



**DETECÇÃO DE ATIVIDADES ILEGAIS DE EMBARCAÇÕES ATRAVÉS DA
FUSÃO DE COMPORTAMENTOS E REGRAS BASEADAS NO
CONHECIMENTO DO ESPECIALISTA**

Vinicius Dalto do Nascimento

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientadores: Diego Leonel Cadette Dutra
Tiago Assumpção de Oliveira Alves
Claudio Miceli de Farias

Rio de Janeiro
Abril de 2025

DETECÇÃO DE ATIVIDADES ILEGAIS DE EMBARCAÇÕES ATRAVÉS DA
FUSÃO DE COMPORTAMENTOS E REGRAS BASEADAS NO
CONHECIMENTO DO ESPECIALISTA

Vinicius Dalto do Nascimento

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Orientadores: Diego Leonel Cadette Dutra
Tiago Assumpção de Oliveira Alves
Claudio Miceli de Farias

Aprovada por: Prof. Diego Leonel Cadette Dutra
Prof. Claudio Miceli de Farias
Prof. Tiago Assumpção de Oliveira Alves
Prof. Karla Tereza Figueiredo Leite
Prof. Jean-David Job Emmanuel Marie Caprace
Prof. Priscila Machado Vieira Lima
Prof. Paulo Cesar Guerreiro da Costa

RIO DE JANEIRO, RJ – BRASIL
ABRIL DE 2025

Nascimento, Vinicius Dalto do

Detecção de Atividades Ilegais de Embarcações Através da Fusão de Comportamentos e Regras baseadas no Conhecimento do Especialista/Vinicius Dalto do Nascimento. – Rio de Janeiro: UFRJ/COPPE, 2025.

XIII, 102 p.: il.; 29, 7cm.

Orientadores: Diego Leonel Cadette Dutra

Tiago Assumpção de Oliveira Alves

Claudio Miceli de Farias

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2025.

Referências Bibliográficas: p. 91 – 100.

1. Monitoramento Marítimo.
 2. Detecção de Atividades Ilegais.
 3. Fusão de Modelos.
 4. Aprendizado Ativo.
- I. Dutra, Diego Leonel Cadette *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

“Toda inovação, todo progresso é o resultado de uma ideia nova, de uma concepção individual. A sociedade se desenvolve por causa do gênio criativo de indivíduos que rompem com a rotina e os caminhos comuns.”

Ludwig von Mises

Agradecimentos

A jornada para a conclusão deste doutorado foi marcada por desafios e, acima de tudo, por momentos de grande aprendizado e crescimento pessoal. Ao olhar para trás, percebo o quanto essa trajetória exigiu resiliência e dedicação para conciliar o trabalho com os estudos, especialmente após 17 anos desde a conclusão da graduação. Este trabalho é, portanto, fruto não apenas de um esforço individual, mas também do apoio e incentivo de pessoas queridas e fundamentais ao meu redor.

Em primeiro lugar, agradeço profundamente à minha mãe, que sempre me apoiou e me proporcionou uma base sólida de educação e valores. Seu apoio incondicional às minhas ideias e sua fé em mim foram fundamentais para que eu tivesse confiança para seguir em frente.

Agradeço, com amor, à minha esposa, pela sua compreensão, dedicação e cumplicidade. A sua paciência e apoio, mesmo nos momentos mais desafiadores, foram essenciais para que eu pudesse avançar em direção aos nossos objetivos comuns. A sua parceria inestimável me proporcionou forças para continuar quando as dificuldades pareciam intransponíveis.

Sou imensamente grato aos meus orientadores pela confiança que depositaram em mim e pela orientação ao longo deste processo. As contribuições deles foram fundamentais para o desenvolvimento deste trabalho, e seu apoio me ajudou a expandir meus limites intelectuais e profissionais. Agradeço pela paciência, pelo incentivo e pelos ensinamentos que levarei para toda a vida.

Por fim, expresso minha gratidão a todos os colegas, amigos e familiares que, de alguma forma, estiveram presentes e colaboraram para que eu pudesse concretizar este sonho. A cada um de vocês, o meu mais sincero “muito obrigado”.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

DETECÇÃO DE ATIVIDADES ILEGAIS DE EMBARCAÇÕES ATRAVÉS DA
FUSÃO DE COMPORTAMENTOS E REGRAS BASEADAS NO
CONHECIMENTO DO ESPECIALISTA

Vinicius Dalto do Nascimento

Abril/2025

Orientadores: Diego Leonel Cadette Dutra
Tiago Assumpção de Oliveira Alves
Claudio Miceli de Farias

Programa: Engenharia de Sistemas e Computação

Este trabalho apresenta um *framework* capaz de detectar atividades ilegais envolvendo embarcações no domínio marítimo, com foco específico na classificação de pesca ilegal, atividades suspeitas, comportamentos anômalos e atividades normais. Ao combinar comportamentos de navegação com regras definidas por especialistas, nossa abordagem integra métodos orientados por dados e conhecimento especializado. O *framework* utiliza um metamodelo baseado em *random forest* para integrar as saídas dos modelos de detecção de comportamento e das regras dos especialistas. Dada a falta de dados rotulados sobre atividades ilegais, implementamos aprendizado ativo, permitindo que especialistas rotulassem trajetórias de maior impacto para otimizar o treinamento do modelo. Treinado com conjuntos de dados sintéticos do GFW e MarineCadastre e testado em dados AIS (*Automatic Identification System*) da Zona Econômica Exclusiva (ZEE) do Brasil, o *framework* alcançou 100% de acurácia para atividades normais, 97% para pesca ilegal e 90% para atividades suspeitas. Esses resultados indicam que o *framework* pode reduzir significativamente a carga de trabalho dos operadores humanos, filtrando trajetórias normais e permitindo o foco em casos críticos. Trabalhos futuros incluem o aumento da escalabilidade pela incorporação de comportamentos, regras e conjuntos de dados adicionais, contribuindo assim para o campo de vigilância marítima com uma solução robusta e escalável.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

DETECTION OF ILLEGAL VESSEL ACTIVITIES THROUGH THE FUSION OF BEHAVIORS AND EXPERT KNOWLEDGE-BASED RULES

Vinicius Dalto do Nascimento

April/2025

Advisors: Diego Leonel Cadette Dutra
Tiago Assumpção de Oliveira Alves
Claudio Miceli de Farias

Department: Systems Engineering and Computer Science

This work presents a framework designed to detect illegal activities involving vessels in the maritime domain, specifically targeting illegal fishing, suspicious activities, anomalous behavior, and normal activity classification. By combining vessel navigation behaviors and expert-defined rules, our approach leverages both data-driven methods and domain-specific knowledge. The framework employs a meta-model based on a random forest ensemble to integrate outputs from behavior detection models and expert rules. Given the lack of labeled data on illegal maritime activities, we implemented active learning, allowing experts to label high-impact trajectories for optimal model training. The framework was trained on synthetic datasets generated from GFW and MarineCadastre data and subsequently tested on AIS (Automatic Identification System) data from Brazil's Exclusive Economic Zone (EEZ). Results demonstrate its effectiveness, achieving 100% accuracy for normal activities, 97% for illegal fishing detection, and 90% for suspicious activities. These outcomes indicate that the framework can significantly reduce the workload of human operators by filtering normal trajectories and enabling a focus on critical cases. Future work includes enhancing scalability by incorporating additional behaviors, rules, and datasets to improve detection and adaptability for other illicit maritime activities, thus contributing to the maritime surveillance field with a robust, scalable solution.

Sumário

Lista de Figuras	x
Lista de Tabelas	xiii
1 Introdução	1
1.1 Abordagem Proposta	3
1.2 Resumos dos Resultados e das Contribuições	5
1.3 Organização do Texto	7
2 Revisão Bibliográfica	9
2.1 Detecção de Comportamentos de Embarcações	9
2.2 Detecção de Comportamentos com base no Conhecimento do Especialista	18
2.3 Detecção de Atividades de Embarcações	20
3 Revisão de Conceitos	23
3.1 Conceitos sobre o Domínio Marítimo	23
3.1.1 Zona Econômica Exclusiva (ZEE)	23
3.1.2 Mar Territorial (MT)	24
3.1.3 Conscientização do Domínio Marítimo (MDA)	25
3.1.4 Monitoramento Marítimo	25
3.1.5 Competências das Autoridades Marítimas	27
3.2 Conceitos Computacionais	28
3.2.1 Ensemble Models	29
3.2.2 Aprendizado Ativo	30
3.2.3 Modelo JDL	32
4 Construção do Framework para Classificação de Atividades Ilegais de Embarcações	34
4.1 Extração das Trajetórias	34
4.2 Detectando Comportamentos de Embarcações	35
4.2.1 AIS Spoofing	36

4.2.2	Dark Ships	37
4.2.3	Encontros entre Embarcações	40
4.2.4	Trajetórias de Pesca	43
4.2.5	Loitering	46
4.2.6	Navegação Anômala	50
4.3	Definindo Regras com Base no Conhecimento do Especialista	50
4.4	Definindo as Atividades Ilegais de Embarcações	53
4.5	Implementando Ensemble Models e a Estratégia de Stacking	55
4.6	Rotulando de Forma Eficiente Utilizando Aprendizado Ativo	57
4.7	Definindo o Framework com Base no JDL	58
4.7.1	Pré processamento	60
4.7.2	Fusão à Nível de Objetos	62
4.7.3	Consciência Situacional	62
4.7.4	Avaliação de Impacto	63
4.7.5	Apoio a Decisão	64
5	Avaliação Experimental	67
5.1	Caso de Estudo: A Zona Econômica Exclusiva do Brasil	67
5.2	Dados Experimentais	68
5.3	Definindo Critérios de Avaliação de Trajetórias	69
5.4	Treinamento do Framework	70
5.5	Teste do Framework	72
5.6	Resultados	74
5.7	Discussão	85
6	Conclusões e Trabalhos Futuros	89
	Referências Bibliográficas	91
A	Abreviações	101

Listas de Figuras

1.1	Processo de detecção de atividade de embarcações	5
4.1	Exemplo de trajetória de embarcação. Cada ponto azul representa uma mensagem AIS recebida da embarcação.	35
4.2	No mapa podemos ver um exemplo de célula geohash. O quadrado azul é uma célula de resolução 5 (4,9 Km) e no seu interior a codificação gerada para a região.	38
4.3	A imagem mostra um exemplo de detecção de lacunas na trajetória de uma embarcação. Os pontos azuis refletem a trajetória da embarcação. Os quadrados vermelhos indicam onde a embarcação deveria ter transmitido o sinal AIS, enquanto os quadrados verdes representam onde a embarcação transmitiu. Os quadrados vermelhos representam lacunas de transmissão.	39
4.4	A imagem mostra um exemplo de encontro entre embarcações. Os pontos verdes representam a localização da embarcação 1, enquanto que os pontos vermelhos representam a embarcação 2. Os quadrados azuis apresentam áreas de exploração de petróleo (FPSO). A imagem mostra que a embarcação realizou um encontro dentro de uma área de FPSO.	41
4.5	Exemplo de célula geohash de resolução 5 (8,54km de aresta) utilizando a biblioteca h3.	42
4.6	Exemplo de trajetórias de embarcações. A imagem a esquerda mostra a trajetória de uma embarcação de pesca, enquanto que a imagem a direita mostra uma trajetória de um navio de carga.	44
4.7	A imagem mostra as trajetórias de embarcações após aplicação do algoritmo do polígono mínimo. Em vermelho temos o polígono construído a partir dos pontos da trajetória em azul. A área do polígono é utilizada como uma das características das trajetórias no modelo de detecção.	45
4.8	Processo de treinamento do metamodelo do framework de detecção de atividades ilegais de embarcações utilizando aprendizado ativo. . .	59

4.9	Framework para classificação de atividades ilegais de embarcações baseado no JDL.	61
4.10	Especialistas avaliam as trajetórias por meio de uma interface no nível de avaliação de impacto. O usuário pode ter uma consciência situacional de uma trajetória com informações sobre a embarcação, comportamentos de navegação apresentados pela trajetória e se a trajetória aciona as regras criadas pelo especialista.	64
4.11	Relatório gerado pelo nosso <i>framework</i> apresentando possíveis casos de pesca ilegal na costa do Brasil. Os pontos amarelos representam embarcações de pesca possivelmente envolvidas em atividades de pesca em locais onde a pesca não é permitida.	65
5.1	A imagem representa o processo de treinamento do <i>framework</i> . No passo 1, o <i>cold start</i> é realizado utilizando consultas SQL seguindo os critérios pré definidos. No passo 2, o algoritmo de aprendizado ativo é executado e as respostas são respondidas pelo bot de critérios. No passo 3, o especialista avalia mais 200 trajetórias de forma a capturar outras características não detectadas pelos critérios.	71
5.2	A imagem apresenta o processo de teste do <i>framework</i> utilizando dados AIS da costa brasileira. No passo 1, todas as trajetórias sem rótulos são inferidas pelo metamodelo (<i>random forest</i>) treinado. No passo 2, são selecionadas 100 trajetórias de cada classe inferida, apresentadas ao especialista para classificação e as métricas são calculadas com base nas respostas do especialista e a predição dada pelo metamodelo.	73
5.3	Matriz de confusão das previsões do metamodelo.	77
5.4	A figura mostra os gráficos SHAP para as previsões do metamodelo. O SHAP exibe as contribuições de cada dimensão para a predição da classe.	78
5.5	Todas as trajetórias inferidas pelo framework representadas no mapa.	81
5.6	A imagem apresenta um exemplo real de uma atividade anômala identificada pelo framework. Uma embarcação como MMSI inválido, informando que é uma embarcação de pesca, apresentando trajetória de pesca e bandeira desconhecida.	82
5.7	A imagem mostra uma atividade suspeita detectada pelo framework. A embarcação representada pelos pontos verdes está encontrando-se com a embarcação representada pelos pontos vermelhos a 61 milhas náuticas da costa. Os quadrados azuis representam áreas de FPSO.	83

5.8 A figura mostra uma possível situação de pesca ilegal. O framework detectou uma embarcação de pesca de bandeira estrangeira apresentando trajetória de pesca dentro da ZEE brasileira na região norte da costa.	84
5.9 A figura mostra uma trajetória suspeita detectada pelo framework. Uma embarcação de pesca de bandeira estrangeira está fundeada há mais de 3 horas dentro da ZEE brasileira na região norte da costa. . .	85

Listas de Tabelas

2.1	Tabela com os trabalhos relacionados com a detecção de comportamentos que utilizam o sensor AIS. Os comportamentos estudados foram: o AIS <i>spoofing</i> , <i>dark ships</i> , encontros, trajetórias de pesca e <i>loitering</i>	18
2.2	Tabela com os trabalhos relacionados que utilizaram regras com base no conhecimento do especialista para a detecção de comportamentos de embarcações.	20
2.3	Tabela com os trabalhos relacionados que tentaram detectar atividades de embarcações.	22
4.1	Tabela com os cálculos dos dados agregados das trajetórias para utilização como dimensão no modelo de detecção de trajetória de pesca.	45
4.2	Tabela com as regras criadas pelo especialista com os respectivos atributos, descrições e regra de disparo.	53
5.3	Métricas de desempenho do metamodelo: precisão, recall e F1-score para cada classe de atividade.	75
5.1	Critérios para a classificação das atividades das embarcações.	87
5.2	Quantitativo de dados e estratégia para rotulação no treinamento.	88
5.4	Números de trajetórias por predição de atividade.	88
A.1	Siglas e Definições	101

Capítulo 1

Introdução

O tráfego marítimo desempenha um papel crucial na conectividade global e no comércio internacional, funcionando como uma artéria vital para a economia mundial. Milhões de toneladas de mercadorias são transportadas pelos oceanos todos os anos, impulsionando o crescimento econômico e o desenvolvimento nos países costeiros [1]. A importância estratégica do transporte marítimo é inegável, pois é responsável por mais de 90% do comércio global [2]. Além disso, as áreas costeiras de alguns países são fontes essenciais de emprego e receita, sustentando comunidades litorâneas e promovendo o desenvolvimento socioeconômico.

No entanto, a manutenção dessa complexa rede de atividades marítimas apresenta desafios significativos, como a necessidade de garantir a segurança das rotas marítimas, proteger o meio ambiente marinho e prevenir atividades ilícitas [3]. Assim, a vigilância eficaz do domínio marítimo pelas Autoridades Marítimas (AM) locais torna-se essencial para assegurar a eficiência, segurança e sustentabilidade das operações marítimas em todo o mundo.

Por meio da Consciência do Domínio Marítimo, do inglês *Maritime Domain Awareness* (MDA), as AMs podem monitorar suas Zonas Econômicas Exclusivas (ZEEs) para garantir um tráfego marítimo seguro e a exploração econômica da costa pelo país [4]. Para melhorar a segurança do tráfego marítimo, é essencial fortalecer os processos de inteligência na vigilância marítima. Assim, detectar atividades ilícitas realizadas por embarcações é crucial para aumentar a segurança da navegação, além de melhorar a eficiência e a proteção das operações marítimas [3]. A detecção de comportamentos de navegação, como trajetórias de pesca, permanência prolongada (*loitering*), falsificação de sinal (*spoofing*), encontros (*rendezvous*) e embarcações com sinais desligados (*dark ships*), é fundamental para a consciência situacional e, consequentemente, para a detecção de atividades ilícitas [5].

Vários trabalhos na literatura abordaram a detecção de comportamentos individuais de embarcações [6–9], com ênfase especial em anomalias. Em [10], discutiu-se a relação entre esses comportamentos de navegação e as atividades realizadas pelas

embarcações, além de como é possível identificar atividades por meio dessa relação. Os comportamentos das embarcações podem ser observados por meio de sensores, enquanto as atividades realizadas refletem a intenção humana que orienta a utilização das embarcações. Em uma revisão da literatura sobre a detecção de comportamentos anômalos de embarcações realizada em [11], concluiu-se que o conhecimento especializado é um auxílio importante na compreensão do comportamento de navegação das embarcações, embora os especialistas com conhecimento na área marítima sejam diversos, dificultando a formação de um padrão unificado. Portanto, estabelecer um método orientado por dados que incorpore o conhecimento especializado representa uma direção relevante para a pesquisa.

Uma parte significativa da literatura sobre vigilância marítima concentra-se, principalmente, na identificação de anomalias, como evidenciado pelos trabalhos de [5, 6, 11, 12]. Na literatura, observa-se uma escassez de pesquisas voltadas à identificação e detecção de atividades ilícitas no domínio marítimo. Como ilustração, [10] discutiu a classificação de atividades, focando exclusivamente na perspectiva ontológica. Em [13], os autores introduziram um método para inferir atividades marítimas ilícitas em tempo real, examinando o comportamento espacial em múltiplas escalas de frotas inteiras, especificamente de grupos de embarcações que transmitem sua localização por meio do sistema AIS. Contudo, em nossa investigação, não encontramos nenhum trabalho que tentasse detectar atividades ilícitas de embarcações utilizando uma metodologia semelhante à nossa. Assim, o objetivo deste trabalho é preencher essa lacuna, apresentando um *framework* capaz de detectar atividades ilícitas realizadas por embarcações por meio da fusão de modelos de detecção de comportamento de embarcações com regras elaboradas a partir de conhecimento especializado.

Muitos sistemas de monitoramento do domínio marítimo utilizados por AMs são fundamentados em regras especializadas elaboradas por especialistas. Essas regras podem ser baseadas em legislação, experiências prévias ou ameaças identificadas nesse domínio. As soluções baseadas em regras têm como objetivo detectar um conjunto restrito de atividades das embarcações, áreas, características dinâmicas ou evoluções das trajetórias das embarcações. O Monitoramento Automatizado de Comportamentos (ABM), desenvolvido pela Agência Europeia de Segurança Marítima (EMSA) [14], aprimora a vigilância marítima ao incorporar uma coleção de cenários predefinidos com base em regras. Esses cenários permitem a ativação automática da detecção para embarcações que se aproximam de áreas específicas, seguem rotas não convencionais ou estão localizadas em zonas designadas. No entanto, um problema significativo relacionado a esses sistemas é a criação de milhares de regras e o consequente acionamento de um grande número de alertas falsos, o que pode sobrecarregar os operadores. Nesse contexto, nosso framework busca aprimorar o

processo de detecção de atividades, reduzindo a necessidade de intervenções por parte dos especialistas e garantindo que eles atuem apenas nas trajetórias das embarcações que realmente requerem atenção. Portanto, nosso framework tem como objetivo alcançar os seguintes principais objetivos:

- Detectar atividades ilícitas no domínio marítimo envolvendo o uso de embarcações;
- Reduzir o número de trajetórias que os especialistas precisam analisar;
- Integrar o conhecimento tácito dos especialistas com métodos orientados por dados;
- Investigar o comportamento da técnica de *ensemble models* para o problema;
- Garantir a escalabilidade do framework para detectar outras atividades, comportamentos e regras.

No entanto, para alcançarmos esses objetivos, contaremos com as seguintes premissas:

- Inicialmente não temos rótulos para atividades ilegais de embarcações;
- Não existem datasets com dados de atividades ilegais de embarcações disponíveis;
- A avaliação manual de algumas trajetórias por um operador humano poderá ser realizada;
- A avaliação manual de uma trajetória é custosa;

1.1 Abordagem Proposta

Para atingir os objetivos previamente delineados, a partir de nossa premissa, e demonstrar a eficácia de nosso framework na detecção de atividades ilícitas no domínio marítimo, utilizaremos uma combinação de métodos dentro de um framework JDL. O framework JDL (*Joint Directors of Laboratories*) é um modelo amplamente reconhecido para o processo de fusão de dados, desenvolvido para lidar com a integração de informações provenientes de diversas fontes, sensores ou sistemas. Ele organiza o processo de fusão de dados em níveis hierárquicos, começando pela aquisição e processamento de dados brutos, passando pela correlação e associação desses dados, até a construção de uma percepção situacional abrangente e a projeção de possíveis evoluções futuras. O objetivo principal do framework JDL é aprimorar a consciência

situacional e a tomada de decisões em tempo real, ao consolidar dados de diferentes fontes em uma visão unificada e coerente. Esse modelo revela-se particularmente útil em áreas como vigilância marítima e segurança, onde a combinação de dados de sensores e o conhecimento especializado são fundamentais para a detecção de comportamentos e atividades anômalas.

Dentro do *framework*, utilizaremos um modelo *ensemble* composto por modelos de primeiro nível que serão responsáveis por detectar os comportamentos de navegação e as regras definidas com base no conhecimento de especialistas. O segundo nível, denominado metamodelo, será incumbido de utilizar os dados provenientes das saídas do primeiro nível para prever as atividades ilegais. Como uma das nossas premissas é que, inicialmente, não dispomos de rótulos para nossa atividade, adotaremos o aprendizado ativo (*active learning*) para rotular de maneira eficiente as trajetórias e treinar o metamodelo. Os modelos de comportamento serão treinados com dados do GFW [15] e do marine cadastre [16]. Por sua vez, o metamodelo será treinado com dados sintéticos gerados a partir de dados do marine cadastre, GFW e SISTRAM (Sistema de Informações sobre o Tráfego Marítimo) [17]. Após o treinamento do metamodelo, o framework será testado com dados AIS do SISTRAM, que abrangem o tráfego marítimo em toda a costa brasileira no período de janeiro de 2019 a dezembro de 2020.

O teste do *framework* consistirá na utilização de dados reais do SISTRAM para detectar indícios de atividades ilegais de embarcações. As atividades que serão identificadas neste trabalho incluem: atividade normal, atividade anômala, atividade suspeita e pesca ilegal. A seleção dessas atividades foi fundamentada nos dados dos sensores disponíveis e nas abordagens apresentadas na literatura sobre a detecção de comportamentos de embarcações. Para cada classe de atividade, selecionaremos uma amostra e a apresentaremos a um especialista humano para rotulagem. Com base nessa avaliação, calcularemos o desempenho, analisaremos o comportamento do metamodelo e apresentaremos os casos relevantes identificados pelo framework.

Para oferecer uma visão geral do processo de detecção, a Figura 1.1 apresenta um resumo de como nosso *framework* para a detecção de atividades ilícitas opera. Primeiramente, as trajetórias são geradas a partir dos dados do Sistema de Identificação Automática (AIS) recebidos de diversas fontes. Subsequentemente, essas trajetórias são aplicadas a modelos de comportamento, assim como a regras definidas com base no conhecimento especializado. Os resultados são então combinados em um modelo de ensemble em pilha (*stacking*), e o aprendizado ativo é utilizado para selecionar as trajetórias que apresentam maior incerteza. Finalmente, o especialista pode avaliar a trajetória e classificá-la, permitindo a atualização do modelo conforme necessário. Nas seções seguintes, forneceremos mais detalhes sobre cada parte do processo e o modo como o treinamento do modelo é conduzido.

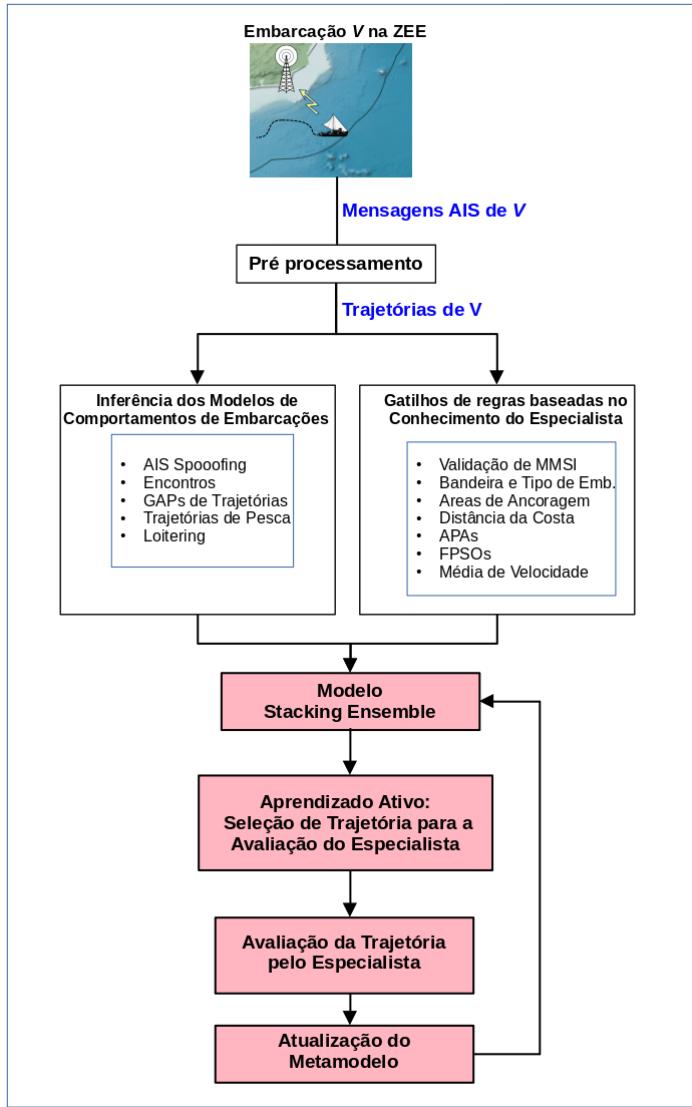


Figura 1.1: Processo de detecção de atividade de embarcações

1.2 Resumos dos Resultados e das Contribuições

Os resultados obtidos nesta pesquisa foram decorrentes de uma avaliação manual de cada uma das 400 trajetórias inferidas pelo *framework*. Selecionamos aleatoriamente 100 trajetórias para cada classe de atividade inferida. Nos testes realizados, destacamos o desempenho superior na detecção de atividades normais, que alcançou 100% de acurácia, seguido pela detecção de pesca ilegal, com 97% de acurácia. O menor desempenho foi observado na detecção de atividades suspeitas, que apresentou 90% de acurácia, seguido pela detecção de atividades anômalas, com 95% de acurácia.

Durante a fase de testes, nosso *framework*, treinado com conjuntos de dados sintéticos, identificou com sucesso casos reais de embarcações envolvidas em encontros suspeitos, qualificando essas atividades como suspeitas. Além disso, foram identificadas embarcações estrangeiras aparentando realizar pesca dentro da Zona

Econômica Exclusiva (ZEE) do Brasil, caracterizando assim atividades de pesca ilegal. Esses resultados demonstram a viabilidade de nosso *framework* para detecção de atividades ilegais de embarcações.

Dessa maneira, nosso trabalho alcançou as seguintes contribuições principais:

- Apresentamos um *framework* híbrido capaz de detectar atividades ilegais de embarcações no domínio marítimo, integrando aprendizado de máquina e regras baseadas no conhecimento de especialistas;
- Destacamos a contribuição original deste trabalho ao propor a fusão explícita entre conhecimento tácito de especialistas e modelos baseados em dados, preenchendo uma lacuna presente na literatura, que geralmente trata essas abordagens de forma isolada;
- Os resultados indicam que o *framework* pode reduzir significativamente a carga de trabalho dos operadores humanos em até 98%, ao filtrar trajetórias normais e permitir o foco em atividades mais severas como a pesca ilegal e atividade suspeita;
- Mostramos que a lacuna apresentada na literatura, referente a soluções que integram métodos baseados em dados e no conhecimento do especialista, pode ser resolvida através da fusão de modelos utilizando a técnica de *stacking ensemble*;
- Demonstramos que os erros propagados pelos modelos de primeiro nível no *stacking ensemble* podem influenciar negativamente o resultado do metamodelo. Dessa forma, utilizar somente os critérios de avaliação pode não ser uma boa estratégia. Por outro lado, a utilização combinada dos critérios com a avaliação do especialista trouxe bons resultados, sendo os erros dos modelos de primeiro nível compensados pelos outros modelos e dimensões desse nível;
- Mostramos que o problema da falta de rótulos para as atividades pode ser solucionado através do aprendizado ativo com a participação de um especialista;
- Atestamos que, mesmo com o treinamento do metamodelo utilizando um conjunto de dados sintético, este foi capaz de identificar possíveis atividades ilegais no conjunto de dados real, demonstrando a viabilidade do uso de dados sintéticos no treinamento de modelos aplicados à vigilância marítima;
- Ressaltamos que, embora algumas marinhas ao redor do mundo possam ter desenvolvido internamente soluções semelhantes, estas não se encontram publicamente indexadas ou descritas na literatura científica, o que reforça a ori-

ginalidade e relevância deste trabalho para a comunidade acadêmica e para instituições de segurança marítima.

