in several iterations until all geographical locations can be analyzed.

7.5 EXPANSION OF RFMO AREAS OF INTEREST

The current model was scoped to only include one RFMO of interest, further model expansion should include all RFMOs to capture a full picture of how and where all fish are regulated. Currently the team only included the IATTC regulations managing Tuna fishing, but there are many more RFMOs out there imposing regulations aimed at protecting and preserving specific species of fish. Some of the additional RFMOs the team found are as follows:

1. IPHC (International Pacific Halibut Commission)
2. NPAFC (North Pacific Anadromous Fish Commission)
3. AIDCP (Agreement on the International Dolphin Conservation Program)
4. Pollock Conservation in the Bering Sea
5. PSC (Pacific Salmon Commission)
6. WCPFC (West and Central Pacific Fisheries Commission)

7.6 ADDITIONAL AREAS OF INTEREST

There are numerous additional areas of interest the Team deems valuable to investigate further; the first being the fishing licensing system. To conduct commercial fishing, and the buying and selling of catch, there are numerous yearly licenses & registrations that must be obtained. An easy way to catch IUU fishing would be to cross-reference licensing databases with AIS location data that indicated fishing activity, if the vessel didn't show up as having the appropriate licenses for the area or type of fishing it was conducting then it should be flagged as "suspicious". Additionally, if the vessel was approaching an area of interest that they weren't licensed for, and then AIS suddenly goes dark, this should raise suspicion and immediately be flagged. At the time of this project the team searched for available licensing system data sources, but was unable to locate any. To further complicate the issue, each U.S. state has its own set of licensing laws, so it would be necessary to acquire multiple data sources (from each state of interest) to properly integrate this piece into the model. An example of licenses needed to fish off the State of California can be seen in the Table 7 below. These are some of the licensing types that should be investigated and cross-referenced with AIS data, once appropriate data sources are identified or become available. In addition to standard licenses to fish and register your vessel, individual licenses are required to fish specific zones, conduct a type of fishing, and to fish for individual fish species.

Table 7: State of California Commercial Fishing License Overview

License	Description
Resident / Non-Resident Commercial Fishing License	--
Commercial Vessel Registration	--
Commercial Passenger Fishing Vessel License	--

Commercial Ocean Enhancement Stamp	Allows for the fishing of white sea bass south of Santa Barbara County
Non-Restrictive Permits	Allows fishing for: Anchovy, shrimp, crayfish, prawn, lobster, crab. Swordfish, etc.
Restricted Access Permits	Different permits allow for particular fishing methods: bottom trawl, gill net, traps, diving, etc.
Nearshore Fishing Permits	Licenses for fishing various coastal regions

Another potential area of interest is looking at whether or not a vessel visited multiple different ports after returning from a commercial fishing trip. Visiting multiple ports could indicate suspicious activity. Typically, you would expect a vessel to return to a single port, offload all of its fish and sell to a licensed buyer. Visiting multiple ports could indicate that either the vessel wasn't disclosing the true size of their catch, and had to visit multiple ports to sell all of their catch, and to not raise suspicion of overfishing. Or it could indicate that they sold part of their catch legally, and sold another part of it illegally at a higher price to an unlicensed buyer. In either scenario, this vessel needs to be tracked further and labeled as "suspicious" so it can be investigated further by law enforcement. The team recommends using AIS location data to look for and flag this type of activity.

Another potential area of further investigation would be the vessel meeting up with another vessel at some location off shore. Here they would likely be conducting the transfer of fish or goods, so that upon their return to port they wouldn't garner any suspicion by having illegal catch onboard. The third party vessel could take ownership of the fish and then sell them to an illegal buyer. Both of these activities were highlighted in the illegal fishing use case diagram, and warrant further exploration and incorporation into the model.

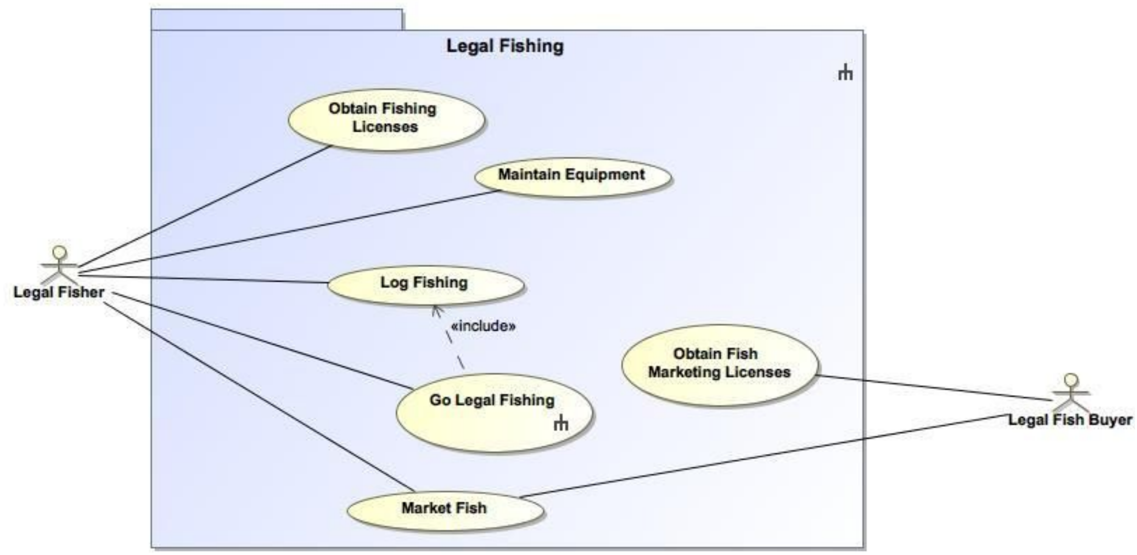
Another area of interest identified by the team is that of purchasing ice to keep fish cold while a vessel is out on an extended fishing trip. While in port a fishing vessel will have to make estimations of how much ice it will need to sustain itself over an extended fishing trip, and these estimations are based on length of the trip, size of the cargo hold, estimated pounds of fish expected, and environmental conditions expected during the trip. Standard guidelines are used to determine how much ice will be needed, a good guideline could be found here: http://www.fao.org/docrep/006/Y5013E/y5013e07.htm . It may prove fruitful to evaluate various vessels to see if the amount of ice they purchased from the port vendor lined up with the conditions they encountered (days at sea, environmental conditions, their cargo hold size), and the amount of fish they returned with for sale. If a vessel is found to be consistently overbuying ice, this could indicate that they are harvesting more fish then they come back to port with, and transferring these to another vessel while still at sea. This would then warrant the dedication of additional resources to monitor the particular vessel for IUU fishing activity. The group was unable to find the necessary data sources at this time, but identified it as a possible area for future model expansion. A similar analysis could be done looking at how much bait or fishing tackle a vessel acquires prior to leaving for an extended fishing trip.

