THE CURIOSITY CUP 2023
A Global SAS® Student Competition

Enhancing Illegal, Unreported, and Unregulated Fishing Detection through Feature Engineering

Jay Gala, Anmaya Agarwal, Saketh Mantha, Yash Katariya, and Prof. Prashasti Kanikar, Department of Computer Science, SVKM's NMIMS Mukesh Patel School of Technology Management & Engineering, Mumbai

Abstract

According to the United Nations Food and Agriculture Organization (FAO), illegal, unreported, and unregulated fishing (IUU) is a major concern for the global fishing industry. IUU fishing results in the loss of over 26 million tonnes of fish each year and has a significant impact on the world's oceans and marine life. The illegal industry supported by IUU fishing is estimated to be worth about 20 billion USD annually. Not only does this illegal activity have financial consequences, but it also leads to ecological losses that put many species at risk due to overfishing and fishing in Marine Protected Areas. IUU fishing not only undermines the sustainability of global fish stocks but also affects the livelihoods of those who depend on fishing for their income. The new dataset presented in this paper will aid in the global efforts to combat IUU fishing and its detrimental effects on the world's oceans and marine life. By providing comprehensive information on IUU fishing activities, this dataset will support the development of Artificial Intelligence enabled systems to address the problem and protect the oceans and marine species for future generations.

Introduction

Suspicious fishing activity or Illegal, Unreported, and Unregulated fishing (IUU fishing) is a worldwide issue that threatens ocean ecosystems and sustainable fisheries. Illegal fishing is the kind of fishing conducted by foreign or national vessels in the waters that are under the jurisdiction of a country or an organization, without the permission of the said country or organization. Unreported fishing is the kind that has not been reported or misreported to the authorities. Unregulated fishing is the kind of fishing that takes place in areas with no conservation or management measures for fishing activities. It happens all around the globe from the high seas to the exclusive economic zone (EEZ), which is a 200-mile band from a nation's coast wherein they have full jurisdiction over the exploration of marine resources. According to the UN Food and Agriculture Organisation (FAO), IUU fishing accounts for 26 million tonnes of fish caught annually. Fisheries are an important part of food sources and employment all around the world but these fisheries are being compromised heavily by IUU fishing activities. IUU fishing is a pressing matter that endangers local fisheries and can harm the local food supply.

Dataset, Cleaning, and Validation

In this study, we utilized the AIS disabling events dataset obtained from the Global Fishing Watch website. The dataset contained the MMSI (Maritime Mobile Service Identity) numbers of ships that disabled their Automatic Identification System (AIS), along with their latitude, longitude, gap hours, and the time when AIS was enabled again.

To further enhance the dataset, we collected vessel identity information from various Regional Fisheries Management Organizations (RFMOs), including the Commission for the Conservation of Antarctic Marine Living Resources(CCAMLR), the Commission for the Conservation of Southern Bluefin Tuna (CCSBT), the Inter-American Tropical Tuna Commission (IATTC), the Indian Ocean Tuna Commission (IOTC), Northwest Atlantic Fisheries Organization(NAFO), and more manually. The RFMO datasets contained details of vessels caught in Illegal, Unreported, and Unregulated (IUU) fishing activities within their respective regions. However, the datasets were disorganized and some lacked the identity information of some of the vessels and a lot of the RFMOs had cross-listed the vessels meaning a lot of the vessels were common for various RFMOs.

We combined the RFMO datasets into a single excel file and used a python script to obtain the MMSI numbers of the respective IMO (International Maritime Organization) numbers from the Global Fishing Watch API. After obtaining the unique MMSI numbers, we combined them with the AIS disabling events dataset and found 11 such MMSI numbers that had been caught in IUU fishing according to various RFMOs and also disabled their AIS according to the Global Fishing Watch dataset.

To generate features from the available attributes, we wrote some python scripts. The features generated included the distance traveled during the AIS disabling period, the determination of whether or not AIS was disabled within the Exclusive Economic Zone (EEZ), and the identification of the ocean in which AIS was disabled. We utilized the spherical distance formula as the world is a globe, and used the start latitude and longitude attributes to identify the ocean. To accurately identify the ocean, we utilized GeoJSON files obtained from the FAO and IHO websites, which contained the Major Fishing Areas and ocean boundaries, respectively. After merging the results from both sources, only 227 out of 55,000 entries were unable to be identified.

