“Illegal, Unreported, Unregulated”

IUU Fishing Detection

Jarred Byrnes
Jonathan Matteson
Edward Kerrigan
Jonathan Gessert

Link to presentation on Google Slides:
https://docs.google.com/presentation/d/16EigEHtQt8Hmfu1er4OkoMwVAMGN1Its1HCmTt4TiWM/edit?usp=sharing

Useful Definitions

Illegal, Unreported, Unregulated Fishing

IUU Fishing is fishing that is Illegal, Unreported, Unregulated

Marine Protected Areas

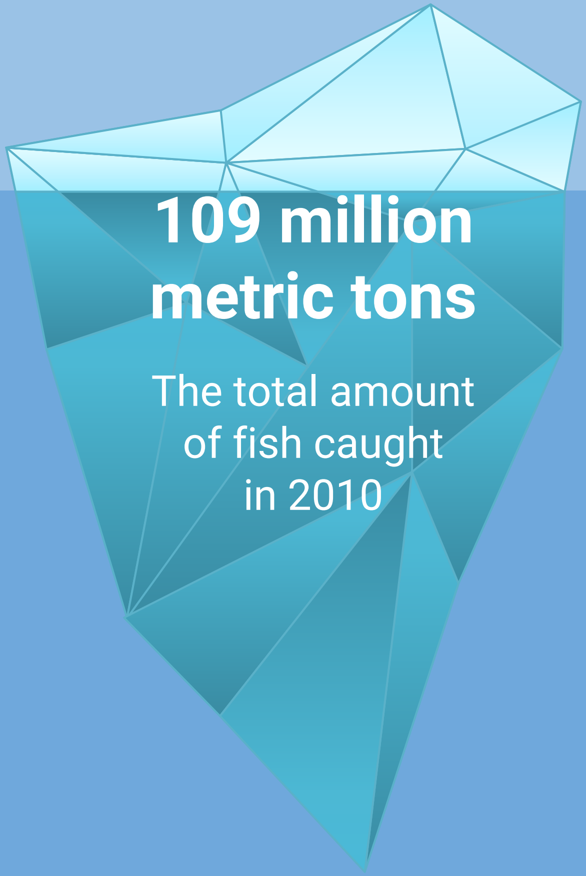
MPAs are protected bodies of water, which restrict human activity to protect natural or cultural resources.

Exclusive Economic Zone

EEZs are sea zones where a governing state has special rights regarding the exploration and use of marine resources.

Regional Fishery Mgmt. Orgs.

RFMOs are international organisations formed by countries to monitor and/or regulate fishing in areas of interest.



77 million metric tons
The reported amount of fish caught in 2010

32 million metric tons
The unreported amount of fish caught in 2010

$10 - $23.5 billion annually
The estimated loss of income to coastal countries and communities caused by IUU fishing in 2009

Illegal, Unreported, and Unregulated (IUU) fishing is unsustainable.

It harms the ecosystem and global economy.

USCG IUU Fishing Countermeasures

- During peak fishing season, huge fleets of foreign vessels encroach upon the US EEZ boundary line from countries of origin such as Russia, Japan, Poland, China and Taiwan.
- The US Coast Guard enforces US EEZ regulations through physically patrolling the boundary line.
- Patrolling encompasses daily C-130 flights, continuous USCG cutter presence, and patrol boats.

Background

Problem & Scope

Methodology/
Technical Approach

- Research
- Systems Modeling
- Data Analytics

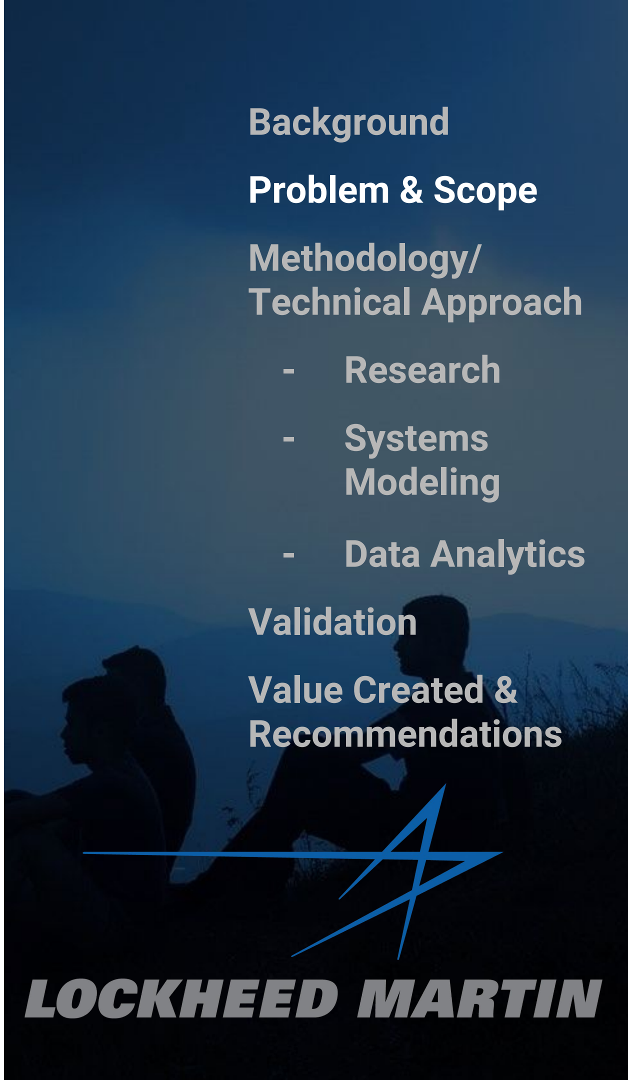
Validation

Value Created &
Recommendations

Problem

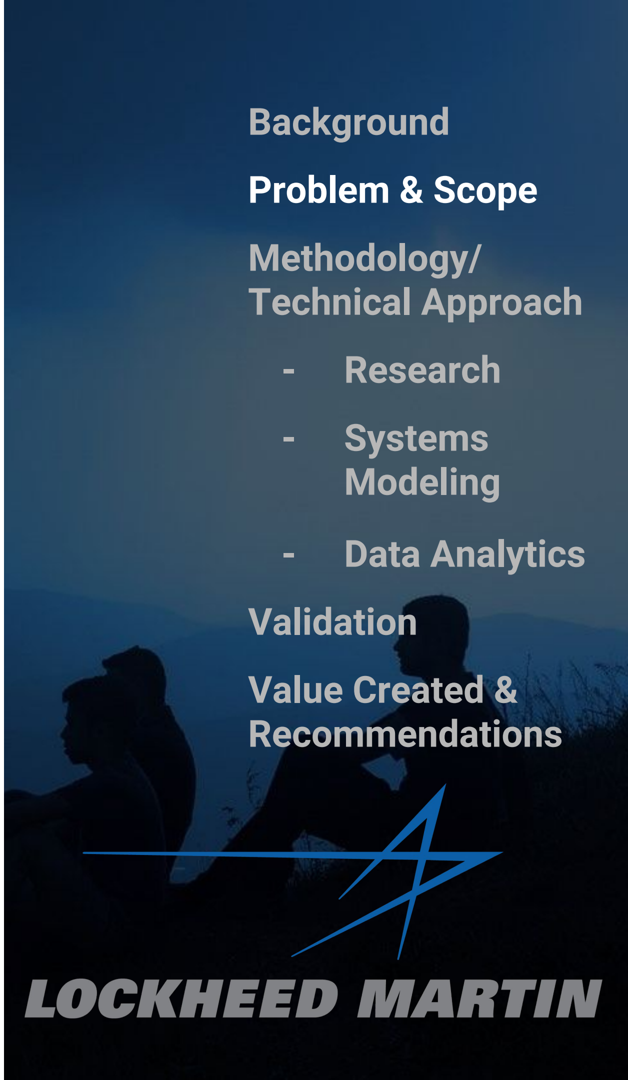
→ Preventing IUU fishing with physical patrolling, investigation, and search and seizure by law enforcement is an expensive and time-consuming process

→ Using geospatially referenced, physics-based sensor intelligence and data analytics techniques, Lockheed Martin sees an opportunity to contribute a solution to IUU fishing detection and enforcement.



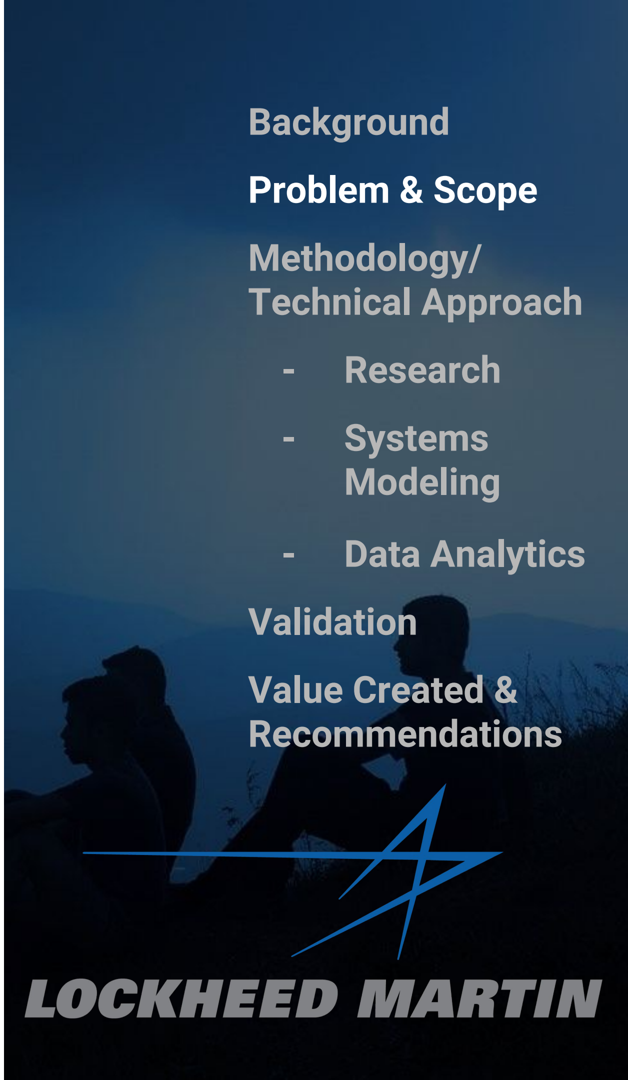
Scope

- Use vessel sensor data to research, develop, and refine an approach to IUU fishing detection
- Develop a series of descriptive and predictive models progressing towards an IUU fishing detection model