8.0 REFERENCES

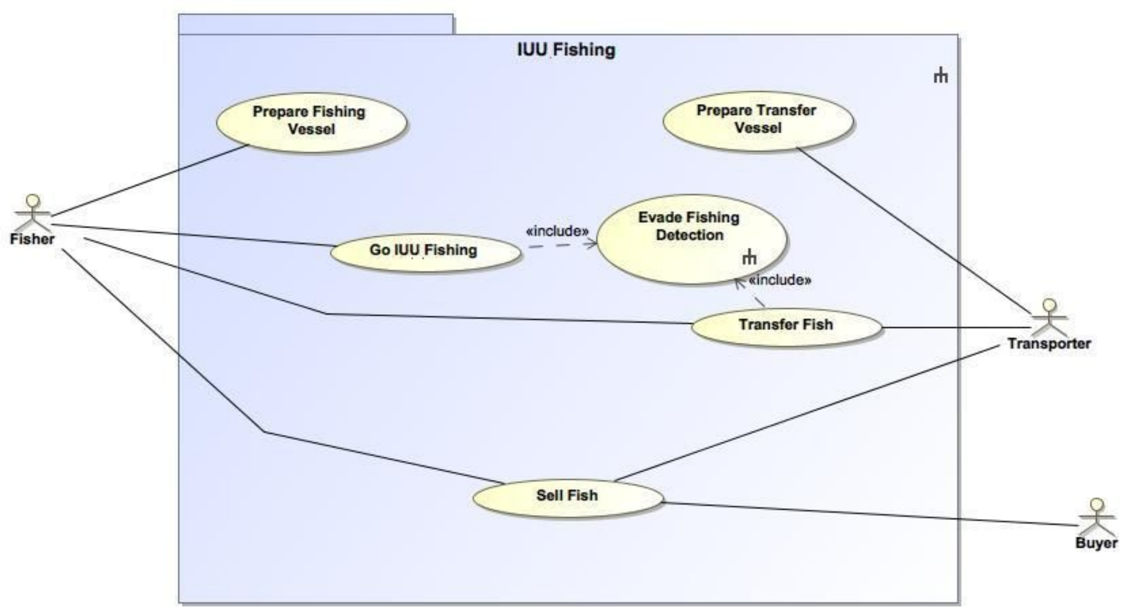
1. World Ocean View. (n.d.). Illegal Fishing. Retrieved February 10, 2017, from http://worldoceanreview.com/en/wor-2/fisheries/illegal-fishing/
2. United States Coastguard. (n.d.). Protecting America's FIsheries. Retrieved April 4, 2017, from https://www.uscg.mil/history/articles/Fisheries.pdf
3. Cheeseman, 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/
4. de Souza EN, Boerder K, Matwin S, Worm B (2016) Improving Fishing Pattern Detection from Satellite AIS Using Data Mining and Machine Learning. PLoS ONE 11(7): e0158248. doi:10.1371/ journal.pone.0158248
5. Natural Earth;. Accessed: 2015-Nov-12. http://www.naturalearthdata.com/http//www.naturalearthdata.com/download/10m/physical/ne_10m_coastline.zip.