The final dataset consisted of 55,268 entries of vessels not caught in IUU fishing activities and 100 entries of vessels caught in IUU fishing activities. The dataset was highly imbalanced. The features used in the study so far included vessel class, gap hours, IUU caught status, distance traveled, EEZ check, and ocean name. This dataset was used to build Machine Learning models to classify vessels as IUU or not.

Feature Engineering for Improved Vessel Identification

In an effort to enhance our machine learning model, we tried plotting a graph of *gap_hours* v/s *spherical_distance* to visualize the data.

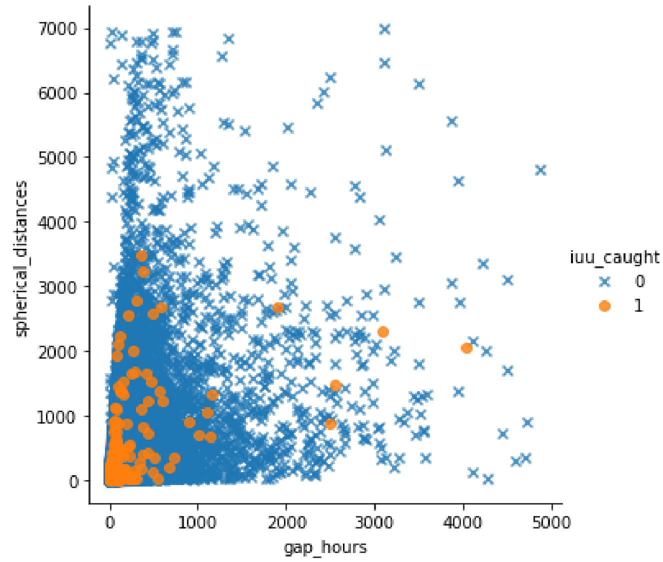

Figure 1. Scatter plot for gap_hours vs. spherical_distances

However, we noticed that the anomalies (vessels caught for illegal, unreported, and unregulated fishing activities) were not necessarily outliers, but instead lay in the middle of the normal data (see Figure 1).

This made it difficult for our algorithms to perform well, as we were expecting them to be anomalies. Thus, we sought to improve our existing features and create new ones to better capture the behavior of these vessels. These are described below:

1. **Vessel Type** - This feature had limited variability, with most vessels classified as "others". To improve this, we wrote code to determine the gear_type of each vessel by using its MMSI number and inputting it into the Global Fishing Watch API. To do this, we first extracted the unique MMSI values from our dataset and found the gear_type information for each one. We then compiled this information into a CSV file and merged it with the original dataset.
2. **Ocean** - In order to merge two datasets and get more detailed information about the oceans in which AIS was disabled, we visualized the data using Tableau. We discovered that the vessels classified as False (i.e., their latitude and longitude did not fall within any ocean) actually belonged to the Pacific Ocean (see Appendix, Figure 5). This was because the international date line crosses the Pacific Ocean, causing some longitudes to be misclassified as False. To resolve this issue, we found the JSON files for all the small seas within the Pacific Ocean and determined which points lay within them. This allowed us to get the exact names of the oceans, making two improvements to the existing attributes.

In addition to these improvements, we also made an effort to find more vessels that had been caught for IUU fishing activities. We conducted internet searches and found a few examples of vessels that had been caught, including Zhe Dai Yu (MMSI 900412888) and Oyang 77, and Lu Rong Yuan Yu 668. We verified their existence in our dataset and updated their IUU_check value to 1. While it is challenging to find vessels caught for IUU fishing in countries without RFMOs, we were able to add 60-70 more data points to our training set through our efforts.

This new and improved dataset showed signs of improvement when fed into different machine learning algorithms with a significant increase in the precision and recall values for the minority class.

