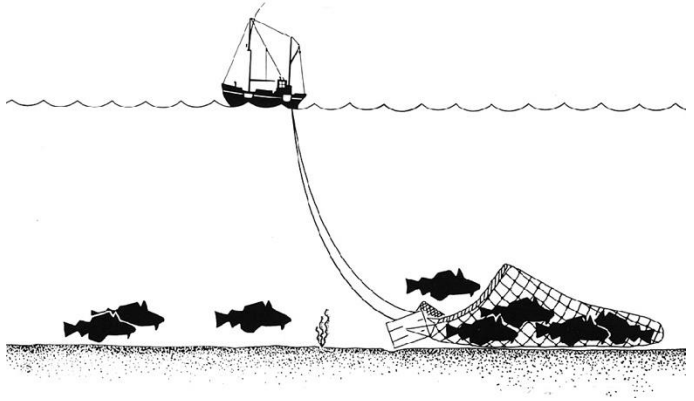
Detecting Fishing Activity in Trawling Vessels with Automatic Identification Systems (AIS) and Machine Learning

David Caspers

15 Sept 2024

Trawling & IUU



Trawling

- Drags funnel-shaped net (trawl)to catch fish
- Targets species near the ocean floor (bottom trawling) or in the water column (midwater trawling).

Effects of Trawling

- Seabed Destruction
- Water Chemistry Alteration
- Bycatch and Waste

Illegal, Unreported, & Unregulated (IUU) Fishing

- Economic Impact:
  - $10-23B estimated annual cost
  - Undermines legitimate fisheries and threatens coastal communities
- Environment Degradation
- Biodiversity Loss

About the Data

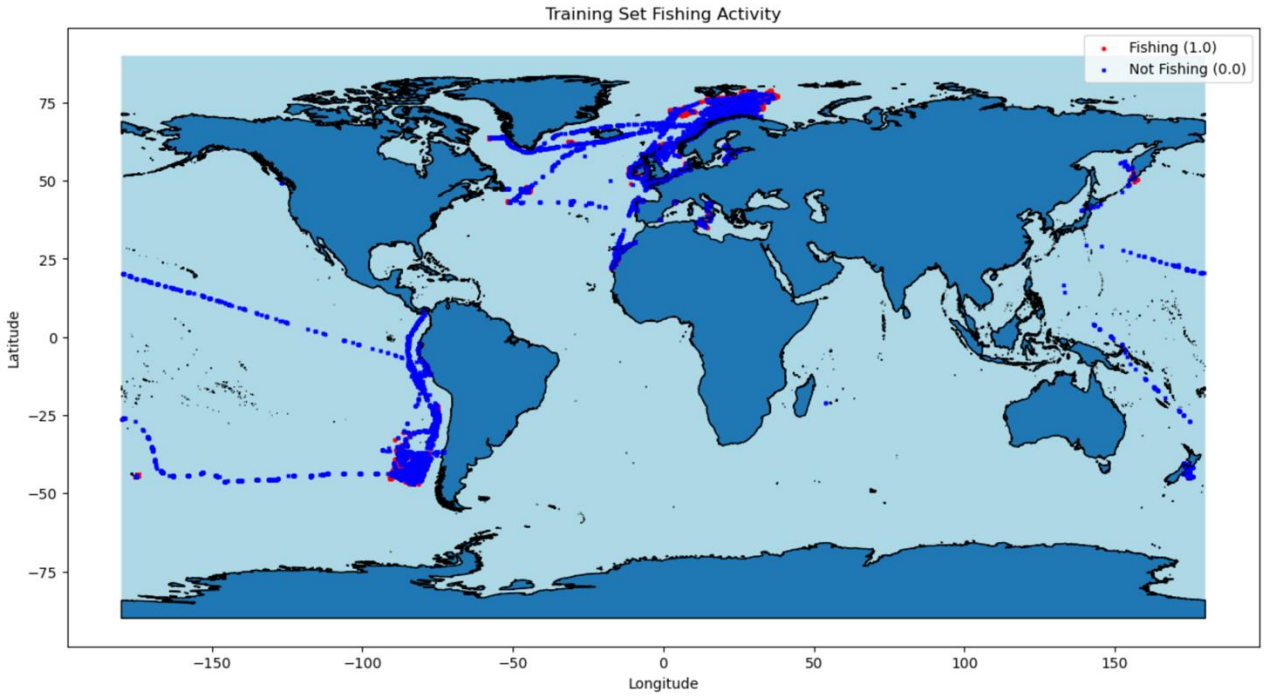
•**Source**: The dataset is sourced from Global Fishing Watch, an organization that provides open-access data on global fishing activity.