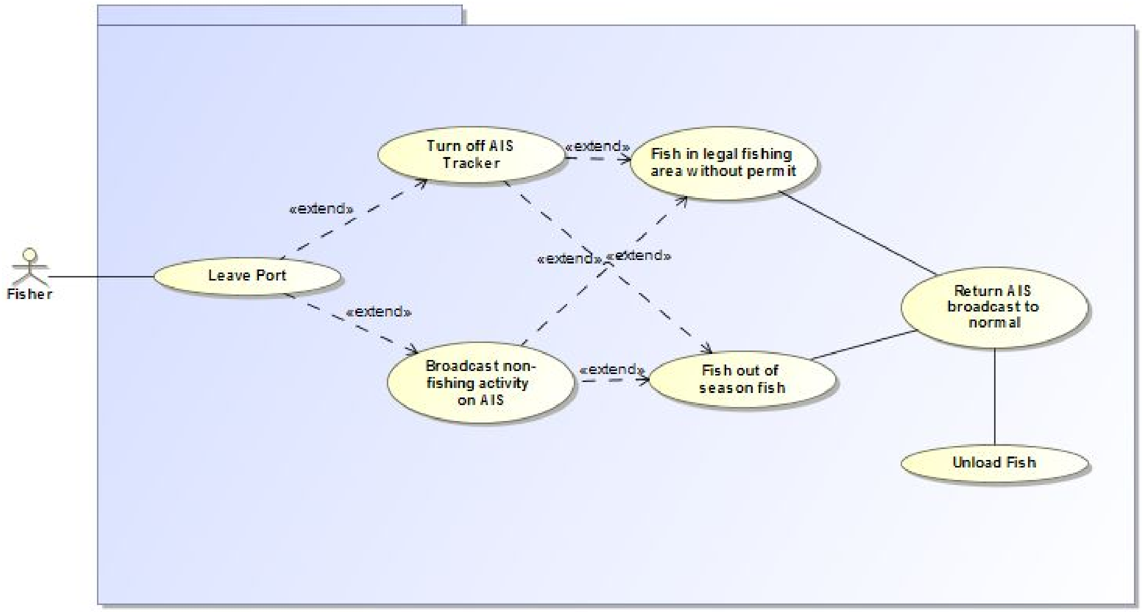
APPENDIX A – SYSTEM ENGINEERING MODELS

Legal Fishing Use Case



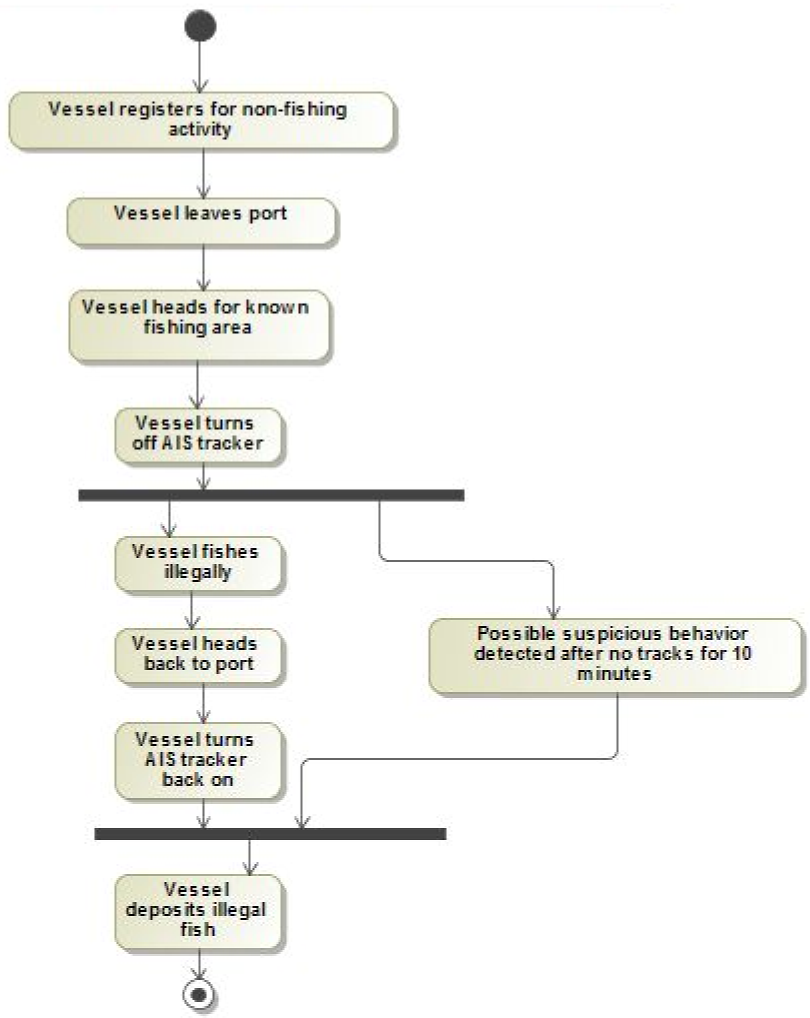
IUU Fishing Use Case



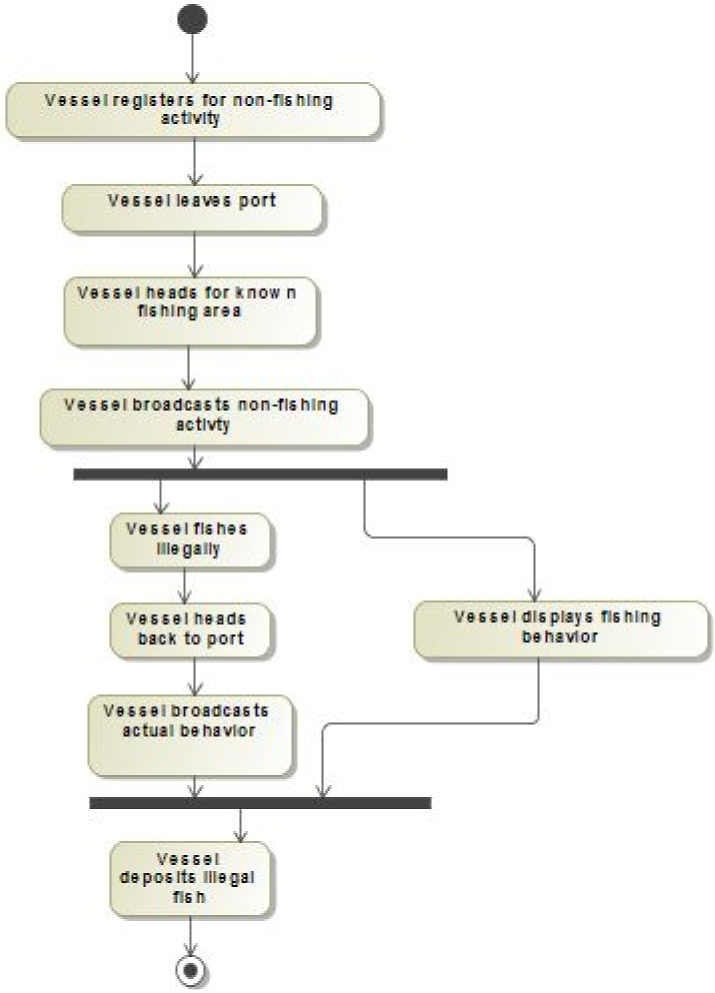


Fisher
Leave Port
«extend»
Turn off AIS Tracker
«extend»
Broadcast non-fishing activity on AIS
«extend»
Fish in legal fishing area without permit
«extend»
«extend»
«extend»
Fish out of season fish
Return AIS broadcast to normal
Unload Fish

