INST 354
Penelope Newcomb

AI Model Implementation Through Logical Regression

Global Shark Catch Risk Hotspots

1.0 Introduction

Using Burns, E.S., et al.'s assessment of tuna regional fisheries management organizations (tRFMO) interactions with sharks occurring at industrial longline fisheries, I evaluated the location of these interactions in combination with species documented. Given the more than sixty-one different data points noted for each shark interaction, I focused on the presence or absence of each species in the region and looked to see if one could predict the likelihood of a species sighting whilst taking the passage of time into consideration. With this data, I looked to answer the following questions:

1. How can interaction trends of currently threatened species predict other species at possible future risk?
2. In which zones/oceans are shark interactions most common?
3. What do these locations [where sharks are encountered] have in common and to what extent are these spots unsustainably exploiting the fish population to the point of current or future habitat degradation?

In my regression and analysis, I examined the rates of interactions between sharks and industrial fisheries to see if any preliminary hypotheses could be drawn. I then divided the data between species to see how accurately my model could predict a species presence given the additional features including year, fishing zone, and latitude and longitude.

2.0 Data Wrangling

With over thirteen thousand encounters documented, large portions of my data needed to be made more accessible for logical regression analysis. It is important to note that large portions of the data used were also culled entirely. This is because of the limited time available to properly sort through variables that could provide more information on the fisheries and oceans where these interactions occurred, including the columns ‘catch’ and ‘catch_units’ and statistical information giving us a better understanding of the water conditions when these fisheries caught the desired tuna and encountered a shark in the process. The columns kept in the data frame included: 'year', 'latitude', 'longitude', 'species_sciname', ‘zone’, and 'pres_abs'. This allowed me to see how features such as location affected the sighting of a specific shark species, particularly over time.

3.0 Classification Report

Analysing longline fishery interactions with sharks allowed for the measurement of rates of presence and absence of sharks in a region. The model had enough data to predict the values of its precision, recall and F1 score.

	Precision	Recall	F1 Score
Absent	0.81	0.93	0.87
Present	0.60	0.33	0.42

Figure 1: tRFMO Absense/Presence Classification Report

For interactions where a shark was present, the precision is 0.60, the recall is 0.33, and the F1 score is 0.42. This means that the proportion of interactions where a shark encounter was predicted is 0.60. However, the proportion of interactions that were accurate in their prediction of sharks being present was only 0.33. The F1 score displays that this model only measured actually at a proportion of 0.42. The model was more successful in measuring interactions where a shark was not present, the precision being 0.81, the recall at 0.93, and the F1 score at 0.87. There is a greater proportion of interactions predicted by the model where a shark will not be present at 0.81. Out of these predictions, the proportion of these encounters where a shark was not present is 0.93. The F1 score shows the model can predict this with an accuracy of 0.87.

4.0 Confusion Matrix

While measuring the rates of presence and absence of sharks in a region, I was also able to compare the true and false, positive and negative rates of the values predicted by the model.

	Predicted Values		
Actual Values		Absent	Present
	Absent	2210	237
	Present	771	157

Figure 2: tRFMO Absense/Presence Confusion Matrix

Out of the 2,447 interactions where sharks were not encountered, ninety percent of these were correctly predicted by the model. Only ten percent of these encounters were incorrectly predicted. This is in stark contrast to the meer 928 interactions fisheries had where sharks were encountered. 771 of these shark encounters were inaccurately predicted as sharks being absent and only 157 of interactions (or 17%), were correctly predicted. The confusion matrix reinforces what was seen in the classification report, showing that the model is more accurate at predicting whether a shark will not be encountered rather than whether a shark will be encountered.

5.1 Classification Reports for Different Shark Species

The precision, recall, and F1 scores for the *alopias* species of sharks are all zero.[1] The same was reported in the following species: carcharhinidae, carcharhinus longimanus, lamna nasus, lamnidae, and sphyrna. These scores are the result of very few interactions in the dataframe, making the model unable to accurately predict presence proportionality for these shark species. As seen in the species unspecified classification report, the model is more accurate for shark absent interactions than shark present interactions.