Além disso, durante a pesquisa tivemos também alguns trabalhos publicados no campo da vigilância marítima. Inicialmente, foi conduzido um estudo comparativo entre diferentes abordagens para a detecção de atividades de pesca, destacando os pontos fortes e limitações de cada metodologia na tarefa de identificar padrões relevantes em dados AIS [18]. Em [19], propomos uma abordagem de fusão multimodal para melhorar o monitoramento de atividades de embarcações, integrando múltiplas fontes de dados e evidências para aumentar a acurácia da inferência. A pesquisa também introduziu modelos baseados em *ensemble learning*, cuja eficácia foi avaliada em diferentes cenários, demonstrando ganhos de desempenho na detecção de comportamentos de pesca [20]. Consolidando esses avanços, foi desenvolvido um *framework* híbrido que combina modelos baseados em aprendizado de máquina com regras baseadas em conhecimento especializado, capaz de detectar atividades ilegais com maior eficiência e escalabilidade [21]. Por fim, a construção e análise de um novo *dataset* focado em encontros de embarcações em alto-mar complementa as contribuições, fornecendo uma base valiosa para estudos futuros e reforçando o papel do trabalho no aprimoramento da segurança e consciência situacional no domínio marítimo [22].

1.3 Organização do Texto

Este trabalho será organizado em seis capítulos, nos quais cada capítulo pode ser resumido a seguir:

- Revisão Bibliográfica: Apresenta os trabalhos relacionados ao tema e as principais abordagens da literatura para a detecção de comportamentos de navegação e atividades de embarcações;
- Revisão de Conceitos: Discute os conceitos fundamentais sobre o domínio marítimo, assim como os conceitos computacionais essenciais para a definição de nossa abordagem;
- Construção do *Framework* para Detecção de Atividades Ilegrais de Embarcações: Detalhamento metodológico da construção do *framework*, incluindo a implementação da extração de trajetórias, detecção de comportamentos, definição de regras baseadas no conhecimento de especialistas, classificação das atividades, modelos de ensemble, aprendizado ativo e a aplicação do modelo JDL;

- Avaliação Experimental: Apresenta um estudo de caso da costa brasileira, detalhando os dados utilizados nos testes, os critérios de avaliação, bem como os parâmetros adotados para o treinamento e teste do *framework*. Além disso, os resultados obtidos são analisados, com destaque para os casos relevantes detectados pelo *framework*;
- Conclusões e Trabalhos Futuros: Nesta seção, sintetizamos as contribuições do trabalho, destacamos suas implicações no campo da vigilância marítima, e sugerimos direções para pesquisas futuras, como a incorporação de novos comportamentos e atividades ilegais, além de melhorias no desempenho e escalabilidade do *framework*.

Capítulo 2

Revisão Bibliográfica

A detecção das atividades de embarcações, um campo que interliga segurança marítima, inteligência artificial e tecnologias de sensoriamento remoto, tem experimentado avanços significativos nas últimas décadas. O escopo deste trabalho abrange metodologias, algoritmos e tecnologias empregados na identificação das atividades das embarcações. Portanto, para chegarmos à detecção das atividades de uma embarcação, é imprescindível estudar a ontologia subjacente à atividade da embarcação, os comportamentos de navegação, os sensores remotos e as técnicas de aprendizado de máquina. Assim, neste capítulo, apresentaremos, organizados em seções, os trabalhos relacionados à detecção de comportamentos das embarcações e à detecção das atividades de embarcações.

2.1 Detecção de Comportamentos de Embarcações

Os comportamentos das embarcações podem ser observados por meio de sensores, como AIS, radares, câmeras e imagens de satélites. A partir dos dados coletados por esses sensores, modelos de aprendizado de máquina podem ser desenvolvidos e treinados para detectar comportamentos específicos, como padrões de trajetórias de pesca, navegação anômala em determinadas regiões e encontros entre embarcações. Durante a pesquisa, realizamos uma revisão bibliográfica sobre os principais comportamentos identificados na literatura, assim como os sensores mais utilizados. Diversas abordagens foram apresentadas na literatura, empregando diferentes sensores ou suas combinações. Entretanto, considerando que o AIS é o sensor mais utilizado e possui o maior conjunto de dados disponível, por questões de escopo, limitaremos nossa análise aos trabalhos que empregaram o sensor AIS. A relação dos trabalhos pode ser observada na tabela 4.2.

Trabalho	Comportamento	Abordagem	Ano
AIS Spoofing: A Tutorial for Researchers [23]	AIS Spoofing	Definição e conceituação sobre AIS Spoofing, principais métodos de spoofing e possível soluções para detecção.	2024
Countering Real-Time Stream Poisoning: An Architecture for Detecting Vessel Spoofing in Streams of AIS Data [24]	AIS Spoofing	Detecção do spoofing através do cálculo do caminho mais curto e do cálculo da velocidade. Caso a velocidade calculada seja considerada impossível para a embarcação, o caso é considerado um spoofing.	2018
Securing Offshore Installations Against Automatic Identification System Spoofing [25]	AIS Spoofing	Detecção do Spoofing através da potência do RSSI da mensagem utilizando a técnica de Krigagem para interpolar dados.	2023
Detection of false AIS messages for the improvement of maritime situational awareness [26]	AIS Spoofing	Detecção pela avaliação de qualidade das mensagens AIS com o objetivo de identificar aquelas que são duvidosas.	2016
Malicious AIS Spoofing and Abnormal Stealth Deviations: A Comprehensive Statistical Framework for Maritime Anomaly Detection [27]	AIS Spoofing	Utiliza a derivação de regras de detecção de anomalias com base no teste da razão de verossimilhança generalizada e nas metodologias de seleção de ordem de modelos.	2021

Trabalho	Comportamento	Abordagem	Ano
Detection of AIS Spoofing in Fishery Scenarios [28]	AIS Spoofing	Deteta se uma embarcação apresenta comportamento típico de pesca por meio de modelos de aprendizado de máquina e, em seguida, verifica-se se o tipo da embarcação informado no AIS corresponde à categoria de pesca. Caso haja divergência entre o comportamento identificado e a informação reportada, o caso é classificado como spoofing.	2019
A novel anomaly detection approach to identify intentional AIS on-off switching [29]	Dark Ship	Explora o Indicador de Intensidade de Sinal Recebido (Received Signal Strength Indicator - RSSI) disponível nas Estações Base (BS) de AIS para indentificar o seu desligamento.	2017
A novel technique to identify AIS transmissions from vessels which attempt to obscure their position by switching their AIS transponder from normal transmit power mode to low transmit power mode [30]	Dark Ship	Analisa o RSSI recebido em combinação com valores de RSSI previstos por meio de modelagem matemática e apresenta um algoritmo de detecção único.	2022

Trabalho	Comportamento	Abordagem	Ano
Real-time Maritime Anomaly Detection: Detecting intentional AIS switch-off [31]	Dark Ship	O trabalho foi fundamentado em uma área de monitoramento fixa, dividida em células. Um limite de trinta minutos foi estabelecido a partir da última transmissão de cada embarcação. Se nenhuma transmissão for recebida nesse intervalo, a arquitetura tenta prever em qual célula a embarcação pode estar e se a célula apresenta algum problema de cobertura de sinal. Se a célula possui um histórico de envio de sinal AIS, o algoritmo marca o último ponto de envio de sinal AIS da embarcação como um desligamento intencional do AIS.	2020
Dark Activity Detection in AIS-Based Maritime Networks [32]	Dark Ship	Abordagem foca em detectar a atividade <i>dark ship</i> de uma embarcação por embarcações próximas, com base nos sinais anteriores. A embarcação de detecção coleta as mensagens AIS no seu entorno e calcula as posições esperadas dessas embarcações no próximo passo de tempo. As posições esperadas foram calculadas utilizando abordagens baseadas em regras e também baseadas em dados.	2023

Trabalho	Comportamento	Abordagem	Ano
Dark-VADER: Detection of Anomalous AIS Message Delays for Maritime Situational Awareness [33]	Dark Ship	Aborda o problema da detecção de eventos de atraso anômalo em transmissões de mensagens AIS. Para esse fim, foi proposto um algoritmo que modela o processo de recepção de mensagens AIS como um processo de Bernoulli e tenta detectar eventos de desligamento de AIS por meio de um teste de hipótese estatístico.	2024
Prediction of Rendezvous in Maritime Situational Awareness [34]	Encontros	Tenta prever algorítmicamente encontros entre embarcações com base em suas previsões de trajetória em um ambiente marítimo. O problema é tratado como um teste de hipótese sobre o valor esperado da distância entre as trajetórias.	2018
Analyzing Trajectory Gaps for Possible Rendezvous: A Summary of Results [35]	Encontros	Dado um conjunto de dados de trajetórias com lacunas de transmissão, essas trajetórias são analisadas em busca de possíveis regiões de encontro. É utilizado inicialmente o algoritmo de varredura planar como linha de base. Um framework de filtragem é proposto utilizando o conceito de <i>space-time grids</i> para detectar possíveis regiões de encontros.	2022

Trabalho	Comportamento	Abordagem	Ano
A semi-supervised deep learning model for ship encounter situation classification [36]	Encontros	Propõe um método chamado Rede de Codificador-Decodificador Convolucional Semi-Supervisionada (SCEDN) para a classificação de situações de encontro de embarcações com base em dados AIS.	2021
Fighting Shadows and Ghosts – A Conceptual Framework for Identification Deception Tactics (IDTs) and MARINE Approach [37]	Encontros	Utiliza uma metodologia chamada MARINE (<i>Methodology for Analyzing Russian oil Naval Exchanges</i>) para detectar operações <i>ship to ship</i> de óleo de países sancionados. Esta metodologia é uma elaboração específica de uma lista semelhante de melhores práticas já desenvolvida pelo OFAC em 2019.	2024
A Deep Learning Streaming Methodology for Trajectory Classification [38]	Trajetórias de Pesca	Apresenta uma abordagem que transforma trajetórias de embarcações em imagens e utiliza algoritmos de aprendizado profundo (<i>Convolutional Neural Networks</i>) para classificar as atividades das embarcações.	2021

Trabalho	Comportamento	Abordagem	Ano
Fishing Vessels Activity Detection from Longitudinal AIS Data [39]	Trajetórias de Pesca	Propõe um modelo de rede neural convolucional (CNN) para a classificação de trajetórias de embarcações. O modelo é treinado utilizando um conjunto de sequências de características espaço-temporais invariantes extraídas dos comportamentos de movimento das embarcações.	2020
A Semi-Supervised Methodology for Fishing Activity Detection Using the Geometry behind the Trajectory of Multiple Vessels [40]	Trajetórias de Pesca	Fez uso dos dados brutos do AIS para criar um método semi-supervisionado de rotulagem das trajetórias das embarcações e uma arquitetura de Rede Neural Recorrente (RNN) para classificar as trajetórias em pesca e navegação.	2022
Fishing Vessels Behavior Identification for Combating IUU Fishing: Enable Traceability at Sea [41]	Trajetórias de Pesca	Utilizou o processo de KDD (<i>Knowledge Discovery in Databases</i>) para classificar as trajetórias das embarcações em pesca, não pesca e transbordo.	2020
Architecture for Trajectory-Based Fishing Ship Classification with AIS Data [42]	Trajetórias de Pesca	Realizou o pré-processamento dos dados, transformando os dados AIS em trajetórias e, em seguida, utilizando os dados baseados em trajetórias (dados agregados) como entrada nos modelos.	2020

Trabalho	Comportamento	Abordagem	Ano
Research on Ship Classification Based on Trajectory Features [43]	Trajetórias de Pesca	Utiliza os dados baseados em trajetórias em uma regressão logística (LR) para classificar as trajetórias das embarcações.	2018
Loitering Behaviour Detection of Boats at Sea [44]	Loitering	Utiliza os dados agregados das trajetórias em uma regressão logística (LR) para classificar as trajetórias das embarcações.	2017
Loitering Behavior Detection and Classification of Vessel Movements Based on Trajectory Shape and Convolutional Neural Networks [45]	Loitering	Investigou a detecção e classificação do comportamento de loitering de embarcações, utilizando a forma das trajetórias e redes neurais convolucionais.	2022
Loitering Behavior detection by Spatiotemporal Characteristics Quantification Based on the Dynamic Features of Automatic Identification System (AIS) Messages [46]	Loitering	A partir de parâmetros obtidos das trajetórias, como a área da trajetória, discrepância entre COG e a direção da proa do navio (<i>heading</i>) e a distância entre os pontos de início e fim da trajetória, o algoritmo de <i>Isolation Forest</i> foi utilizado para detectar o comportamento de loitering.	2023

Trabalho	Comportamento	Abordagem	Ano
AIS-based kinematic anomaly classification for maritime surveillance [47]	Loitering	O trabalho utiliza a classificação de anomalias cinemáticas das embarcações em três categorias: Anomalia de Velocidade e Rumo (SCA), Anomalia de Curva (TA) e Anomalia de Loitering (LA), juntamente com métodos de detecção adaptados a cada tipo de anomalia para facilitar a produção de rótulos de anomalias de alta qualidade. Após a obtenção dos rótulos, o treinamento supervisionado é utilizado para a classificação das anomalias sendo utilizados modelos como LSTM, GRU, MLP e FCN.	2024

Trabalho	Comportamento	Abordagem	Ano
Vessel trajectory classification via transfer learning with Deep Convolutional Neural Networks [48]	Loitering	Foi proposto um framework chamado Dense121-VMC, utilizando modelos de DCNN pré-treinados para a detecção e classificação de trajetórias de embarcações. Os padrões de navegação e loitering são extraídos de pontos AIS com base nas distribuições de ângulos e na razão de mudança de curso. Em seguida, esses padrões de trajetória são convertidos em imagens de 120×120 pixels. O framework Dense121-VMC demonstrou superioridade em relação às abordagens anteriores baseadas em CNN para a classificação de trajetórias de navios.	2024

Tabela 2.1: Tabela com os trabalhos relacionados com a detecção de comportamentos que utilizam o sensor AIS. Os comportamentos estudados foram: o AIS *spoofing*, *dark ships*, encontros, trajetórias de pesca e *loitering*.

2.2 Detecção de Comportamentos com base no Conhecimento do Especialista

Em diversos sistemas de monitoramento do domínio marítimo, é comum a utilização de conjuntos de regras baseadas em conhecimento, desenvolvidas por especialistas na área, com o intuito de detectar comportamentos ou atividades de embarcações. Essas regras são formuladas a partir de anos de experiência e de um profundo entendimento dos padrões de navegação, dos comportamentos esperados das embarcações, dos indicadores de atividades suspeitas e das regulamentações pertinentes. Elas são tipicamente elaboradas para identificar desvios nos padrões normais de operação

marítima, abrangendo critérios como trajetórias inesperadas, variações anômalas de velocidade, uso de identificadores inválidos (MMSI), além de comportamentos indicativos de *spoofing* ou operações ilícitas.

Comportamentos cinemáticos referem-se ao padrão de movimento das embarcações, abrangendo aspectos como velocidade, direção e suas variações ao longo do tempo. Agências como a Agência Europeia de Segurança Marítima (EMSA) [14] desenvolvem modelos de Monitoramento Automático de Comportamento (ABM), cujo objetivo é detectar e destacar automaticamente comportamentos que sejam de interesse ou preocupação. Entretanto, o uso excessivo dessas regras pode se tornar inviável, uma vez que demanda a verificação de cada acionamento por especialistas. Em sistemas de monitoramento marítimo, é viável criar milhares de regras cinemáticas, sendo responsabilidade do operador revisar cada acionamento e atualizar essas regras conforme necessário. Diversos trabalhos na literatura adotaram esse tipo de solução para identificar comportamentos ou atividades de embarcações. Na tabela 2.2 podemos observar alguns trabalhos que empregam essa abordagem.

Trabalho	Comportamento	Abordagem	Ano
Fitting Empirical Distributions for Vessels Behavioural Analysis and Maritime Anomaly Detection [49]	Dark ship, pesca e outras	Este estudo analisa múltiplas atividades marítimas complexas (como pesca, deriva e atividades ocultas), oriundas do processamento de mensagens AIS coletadas na região do Mar Negro. A identificação das atividades foram feitas com base nas distribuições paramétricas conhecidas (como Gaussiana, Gama, Rayleigh, Log-normal), identificando os parâmetros que melhor se ajustaram a cada atividade.	2023

Trabalho	Comportamento	Abordagem	Ano
Knowledge-based vessel position prediction using historical AIS data [50]	posição futura	um algoritmo Bayesiano baseado em Filtro de Partículas (PF) é proposto para prever a posição de navios dentro de um contexto local onde a rota de navegação é conhecida.	2015
Composite Event Recognition for Maritime Monitoring [51]	diferentes eventos	O trabalho utiliza diversas regras cinemáticas criadas em conjunto com um especialista para detectar diferentes eventos praticados por embarcações como: ancoragem, rendezvous, rebocagem e etc.	2018

Tabela 2.2: Tabela com os trabalhos relacionados que utilizaram regras com base no conhecimento do especialista para a detecção de comportamentos de embarcações.

2.3 Detecção de Atividades de Embarcações

Na literatura, a área de detecção de atividades de embarcações muitas vezes se confunde com a de detecção de comportamentos. No nosso caso, buscamos a detecção de atividades ilegais de embarcações onde somente detectar a atividade não é suficiente, precisamos também conhecer o contexto da situação onde a atividade está sendo realizada. Na tabela 2.3 relacionamos os principais trabalhos que tentaram detectar atividades de embarcações.

Trabalho	Comportamento	Abordagem	Ano
Ontology-based Approach for Vessel Activity Recognition [10]	Todas as Atividades	<p>Este artigo propõe uma abordagem baseada em ontologia para o reconhecimento de atividades de embarcações. O trabalho faz a correlação entre comportamentos e atividades de embarcações na detecção. A descrição semântica é independente de tecnologia, suportando portabilidade e permitindo um primeiro passo em direção a um reconhecimento de atividades de embarcações baseado em dados e conhecimento, aumentando a interpretabilidade das fontes de informação.</p>	2021
Detecting Illegal Maritime Activities from Anomalous Multiscale Fleet Behaviours [13]	Pesca Ilegal	<p>Apresentou uma abordagem para inferir atividades marítimas ilegais em tempo real, examinando o comportamento espacial multiscala de frotas inteiras, ou seja, grupos de embarcações que transmitem suas localizações via AIS. A hipótese é de que atividades ilegais induzem uma anomalia espacial multiescala, que pode ser observada usando essa abordagem. O método foi desenvolvido utilizando um modelo baseado em agentes.</p>	2022

Trabalho	Comportamento	Abordagem	Ano
Maritime Anomaly Detection using Gaussian Process Active Learning [52]	Atividades Anômalas	Apresenta um modelo Bayesiano não paramétrico orientado por dados, baseado em Processos Gaussianos, para modelar o comportamento normal de navegação. O modelo é aprendido a partir de dados AIS e utiliza Aprendizado Ativo para treinamento. Com base na normalidade, embarcações podem ser identificadas como potencialmente anômalas e priorizadas para investigações adicionais. O desempenho do modelo é avaliado por sua capacidade de detectar anomalias geradas artificialmente em locais ao redor do Reino Unido.	2012

Tabela 2.3: Tabela com os trabalhos relacionados que tentaram detectar atividades de embarcações.

Capítulo 3

Revisão de Conceitos

Neste capítulo, apresentaremos alguns conceitos relacionados ao domínio marítimo, bem como os principais conceitos computacionais necessários para a nossa solução.

3.1 Conceitos sobre o Domínio Marítimo

Nesta seção, abordaremos os principais conceitos relacionados ao domínio marítimo, os quais são essenciais para a compreensão do presente trabalho e de sua aplicação prática. Ademais, a compreensão das áreas marítimas é crucial para a aplicação das legislações pertinentes e para a proteção da soberania nacional no ambiente marítimo. Dentro desse contexto, é fundamental distinguir entre duas áreas de grande relevância: a Zona Econômica Exclusiva (ZEE) e o Mar Territorial (MT).

3.1.1 Zona Econômica Exclusiva (ZEE)

A Zona Econômica Exclusiva (ZEE) é uma área marítima que se estende até 200 milhas náuticas (MN) a partir da linha de base costeira de um país. Dentro dessa zona, o país costeiro detém direitos soberanos para a exploração e utilização dos recursos naturais, tanto os vivos (como a pesca) quanto os não vivos (como petróleo, gás e minerais). Esses direitos também abrangem a geração de energia a partir das águas, correntes e ventos [53].

Embora o país possua direitos exclusivos sobre os recursos naturais, a ZEE não se equipara a um território nacional pleno. Outros países ainda detêm liberdade de navegação e de sobrevoo, além do direito de instalar cabos e dutos submarinos. Entretanto, o país costeiro pode regulamentar determinadas atividades para proteger seus recursos, assegurar a conservação ambiental e prevenir a poluição marinha.

As principais leis e convenções que regulam a ZEE incluem:

- **Convenção das Nações Unidas sobre o Direito do Mar (UNCLOS):**

Estabelece os direitos e responsabilidades dos estados costeiros sobre a exploração e conservação dos recursos na ZEE [54].

- **Leis nacionais de pesca e conservação:** Determinam os limites de captura, técnicas permitidas e proibições para proteger os recursos pesqueiros e os ecossistemas marinhos.
- **Regulações ambientais:** Incluem a prevenção de poluição por navios, medidas de proteção de espécies ameaçadas e restrições em áreas sensíveis, como recifes de corais e habitats de espécies marinhas.

3.1.2 Mar Territorial (MT)

O Mar Territorial (MT) é a zona marítima que se estende até 12 milhas náuticas (MN) a partir da linha de base costeira de um país. Dentro dessa área, o Estado costeiro exerce soberania plena, similar àquela exercida sobre o território terrestre, incluindo a aplicação das leis civis, criminais e aduaneiras, entre outras [54].

Apesar de ser uma área sob jurisdição plena, outros estados possuem o direito de realizar a “passagem inocente” no mar territorial, desde que a navegação seja contínua, rápida e não comprometa a paz, a boa ordem ou a segurança do estado costeiro. A passagem inocente proíbe atividades como pesca, coleta de dados ou manobras militares sem a permissão do estado costeiro.

As principais legislações que se aplicam no MT incluem:

- **Leis nacionais de soberania e segurança:** O estado costeiro pode impor suas leis de imigração, alfândega e saúde, além de realizar operações de fiscalização para proteger a segurança nacional e a ordem pública.
- **Leis de navegação e trânsito:** O estado costeiro pode estabelecer regras de navegação, como rotas marítimas específicas, para garantir a segurança e a proteção ambiental.
- **Leis ambientais:** Incluem a prevenção de poluição e medidas de proteção ambiental mais rigorosas, dado que o MT está mais próximo das áreas costeiras e ecossistemas frágeis.

O respeito às legislações internacionais e nacionais dentro da ZEE e do MT é essencial para a governança marítima e para a manutenção da ordem no ambiente marinho, assegurando a exploração sustentável dos recursos e a segurança das águas territoriais.

3.1.3 Conscientização do Domínio Marítimo (MDA)

A Conscientização do Domínio Marítimo (MDA, do inglês *Maritime Domain Awareness*) refere-se à compreensão efetiva de qualquer atividade associada ao domínio marítimo que possa impactar a segurança, a economia ou o meio ambiente [55]. Essa conscientização é fundamental para a segurança das atividades marítimas, pois permite a detecção de operações legais e ilegais, como pesca, transporte de carga, tráfico de drogas e outras atividades ilícitas [3]. A MDA emprega uma combinação de sistemas de vigilância que podem utilizar radares, dados do AIS e imagens de satélite, para fornecer informações em tempo real sobre as condições e atividades no mar. No âmbito de nosso trabalho, o foco será na utilização do AIS, considerando que se trata de um sensor amplamente utilizado e que possui conjuntos de dados disponíveis para fins acadêmicos.

3.1.4 Monitoramento Marítimo

O monitoramento marítimo é efetuado por diversas autoridades governamentais, militares e agências de segurança ao redor do mundo, com a finalidade de assegurar a segurança das atividades no mar e a soberania nacional sobre as águas territoriais. Adicionalmente, organizações internacionais e entidades privadas, como empresas de transporte marítimo e pesca, também contribuem para o processo de monitoramento, especialmente em áreas de relevância econômica ou ambiental.

As regiões monitoradas geralmente incluem áreas de intenso tráfego marítimo dentro do MT e da ZEE. O principal foco é nas rotas de navegação internacional, portos e áreas costeiras. Áreas sensíveis, como regiões de exploração de petróleo e gás, zonas de conservação ambiental e áreas com alto risco de pirataria, também são monitoradas de maneira intensiva.

As situações monitoradas englobam uma ampla variedade de atividades, tais como:

- Tráfego de embarcações comerciais e pesqueiras.
- Atividades ilegais, como pesca não autorizada, contrabando, tráfico de drogas e pirataria.
- Operações de busca e salvamento, em caso de acidentes ou desaparecimentos no mar.
- Áreas de proteção ambiental (APA) e possíveis casos de poluição hídrica.
- Navegação em áreas protegidas ou de segurança nacional, visando prevenir colisões e outras ameaças à segurança.

Para proceder com esse monitoramento, é possível empregar diversos tipos de sistemas e tecnologias, incluindo:

- **Sistema de Identificação Automática (AIS):** Uma tecnologia de rastreamento usada para monitorar a posição, velocidade e curso das embarcações em tempo real. Trata-se de um sensor colaborativo, ou seja, o navio pode ligá-lo ou desligá-lo a qualquer momento. O Sistema de Identificação Automática (AIS) utiliza frequências de rádio VHF para transmitir informações entre embarcações e estações receptoras. As mensagens transmitidas via AIS incluem dados como posição, velocidade e direção da embarcação e são captadas por estações costeiras dentro do alcance do VHF, que geralmente é de até 40 milhas náuticas. Para ampliar o monitoramento de embarcações em áreas mais remotas e oceânicas, satélites de baixa órbita (*Low Earth Orbit*) também recebem sinais AIS. Esses satélites cobrem grandes áreas e permitem que informações das embarcações sejam captadas e retransmitidas para as autoridades em terra, mesmo fora do alcance das estações costeiras. A combinação de recepção por VHF em estações terrestres e satélites possibilita uma cobertura global do tráfego marítimo, aumentando a segurança e a eficiência da vigilância marítima.
- **Radares Costeiros:** O radar é um sensor ativo, o que significa que ele emite ondas eletromagnéticas e detecta os ecos refletidos pelos objetos para determinar sua posição, velocidade e trajetória. Os radares costeiros geralmente operam em frequências que permitem a detecção de embarcações e outros alvos em distâncias de até 60 milhas náuticas (MN), cobrindo áreas próximas ao litoral. Já radares de longo alcance são projetados para monitorar áreas mais distantes, podem ter um alcance de até 250 MN ou mais. Esses radares são essenciais para a detecção de embarcações em alto-mar e oferecem capacidade de vigilância além da linha do horizonte, conhecida como Over-the-Horizon (OTH). O radar OTH, por exemplo, é um radar de alta frequência (HF) que usa a refração da ionosfera para detectar alvos a distâncias superiores a 1.000 MN, sendo particularmente útil para vigilância estratégica e monitoramento de ameaças que se aproximam das águas territoriais de um país.
- **Satélites de Observação Marítima:** Os satélites desempenham um papel crucial no monitoramento de grandes áreas marítimas, fornecendo imagens detalhadas e contínuas do tráfego no mar. Entre os tipos de sensores utilizados, destaca-se o **Radar de Abertura Sintética (SAR, do inglês *Synthetic Aperture Radar*)**, que é uma tecnologia de radar ativa capaz de operar independentemente das condições meteorológicas e de luminosidade, o que a torna ideal para vigilância marítima.

- **Sistemas de Vigilância Não Tripulados (Drones):** Usados para obter imagens aéreas em tempo real, cobrindo grandes áreas com eficiência. No entanto, esse tipo de equipamento possuí uma autonomia limitada e além de ser vulnerável a condições climáticas.

Além disso, um dos esforços mais notáveis no campo do monitoramento marítimo é o **Global Maritime Situational Awareness** (GMSA), um programa estabelecido pelos Estados Unidos, cujo objetivo é promover a cooperação internacional no compartilhamento de informações sobre o tráfego marítimo. O GMSA facilita a troca de dados AIS entre os países participantes, permitindo uma visão mais abrangente e integrada das atividades no mar. O compartilhamento de dados AIS aprimora significativamente a capacidade de detectar e rastrear embarcações suspeitas, prevenindo atividades ilegais e fortalecendo a segurança marítima global. Esse tipo de cooperação internacional é essencial para enfrentar ameaças transnacionais, como o tráfico de drogas, contrabando e pirataria, que frequentemente envolvem operações em águas de múltiplos países. Para isso, diversos sistemas do GMSA são utilizados, tais como o sistema SeaVision [56]. O SeaVision é uma ferramenta de consciência situacional marítima baseada na web, que permite aos usuários visualizar e compartilhar uma ampla gama de informações marítimas, com o propósito de aprimorar as operações, aumentar a segurança marítima e consolidar parcerias dentro da comunidade marítima.

Para auxiliar o usuário na gestão das informações, o SeaVision oferece análises fundamentadas em regras estabelecidas pelo próprio usuário, avaliando e notificando sobre atividades ou eventos marítimos específicos. Esse sistema pode servir como um exemplo do que está sendo atualmente utilizado pelas autoridades marítimas ao redor do mundo.