Planned Start with Sponsor

- A scheduled weekly meeting with the sponsor
- A fundamental goal to advance data analytics efforts in IUU detection
- Creative freedom to present potential solutions



Methodology / Technical Approach

research → systems modeling → data analytics lifecycle

research

- relevant code repositories
- relevant research
- training data sources
- vessel intelligence
- relevant regulations
- vessel fishing behavior
- IUU fishing behavior

systems modeling

- fishing behavior architecture
- IUU fishing use cases and state diagrams
- IUU fishing architecture

data analytics lifecycle

- data preparation
- descriptive analytics
- predictive modeling
- model validation

Background

Problem & Scope

Methodology/ Technical Approach

- Research
- Systems Modeling
- Data Analytics

Validation

Value Created & Recommendations

[1]Improving Fishing Pattern Detection from Satellite AIS Using Data Mining and Machine Learning

Systems Modeling

- A series of system engineering models were developed to build a framework for which the data analytics model could be built
- Use Case diagrams were developed to define which scenarios the analytical model would examine
- Activity and State Machine diagrams provided in-depth, step-by-step activities and states the model would investigate to detect IUU Fishing
- A high level architecture diagram shows the overarching schema of the system

Background

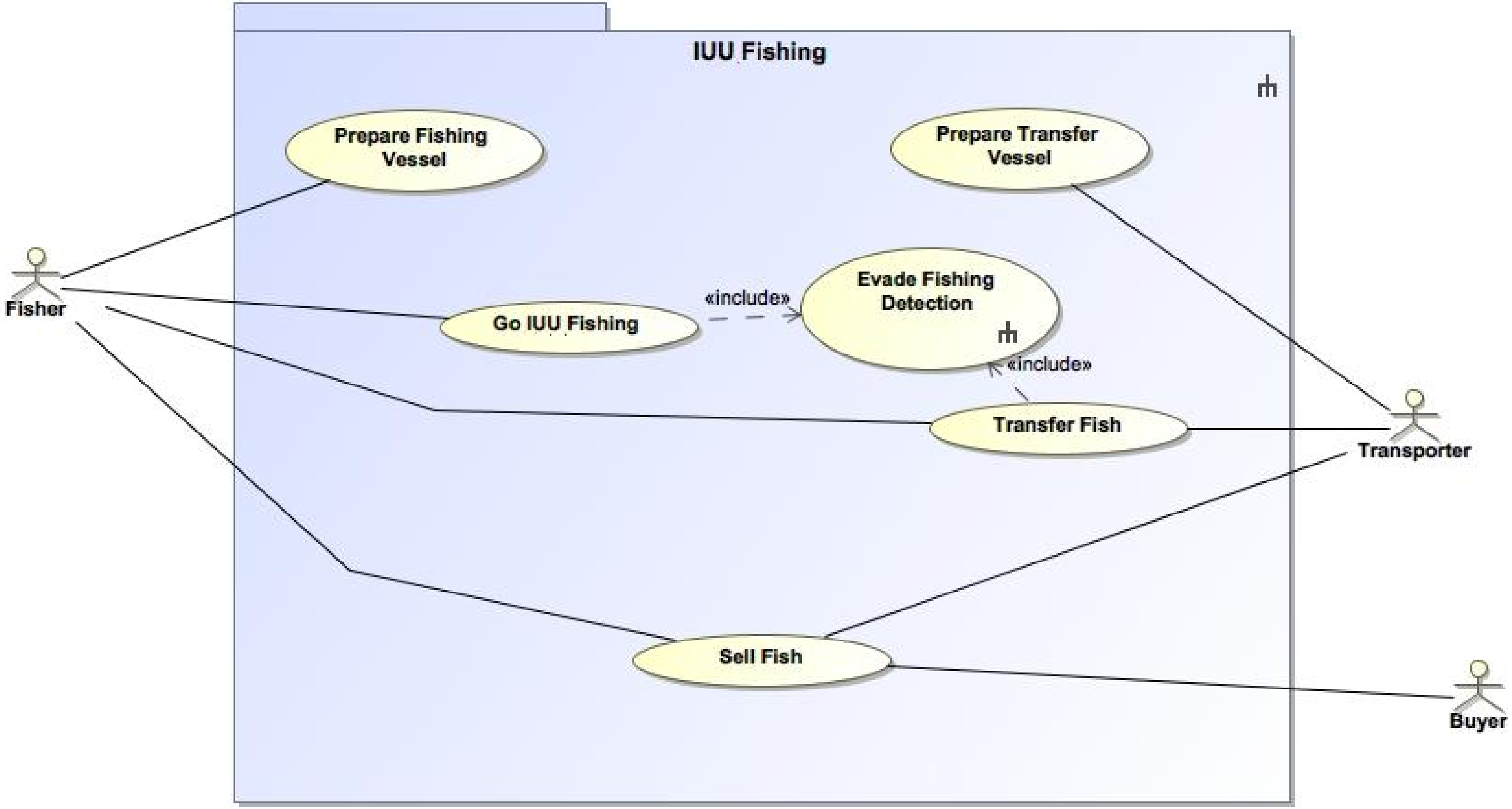
Problem & Scope

Methodology/
Technical Approach

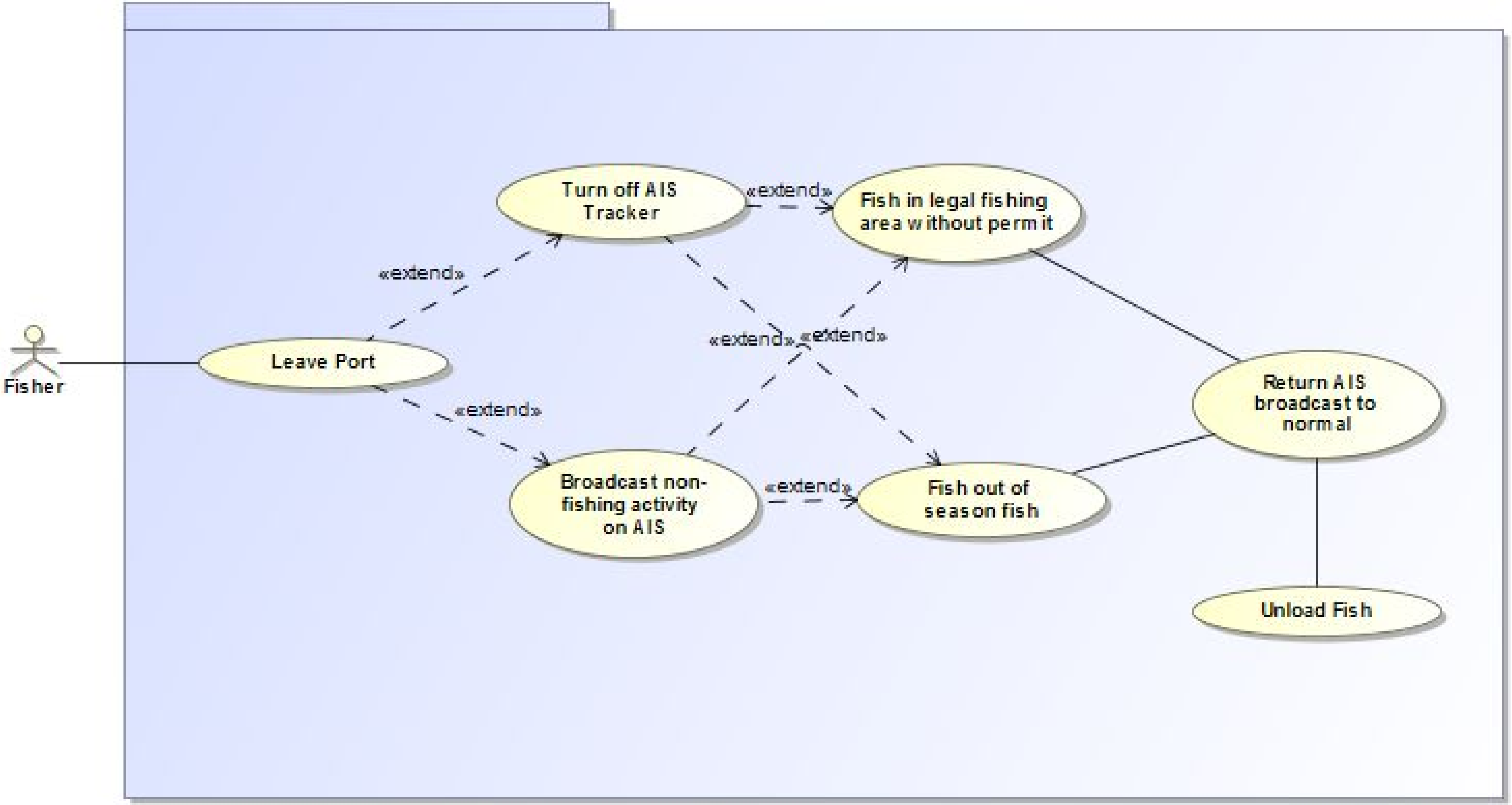
- Research
- Systems Modeling
- Data Analytics