For interactions where *alopias* are not present; the precision is 0.96, the recall is 1.00, and the F1 score is 0.98. This means that 0.96 proportion of interactions with no *alopias* identified predicted where no *alopias* were actually identified. The proportion of predicted interactions with no *alopias* who actually had no identified *alopias* was 1.00. The F1 score shows that the model can predict this with a measured accuracy of 0.98.

For interactions where *carcharhinidae* are not present; the precision, recall, and F1 score were all 1.00. This means that for all interactions with no *carcharhinidae* identified predicted, no *carcharhinidae* were actually identified. The F1 score shows that the model can predict this with a measured accuracy of 1.00.

The *carcharhinus falciformis* provides a more diverse range of proportions in their classification report. The precision, recall, and F1 score for when *carcharhinus falciformis* were present in an interaction and predicted to be present are 0.73, 0.64, and 0.68 accordingly. This shows that the proportion of *carcharhinus falciformis* interactions where *carcharhinus falciformis* were predicted to be present is 0.73, while the proportion of *carcharhinus falciformis* predicted to be present and actually were present is 0.64. The F1 score shows that the model can predict this with a measured accuracy of 0.68. The precision, recall, and F1 score for when

[1] Given the focus on the differences between nine distinct shark species, all classification reports will be located at the end of this section to enhance readability.

carcharhinus falciformis were absent in an interaction and predicted to be absent are 0.75, 0.82, and 0.79 accordingly. This shows that the proportion of *carcharhinus falciformis* interactions where *carcharhinus falciformis* were predicted to be present is 0.75, while the proportion of *carcharhinus falciformis* predicted to be present and actually were present is 0.82. The F1 score shows that the model can predict this with a measured accuracy of 0.79. Once again our absent predictions are more accurate, although less so in interactions with *carcharhinus falciformis*.

The precision, recall, and F1 score for when *carcharhinus longimanus* were absent in an interaction and predicted to be absent are 0.95, 1.00, and 0.97. This shows that the proportion of *carcharhinus longimanus* interactions where *carcharhinus longimanus* were predicted to be present is 0.95, while the proportion of *carcharhinus longimanus* predicted to be present and actually were present is 1.00. The F1 score shows that the model can predict this with a measured accuracy of 0.97.

The precision, recall, and F1 score for when *lamna nasus* were absent in an interaction and predicted to be absent are 0.94, 1.00, and 0.97. This shows that the proportion of *lamna nasus* interactions where *lamna nasus* were predicted to be present is 0.94, while the proportion of *lamna nasus* predicted to be present and actually were present is 1.00. The F1 score shows that the model can predict this with a measured accuracy of 0.97.

The *lamnidae* also has a more diverse range of proportions in their classification report. The precision, recall, and F1 score for when *lamnidae* were present in an interaction and predicted to be present are 0.80, 0.99, and 0.88 accordingly. This shows that the proportion of *lamnidae* interactions where *lamnidae* were predicted to be present is 0.80, while the proportion of *lamnidae* predicted to be present and actually were present is 0.99. These are some of the highest proportions seen in any of the present variables of the nine shark species. The F1 score shows that the model can predict this with a measured accuracy of 0.88. The precision, recall, and F1 score for when *lamnidae* were absent in an interaction and predicted to be absent are 0.93, 0.27, 0.42. This shows that the proportion of *lamnidae* interactions where *lamnidae* were predicted to be present is 0.93, while the proportion of *lamnidae* predicted to be present and actually were present is 0.27. Having learned about the likelihood of sharks not being interacted with, the model seems less likely to assume that a shark will be seen. The F1 score shows that the model can predict this with a measured accuracy of 0.88.