•**Key Data Points**:

- **MMSI**: Unique vessel identifier.
- **Timestamp**: Unix timestamp indicating the precise time of transmission.
- **Geolocation**: Latitude and longitude coordinates of the vessel at the time of transmission.
- **Speed and Course**: Vessel speed (in knots) and course (heading).
- **Distance from Shore/Port**: Additional geographic data indicating proximity to shorelines and ports.
- **Fishing Status**: Labels indicating whether a vessel was engaged in fishing at the time of transmission (e.g., 0 = not fishing, >0 = fishing, -1 = no data).

Observations:

4,369,101 labeled AIS signals taken from 49 trawling vessels across the word

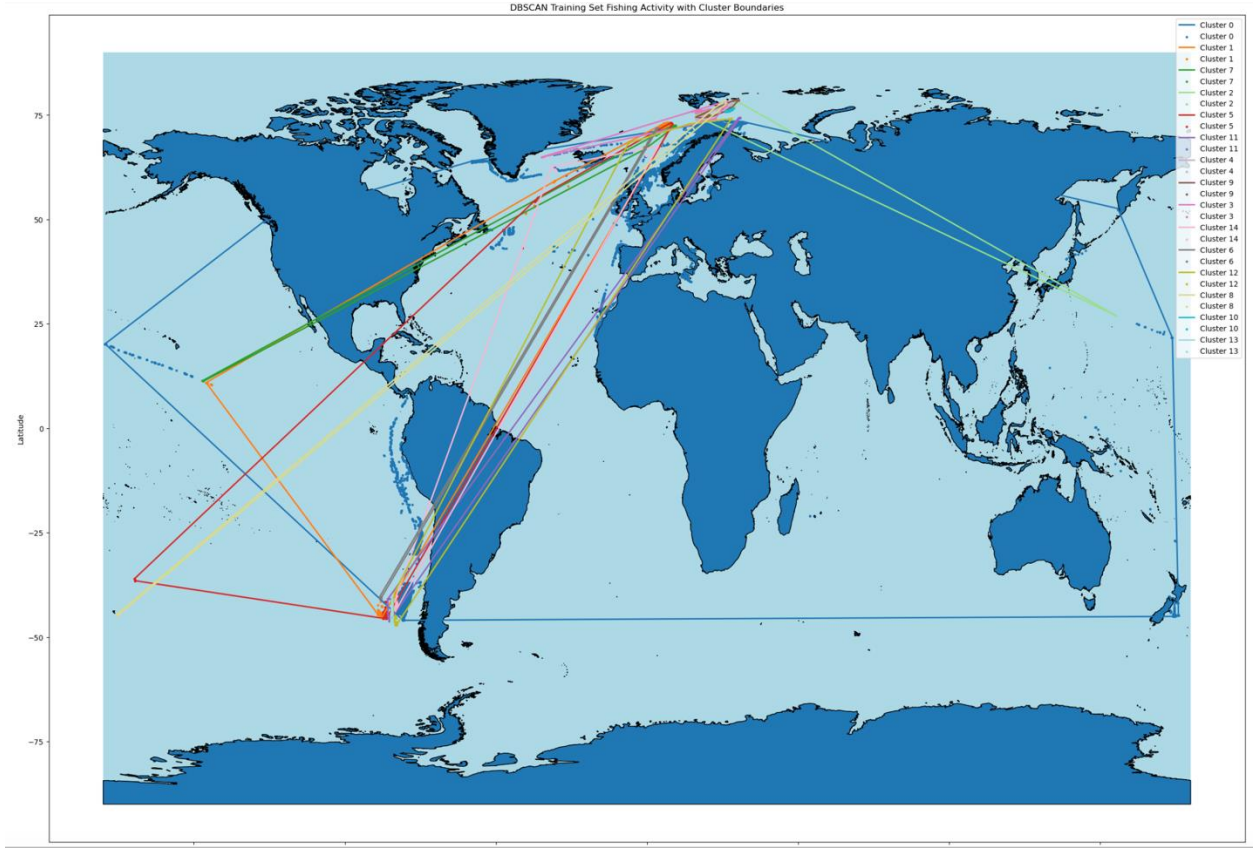


Project Goal

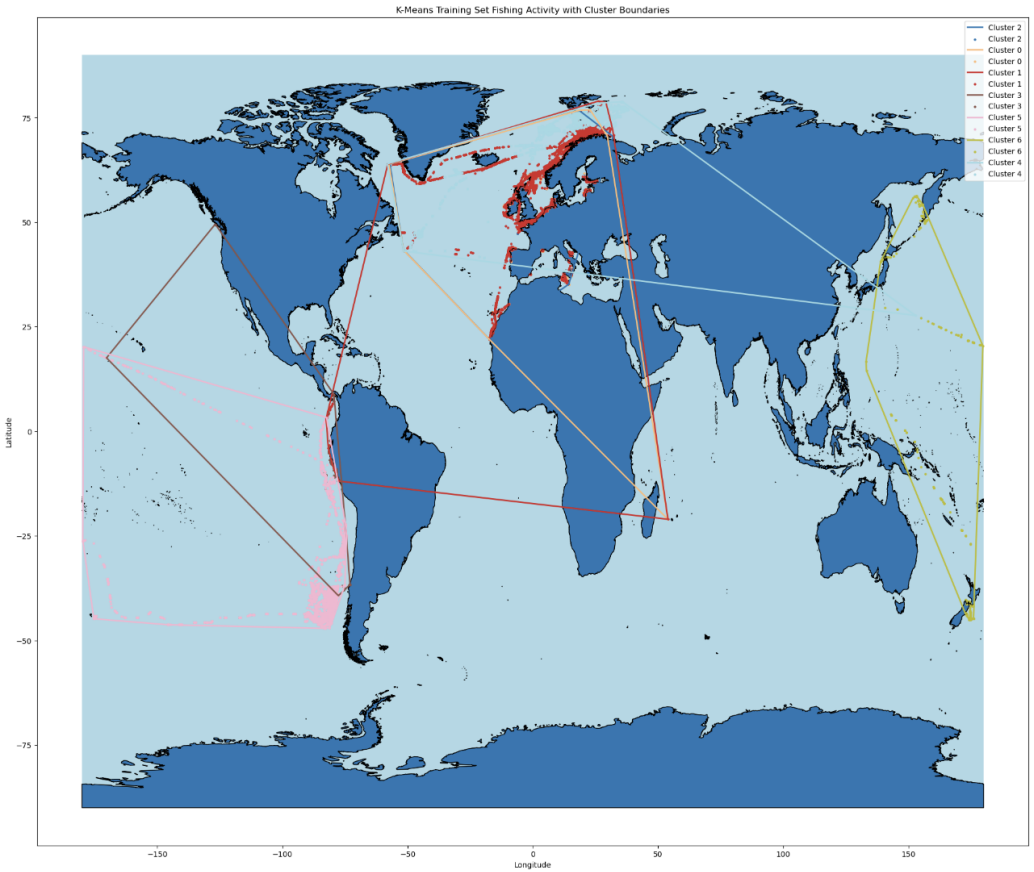
- **Objective:** develop a robust model for detecting fishing activities of trawling vessels using AIS data.
  - Contribute to detection of IUU Fishing detection models by refining fishing detection

