**LM StarGazer - Design of a Decision Support Tool and Data Science to Predict Illegal Fishing**

1. Mityushina, R. Skaddan, S. Jagtap, E. Taylor, A. Mirsa

**Abstract.** Illegal, unreported and unregulated (IUU) fishing is a worldwide issue contributing to ecological devastation. It is estimated that 20% of the total fishing yield is due to IUU, making it a top priority for global law enforcement services. Fishing in restricted areas can have detrimental effects on the ocean floor and fish populations pushing species towards extinction and causing damage to ecosystems on a global scale. Enforcement agencies vary depending on the country; the primary stakeholder in the United States is the Coast Guard. One of the primary inefficiencies of detecting a vessel’s location/activity is the lack of data. The Automated Identification System (AIS) which broadcasts ship information is mandatory on all large vessels but is inconsistent, easy to modify, and not continuous. This project has developed a decision support system to assist law enforcement in the identification and apprehension of vessels who are partaking in illegal fishing by observing vessel behavior, predicting the likelihood of a ship crossing into a restricted area, and providing a satellite image of the area. The system, LM Stargazer, supplements law enforcement agencies with evidence and additional information of vessel activity in a condensed report which is generated for every flagged vessel with a rank of .6 (0 to 1) or more. The data analysis and machine learning of vessel behavior can be extrapolated to other vessel identification scenarios, making this project valuable to situations with other similar actors. Such as, perhaps, vessels engaged in drug or human trafficking.

Introduction

Currently, there are no tools which can make a prediction about a vessel’s ‘status’ (fishing or not fishing), determine if the vessel is in a restricted region, and recommend a course of action. There is a need for a vessel identification system that uses available data and quickly provides recommendations to the personnel monitoring an identified region. LM Stargazer improves the identification process, enabling enforcement agencies to make decisions about suspicious vessels where no analysis was available before.

The Automated Identification System (AIS) is a transponder that broadcasts ship information to a centralized database. Vessels larger than XX tons are required by law to be equipped with this transponder. The data communicated through the transponder includes a longitude and latitude position, a timestamp, vessel ID, speed and direction. Broadcasts are made in intervals of roughly 10 minutes, but can be turned off, be outside of coverage, or even modified causing gaps or anomalies in the dataset.

AIS data is used as a baseline for the identification of illegal fishing, but there are a few other methods used to supplement those datasets. The Coast Guard currently detects a ship’s location through radar towers and patrol missions. The radar detection provides more precise and continuous information than the AIS feed. A drawback is that the tower has a limited range and is primarily restricted to land. Another method for catching vessels engaged in illegal fishing is by having patrol boats conduct random walks around protected areas, but this is severely limited because the typical line of sight is less than 10 miles. Our system will not replace the existing methods but can aid in the identification process. By creating a tool that adapts and ‘learns’ vessel behavior, the prediction component for vessel status becomes more accurate as data is fed into the system. With the ability to coordinate behavior and location, it is possible to predict a likelihood of illegal activity. If a likelihood exceeds a value of .6 (on a scale of 0 to 1) then the system checks for available satellite coverage. If a satellite image is available, then a portion of the image with the vessel’s position is captured and added to the output file. With a likelihood of above .6, a condensed report will be displayed to the user containing all available information about the vessel in a comprehensive manner.

LM Stargazer can be partitioned into two key models. The first model is a logistic regression model with gradient boosting based on nine variables including AIS location data, speed of the vessel, and type of the vessel. This model predicts if a vessel is fishing or not fishing at a given data point. The second model uses a subset of that dataset and determines if a vessel has crossed into a restricted region. There are three parts to the position model: a position check against identified points of interest, an estimation of if a vessel crossed into a restricted region between received positions, and a predictive component that estimates the next location a vessel could be in, turning the tool into a proactive analysis, instead of being purely a reactive.

The system first recognizes and flags a vessel operating with an identified behavior, and then the data is enriched by coordinating satellite imagery. This process of vessel identification through AIS tracking, and a confirmation through satellite imagery improves the analysis for vessel apprehension and improves the response time of the organization. Vessels identified with a high risk of illegal activity are communicated to the user through a display. The identification of a vessel engaged in illegal behavior is identified through the support tool, but the ultimate decision to engage in apprehending the vessel will be left to the individual.

Context Analysis

Illegal fishing is a worldwide concern, through global efforts, stricter regulations and improved detection methodologies its severity is being reduced every year. Nonetheless, The impact of overfishing a species can change global ecosystems and have unexpected consequences by damaging other dependent species and the environments they live in. Roughly 11 to 20 million tons of fish are caught illegally, equaling about 30% of the fishing market [1]. By sectioning off and protecting regions, the ability for vessels to dredge, deplete and damage the environment is reduced. Unfortunately, protection of the identified areas is complex, costly and potentially dangerous. It is dangerous because the trip to arrest a vessel can be a long, encounter rough weather, and met with hostile resistance. With minimal data and few resources to allocate patrolling large bodies of water, illegal fishing continues to thrive.

