



# **COLEGIO DE CIENCIAS E INGENIERÍAS**

## **INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN**

**Planificación para el Desarrollo del Proyecto Integrador**

**Tutor: Felipe Grijalva**

**Autor: Erick Suárez**

**Quito – Ecuador**

**2025**



### 1. Título del Proyecto:

Detección no supervisada de anomalías, usando series de tiempo, en trayectorias AIS de pesqueros para vigilancia de la Reserva Marina de Galápagos.

### 2. Relevancia y Justificación:

La Reserva Marina de Galápagos (GMR o RMG) es uno de los ecosistemas marinos más sensibles del planeta, y representa para nuestro país el área protegida más importante. Durante décadas se ha luchado contra la presión pesquera y actividades de pesca ilegales o no reportadas (IUU) que han amenazado su integridad ecológica. Hoy en día existen datos AIS (Auto Identification System) de cobertura global que permiten observar el comportamiento de embarcaciones a alta resolución temporal; sin embargo, convertir estos datos masivos en inteligencia operativa sigue siendo un reto. Mi motivación por realizar este proyecto es que soy consciente de la problemática y del daño que este tipo de actividades hace al ecosistema de la GMR, por lo que me interesa realizar una herramienta para identificar patrones anómalos y contribuir a la seguridad y bienestar de nuestros bienes como país.

Este proyecto propone un pipeline de detección no supervisada de anomalías sobre series de tiempo de trayectorias AIS de buques pesqueros alrededor de Galápagos. La propuesta es relevante porque:

- 1) Aprende comportamiento normal a partir de datos históricos (puntos con `is_suspicious=0`) y detecta desviaciones sin requerir etiquetas exhaustivas.
- 2) Integra contexto geoespacial y ambiental (polígonos de la GMR, distancia a costa/puertos, batimetría GEBCO).
- 3) Genera señales accionables (p. ej., ventanas y trayectorias anómalas) que pueden priorizar vigilancia.

### 3. Objetivos

#### a. Generales

Desarrollar y evaluar un sistema no supervisado que modele trayectorias AIS como series de tiempo y detecte anomalías indicativas de posible pesca ilegal o comportamientos atípicos en torno a la GMR.

#### b. Específicos

Integrar datos AIS con capas geoespaciales y batimetría para generar un dataset enriquecido.

Diseñar un proceso de serialización temporal por embarcación para representar trayectorias como series de tiempo.

Construir ventanas de observación que permitan modelar dinámicas de navegación a corto plazo.



Aplicar modelos de detección de anomalías no supervisados sobre comportamientos normales de navegación.

Evaluar el desempeño de los modelos utilizando métricas robustas de detección de anomalías.

Desarrollar un prototipo reproducible con scripts de preprocesamiento y reporte técnico.

#### 4. Estado del Arte

El presente estado del arte está realizado con el objetivo de presentar los principales y mas recientes artículos de investigación relacionados al tema de “detección de pesca ilegal” usando Machine Learning. Además se detallarán los principales conceptos relacionados con el área de estudio.

La detección de pesca ilegal es un área crítica para la conservación marina, especialmente en reservas protegidas como la Reserva Marina de Galápagos. Los Sistemas de Identificación Automática (AIS) constituyen la fuente principal de información, ya que registran datos de posición, velocidad, rumbo e identificadores únicos de embarcaciones (MMSI). Estos registros permiten reconstruir trayectorias marítimas y, al combinarse con capas geoespaciales y batimetría, se puede contextualizar el comportamiento de los barcos en relación con zonas protegidas, profundidades y rutas habituales de pesca.

En el ámbito de machine learning aplicado a series de tiempo AIS, los modelos de detección de anomalías permiten diferenciar comportamientos regulares de posibles actividades ilegales. Los enfoques supervisados, como Random Forest, SVM, KNN, regresión logística y redes neuronales (RNN y CNN), requieren datos previamente etiquetados como legales o ilegales y han mostrado altos niveles de precisión en escenarios locales. Sin embargo, su dependencia de etiquetas limita la aplicabilidad en regiones con información incompleta. Por otro lado, los modelos no supervisados y semi-supervisados (HDBSCAN, One-Class SVM, Isolation Forest, pseudo-labelling) buscan patrones inusuales sin necesidad de datos etiquetados, lo que los hace prometedores para detectar anomalías en contextos con información limitada. Los enfoques híbridos combinan heurísticas, como geofencing o reglas basadas en velocidad y rumbo, con algoritmos de machine learning para aumentar la robustez del sistema de detección.