Validation

Value Created &
Recommendations

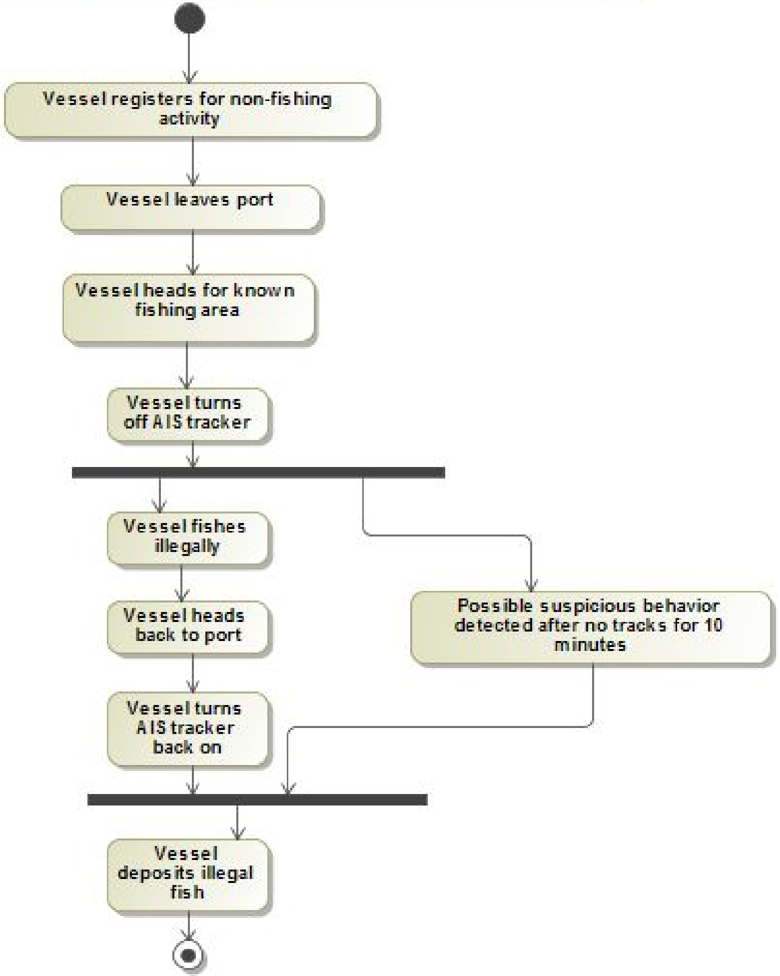


IUU Fishing Use Case Diagrams

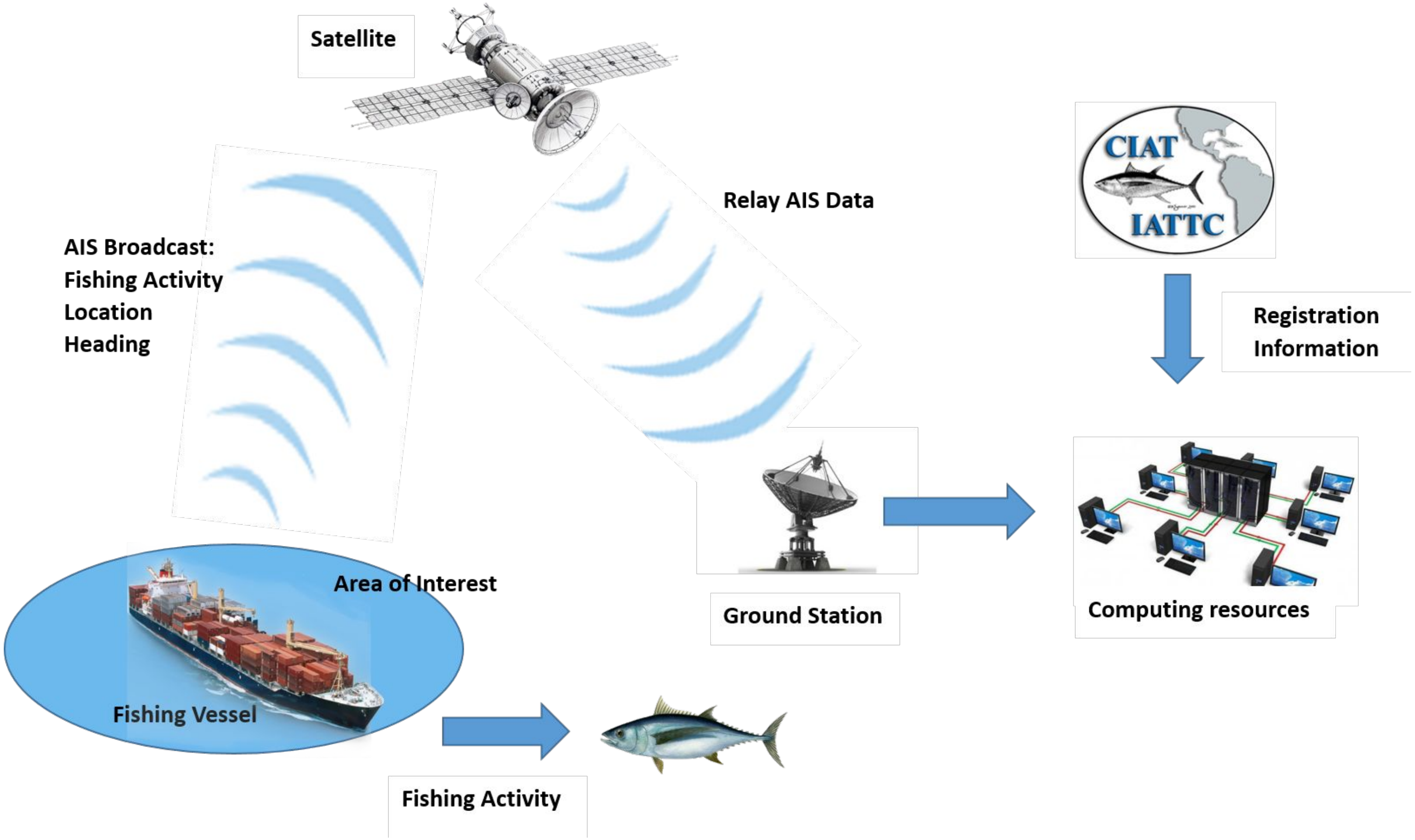


IUU Fishing Use Case Diagrams

- Turns off AIS
- Spoofing AIS
- Fishing Out of Season
- Vessels Too Close
- Illegal Sale of Fish



IUU Fishing Activity Diagrams



IUU Fishing High Level Concept of Operation

Limitations

→ Validated training data was limited to data which was hand labeled by Kristina Boerder

→ No access to IUU fishing subject matter experts

→ Missing items referred to in Global Fishing Watch's training-data and vessel-scoring repositories

- Datasets
- Dataset documentation
- Code
- Code documentation

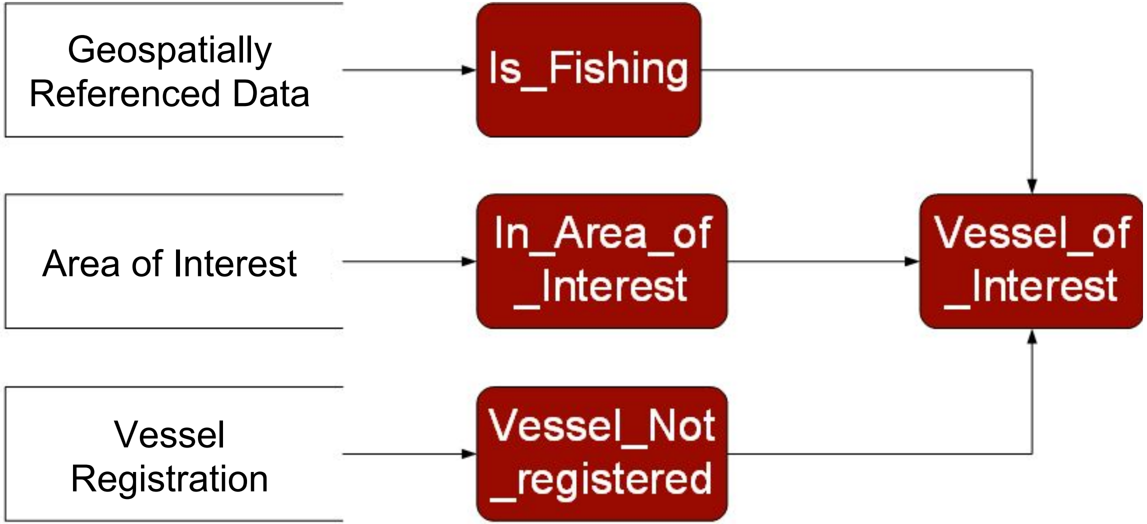
Background
Problem & Scope
Methodology/ Technical Approach
- Research
- Systems Modeling
- Data Analytics
Validation
Value Created & Recommendations

Assumptions

- Not all vessels turn off their AIS transmission while IUU fishing
- Models using AIS data are data source agnostic, and can be used with other signal intelligence providing similar data
- Vessels with similar fishing gear exhibit the same fishing behaviour regardless of the following:
  - Legal vs. IUU fishing
  - Location of operation

Background
Problem & Scope
Methodology/ Technical Approach
- Research
- Systems Modeling
- Data Analytics
Validation
Value Created & Recommendations

Data Model



Background

Problem & Scope

Methodology/ Technical Approach

- Research
- Systems Modeling
- Data Analytics

Validation

Value Created & Recommendations

Data Analytics

Lifecycle

- → Data Preparation
- → Descriptive Analytics
- → Predictive Modeling
- → Model Validation

Tools Used

- → Python 2.7
- → Jupyter Notebook / iPython
- → SciPy
- → Pandas
- → MatPlotLib
- → SciKit Learn

Background
Problem & Scope
Methodology/ Technical Approach
- Research
- Systems Modeling
- Data Analytics
Validation
Value Created & Recommendations

Data Preparation

Data collection

→ GFW's training-data From GitHub Repo

Data exploration (see table)

Data validation

→ Cleaned data to work with GFW derivation Scripts

KB's Training Data	
Features	**KB's Source**
Timestamp	AIS data
Vessel ID	AIS data
Latitude	AIS data
Longitude	AIS data
Course	AIS data
Speed	AIS data
Distance from shore	Derived (KB)
Distance from port	Derived (KB)
Fishing flag	Self Classification (KB)

Background

Problem & Scope

Methodology/ Technical Approach

- Research
- Systems Modeling
- **Data Analytics**

Validation

Value Created & Recommendations

Descriptive Analytics Performed

- → Descriptive statistics
- → Univariate analysis
- → Visualizations and insights
- → Created derived data

