Detailed reference : Detecting Illegal Fishing Activities Using Satellite AIS Data

1. Problem Statement

Illegal, Unreported, and Unregulated (IUU) fishing causes:

- Loss of marine resources,
- Economic damage to coastal countries,
- Ecological imbalance.

Many fishing vessels intentionally hide their identity or operate in restricted waters.
This project uses **satellite AIS (Automatic Identification System) data** to **detect suspicious vessel behavior**.

Goal:
Classify vessel behavior as **normal fishing activity** or **potentially illegal fishing** using machine learning.

---

2. What is AIS Data?

Ships continuously broadcast AIS signals containing:

- Vessel identity (MMSI),
- GPS coordinates (latitude, longitude),
- Timestamp,
- Speed & movement behavior.

This data is captured by satellites and processed to identify **when the vessel is fishing vs traveling**.

AIS is legally required for large fishing vessels.
Turning off AIS is a major red flag → indicates possible illegal activity.

---

3. Dataset Description

We used **processed monthly AIS activity summaries** (not raw GPS tracks).

Each row represents a vessel's activity in a specific grid cell on a specific day.

Column	Description	Reason Used
date	Activity date	Time-based signals
year, month	Derived from date	Seasonality pattern
cell_ll_lat	Latitude	Spatial behavior
cell_ll_lon	Longitude	Spatial behavior
flag	Country of vessel registration	Jurisdiction check
geartype	Type of fishing gear used	High-risk gear detection
hours	Total hours in area	Presence intensity
fishing_hours	Active fishing hours	Core fishing behavior indicator
mmsi_present	1 if vessel identity broadcasted, 0 if not	**Turning off AIS is suspicious**

Key Risk Indicators

Behavior	Interpretation
mmsi_present = 0	Vessel turned off tracking → **Very high illegal suspicion**
High fishing_hours far offshore	Deep ocean unlicensed fishing
Foreign flag inside another country's EEZ	Territorial violation
High-risk geartype (e.g., drifting longlines)	Typically associated with IUU fishing

---

4. Data Preprocessing

1. Checked missing values (none).
2. Converted date to datetime format.
3. Created time-based features:
    - day_of_week
    - is_weekend
    - day_of_month
4. Created geographic features:
    - distance_from_equator = abs(lat)
    - hemisphere = north / south
5. Prepared target variable: suspicious (0 = normal, 1 = suspicious).

---

5. Feature Engineering

To evaluate vessel behavior, we used:

Feature Name	**Interpretation**
fishing_efficiency = fishing_hours / hours	How aggressively vessel fishes
mmsi_present	Identity hiding behavior
geartype	Operational fishing method
day_of_week + is_weekend	Human fishing scheduling patterns
distance_from_equator	Fishing zone type (tropical vs deep ocean)

---

6. Handling Imbalanced Data

Often, **normal vessels > suspicious vessels**.
We use **SMOTE (Synthetic Minority Oversampling Technique)** to balance both classes.

This ensures the model does **not become biased toward normal class**.

---

7. Model Training

We tested multiple classification models:

Model	Reason
Logistic Regression	Baseline explainable model
Decision Tree	Captures nonlinear patterns
Random Forest	Robust, handles complex feature interactions
XGBoost	High-performance boosted decision trees
SVM	Boundary-based classification
KNN	Distance-based similarity detection
Naive Bayes	Probabilistic baseline model

---

8. Performance Metrics

Used:

- Accuracy
- Precision (false positive control)
- Recall (illegal detection sensitivity)
- F1-score (balanced evaluation)
- ROC-AUC (separation power)

Important Goal:
High **recall** → avoid missing illegal activity.

---

9. Interpretation of Model Output

If model predicts **suspicious**:

- Vessel may be fishing in restricted waters,
- Or using high-risk gear in protected zones,
- Or turning off AIS to hide identity.

These vessels get **flagged for inspection** by maritime authorities.

---

10. Real-World Application

Government agencies can use this model to:

- Detect illegal fishing in real time,
- Deploy coast guard patrols efficiently,
- Protect marine resources,
- Reduce economic and ecological damage.

NGOs (e.g., Global Fishing Watch) use similar logic to **publish illegal activity heatmaps**.

---

11. Core Explanation to Say Verbally

"We analyzed satellite AIS vessel activity to detect suspicious fishing patterns.
The model identifies vessels that turn off AIS, spend abnormally high time fishing, or operate outside permitted regions.
By classifying vessel behavior, we can support enforcement against illegal fishing."

For self-reference :

12. How do you know your model makes sense?

- Features used have **direct real-world meaning** (not abstract).
- Illegal vessels **reliably** show behavior patterns (AIS-off, deep-sea, high-effort).
- Model is **validated using real fishing behavior rules**.

---

13. What is the strongest illegal indicator?

mmsi_present = 0 (AIS turned off)
because legal vessels have no reason to hide identity.

---

14. What improvements can be made?

- Use **AIS trajectory time-series**, not just aggregated rows.
- Integrate **Exclusive Economic Zone boundary datasets**.
- Apply **LSTM / GRU models** for movement pattern recognition.
- Add satellite optical/radar verification (SAR overlay).