This trend continues with *prionace glauca*'s classification report. The proportion of *prionace glauca* interactions where *prionace glauca* were predicted to be present is 0.86, while the proportion of *prionace glauca* predicted to be present and actually were present is 1.00.The F1 score shows that the model can predict this with a measured accuracy of 0.92. The precision, recall, and F1 score for when *prionace glauca* were absent in an interaction and predicted to be absent are 1.00, 0.12, and 0.21. This shows that the proportion of *prionace glauca* interactions where *prionace glauca* were predicted to be present is 1.00, while the proportion of *prionace glauca* predicted to be present and actually were present is 0.12 at a rate of only 0.21 accuracy.

The *sharks nei* classification report appears similar to the *carcharhinus falciformis* one as the preportioned values are quite similar for both variables. The proportion of *sharks nei* interactions where *sharks nei* were predicted to be present is 0.73, while the proportion of *sharks nei* predicted to be present and actually were present is 0.70.The F1 score shows that the model can predict this with a measured accuracy of 0.72. The precision, recall, and F1 score for when *sharks nei* were absent in an interaction and predicted to be absent are 0.73, 0.76, and 0.74. This shows that the proportion of *sharks nei* interactions where *sharks nei* were predicted to be present is 0.73, while the proportion of *sharks nei* predicted to be present and actually were present is 0.76 at a rate of 0.74 accuracy.

For interactions where *sphyrna* are not present; the precision is 0.97, the recall is 1.00, and the F1 score is 0.98. This means that 0.97 proportion of interactions with no *sphyrna* identified predicted where no *sphyrna* were actually identified. The proportion of predicted interactions with no *sphyrna* who actually had no identified *sphyrna* was 1.00. The F1 score shows that the model can predict this with a measured accuracy of 0.98.

5.2 Classification Report Figures for Different Shark Species

	Precision	Recall	F1 Score
Absent	0.96	1.00	0.98
Present	0.00	0.00	0.00

Figure 3.1: Alopias Classification Report

	Precision	Recall	F1 Score
Absent	1.00	1.00	1.00
Present	0.00	0.00	0.00

Figure 3.2: Carcharhinidae Classification Report

	Precision	Recall	F1 Score
Absent	0.75	0.82	0.79
Present	0.73	0.64	0.68

Figure 3.3: Carcharhinus Falciformis Classification Report

	Precision	Recall	F1 Score
Absent	0.95	1.00	0.97
Present	0.00	0.00	0.00

Figure 3.4: Carcharhinus Longimanus Classification Report

	Precision	Recall	F1 Score
Absent	0.94	1.00	0.97
Present	0.00	0.00	0.00

Figure 3.5: Lamna Nasus Classification Report

	Precision	Recall	F1 Score
Absent	0.93	0.27	0.42
Present	0.80	0.99	0.88

Figure 3.6: Lamnidae Classification Report

	Precision	Recall	F1 Score
Absent	1.00	0.12	0.21
Present	0.86	1.00	0.92

Figure 3.7: Prionace Glauca Classification Report

	Precision	Recall	F1 Score
Absent	0.73	0.76	0.74
Present	0.73	0.70	0.72

Figure 3.8: Sharks Nei Classification Report

	Precision	Recall	F1 Score
Absent	0.97	1.00	0.98
Present	0.00	0.00	0.00

Figure 3.9: Sphyrna Classification Report

6.1 Confusion Matrices for Different Shark Species

Declining shark populations can be visualized in many of these confusion matrices as little to no sharks were predicted.[2] For *alopias* this means that for the 361 encounters, none of them were predicted to be interactions with *alopias*. For the correctly predicted zero interactions, there were only fourteen actual encounters, making the numbers too small for the model to accurately represent and predict. Although the exact number of actual encounters with shark species vaires, this trend is replicated in the *carcharhinidae*, *carcharhinus longimanus*, *lamna nasus*, and *sphyrna*.

The *carcharhinus falciformis* however, sees a different pattern of predictions. For the 215 fishery interactions in waters known to be habitated by *carcharhinus falciformis* where this species was not encountered, 177 of them were predicted to be absent while only 38 were predicted to be present. For the 160 fishery interactions in waters known to be habitated by *carcharhinus falciformis* where this species was encountered, 58 of them were predicted to be absent while 102 were predicted to be present.