Background

Problem & Scope

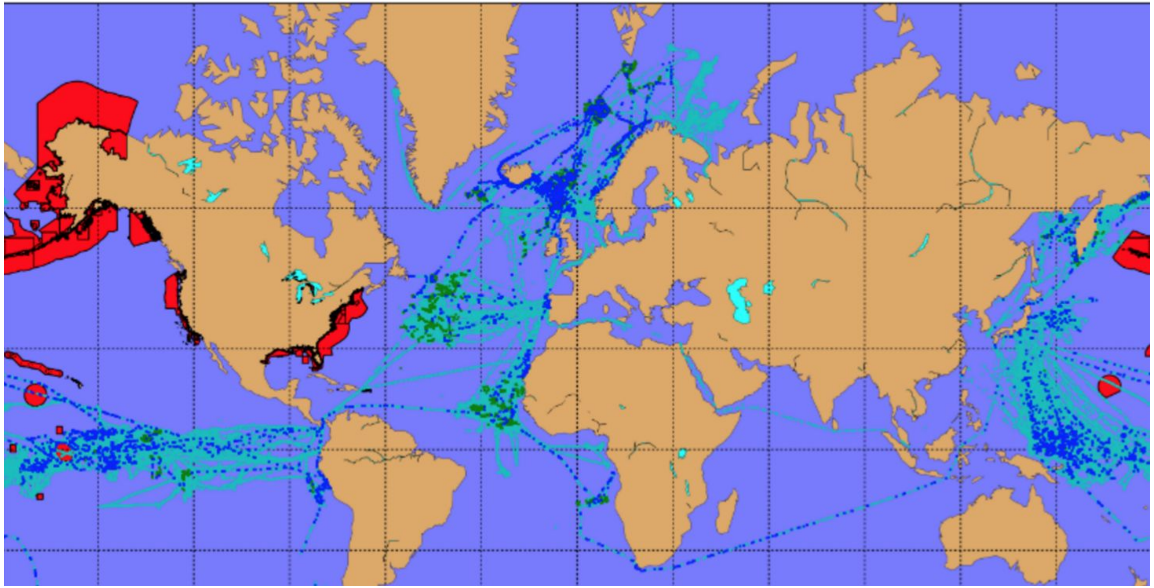
Methodology/ Technical Approach

- Research
- Systems Modeling
- Data Analytics

Validation

Value Created & Recommendations

Vessel Tracks and MPAs



Fishing
Unclassified
Not Fishing
US MPA

Background

Problem & Scope

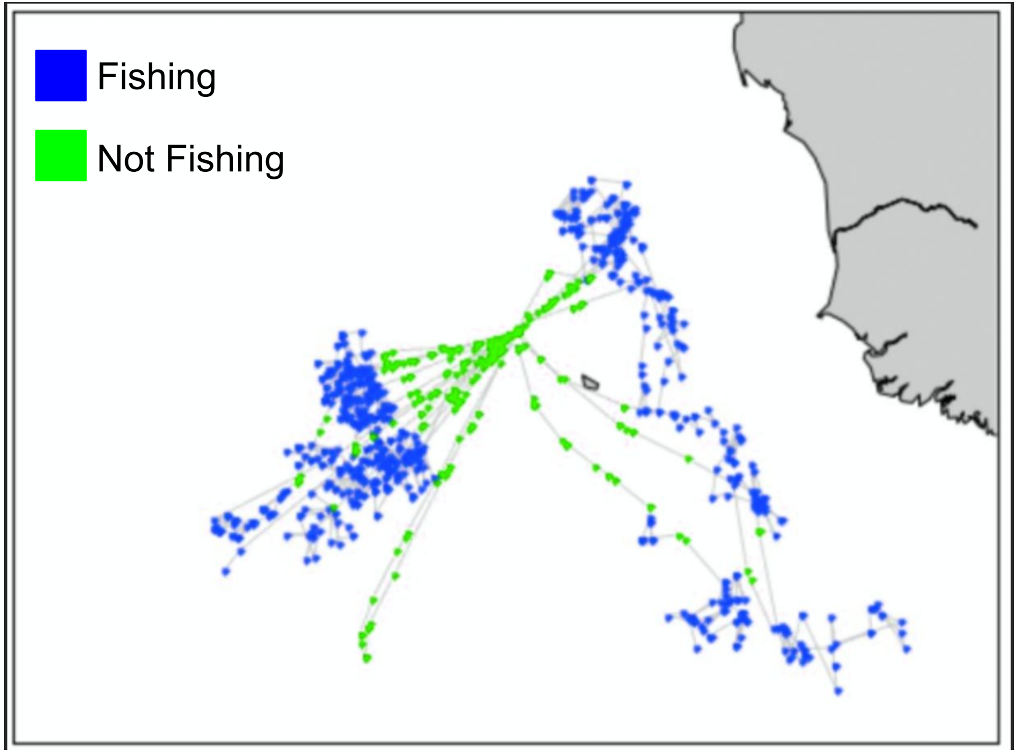
Methodology/
Technical Approach

- Research
- Systems Modeling
- Data Analytics

Validation

Value Created &
Recommendations

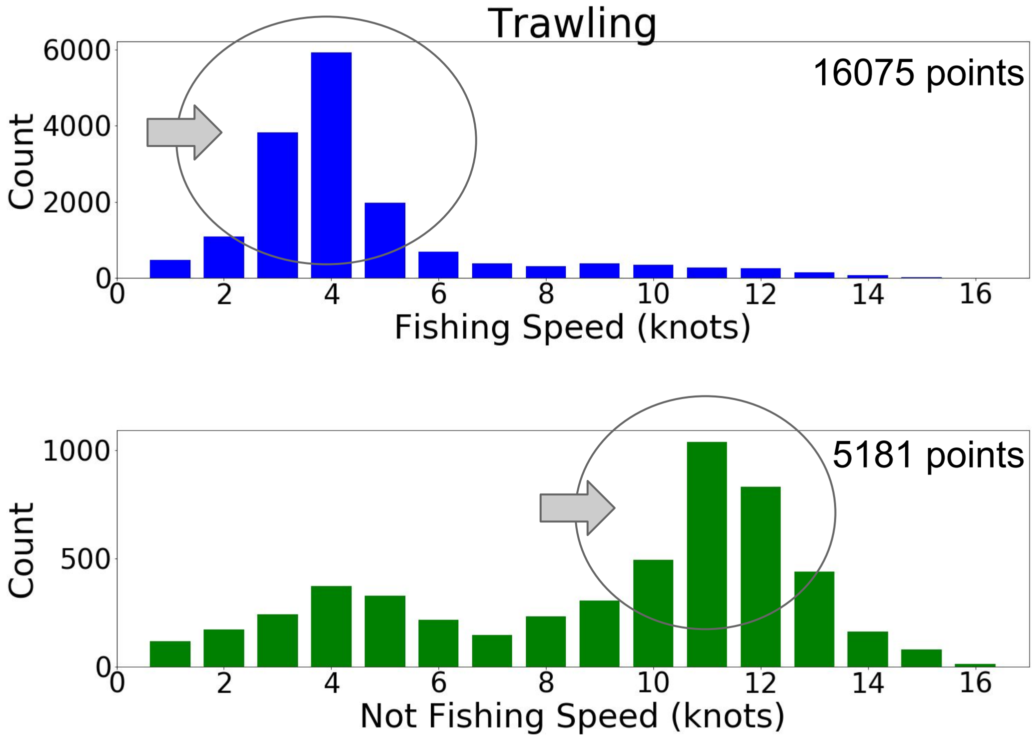
Ex. Course Deviation While Fishing



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

Background
Problem & Scope
Methodology/ Technical Approach
- Research
- Systems Modeling
- Data Analytics
Validation
Value Created & Recommendations

Ex. Speed Univariate Analysis



Background

Problem & Scope

Methodology/
Technical Approach

- Research
- Systems Modeling
- Data Analytics

Validation

Value Created & Recommendations

Creating Derived Data

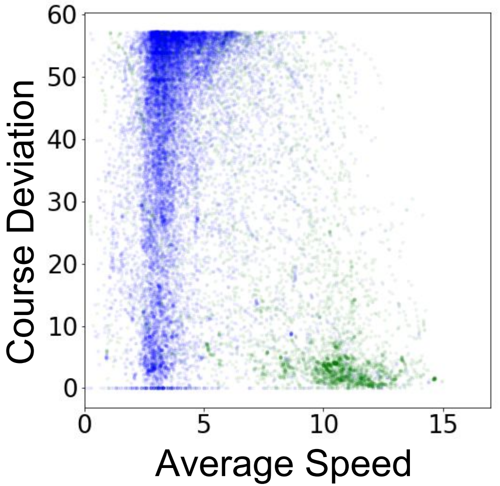
KB's Training Data	
Features	**KB's Source**
Timestamp	AIS data
Vessel ID	AIS data
Latitude	AIS data
Longitude	AIS data
Course	AIS data
Speed	AIS data
Distance from shore	Derived (KB)
Distance from port	Derived (KB)
Fishing flag	Self Classification (KB)

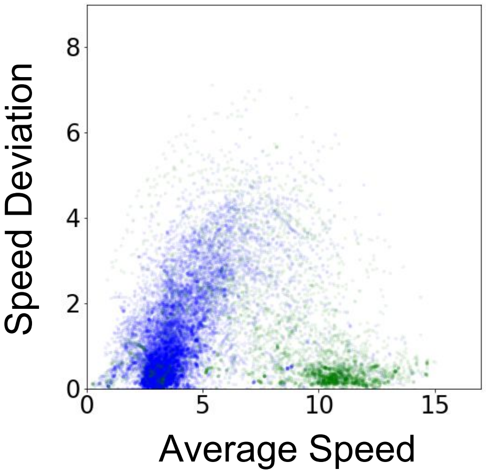
Training Data Features
Speed Deviation in Time Window
Normalized Speed in Time Window
Course Deviation in Time Window
Normalized Course in Time Window

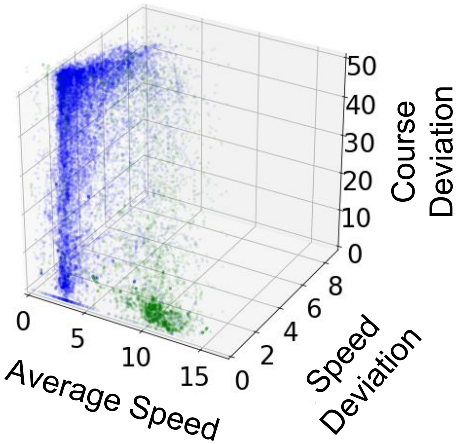
Time Windows	
Min	**Hours**
15	0.25
30	0.50
60	1
180	3
360	6
720	12
1440	24

Background
Problem & Scope
Methodology/ Technical Approach
- Research
- Systems Modeling
- Data Analytics
Validation
Value Created & Recommendations