3.1.5 Competências das Autoridades Marítimas

As autoridades marítimas (AM) desempenham um papel central na proteção da soberania e da segurança marítima, sendo responsáveis por uma série de atividades, incluindo a fiscalização de embarcações, o combate a atividades ilícitas e a garantia da segurança na navegação. Para cumprir essas responsabilidades, as autoridades necessitam realizar inspeções navais de rotina e operações de vigilância específicas, o que requer um esforço significativo em termos de recursos e logística.

As inspeções navais, especialmente aquelas direcionadas a embarcações suspeitas de atividades ilícitas, como tráfico de drogas, demandam a mobilização de navios, equipes especializadas e equipamentos de inspeção e vigilância. Por exemplo, para inspecionar um navio suspeito de envolvimento em tráfico, a Marinha pode necessitar mobilizar uma fragata ou um navio patrulha, os quais devem estar equipados

com embarcações rápidas e helicópteros para apoiar a operação. Ademais, é imprescindível um contingente de pessoal especializado, incluindo equipes de abordagem, inspetores de órgãos competentes e profissionais de saúde.

O planejamento e a execução dessas operações podem envolver:

- **Mobilização de meios navais:** Navios-patrulha, fragatas ou corvetas, além de embarcações menores para abordagem.
- **Equipes de bordo:** Tripulação, inspetores de outros órgãos, pessoal de abordagem, peritos, médicos para lidar com situações imprevistas e entre outros.
- **Materiais:** Helicópteros, drones, câmeras especiais, detectores de substâncias ilícitas, equipamentos para comunicação, armamento específicos e etc.

No entanto, um erro de predição, como a classificação incorreta de uma embarcação como suspeita, pode resultar em um desperdício considerável de recursos e esforços. Por exemplo, se uma operação de inspeção é iniciada com base em informações equivocadas, os custos relacionados ao combustível, às horas de trabalho da tripulação, ao desgaste dos equipamentos e ao impacto no cronograma das operações navais em andamento representam desperdícios significativos. Além disso, a mobilização inadequada de meios navais pode afetar a prontidão para responder a outras emergências ou ameaças reais.

Um exemplo prático seria a Marinha mobilizar um navio-patrulha para interceptar um cargueiro suspeito de tráfico de drogas, identificado por meio de um sistema de monitoramento automatizado. A operação pode incluir o envio de um helicóptero para reconhecimento aéreo, barcos de abordagem rápida, e uma equipe completa de inspetores e pessoal de segurança. Se a suspeita for infundada, todo esse esforço resultará em um gasto de recursos financeiros e humanos sem um retorno operacional, além de diminuir a confiança nos sistemas de predição empregados.

Assim, a precisão na identificação de embarcações suspeitas é crucial para otimizar os recursos das autoridades marítimas e aumentar a eficiência das operações de fiscalização. A integração de soluções tecnológicas confiáveis, como sistemas de aprendizado de máquina e fusão de dados, é essencial para minimizar os riscos de erro e evitar desperdícios nas operações de segurança marítima.

3.2 Conceitos Computacionais

Nesta seção, serão apresentados os conceitos computacionais que fundamentam a metodologia proposta neste trabalho.

3.2.1 Ensemble Models

Os *ensembles* são técnicas de aprendizado de máquina que combinam previsões de múltiplos modelos individuais para alcançar um desempenho mais robusto e preciso. Essas técnicas podem ser classificadas em duas categorias principais:

- **Ensembles homogêneos:** Utilizam múltiplas instâncias do mesmo tipo de modelo, como vários classificadores de decisão (por exemplo, árvores de decisão) treinados em diferentes subconjuntos de dados ou com parâmetros diferentes. Exemplos comuns de ensembles homogêneos incluem *bagging* (como o *Random Forest*) e *boosting* (como o *XGBoost*).
- **Ensembles heterogêneos:** Combinam diferentes tipos de modelos, como redes neurais, regressões logísticas e máquinas de vetores de suporte, para aproveitar a diversidade de abordagens. Esse tipo de ensemble é eficaz quando modelos diferentes capturam aspectos distintos do problema.

A combinação dos modelos em um ensemble pode ser efetuada de diversas maneiras:

- **Média (*Averaging*):** Os modelos individuais geram previsões, e o resultado final é obtido pela média aritmética das previsões, o que é eficaz em problemas de regressão.
- **Votação (*Voting*):** É utilizada em problemas de classificação, onde os modelos individuais votam em uma classe, e a classe com mais votos é selecionada. Pode ser implementada de forma simples (votação majoritária) ou ponderada (quando modelos com melhor desempenho têm maior influência).
- **Stacking:** Trata-se de uma técnica mais avançada que combina os modelos por meio de um meta-modelo, ou seja, um segundo modelo de aprendizado de máquina que aprende a partir das saídas dos modelos de base. O *stacking* usa as previsões dos modelos de base como características de entrada, buscando capturar relações entre os erros residuais dos modelos e ajustar a predição final de maneira mais precisa.

No presente trabalho, optamos por utilizar o *stacking* como método de ensemble, devido à sua capacidade de combinar de forma eficiente modelos de comportamento e regras elaboradas por especialistas. O *stacking* é considerado o método mais apropriado para a fusão de dados neste contexto, pois atua como um *meta-modelo*, sendo apto a aprender como combinar as previsões dos modelos de comportamento com as saídas das regras desenvolvidas por especialistas.

Matematicamente, o *stacking* pode ser descrito da seguinte maneira em relação ao nosso problema:

$$F(x) = H(M_1(x), M_2(x), \dots, M_n(x), R_1(x), R_2(x), \dots, R_n(x)), \quad (3.1)$$

onde:

- $F(x)$ é a predição final do ensemble.
- $M_1(x), M_2(x), \dots, M_n(x)$ representam as predições dos modelos de comportamento individuais.
- $R_1(x), R_2(x), \dots, R_n(x)$ são as predição baseadas nas regras definidas por especialistas.
- H é o meta-modelo, que recebe as saídas dos modelos de base e aprende a combinação ótima das predições para minimizar o erro final.

O uso de *stacking* é particularmente benéfico no contexto deste trabalho, pois permite ajustar a predição final de forma mais sofisticada, capturando não apenas a diversidade dos modelos de comportamento, mas também incorporando o conhecimento especializado das regras. Essa abordagem reduz a taxa de falsos positivos e melhora a detecção de atividades anômalas no domínio marítimo. Maiores detalhes sobre métodos de ensembles e *stacking* podem ser encontradas em [57].

3.2.2 Aprendizado Ativo

Segundo [58] o Aprendizado Ativo (AL, do inglês *Active Learning*) é uma técnica de aprendizado de máquina que tem como objetivo otimizar o processo de treinamento de modelos, selecionando amostras de dados que sejam mais informativas para o modelo e apresentando-as ao usuário para rotulagem. Essa abordagem é particularmente eficaz em cenários onde a rotulagem de dados é cara ou demorada, como no caso de dados de trajetórias marítimas, onde especialistas necessitam identificar manualmente comportamentos suspeitos ou anômalos.

O funcionamento do Aprendizado Ativo pode ser descrito em três etapas principais:

- **Seleção de Amostras:** O modelo de AL escolhe as amostras de dados mais informativas para rotulagem. A seleção pode ser feita com base em critérios como incerteza (*uncertainty sampling*), divergência (*divergence-based sampling*) ou representatividade (*representativeness sampling*). O objetivo é identificar as amostras que mais contribuiriam para a melhoria do modelo se rotuladas.

- **Rotulagem por Especialistas:** As amostras selecionadas são apresentadas ao usuário (especialista), que fornece rótulos precisos. Esses rótulos são então usados para treinar o modelo, ajustando seus parâmetros e aumentando sua precisão.
- **Treinamento Iterativo:** O processo de seleção de amostras e rotulagem é repetido iterativamente, com o modelo se tornando progressivamente mais preciso a cada iteração.

No caso de um conjunto de dados que não possui rótulos iniciais, é necessário realizar um *cold start*, onde um conjunto inicial de amostras é rotulado manualmente para dar início ao treinamento do modelo. Essa etapa é fundamental para fornecer uma base mínima de dados rotulados, a partir da qual o Aprendizado Ativo pode começar a identificar amostras mais informativas. No algoritmo 1 podemos observar o funcionamento do aprendizado ativo.

Algorithm 1: Aprendizado Ativo para Detecção de Atividades Ilegais

```

Result: Modelo treinado para detectar atividades ilegais
Cold start: Selecionar um conjunto inicial de amostras;
amostras_rotuladas ← selecionar_amostras_iniciais(Dados);
rotulos_iniciais ← obter_rotulos(usuario, amostras_rotuladas);
Modelo.treinar(amostras_rotuladas, rotulos_iniciais);
for iter = 1 até Iterações do
    amostras_candidatas ← selecionar_amostras(Dados, tamanho_batch,
        Modelo);
    rotulos_candidatos ← obter_rotulos(Usuario, amostras_candidatas);
    amostras_rotuladas ← amostras_rotuladas + amostras_candidatas;
    Modelo.treinar(amostras_rotuladas, rotulos_candidatos);
end
return Modelo;
```

No algoritmo 1 podemos fazer as seguintes observações:

- A função `selecionar_amostras_iniciais(Dados)` escolhe um conjunto de amostras iniciais para o *cold start*, que são rotuladas manualmente.
- A função `selecionar_amostras(Dados, Tamanho_Batch, Modelo)` identifica as amostras mais informativas a cada iteração.
- A função `obter_rotulos(Usuario, Amostras)` representa o processo de rotulagem pelo especialista.
- O modelo é treinado iterativamente com as amostras rotuladas, refinando a detecção de comportamentos anômalos.

Dessa forma, o aprendizado ativo permite que o modelo de detecção de atividades ilegais seja refinado com uma quantidade limitada de dados rotulados, aumentando a eficiência e reduzindo o esforço de rotulagem manual.

3.2.3 Modelo JDL

O Modelo JDL (*Joint Directors of Laboratories*) de fusão de dados é um *framework* conceito amplamente utilizado em sistemas de vigilância e monitoramento. Criado nos Estados Unidos na década de 1980 sob a supervisão do Departamento de Defesa dos EUA, o modelo JDL foi inicialmente desenvolvido para apoiar a integração de informações em ambientes de defesa e inteligência. Desde então, ele tem sido aplicado em diversas áreas, incluindo vigilância marítima, segurança pública e gestão de crises, devido à sua capacidade de estruturar de maneira eficaz o processo de fusão de dados [59].

O modelo estrutura a fusão de dados em cinco níveis distintos:

- **Nível 0 (Pré-processamento):** Refere-se ao processamento inicial dos dados brutos, como filtragem, limpeza e transformação para facilitar a análise subsequente.
- **Nível 1 (Avaliação de Objetos):** Foca na identificação e rastreamento de objetos individuais (como embarcações) nos dados processados. É responsável por correlacionar dados de diferentes sensores para determinar características de localização e movimento.
- **Nível 2 (Avaliação de Situação):** Integra as informações dos objetos identificados para compreender a situação geral, avaliando a relação entre os objetos e seu contexto operacional. Esse nível é crucial para identificar padrões de comportamento e possíveis ameaças no domínio marítimo.
- **Nível 3 (Avaliação de Impacto):** Envolve a análise das possíveis implicações das situações identificadas, como avaliar a probabilidade de uma atividade ser ilícita ou de alto risco. Isso inclui a previsão das consequências de ações ou eventos observados.
- **Nível 4 (Gerenciamento de Processos):** Ajustes no processo de fusão de dados para otimizar o desempenho do sistema. Este nível foca também em oferecer suporte à tomada de decisão, fornecendo recomendações baseadas nas avaliações anteriores. O objetivo é facilitar a ação por parte dos operadores ou autoridades, apresentando informações de forma clara e açãoável.

O Modelo JDL é reconhecido como um marco no campo de fusão de dados, oferecendo uma estrutura robusta para integração de múltiplas fontes de informação

em ambientes complexos. Essa abordagem permite otimizar a detecção de atividades ilícitas e a resposta a ameaças em tempo real, tornando-se essencial para a vigilância marítima e outras aplicações de segurança.

Capítulo 4

Construção do Framework para Classificação de Atividades Ilegais de Embarcações

Neste capítulo, apresentaremos a metodologia adotada neste trabalho, juntamente com a aplicação dos conceitos discutidos no capítulo anterior. Serão detalhadas a construção de cada modelo de comportamento, as regras do especialista, as atividades a serem detectadas, a estrutura do *ensemble* e o aprendizado ativo (*active learning*). Por fim, organizaremos essas técnicas em um *framework* baseado no modelo JDL.

4.1 Extração das Trajetórias

Uma trajetória é uma sequência de pontos no espaço-tempo que descreve o deslocamento de uma embarcação. Cada ponto da trajetória é caracterizado por coordenadas espaciais (latitude e longitude) e um registro temporal (*timestamp*). A trajetória proporciona uma visão detalhada do percurso da embarcação ao longo do tempo. Além disso, temos à disposição dados de velocidade em relação ao solo (*Speed Over Ground* - SOG) e rumo em relação ao solo (*Course Over Ground* - COG), os quais indicam a velocidade e a direção da embarcação em relação ao fundo marinho, conforme ilustrado na Figura 4.1. Uma trajetória normal segue um padrão de movimento esperado, fundamentado em comportamentos históricos ou regras específicas. Por exemplo, visando à economia de combustível, navios de carga geralmente seguem uma rota comercial predefinida a uma velocidade constante, evitando desvios abruptos. No caso de embarcações pesqueiras, estas podem apresentar trajetórias específicas de atividades pesqueiras nos locais onde estão efetivamente operando. Em contrapartida, uma trajetória anômala pode revelar desvios significativos do padrão

esperado, indicando comportamentos atípicos ou suspeitos. A próxima seção aborda esses comportamentos e as metodologias para sua detecção.

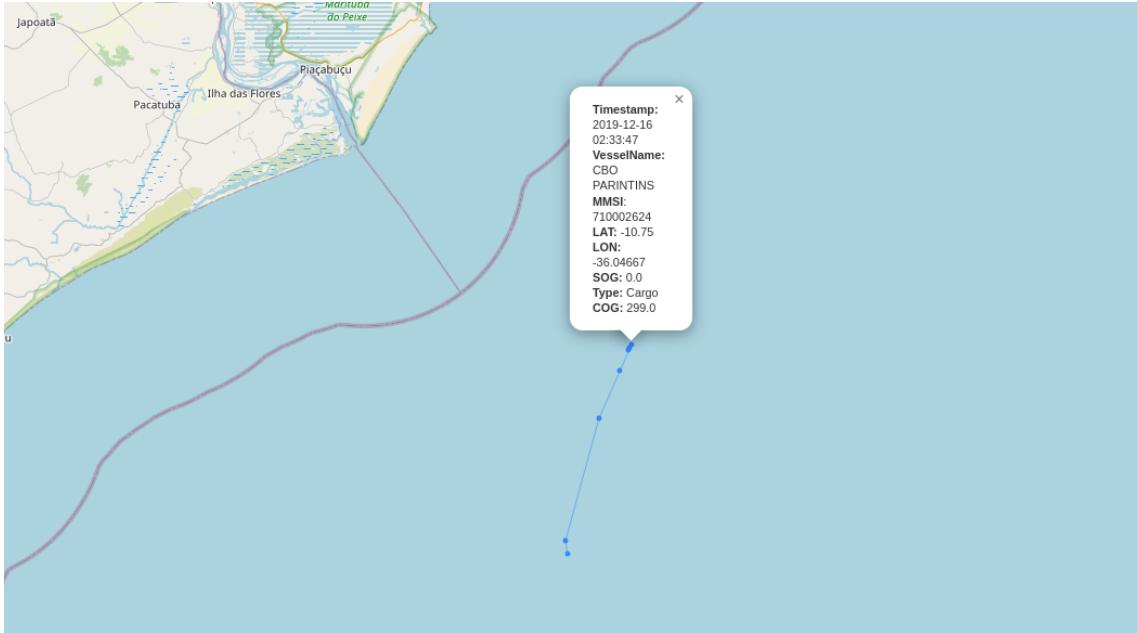


Figura 4.1: Exemplo de trajetória de embarcação. Cada ponto azul representa uma mensagem AIS recebida da embarcação.

Para a construção das trajetórias, os dados AIS foram inicialmente convertidos em *GeoDataFrames* por meio da biblioteca `geopandas` [60], permitindo o tratamento espacial das informações. Em seguida, utilizou-se a biblioteca `movingpandas` [61] para a modelagem das trajetórias. A segmentação das trajetórias de cada embarcação em subtrajetórias foi realizada por meio da classe `ObservationGapSplitter`, disponível no `movingpandas`, com o parâmetro `gap` definido em 90 minutos. Com essa configuração, períodos de inatividade superiores a 90 minutos foram interpretados como paradas (por exemplo, em áreas de fundeio), levando à fragmentação da trajetória em múltiplas partes. Assim, o algoritmo considera o fim de uma trajetória e o início de outra a partir do momento em que a embarcação retoma seu deslocamento.

4.2 Detectando Comportamentos de Embarcações

Os comportamentos das embarcações representam as respostas observáveis de um navio a estímulos externos e internos, os quais não estão necessariamente associados a um objetivo específico. Esses comportamentos podem ser rotineiros, como a realização de uma rota comercial previamente estabelecida, ou podem sinalizar atividades específicas, tanto legais quanto ilegais. A compreensão e a detecção desses comportamentos são cruciais para a segurança, proteção e vigilância marítima.

A detecção e o monitoramento dos comportamentos das embarcações fundamentam-se, primordialmente, em diversos sensores. O mais proeminente dentre eles é o AIS, que, como mencionado nas seções anteriores, constitui um sistema de rastreamento colaborativo amplamente utilizado em navios e pelos serviços de tráfego marítimo. Os transponders AIS transmitem automaticamente informações, como posição, velocidade e estado de navegação, em intervalos regulares por meio de um transmissor VHF. Outros sensores incluem radares, que detectam e mapeiam as posições dos navios com base em ondas de rádio refletidas, e imagens de satélite, que fornecem uma visão abrangente das regiões marítimas e podem capturar movimentos das embarcações.

Diversos comportamentos têm sido identificados e estudados na literatura marítima. Neste estudo, abordaremos os comportamentos mais recorrentes, apresentando uma breve explicação de cada um nas subseções subsequentes. Implementamos cada modelo de comportamento descrito a seguir; estes modelos recebem as trajetórias das embarcações como entrada e produzem um valor em ponto flutuante que representa a probabilidade do comportamento detectado. Nas próximas subseções, forneceremos explicações detalhadas de cada comportamento implementado em nosso framework.

4.2.1 AIS Spoofing

A manipulação maliciosa dos dados AIS transmitidos pelas embarcações, com o intuito de disfarçar sua identidade ou localização, foi comprovada como possível em diversos estudos [62, 63]. Essa prática pode ocultar ou facilitar atividades ilícitas, comprometer sistemas de monitoramento e gerar riscos à navegação. Vários estudos têm se dedicado à detecção da manipulação *spoofing* dos dados do AIS, como em [24, 27, 30]. Em [24], os autores propuseram uma arquitetura para detectar *spoofing* dos dados do AIS em fluxos de dados online. A referida arquitetura armazena o MMSI da embarcação em uma tabela de dispersão (*hash*) e, para cada trajetória associada, computa a distância haversine entre pontos consecutivos e a velocidade média. Se a velocidade média exceder 50 nós, a trajetória é considerada como tendo sofrido *spoofing*. O valor de 50 nós foi utilizado uma vez que poucas embarcações conseguem atingir essa velocidade, com exceção de embarcações militares e de alta velocidade.

Dessa maneira, para detectar o comportamento de *spoofing* no AIS, calculamos a velocidade em nós entre cada par de pontos consecutivos da trajetória. Caso em qualquer segmento a velocidade exceder 50 nós, essa será considerada fisicamente impossível, e a trajetória será classificada como um caso de *spoofing*. Ademais, trajetórias que apresentam pontos em locais impossíveis, como ao atravessar áreas terrestres, também serão consideradas como tendo sofrido *spoofing*.

Em nosso framework, a detecção de trajetórias utilizando *spoofing* de AIS é realizada conforme representado no algoritmo 2.

Algorithm 2: Detecção de *spoofing* em trajetórias AIS

Result: Identificar trajetórias com possível *spoofing*

```

foreach trajetória do
    foreach ponto consecutivo na trajetória do
        calcular a velocidade entre os pontos;
        if velocidade > 50 nós then
            marcar a trajetória como spoofed;
            break
        end
    end
    if trajetória cruza terra firme then
        marcar a trajetória como spoofed;
    end
end
```

Para verificar se a trajetória cruzou terra, utilizamos um *shapefile* da costa brasileira, disponível em [64]. Empregamos a biblioteca *Shapely* [65] para carregar o arquivo e a biblioteca *GeoPandas* [60] para realizar operações geoinformacionais. Por fim, utilizamos a função `contains` do *GeoPandas* para determinar se algum ponto da trajetória se encontrava dentro do *shapefile*.

4.2.2 Dark Ships

Embarcações que desativam deliberadamente seus transponders AIS para evitar detecção frequentemente indicam uma intenção de se envolver em atividades ilícitas. Entretanto, essa prática pode, em algumas ocasiões, gerar ambiguidade quando a embarcação navega por áreas com interferência de sinal ou condições meteorológicas adversas. Portanto, é essencial conhecer essas áreas para que possamos inferir com precisão esse tipo de comportamento.

Na literatura, encontramos diversas abordagens para solucionar esse problema, principalmente por meio do uso de sensores ativos para detecção. Para sistemas não colaborativos, trabalhos como [66, 67] empregaram radares que, quando integrados ao AIS, permitem associar as trajetórias do AIS e do radar, possibilitando a detecção de anomalias quando estas divergem. Outros estudos buscaram resolver o problema utilizando apenas o sensor AIS [29, 31]. Contudo, utilizando somente o AIS, podemos detectar lacunas de transmissão na trajetória e presumir que a embarcação pode ter desativado o transponder AIS. A Figura 4.3 ilustra um exemplo de lacunas em trajetórias. Os pontos azuis representam onde a embarcação transmitiu o sinal AIS. Os quadrados verdes indicam regiões onde a embarcação transmitiu dentro da



Figura 4.2: No mapa podemos ver um exemplo de célula geohash. O quadrado azul é uma célula de resolução 5 (4,9 Km) e no seu interior a codificação gerada para a região.

área, enquanto os quadrados vermelhos sinalizam lacunas de transmissão, indicando que a embarcação deveria ter transmitido dentro daquela região. Os quadrados verdes e vermelhos foram codificados com o sistema Geohash. O Geohash é uma técnica de codificação geoespacial que converte coordenadas de latitude e longitude em uma sequência curta de caracteres alfanuméricicos [68]. Cada caractere adicional na sequência aumenta a precisão da localização. Essa codificação permite a divisão do espaço geográfico em células hierárquicas, facilitando a indexação, armazenamento e consulta eficiente de dados espaciais, além de simplificar a comparação de proximidade entre diferentes pontos. Na figura 4.2 podemos observar um exemplo.

Para calcular essas lacunas nas trajetórias, inicialmente, estabelecemos uma trajetória única para cada embarcação, independentemente de permanecer estacionária ou não. Em seguida, geramos *geohashes* com precisão de 3 (156 km) ao longo da trajetória. Considerando que uma embarcação transmite um sinal AIS pelo menos a cada 2 horas e extrapolando sua velocidade para 40 nós, a embarcação deveria emitir um sinal a cada 80 milhas náuticas (aproximadamente 148 km). Dessa forma, mapeando todos os quadrados de *geohash* pelos quais a embarcação passou, como ilustrado na Figura 4.3, identificamos em vermelho todos os quadrados onde a trajetória da embarcação cruzou, mas não houve transmissão.

Outro fator importante a ser verificado é se a área onde a embarcação deveria ter transmitido apresenta problemas de interferência eletromagnética. Para isso, armazenamos um histórico de *geohashes* com as respectivas quantidades de transmissões já realizadas a partir daquela região. Quando identificamos uma lacuna na trajetória de uma embarcação, verificamos se transmissões AIS ocorreram previamente nas áreas indicadas em vermelho. Se não houver transmissões nessa região, ela não

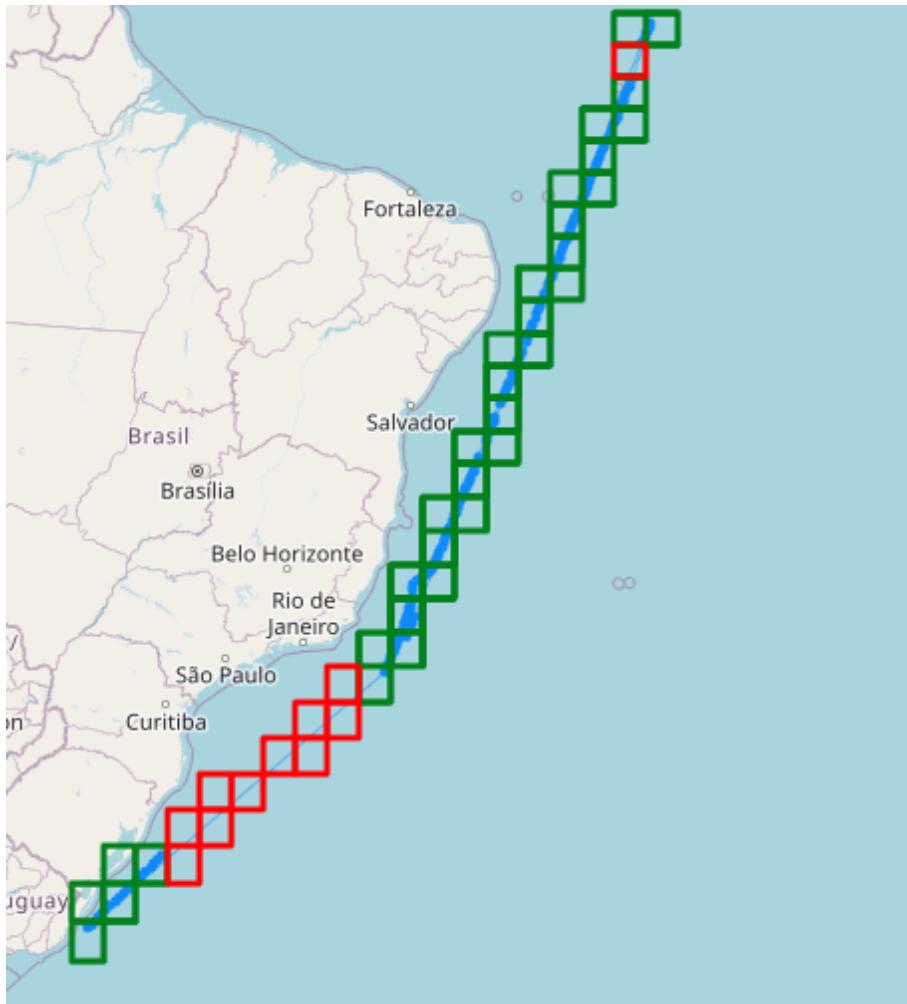


Figura 4.3: A imagem mostra um exemplo de detecção de lacunas na trajetória de uma embarcação. Os pontos azuis refletem a trajetória da embarcação. Os quadrados vermelhos indicam onde a embarcação deveria ter transmitido o sinal AIS, enquanto os quadrados verdes representam onde a embarcação transmitiu. Os quadrados vermelhos representam lacunas de transmissão.

é considerada uma lacuna na trajetória. Portanto, em nosso trabalho, consideramos essas lacunas de transmissão juntamente com os históricos de transmissões dos *geohashes* para determinar se a embarcação deveria ter transmitido um sinal AIS a partir daquela posição. No algoritmo 3 podemos observar uma lógica simplificada para a detecção de lacunas de transmissão de AIS nas trajetórias dos navios.

Algorithm 3: Detecção de GAP de transmissão AIS em trajetórias de embarcações.

Result: Identificar trajetórias com possível *gaps* de transmissão AIS

```
trajetorias = cria_trajetorias_unicas_por_mmsi()
foreach trajetorias do
    linestring = transforma trajetória em line string;
    lista_geohashes = geohashes_do_caminho( linestring ) {todos verdes};
    lista_geohashes_pontos = geohashes_lista_dePontos( trajetoria );
    foreach geohash in lista_geohashes do
        if geohash não está em lista_geohashes_pontos then
            if geohash em histórico de transmissão then
                | marcar a trajetória como gap de transmissão {vermelho};
            end
        end
    end
    insere lista_geohashes_pontos em histórico de transmissão;
end
```

4.2.3 Encontros entre Embarcações

Em um contexto marítimo, o termo “encontro” refere-se a um engajamento não planejado ou acidental entre embarcações, ocorrendo frequentemente caracterizado por sua ocorrência fortuita, sem qualquer coordenação prévia. Em contrapartida, um *rendezvous* denota um encontro planejado e intencional em um momento e local específicos, envolvendo coordenação e acordo prévios entre as partes envolvidas. Enquanto um encontro é espontâneo e pode ocorrer inesperadamente, um *rendezvous* é deliberado e organizado, comumente empregado para atividades ilícitas ou propósitos específicos nos quais embarcações necessitam se reunir em um ponto predeterminado. Para detectar um *rendezvous*, é necessário primeiramente identificar o encontro. Para inferir a intenção subjacente a um encontro, é imprescindível analisar o contexto da situação. Portanto, para este comportamento específico de embarcações, focaremos inicialmente na detecção de encontros.