For the species *lamnidae*, there were 97 fishery interactions in waters known to be habitated by *lamnidae* where this species was not encountered, 26 of them were predicted to be absent while 71 were predicted to be present. For the 278 fishery interactions in waters known to be habitated by *lamnidae* where this species was encountered, only 2 of them were predicted to be absent while 276 were predicted to be present.

Prionace glauca had 59 fishery interactions in waters known to be habitated by *prionace glauca* where this species was not encountered, 7 of them were predicted to be absent while 52 were predicted to be present. For the 316 fishery interactions in waters known to be habitated by

[2] Given the focus on the differences between nine distinct shark species, all confusion matrices will be located at the end of this section to enhance readability.

prionace glauca where this species was encountered, 0 of them were predicted to be absent while 316 were predicted to be present.

Finally, *sphyrna* had 192 fishery interactions in waters known to be habitated by *sphyrna* where this species was not encountered, 145 of them were predicted to be absent while 47 were predicted to be present. For the 183 fishery interactions in waters known to be habitated by *prionace glauca* where this species was encountered, 54 of them were predicted to be absent while 129 were predicted to be present.

6.2 Confusion Matrix Figures for Different Shark Species

	Predicted Values		
Actual Values		Absent	Present
	Absent	361	0
	Present	14	0

Figure 4.1: Alopias Confusion Matrix

	Predicted Values		
Actual Values		Absent	Present
	Absent	374	0
	Present	1	0

Figure 4.2: Carcharhinidae Confusion Matrix

	Predicted Values		
Actual Values		Absent	Present
	Absent	177	38
	Present	58	102

Figure 4.3: Carcharhinus Falciformis Confusion Matrix

	Predicted Values		
Actual Values		Absent	Present
	Absent	356	0
	Present	19	0

Figure 4.4: Carcharhinus Longimanus Confusion Matrix

	Predicted Values		
Actual Values		Absent	Present
	Absent	351	0
	Present	24	0

Figure 4.5: Lamna Nasus Confusion Matrix

	Predicted Values		
Actual Values		Absent	Present
	Absent	26	71
	Present	2	276

Figure 4.6: Lamnidae Confusion Matrix

	Predicted Values		
Actual Values		Absent	Present
	Absent	7	52
	Present	0	316

Figure 4.7: Prionace Glauca Confusion Matrix

	Predicted Values		
Actual Values		Absent	Present
	Absent	145	47
	Present	54	129

Figure 4.8: Sharks Nei Confusion Matrix

	Predicted Values		
Actual Values		Absent	Present
	Absent	363	0
	Present	12	0

Figure 4.9: Sphyrna Confusion Matrix

7.0 Fairness Analysis

With this dataset I was able to evaluate the differences in precision, recall, and F1 score between nine different shark species and their presence in waters where these sharks had previously been seen. Some present shark species had precision, recall, and F1 score all at zero. This was because of two reasons; one: there were very few encounters with this shark species and two: there was too little data for this model to make an accurate prediction.

The classification reports and confusion matrices show disproportionate numbers of sharks being predicted as absent. This can be seen in the results of *alopias*, *carcharhinidae*, *carcharhinus longimanus*, *lamna nasus*, and *sphyrna*. The model skews to predict that a shark will not be encountered in an interaction, to the detriment of accuracy in *lamnidae* and *prionace glauca* results.

With this in mind the logical regression model created is not fair. However, this does not mean that the model's results should be dismissed completely. In comparing these different species it is interesting to also put conservation status into consideration. *Lamnidae* and *prionace glauca* were both skewed to be more likely present than absent. These species are only classified as vulnerable or near threatened by the International Union for Conservation of Nature Red List of Threatened Species, whereas all other species with a skewed distribution towards being absent are classified as endangered or critically endangered. Without time constraints and taking more factors into consideration, I believe that this model could be developed into a more accurate prediction model for a shark species presence or absence in a region over time.