The primary apprehension unit for the United States is the Coast Guard, who attempts to regulate vessels operating in restricted areas within the US economic zones (roughly 100,000 miles around the coast). Fishing detection and apprehension is heavily dependent on data. The decision to engage with a suspect vessel is costly, potentially dangerous, and should not be done without a high probability of illegal activity. While it is possible to track a vessel and know it’s relative position through the Automatic Identification System (AIS), the Coast Guard supplements the data source with radar detection and patrolling units. The radar detection provides a precise and continuous track of vessels, but only covers major ports and is unable to detect every vessel on the coastline. Therefore, adding an analysis to the available data strengthens the detection process.

Previous work was conducted by another George Mason University Capstone project who built a regression model to predict if a vessel is fishing or not-fishing [2]. Our system incorporates the previous work and furthers the analysis by relating the location points and the fishing status to create a prediction of if the vessel is conducting illegal fishing or not.

Lastly, illegal, unreported and unregulated fishing is not the only application of this methodology. It is possible to relate a regression model and a position prediction model to other scenarios such as drug and human trafficking with boat transportation. While this model is outfitted to detect illegal fishing, the process of a model that learns vessel behavior and can make an analysis about its known path, current status, and predict its next location is valuable to different enforcement agencies.

Need Statement

As stated above, IUU can have detrimental effects on the ocean floor and fish populations, pushing species towards extinction and causing damage to ecosystems on a global scale. Vessels conducting IUU fishing do not adhere to global rules regarding catch limits or quotas, further harming fish populations. The current methodologies for detecting and apprehending vessels engaged in illegal behavior still permits roughly 20% of the illegal market.

Therefore, this project fulfills the need to improve the detection process and determine which boats are conducting IUU fishing. The faster and more thorough the analysis in the detection of illegal fishing results in more boats being stopped, and in illegal fishing to be reduced.

Stakeholder Analysis

The operator of this system is the United States Coast Guard or another country’s vessel apprehension organization. Therefore, not only does the analysis have to be thorough but the user interface and output must be comprehensible so that anyone can understand the situation. The Coast Guard is a branch of the US military that is tasked with enforcing US laws in open waters. The commitment to arrest, apprehend, or punish a captain/vessel can be a cumbersome task. The proper team must be assembled with the correct vehicle and adequate supplies, therefore the decision to engage with a large vessel is taken under thorough consideration. To achieve the decision to board and apprehend a vessel, there must be an overwhelming circumstance with supportive data to formulate the foundation to risk engagement. LM Stargazer could aid the Coast Guard by continually running the analysis which prioritizes active vessels with the highest likelihood of illegal fishing and produces an automated report with all available data.

To achieve maximum capability of the system, the system must be connected to the appropriate data sources. A live AIS feed that supplies, MMSI values, position and speed are necessary to have a real time analysis. Another data source are satellite images over the ocean, which currently are sparse and inconsistent. There are commercial services, such as SkyTruth and Planet Labs, as well as government organizations who provide satellite imagery. To obtain the full report, the tool must connect and retrieve available images from the third-party provider.

Fishing vessels are a primary actor in the system. They must be equipped with the required technology, and they are the ones who potentially engage in illegal activity. They want as little government interference as possible and the fact that they must register, request to leave port, and have a tracker on their ship is viewed as a hassle and a waste of time [2]. Then, whenever the fishermen encounter the Coast Guard it has the same viewpoint as other forms of law enforcement agencies: general feelings of disgust (until they are in a situation they need to be saved). Also, not all illegal fishing is intentional. It is possible to set nets in a legal location and then the current changes causing the nets to drift into a protected region. Then upon retrieval, the ship must travel into the restricted area and run the risk of being caught or losing their equipment.

Because the law enforcement agencies would adapt to this new system, there is little opposition to resist the technological development of detecting IUU fishing. Other corporations provide portions of this project, such as providing satellite analysis or a position model, but no current system combines logistic regression, position model, and satellite imagery to produce a single comprehensive result.

**Concept of Operations.** Throughout the day, a Coast Guard user would conduct their standard surveillance operations but is able to switch the computer monitor to the systems user interface. The interface provides a map of the monitored region with vessels recorded tracks, a list of unique vessel identification numbers with the likelihood of illegal activity, and condensed reports for selected vessels; the system defaults to the vessel with the highest rank. A user could select the unique vessel id and view the condensed report containing the logic for the illegal activity detection and a satellite image if available. The report can be saved and used as evidence to initiate an investigation.