Diversos estudios recientes evidencian la eficacia de estas metodologías. Dumpit & Monreal (2025) aplicaron RNN para detectar pesca ilegal en Filipinas, demostrando la viabilidad de los enfoques supervisados a nivel local. Motiramani et al. (2023) utilizaron Random Forest y SVM sobre datos de Global Fishing Watch, logrando hasta un 99% de precisión, aunque con riesgo de sobreajuste. Huang et al. (2021) implementaron un sistema en tiempo real basado en edge computing y clustering espaciotemporal, mejorando la identificación de anomalías. Brown et al. (2024) mostraron que la combinación de pseudo-labelling y modelos semi-supervisados permite detectar actividades sospechosas incluso con datos incompletos. Xu (2023) alcanzó altos niveles de accuracy y recall en longliners usando Random Forest y Gradient Boosting. Sharma & Panjkar (2025) desarrollaron Fishing Forecast Guardian, un sistema híbrido que combinó CNN y Random Forest sobre AIS, alcanzando hasta 90% de precisión y validando casos reales de pesca ilegal en Asia.



A partir de estos antecedentes, se observa un predominio de enfoques supervisados, con limitaciones en escalabilidad a regiones con información incompleta, mientras que los métodos no supervisados ofrecen un campo prometedor pero con retos interpretativos. Además, la mayoría de investigaciones se centran en contextos regionales como Asia y Filipinas, dejando vacíos en zonas ecológicamente sensibles como Galápagos. El proyecto integrador busca aportar un enfoque replicable mediante la construcción de un dataset enriquecido con AIS y capas geoespaciales, la serialización de trayectorias en ventanas temporales y la aplicación de modelos no supervisados para detectar anomalías, contribuyendo así a la vigilancia de la pesca ilegal en esta área protegida.

**Machine Learning (Aprendizaje Automático):** Es un área de la inteligencia artificial que permite a los sistemas aprender patrones a partir de datos y realizar predicciones o clasificaciones sin necesidad de ser programados explícitamente para cada tarea. En la detección de pesca ilegal, los algoritmos de machine learning permiten identificar comportamientos atípicos de embarcaciones a partir de sus trayectorias y características de movimiento registradas por AIS, clasificando actividades como legales o sospechosas.

**Series de Tiempo:** Se refiere a conjuntos de datos secuenciales registrados en intervalos de tiempo específicos, donde cada registro depende del anterior y puede reflejar patrones temporales. En el contexto de AIS, cada embarcación genera una serie de tiempo con sus posiciones, velocidades y rumbos, lo que permite reconstruir trayectorias y analizar cambios en el comportamiento de navegación a lo largo del tiempo.

**Detección de Anomalías en AIS:** Consiste en identificar comportamientos que se desvían de lo esperado o de la norma, lo que podría indicar actividades ilegales o riesgosas. Esto se logra mediante modelos supervisados, no supervisados o híbridos, que comparan las trayectorias y características de movimiento de las embarcaciones contra patrones típicos de navegación. La detección de anomalías es fundamental para monitorear áreas protegidas, optimizar la vigilancia y priorizar la acción de las autoridades marítimas.

## 5. Metodología de Trabajo

### 1. Datos y limpieza

- AIS (lat, lon, timestamp, speed, course, is\_fishing, distancias a costa/puertos).
- Shapefile World Marine Heritage Sites (filtrado “Galápagos Islands”) y unión espacial para in\_protected\_area.
- Batimetría GEBCO → depth por muestreo raster.
- Reglas base para construir is\_suspicious inicial (MPA, <12MN, baja profundidad + baja velocidad, etc.) solo para validación, no para entrenamiento.

### 2. Ingeniería de variables (por MMSI, orden temporal)

- Lags (t-1, t-2), deltas (diff1/diff2).
- SMA / EMA (ventanas cortas).
- segment\_distance (Haversine), delta\_time, acceleration, angular\_velocity.
- Ratios (p. ej., distancia\_a\_costa / distancia\_a\_puerto, speed/(depth+1)).
- Agregados por barco (media/desv/std/max de velocidad, distancia segmentaria, aceleración, giro).