Trawler Data View @ 6 hour Window







Fishing

Not Fishing

Background

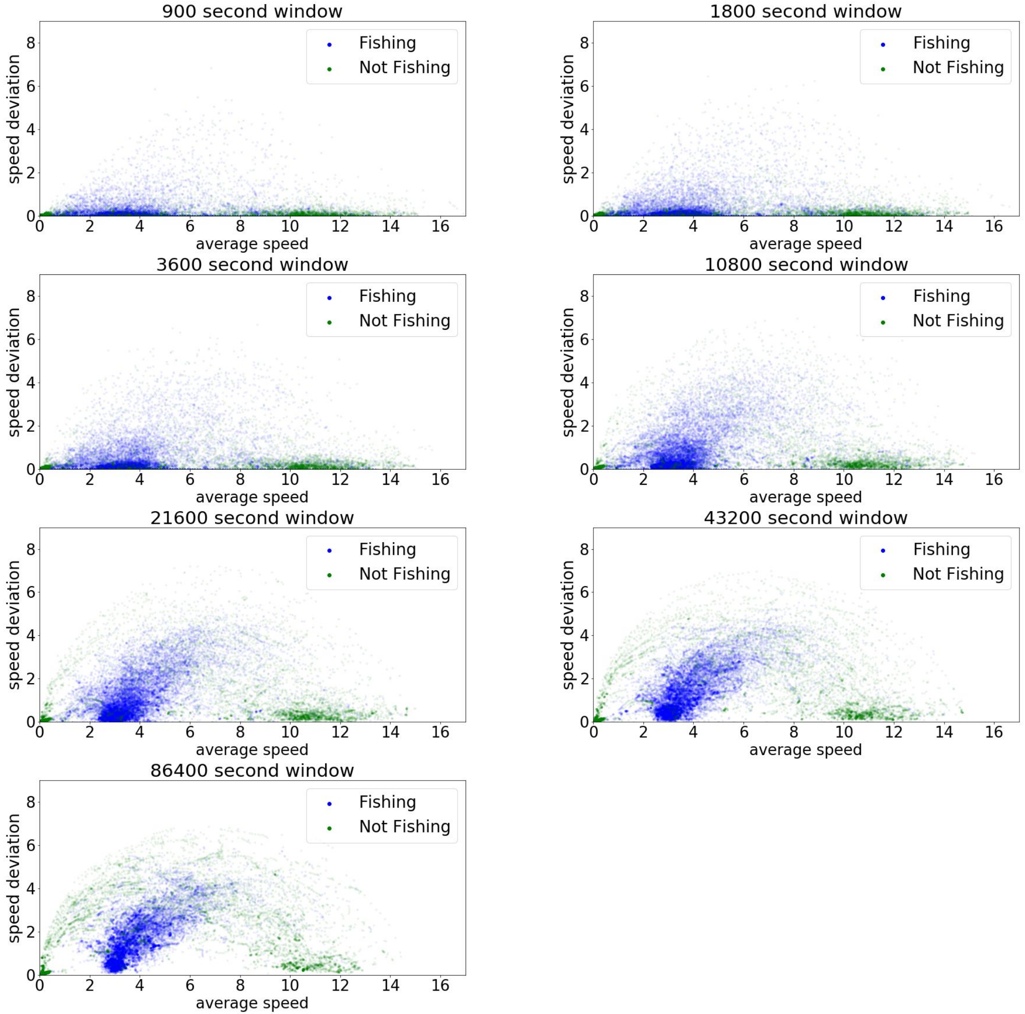
Problem & Scope

Methodology/ Technical Approach

- Research
- Systems Modeling
- Data Analytics

Validation

Value Created & Recommendations



Background

Problem & Scope

Methodology/ Technical Approach

- Research
- Systems Modeling
- **Data Analytics**

Validation

Value Created & Recommendations

Definitions for Validation

Term	Definition
Accuracy	Percentage of Correct Predictions
Precision	Percentage of Correct Positive (Fishing) Predictions
Recall	Percentage of Positive (Fishing) Predictions found
Receiver Operating Characteristic (ROC) Curve	True Positive Rate vs. False Positive Rate Used to determine predictiveness of model using percentage of Area Under Curve.
Precision-Recall (PR) Curve	Precision vs. Recall Used to determine predictiveness of correct positive predictions using percentage of Area Under Curve. Best used for cases of class imbalance.

Background

Problem & Scope

Methodology/ Technical Approach

- Research
- Systems Modeling
- **Data Analytics**

Validation

Value Created & Recommendations

Predictive Modeling

Modeling Technique Identification: Logistic Regression

- Variable was discrete with two classification values
- GFW's documentation identified the following as the most predictive model:
  Logistic regression over multiple time windows and individual gear types

Model Validation

- Accuracy, Precision and Recall
- ROC Curve & Precision-Recall Curve
- k-Fold Cross-Validation

Background
Problem & Scope
Methodology/ Technical Approach
- Research
- Systems Modeling
- Data Analytics
Validation
Value Created & Recommendations

Predictive Analytics Process Used

1. Split into training and test data sets
   - 70% and 30% respectively
2. Instantiate a logistic regression model and fit to training data
3. Evaluate model using testing data
   - Accuracy Score
   - Precision Score
   - ROC & PR AUC Scores
4. Evaluate model accuracy with 10-fold cross-validation

Background

Problem & Scope

Methodology/
Technical Approach

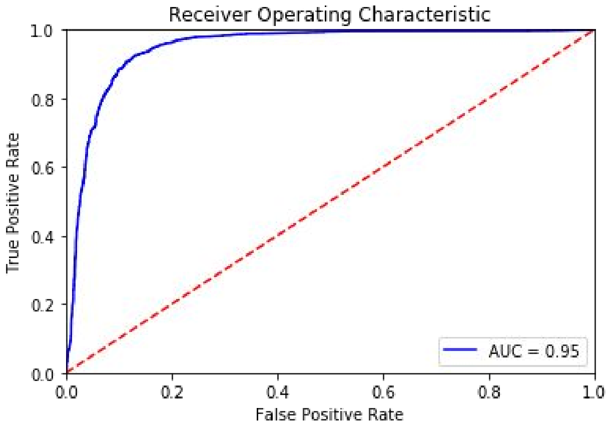
- Research
- Systems Modeling
- Data Analytics

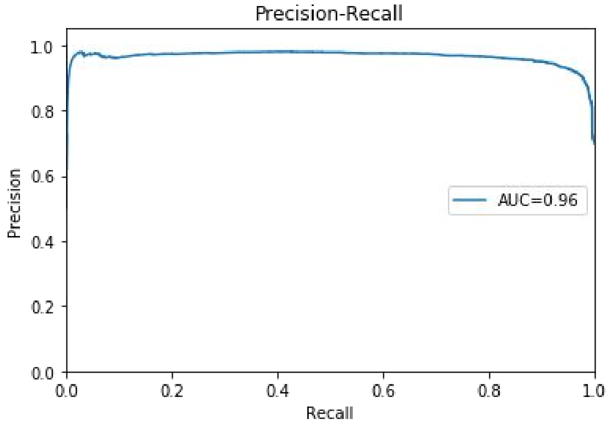
Validation

Value Created &
Recommendations

Predictive Model: Longliner

Name	Value
Pos. (Fishing) Values	9007
Neg. (Not Fishing) Values	3890
Accuracy	91.4%
Null Accuracy	69.8%
True Positive Precision	92%
True Negative Precision	91%
ROC AUC Score	94.7%
Precision-Recall AUC Score	96%
10-Fold Cross Validation Accuracy Scores Mean	90.6%





Background

Problem & Scope

Methodology/ Technical Approach

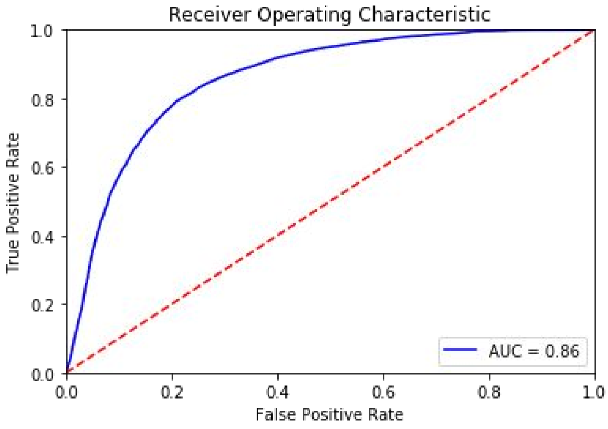
- Research
- Systems Modeling
- Data Analytics

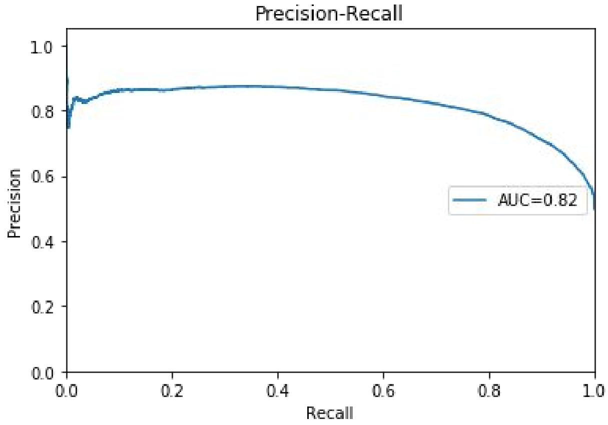
Validation

Value Created & Recommendations

Predictive Model: Trawler

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

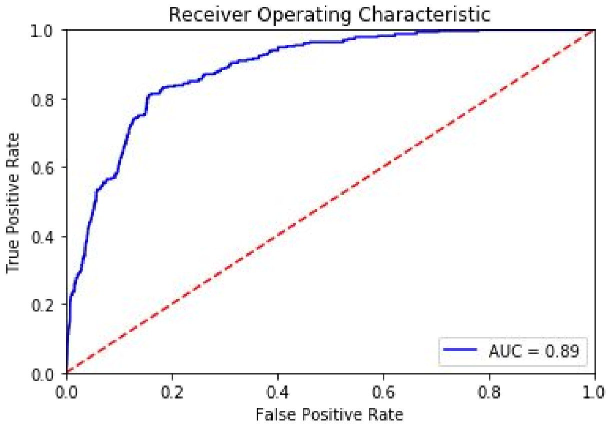


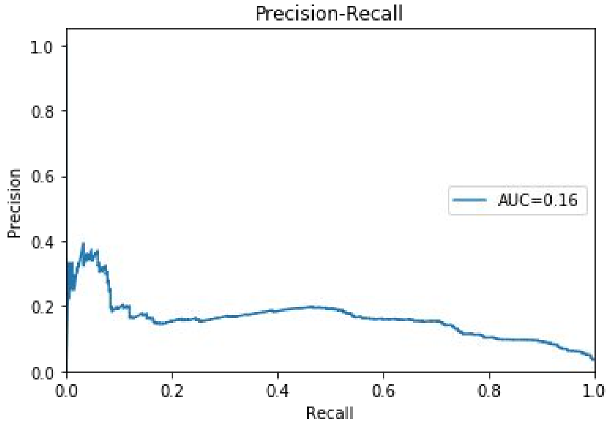


Background
Problem & Scope
Methodology/ Technical Approach
- Research
- Systems Modeling
- Data Analytics
Validation
Value Created & Recommendations