Duas ou mais embarcações que se encontram em mar aberto podem estar envolvidas na transferência de mercadorias ou pessoas. Isso pode ser indicativo de comércio legítimo, mas também de atividades como contrabando ou tráfico. Embora esse encontro possa ser considerado um comportamento suspeito, nem sempre essa suposição é válida. Em áreas de exploração petrolífera, numerosas embarcações de apoio transportam pessoal, petróleo e materiais de suporte. Nessas regiões, a prática de transferência entre navios (*ship-to-ship transfer*), que consiste na movimentação de carga entre embarcações no mar, é comum. Nessas situações, a proximidade

entre embarcações é considerada normal, uma vez que estão localizadas próximas a portos onde podem atracar uma perto da outra. A Figura 4.4 ilustra um exemplo de um encontro entre embarcações. Em verde, temos a embarcação 1; em vermelho, a embarcação 2; e os quadrados azuis representam áreas de exploração petrolífera.

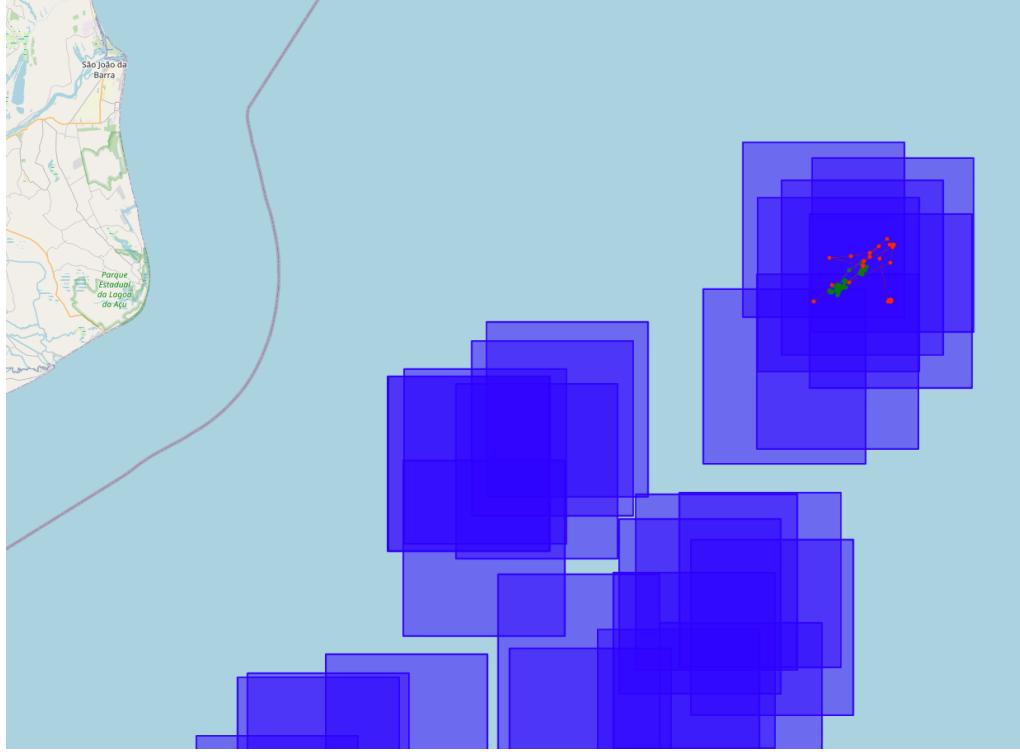


Figura 4.4: A imagem mostra um exemplo de encontro entre embarcações. Os pontos verdes representam a localização da embarcação 1, enquanto que os pontos vermelhos representam a embarcação 2. Os quadrados azuis apresentam áreas de exploração de petróleo (FPSO). A imagem mostra que a embarcação realizou um encontro dentro de uma área de FPSO.

Diversos estudos, como [34, 35, 69, 70], propuseram soluções para a detecção desses encontros, levando em consideração sua intencionalidade. No entanto, para simplificar, neste trabalho utilizaremos apenas as posições AIS e os registros temporais (*timestamps*) para detectar embarcações que se posicionaram a menos de 76m de distância. Para tal, empregamos um sistema de codificação baseado em *geohash*.

A biblioteca H3 [71] é uma implementação em Python do *geohash*, desenvolvida pela Uber, que divide o mundo em uma grade hexagonal hierárquica. O H3 incorpora os conceitos de *geohashing*, combinando a flexibilidade e a eficiência das grades hexagonais, que são preferíveis devido às diversas propriedades geométricas vantajosas em comparação às grades quadradas. Na Figura 4.5, apresentamos um exemplo da aplicação da biblioteca H3.

Para a detecção de encontros neste estudo, agrupamos os *hashes* gerados para cada posição AIS usando a resolução 10 (76 m) e empregamos uma janela temporal

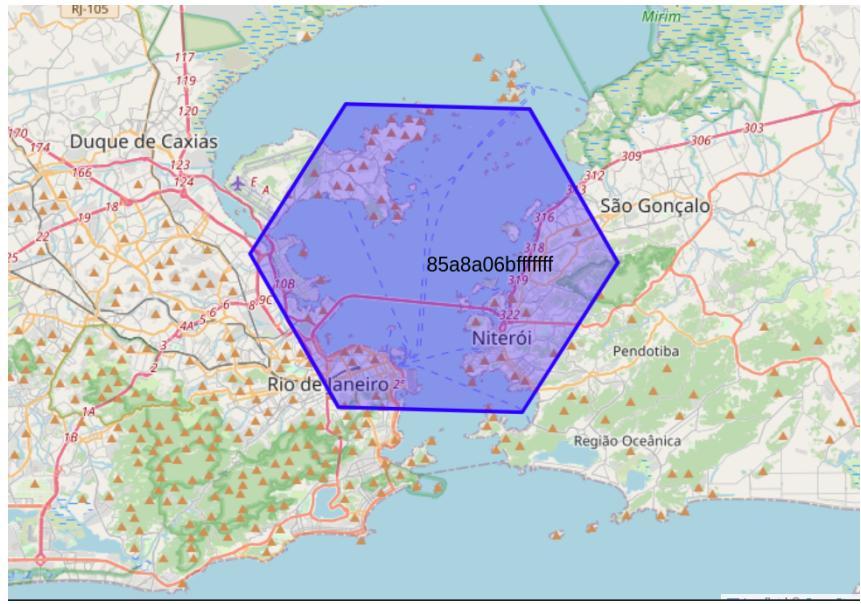


Figura 4.5: Exemplo de célula geohash de resolução 5 (8,54km de aresta) utilizando a biblioteca h3.

de 4 horas. Assim, para que uma ou mais embarcações sejam consideradas em um encontro, elas devem estar na mesma célula de *geohash* dentro da janela de 4 horas e a mais de 10 milhas náuticas da costa. A escolha da resolução 10 foi fundamentada na necessidade de uma precisão espacial que permita capturar encontros relevantes entre embarcações evitando a geração excessiva de dados. Optamos pela janela temporal de 4 horas para detectar proximidade, mesmo que ambas as embarcações possam levar até 2 horas para emitir um sinal AIS.

No Algoritmo 4, apresentamos uma versão simplificada do processo de detecção de encontros entre embarcações.

Algorithm 4: Detecção de encontros entre embarcações. Dada uma trajetória, verificar a existência de outra trajetória próxima dentro da janela de tempo.

Result: Encontros da trajetória, se houver.

```
pontos_ais = insere geohash h3 com resolução 10 para cada ponto AIS;  
lista_traj_id = []  
adiciona traj_id de trajetoria em lista_traj_id;  
foreach ponto in trajetória do  
    pontos_geohash = consultaPontosMesmoGeohash( ponto.geohash,  
        pontos_ais );  
    foreach ponto_geohash in pontos_geohash do  
        if |ponto_geohash.timestamp - ponto.timestamp| < 4 then  
            adiciona ponto_geohash.traj_id em lista_traj_id;  
            break;  
        end  
    end  
end  
retorna lista_traj_id;
```

4.2.4 Trajetórias de Pesca

Embarcações de pesca, ao se deslocarem de uma região para outra, normalmente se comportam como navios de carga, buscando seguir uma trajetória otimizada. Contudo, ao chegarem aos locais de pesca e iniciarem sua atividade principal, essas embarcações realizam manobras específicas. Para cada tipo de atividade pesqueira, existem manobras características, conforme exemplificado na Figura 4.6. Os padrões de movimento revelam a atividade pesqueira de uma embarcação, permitindo o monitoramento em áreas restritas.

A detecção de trajetórias de pesca tem sido abordada em diversos trabalhos utilizando dados de sensores AIS em modelos de aprendizado de máquina, conforme discutido na Seção 2.1. Neste trabalho, adotamos uma solução similar à proposta por [43] e [42], na qual as trajetórias são inicialmente pré-processadas. Nesse processo, extraem-se características como a média e a variância da velocidade, variância da variação angular e a área formada pelos pontos da trajetória após a aplicação do algoritmo do mínimo polígono convexo. Os dados agregados das trajetórias são, então, utilizados para treinar um modelo de *gradient boost* (GB). A escolha do modelo GB, hiperparâmetros e da estratégia de agregação dos dados da trajetória é justificada pela boa performance demonstrada por esse tipo de abordagem, conforme evidenciado em [72].

A utilização de dados agregados de trajetória nem sempre constitui uma abor-

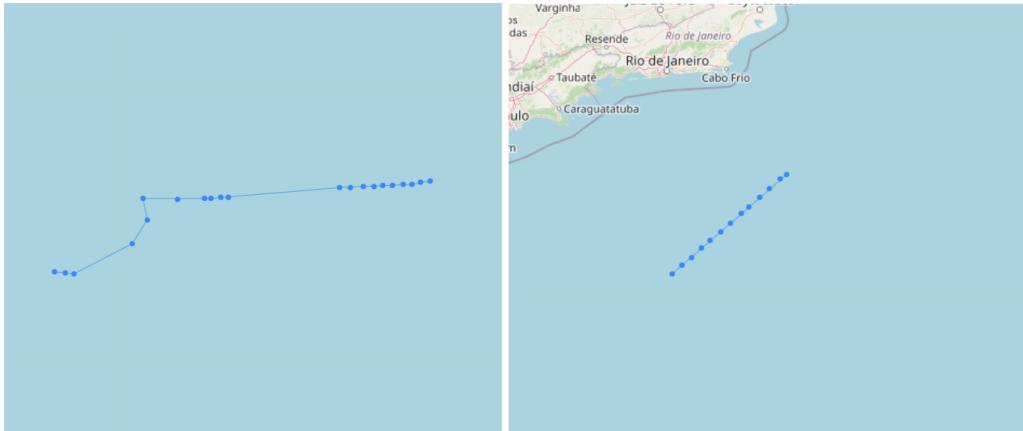


Figura 4.6: Exemplo de trajetórias de embarcações. A imagem a esquerda mostra a trajetória de uma embarcação de pesca, enquanto que a imagem a direita mostra uma trajetória de um navio de carga.

dagem óbvia nos modelos. Na Figura 4.6, observamos que embarcações de pesca tendem a apresentar maiores variações de curso. Calculando a média e a variância da velocidade dos pontos AIS ao longo das trajetórias, essas características podem ser incorporadas nos modelos. De maneira semelhante, ao calcularmos a diferença de curso entre um ponto e o próximo, podemos determinar a variância dessas diferenças ao longo de toda a trajetória. Na Equação 4.1, apresentamos o cálculo da variância utilizado, onde μ representa a média de todos os pontos, N indica o número total de pontos, e x_i corresponde ao valor no i -ésimo ponto.

$$\sigma = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (4.1)$$

Para calcular a diferença angular entre os pontos, podemos utilizar a equação (4.2).

$$\Delta\theta = \arctan 2(\sin(\theta_1 - \theta_2), \cos(\theta_1 - \theta_2)) \quad (4.2)$$

onde θ_1 e θ_2 representam as direções em cada ponto, e $\arctan 2$ é uma variação da função arco tangente que considera o sinal tanto do seno quanto do cosseno para determinar o quadrante do ângulo resultante. Outra característica relevante extraída das trajetórias é a área do polígono mínimo formada por essas trajetórias. As trajetórias de pesca tendem a formar áreas maiores devido aos seus frequentes desvios de direção, o que não ocorre com navios que seguem trajetórias retilíneas. Na figura 4.7 podemos observar a diferença entre as áreas formadas pelas trajetórias de pesca e de um navio de carga. A linha em vermelho denota a delimitação do polígono mínimo a partir dos pontos das trajetórias. A área do polígono é utilizada como uma das características que descrevem a trajetória.

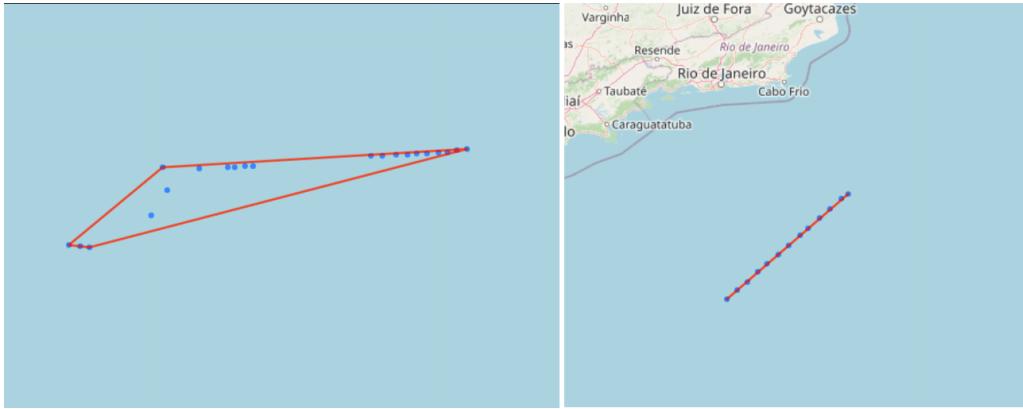


Figura 4.7: A imagem mostra as trajetórias de embarcações após aplicação do algoritmo do polígono mínimo. Em vermelho temos o polígono construído a partir dos pontos da trajetória em azul. A área do polígono é utilizada como uma das características das trajetórias no modelo de detecção.

Dimensão	Forma de Cálculo
Média da Velocidade	$\mu = \frac{1}{N} \sum_{i=1}^N (sog_i)$
Variância da Velocidade	$\sigma_{sog}^2 = \frac{1}{N} \sum_{i=1}^N (sog_i - \mu)^2$
Variância da Diferença de Curso	$\sigma_{\Delta\theta}^2 = \frac{1}{N-1} \sum_{i=1}^{N-1} (\arctan 2(\sin(\theta_{i+1} - \theta_i), \cos(\theta_{i+1} - \theta_i)) - \mu_{\Delta\theta})^2$
Área Mínima Convexa da Trajetória	Dado um conjunto de pontos $S = \{P_1, P_2, \dots, P_n\}$ do polígono convexo mínimo. A área pode ser calculada por $A = \frac{1}{2} \sum_{i=1}^n (x_i y_{i+1} - x_{i+1} y_i) $

Tabela 4.1: Tabela com os cálculos dos dados agregados das trajetórias para utilização como dimensão no modelo de detecção de trajetória de pesca.

Portanto, de forma resumida, temos as seguintes dimensões utilizadas para treinar o nosso modelo de detecção de trajetórias de pesca, conforme representado na tabela 4.1.

Após a extração das características das trajetórias apresentadas na tabela 4.1, o modelo *gradient boost* foi treinado com os mesmos dados utilizados no trabalho de [73]. Em seguida, o modelo treinado foi salvo em um arquivo *joblib* para futura utilização no *framework*. Na predição de uma trajetória, o modelo infere um valor de probabilidade entre 0 e 1, onde 1 indica 100% de probabilidade de a trajetória ser de pesca, enquanto 0 indica o contrário.

4.2.5 Loitering

O comportamento de *loitering* envolve uma embarcação que permanece por um período anormalmente prolongado na mesma área ou que realiza movimentos lentos ao redor de uma área maior sem deslocar-se uma distância significativa. Embarcações que permanecem estacionárias ou que se movem em círculos podem indicar que estão possivelmente aguardando outra embarcação ou envolvidas em atividades como pesca ilegal.

Diversos estudos sobre detecção de *loitering* geralmente a descrevem como a identificação de um objeto em uma cena durante um período que excede um determinado limiar [44, 74, 75]. No entanto, para detectar uma trajetória de *loitering*, desenvolvemos de forma inovadora um método que utiliza um modelo de LSTM (Long Short-term Memory), semelhante à detecção de trajetórias de pesca. Inicialmente, utilizando o conjunto de dados do MarineCadastre [16] que contém dados AIS da costa dos Estados Unidos, detectamos encontros entre embarcações. Para que as embarcações se encontrem, elas precisam primeiro desacelerar até parar completamente. Essa desaceleração inicial, seguida de um movimento em baixa velocidade por algumas horas em um determinado local, caracteriza as trajetórias de *loitering* que podem preceder um encontro entre embarcações. Ao detectar as trajetórias que resultaram nesses encontros, extraímos os dados necessários para treinar nosso modelo LSTM; assim, no algoritmo 5, podemos obter uma compreensão mais profunda do processo. Em seguida, criamos uma arquitetura LSTM com as seguintes dimensões: velocidade, diferença de distância, diferença de tempo, diferença angular e aceleração. Por fim, treinamos o modelo com as trajetórias extraídas dos encontros presentes no conjunto de dados [16].

Algorithm 5: Detecção de Trajetórias de Loitering

Result: Modelo LSTM treinado para detectar trajetórias de loitering

Passo 1: Ler o dataset com as trajetórias;

Dataset_Trajetorias \leftarrow carregar_dataset("trajetorias.csv");

Passo 2: Detectar trajetórias que se encontram;

Trajetorias_Encontros \leftarrow [];

foreach trajetória T_i em Dataset_Trajetorias **do**

foreach trajetória T_j em Dataset_Trajetorias **do**

if T_i e T_j se encontram **then**

 Trajetorias_Encontros.append(T_i);

 Trajetorias_Encontros.append(T_j);

end

end

end

Passo 3: Criar dataset balanceado;

Trajetorias_Sem_Encontros \leftarrow [];

foreach trajetória T_k em Dataset_Trajetorias **do**

if T_k não está em Trajetorias_Encontros **then**

 Trajetorias_Sem_Encontros.append(T_k);

end

end

Dataset_Balanceado \leftarrow balancear(Trajetorias_Encontros,

 Trajetorias_Sem_Encontros);

Passo 4: Treinar o modelo LSTM;

Modelo_LSTM \leftarrow inicializar_LSTM();

Modelo_LSTM.treinar(Dataset_Balanceado);

return Modelo_LSTM;

Na arquitetura LSTM, utilizamos as seguintes dimensões: velocidade, diferença angular, diferença de tempo, diferença de distância e aceleração. A velocidade é calculada pela relação entre a distância e o tempo; a diferença angular corresponde à variação de rumo entre o ponto atual e o ponto anterior; a distância é medida entre o ponto atual e o ponto anterior; e a aceleração é obtida pela variação da velocidade em relação ao tempo. O modelo foi configurado com os seguintes hiperparâmetros:

- **dropout:** Taxa de *dropout* aplicada à entrada da célula LSTM, definida como 0,2. Essa técnica visa reduzir o risco de *overfitting*, promovendo a regularização do modelo.
- **units:** Número de unidades da camada LSTM, configurado como 50. Esse valor representa a dimensionalidade do espaço latente interno à célula LSTM, influenciando diretamente a capacidade de memorização e aprendizado de pa-

drões temporais complexos.

- **recurrent_dropout**: Taxa de *dropout* aplicada às conexões recorrentes da célula LSTM, também definida como 0,2. Essa configuração contribui para a regularização do fluxo de memória ao longo da sequência.
- **input_shape**: A forma da entrada fornecida ao modelo corresponde a $(T, 3)$, onde T representa o comprimento máximo da trajetória (`max_trajectory_length`) e cada ponto da trajetória é representado por três atributos.
- **camada de saída**: Utiliza uma camada `Dense` com duas unidades e função de ativação `softmax`, apropriada para tarefas de classificação binária multiclasse (duas classes com codificação *one-hot*).
- **otimizador**: O modelo foi treinado com o otimizador `RMSProp`, recomendado para redes recorrentes por sua capacidade de adaptar a taxa de aprendizado com base na média móvel dos gradientes.
- **função de perda**: Utilizou-se a função de perda `categorical_crossentropy`, adequada para problemas de classificação com rótulos codificados em *one-hot*.
- **métrica de avaliação**: A acurácia (`acc`) foi utilizada como métrica principal de desempenho durante o treinamento e validação do modelo.

No algoritmo 6, apresentamos detalhes sobre o processo de treinamento da LSTM.

Algorithm 6: Treinamento de LSTM para Detecção de Loitering

Result: Modelo LSTM treinado para detectar trajetórias de loitering

Inicialização:

- Carregar dados de treinamento x e y e dados de teste $test_x$ e $test_y$.
- Padronizar sequências com `pad_sequences()` para um comprimento fixo.
- Inicializar variáveis $best_accuracy = 0.0$, $mean_accuracy = []$, $n_rounds = 0$.

Preparação dos Dados:

- Aplicar `LabelEncoder()` para transformar rótulos em formato categórico.
- Criar validação cruzada estratificada com 5 *folds* (`StratifiedKFold()`).

foreach $fold$ na validação cruzada (5 *folds*) **do**

Dividir dados de treino e validação:

- Separar dados de treinamento e validação usando os índices do $fold$.
- Padronizar sequências de X_train e X_val com `pad_sequences()`.
- Transformar rótulos em formato categórico.

Definir a Arquitetura da LSTM:

- Criar modelo sequencial.
- Adicionar uma camada LSTM com 100 unidades e *input_shape*.
- Adicionar camada `Dropout(0.2)`.
- Adicionar camada `Dense(2)` com ativação `softmax`.
- Compilar modelo com `categorical_crossentropy` e métrica de acurácia.

Treinar o Modelo:

- Criar `ModelCheckpoint()` para salvar os melhores pesos.
- Treinar o modelo por *epochs* épocas com dados de treinamento e validação.
- Carregar os melhores pesos do modelo salvos.

Avaliar o Modelo:

- Prever resultados em X_val e calcular tempo de execução.
- Calcular acurácia usando y_true e y_pred .
- Se acurácia atual > $best_accuracy$, salvar os pesos do modelo.

end

return Melhor Modelo LSTM e Estatísticas de Treinamento;

Finalmente, após o treinamento do modelo de detecção de comportamento de *loitering* este pode ser utilizado para inferir, de maneira contínua, as trajetórias que são recebidas.

4.2.6 Navegação Anômala

Desvios em rotas padrão típicas de navegação podem indicar que a embarcação está perdida, enfrentando problemas técnicos ou envolvida em atividades ilícitas. A maioria dos estudos na literatura aborda anomalias nas rotas de embarcações de forma binária, analisando as rotas mais comuns e classificando trajetórias que se desviam desses padrões como anômalas [76–79].

Neste trabalho, optamos por abranger toda a costa brasileira, em vez de monitorar uma área específica. Portanto, para detectar trajetórias anômalas, seria necessário detectar o padrão normal de cada canal de navegação ao longo da costa para utilizar essa abordagem. Então, para simplificar, não utilizaremos o conceito de anomalia como um comportamento, mas sim como uma atividade.

Algumas combinações de comportamentos descritas anteriormente, quando associadas a determinadas regras estabelecidas por especialistas, podem ser classificadas como atividades anômalas. Por exemplo, podemos classificar uma embarcação que navega com um MMSI inválido, sem identificação e que exibe um comportamento de trajetória típico de pesca, como engajada em uma atividade anômala. Como resultado, adotaremos a classificação de atividade anômala para situações que estão fora do padrão, embora menos severas, nas quais uma combinação de comportamentos e regras é considerada anômala pelos especialistas.

4.3 Definindo Regras com Base no Conhecimento do Especialista

O especialista em monitoramento marítimo possui ampla experiência em navegação, legislações e normas relativas ao domínio marítimo, bem como sobre as principais ameaças nesse contexto. Dessa forma, seu conhecimento empírico pode ser utilizado para estabelecer regras cinemáticas e de validação em situações envolvendo embarcações. Em sistemas de Consciência Situacional Marítima (MSA), é comum a utilização dessas regras para gerar alertas, visando à detecção de comportamentos ou atividades suspeitas de embarcações. Assim, desenvolveremos algumas regras baseadas no conhecimento do especialista para integrá-las ao nosso *framework*, em conjunto com os comportamentos das embarcações. Na Tabela 2.2, apresentamos as regras estabelecidas pelo especialista, com os respectivos atributos utilizados no

metamodelo, a descrição de cada atributo e a regra aplicada para determinar seu valor.

Atributo	Descrição	Regra
dist_costa	Distância da trajetória com relação a costa.	<i>Haversine distance</i> entre a trajetória e o <i>shape file</i> do mapa do Brasil.
dentro_zee	Se a trajetória está dentro da ZEE.	se dist_costa > 200 entao 0 senão 1.
dentro_mt	Se a trajetória está dentro do mar territorial.	se dist_costa > 12 entao 0 senão 1.
dentro_apa	Se a trajetória cruza uma área de proteção ambiental.	se a trajetória cruza os <i>shape files</i> das APAs brasileiras então 1 senão 0.
fora_zn_ancoragem	Se a trajetória cruza uma zona de ancoragem.	se a trajetória cruza os polígonos de zonas de ancoragem detectados pelo algoritmo então 0 senão 1.
dentro_fpsو	Se a trajetória cruza os quadrados que delimitam as zonas de operação de plataformas de petróleo disponibilizadas pela ANP.	Se cruza a área de FPSO então 1 senão 0.
bandeira_brasil	Informa se a embarcação é de bandeira brasileira.	Se a embarcação é de bandeira brasileira então 1 senão 0.
bandeira_outra	Informa se a embarcação é de outro país que não o Brasil.	Se a embarcação é de bandeira estrangeira então 1 senão 0.

Atributo	Descrição	Regra
bandeira_desconhecida	Informa se a bandeira da embarcação é desconhecida	Se a embarcação possuir bandeira desconhecida então 1 senão 0.
tipo_fishing tipo_offshore tipo_tanker tipo_tug tipo_anti_pollution tipo_cargo tipo_research tipo_buoy tipo_outro tipo_desconhecido	Informa o tipo da embarcação.	Se for o tipo da embarcação informado no atributo então 1 senão 0.
media_var_rumo	Média de variação de rumo.	A média de variação de rumo ao longo da trajetória.
media_velocidade	Média de velocidade.	A média de velocidade ao longo da trajetória. A velocidade foi calculada utilizando as diferença das posições e de tempo entre os pontos.
mmsi_valido	Informa se o mmsi é valido, se é brasileiro ou estrangeiro.	Se o mmsi for inválido o valor é 0. Caso seja válido para um país que não o Brasil o valor é 0,5. Caso seja válido para o Brasil o valor será 1.

Atributo	Descrição	Regra
arriving	Informa se o navio está indo em direção à terra, em paralelo ou contrário.	Se o navio estiver indo na direção terra o valor será 1. Caso navio esteja em direção contrária a terra o valor será -1. Caso navio esteja navegando direção paralela a terra o valore será 0.
tempo_parado	Informa o tempo que o navio ficou parado durante a trajetória.	Informa o tempo (em horas) em que navio ficou com a velocidade abaixo de 0,5 nó durante a trajetória.

Tabela 4.2: Tabela com as regras criadas pelo especialista com os respectivos atributos, descrições e regra de disparo.

Antes da utilização no metamodelo, todos os valores da Tabela 2.2 foram normalizados para o intervalo de 0 a 1. Dessa forma, as saídas obtidas das regras definidas pelos especialistas podem agora ser combinadas com as saídas dos modelos de detecção de comportamentos.

Além das soluções apresentadas neste trabalho, existem abordagens na literatura que exploram a utilização de cartas náuticas, dados batimétricos e informações geoespaciais detalhadas para enriquecer a análise de trajetórias de embarcações. Essas abordagens costumam empregar cartas no formato S-57 ou S-101 para identificar zonas de navegação restritas, áreas de fundeio, profundidades e perigos à navegação, contribuindo para uma compreensão mais precisa do contexto marítimo. No entanto, por questões de escopo, optamos por não incorporar essas fontes complementares neste estudo. Nosso foco permanece na análise comportamental de embarcações com base em dados AIS e na fusão de modelos baseados em aprendizado de máquina com regras derivadas do conhecimento especializado, mantendo o escopo metodológico consistente e centrado na detecção de atividades ilegais.

4.4 Definindo as Atividades Ilegais de Embarcações

As Autoridades Marítimas (AMs) são tipicamente responsáveis por uma variedade de atividades em suas águas jurisdicionais, incluindo o monitoramento costeiro, a proteção da vida humana no mar, a aplicação da lei, a proteção de Áreas Marinhas

Protegidas (AMPs), a proteção de infraestrutura crítica e a resposta a crimes ambientais [80]. Soluções em larga escala são necessárias devido ao grande número de atividades que requerem monitoramento, uma vez que agentes humanos não podem desempenhar essas tarefas de forma eficaz em toda a ZEE. Por exemplo, a ZEE do Brasil é a décima maior do mundo, com 3,5 milhões de quilômetros quadrados [53], e, consequentemente, o uso de sensores e aplicações automatizadas para alertar operadores humanos tornou-se essencial.

As AMs ao redor do mundo têm uma grande preocupação com atividades ilegais realizadas por embarcações em seus domínios marítimos. Alguns exemplos incluem pesca ilegal, proteção de áreas marinhas protegidas (AMPs), proteção de infraestrutura crítica, combate ao tráfico (de armas, drogas, pessoas e dinheiro) e situações de segurança na navegação. No entanto, a automatização na identificação dessas atividades representa um desafio significativo, pois exige não apenas a identificação do comportamento da embarcação, mas também a compreensão da intenção humana (como a do comandante, tripulação ou do armador) por trás do uso da embarcação.