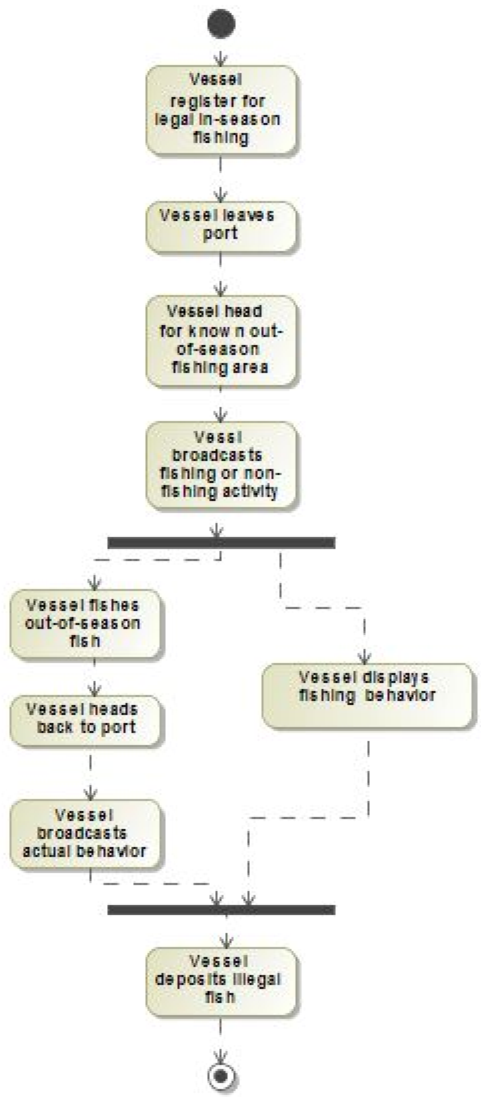
IUU vessel turns off AIS tracker



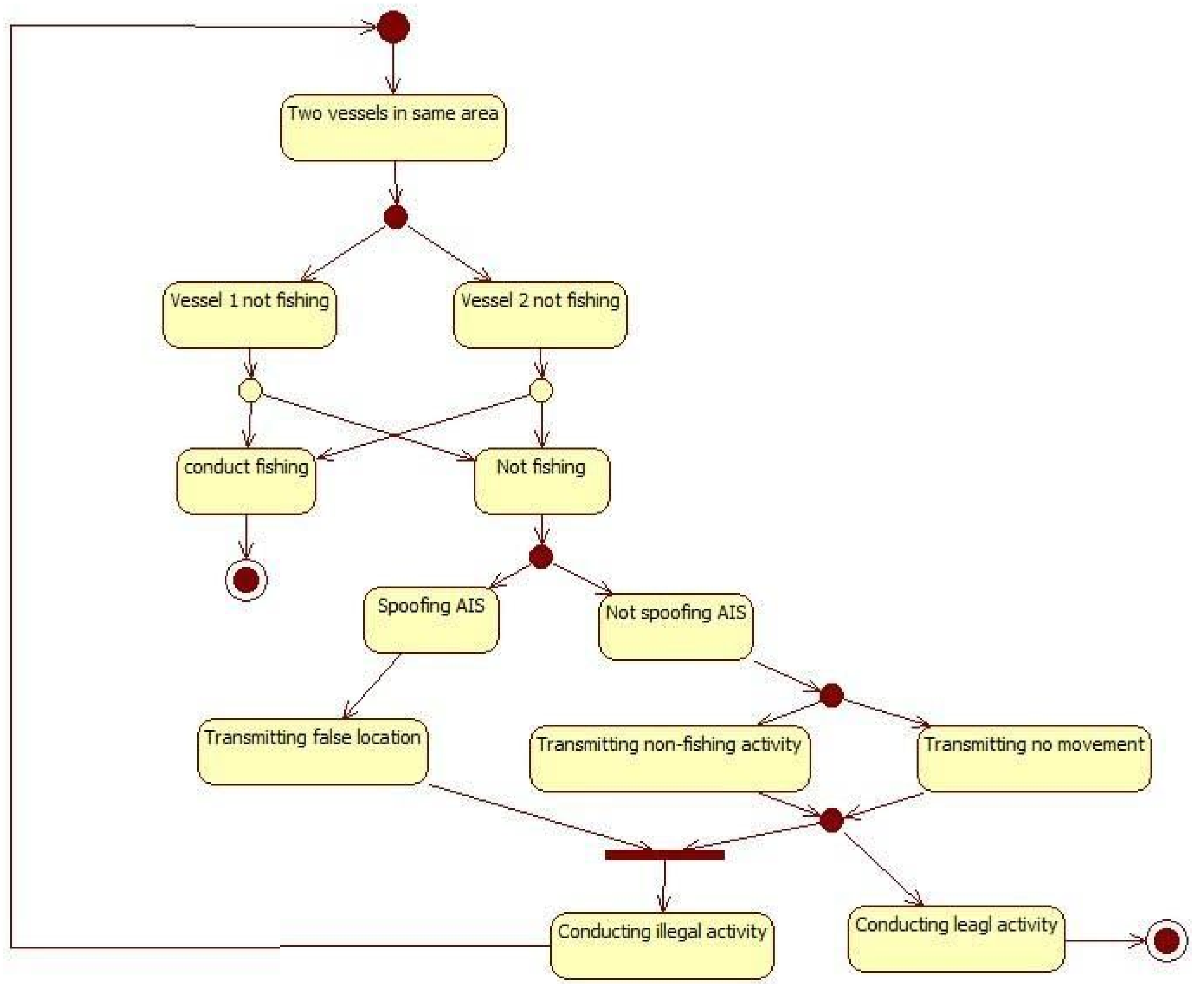
IUU fishing vessel broadcasts non-fishing behavior