Predictive Model: Purse Seine

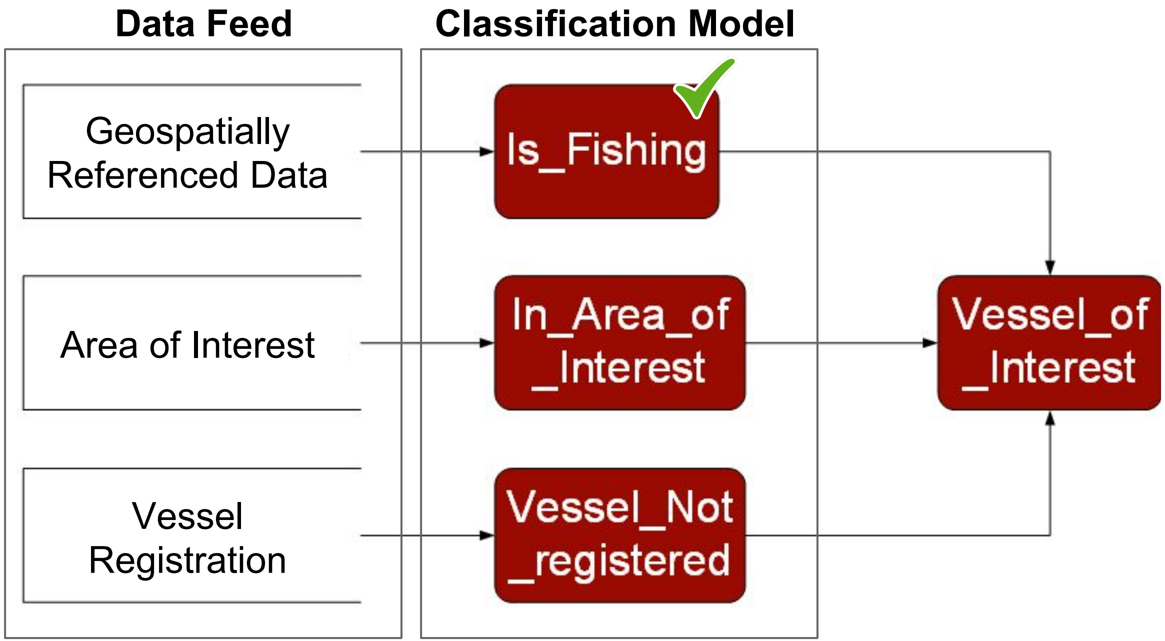
Name	Value
Pos. (Fishing) Values	333
Neg. (Not Fishing) Values	12254
Accuracy	97.3%
Null Accuracy	97.3%
True Positive Precision	27%
True Negative Precision	97%
ROC AUC Score	88.3%
Precision-Recall AUC Score	16%
10-Fold Cross Validation Accuracy Scores Mean	97.0%





Background
Problem & Scope
Methodology/ Technical Approach
- Research
- Systems Modeling
- Data Analytics
Validation
Value Created & Recommendations

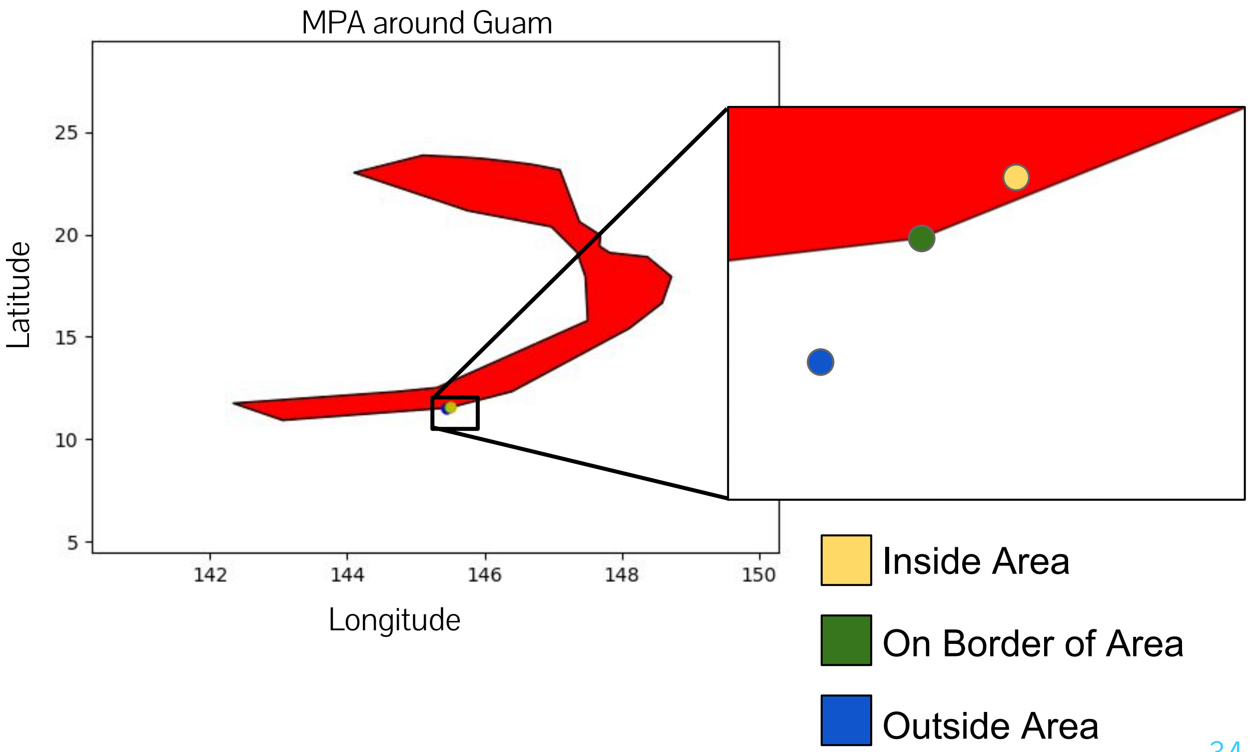
Scoring Model Revisited



Vessel of Interest Scoring Model

Background
Problem & Scope
Methodology/ Technical Approach
- Research
- Systems Modeling
- Data Analytics
Validation
Value Created & Recommendations

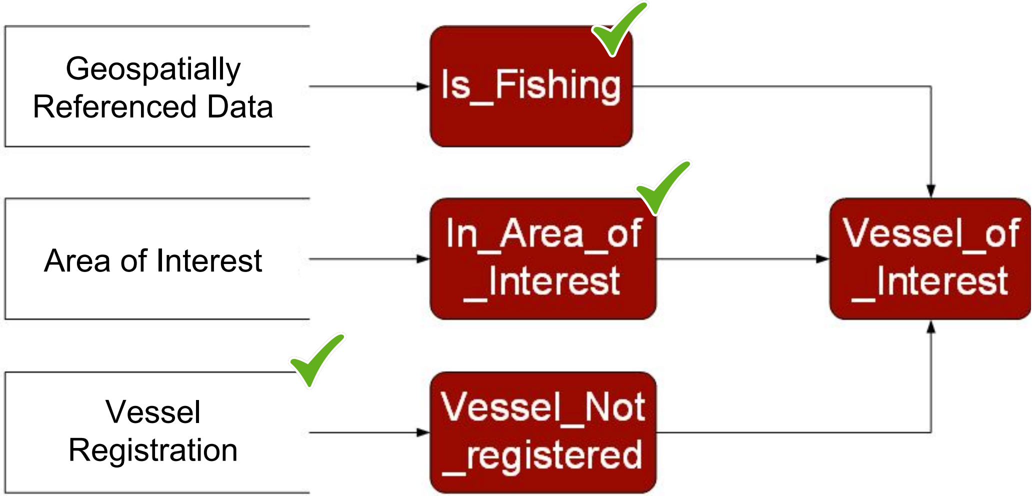
Ex. In_Area_of_Interest Code Output



Background
Problem & Scope
Methodology/ Technical Approach
- Research
- Systems Modeling
- Data Analytics
Validation
Value Created & Recommendations

Summary of Value Created

- IUU Fishing Research
- IUU Fishing Framework
- Partially Developed Vessel of Interest Scoring Model Prototype



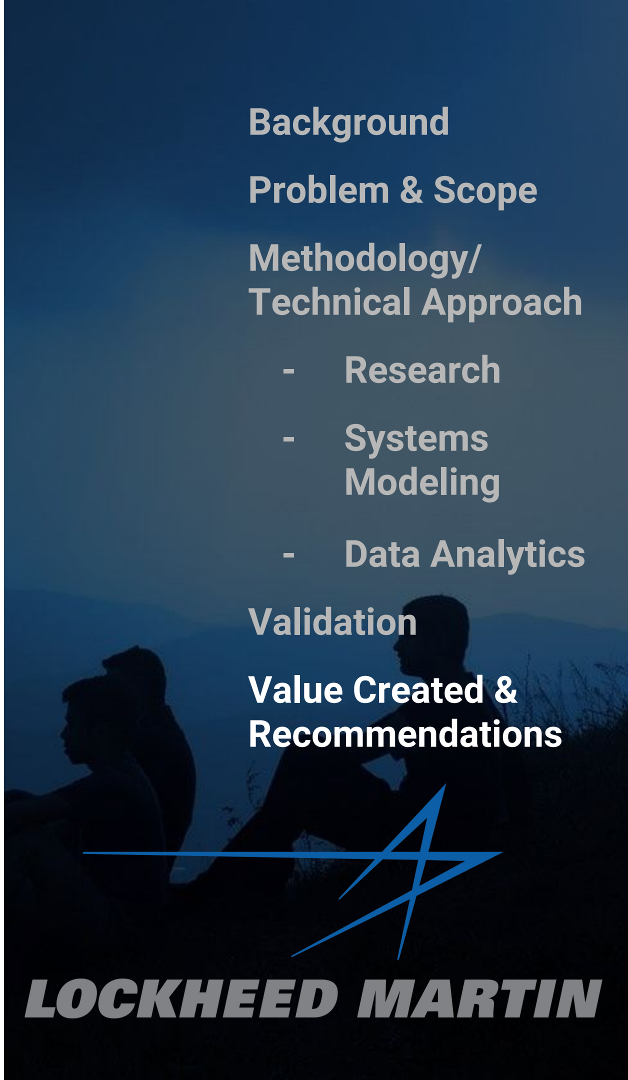
Vessel of Interest Scoring Model

Summary of Value Created (Cont'd)