Em [10], uma abordagem baseada em ontologia foi proposta para detectar atividades de embarcações com base em seu comportamento. O autor argumenta que a modelagem e exploração de informações incertas são fundamentais para o reconhecimento das atividades das embarcações, dada a alta variabilidade de comportamento das embarcações e a incerteza inerente às suas intenções. Como resultado, o comportamento das embarcações é parte da atividade observada. Neste trabalho, consideramos os comportamentos detectados das embarcações em conjunto com as regras criadas por especialistas para detectar atividades de embarcações.

Detectar todas essas atividades seria um escopo inviável para este trabalho, tanto pela falta de dados históricos e sensores quanto pelo amplo escopo. Portanto, focaremos em atividades para as quais existem dados acadêmicos, permitindo o treinamento de nossos modelos e a apresentação de testes que demonstrem a eficácia e escalabilidade de nosso *framework*. Assim, este trabalho classificará as atividades nas seguintes categorias:

- *Pesca ilegal*: a pesca é considerada ilegal quando uma embarcação realiza atividades de pesca em uma área onde essa atividade não é permitida ou quando a pesca é realizada por embarcação estrangeira dentro da ZEE [81];
- *Atividade suspeita*: atividade que segundo o julgamento do especialista pode estar relacionada à prática de atividades ilícitas. Contudo, faltam algumas evidências devido à incerteza em certos dados para classificar como ilegal. Baseado na experiência do especialista, essas atividades devem ser consideradas com um nível de severidade superior ao de uma atividade anômala típica. Por exemplo, encontros entre embarcações situadas a uma distância de 20 milhas

náuticas da costa podem indicar transporte de mercadorias ou transferência de armas. Outro exemplo seria embarcações não identificadas exibindo padrões de pesca entre 180 e 200 milhas náuticas, o que muitas vezes é indicativo de pesca ilegal;

- *Atividade anômala*: um especialista classifica uma atividade como anômala quando um conjunto específico de comportamentos e/ou regras se mostra incomum, mas não sugere de imediato uma atividade ilegal. Por exemplo, se uma embarcação realiza manobras atípicas em uma região que normalmente é desprovida de atividade, o especialista pode classificá-la como anômala. No entanto, nessa situação, não há evidências de conduta ilícita, apenas uma variação em relação ao padrão;
- *Atividade normal*: quando uma atividade de navegação é considerada normal para a localidade em questão.

A princípio, não há uma associação pré-estabelecida entre comportamentos e as atividades das embarcações. Essa correlação existe empiricamente na mente do especialista e será estabelecida por meio de sua avaliação das trajetórias, além de critérios previamente definidos. Nas seções subsequentes, delinearemos a estratégia de seleção de trajetórias para que o especialista possa avaliá-las, permitindo, assim, essa correlação seja desenvolvida. Neste trabalho, utilizaremos essas quatro atividades como os principais rótulos para a classificação. Na seção do estudo de caso, definiremos critérios para que o especialista classifique cada atividade.

4.5 Implementando Ensemble Models e a Estratégia de Stacking

Como explicado na seção 3.2.1, modelos de *ensemble* são uma técnica de aprendizado de máquina que combina vários modelos para aumentar a acurácia e o desempenho na resolução de problemas complexos. Dentro da técnica de ensemble, utilizaremos a estratégia de *stacking* para combinar os dados obtidos dos modelos de detecção de comportamento com as informações geradas pelas regras elaboradas pelo especialista. O *stacking* combina as previsões de diferentes modelos de aprendizado de máquina para melhorar a acurácia preditiva geral. O processo de *stacking* pode ser dividido em duas fases principais:

1. **Treinamento dos modelos base**: vários modelos, chamados de *base learners*, são treinados com o conjunto de dados de treinamento. Esses modelos podem ser de diferentes tipos, como regressão logística, árvores de decisão,

random forests, redes neurais, etc. Cada *base learner* captura um aspecto diferente dos padrões nos dados.

2. **Combinação de previsões:** após o treinamento, as previsões dos modelos base são combinadas para formar um novo conjunto de características. Esse conjunto de características é então utilizado para treinar um metamodelo (*meta-learner*), que é responsável por aprender a maneira ideal de combinar as previsões dos modelos base para produzir a previsão final. O metamodelo pode ser, por exemplo, uma regressão linear, *gradient boosting* ou qualquer outro modelo de aprendizado supervisionado.

Matematicamente, o *stacking* pode ser descrito da seguinte maneira em relação ao nosso problema:

$$F(x) = H(M_1(x), M_2(x), \dots, M_n(x), R_1(x), R_2(x), \dots, R_n(x)), \quad (4.3)$$

onde:

- $F(x)$ é a predição final do ensemble.
- $M_1(x), M_2(x), \dots, M_n(x)$ representam as previsões dos modelos de comportamento individuais.
- $R_1(x), R_2(x), \dots, R_n(x)$ são as previsão baseadas nas regras definidas por especialistas.
- H é o meta-modelo, que recebe as saídas dos modelos de base e aprende a combinação ótima das previsões para minimizar o erro final.

Seja $D = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ o conjunto de dados de treinamento, onde x_i são as características de entrada e y_i representam as saídas correspondentes. Primeiramente, cada modelo base, M_j , é treinado em D , onde $M_j(x)$ representa a previsão do modelo M_j para a entrada x_i . Em seguida, utilizamos o conjunto de dados D em R_j , onde $R_j(x)$ são as saídas produzidas pelas regras dos especialistas.

Por fim, seja $\hat{x} = (M_1(x), M_2(x), \dots, M_k(x), R_1(x), R_2(x), \dots, R_m(x))$, onde o metamodelo H é treinado em \hat{x} e $H(\hat{x})$ representa a previsão combinada final.

A técnica de *stacking* possibilita o aproveitamento das forças de diferentes modelos de aprendizado, o que pode resultar em uma melhora significativa no desempenho preditivo. A eficácia do *stacking* é particularmente notável em cenários onde há uma grande heterogeneidade entre os modelos base. Essa diversidade possibilita que o metamodelo desenvolva uma habilidade aprimorada para reconhecer a forma mais eficiente de integrar as previsões individuais dos modelos base, resultando em uma

maior acurácia na previsão final. Como metamodelo H , utilizaremos o modelo *random forest* em decorrência de sua robustez na redução de *overfitting* e na melhoria da acurácia preditiva. O *random forest*, que se configura como um *ensemble* de múltiplas árvores de decisão, combina a previsibilidade de diferentes modelos base e capta diversas nuances nos dados. Sua natureza não paramétrica permite lidar de forma eficiente com dados de alta dimensionalidade e heterogêneos, os quais são comuns na detecção de atividades ilegais de embarcações.

Com relação aos modelos de comportamento M , utilizaremos os diferentes modelos apresentados na Seção 4.2. Quanto às regras elaboradas por especialistas, representadas por R , faremos uso das que estão descritas na Seção 4.3. Por fim, mesmo ao combinar modelos de comportamento com regras de especialistas, ainda nos depararemos com o problema da falta de rótulos para as atividades. Na próxima seção, onde discutiremos o aprendizado ativo, abordaremos essa questão com mais detalhes.

4.6 Rotulando de Forma Eficiente Utilizando Aprendizado Ativo

Conforme mencionado previamente, uma de nossas premissas é que não temos rótulos para atividades ilegais. Portanto, é necessário desenvolver uma solução que otimize o trabalho do especialista na rotulagem das trajetórias. Para isso, empregaremos a técnica de aprendizado ativo, conforme apresentado na Seção 3.2.2. Na Figura 4.8, podemos observar o processo de rotulagem e treinamento do metamodelo utilizando aprendizado ativo dentro do nosso *framework*. Nesta figura, verificamos que o processo se inicia após as inferências dos modelos de primeiro nível e as saídas das regras dos especialistas, que compõem os dados de treinamento do metamodelo.

Inicialmente, não dispomos de rótulos para os dados do metamodelo, sendo necessário um *cold start* para que o modelo possa realizar uma primeira rodada de treinamento. Esse *cold start* pode ser obtido por meio da rotulagem aleatória de trajetórias realizada por um usuário. Entretanto, empregamos uma abordagem automatizada, na qual critérios são estabelecidos em um *bot* escrito em *Python*, o qual classifica rigorosamente as trajetórias sem considerar o contexto; esses critérios serão detalhados na Seção 5.3. Subsequentemente, durante a inicialização, os dados são normalizados para o intervalo de 0 a 1.

Finalizada a inicialização, o metamodelo pode ser treinado utilizando o modelo *random forest* e, em seguida, realizar inferências sobre os dados do metamodelo que não possuem rótulos. Após essa etapa, os dados que apresentem maior incerteza na predição são selecionados e apresentados ao especialista para rotulagem. Finalmente,

o modelo é atualizado, possibilitando uma nova rodada de treinamento e rotulagem, ou o processo de treinamento pode ser encerrado.

Para implementar a técnica de aprendizado ativo em nosso framework, utilizamos a biblioteca *modAL* [82]. A *modAL* é uma biblioteca de aprendizado ativo para Python que facilita a implementação de algoritmos de aprendizado ativo em modelos de aprendizado de máquina. Ela é desenvolvida sobre o Scikit-Learn, permitindo integrar de forma simples métodos de aprendizado ativo a pipelines já existentes. A *modAL* suporta diversas estratégias de amostragem, como incerteza, margem e entropia, além de permitir que o usuário crie estratégias personalizadas para selecionar os dados mais informativos e reduzir a quantidade de rotulagem necessária. Como o tempo do especialista é um recurso escasso, a rotulagem torna-se cara e demorada. Nesse contexto, a técnica de *uncertainty sampling* é amplamente aplicada, especialmente em problemas de reconhecimento de imagem, processamento de linguagem natural e detecção de anomalias. *Uncertainty sampling* é uma técnica de amostragem utilizada em aprendizado ativo para selecionar os dados mais informativos para rotulação, com o objetivo de otimizar o treinamento de um modelo utilizando o menor número possível de amostras rotuladas. Na *uncertainty sampling*, o modelo identifica e seleciona as amostras nas quais possui maior incerteza, ou seja, onde suas previsões são menos confiáveis, para que essas amostras sejam rotuladas por um especialista. Esse processo permite que o modelo melhore sua acurácia de forma mais eficiente, concentrando o esforço de rotulagem em dados que ainda não comprehende bem. Na biblioteca *modAL* é possível empregar essa técnica, a qual também utilizamos. Na próxima seção, organizaremos todas as técnicas utilizadas em nosso *framework* baseando-nos no JDL, incluindo o aprendizado ativo.

4.7 Definindo o Framework com Base no JDL

Como apresentado na seção 3.2.3, o modelo de fusão de dados do *Joint Directors of Laboratories* (JDL) serve como um referencial conceitual que orienta a integração e fusão de dados provenientes de várias fontes. Dado que nosso problema envolve o uso de sensores e a fusão de modelos de comportamento e regras criadas por especialistas, emerge uma complexidade que requer a utilização formal de um referencial para organizar nosso conhecimento. Baseamos nossa abordagem no modelo JDL em razão de sua capacidade de fornecer uma estrutura hierárquica e organizada para a fusão de dados. Esse modelo integra informações de múltiplas fontes de maneira eficiente, resultando em uma visão situacional mais abrangente e precisa. Em contextos onde decisões críticas dependem de dados coletados de diversos sensores e sistemas, essa característica torna-se essencial. Assim, nosso *framework* apresenta cinco níveis: pré-processamento, fusão em nível de objeto, consciência situacional, avaliação de

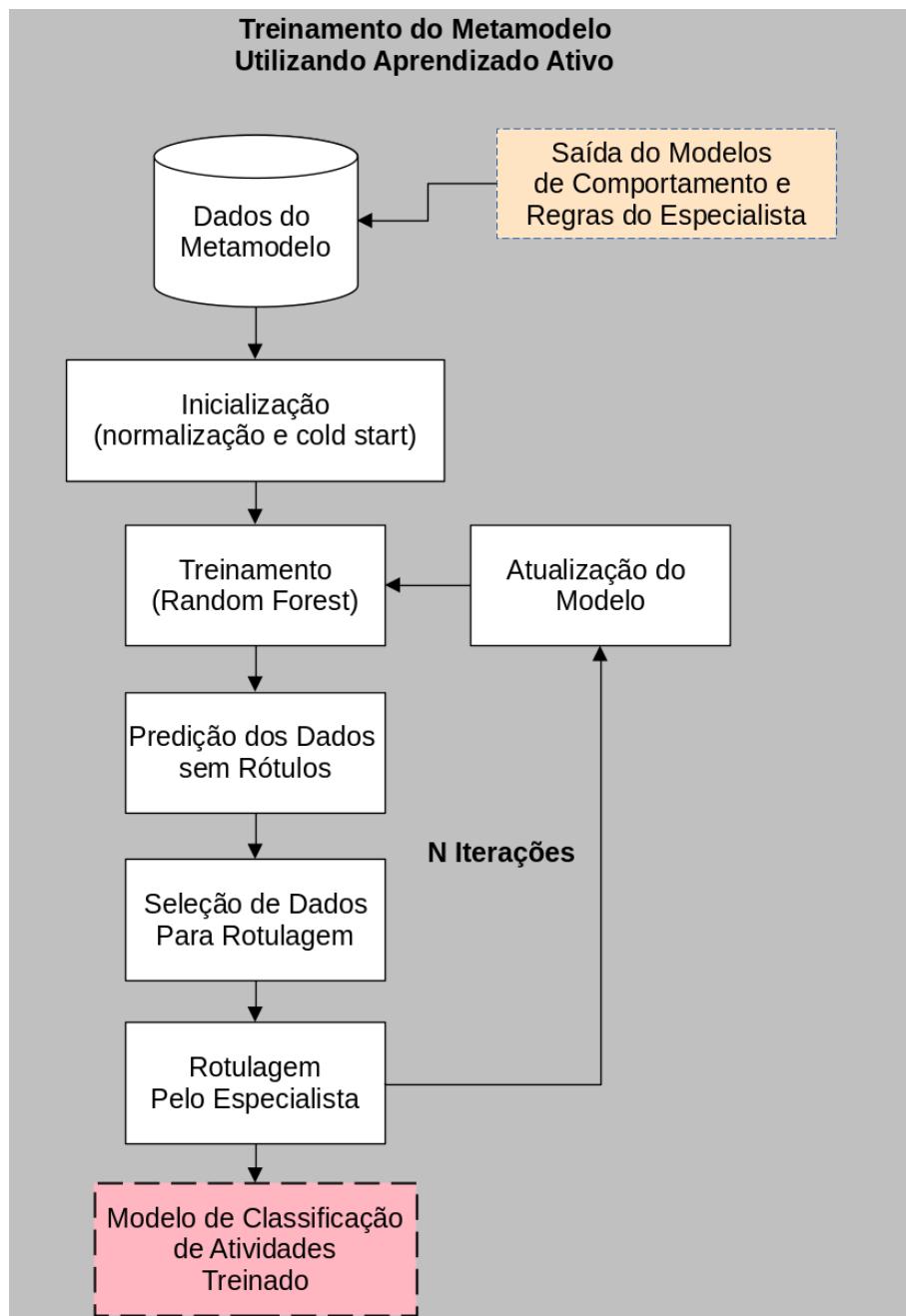


Figura 4.8: Processo de treinamento do metamodelo do framework de detecção de atividades ilegais de embarcações utilizando aprendizado ativo.

impacto e apoio à decisão. A Figura 4.9 ilustra os níveis do *framework* com base no modelo JDL. Nas seções a seguir, apresentaremos detalhes sobre cada nível do nosso *framework*.

4.7.1 Pré processamento

O pré-processamento de dados é uma etapa crucial para garantir a qualidade e integridade das análises subsequentes. Os dados podem ser obtidos de diferentes fontes e sensores, como AIS, radares, imagens de satélite, entre outros. Os dados utilizados neste estudo são provenientes de dados brutos do AIS. Cada linha do conjunto de dados contém informações de latitude, longitude, velocidade sobre o solo (SOG), rumo sobre o solo (COG), MMSI, nome da embarcação e um carimbo de data e hora. Consequentemente, as etapas primárias do pré-processamento são:

- Carregamento de dados brutos do AIS: os dados brutos do AIS são inicialmente carregados diretamente dos arquivos disponíveis, utilizando a biblioteca Pandas do Python;
- Limpeza de dados ausentes ou inválidos: após o carregamento, as linhas com dados ausentes ou incorretos são removidas. Uma trajetória deve consistir em pelo menos cinco pontos, de modo que casos incompletos são eliminados;
- Transformação dos dados AIS em *geo data frames*: os dados limpos são transformados em *geo data frames* utilizando a biblioteca Geopandas [60]. Essa transformação permite uma manipulação mais eficiente e precisa das informações geoespaciais, facilitando a análise e visualização dos dados;
- Conversão para trajetórias com o MovingPandas: por fim, os *geo data frames* são convertidos em trajetórias utilizando a biblioteca MovingPandas [61]. Essa conversão é essencial para manipular os movimentos das embarcações no tempo e no espaço, permitindo análises detalhadas de suas trajetórias e comportamentos de navegação.

Esse pré-processamento assegura que a entrada para os processos subsequentes esteja no formato de trajetórias e *geo data frames* do Geopandas. Por fim, os dados das trajetórias são armazenados como objetos serializados em um banco de dados SQLite, o que facilita o acesso. Esse formato de armazenamento é eficiente e permite consultas rápidas e flexíveis. Cada trajetória é convertida em um formato de objeto, encapsulando todos os pontos e informações associadas, que é então serializado e armazenado no banco de dados.

Framework para Classificação de Atividades Ilegais de Embarcações

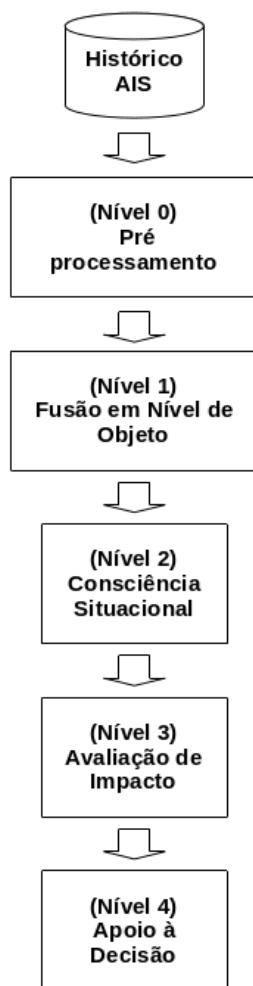


Figura 4.9: Framework para classificação de atividades ilegais de embarcações baseado no JDL.

4.7.2 Fusão à Nível de Objetos

Neste nível, recebemos as trajetórias das embarcações como entrada e aplicamos os modelos para detectar diversos comportamentos descritos na Seção 4.2, assim como as regras dos especialistas apresentadas na Seção 4.3. Caso um outro tipo de sensor, como o radar, estivesse em uso, poderíamos realizar a fusão em nível de objeto, integrando os dados AIS com os dados do radar para o mesmo contato.

Aqui, instanciamos as classes utilizadas para detectar cada comportamento e regra. Essas classes empregam as listas de trajetórias construídas durante o pré-processamento como seus parâmetros. Após o processamento das trajetórias, as saídas são armazenadas em um *dataframe*, que será utilizado posteriormente no metamodelo.

Esta é uma das partes mais intensivas em processamento do *framework*, pois as trajetórias não são apenas processadas por modelos pré-carregados, mas também por algoritmos que examinam cada ponto da trajetória, como na detecção de proximidade entre trajetórias e na identificação de *gaps* de transmissão nas trajetórias. Assim, neste nível do *framework*, as saídas de cada modelo de detecção de comportamento e as regras elaboradas pelos especialistas funcionam como fontes de dados distintas, que correlacionaremos no próximo nível.

4.7.3 Consciência Situacional

Este nível é responsável por combinar as saídas dos comportamentos de navegação com as regras de especialistas, fornecendo um contexto global para a situação da trajetória. Adicionalmente, este nível também é responsável por armazenar os dados em um banco de dados, preparar os dados para utilização no aprendizado ativo e apresentar as informações de maneira intuitiva ao usuário.

Para combinar os dados, inicialmente criamos um *dataframe* no qual cada linha contém colunas com os dados das saídas de cada modelo e regra de especialista, correlacionados com a respectiva trajetória. Em seguida, esses dados são armazenados em um banco de dados (SQLite). A utilização de um banco de dados facilita a manipulação e acelera o carregamento dos dados para a execução do modelo. Após essa etapa, o próximo nível pode utilizar esses dados como entrada para sua execução.

De maneira prática, uma classe denominada *Situational Awareness* é instanciada, e uma *Object Level Fusion* é fornecida como argumento ao construtor. Subsequentemente, o método *fuse* é invocado para agregar todos os comportamentos, centralizar todas as informações pertinentes a cada trajetória e prepará-las para a etapa subsequente.

4.7.4 Avaliação de Impacto

O nível 3 do *framework*, denominado avaliação de impacto, foca na interpretação dos dados pelo usuário, no treinamento/atualização do metamodelo e na predição de atividades. Primeiramente, este nível é responsável por instanciar a classe de aprendizado ativo e sua interface gráfica para possibilitar a interação com o especialista. Além disso, os dados do metamodelo são carregados a partir do banco de dados neste nível. A estratégia de *cold start* é implementada nesta fase, assim como a interface gráfica que utiliza aprendizado ativo para treinar o modelo *random forest* com os dados rotulados pelo usuário.

Inicialmente, o algoritmo de aprendizado ativo seleciona trajetórias com a maior probabilidade de aumentar a precisão do sistema. A estratégia de consulta utilizada para selecionar os dados mais informativos no aprendizado ativo é a *uncertainty sampling*, conforme mencionado na Seção 4.6. Essa estratégia tem como objetivo selecionar, a cada iteração, as amostras do conjunto não rotulado que o modelo classificador considera mais incertas, com base nas probabilidades preditas para cada classe. Dentre as variações disponíveis, empregamos a métrica de *entropy sampling*, que calcula a entropia da distribuição de probabilidade das classes para cada amostra, priorizando aquelas cuja incerteza preditiva é mais alta. A entropia é uma medida de dispersão que atinge seu valor máximo quando todas as classes possuem probabilidades iguais, o que reflete a indecisão do modelo. É importante destacar que essa abordagem não utiliza um *threshold* fixo para selecionar as amostras; ao contrário, ela seleciona, de forma relativa, aquelas com maior incerteza dentro do conjunto de dados disponíveis. Esse critério adaptativo permite que o processo de amostragem seja mais flexível e guiado pelas características específicas do modelo e dos dados em cada rodada de treinamento.

A Figura 4.10 ilustra um exemplo da interface do usuário. No nível de avaliação de impacto, os especialistas analisam as trajetórias por meio de uma interface. O usuário pode ter uma consciência situacional da trajetória, com informações sobre a embarcação, os comportamentos de navegação apresentados pela trajetória e se esta aciona as regras criadas pelo especialista.

O modelo é atualizado após cada rotulagem realizada pelo usuário. Após um número suficiente de classificações (avaliado na Seção 5.4), o modelo pode inferir atividades para cada trajetória recebida. Dessa forma, o especialista pode concentrar-se apenas nas trajetórias que geram previsões relevantes, como pesca ilegal e atividades suspeitas. Na próxima seção, iremos abordar esse tema.

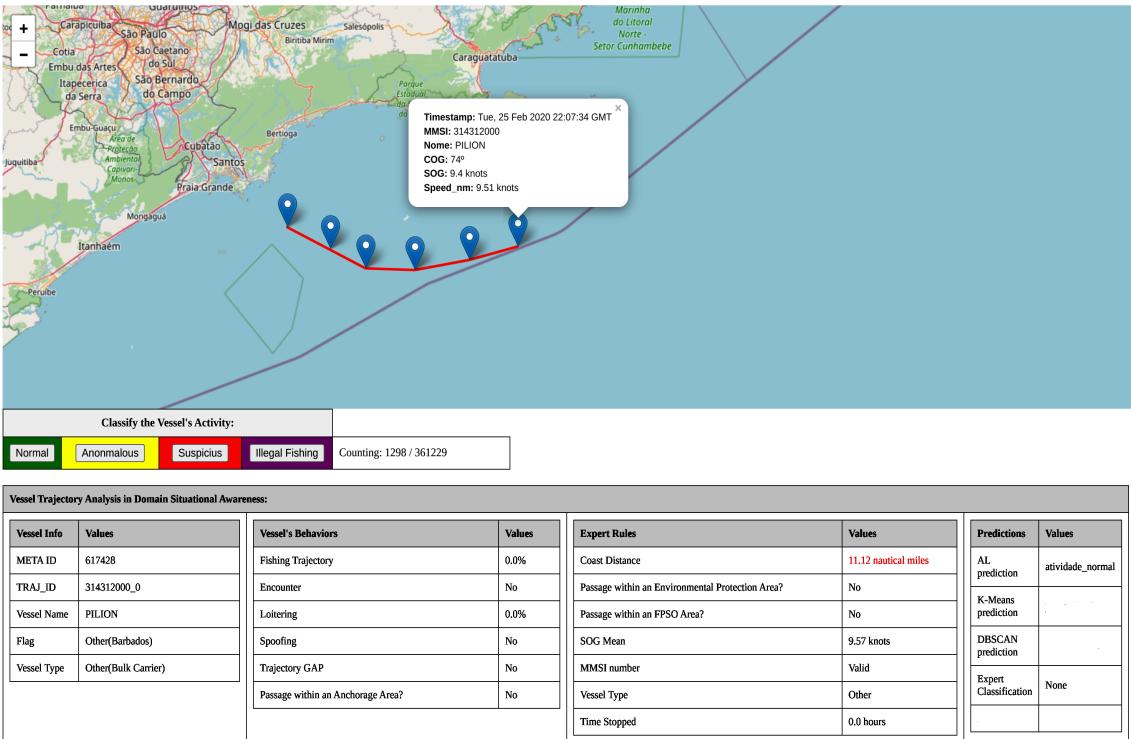


Figura 4.10: Especialistas avaliam as trajetórias por meio de uma interface no nível de avaliação de impacto. O usuário pode ter uma consciência situacional de uma trajetória com informações sobre a embarcação, comportamentos de navegação apresentados pela trajetória e se a trajetória aciona as regras criadas pelo especialista.

4.7.5 Apoio a Decisão

O nível 4 do *framework*, denominado apoio à decisão, é responsável por utilizar as informações processadas para fornecer *insights* e apoiar a tomada de decisão. As informações obtidas nos níveis anteriores são integradas para auxiliar na tomada de decisões. Inicialmente, com base nas respostas do especialista, podemos calcular o desempenho das previsões do *framework* utilizando métricas de *precision*, *recall* e F1-score. Além disso, com o metamodelo treinado, é possível gerar previsões para todas as trajetórias não rotuladas e apresentá-las em um mapa para que o especialista avalie por atividade. Isso otimizará o tempo do especialista, permitindo que ele se concentre apenas nas atividades de seu interesse.

Este nível do *framework* é capaz de gerar relatórios suplementares que aprimoram a consciência situacional na área marítima. As informações específicas incluem, por exemplo, o mapeamento das interações entre embarcações, a identificação de pontos onde navios comumente ancoram, bem como a detecção de áreas onde podem ocorrer atividades ilegais, entre outras situações. Na Figura 4.11, por exemplo, apresentamos um relatório gerado com base na previsão do nosso *framework*, que ilustra possíveis situações de pesca ilegal. Os pontos amarelos representam embarcações de pesca que estão possivelmente envolvidas em atividades de pesca em locais

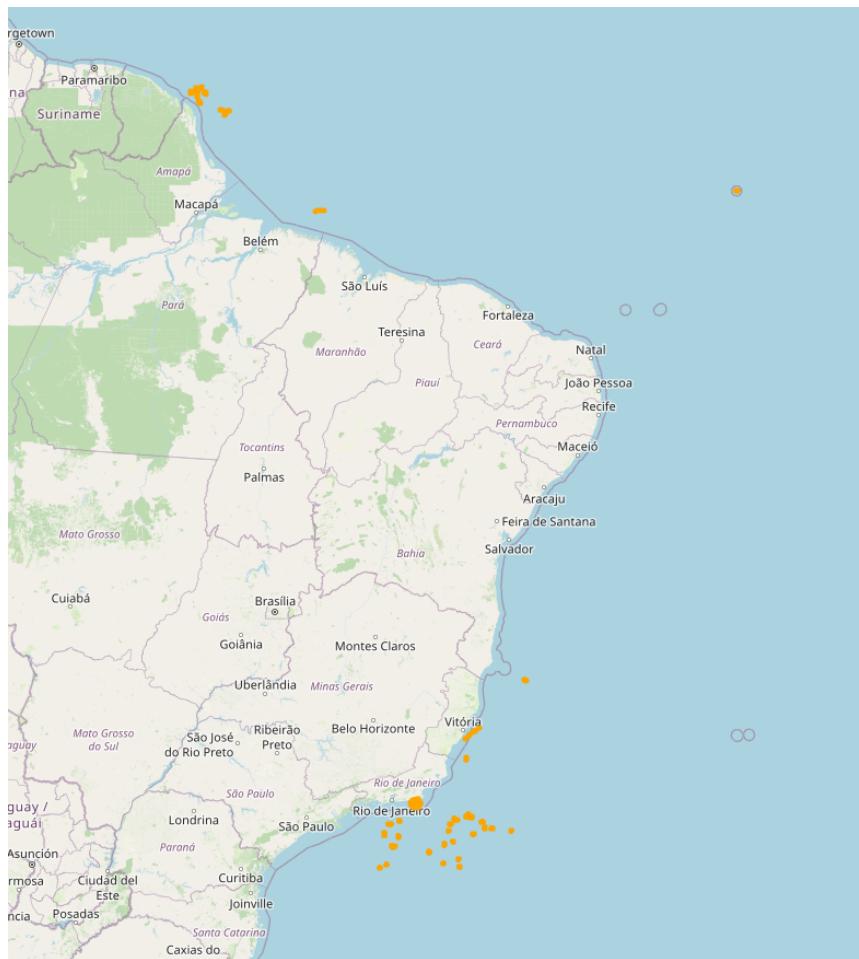


Figura 4.11: Relatório gerado pelo nosso *framework* apresentando possíveis casos de pesca ilegal na costa do Brasil. Os pontos amarelos representam embarcações de pesca possivelmente envolvidas em atividades de pesca em locais onde a pesca não é permitida.

não permitidos. Portanto, é responsabilidade do especialista realizar uma verificação detalhada de cada caso.