LM Stargazer will be incorporated into the surveillance component of the existing system. The user would identify a region to monitor and run the system throughout the day. Detection of illegal fishing vessels is identified by surveillance of designated areas and collecting data. The system continually pulls data from public data sources, such as Global Fishing Watch, and identifies a list of recommended vessels to investigate within the identified region.

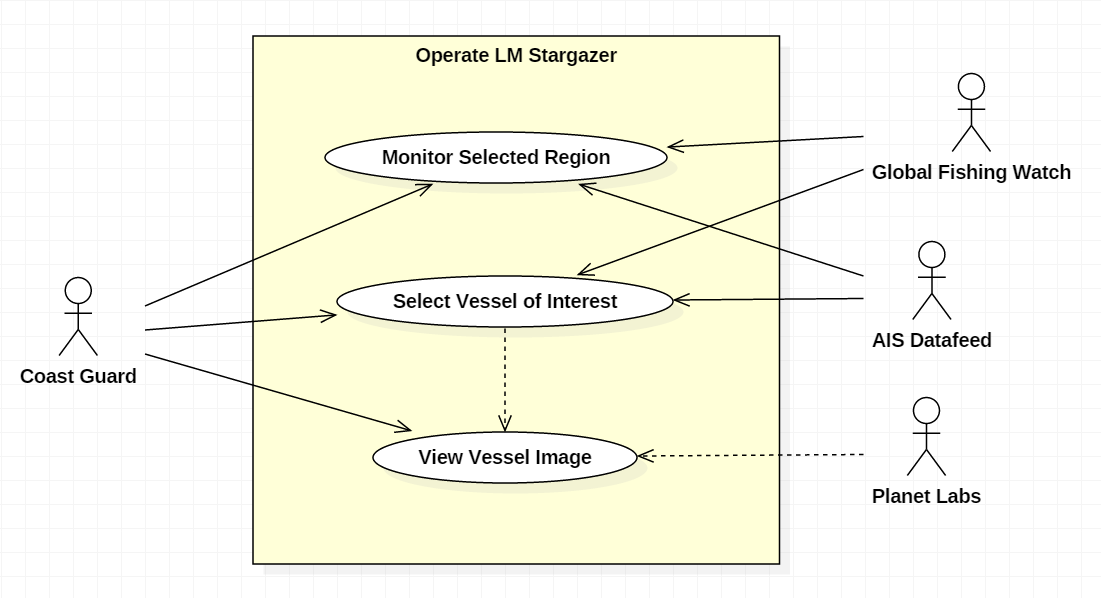


Figure 1. Use Case Diagram

In the image above, it describes the actors involved with the basic functionality of the system. The Coast Guard only interacts with the system and does not communicate with the data sources directly. The Coast Guard user has the ability to monitor the selected region or select a vessel and find out supplementary information about its path. To incorporate the system into the detection process, the current methodologies used by the Coast Guard can be improved by prioritizing the list of predicted vessels fishing in an illegal location.

Data analysis

**Preparation.** Data preparation included data collection, data exploration, data validation, and data cleaning. The major sources of the data used in this project are Global Fishing Watch (GFW) AIS, Point of Interest, and Planet Labs Satellite imagery data.

**Data Sources**

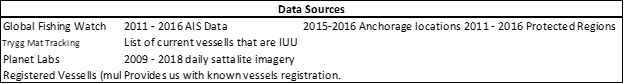
****

Table 1. Data Sources

**Data Preparation.** Global Fishing Watch (GFW) is a new technology platform developed by Google, SkyTruth, and Oceana that uses satellite Automated Information Systems (AIS) data to monitor fishing activity around the world in near real-time. GFW provides the user with vessel identity, fishing activity, transshipment information, and anonymized AIS data.

The fishing hours of vessels on each day have been combined to get the fishing hours for gear types. Fishing vessels are categorized into four groups of fishing gear type: longliner vessels, trawler vessels, Fixed Gear vessels and seiner vessels. Every vessel with different gear type has a distinct behavior: A longliner fishing vessel is described to use long line, with baited hooks attached at intervals by means of branch lines called snoods to catch fish and once the lines are laid the vessel moves at very slow speed. The time required to fish in this manner varies but may take up to a full day. A trawler fishing vessel is 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. A purse seiner fishing vessel searches for large schools of fish, and once identified this method uses a net called a seine, that hangs vertically in the water with its bottom edge held down by weights and its top edge buoyed by floats. Purse seine fishing vessels move sporadically and at inconsistent speeds, following schools of fish. A fixed gear vessel is a category of vessels that includes set longlines, set gillnets, and pots and traps. They can move to various locations, with expected rest points where the equipment is set. This behavior and inputs are used 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.