Analysis of the Dataset

Upon visualizing the final dataset using SAS® Viya For Learners, we understood the following:

1. **Oceans with Maximum IUU activity** - The South Atlantic Ocean is where most of the IUU activity takes place. The North Atlantic Ocean and the South Pacific Ocean follow this. This insight highlights the need for increased monitoring and enforcement efforts in the Atlantic and Pacific Oceans to curb IUU activities (see Appendix, Figure 2).
2. **Gear types** - We can see that there is a clear correlation between some vessel gear types and IUU activity. Particularly, vessels with tune_purse_seines, drifting_longlinges, and set_longlines are caught more often than others. This information can be used by authorities to better understand IUU patterns and target their enforcement patterns. For example. vessels with these gear types must go through stricter background checks before getting their fishing licenses (see Appendix, Figure 3).
3. **Time of Day** - More than 12,000 fishing vessels have disabled their AIS signals during dawn, and 11,000+ vessels during the twilight hours (see Appendix, Figure 4). Although not all AIS disabling events indicate IUU activity, it is still a cause of concern and must be taken into consideration when studying IUU activity patterns. AIS can be disabled in the following cases: faulty hardware, pirate activity, bad weather, etc.

Suggestions for Future Studies

Future research in this area could focus on further refining and expanding the dataset, exploring alternative and more advanced machine learning techniques, and incorporating additional information sources to improve the accuracy and reliability of IUU fishing detection. Another direction could be to extend the study to include more diverse geographical locations and investigate the cultural, economic, and political factors contributing to IUU fishing in different regions.

Conclusion

In conclusion, the preparation of a novel dataset for IUU fishing detection is a crucial step in advancing the field of sustainable fishing practices. By gathering high-quality and diverse data from various sources, this dataset provides a valuable resource for researchers and practitioners working on IUU fishing detection. By analyzing the results from this dataset, we have been able to gain a deeper understanding of the challenges and limitations of IUU fishing detection and to develop more effective methods for addressing this important issue. We hope that this work will serve as a foundation for future research in the field and that it will encourage other researchers to build upon our findings and contribute to the development of more sophisticated solutions to IUU fishing. Ultimately, the creation of such a dataset is an important step in promoting sustainable fishing practices and ensuring the protection of our oceans and marine life.

References

Website What is IUU fishing? Available at https://www.fao.org/iuu-fishing/background/what-is-iuu-fishing/en/

Website The toll of illegal, unreported and unregulated fishing Available at https://www.un.org/en/observances/end-illegal-fishing-day#:~:text=According%20to%20the%20UN%20Food,of%20US%2410%E2%80%9323%20billion

Article Merten, Wessley, Adam Reyer, Jackie Savitz, John Amos, Paul Woods, and Brian Sullivan. "Global Fishing Watch: Bringing transparency to global commercial fisheries." arXiv preprint arXiv:1609.08756 (2016).

Dataset Welch et al. (2022) - Hotspots of unseen fishing vessels Available at https://globalfishingwatch.org/data-download/datasets/public-welch-et-al-disabling-events:v20221102

RFMO Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR). Available at http://www.ccamlr.org/en/compliance/illegal-unreported-and-unregulated-iuu-fishing.

RFMO The Commission for the Conservation of Southern Bluefin Tuna (CCSBT) Available at https://www.ccsbt.org/en/content/iuu-vessel-lists.

RFMO The Inter-American Tropical Tuna Commission (IATTC) Available at https://www.iattc.org/VesselRegister/IUU.aspx?Lang=en.

RFMO Indian Ocean Tuna Commission (IOTC). Available at http://www.iotc.org/vessels#iuu.

RFMO Northwest Atlantic Fisheries Organization (NAFO). Available at https://www.nafo.int/Fisheries/IUU.

RFMO North East Atlantic Fisheries Commission (NEAFC). Available at http://www.neafc.org/mcs/iuu

RFMO The North Pacific Fisheries Commission (NPFC). Available at https://www.npfc.int/npfc-iuu-vessel-list.

RFMO The South East Atlantic Fisheries Organisation (SEAFO). Available at http://www.seafo.org/Management/IUU.

RFMO South Indian Ocean Fisheries Agreement (SIOFA). Available at http://www.apsoi.org/mcs/iuu-vessels.

RFMO Western and Central Pacific Fisheries Commission (WCPFC). Available at http://www.wcpfc.int/wcpfc-iuu-vessel-list.