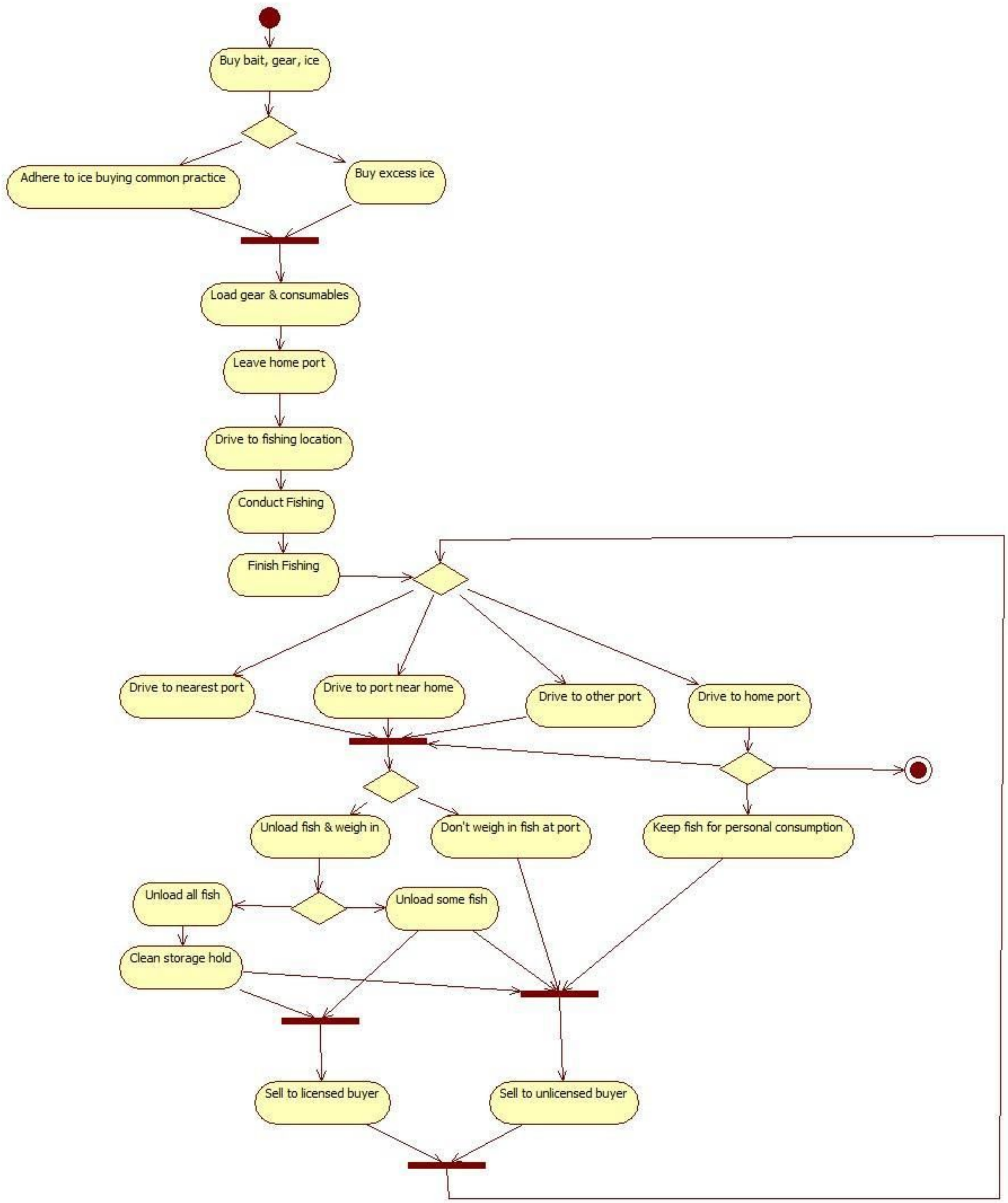
IUU fishing vessel is fishing out-of-season fish



Vessels Too Close Activity Diagram



Illegal Sale of Fish Activity Diagram



APPENDIX B – ABBREVIATIONS

AIDCP – Agreement on the International Dolphin Conservation Program

AIS – Automatic Identification System

AUC – Area Under Curve

EEZ – Exclusive Economic Zone

ESRI – Environmental Systems Research Institute

GFW – Global Fishing Watch

GMU – George Mason University

IATTC – Inter-American Tropical Tuna Commission

IPHC – International Pacific Halibut Commission

IUU – Illegal, Unreported, or Unregulated

LM – Lockheed Martin

MMSI – Maritime Mobile Service Identity

MPA – Marine Protected Area

NOAA – National Oceanic and Atmospheric Administration

NPAFC – North Pacific Anadromous Fish Commission

PSC – Pacific Salmon Commission

RFMO – Regional Fisheries Management Organization

ROC – Receiver Operating Characteristic

USCG – U.S. Coast Guard

WCPFC – West and Central Pacific Fisheries Commission

APPENDIX C – PROJECT MANAGEMENT

Organization Chart – Roles & Responsibilities

Team Member	Role	Responsibilities
Jarred Byrnes	Systems Engineer	Systems Engineering tasks: system architecture model development, scoring system, IUU background and initial research, all team deliverables
Jonathan Gessert	Lead Data Analyst	Development of data analytics model, descriptive analytics, predictive modeling, model validation, website creation, all team deliverables
Edward Kerrigan	Data Analyst / Systems Engineer	Development of data analytics model, descriptive analytics, predictive modeling, model validation, gathering datasets, all team deliverables
Jonathan Matteson	Project Manager / Systems Engineer	Organized team & sponsor meetings, handled communication with sponsor & professor, EVM, project management responsibilities, developed system architecture models, gathered various datasets, all team deliverables