Sponsor stated the following

→ Interested in applying concept of deriving time windows in predictive models used in other defense contracting research

→ Research work product and will be used, but there are restrictions on what can be discussed



Recommendations

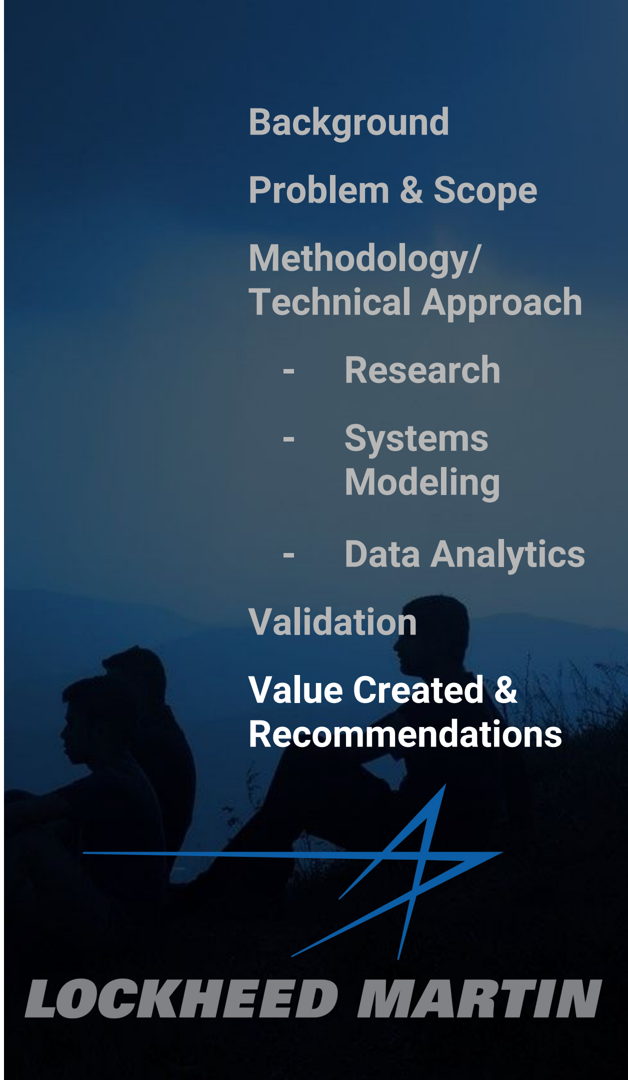
Primary Recommendations

Use validated work and continue development of Vessel of Interest Scoring Model Prototype

- → Improve Purse Seine Classification Model through data collection
- → Improve In-Area-of-Interest Scoring Model
- → Using data sources, create database or live data feed

Secondary Recommendations

- → Incorporate satellite imagery data as it becomes affordable
- → Predict probability of transshipment
- → Quantify suspicious behavior around port cities
  - Flag vessels visiting multiple ports after a fishing trip
  - Excess purchasing of ice



A Special Thanks

Sponsor

- David Cabelly
- Brian Hillanbrand
- John Luster
- Jonathan Brant
- Tim Parker

IUU Subject Matter Experts

- Craig Nilson
- Kristina Boerder
- Global Fishing Watch

GMU Mentors & Associates

- Dr. Laskey
- Class Peers
- Previous Instructors

LOCKHEED MARTIN

Thanks!

Any questions?

Backup slides

Useful Definitions: Gear Types

Trawler

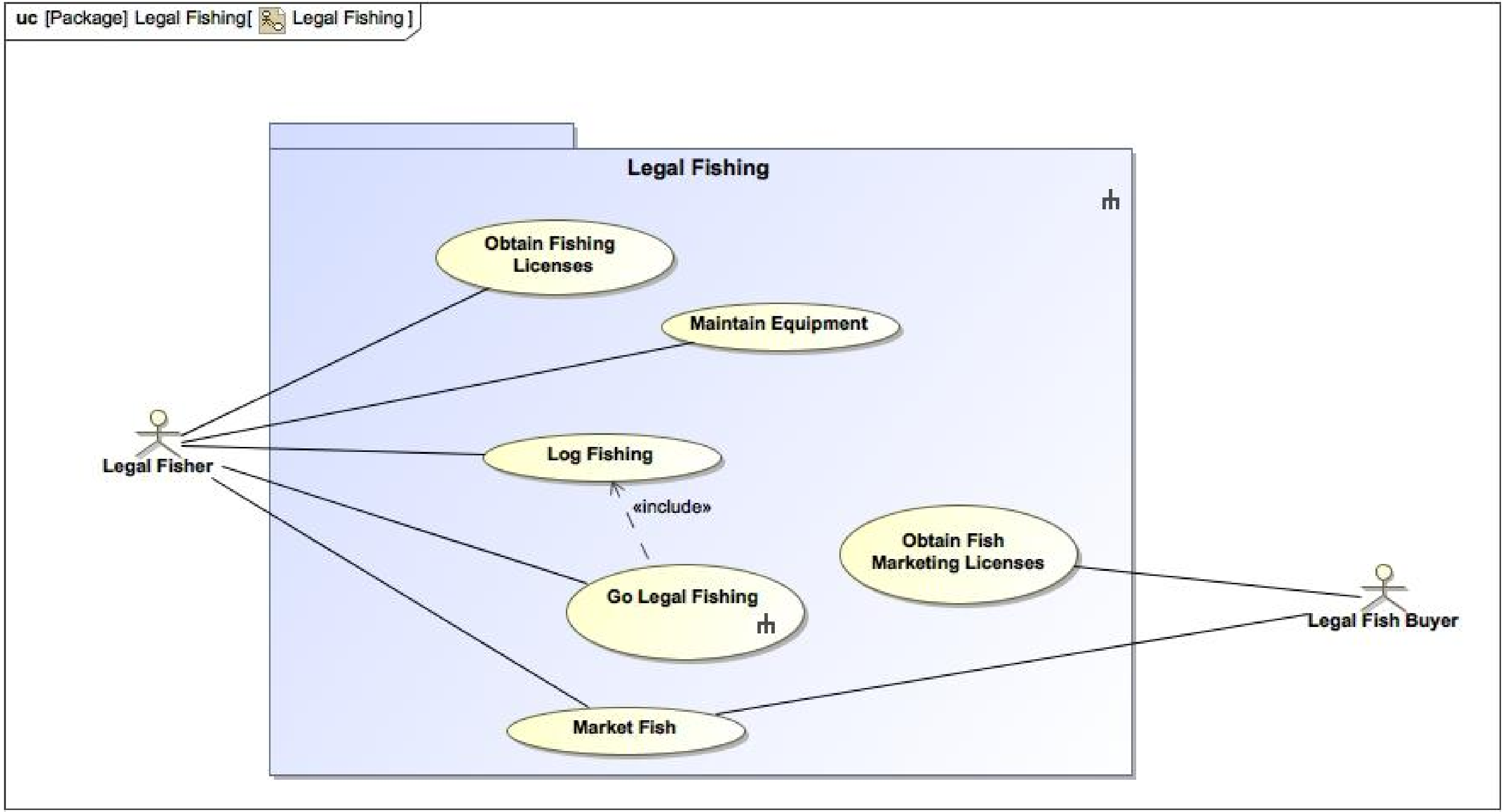
A trawling vessel captures fish by dragging a net behind the ship while moving at a very slow speed. These ships will typically fish from 3-5 hours at a time travelling at around 2.5-5.5 knots

Longliner

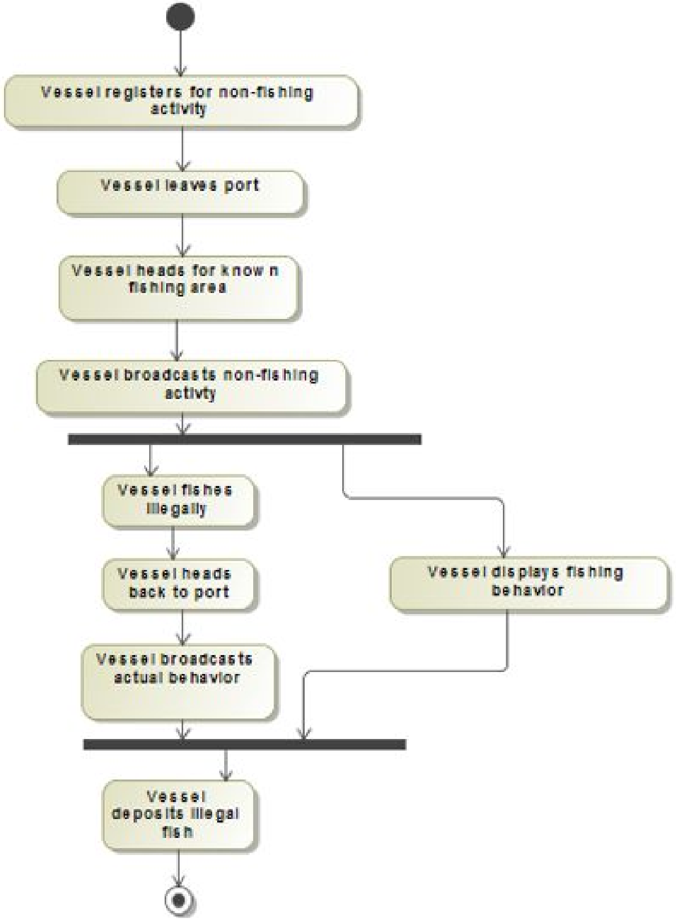
Longliners lay long lines with hooks attached to catch fish. The ship travels at about cruising speed while laying the line. The vessel then drifts for several hours before reversing to haul in the line. The process may take up to a full day.