**Data Conditioning.** The training Data accessed from the Global Fishing Watch GitHub repository, which included several merged datasets. To access the merged training datasets the "prepare.sh" script inside the repository must be run. The "prepare.sh" script is used to run the python codes that merge different datasets to combine separate vessel track files into a combined gear-specific file; i.e., one file for each longliner, trawler, purse seine, fixed\_gear, squid\_jigger and other\_fishing. After obtaining all the training datasets, the data is cleaned by removing all the rows with missing or duplicate data.

**Predictive Modeling.** Three models were re-developed from prior work to determine if a vessel is fishing or not fishing based on various attributes. The models needed redevelopment as the original build included the majority of all attributes and overfit the model.

For models 2.0 Gradient Boosting was used to reduce dimensionality from 94 attributes down to 9 in each model. Further, each of the final variables were reviewed for variable correlation and transformed to ensure the best linear fit in the final model.

These three models will be incorporated into the design as a first basis to determine if a vessel is fishing illegally. Initial model performance indicates that more variable transformation is needed to prevent overfitting.

**Satellite Imagery.** The team has partnered with Planet Labs to acquire Satellite imagery from 2009 to 2018. The data is accessible to the team via an API call which allows the final system to request an image with determined end points to be clipped and shipped to the user interface. At a later time, this image will be analyzed to determine the position of a protected are in comparison to the boat.

System Development

There are five components to the model. A position model – to predict overlap between a ship and identified regions, a regression model – to predict the behavior of a vessel and predict if it is fishing or not fishing, a Monte Carlo prediction – to predict the next location the vessel will be, a Bayesian net – to use the generated information to form a prediction about the likelihood of a vessel engaged in illegal fishing. And finally, when the image is available, an image extraction – to accesses Plant Labs API and collect an image from a selected region and specified timeframe.

**Position Model.** The position model has been coded in python and uses the AIS data to determine how close the recorded points are to a point of interest. Points of interest include restricted fishing regions, ports and anchorage zones. Because the AIS data doesn’t provide a continuous data feed and there is a significant amount of time between data points, uniform distributions are used to predict the positions a vessel could be in between data points. This was done by using the midpoint between the intervals and associating an area of a circle as a uniform distribution. Uniform distributions are assumed due to the lack of information and uncertainty of a vessels path between points.

To calculate the area overlap, the points of interest are estimated by finding the center of the identified region and estimating the area of the circle to match size of the identified region. To calculate the area overlap between the vessel the point of interest, the equation to calculate different sized circles overlap, equation (1), was used. This prediction provides a percentage of how likely a vessel was to cross into a protected region. To achieve an overlap greater than 50%, the calculated midpoint must be within the radius of the protected area. Therefore, midpoint values which are outside the circle, yet yield an overlap greater than 30%, are flagged for consideration.

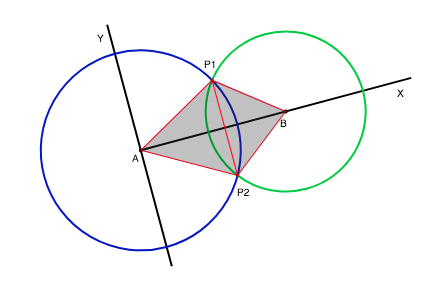


Figure 4. Area Overlap

Equation (1)

**Regression Model.** The regression model is built off of 5 years of recorded data with verified fishing status. Logistic regression is a heavily trusted and used tool in risk modeling, prediction models (Quantitative Structure-Activity Relationship), and discovering gene interactions. It has shortcomings as it can overfit and it can predict poorly on multi-level factor datasets. It has advantages over classifiers such as random forest models when needing ranking and probability estimation. For our purposes, ranking and probability estimation is needed. The team used gradient boosted mechanism for variable reduction with glm logit as the end model.

The gbm package is initially ran with 0.05 step size (i.e. shrinkage), 800 trees, bernouilli distribution to get the probabilities, interaction depth of 2, 5 cross validation folds, and The gbm model starts to overfit around 700 trees in that we no longer see an improvement form additional tress, but we do see a reduction in TrainingDeviance (i.e. 20% hold out) and ValidDeviance (5 fold cross validation).   
Then we run the gbm function for summarization which allows us to view the top contributors in explaining our variance in the dataset. For longliner, we find that the following 9 attributes contribute to 92.96% of variance in the data. We could select more attributes, but we want to make a model with the least number of attributes that will reduce its chances of being susceptible to noise.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **AUC** | **BestIteration** | **Learning Rate** | **Training Error** | **Validation Error** |
| **LL** | **0.998** | **700** | **0.05** | **0.1525** | **0.1638** |
| **PS** | **0.990** | **500** | **0.05** | **0.0803** | **0.0814** |
| **TR** | **0.998** | **700** | **0.05** | **0.3989** | **0.4022** |

The final model confirms our hypothesis that hat gear type is an important variable to determine the true fishing behavior of the vessel. We will add variables one by one to see how the lift chart improves. After we assess single variables in the model, we then add interactions of vessel flag plus the predictive variables to see if the performance improves.