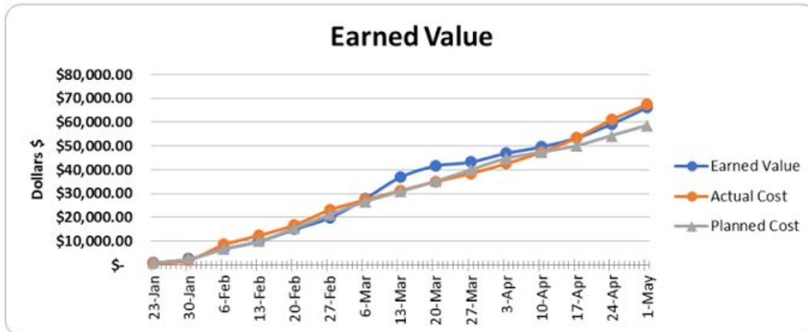
Project Schedule

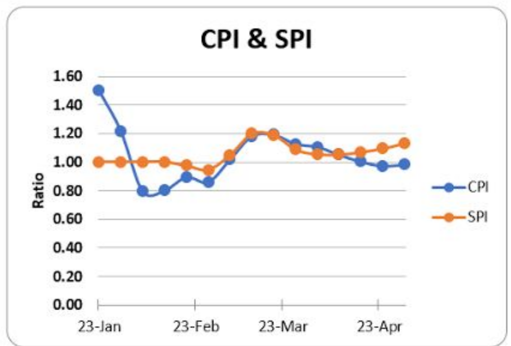
ID	Info	Task Mode	WBS	Task Name	Duration	Start	Finish
1			**1**	**Project Management**	**94 days**	**Sun 1/29/17**	**Tue 5/2/17**
2			**1.1**	**Project Initiation**	**19 days**	**Sun 1/29/17**	**Thu 2/16/17**
14			**1.2**	**Reviews**	**49 days**	**Fri 3/3/17**	**Thu 4/20/17**
19			**1.3**	**Sponsor Synchronization Meeting**	**92 days**	**Tue 1/31/17**	**Tue 5/2/17**
34			**2**	**Research**	**61 days**	**Sun 1/29/17**	**Thu 3/30/17**
35			2.1	Illegal Fishing	61 days	Sun 1/29/17	Thu 3/30/17
36			2.2	Tools Research	11 days	Fri 2/17/17	Mon 2/27/17
37			**3**	**Model Development**	**86 days**	**Sun 1/29/17**	**Mon 4/24/17**
38			**3.1**	**Systems Engineering**	**47 days**	**Fri 2/17/17**	**Tue 4/4/17**
39			3.1.1	Develop Illegal Fishing Use Case	15 days	Fri 2/17/17	Fri 3/3/17
40			3.1.2	Develop Illegal Fishing Behavior Diagrams	40 days	Fri 2/24/17	Tue 4/4/17
41			**3.2**	**Spiral 1 - Fishing Identification**	**35 days**	**Sun 1/29/17**	**Sat 3/4/17**
42			3.2.1	Spiral 1 - Objectives	3 days	Fri 2/17/17	Sun 2/19/17
43			3.2.2	Spiral 1 - Data Preparation	3 days	Fri 2/17/17	Sun 2/19/17
44			3.2.3	Spiral 1 - Develop Data Model	8 days	Fri 2/17/17	Fri 2/24/17
45			3.2.4	Spiral 1 - Descriptive Analysis	5 days	Mon 2/20/17	Fri 2/24/17
46			3.2.5	Spiral 1 - Predictive Modeling Phase I	8 days	Sat 2/25/17	Sat 3/4/17
47			3.2.6	Spiral 1 - Model Validation	7 days	Sun 1/29/17	Sat 2/4/17
48			3.2.7	Spiral 1 - Complete	0 days	Sat 2/4/17	Sat 2/4/17
49			**3.3**	**Spiral 2 - Flagging Vessels**	**40 days**	**Thu 3/16/17**	**Mon 4/24/17**
50			3.3.1	Spiral 2 - Objectives	4 days	Thu 3/16/17	Sun 3/19/17
51			3.3.2	Spiral 2 - Data Preparation	5 days	Mon 3/20/17	Fri 3/24/17
52			3.3.3	Spiral 2 - Develop Data Model	17 days	Sat 4/8/17	Mon 4/24/17
53			3.3.4	Spiral 2 - Predictive Modeling	22 days	Sat 3/25/17	Sat 4/15/17
54			3.3.5	Spiral 2 - Model Validation	7 days	Sun 4/16/17	Sat 4/22/17
55			3.3.6	Spiral 2 - Complete	0 days	Mon 4/24/17	Mon 4/24/17
56			3.4	Model Development Complete	0 days	Mon 4/24/17	Mon 4/24/17
57			**4**	**Deliverables**	**97 days**	**Sun 2/5/17**	**Fri 5/12/17**
58			4.1	High Level Architecture	7 days	Mon 4/24/17	Sun 4/30/17
59			4.2	Library of Data Sets	7 days	Mon 4/24/17	Sun 4/30/17
60			4.3	Library of Algorithms	7 days	Mon 4/24/17	Sun 4/30/17
61			**4.4**	**Final Presentation**	**97 days**	**Sun 2/5/17**	**Fri 5/12/17**
74			**4.5**	**Final Report**	**95 days**	**Sun 2/5/17**	**Wed 5/10/17**
83			**4.6**	**Web Site**	**7 days**	**Thu 5/4/17**	**Wed 5/10/17**
85			4.7	Final Presentations	0 days	Fri 5/12/17	Fri 5/12/17
86			4.8	Models Developed	7 days	Mon 4/24/17	Sun 4/30/17
87			4.9	Deliverables Complete	0 days	Fri 5/12/17	Fri 5/12/17