Capítulo 5

Avaliação Experimental

Neste capítulo, fundamentados nos conceitos e técnicas abordados nos capítulos anteriores, aplicaremos o *framework* em dados sintéticos e em dados reais da costa brasileira, visando realizar uma avaliação experimental. Para uniformizar o conhecimento dos especialistas, alcançar um nível inicial de automação e permitir a reproduzibilidade dos experimentos, adotaremos critérios para a classificação das atividades, os quais serão apresentados nas próximas seções. Esses critérios poderão ser utilizados em parte do treinamento do *framework*, poupando o esforço do especialista. Após o treinamento do *framework* com dados sintéticos da costa sudeste, utilizaremos dados AIS reais de toda costa do Brasil para os testes, inferindo todas as trajetórias não rotuladas ao longo dessa região. Por fim, apresentaremos os resultados, alguns casos reais detectados e uma discussão a respeito dos resultados obtidos.

5.1 Caso de Estudo: A Zona Econômica Exclusiva do Brasil

Para testar nosso *framework*, utilizamos um caso prático na Zona Econômica Exclusiva (ZEE) do Brasil, que abrange um total de 3,6 milhões de quilômetros quadrados. Essa área marítima inclui rotas de navegação complexas e extensas que conectam diversos portos. Para os testes, utilizamos um conjunto de dados de AIS coletados de janeiro de 2019 a dezembro de 2020. Esses dados incluem mensagens dos tipos A e B: o AIS tipo A é obrigatório para embarcações maiores (geralmente com mais de 300 toneladas brutas ou envolvidas em viagens internacionais) e transmite com maior potência e frequência; já o AIS tipo B é utilizado por embarcações menores, como barcos de pesca e de recreio, e opera com menor potência e menor taxa de transmissão, sendo, portanto, mais suscetível a perda de dados em regiões afastadas da costa.

Primeiramente, foi necessário realizar a limpeza deste conjunto de dados. Para isso, adotamos os seguintes critérios: para eliminar dados ausentes ou inválidos, consideramos que uma trajetória deve conter pelo menos cinco pontos, que a velocidade média da trajetória deve variar entre 1 e 50 nós, e que as trajetórias devem situar-se dentro da ZEE do Brasil. Com base nesses critérios, obtivemos 361.229 trajetórias de embarcações, considerando que cada trajetória pode conter um número diferente de pontos.

Devido à dificuldade em encontrar conjuntos de dados que contemplam os tipos de atividades ilegais que estamos tentando classificar, tornou-se necessário criar dados sintéticos para treinar nosso framework. Para isso, utilizamos trajetórias de pesca extraídas do conjunto de dados GFW [83] e encontros de embarcações obtidos do conjunto de dados MarineCadastre [16]. Em seguida, realizamos a translação das trajetórias para posições aleatórias dentro da ZEE brasileira, especificamente dentro da região sudeste da costa brasileira. Dessa forma, pudemos realizar o treinamento e o teste de nosso framework. Para o treinamento, utilizamos as trajetórias da região sudeste da costa brasileira, enquanto que, para o teste, empregamos as trajetórias das regiões sul, nordeste e norte da costa. Nas seções seguintes, detalharemos a construção do conjunto de dados sintético, o processo de treinamento e teste do framework e os resultados.

5.2 Dados Experimentais

Um conjunto de dados sintético é um conjunto de dados gerado artificialmente, em vez de coletado diretamente do mundo real. Esses dados podem ser criados usando algoritmos e simulações que modelam o comportamento de sistemas ou processos reais. A criação de conjuntos de dados sintéticos permite que pesquisadores e desenvolvedores testem e validem algoritmos em situações onde os dados reais são escassos, indisponíveis ou incompletos. No nosso caso, usamos um conjunto de dados sintético devido à falta de rótulos para classificar atividades ilegais de embarcações. Os dados reais disponíveis não incluem informações suficientes para identificar e rotular com precisão essas atividades, tornando necessário criar um conjunto de dados sintético que simule essas situações.

Nosso conjunto de dados sintético foi construído por meio de diversas etapas:

- *Trajetórias de pesca do GFW transladadas aleatoriamente para a costa sudeste do Brasil:* utilizamos dados de trajetórias de pesca do conjunto de dados GFW [83] e realizamos translações aleatórias dessas trajetórias para a costa sudeste do Brasil. As regiões onde essas trajetórias foram introduzidas estavam dentro da ZEE e em áreas ambientais da costa sudeste do Brasil;

- *Modificação da bandeira e do tipo da embarcação:* para aumentar a heterogeneidade dos dados, alteramos a bandeira e o tipo de algumas embarcações nas trajetórias. A bandeira da embarcação pode ser *brasil*, *outro* ou *desconhecido*. O tipo de embarcação pode ser: *fishing*, *off shore*, *tanker*, *tug*, *anti pollution*, *cargo*, *research*, *buoy*, *outro* ou *desconhecido*;
- *Trajetórias que participaram de encontros de embarcações do MarineCadastre transladadas para a costa sudeste do Brasil:* dados de encontros de embarcações foram obtidos do conjunto de dados MarineCadastre [16] e transladados para a mesma região sudeste, mantendo a mesma coerência geográfica dos dados de pesca. As regiões onde essas trajetórias foram inseridas estavam dentro da ZEE e em áreas de exploração offshore da região sudeste do Brasil;
- *Atualização do timestamp do AIS:* atualizamos os timestamps do AIS para identificar os dados sintéticos no conjunto de dados;
- *Concatenação de dados sintéticos com dados reais do AIS da costa do Brasil (2019 a 2020):* finalmente, combinamos os dados sintéticos com os dados reais do AIS coletados da costa do Brasil de 2019 a 2020, criando um conjunto de dados abrangente e representativo.

Dessa forma, treinamos nosso framework utilizando dados da região Sudeste e o testamos ao longo de toda a costa para detectar diversas situações reais. Além disso, ao elaborar um conjunto de dados sintético, conseguimos treinar e validar nosso framework em condições mais variáveis e desafiadoras, o que contribuiu para aumentar sua precisão e eficácia na detecção de atividades suspeitas.

5.3 Definindo Critérios de Avaliação de Trajetórias

Os analistas de segurança marítima possuem como uma de suas funções a classificação manual das trajetórias de embarcações, examinando cada caso para determinar se a embarcação está ou não incorrendo em algum crime. No entanto, o conhecimento dos analistas não é homogêneo; dessa forma, em alguns casos, esses analistas podem discordar entre si. Portanto, a classificação levando em consideração somente a experiência individual do especialista pode gerar divergências no treinamento do modelo. Para uniformizar essas avaliações, definiremos alguns critérios objetivos que servirão como base para a concordância entre os analistas e serão utilizados na avaliação das trajetórias. Estes critérios foram definidos com base na experiência do especialista em segurança marítima e que podem mudar ao longo do tempo, visto que o ambiente marítimo está em constante mudança. Para uniformizar a avaliação

das trajetórias de forma que possamos treinar e testar nosso framework, adotaremos os critérios da tabela 5.1 para classificar cada atividade.

5.4 Treinamento do Framework

Para o treinamento de nosso metamodelo, executamos os três primeiros níveis do nosso *framework* (Figura 4.9) utilizando o conjunto de dados sintético. Nessas etapas, realizamos o pré-processamento dos dados, a fusão de modelos e regras, além da preparação e correlação dos dados para apresentação ao especialista. O conjunto de dados sintético gerado para a região sudeste da costa do Brasil compreende aproximadamente 65.551 pontos AIS, que foram reduzidos para 2.777 trajetórias após o pré-processamento. Esses dados sintéticos foram concatenados com dados AIS reais da mesma região, totalizando 2,5 milhões de pontos e 43.744 trajetórias.

Inicialmente, os dados do nosso metamodelo são compostos pelas saídas dos modelos de comportamento e pelas regras do especialista, não possuindo rótulos que classifiquem as atividades das embarcações. Portanto, para realizarmos o treinamento por meio de aprendizado ativo, é imprescindível dispor de alguns dados rotulados para que o metamodelo seja treinado e, possa inferir novos dados. Dado que definimos critérios de avaliação, utilizaremos esses critérios para realizar o *cold start* do conjunto de dados. Primeiramente, criamos consultas SQL para cada critério a fim de verificar as linhas do metamodelo que atendiam a esses critérios. Em seguida, selecionamos aleatoriamente 10 linhas que atendiam a cada critério e inserimos um rótulo correspondente à respectiva atividade à qual o critério se referia. Dessa forma, podemos agora iniciar a execução do algoritmo de aprendizado ativo, pois já dispomos de dados rotulados para cada classe. Esse processo pode ser observado no passo 1 da figura 5.1.

Em seguida, para avaliar as trajetórias por meio do aprendizado ativo, criamos uma classe em Python com os critérios estabelecidos para cada tipo de atividade, a qual denominamos de “bot de critérios”. O aprendizado ativo, com base na incerteza dos dados seleciona a amostra com a maior probabilidade de aumentar a performance do modelo. Após a classificação da amostra selecionada, o modelo, que no nosso caso é um *random forest*, é então atualizado. Para cada solicitação de classificação realizada pelo aprendizado ativo, fazemos uma chamada ao bot de critérios para obter uma resposta com base nos critérios que definimos. Empregamos essa estratégia para 380 trajetórias por meio do aprendizado ativo, contabilizando um total de 600 trajetórias, somando as 220 que foram utilizadas no *cold start* e 380 no aprendizado ativo. O desenho dessa estratégia para o treinamento do framework pode ser observado no passo 2 da figura 5.1.

A utilização exclusiva do bot de critérios para classificar as atividades durante o

Processo de Treinamento do Framework

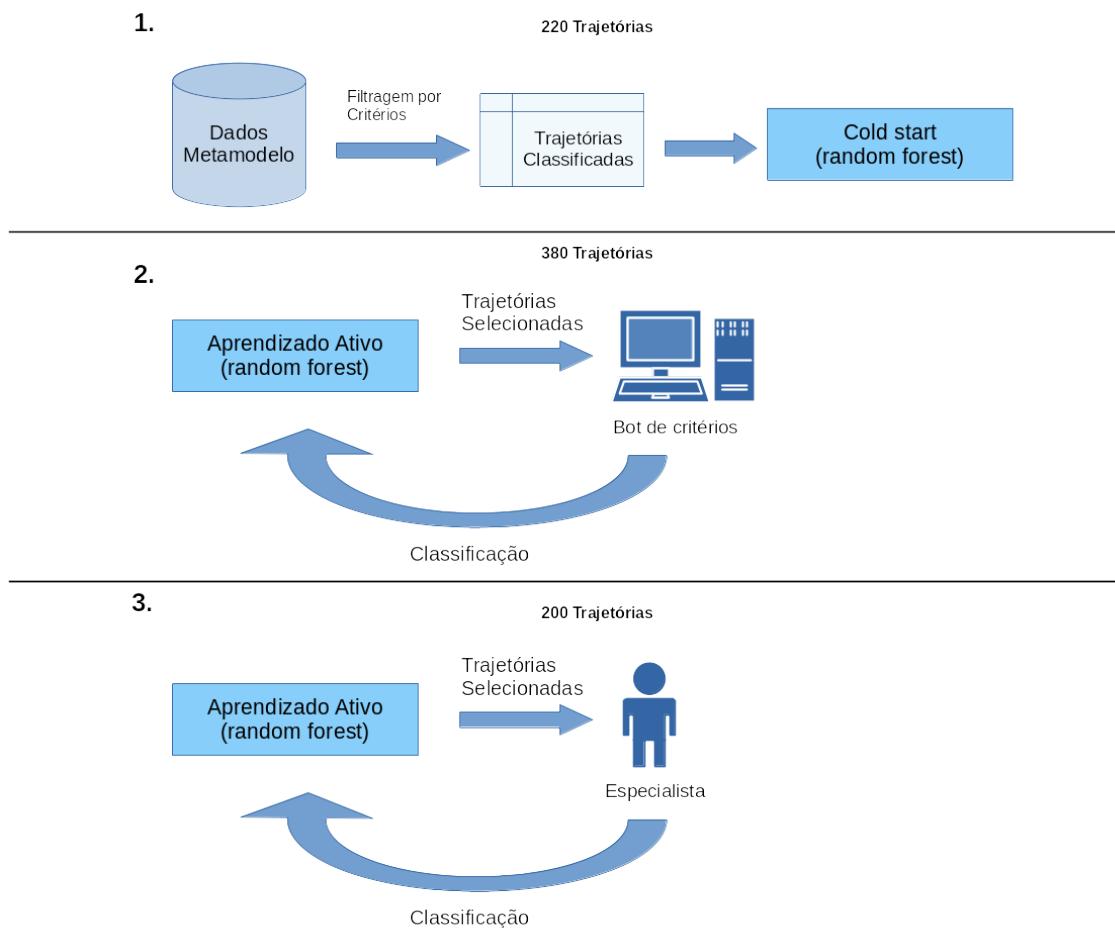


Figura 5.1: A imagem representa o processo de treinamento do *framework*. No passo 1, o *cold start* é realizado utilizando consultas SQL seguindo os critérios pré definidos. No passo 2, o algoritmo de aprendizado ativo é executado e as respostas são respondidas pelo bot de critérios. No passo 3, o especialista avalia mais 200 trajetórias de forma a capturar outras características não detectadas pelos critérios.

treinamento não produziu resultados satisfatórios. Essa situação pode ser explicada pelo fato de que, embora os critérios para classificar as trajetórias sejam bem definidos, eles não conseguem abranger todas as situações. Por exemplo, um dos critérios para classificar uma atividade como pesca ilegal seria quando uma embarcação de pesca apresentar comportamento de pesca dentro de uma APA. No entanto, alguns portos estão localizados dentro de APAs, e determinados comportamentos que ocorrem antes da atracação podem ser confundidos com atividade de pesca, resultando em falsos positivos. Esse tipo de caso pode ser detectado visualmente pelo especialista, mas causaria um falso positivo se apenas os critérios fossem levados em conta. Outro fator a ser destacado é que estamos treinando um metamodelo, ou seja, os dados que serão treinados são saídas de outros modelos. Esses modelos podem conter erros em suas previsões e estaríamos assumindo que esses modelos não erram. O operador, ao classificar a trajetória manualmente, pode perceber que um comportamento específico não é condizente e, portanto, optar por ignorá-lo. Portanto, ao treinar o metamodelo também com as respostas do operador, o modelo aprenderá a lidar com os erros dos modelos de primeiro nível e preencher os critérios ausentes. Esses erros podem ser “compensados” por outras dimensões do metamodelo, como por exemplo, distância da costa, tempo parado, velocidade média e etc. Assim, após a avaliação automatizada de 600 trajetórias, realizamos a avaliação manual de mais 200 trajetórias para que o metamodelo se ajustasse adequadamente aos possíveis erros dos modelos de primeiro nível. Esse processo é demonstrado no passo 3 da figura 5.1, enquanto os dados quantitativos mencionados, juntamente com as respectivas estratégias, estão relacionados na tabela 5.2.

No caso do metamodelo, como o número de linhas do conjunto de dados de treinamento corresponde ao número de trajetórias avaliadas pelo especialista e pelo nosso bot de critérios, o treinamento do metamodelo leva apenas alguns segundos. A maior parte do tempo de processamento foi despendida no pré-processamento, na construção das fontes de comportamentos e no treinamento dos comportamentos. Além disso, para identificar os parâmetros ótimos dos modelos nos comportamentos, foi necessário executar treinamentos e testes para cada parâmetro avaliado. Em nossa abordagem, pré-treinamos os comportamentos e os salvamos em formato `joblib`. Durante a execução do *framework*, simplesmente carregamos os modelos dos arquivos `joblib` e os aplicamos às trajetórias pré-processadas. Após treinar o *framework*, realizamos o processo de teste, que será detalhado na próxima seção.

5.5 Teste do Framework

Após o treinamento do *framework*, todas as trajetórias não utilizadas no conjunto de dados de treinamento puderam ser inferidas, incluindo as trajetórias das regiões

Processo de Teste do Framework

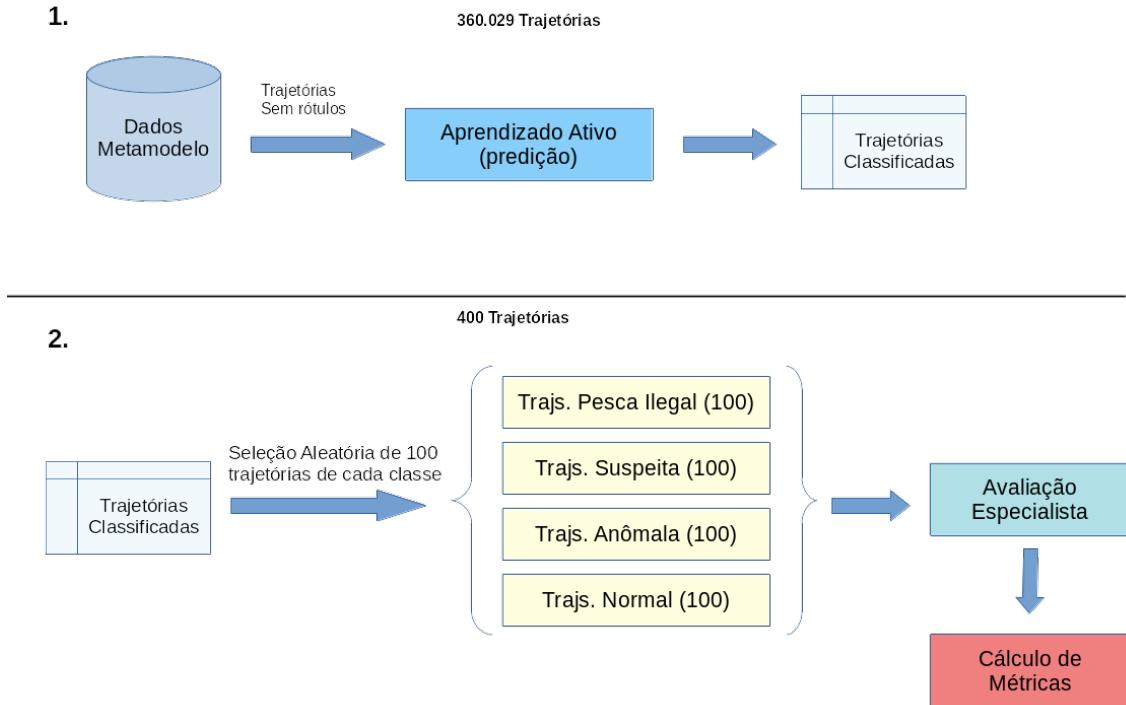


Figura 5.2: A imagem apresenta o processo de teste do *framework* utilizando dados AIS da costa brasileira. No passo 1, todas as trajetórias sem rótulos são inferidas pelo metamodelo (*random forest*) treinado. No passo 2, são selecionadas 100 trajetórias de cada classe inferida, apresentadas ao especialista para classificação e as métricas são calculadas com base nas respostas do especialista e a predição dada pelo metamodelo.

Sul, Nordeste e Norte. Para isso, essas trajetórias precisaram inicialmente passar pelos níveis 0 a 2 do *framework*, conforme ilustrado na Figura 4.9. No passo 1 da figura 5.2 é possível observar o metamodelo inferindo as atividades das trajetórias que ainda não possuem rótulos. Em seguida, para o nosso teste de validação, foram selecionadas 100 trajetórias de cada classe (normal, pesca ilegal, atividade suspeita e atividade anômala) que foram inferidas pelo *framework* e apresentadas ao especialista.

Durante o processo de teste, o especialista avaliou um total de 400 trajetórias, analisando as saídas dos comportamentos e as regras de cada trajetória, e, por fim, atribuindo suas classificações. Na Figura 5.2, no passo 2, esse processo de avaliação pode ser observado. Assim, com base nas inferências das atividades do metamodelo e nas respostas do especialista, pudemos avaliar o desempenho de nosso framework.

Para compreender melhor as previsões do nosso metamodelo, empregamos a ferramenta SHAP (*SHapley Additive exPlanations*). O SHAP é uma ferramenta de interpretação de modelos de aprendizado de máquina, fundamentada nos valores

de Shapley, que têm suas raízes na teoria dos jogos [84]. Os valores de Shapley proporcionam uma forma justa de atribuir a contribuição de cada característica à previsão de um modelo, considerando todas as combinações possíveis de características. O SHAP permite tanto a interpretação global, fornecendo uma visão geral da importância das características em todo o conjunto de dados, quanto a interpretação local, explicando a previsão de um único ponto de dados. Os valores de Shapley nos auxiliam a compreender quanto cada característica contribui para a previsão de um modelo, assegurando consistência e uma interpretação intuitiva dos resultados. Na próxima seção, apresentaremos e analisaremos os resultados com maior detalhamento.

5.6 Resultados

Conforme mencionado anteriormente, os resultados obtidos nesta análise foram fruto de uma avaliação manual de cada uma das 400 trajetórias inferidas pelo framework no processo de teste. Selecioneamos 100 trajetórias aleatórias para cada classe de atividade inferida. Para mensurar o desempenho do framework, utilizamos as métricas de *precision*, *recall*, *f1-score* e *support*. Cada métrica pode ser definida da seguinte maneira:

- **Precision:** é a proporção de verdadeiros positivos entre as instâncias classificadas como positivas. Mede a precisão das previsões positivas do modelo;
- **Recall:** é a proporção de verdadeiros positivos entre todas as instâncias que são realmente positivas. Mede a capacidade do modelo de encontrar todas as instâncias positivas;
- **F1-score:** é a média harmônica de *precision* e *recall*, fornecendo uma métrica única que equilibra ambas as considerações. É especialmente útil quando há um desequilíbrio entre as classes; e
- **Support:** refere-se ao número de ocorrências reais para cada classe no conjunto de dados. É uma medida que indica quantas instâncias estão presentes para cada classe.

Além disso, foram calculadas as médias **macro** e **weighted**:

- **Macro avg:** calcula a média das métricas para todas as classes, tratando cada classe igualmente, independentemente do número de instâncias; e
- **Weighted avg:** calcula a média das métricas para todas as classes, ponderada pelo número de instâncias em cada classe.

Na Tabela 5.3 podemos verificar o desempenho do *framework* nas quatro classes, destacando a maior precisão na detecção de atividades normais. A classe “pesca ilegal” alcançou uma precisão de 97%, *recall* de 100% e F1-score de 98%. Isso indica que o modelo é altamente eficaz na detecção de casos de pesca ilegal (com *recall* de 100%), embora ainda haja uma pequena taxa de falsos positivos, dada a precisão ligeiramente inferior a 1.

Para atividades classificadas como “suspeita”, o modelo alcançou uma precisão de 90% e *recall* de 99%, com F1-score de 94%. Esse resultado indica um bom equilíbrio, especialmente em termos de *recall*, o que sugere que o modelo está identificando quase todas as atividades suspeitas. A ligeira redução na precisão implica que alguns casos podem estar sendo classificados como suspeitos, embora não o devam ser.

Na classe “anômala”, o modelo obteve uma precisão de 95% e um *recall* de 100%, com um F1-score de 97%. Esses resultados indicam um bom desempenho na identificação de atividades atípicas, caracterizado por uma baixa taxa de falsos positivos e pela ausência de falsos negativos.

Por fim, a classe “normal” apresentou uma precisão de 100%, mas um *recall* de 85%, resultando em F1-score de 92%. Isso sugere que o modelo é bastante conservador ao classificar uma atividade como normal, o que pode ser positivo do ponto de vista da segurança, pois reduz o risco de atividades ilegais passarem despercebidas. No entanto, o *recall* indica que algumas atividades normais estão sendo classificadas incorretamente como suspeitas ou anômalas.

Também calculamos as médias *macro* e *ponderada*. Conforme apresentado na Tabela 5.3, mesmo com o suporte ligeiramente diferente em algumas classes, as médias ficaram praticamente em 96%.

Tabela 5.3: Métricas de desempenho do metamodelo: precisão, *recall* e F1-score para cada classe de atividade.

Atividade	Precisão	Recall	F1-score	Suporte
Pesca Ilegal	0,97	1,00	0,98	97
Suspeita	0,90	0,99	0,94	91
Normal	1,00	0,85	0,92	117
Anômala	0,95	1,00	0,97	95
Acurácia			0,95	400
Média macro	0,96	0,96	0,96	400
Média ponderada	0,96	0,96	0,96	400

Na Figura 5.3, analisamos a matriz de confusão das classes, a qual fornece infor-

mações sobre o desempenho do modelo em cada classe ao comparar as previsões do aprendizado ativo com os rótulos fornecidos pelo especialista.

Para a classe “pesca ilegal”, o modelo classificou corretamente 97 casos, mas cometeu três erros ao rotular atividades que deveriam ser classificadas como “normal” como “pesca ilegal”. Isso sugere que o modelo possui uma tendência conservadora ou rigorosa, identificando algumas atividades normais como ilegais. Esse tipo de falso positivo é prejudicial para a autoridade marítima, pois pode levar a ações desnecessárias e no desperdício de recursos.

Na classe “normal”, o modelo classificou corretamente 100 dos 117 casos. Os erros incluem os três já mencionados anteriormente (atividades normais classificadas como “pesca ilegal”), além de 10 atividades normais classificadas como “suspeita” e quatro como “anômala”. Esses erros indicam que o modelo adota uma abordagem cautelosa ao classificar atividades como “normal”, ou seja, quando o modelo rotula uma atividade como normal, há uma alta probabilidade de acerto.

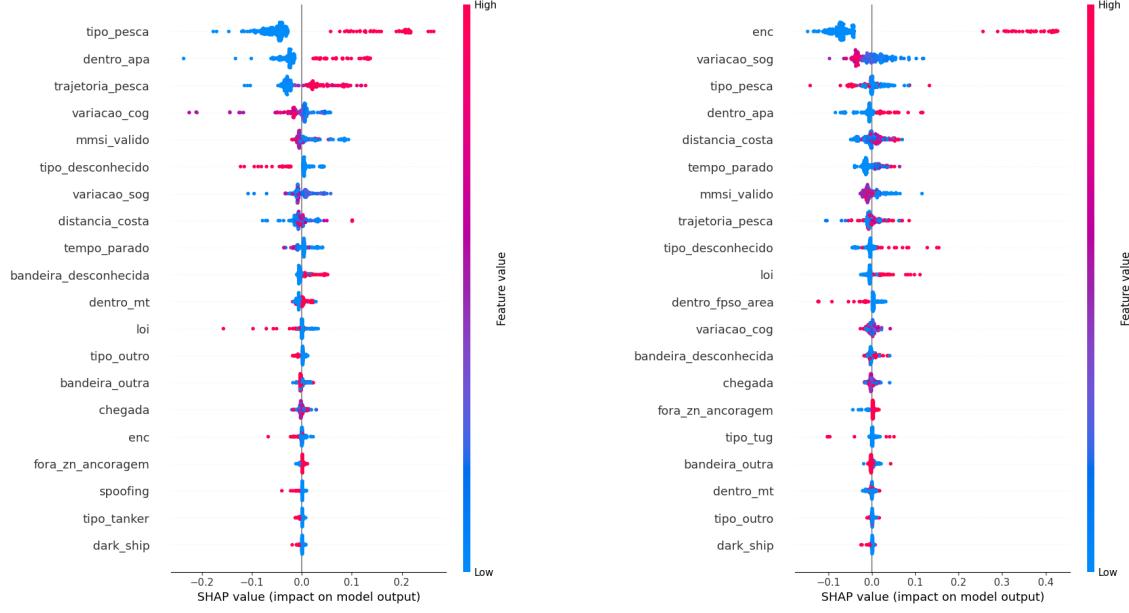
Para a classe “suspeita”, o modelo alcançou 90 classificações corretas; no entanto, cometeu 10 erros ao classificar esses casos incorretamente como “normal”. Esses falsos negativos indicam que algumas atividades suspeitas podem não estar sendo identificadas adequadamente, representando uma área de melhoria significativa para o modelo.

Por fim, na classe “anômala”, o modelo classificou corretamente 95 casos, mas cometeu cinco erros: quatro atividades anômalas foram erroneamente classificadas como “normais” e uma como “suspeita”. Esses erros sugerem que, em algumas ocasiões, as atividades anômalas estão sendo subestimadas como normais, o que pode levar à falta de atenção adequada a comportamentos fora do padrão.