**Future Predictive Model.** A Monte Carlo analysis varies the direction and speed to extrapolate possible future positions and calculates how many of the next positions are located within an identified region. This model is intended to provide a proactive component to the identification system, instead of being purely reactive. Because it is impossible to predict the decision making of the captain of various ships, uniform distributions are used to replicate possible course of actions. One thousand various paths are generated for each vessel point and the number of positions that are within a protected region returns the prediction of a vessel to move into a restricted region. One thousand replications were chosen to maintain a usable processing time of under one minute.

**Bayesian Net.** A Netica diagram was built to give a high-level overview of various attributes that feed the illegal prediction model, figure 1. As probabilities are derived from the previous models, the likelihood of a vessel conducting illegal fishing change.

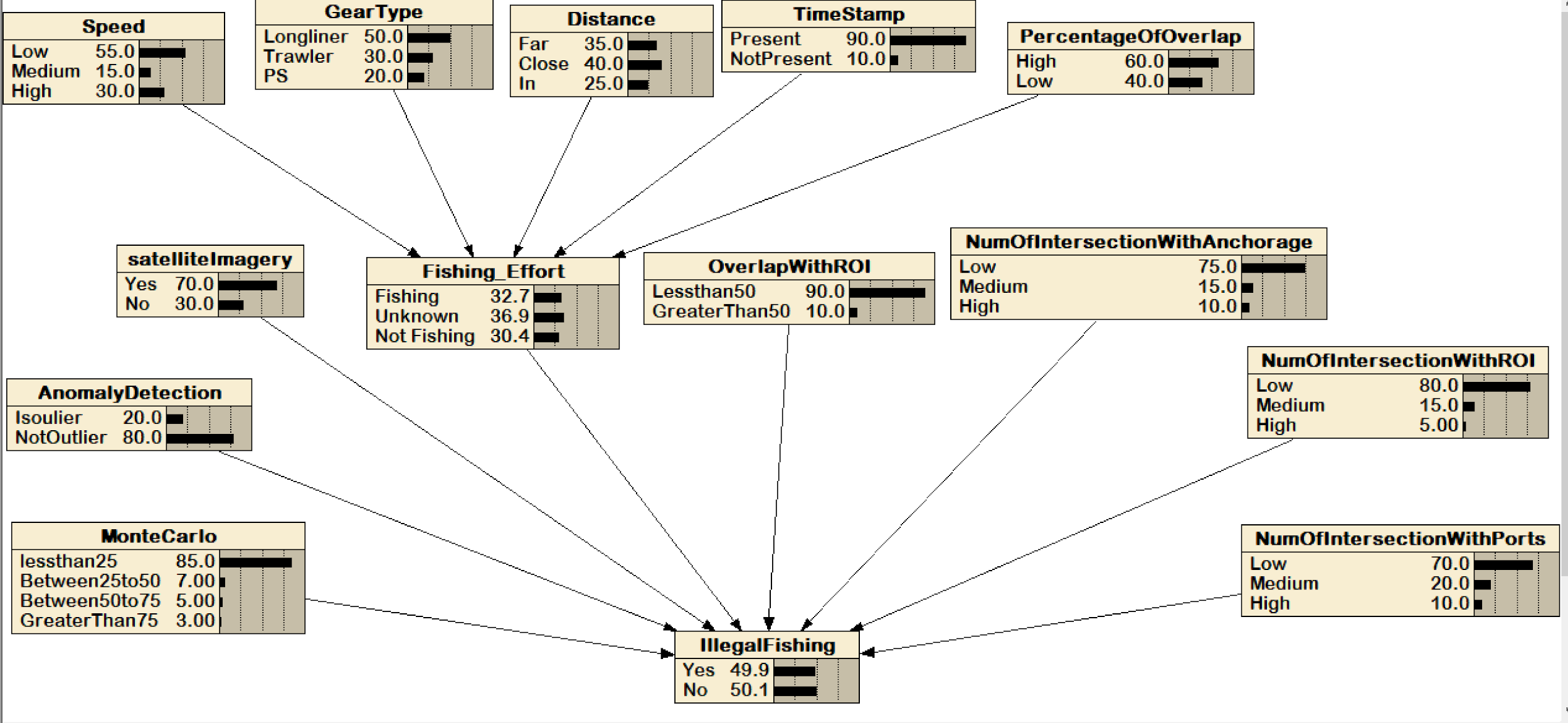


Figure 5. Netica Diagram of Illegal Fishing Rank

The attributes i.e., the Gear Type, Satellite Imagery, Fishing effort, Percentage of overlap and Distance formulate dependent probabilities to provide a likelihood of illegal fishing. Probabilities were derived from research, subject matter experts and verified through our sponsors.