Page 1

Earned Value Management

Rate	50
Overhead	106.383
Week	Planned Hours
Week 1	9
Week 2	15
Week 3	40
Week 4	30
Week 5	52
Week 6	52
Week 7	52
Week 8	40
Week 9	40
Week 10	45
Week 11	45
Week 12	25
Week 13	25
Week 14	40
Week 15	40
Week 16	20





Week	Week Starting	Planned Budg	Cumulative Bud	Actual Spent	Cumulative Sp	Planned Cost	Cumulative	Cost Variance	Schedule Variar	CPI	SPI
Week 1	23-Jan	$ 957.45	$ 957.45	$ 638.30	$ 638.30	$ 957.45	$ 957.45	$ 319.15	$ -	1.50	1.00
Week 2	30-Jan	$ 1,595.75	$ 2,553.19	$ 1,462.77	$ 2,101.06	$ 1,595.75	$ 2,553.19	$ 132.98	$ -	1.22	1.00
Week 3	6-Feb	$ 4,255.32	$ 6,808.51	$ 6,489.36	$ 8,590.43	$ 4,255.32	$ 6,808.51	$ (2,234.04)	$ -	0.79	1.00
Week 4	13-Feb	$ 3,191.49	$ 10,000.00	$ 3,909.58	$ 12,500.00	$ 3,191.49	$10,000.00	$ (718.09)	$ -	0.80	1.00
Week 5	20-Feb	$ 5,531.92	$ 15,531.92	$ 4,414.89	$ 16,914.90	$ 5,106.38	$15,106.39	$ 691.49	$ (425.53)	0.89	0.97
Week 6	27-Feb	$ 5,531.92	$ 21,063.83	$ 6,223.41	$ 23,138.30	$ 4,787.24	$19,893.62	$ (1,436.17)	$ (744.68)	0.86	0.94
Week 7	6-Mar	$ 5,531.92	$ 26,595.75	$ 4,255.32	$ 27,393.62	$ 7,978.73	$27,872.35	$ 3,723.41	$ 2,446.81	1.02	1.05
Week 8	13-Mar	$ 4,255.32	$ 30,851.07	$ 3,882.98	$ 31,276.60	$ 9,042.56	$36,914.90	$ 5,159.58	$ 4,787.24	1.18	1.20
Week 9	20-Mar	$ 4,255.32	$ 35,106.39	$ 3,776.60	$ 35,053.20	$ 4,787.24	$41,702.14	$ 1,010.64	$ 531.92	1.19	1.19
Week 10	27-Mar	$ 4,787.24	$ 39,893.63	$ 3,404.26	$ 38,457.45	$ 1,595.75	$43,297.88	$ (1,808.51)	$ (3,191.49)	1.13	1.09
Week 11	3-Apr	$ 4,787.24	$ 44,680.86	$ 4,095.75	$ 42,553.20	$ 3,723.41	$47,021.29	$ (372.34)	$ (1,063.83)	1.11	1.05
Week 12	10-Apr	$ 2,659.58	$ 47,340.44	$ 4,787.24	$ 47,340.44	$ 2,659.58	$49,680.86	$ (2,127.66)	$ -	1.05	1.05
Week 13	17-Apr	$ 2,659.58	$ 50,000.01	$ 6,010.64	$ 53,351.07	$ 3,723.41	$53,404.27	$ (2,287.23)	$ 1,063.83	1.00	1.07
Week 14	24-Apr	$ 4,255.32	$ 54,255.33	$ 7,659.58	$ 61,010.65	$ 5,851.07	$59,255.33	$ (1,808.51)	$ 1,595.75	0.97	1.09
Week 15	1-May	$ 4,255.32	$ 58,510.65	$ 6,382.98	$ 67,393.63	$ 6,914.90	$66,170.23	$ 531.92	$ 2,659.58	0.98	1.13
Week 16	8-May	$ 2,127.66	$ 60,638.31	$ -	$ 67,393.63			$ -	$ (2,127.66)	0.00	0.00

APPENDIX D - DATA DICTIONARIES

Table 8 Global Fishing Watch Data Dictionary

Attribute	Description
Mmsi	Vessel Identification
Timestamp	Time in UTC (seconds)
distance_from_shore	Haversine distance from point to shoreline; data provided by Natural Earth [5]
distance_from_port	Haversine distance from point to port; data provided by Natural Earth [5]
Speed	AIS reported speed
Course	AIS reported course
Lat	AIS reported latitude
Lon	AIS reported longitude
is_fishing	Classification of the data point 0 = Not Fishing 1 = Fishing -1 = Not Labeled