Contact Information

Your comments and questions are valued and encouraged. Contact the author at:

Jay Gala
SVKM's NMIMS Mukesh Patel School of Technology Management & Engineering, Mumbai
+91 9004942831
jay.gala78@nmims.edu.in
www.linkedin.com/in/jaykishorgala

Appendix

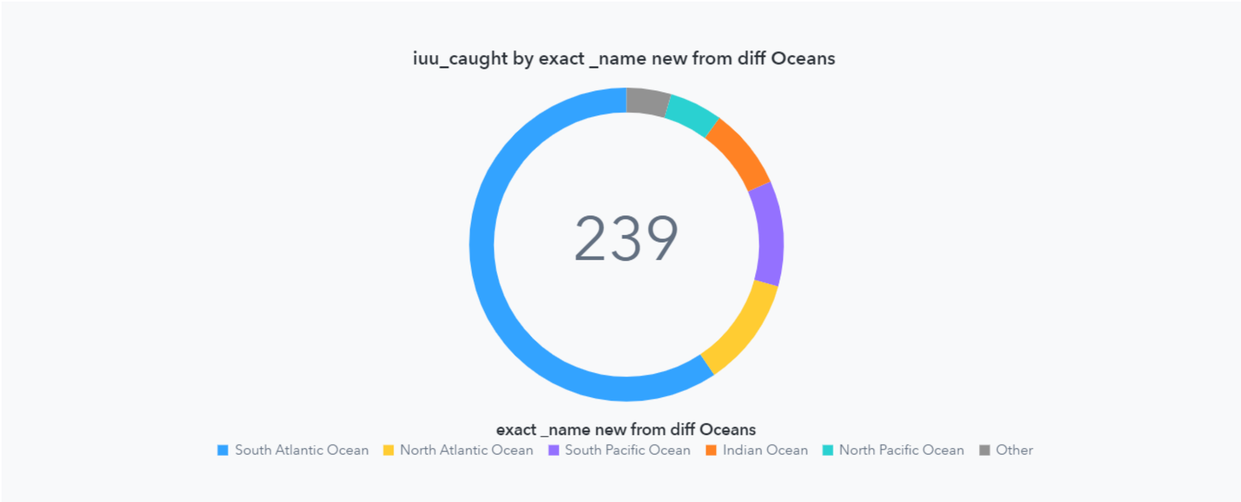