Clustering

Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

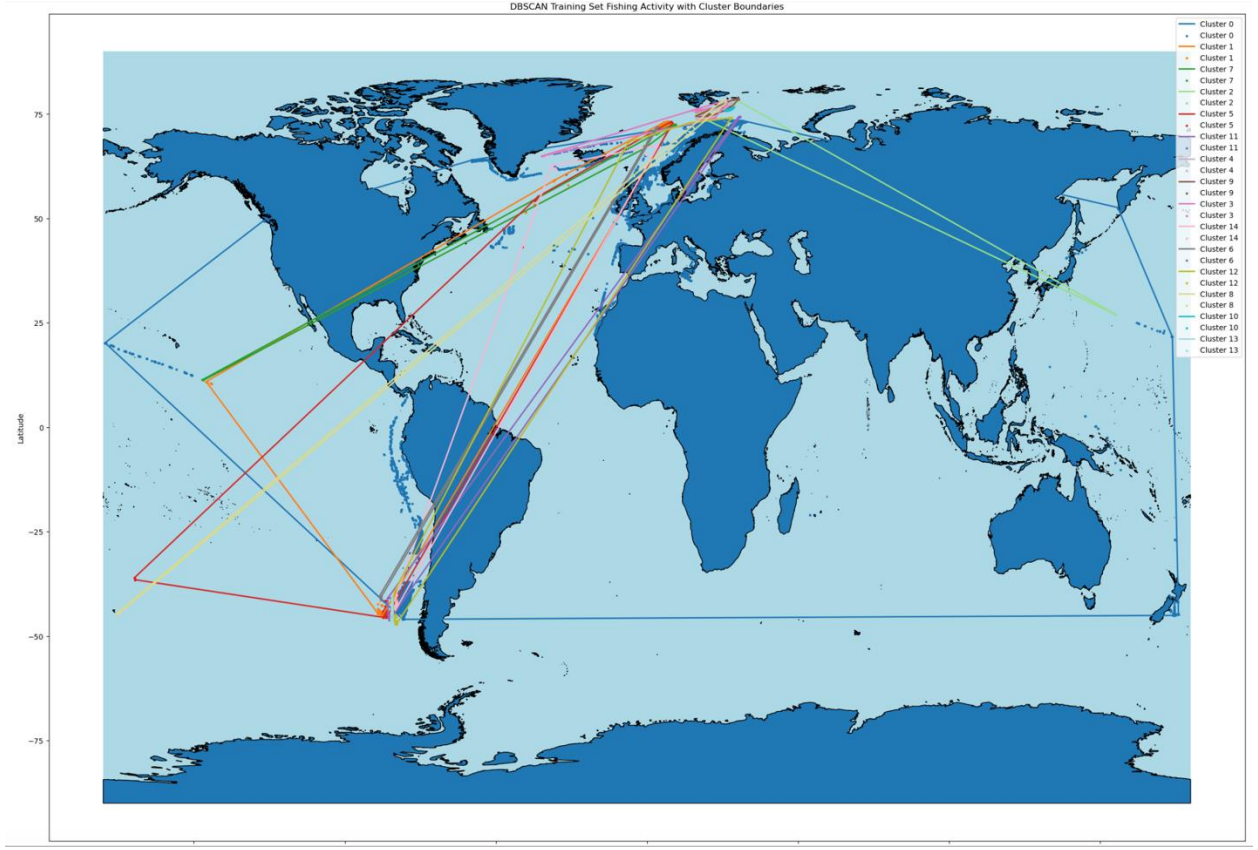


K-Means Clustering



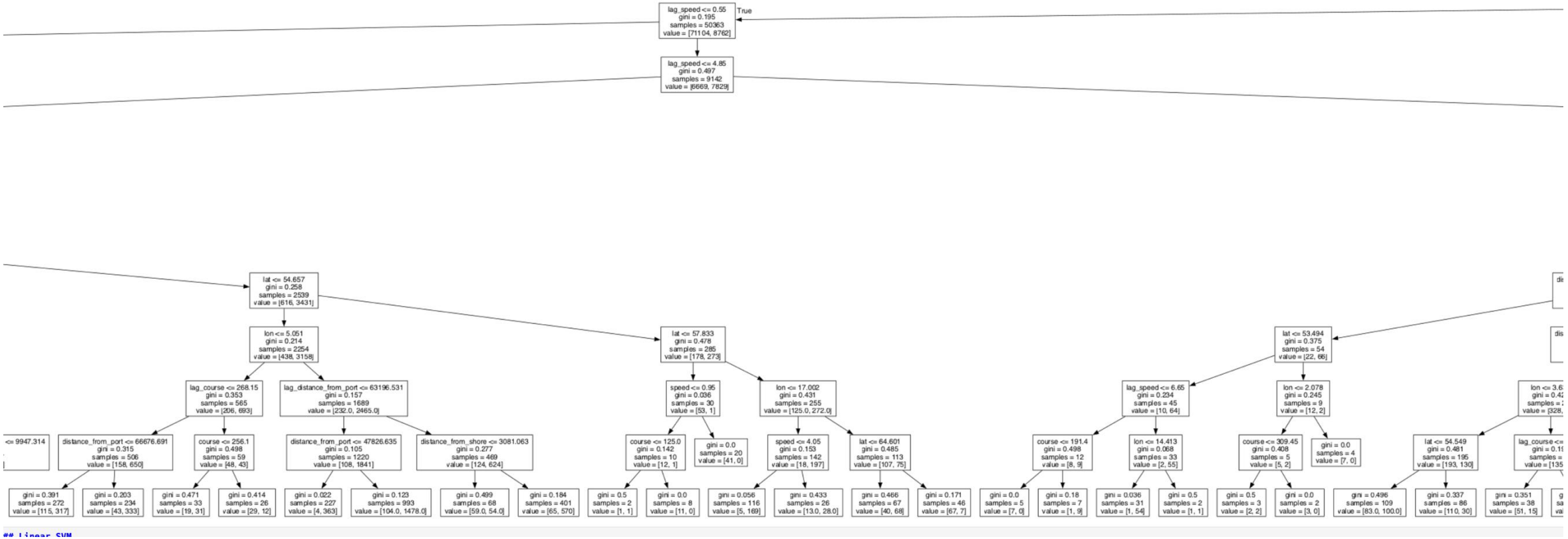
Clustering

Density-Based Spatial Clustering of Applications with Noise (DBSCAN)



Cluster Distribution: 16 Distinct Clusters. Fishing activity concentrated in clusters 1, 2, 3, 4, and 7.
Cluster Characteristics: Clusters 2, 3, and 4 show high fishing percentages, with over 88% of the observations identified as fishing. Cluster 7 is unique with a 100% fishing rate, indicating a specialized area for fishing possibly supported by nearby resources or fish populations.
Noise Points: Cluster -1 (noise) accounts for 4,981 points, approximately 5% of the total dataset, with 53.6% of those points classified as fishing.
Geographical Spread: Clusters with high fishing activity (e.g., 2, 3, and 4) are located far from the port and shore, indicating deep-sea fishing operations.
Variability in Distance and Speed: Significant variations in average distances from port and shore across clusters highlight different fishing strategies and vessel capabilities.