		Matriz de Confusão: AL vs. Humano			
		pesca ilegal	normal	suspeita	anômala
Rótulos Verdadeiros (humano)	pesca ilegal	97	0	0	0
	normal	3	100	10	4
	suspeita	0	0	90	1
	anômala	0	0	0	95

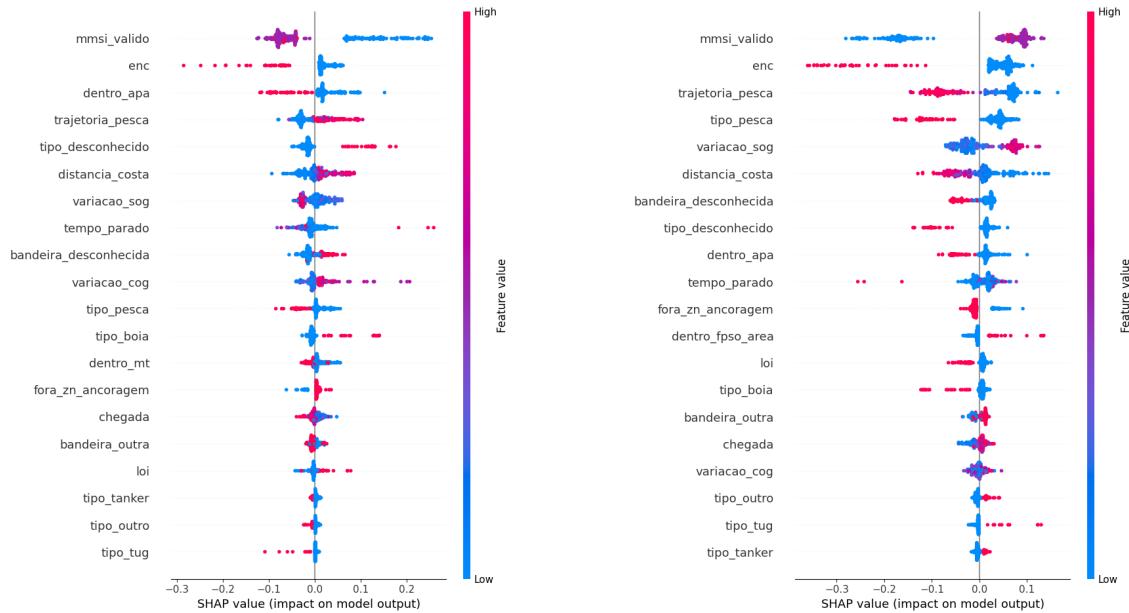
Figura 5.3: Matriz de confusão das previsões do metamodelo.

Na Figura 5.4, são apresentados os gráficos SHAP correspondentes às quatro atividades, nos quais o eixo y representa os atributos utilizados no modelo, enquanto o eixo x indica a extensão do impacto de cada atributo na previsão. Valores SHAP positivos (em vermelho) indicam que a variável contribui de forma positiva para a previsão da classe, enquanto valores negativos (em azul) indicam contribuição negativa para a classe. O valor absoluto do SHAP expressa a magnitude da contribuição.



(a) SHAP para a classe pesca ilegal.

(b) SHAP para a atividade suspeita.



(c) SHAP para a atividade anômala.

(d) SHAP para a atividade normal.

Figura 5.4: A figura mostra os gráficos SHAP para as previsões do metamodelo. O SHAP exibe as contribuições de cada dimensão para a predição da classe.

Na Figura 5.4a, é possível observar o comportamento dos valores das dimensões durante a inferência da classe de pesca ilegal em nosso metamodelo. Esse gráfico demonstra que as dimensões mais influentes são a *tipo_pesca* que indica se a embarcação se dedica à pesca, a *dentro_apo* que indica se a embarcação navegou dentro de uma APA e a *trajetoria_pesca*, que corresponde à saída do modelo responsável pela detecção do comportamento de pesca.

Nesse contexto, quanto mais elevados forem os valores (indicado em vermelho) de *tipo_pesca*, *dentro_apo* e *trajetoria_pesca*, maior será o impacto na inferência de pesca ilegal pelo modelo. Uma vez que empregamos critérios e trajetórias sintéticas, as dimensões com maior impacto se evidenciam claramente, em virtude do padrão bem definido estabelecido nos critérios e nos dados sintéticos, o que pode ter facilitado a identificação desses casos mais simples pelo modelo.

Com relação à Figura 5.4b, apresentamos os valores das dimensões quando o metamodelo infere a classe de atividade suspeita. Nesse caso, a dimensão mais impactante é a *enc* (encontros). Especificamente, a ocorrência de um encontro é altamente indicativa de uma atividade suspeita, sendo claramente evidenciada na figura 5.4b. Entretanto, observamos que a dimensão *dentro_fpsos_area* contribui negativamente para a predição da classe, o que se alinha aos critérios que adotamos.

Nas atividades suspeitas relacionadas à pesca, podemos destacar outras dimensões que se mostraram bem definidas em relação à sua distribuição no gráfico, contribuindo positivamente para a predição de atividades suspeitas, como é o caso das dimensões *dentro_apo*, *tipo_desconhecido*, *mmsi_valido* e *loi*. Observa-se que a dimensão *trajetoria_pesca* não se apresenta de forma clara, o que pode ser atribuído a possíveis erros na predição do modelo de primeiro nível. Por exemplo, navios que não eram destinados à pesca foram erroneamente classificados como apresentando trajetórias associadas a atividades pesqueiras.

Na Figura 5.4c, são apresentados os casos de atividades classificadas como anômalas. Nestes casos, observamos, a partir das cinco dimensões mais impactantes, que, para ser classificada como uma atividade anômala, o MMSI (*mmsi_valido* = 0) deve ser inválido, podendo haver uma trajetória de pesca (*trajetoria_pesca* próximo de 1), não deve haver encontros (*enc* = 0) e (*tipo_desconhecido* = 1). É importante destacar que não é necessário que todas as condições sejam atendidas, mas sim que exista uma combinação que satisfaça os critérios que definimos. Por exemplo, o critério referente a embarcações sem identificação que apresentem comportamento de pesca dentro da ZEE é considerado uma atividade anômala. Podemos observar na Figura 5.4c que esse critério é atendido pelas dimensões *mmsi_valido*, *trajetoria_pesca* e *distancia_costa*.

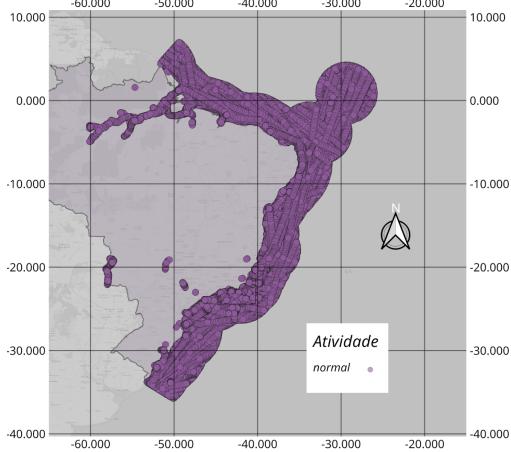
Por fim, na Figura 5.4d, podemos destacar as dimensões mais impactantes para a classificação de uma atividade considerada normal. As dimensões *mmsi_valido* superiores a zero, *enc* = 0, *trajetoria_pesca* próximas de zero e o tipo de embarcação diferente de pesca (*tipo_pesca* = 0) influenciam o metamodelo na classificação da atividade como normal. Portanto, conforme demonstrado pelos gráficos SHAP, o modelo foi capaz de capturar com precisão as características de cada atividade, as quais estavam inicialmente restritas ao conhecimento do especialista.

Em um cenário real no qual aplicamos o framework a dados reais de AIS, identi-

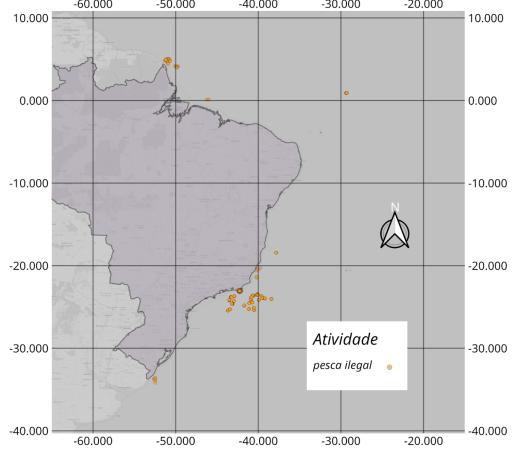
ficamos alguns potenciais casos de atividade de pesca ilegal. Ademais, em situações de atividades suspeitas, observamos encontros reais entre embarcações na ZEE que poderiam estar associados a atividades de tráfico ou contrabando. Adicionalmente, identificamos casos de embarcações pesqueiras operando em áreas de proteção ambiental marinha, possivelmente envolvidas em atividades de pesca.

No caso de atividades anômalas, é necessário analisar padrões recorrentes de anomalias e investigar suas causas. Por exemplo, no que diz respeito a um MMSI inválido, muitas embarcações menores utilizam transponders AIS sem registro na autoridade competente e atribuem qualquer número ao MMSI. Isso poderia ser interpretado como uma pequena embarcação se afastando da costa para pescar, por exemplo. No pior cenário, essa mesma embarcação pode optar por não se identificar devido à prática de atividades ilegais. Portanto, é imperativo estudar esses padrões de anomalias e incorporar outras formas de detecção de comportamentos no framework para identificar essas situações.

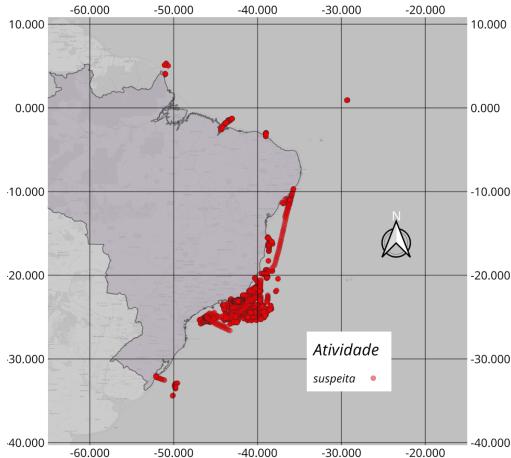
Na Figura 5.5, podemos observar todas as trajetórias inferidas pelo framework e plotadas no mapa. Na Figura 5.5a, são apresentadas as trajetórias consideradas como atividades normais; na Figura 5.5b, as trajetórias classificadas como pesca ilegal; na Figura 5.5c, as atividades consideradas suspeitas; e na Figura 5.5d, as atividades inferidas como anômalas. Na Figura 5.5a, notamos um grande número de trajetórias, representando cerca de 98% do total, ocupando quase todo o espaço da ZEE. Além disso, constatamos a presença de trajetórias em rios, que não estão incluídas em nosso estudo.



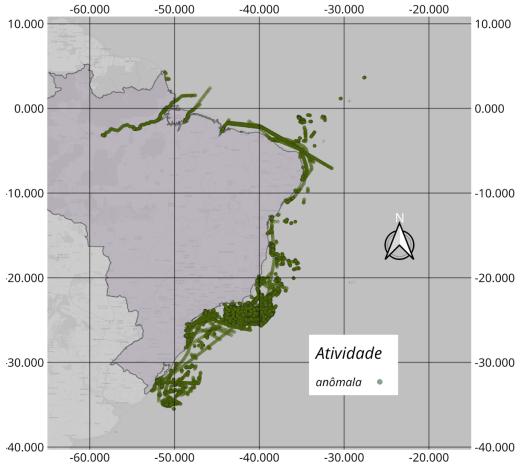
(a) Trajetórias identificadas como atividades normais pelo framework.



(b) Trajetórias identificadas como atividades de pesca ilegal pelo framework.



(c) Trajetórias identificadas como atividades suspeitas pelo framework.



(d) Trajetórias identificadas como atividades anômalas pelo framework.

Figura 5.5: Todas as trajetórias inferidas pelo framework representadas no mapa.

Em amarelo, na Figura 5.5b, observamos que as trajetórias consideradas como pesca ilegal estão predominantemente concentradas na região sudeste. De fato, todas essas trajetórias são sintéticas. No entanto, o framework evidenciou algumas atividades de pesca ilegal nas regiões norte, nordeste e sul, as quais poderemos analisar pontualmente mais adiante. Em verde, na Figura 5.5d, diversas trajetórias em rios foram consideradas anômalas, uma vez que não faziam parte do nosso estudo e não foram utilizadas no treinamento. Além disso, observam-se trajetórias anômalas ao longo de toda a costa, principalmente nas regiões sudeste e sul.

Em vermelho, na Figura 5.5c, observamos que a maioria dessas trajetórias está concentrada na região sudeste, que é precisamente onde inserimos as trajetórias sintéticas de encontros entre embarcações. Entretanto, o framework também iden-

tificou outras trajetórias do conjunto de dados real como potencialmente suspeitas, conforme evidenciado em diferentes regiões costeiras.

Durante as avaliações manuais dos testes, identificamos alguns cenários do mundo real que merecem destaque como exemplos. Na Figura 5.6, temos uma trajetória que nosso framework classificou como anômala. Na imagem, é possível observar as saídas do modelo e as regras de especialistas, onde a embarcação apresenta um MMSI inválido, bandeira desconhecida e é classificada como tipo pesca. Além disso, mostrou uma trajetória de pesca caracterizada por um padrão sinuoso, com velocidade média de 0,96 nós. Adicionalmente, a embarcação está localizada a apenas 2,86 milhas náuticas da costa. Em um cenário mais provável, poderíamos supor que esta seria uma embarcação de pesca local de menor porte que não possui registro. No entanto, para confirmar essa hipótese, precisaríamos de outro tipo de comportamento ou regra adicional. Assim, neste caso, a atividade foi identificada como anômala, uma vez que não se trata de uma atividade normal e não apresenta indícios de maior severidade.

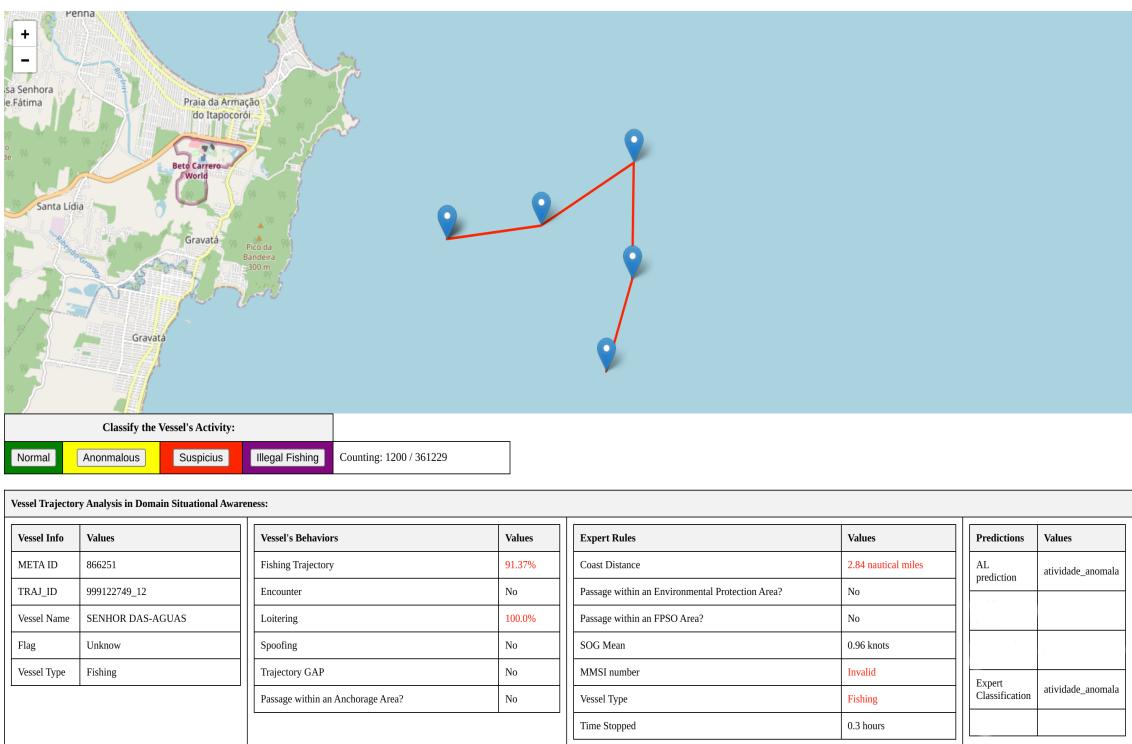


Figura 5.6: A imagem apresenta um exemplo real de uma atividade anômala identificada pelo framework. Uma embarcação como MMSI inválido, informando que é uma embarcação de pesca, apresentando trajetória de pesca e bandeira desconhecida.

Na Figura 5.7, apresentamos um exemplo de atividade suspeita classificada pelo nosso framework, no qual duas embarcações se encontram a aproximadamente 62 milhas náuticas da costa. Encontros entre embarcações não são ilegais, mas podem ser indicativos de atividades ilícitas. Para minimizar falsos positivos na detecção de

atividades, é fundamental criar modelos que identifiquem comportamentos, permitindo uma melhor compreensão da situação. Por exemplo, plataformas de petróleo necessitam de embarcações de apoio para o transporte de óleo, pessoal e materiais. Consequentemente, encontros entre embarcações nessas regiões são comuns. A área onde o encontro ocorreu, conforme ilustrado na Figura 5.7, não apresenta áreas de exploração de petróleo, representadas pelos quadrados azuis. Dessa forma, a embarcação tinha um outro motivo para o encontro, além da atividade offshore. Portanto, o especialista deve dedicar atenção especial a esses casos.

Observando o gráfico SHAP da classe de atividade suspeita (Figura 5.4b), verificamos que, quando um encontro entre embarcações é detectado dentro de uma área de FPSO, ou quando o navio é do tipo *tug* (rebocador) ou de pesca, essas dimensões têm uma contribuição negativa para a previsão da classe como atividade suspeita. Essas regras foram implicitamente incorporadas pelo especialista durante o treinamento utilizando active learning. Essa é uma vantagem significativa do framework, que transforma o conhecimento tácito do especialista em modelos treinados.

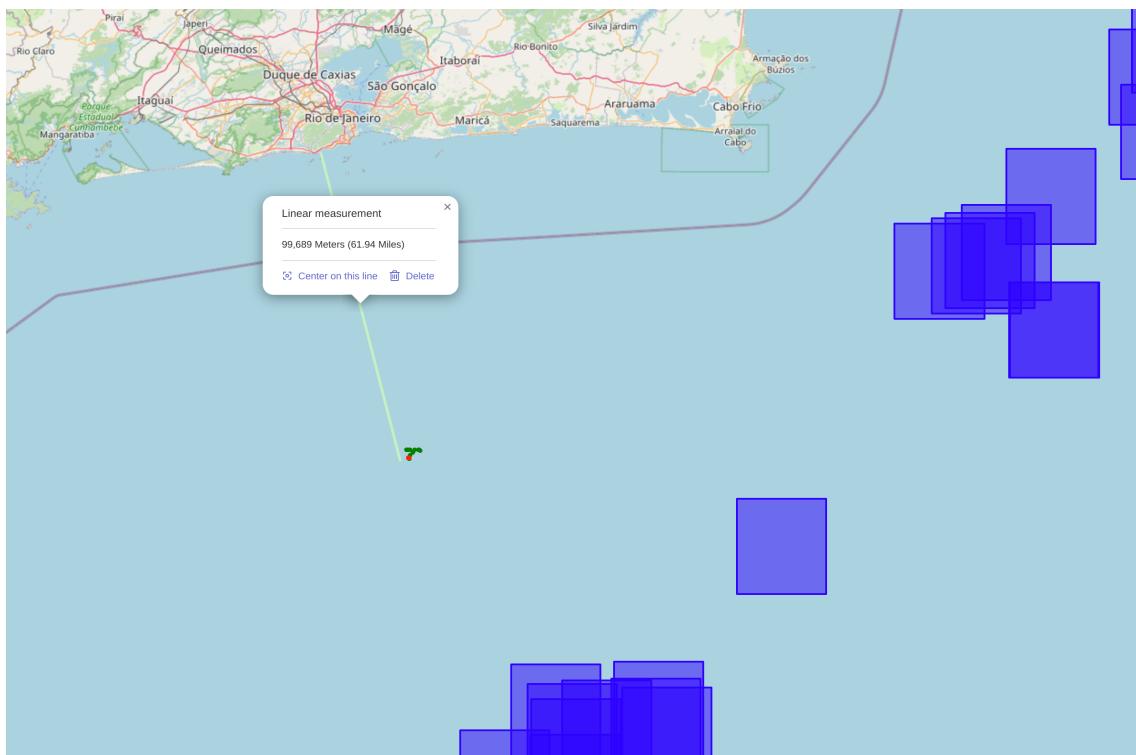


Figura 5.7: A imagem mostra uma atividade suspeita detectada pelo framework. A embarcação representada pelos pontos verdes está encontrando-se com a embarcação representada pelos pontos vermelhos a 61 milhas náuticas da costa. Os quadrados azuis representam áreas de FPSO.

No norte do Brasil, o framework identificou atividades de pesca ilegal relacionadas às trajetórias de embarcações pesqueiras de um país estrangeiro. Na Figura 5.8 podemos observar que uma embarcação de pesca de bandeira estrangeira exibiu

comportamento de pesca dentro da ZEE brasileira. Esse não foi um caso isolado; outras embarcações do mesmo país também apresentaram comportamento similar na referida localidade no norte do Brasil. Em virtude da localização e do fato de as embarcações pertencerem ao mesmo país, essa atividade é considerada severa e precisa ser analisada com mais rigor por um especialista.

Vessel Trajectory Impact Assessment:

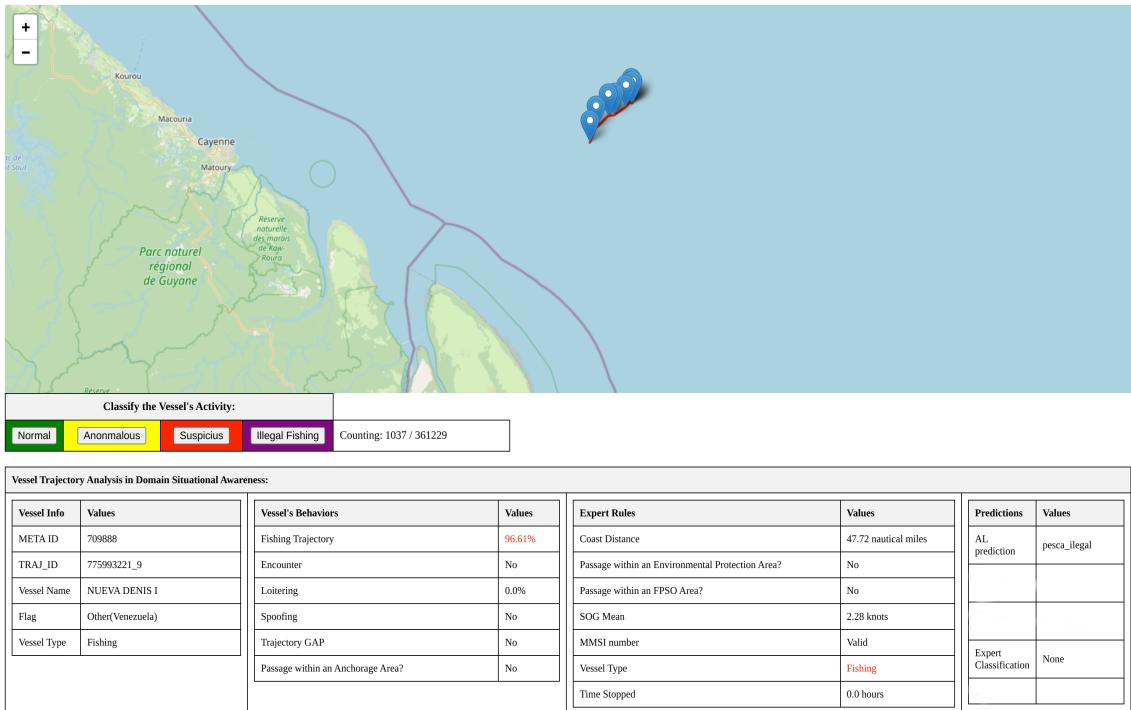


Figura 5.8: A figura mostra uma possível situação de pesca ilegal. O framework detectou uma embarcação de pesca de bandeira estrangeira apresentando trajetória de pesca dentro da ZEE brasileira na região norte da costa.

Um outro caso detectado pelo framework que vale a pena ilustrar é o de embarcações de pesca de bandeira estrangeira fundeadas por um longo período na ZEE brasileira, ao longo da costa norte. Na Figura 5.9 podemos observar uma embarcação de pesca estrangeira fundeada na ZEE brasileira por mais de 3 horas. Por não apresentar uma trajetória de pesca, a atividade foi considerada suspeita. Contudo, mesmo quando fundeada, a embarcação pode estar realizando outra modalidade de pesca. Neste caso, o especialista deve avaliar com cautela, utilizando outras fontes de dados disponíveis.

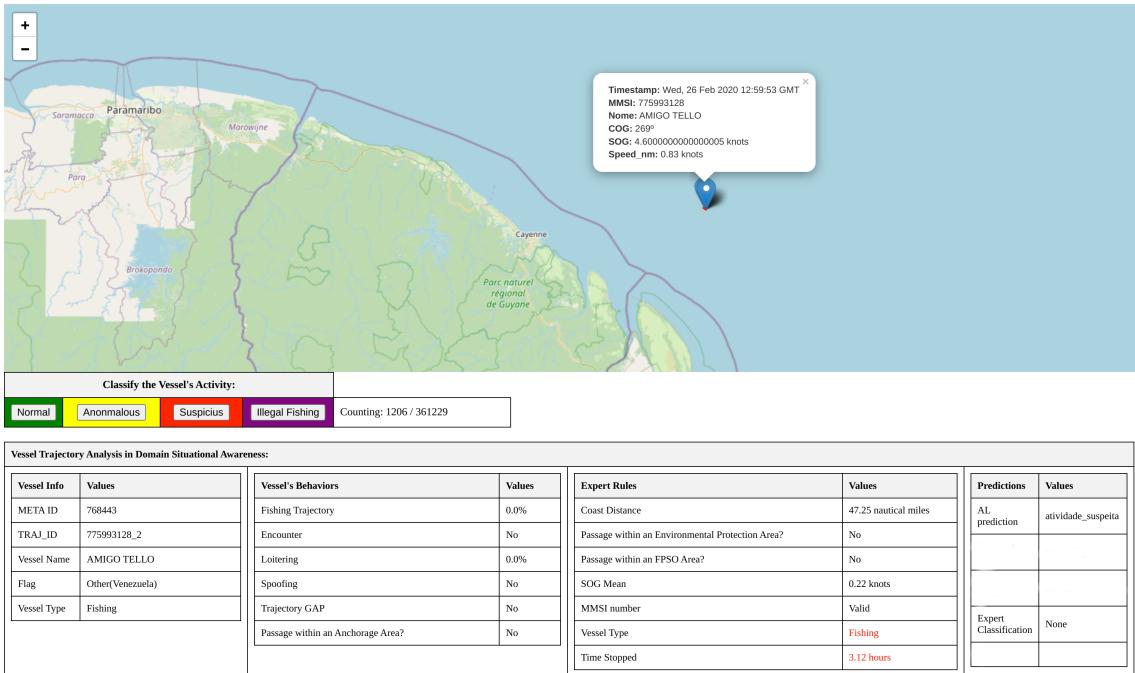


Figura 5.9: A figura mostra uma trajetória suspeita detectada pelo framework. Uma embarcação de pesca de bandeira estrangeira está fundeada há mais de 3 horas dentro da ZEE brasileira na região norte da costa.

Um dos principais resultados desta pesquisa é a possibilidade de aumentar a eficiência dos operadores humanos ao eliminar a necessidade de examinar trajetórias consideradas normais. Podemos realizar uma análise quantitativa para avaliar a carga de trabalho exigida dos operadores humanos neste estudo de caso, na ausência de nosso framework. Na Tabela 5.4, apresentamos a contagem agregada de trajetórias analisadas pelo framework.

Na Tabela 5.4, são apresentados o número de atividades inferidas pelo nosso framework, sendo que 1.200 trajetórias de um total de 361.229 foram utilizadas para treinamento e teste. Inicialmente, em uma situação de monitoramento do domínio marítimo por especialistas, seria possível evitar a verificação de 355.465 trajetórias consideradas como atividades normais, o que representa aproximadamente 98% do total de trajetórias. Isso faz sentido, uma vez que a maioria das trajetórias de embarcações são normais. Posteriormente, os especialistas poderiam priorizar atividades mais severas como as relacionadas a atividades suspeitas e à pesca ilegal, resultando em uma significativa economia de esforço.

5.7 Discussão

Os resultados apresentados destacaram a robustez e a eficácia do *framework* proposto na classificação de atividades ilegais no domínio marítimo, utilizando dados de trajetórias de embarcações. O elevado valor de *precision* e *recall*, especialmente na

detecção de pesca ilegal e atividades normais, demonstram a capacidade do modelo em identificar essas atividades com precisão. No entanto, identificaram-se algumas limitações na classificação de atividades suspeitas, as quais foram ocasionalmente confundidas com atividades normais, indicando uma necessidade de aprimoramento do modelo. Esse refinamento poderia ser alcançado por meio do aumento no número de trajetórias nos dados de treinamento e/ou da inclusão de outros comportamentos e/ou regras que possam caracterizar de maneira mais precisa essas atividades.

A análise SHAP confirmou que as previsões do modelo estão alinhadas com o conhecimento dos especialistas, evidenciando que variáveis como trajetória de pesca e tipo de embarcação sendo determinantes principais na detecção de pesca ilegal. A identificação de atividades suspeitas em cenários reais depende da detecção de encontros entre embarcações em áreas incomuns e de operações em áreas protegidas. A detecção dessas atividades ilustra a aplicabilidade prática do *framework*. Esses resultados sugerem que o *framework* pode reduzir significativamente a carga de trabalho dos operadores humanos em até 98%, ao filtrar trajetórias normais e permitir o foco em casos críticos. Embora tenha sido treinado em dados sintéticos, o *framework* demonstrou capacidade de identificar com precisão atividades suspeitas e anômalas em cenários reais, conforme evidenciado na seção anterior.

Outra consideração importante foi a utilização do *bot* de critérios para avaliar as trajetórias utilizando aprendizado ativo. Inicialmente, tentamos aplicar apenas o *bot* de critérios para avaliar as trajetórias durante o treinamento, sem a participação do especialista. No entanto, essa abordagem não foi bem-sucedida, pois resultou em uma baixa acurácia. Esse resultado pode ser explicado pelo fato de que os dados utilizados no metamodelo são saídas de outros modelos, os quais podem conter erros. Se utilizássemos apenas o *bot* de critérios, estaríamos assumindo que os modelos de primeiro nível não apresentariam falhas. Ao incluir o especialista no treinamento, este, ao analisar a trajetória, pode identificar claramente o erro do modelo, desconsiderar a predição incorreta e classificar manualmente a trajetória de maneira adequada. Dessa forma, o metamodelo pode ajustar os pesos para “compensar” os erros dos modelos de primeiro nível.