Figure 2. IUU Vessels by Oceans

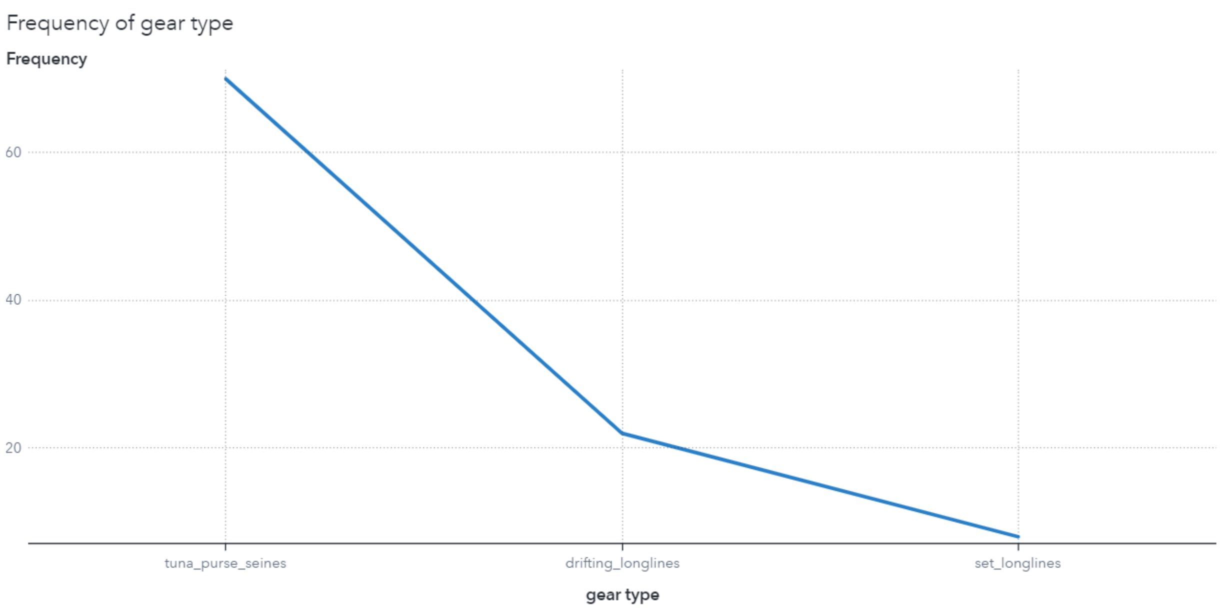

Figure 3. Frequency of Gear Type for IUU vessels

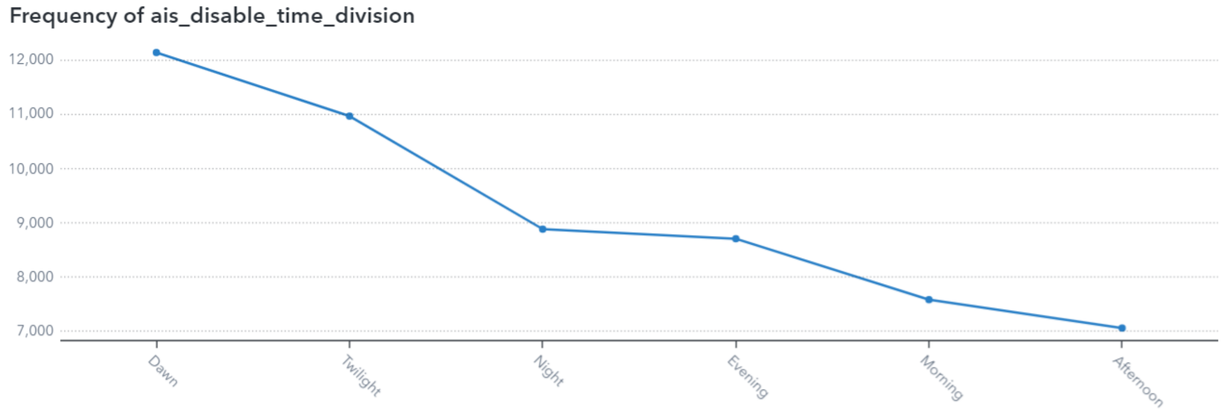

Figure 4. Frequency of Ais_disable_time vs. Time of day

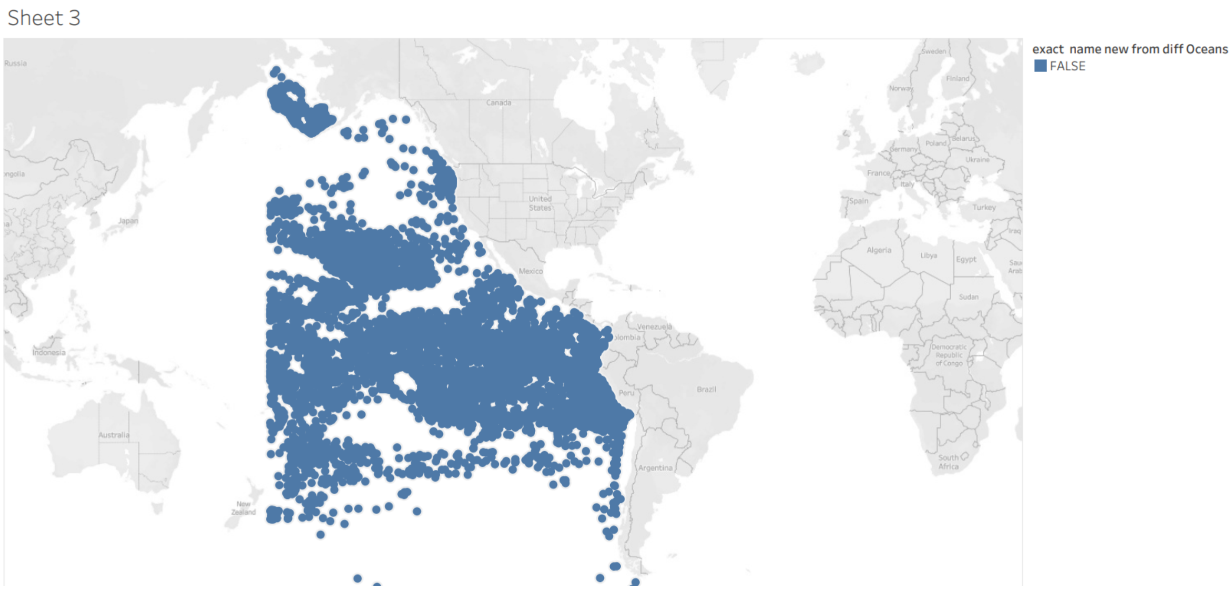

Figure 5. False entries in the Pacific Ocean