Purse Seine

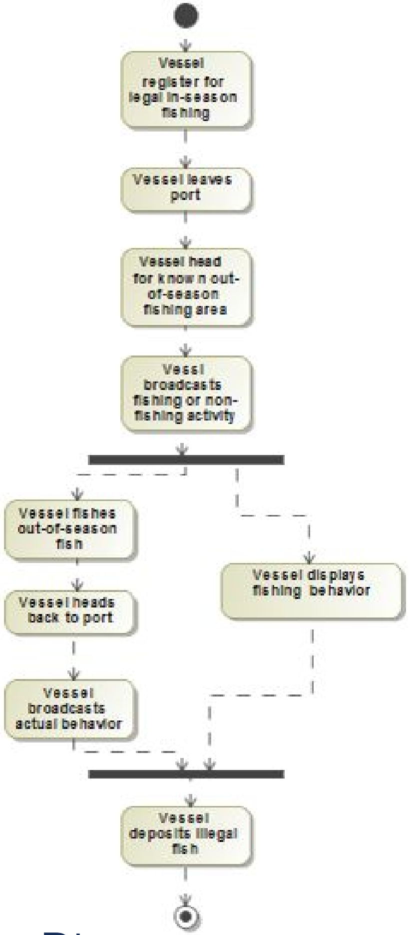
Purse Seine search for large schools of fish then deploy large nets attached to floats are deployed. The ship then moves at fast speeds to capture the fish. The ship begins to drift to real in the haul.



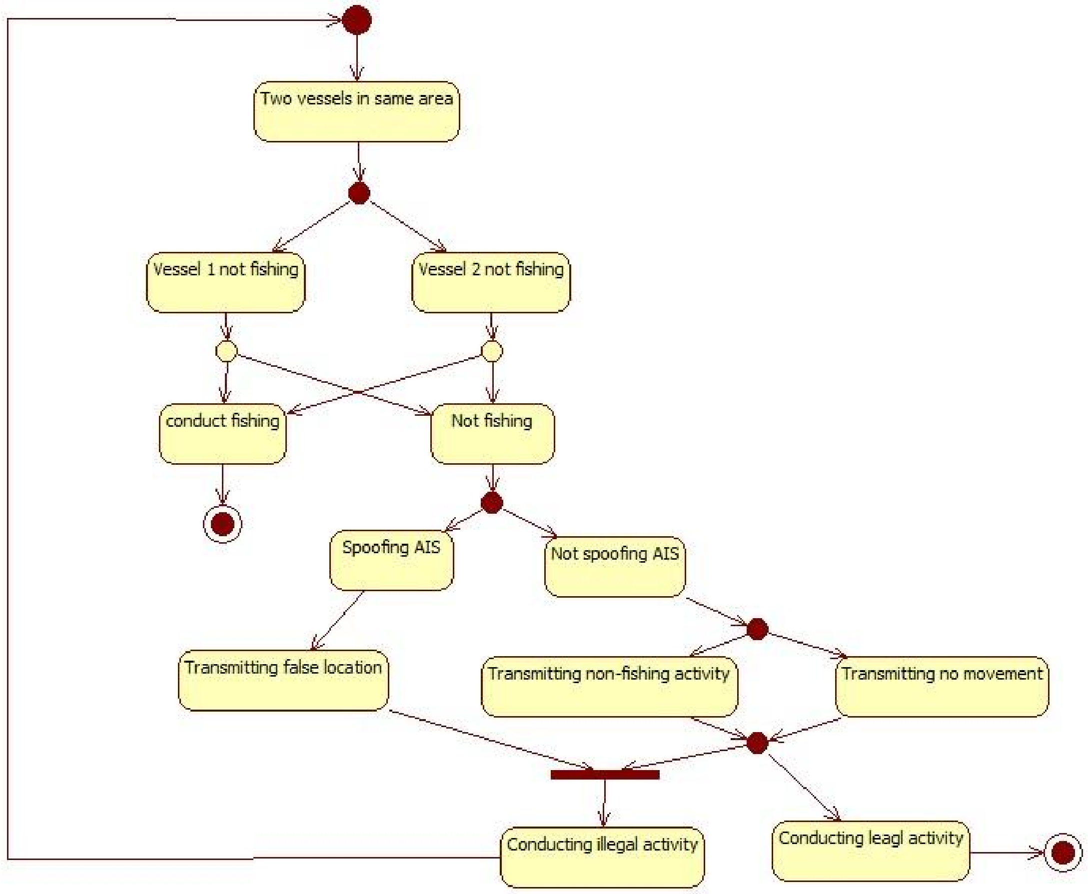
Legal Fishing Use Case Diagram



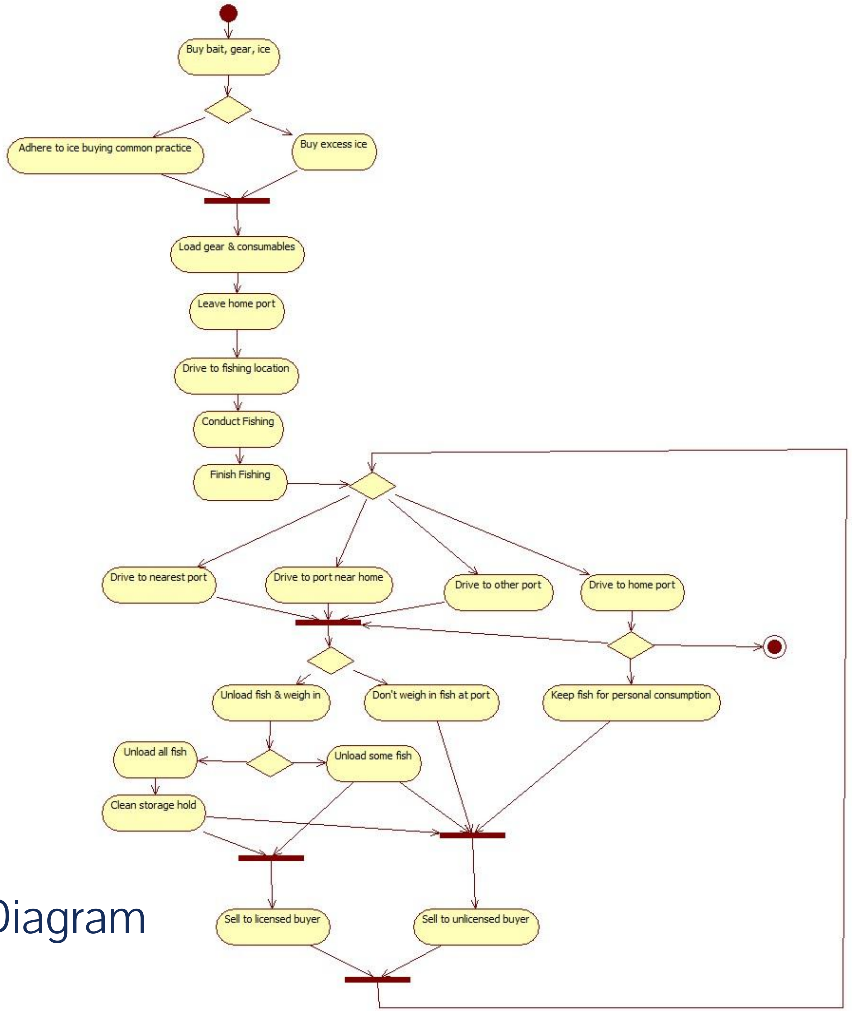
Broadcasting Non-Fishing Activity Diagram



Fishing Out of Season Activity Diagram



Vessels Too Close State Diagram



Illegal Sale of Fish State Diagram

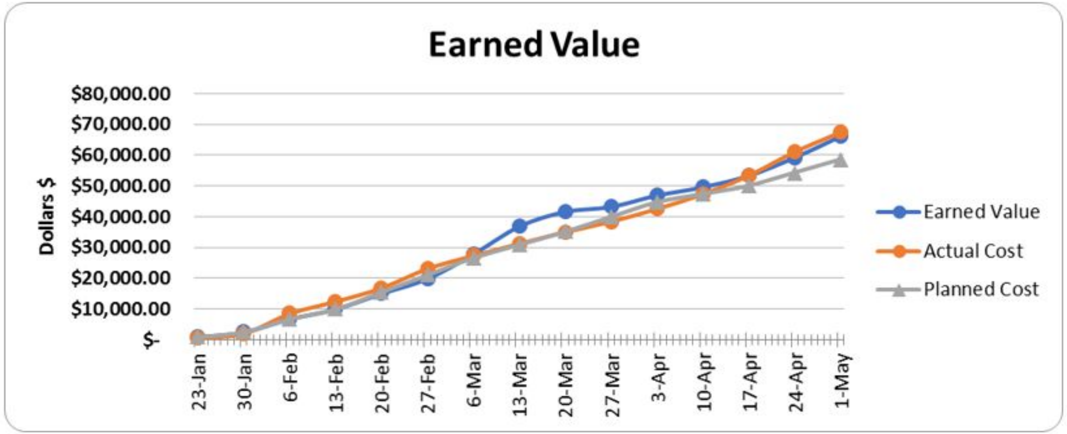
Pre-Derived Training Set

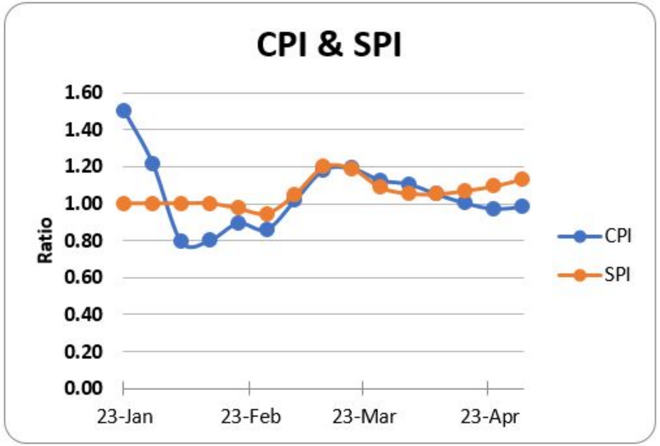
Attribute	Description
Mmsi	Vessel Identification
Timestamp	Time in UTC (seconds)
distance_from_shore	Haversine distance from point to shoreline; data provided by Natural Earth [2]
distance_from_port	Haversine distance from point to port; data provided by Natural Earth [2]
Speed	AIS reported speed
Course	AIS reported course; compass direction
Lat	AIS reported latitude
Lon	AIS reported longitude
is_fishing	Classification of the data point 0 = Not Fishing 1 = Fishing -1 = Not Labeled

Post-Derived Training Set

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 window)

Window (seconds)
900
1800
3600
10800
21600
43200
86400





Earned Value Management

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

Roles and Responsibilities

ID		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

Page 1

ID		Task Mode	WBS	Task Name	Duration	Start	Finish
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

Schedule