**Image Extraction.** Using Planet Labs’ library of images, an API is utilized to feed coordinates and save an image to a storage device. The image retrieval is initiated when the likelihood of illegal fishing exceeds a rating of 0.6. A region containing the last known location of the vessel is provided and the API searches for an available image. If an image is found, the image is saved for the user to view. Future work would be to add a layer of the location of protected areas in red. This would provide the user context to the image, allowing them to internalize the image faster. This modified image is linked to the condensed report that is available to the user to inspect.

The report generation is a summary of the ship that is selected, and includes information such as: regression model results, position overlap prediction model, the Monte Carlo analysis, and the IUU prediction. Below the report is an image dock, which shows real satellite images, of which are in the latitude/longitude bin where the selected ship is and/or was.

**Visualization.** Our finalized UI is in the form of a Tableau project. The best way to describe the setup would be to describe its left side and then the right side as shown in the figure below.

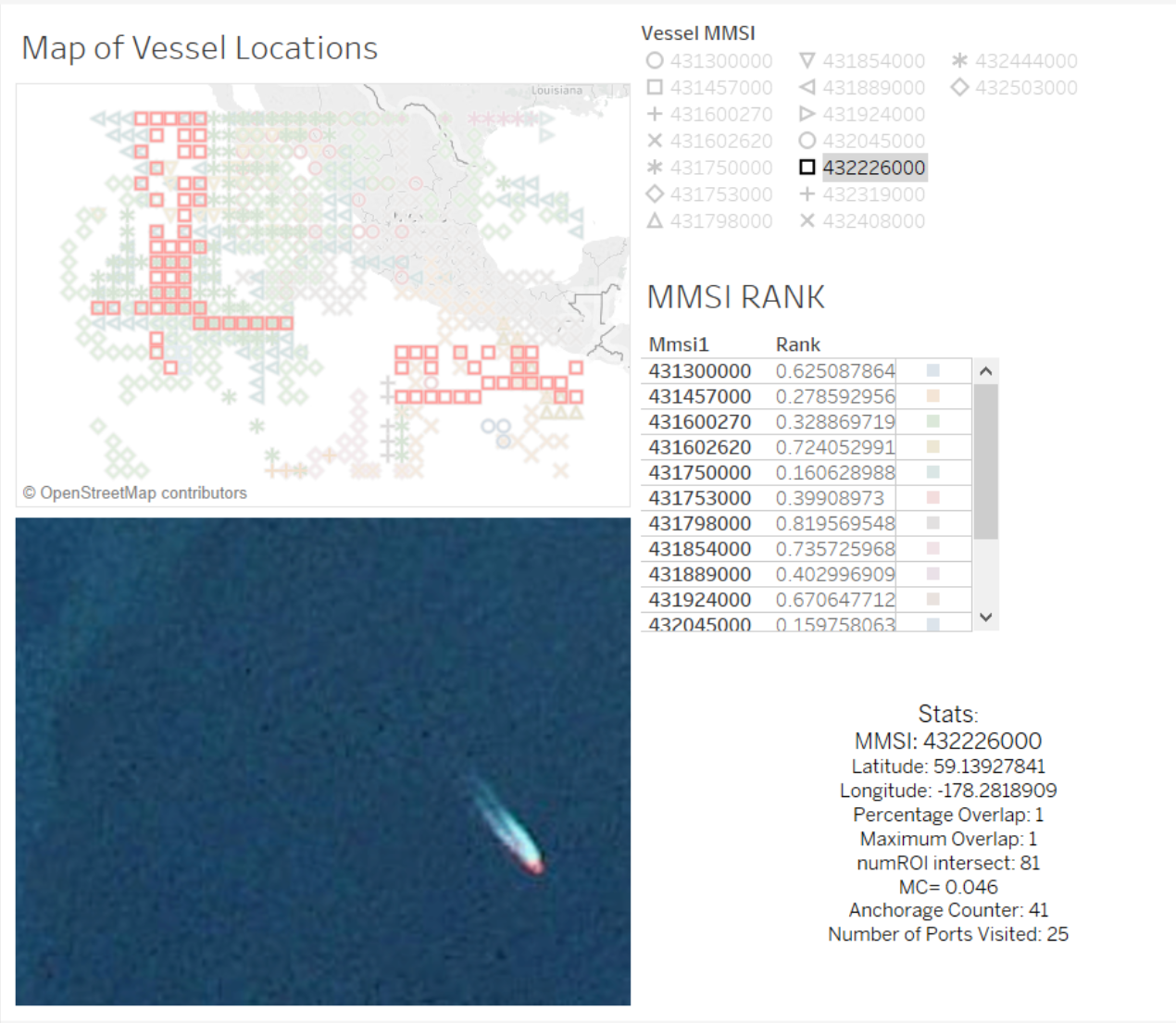


Figure 6. Tableau Dashboard

The left side of the user interface contains three parts: A map with color-coded density squares, a region selector, and a static key. The process starts with the region selector, where the user will select which region to observe (it is automatically set to Pacific Northwest at loading). Once that happens, the map will populate and display a shape density plot of boats out in the area, showing the latitude/longitude bins. Having a static key at the bottom of the left-hand panel is useful for first-time users so they become familiar with the vocabulary as well as the basic functionality of the UI.

The right side of the interface (top to bottom) has a ranking table, a mini-report, and then an image deck. For the region that was selected on the left-hand side of the UI, the table at the top of the right-hand side displays the top ten MMSI and their respective rank. If a person clicks on the MMSI, it will activate the results.

**Proof of Concept.** To assess and verify the abilities of the system a sample dataset has been identified and used to test the performance of the system. A vessel was identified as fishing on Jan 4th, 2013 off the coast of California. In a Planet Labs database, a satellite image was identified to contain the longlineer vessel. The vessel has recorded longitude and latitude points from the AIS transponder, and some of the potential vessel paths intersect with protected regions making this an optimal trial dataset. The system used the AIS information and matches the points to the linear regression to provides a likelihood of if the vessel is fishing or not fishing. Then the system creates additional data points by calculating the midpoints and determining a prediction of how likely the vessel went into the protected area. This is done by assigning the vessel a uniform distribution of the area of a circle, diameter of 2.5 miles, and assigning the protected area a circle with the respective size of the area. By calculating the overlap, it provides an estimate of how likely the vessel crossed into the restricted area. Because there were sections of the path that exceeded an overlap of 50%, then the model searched Planet Labs for an image within the respective timestamp, and location. The model was able to identify the vessel and retrieve the satellite image from Planet Labs.

Analysis