Random Forest Classifier



Random Forest Classifier

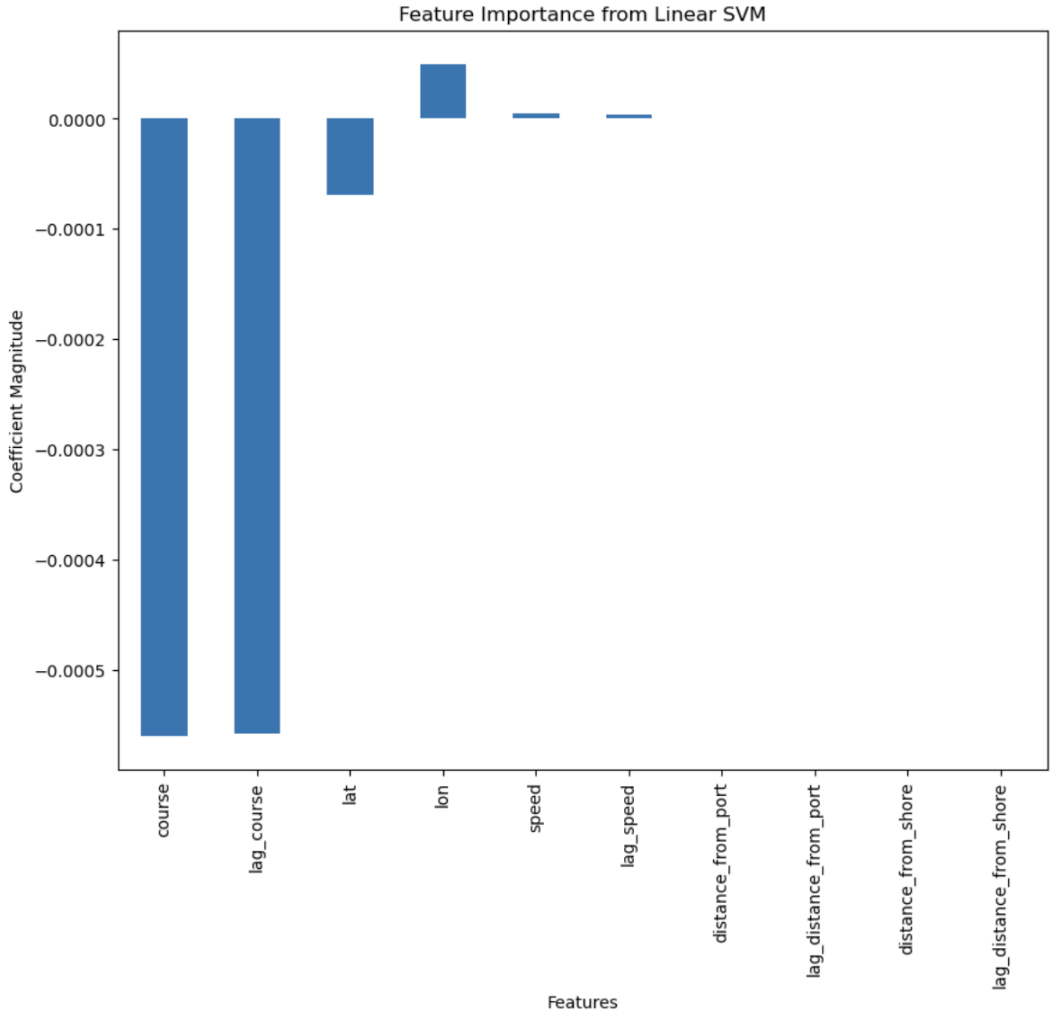
Results Without Lag

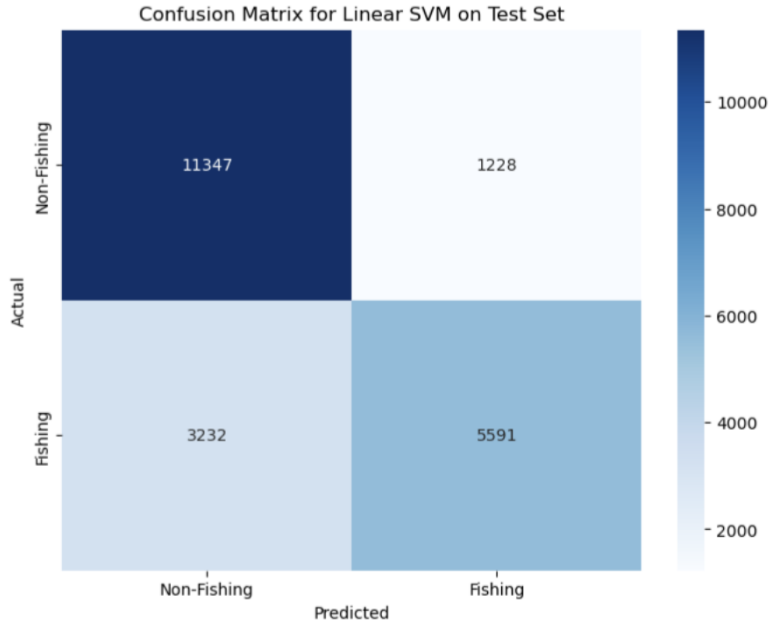
- **Accuracy**: 84.88%
- **Precision**:
  - Non-Fishing: 0.84
  - Fishing: 0.86
- **Recall**:
  - Non-Fishing: 0.92
  - Fishing: 0.75
- **Top 5 Features**:
  - **Speed**: 37.57%
  - **Distance from Shore**: 26.06%
  - **Distance from Port**: 16.09%
  - **Latitude**: 11.95%
  - **Longitude**: 6.60%

Results With Lag

- **Accuracy**: 87.69%
- **Precision**:
  - Non-Fishing: 0.88
  - Fishing: 0.87
- **Recall**:
  - Non-Fishing: 0.91
  - Fishing: 0.80
- **Top 5 Features**:
  - **Lag Speed**: 31.00%
  - **Speed**: 21.54%
  - **Lag Distance from Shore**: 16.94%
  - **Distance from Shore**: 8.80%
  - **Latitude**: 6.65%

Support Vector Machine (Linear Kernal)





Classification Report on test set:

	precision	recall	f1-score	support
0.0	0.78	0.90	0.84	12575
1.0	0.82	0.63	0.71	8823
accuracy			0.79	21398
macro avg	0.80	0.77	0.78	21398
weighted avg	0.80	0.79	0.79	21398

Conclusion

- **Vessel activity is highly variable and noisy**
  - This introduces challenges in accurately identifying consistent fishing patterns, as vessel behavior is not uniform.
- **Time-series data is critical for detecting fishing activity**
  - Sequential changes in vessel behavior, such as speed and movement over time, are key for accurate classification.
- **Speed and distance from shore are the strongest predictors of fishing activity**
  - Speed variations and proximity to shore or deep-sea areas offer clear indicators for distinguishing fishing from non-fishing activities.
- **Future improvements should focus on handling sequential data**
  - Models that can account for the sequential nature of the data such as Long Short-Term Memory (LSTM) Networks or Recurrent Neural Networks (RNN)