Table 9 Added Measures Data Dictionary

Attribute	Description
measure_course	Normalized course; course / 360.0
measure_cos_course	cos(course) / sqrt(2)
measure_sin_course	sin(course) / sqrt(2)
measure_courseavg_(window)	rolling average of measure_course using the specified window
measure_coursestddev_(window)	sum over the window; stddev(measure_cos_course) + stddev(measure_sin_course)
measure_coursestddev_(window)_log	EPSILON = 1e-3 log10(measure_coursestddev + EPSILON)
measure_speed	1.0 - min(1.0, speed / 17.0)
measure_speedavg_(window)	average of measure_speed over the window
measure_speedstddev_(window)	stddev of measure_speed over the window
measure_speedstddev_(window)_log	EPSILON = 1e-3 log10(measure_speedstddev + EPSILON)
measure_pos_(window)	sum over the window; stddev(lat) + stddev(lon)
measure_latavg_(window)	average of the latitude over the window
measure_lonavg_(window)	average of the longitude over the window
measure_count_(window)	number of datapoints in the window
measure_daylight	0 = before noon local time; 1 = after noon local time
measure_daylightavg_(window)	average of measure_daylight over the window

APPENDIX E - MODEL COEFFICIENTS

Longliner Logistic Regression Model Feature Coefficients

Feature	Coefficient
Intercept	[-0.0709763585979]
measure_courseavg_10800	[-0.310638360677]
measure_courseavg_1800	[-1.26855417735]
measure_courseavg_21600	[0.507362025251]
measure_courseavg_3600	[0.10398749559]
measure_courseavg_43200	[-0.62838968908]
measure_courseavg_86400	[0.717604065818]
measure_coursestddev_10800	[-0.772046713269]
measure_coursestddev_1800	[0.483546013088]
measure_coursestddev_21600	[0.661828446134]
measure_coursestddev_3600	[-5.13415260403]
measure_coursestddev_43200	[1.54206315925]
measure_coursestddev_86400	[7.25709989241]
measure_pos_10800	[-6.07633115579]
measure_pos_1800	[-3.23712211457]
measure_pos_21600	[1.35246345489]
measure_pos_3600	[-5.73965118962]
measure_pos_43200	[-2.59047628565]
measure_pos_86400	[-0.0239622217355]
measure_speedavg_10800	[-0.677796517851]
measure_speedavg_1800	[-0.930341079527]
measure_speedavg_21600	[3.5504488374]
measure_speedavg_3600	[0.403979655647]
measure_speedavg_43200	[-1.78554989866]
measure_speedavg_86400	[-2.883236906]
measure_speedstddev_10800	[4.05860157298]
measure_speedstddev_1800	[8.49517791511]
measure_speedstddev_21600	[6.61554102422]
measure_speedstddev_3600	[8.0539460381]
measure_speedstddev_43200	[1.54456490253]
measure_speedstddev_86400	[-3.81445360924]

Trawler Logistic Regression Model Feature Coefficients

Feature	Coefficient
Intercept	[0.122378159662]
measure_courseavg_10800	[-0.371817739163]
measure_courseavg_1800	[-0.327048906864]
measure_courseavg_21600	[-0.340147614152]
measure_courseavg_3600	[0.0695316805656]
measure_courseavg_43200	[0.82493037292]
measure_courseavg_86400	[-2.06590151948]
measure_coursestddev_10800	[0.113123912787]
measure_coursestddev_1800	[-0.998654517579]
measure_coursestddev_21600	[1.23564316028]
measure_coursestddev_3600	[-1.47168785955]
measure_coursestddev_43200	[1.98460876768]
measure_coursestddev_86400	[4.82410498093]
measure_pos_10800	[0.0584962969648]
measure_pos_1800	[0.0602282017262]
measure_pos_21600	[0.255553041308]
measure_pos_3600	[0.00752593758605]
measure_pos_43200	[0.0656509574289]
measure_pos_86400	[-0.692917635983]
measure_speedavg_10800	[-3.03165961928]
measure_speedavg_1800	[-0.687794978538]
measure_speedavg_21600	[5.44965844929]
measure_speedavg_3600	[-0.253625285309]
measure_speedavg_43200	[2.62251481962]
measure_speedavg_86400	[-7.4425260421]
measure_speedstddev_10800	[-0.596941032847]
measure_speedstddev_1800	[5.01508713974]
measure_speedstddev_21600	[-0.028648416638]
measure_speedstddev_3600	[2.53510341058]
measure_speedstddev_43200	[0.0422633750794]
measure_speedstddev_86400	[-2.68160804699]

Purse Seine Logistic Regression Model Feature Coefficients

Feature	Coefficient
Intercept	[1.52920900439]
measure_courseavg_10800	[2.91808240418]
measure_courseavg_1800	[0.647206755472]
measure_courseavg_21600	[-0.516238678091]
measure_courseavg_3600	[-2.86605017966]
measure_courseavg_43200	[0.741876180077]
measure_courseavg_86400	[0.172226803733]
measure_coursestddev_10800	[-2.04435534195]
measure_coursestddev_1800	[-3.65673675547]
measure_coursestddev_21600	[1.64843703349]
measure_coursestddev_3600	[-1.8254731408]
measure_coursestddev_43200	[0.917804274474]
measure_coursestddev_86400	[-1.29320977568]
measure_pos_10800	[4.77542626957]
measure_pos_1800	[-2.30291485481]
measure_pos_21600	[-13.4899512348]
measure_pos_3600	[-0.687327594403]
measure_pos_43200	[-1.9595984517]
measure_pos_86400	[-0.877992596922]
measure_speedavg_10800	[-7.97726932171]
measure_speedavg_1800	[0.101202610291]
measure_speedavg_21600	[4.2639035885]
measure_speedavg_3600	[1.86327969681]
measure_speedavg_43200	[-2.74999262585]
measure_speedavg_86400	[0.596345108224]
measure_speedstddev_10800	[0.546900619392]
measure_speedstddev_1800	[-4.15947910381]
measure_speedstddev_21600	[6.35174570614]
measure_speedstddev_3600	[4.70268339437]
measure_speedstddev_43200	[-3.36176153512]
measure_speedstddev_86400	[6.06718858569]