The training data extracted from Global Fishing Watch was provided by a researcher named Kristina Boerder, from Dalhousie University, Canada. The provided data includes the measures used in LM Stargazers logistic regression, but also includes the vessel’s status, providing a known datasource of actual fishing and not fishing behavior. There are three types of vessels identified, longliners, trawlers, and purse seine vessels, each having their own behavior to model. More training data would improve the accuracy of the model although the regression has an accuracy of over 85% for each type of vessel.

For the overlap prediction portion of the position model, calculating the midpoints between recorded locations and associating a circle’s area as a uniform distribution of where a vessel could have traveled is a crude estimate. Uniform distributions are an acceptable estimation due to the unpredictability of a vessel’s path between points. Using circles to measure the overlap was the best method of calculating shape overlap. Future work should include the development of polygon shape analysis to provide precise depiction of a protected area. The vessel’s uniform distribution represents a prediction of how likely the vessel traveled into the identified region. This prediction provides one piece of the illegal fishing detection prediction.

The Monte Carlo analysis currently runs 1,000 iterations to vary the possible locations the vessel’s next position will be. This takes about 3 seconds to process per vessel, therefore if the number of vessels being observed at one time is greater than 40 vessels, a two minute delay may not be acceptable to our system users. The when the number of active vessels exceeds 40, the number of iterations should be reduced to 100 iterations, allowing up to 400 vessels to be processed in the same amount of time.

The image extraction from Plant Labs currently has limited applicability. This is due to the lack of satellite coverage over the ocean. Companies, such as SkyTruth and HawkEye, are beginning to expand their coverage to the ocean to aid in the identification of illegal fishing activity, but is still lacking full coverage. Unfortunately, satellite images are similar to AIS data feeds, they do not provide a continuous source of data and take images in intervals. This means that there will be times when an image is taken but the vessel is no longer within a restricted region. Future work can be expanded to adding layers to the retrieved images and performing image analysis to identify how close the vessel is to a restricted region and where it was coming from. Future work may be to outsource this portion of the analysis to a company who performs image analysis for vessels as a singular task and incorporate the analysis into the system.