### 3. Serialización por ventanas



- a. Orden por mmsi, timestamp y construcción de ventanas deslizantes (p. ej., 10 pasos) → tensores 3D (N, T, F) o vectores aplanados (N, T×F) según el modelo.
  4. Modelos no supervisados
    - a. Baseline tabular/ventana: Isolation Forest, One-Class SVM, LOF/COPOD.
    - b. Densidad/cluster: HDBSCAN + puntuación de outlier.
    - c. (Opcional) LSTM Autoencoder para reconstrucción de secuencias.???
  5. Evaluación
    - a. Entrenar con solo is\_suspicious=0; validar contra eventos con is\_suspicious=1 y cruces espaciales (dentro de GMR).
    - b. Métricas: PR-AUC, ROC-AUC, Precision@k, Recall@k, F1@k, tasa de detección por ventana y por trayectoria, ablation por familias de features.
    - c. Validación por barco (split a nivel MMSI) para evitar fuga temporal o de identidad.
  6. Entrega
    - a. Scripts productivos (procesar\_ais\_gal.py, \*\_features.py, \*\_windowing.py), datasets finales y notebooks de entrenamiento/evaluación, con semilla y especificaciones de hardware.
- 
6. Sumario de Contenidos
    - a. Introducción
      - 1.1 Motivación y problema
      - 1.2 Aportes y alcance
      - 1.3 Descripción de datos y contexto Galápagos
    - b. Estado del Arte
      - 2.1 AIS y pesca
      - 2.2 Detección no supervisada y series de tiempo
      - 2.3 Limitaciones y oportunidades
    - c. Descripción de la Propuesta
      - 3.1 Arquitectura del pipeline
      - 3.2 Ingeniería de variables (geoespacial y temporal)
      - 3.3 Serialización por ventanas
    - d. Desarrollo del Prototipo
      - 4.1 Preparación de datos y escalabilidad (procesamiento por chunks)
      - 4.2 Implementación de detectores:
        - Isolation Forest (baseline rápido, tabular).
        - LOF o COPOD (para anomalías locales).
        - HDBSCAN (detección basada en clusters).
        - LSTM Autoencoder (para dependencias temporales).
        - Matrix Profile (para subsecuencias raras).
    - e. Experimentos y Resultados
      - 5.1 Diseño experimental (split por MMSI, métricas)
      - 5.2 Resultados principales
      - 5.3 Análisis de sensibilidad (features, tamaño de ventana)



## 5.4 Casos de estudio en GMR (mapas)

### f. Conclusiones y Trabajo Futuro

#### 6.1 Hallazgos

#### 6.2 Limitaciones

#### 6.3 Extensiones (detección online, fusión con SAR, reglas regulatorias)

## 7. Recursos

### a. Humanos

Estudiante, tutor, profesor de la clase Proyecto Integrador.

### b. Materiales

MacBook Pro (compu personal), hardware de laboratorio y Lightning studio.

### c. Económicos

No por ahora.

## 8. Cronograma de Actividades

Actividades	2025				
	8	9	10	11	12
<b>A1:</b> Planificación y aprobación	X	X			
<b>A2:</b> Estudio del estado del arte	X	X	X		
A3: Integración y limpieza de datos (GMR/GEBCO/AIS)		X			
A4: Ingeniería de variables y dataset enriquecido		X			
A5: Serialización por ventanas y muestreos		X			
A6: Implementación de modelos		X	X	X	
A7: Validación y análisis			X	X	
A8: Recomendaciones y documentación				X	
A9: Redacción Final				X	X

## 9. Entregables

- Documento de Planificación del Proyecto Integrador
- Muestras del Dataset y validación de limpieza
- Prototipo funcional del/los modelo/s.



- Web -> ngrok y streamlit
- Documento Final

## 10. Referencias

- [1] T. L. Dumpit and R. Monreal, "A Machine Learning Approach on Illegal Fishing Detection Using RNN," La Union, Philippines, 2025.
- [2] R. Motiramani, A. Mody, and S. Sejjal, "Identifying and Combating Unlawful Fishing Activities: A Classification-Based Approach," NMIMS, Mumbai, India, 2023.
- [3] C. Huang, L. Zhu, R. Huang, X. Wan, and Y. Ren, "Research on Real Time Anomaly Detection of Fishing Vessels in a Marine Edge Computing Environment," *Mobile Information Systems*, 2021.
- [4] T. L. Dumpit Jr. and R. Monreal, "A Machine Learning Approach on Illegal Fishing Detection Using RNN for the Area of Bauang, La Union, Philippines," *International Journal of Computing Sciences Research*, 2025.
- [5] B. Brown, K. Katz, A. Korotovskikh, and D. Kullman, "Detecting Illegal Fishing with Automatic Identification Systems and Machine Learning," GA-CCRi Capstone Team, 2024.
- [6] Y. Xu, "Illegal, Unreported and Unregulated Fishing Detection with Machine Learning," in *Proceedings of the 2023 International Conference on Software Engineering and Machine Learning*, 2023.
- [7] P. Sharma and M. Panjkar, "Fishing Forecast Guardian," 2025.

## 11. Revisión y firma del tutor del proyecto

Yo, Felipe Grijalva, profesor de la carrera de Ingeniería en Ciencias de la Computación, hago constar que he revisado y, por lo tanto, apruebo el documento de planificación del proyecto titulado "Detección no supervisada de anomalías, usando series de tiempo, en trayectorias AIS de pesqueros para vigilancia de la Reserva Marina de Galápagos" propuesto por el estudiante Erick Suárez. Por otra parte, me comprometo a proporcionar al estudiante el soporte necesario y oportuno para el buen desarrollo del proyecto antes mencionado.

---

Fdo: Felipe Grijalva

Quito, 6 de septiembre de 2025