Conclusion

The models used in the identification of illegal fishing are robust enough to handle various types. The logistic regression and Monte Carlo analysis adjust their functions to the vessel type allowing a more accurate analysis versus a broad generalization of boats. The logistic regression model has sacrificed some accuracy to reduce the likelihood of overfitting yet maintains a high level of accuracy. Then, by extrapolating predicted positions from recorded locations the position model is able to increase its capability and transform the model into a proactive analysis. Through the combination of a logistic regression, position model, bayesian net, and image extraction, an analysis on a vessel’s status is able to be predicted unlike any system currently available. This system does not replace the current methods of identification, but strengthens the analysis of the collected data to support the decision process of apprehending a vessel.

References

Agnew, David J., et al. “Estimating the Worldwide Extent of Illegal Fishing.” *PLOS ONE*, Public Library of Science, journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0004570.

Byrnes, Jarred, et al. “IUU Fishing Detection Final Report.” *IUU\_Fishing*, 2017, seor.vse.gmu.edu/~klaskey/Capstone/IUU\_Fishing/index.html.

Global Fishing Watch. [2018]. ​www.globalfishingwatch.org​

Jsoma. “Greek PM Sentiment Analysis before and after Signing the MoU with EU[Project] · Issue #100 · Jsoma/Data-Studio-Projects.” *GitHub*, github.com/jsoma/data-studio-projects/issues/100.

Planetlabs.com, planetlabs.com/.

Revkin, Andrew C. “How Digital Tracking of Rogue Fishing Can Safeguard Vast Ocean Reserves.” *The New York Times*, The New York Times, 15 Sept. 2016, dotearth.blogs.nytimes.com/2016/09/15/how-digital-tracking-of-rogue-fishing-can-safeguard-vast-ocean-reserves/.

Souza, Erico N. de, et al. “Correction: Improving Fishing Pattern Detection from Satellite AIS Using Data Mining and Machine Learning.” *PLOS ONE*, Public Library of Science, doi.org/10.1371/journal.pone.0163760.

Team Bio:

### Anya

**Education:** Finishing last semester of the Master of Science in Data Analytics Engineering with a Predictive Analytics Concentration. Hold a Bachelor of Science in Criminal Justice from Texas State University.

**Work Experience:** Data and Decision Scientist at Fundation (FinTech). Work cooperatively on the design, development and testing of statistically grounded insights for various business needs including risk modeling. Prior work experience as a Data Analyst at a Non-Profit working on reporting, data processing, database migrations, and segmentation analysis.

**Tools:** Microsoft SQL Server, SQL language, R Programming, Excel (Advanced). Learning and eager to use on a project: Python, AWS.

**Interests:** Behavior and financial modeling, NLP, Image analysis.

***LinkedIn****:* https://www.linkedin.com/in/mityushina/

### Spandana

Is a graduate student, pursuing her Master’s in Data Analytics Engineering at George Mason University. Also, a Graduate Teaching Assistant with the AIT department under Dr. Ioulia Rytikova and with the SEOR department under Dr. Paulo costa and Kathy laskey. Holds a bachelor's degree in Information Technology from Osmania university in India.

**Tools**: MySQL, R, Weka and Tableau.

**Interests**: Predictive analytics, NPL, Sentiment analysis.

***LinkedIn****:* <https://www.linkedin.com/in/spandana-jagtap-4915a980/>

### Emma:

BS Science in Physics from Guilford College, with departmental honors, and a Mathematics minor. Undergraduate thesis: The Evolution of the Relationship Between the Temporal Decay Index and the Spectral Index of Gamma-Ray Bursts.

**Tools**: MatLab, Unix shell, LabView, various other physics-oriented tools, Python, R, SQL, Tableau, Spark and now ElasticSearch.

**Interests**: Numbers, learning new things, challenges, tea

### Abhishek:

Is pursuing his degree in Master’s Data Analytics Engineering Program. He is a TA to IT 213, IT 315, IT 431 courses of the Information Sciences and Technology Department of Volgenau School of Engineering. He hold a bachelor’s degree in Biomedical Engineering. Prior experience in molecular mand eLearning, fabricating microneedles for a Bill and Melinda Gates Foundation project, and developing e-learning products for Microsoft and Symantec and Symantec.

***LinkedIn****:* <https://www.linkedin.com/in/abhishekmishra11/>

Raymond Skaddan:

He graduated in 2017 from George Mason University with a Bachelor of Science in Systems Engineering and is also a recipient of the Carl M. Harris Memorial Scholarship. While completing his Masters of Science in Systems engineering, he has been the Teacher Assistant for the undergraduate capstone senior design course (Syst 490 & 495). Outside of school, he has worked at Zodiac Aerospace for the past four years where he has been able to supplement his degree with mechanical, electrical and software engineering. In July of 2018, he will begin working at Northrop Grumman in Melborne Florida to apply his Systems Engineering background to the aerospace industry.