Para fins de comparação, poderíamos comparar o desempenho de nossos resultados com a literatura existente. No entanto, até onde sabemos, não há estudos anteriores que abordem nosso problema específico utilizando a metodologia que empregamos. A maioria dos trabalhos na literatura foca especificamente na detecção de comportamentos, em vez de considerar a atividade geral da embarcação.

Atividade	Critério
Normal	<ul style="list-style-type: none"> • Embarcações em rotas normais, com dados completos (MMSI, nome, bandeira, etc.) que não apresentem comportamentos atípicos para a localidade onde estão navegando; • Embarcações do tipo offshore realizando encontros dentro de áreas de FPSO; • Contatos do tipo boia com velocidade média abaixo de 2 nós; • Embarcações do tipo rebocador realizando encontros próximo a costa.
Anômala	<ul style="list-style-type: none"> • Embarcações sem identificação dentro do MT; • Embarcações dentro do MT fundeadas em locais próximos à costa e afastadas de portos; • Comportamento de embarcações inconsistentes com seu tipo (ex.: navios de carga em trajetórias sinuosas); • Embarcações sem identificação apresentando comportamento de pesca dentro do MT ou ZEE; • Embarcações de pesca com nome em português, sem bandeira e próximo a costa; • Embarcações de tipo diferente de pesca ou sem identificação, apresentando comportamento de <i>loitering</i> ou fundeadas dentro de APAs por mais de 2h; • Embarcações sem identificação apresentando comportamentos de <i>spoofing</i> e <i>gaps</i> em trajetórias; • Contatos do tipo boia com velocidade média acima de 2 nós.
Suspeita	<ul style="list-style-type: none"> • Embarcações realizando encontros a mais de 12 MN fora da área de FPSO; • Embarcações que não sejam do tipo rebocador realizando encontros dentro do MT; • Embarcações de pesca apresentando comportamento de <i>loitering</i> ou fundeadas dentro de APAs; • Embarcações sem identificação apresentando comportamentos de pesca dentro de APAs; • Embarcações de pesca estrangeiras apresentando comportamentos de <i>gaps</i> em trajetórias, <i>spoofing</i> ou <i>loitering</i> dentro da ZEE ou MT; • Embarcações estrangeiras de pesca fundeadas dentro da ZEE por mais de 2h.
Pesca Ilegal	<ul style="list-style-type: none"> • Embarcações de pesca apresentando comportamento de pesca dentro de APAs; • Embarcações de pesca de bandeira estrangeira apresentando comportamento de pesca dentro da ZEE ou MT.

Tabela 5.1: Critérios para a classificação das atividades das embarcações.

Etapa do Treinamento	Quantidade de Trajetórias	Tipo de Inserção
Cold start	220	Inserções SQL utilizando os critérios definidos da tabela 5.1.
Aprendizado Ativo	380	Respostas automatizadas utilizando os critérios definidos da tabela 5.1.
Aprendizado Ativo	200	Avaliação manual do especialista.

Tabela 5.2: Quantitativo de dados e estratégia para rotulação no treinamento.

Atividade Predita	Trajetórias
Pesca Ilegal	239
Suspeita	783
Anômala	3.542
Normal	355.465
Total	360.029

Tabela 5.4: Números de trajetórias por predição de atividade.

Capítulo 6

Conclusões e Trabalhos Futuros

Este trabalho apresentou um *framework* capaz de classificar as atividades de embarcações em quatro categorias: pesca ilegal, atividade suspeita, atividade anômala e atividade normal. O *framework* combinou os comportamentos de navegação apresentados por cada trajetória com regras criadas com base no conhecimento do especialista. Para tal, desenvolvemos um metamodelo utilizando o modelo de *random forest* para realizar a fusão de dados. Para lidar com a escassez de rótulos, empregamos técnicas de aprendizado ativo, permitindo que o especialista classificasse as trajetórias com maior probabilidade de aprimorar o desempenho do modelo. Além disso, nosso *framework* foi estruturado com base no modelo JDL, consistindo em cinco níveis.

Para o treinamento dos modelos de detecção de comportamento, utilizamos os conjuntos de dados GFW [83] e MarineCadastre [16]. Para o treinamento e a avaliação de nosso metamodelo com aprendizado ativo, empregamos um conjunto de dados sintético, desenvolvido a partir da translação de trajetórias de pesca e de encontros dos conjuntos GFW e MarineCadastre para um conjunto de dados real da costa brasileira. Nos testes realizados, destacamos o desempenho mais significativo, que ocorreu na detecção de atividades normais, alcançando 100% de acurácia. A detecção de pesca ilegal obteve um desempenho de 97% de acurácia, enquanto a detecção de atividades suspeitas apresentou o desempenho mais baixo, com 90% de acurácia.

Ao longo dos testes, nosso *framework*, treinado com conjuntos de dados sintéticos, identificou com sucesso casos reais de embarcações envolvidas em encontros suspeitos, bem como embarcações estrangeiras pescando dentro da ZEE do Brasil. Isso demonstra que nosso *framework* para a detecção de atividades ilegais de embarcações é viável e flexível para a incorporação de novos comportamentos e atividades, além de reduzir significativamente o número de trajetórias que requerem a atenção dos especialistas. Os algoritmos, dados e códigos-fonte podem ser encontrados em https://github.com/vindn/vad_framework.

Como trabalho futuro, testes com conjuntos de dados maiores e a inclusão de novos comportamentos, regras e atividades poderiam avaliar ainda mais a escalabilidade do *framework*. A especialização na identificação de atividades suspeitas poderia ser aprimorada para abranger atividades mais específicas, como o tráfico de drogas ou a poluição hídrica. A especificação dessas atividades pode ser alcançada pela adição de novos comportamentos/regras que detectem as características dessas atividades. Por fim, este trabalho representa uma contribuição significativa para o campo da vigilância marítima, oferecendo uma solução escalável e eficiente para a detecção de atividades ilegais no domínio marítimo.

Referências Bibliográficas

- [1] STOPFORD, M. *Maritime Economics*. 3rd ed. London and New York, Routledge, 2009. ISBN: 9780415275576.
- [2] FORUM, W. E. “Worlds busiest ocean shipping routes”. 2024. Disponível em: <<https://www.weforum.org/agenda/2024/02/worlds-busiest-ocean-shipping-routes-trade>>.
- [3] CRIMES, M. “MICA Center - Annual Report 2021”. <https://maritimescrimes.com/2022/01/10/mica-center-annual-report-2021/>, 2022. Accessed: 2024-08-13.
- [4] SHARDA. “A General Overview of Maritime Domain Awareness (MDA)”. <https://www.marineinsight.com/maritime-law/a-general-overview-of-maritime-domain-awareness-md/>, 2024. Accessed: 2024-08-13.
- [5] WOLSING, K., ROEPERT, L., BAUER, J., et al. “Anomaly Detection in Maritime AIS Tracks: A Review of Recent Approaches”. 1 2022. ISSN: 20771312.
- [6] DOGANCAY, K., TU, Z., IBAL, G. “Research into vessel behaviour pattern recognition in the maritime domain: Past, present and future”, *Digital Signal Processing: A Review Journal*, v. 119, 12 2021. ISSN: 10512004. doi: 10.1016/j.dsp.2021.103191.
- [7] RONG, H., TEIXEIRA, A., SOARES, C. G. “A framework for ship abnormal behaviour detection and classification using AIS data”, *Reliability Engineering & System Safety*, p. 110105, 3 2024. ISSN: 09518320. doi: 10.1016/j.ress.2024.110105. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0951832024001790>>.
- [8] SHI, Y., LONG, C., YANG, X., et al. “Abnormal Ship Behavior Detection Based on AIS Data”, *Applied Sciences*, v. 12, pp. 4635, 5 2022. doi: 10.3390/app12094635.

- [9] FAHN, C.-S., LING, J., YEH, M.-Y., et al. “Abnormal Maritime Activity Detection in Satellite Image Sequences Using Trajectory Features”, *International Journal of Future Computer and Communication*, v. 8, pp. 29–33, 3 2019. ISSN: 20103751. doi: 10.18178/ijfcc.2019.8.1.535. Disponível em: <<http://www.ijfcc.org/index.php?m=content&c=index&a=show&catid=80&id=886>>.
- [10] ZOCHOLL, M., IPHAR, C., JOUSSELME, A. L., et al. “Ontology-based approach for vessel activity recognition”. v. 2021-September. Institute of Electrical and Electronics Engineers Inc., 2021. ISBN: 9780692935590. doi: 10.23919/OCEANS44145.2021.9705824.
- [11] WANG, Y., LIU, J., LIU, R. W., et al. “Data-driven methods for detection of abnormal ship behavior: Progress and trends”, *Ocean Engineering*, v. 271, pp. 113673, 3 2023. ISSN: 00298018. doi: 10.1016/j.oceaneng.2023.113673.
- [12] FERLANSYAH, N. M., SUHARJITO. “A systematic literature review of vessel anomaly behavior detection methods based on Automatic Identification System (AIS) and another sensor fusion”, *Advances in Science, Technology and Engineering Systems*, v. 5, pp. 287–292, 2020. ISSN: 24156698. doi: 10.25046/aj050237.
- [13] WATSON, J. R., WOODILL, A. J. “Detecting illegal maritime activities from anomalous multiscale fleet behaviours”, *Fish and Fisheries*, v. 23, pp. 1055–1069, 9 2022. ISSN: 14672979. doi: 10.1111/faf.12669.
- [14] “European Maritime Safety Agency (EMSA)”. <https://www.emsa.europa.eu>. Accessed: 2024-05-24.
- [15] GFW. “Global Fishing Watch”. <https://globalfishingwatch.org/>, 2023. [Online; accessed 07-March-2023].
- [16] CADASTRE, M. “Marine Cadastre National Viewer Datasets”. <https://marinecadastre.gov/data/>, 2024. Accessed: 2024-05-20.
- [17] MB. “Sistema de Informações Sobre o Tráfego Marítimo”. 2024. Disponível em: <<https://www.marinha.mil.br/salvamarbrasil/Sistema/sistram>>.
- [18] DO NASCIMENTO, V. D., ALVES, T. A. O., DUTRA, D. L. C., et al. “A Comparative Study of Fishing Activity Detection Approaches in Maritime Surveillance”. In: *2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE)*, pp. 173–179, 2023. doi: 10.1109/CSCE60160.2023.00033.

- [19] DO NASCIMENTO, V. D., ALVES, T. A. O., DUTRA, D. L. C. “Multi-Modal Fusion Approach for Enhancing Vessel Activity Monitoring”. In: *2023 1st Latin American Workshop on Information Fusion (LAFUSION)*, Rio de Janeiro, Brazil, 2023.
- [20] DO NASCIMENTO, V. D., FARIAS, C. M. D., DUTRA, D. L. C., et al. “Ensemble Learning Approaches for Detecting Fishing Activity in Maritime Surveillance: A Performance Evaluation”. In: *2024 27th International Conference on Information Fusion (FUSION)*, pp. 1–8, 2024. doi: 10.23919/FUSION59988.2024.10706287.
- [21] DO NASCIMENTO, V. D., ALVES, T. A. O., DE FARIAS, C. M., et al. “A Hybrid Framework for Maritime Surveillance: Detecting Illegal Activities through Vessel Behaviors and Expert Rules Fusion”, *Sensors*, v. 24, n. 17, pp. 5623, 2024. doi: 10.3390/s24175623.
- [22] DO NASCIMENTO, V. D., FARIAS, C. M. D., DUTRA, D. L. C., et al. “Analyzing Offshore Vessel Encounters: A Dataset for Enhancing Maritime Security and Monitoring”. In: *2025 28th International Conference on Information Fusion (FUSION)*, 2025. Submetido.
- [23] KESSLER, G. C., ZORRI, D. M. “AIS Spoofing: A Tutorial for Researchers”. In: *2024 IEEE 49th Conference on Local Computer Networks (LCN)*, pp. 1–7, 2024. doi: 10.1109/LCN60385.2024.10639747.
- [24] KONTOPOULOS, I., SPILIOPOULOS, G., ZISSIS, D., et al. “Countering Real-Time Stream Poisoning: An Architecture for Detecting Vessel Spoofing in Streams of AIS Data”. pp. 981–986. IEEE, 8 2018. ISBN: 978-1-5386-7518-2. doi: 10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00139. Disponível em: <<https://ieeexplore.ieee.org/document/8512006/>>.
- [25] KOT, G. J., MENG, W., LAMPE, B. “Securing Offshore Installations Against Automatic Identification System Spoofing”. In: *GLOBECOM 2023 - 2023 IEEE Global Communications Conference*, pp. 4473–4478, 2023. doi: 10.1109/GLOBECOM54140.2023.10437123.
- [26] IPHAR, C., NAPOLI, A., RAY, C. “Detection of false AIS messages for the improvement of maritime situational awareness”. In: *OCEANS 2015 - MTS/IEEE Washington*. Institute of Electrical and Electronics Engineers Inc., 2 2016. ISBN: 9780933957435. doi: 10.23919/oceans.2015.7401841.

- [27] D’AFFLISIO, E., BRACA, P., WILLETT, P. “Malicious AIS Spoofing and Abnormal Stealth Deviations: A Comprehensive Statistical Framework for Maritime Anomaly Detection”, *IEEE Transactions on Aerospace and Electronic Systems*, v. 57, pp. 2093–2108, 8 2021. ISSN: 15579603. doi: 10.1109/TAES.2021.3083466.
- [28] KRUGER, M. “Detection of AIS Spoofing in Fishery Scenarios”. In: *2019 22th International Conference on Information Fusion (FUSION)*, pp. 1–7, 2019. doi: 10.23919/FUSION43075.2019.9011328.
- [29] MAZZARELLA, F., VESPE, M., ALESSANDRINI, A., et al. “A novel anomaly detection approach to identify intentional AIS on-off switching”, *Expert Systems with Applications*, v. 78, pp. 110–123, 7 2017. ISSN: 09574174. doi: 10.1016/j.eswa.2017.02.011.
- [30] KELLY, P. “A novel technique to identify AIS transmissions from vessels which attempt to obscure their position by switching their AIS transponder from normal transmit power mode to low transmit power mode”, *Expert Systems with Applications*, v. 202, 9 2022. ISSN: 09574174. doi: 10.1016/j.eswa.2022.117205.
- [31] KONTOPOULOS, I., CHATZIKOKOLAKIS, K., ZISSIS, D., et al. “Real-time maritime anomaly detection: detecting intentional AIS switch-off”, *International Journal of Big Data Intelligence*, v. 7, n. 2, pp. 85–96, 2020. doi: 10.1504/IJBDI.2020.107375. Disponível em: <<https://www.inderscienceonline.com/doi/abs/10.1504/IJBDI.2020.107375>>.
- [32] GÖRKEM, B. N., ÇAĞLAYAN, B., KARACA, E., et al. “Dark Activity Detection in AIS-Based Maritime Networks”. In: *2023 34th Conference of Open Innovations Association (FRUCT)*, pp. 35–40, 2023. doi: 10.23919/FRUCT60429.2023.10328171.
- [33] IOANNOU, G., GAGLIONE, D., MILLEFIORI, L. M., et al. “Dark-VADER: Detection of Anomalous AIS Message Delays for Maritime Situational Awareness”. In: *2024 27th International Conference on Information Fusion (FUSION)*, pp. 1–7, 2024. doi: 10.23919/FUSION59988.2024.10706391.
- [34] UNEY, M., MILLEFIORI, L. M., BRACA, P. “Prediction of Rendezvous in Maritime Situational Awareness”. In: *2018 21st International Conference on Information Fusion (FUSION)*, pp. 622–628, 2018. doi: 10.23919/ICIF.2018.8455816.

- [35] SHARMA, A., SHEKHAR, S. “Analyzing Trajectory Gaps to Find Possible Rendezvous Region”, *ACM Trans. Intell. Syst. Technol.*, v. 13, n. 3, jan 2022. ISSN: 2157-6904. doi: 10.1145/3467977. Disponível em: <<https://doi.org/10.1145/3467977>>.
- [36] CHEN, X., LIU, Y., ACHUTHAN, K., et al. “A semi-supervised deep learning model for ship encounter situation classification”, *Ocean Engineering*, v. 239, 11 2021. ISSN: 00298018. doi: 10.1016/j.oceaneng.2021.109824.
- [37] PILI, G., ARMENZONI, A. “Fighting Shadows and Ghosts – A Conceptual Framework for Identification Deception Tactics (IDTs) and MARINE Approach”. pp. 1–7. Institute of Electrical and Electronics Engineers (IEEE), 9 2024. doi: 10.1109/lcn60385.2024.10639801.
- [38] KONTOPOULOS, I., MAKRIS, A., TSERPES, K. “A deep learning streaming methodology for trajectory classification”, *ISPRS International Journal of Geo-Information*, v. 10, 4 2021. ISSN: 22209964. doi: 10.3390/ijgi10040250.
- [39] ARASTEH, S., TAYEBI, M. A., ZOHREVAND, Z., et al. “Fishing Vessels Activity Detection from Longitudinal AIS Data”. pp. 347–356. Association for Computing Machinery, 11 2020. ISBN: 9781450380195. doi: 10.1145/3397536.3422267.
- [40] FERREIRA, M. D., SPADON, G., SOARES, A., et al. “A Semi-Supervised Methodology for Fishing Activity Detection Using the Geometry behind the Trajectory of Multiple Vessels”, *Sensors*, v. 22, 8 2022. ISSN: 14248220. doi: 10.3390/s22166063.
- [41] CHUAYSI, B., KIATTISIN, S. “Fishing Vessels Behavior Identification for Combating IUU Fishing: Enable Traceability at Sea”, *Wireless Personal Communications*, v. 115, pp. 2971–2993, 12 2020. ISSN: 1572834X. doi: 10.1007/s11277-020-07200-w.
- [42] PEDROCHE, D. S., AMIGO, D., GARCÍA, J., et al. “Architecture for trajectory-based fishing ship classification with AIS data”, *Sensors (Switzerland)*, v. 20, pp. 1–21, 7 2020. ISSN: 14248220. doi: 10.3390/s20133782.
- [43] SHENG, K., LIU, Z., ZHOU, D., et al. “Research on Ship Classification Based on Trajectory Features”, *Journal of Navigation*, v. 71, pp. 100–116, 1 2018. ISSN: 14697785. doi: 10.1017/S0373463317000546.

- [44] PATINO, L., FERRYMAN, J. “Loitering Behaviour Detection of Boats at Sea”. v. 2017-July, pp. 2169–2175. IEEE Computer Society, 8 2017. ISBN: 9781538607336. doi: 10.1109/CVPRW.2017.269.
- [45] ZHANG, Z., HUANG, L., PENG, X., et al. “Loitering behavior detection and classification of vessel movements based on trajectory shape and Convolutional Neural Networks”, *Ocean Engineering*, v. 258, 8 2022. ISSN: 00298018. doi: 10.1016/j.oceaneng.2022.111852.
- [46] WIJAYA, W. M., NAKAMURA, Y. “Loitering behavior detection by spatio-temporal characteristics quantification based on the dynamic features of Automatic Identification System (AIS) messages”, *PeerJ Computer Science*, v. 9, 2023. ISSN: 23765992. doi: 10.7717/peerj-cs.1572.
- [47] LIU, J., LI, J., LIU, C. “AIS-based kinematic anomaly classification for maritime surveillance”, *Ocean Engineering*, v. 305, 8 2024. ISSN: 00298018. doi: 10.1016/j.oceaneng.2024.118026.
- [48] KIM, H., CHOI, M., PARK, S., et al. “Vessel trajectory classification via transfer learning with Deep Convolutional Neural Networks”, *PLoS ONE*, v. 19, 8 2024. ISSN: 19326203. doi: 10.1371/journal.pone.0308934.
- [49] POHONTU, A., NEDELCU, L., VERTAN, C. “Fitting Empirical Distributions for Vessels Behavioural Analysis and Maritime Anomaly Detection”. In: *2023 17th International Conference on Engineering of Modern Electric Systems, EMES 2023*. Institute of Electrical and Electronics Engineers Inc., 2023. ISBN: 9798350310634. doi: 10.1109/EMES58375.2023.10171682.
- [50] MAZZARELLA, F., ARGUEDAS, V. F., VESPE, M. “Knowledge-based vessel position prediction using historical AIS data”. In: *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pp. 1–6, 2015. doi: 10.1109/SDF.2015.7347707.
- [51] PITSIKALIS, M., KONTOPOULOS, I., ARTIKIS, A., et al. “Composite Event Patterns for Maritime Monitoring”. In: *Proceedings of the 10th Hellenic Conference on Artificial Intelligence*, SETN ’18, New York, NY, USA, 2018. Association for Computing Machinery. ISBN: 9781450364331. doi: 10.1145/3200947.3201042. Disponível em: <<https://doi.org/10.1145/3200947.3201042>>.

- [52] KOWALSKA, K., PEEL, L. “Maritime anomaly detection using Gaussian Process active learning”. In: *2012 15th International Conference on Information Fusion*, pp. 1164–1171, 2012.
- [53] “Amazonia Azul”. <https://www.marinha.mil.br/delareis/?q=amazoniazul>. Accessed: 2024-05-27.
- [54] UNITED NATIONS. “United Nations Convention on the Law of the Sea”. https://www.un.org/depts/los/convention_agreements/texts/unclos/unclos_e.pdf, 1982. Acesso em: 28 de outubro de 2024.
- [55] U.S. DEPARTMENT OF HOMELAND SECURITY. “National Plan to Achieve Maritime Domain Awareness”. https://www.dhs.gov/sites/default/files/publications/HSPD_MDAPlan_0.pdf, 2005. Accessed: 2024-10-28.
- [56] U.S. MARITIME ADMINISTRATION. “SeaVision Program”. <https://info.seavision.volpe.dot.gov/#overview>, 2023. Acesso em: 28 de outubro de 2024.
- [57] ZHOU, Z.-H. *Ensemble Methods: Foundations and Algorithms*. Chapman and Hall/CRC, 2012.
- [58] SETTLES, B. *Active Learning Literature Survey*. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.
- [59] STEINBERG, A. N., BOWMAN, C. L., WHITE, F. E. “Revisions to the JDL Data Fusion Model”.
- [60] TEAM, G. D. “GeoPandas: Python tools for geographic data”, *Journal of Open Source Software*, v. 3, n. 24, pp. 3522, 2018.
- [61] GRASER, A. “MovingPandas: Efficient Structures for Movement Data in Python”. <https://github.com/anitagraser/movingpandas>, 2021. Accessed: 2023-03-08.
- [62] ANDROJNA, A., PERKOVIC, M., PAVIC, I., et al. “Ais data vulnerability indicated by a spoofing case-study”, *Applied Sciences (Switzerland)*, v. 11, 6 2021. ISSN: 20763417. doi: 10.3390/app11115015.
- [63] ANDROJNA, A., PERKOVIĆ, M. “Impact of Spoofing of Navigation Systems on Maritime Situational Awareness”, *Transactions on Maritime Science*, v. 10, pp. 361–373, 10 2021. ISSN: 1848-3305. doi: 10.7225/toms.v10.n02.w08.

- [64] INSTITUTE, F. M. “Union of the ESRI Country shapefile and the Exclusive Economic Zones (version 3)”. 2020. Disponível em: <<https://www.marineregions.org/>>. Available online at <https://www.marineregions.org/>.
- [65] GILLIES, S., CONTRIBUTORS. “Shapely: Manipulation and analysis of geometric objects”. 2007. Disponível em: <<https://shapely.readthedocs.io/>>. Disponível em Python Package Index (PyPI): <https://pypi.org/project/Shapely/>.
- [66] ZHANG, T., ZHAO, S., CHENG, B., et al. “Detection of AIS closing behavior and MMSI spoofing behavior of ships based on spatiotemporal data”, *Remote Sensing*, v. 12, 2 2020. ISSN: 20724292. doi: 10.3390/rs12040702.
- [67] NIKOLIC, D., STOJKOVIC, N., POPOVIC, Z., et al. “Maritime over the horizon sensor integration: HFSWR data fusion algorithm”, *Remote Sensing*, v. 11, 4 2019. ISSN: 20724292. doi: 10.3390/RS11070852.
- [68] NIEMEYER, G. “Geohash”. <http://geohash.org>, 2008. Accessed: 2024-05-20.
- [69] SHAHIR, H. Y., GLASSER, U., SHAHIR, A. Y., et al. “Maritime situation analysis framework: Vessel interaction classification and anomaly detection”. pp. 1279–1289. Institute of Electrical and Electronics Engineers Inc., 12 2015. ISBN: 9781479999255. doi: 10.1109/BigData.2015.7363883.
- [70] MATOSSIAN, M., LAURILA, P., BLANCHET, C. “Detecting dark vessels: Radar satellite-based monitoring of illegal activities at sea”, *Sea Technology*, v. 61, n. 6, pp. 30 – 32, 2020. ISSN: 00933651. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85092525105&partnerID=40&md5=6f9cba80ebf21\3a2defd9e8a8a303047>>. Cited by: 0.
- [71] UBER TECHNOLOGIES, I. “H3: A Hexagonal Hierarchical Spatial Index”. <https://h3geo.org>, 2018. Accessed: 2024-05-20.
- [72] DO NASCIMENTO, V. D., DE FARIAS, C. M., DUTRA, D. L. C., et al. “Ensemble Learning Approaches for Detecting Fishing Activity in Maritime Surveillance: A Performance Evaluation”. In: *2024 27th International Conference on Information Fusion (FUSION)*, pp. 1–8, 2024. doi: 10.23919/FUSION59988.2024.10706287.

- [73] DO NASCIMENTO, V. D., ALVES, T. A. O., DUTRA, D. L. C., et al. “A Comparative Study of Fishing Activity Detection Approaches in Maritime Surveillance”. In: *2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE)*, pp. 173–179, 2023. doi: 10.1109/CSCE60160.2023.00033.
- [74] LI, W., ZHANG, D., SUN, M., et al. “Loitering Detection Based on Trajectory Analysis”. pp. 530–533. Institute of Electrical and Electronics Engineers Inc., 5 2016. ISBN: 9781467376440. doi: 10.1109/ICICTA.2015.136.
- [75] LU, R., YANG, H., ZHU, J., et al. “Hierarchical video summarization with loitering indication”. Institute of Electrical and Electronics Engineers Inc., 4 2016. ISBN: 9781467373142. doi: 10.1109/VCIP.2015.7457818.
- [76] SIDIBÉ, A., SHU, G. “Study of automatic anomalous behaviour detection techniques for maritime vessels”, *Journal of Navigation*, v. 70, pp. 847–858, 7 2017. ISSN: 14697785. doi: 10.1017/S0373463317000066.
- [77] PAN, X., WANG, H., CHENG, X., et al. “Online detection of anomaly behaviors based on multidimensional trajectories”, *Information Fusion*, v. 58, pp. 40–51, 2020. ISSN: 1566-2535. doi: <https://doi.org/10.1016/j.inffus.2019.12.009>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1566253518307498>>.
- [78] LAXHAMMAR, R., FALKMAN, G. “Conformal Prediction for Distribution-Independent Anomaly Detection in Streaming Vessel Data”. In: *Proceedings of the First International Workshop on Novel Data Stream Pattern Mining Techniques*, StreamKDD ’10, p. 47–55, New York, NY, USA, 2010. Association for Computing Machinery. ISBN: 9781450302265. doi: 10.1145/1833280.1833287. Disponível em: <<https://doi.org/10.1145/1833280.1833287>>.
- [79] LU, N., LIANG, M., YANG, L., et al. “Shape-Based Vessel Trajectory Similarity Computing and Clustering: A Brief Review”. In: *2020 5th IEEE International Conference on Big Data Analytics (ICBDA)*, pp. 186–192, 2020. doi: 10.1109/ICBDA49040.2020.9101322.
- [80] DE FARIA, J. A. P. M. “A consciência situacional marítima (CSM) e a Marinha do Brasil”, *NAVAL WAR COLLEGE JOURNAL; Vol 18, No 1 (2012): REVISTA DA ESCOLA DE GUERRA NAVAL*, 2012.
- [81] NOAA FISHERIES. “Illegal, Unreported, and Unregulated Fishing”. 2020. Disponível em: <<https://www>.

<fisheries.noaa.gov/national/international-affairs/illegal-unreported-and-unregulated-fishing>. Accessed: 2024-07-02.

[82] BOSKOVITZ, D. D. E., PAOLO, C. I. D. “modAL: A modular active learning framework for Python”. <https://github.com/modAL-python/modAL>, 2018. Accessed: 2024-10-31.

[83] WATCH, G. F. “Global Fishing Watch Dataset”. <https://globalfishingwatch.org/data-download/datasets/public-training-data-v1>, 2023. [Online; accessed 07-March-2023].

[84] LUNDBERG, S. M., LEE, S.-I. “A Unified Approach to Interpreting Model Predictions”, *Advances in Neural Information Processing Systems*, v. 30, 2017. Disponível em: <<https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf>>.

Apêndice A

Abreviações

As seguintes abreviações são utilizadas neste texto:

Tabela A.1: Siglas e Definições

AIS	Automatic Identification System
APA	Área de Proteção Ambiental
MA	Maritime authority
JDL	Joint Directors of Laboratories Data Fusion Model
SHAP	SHapley Additive exPlanations
AL	Active learning
HF	High Frequency
EEZ	Economic Exclusive Zone
ZEE	Zona Econômica Exclusiva
MT	Mar Territorial
MN	Milhas Náuticas
MPA	Marine protected area
MDA	Maritime Domain Awareness
ABM	Automated Behaviour Monitoring
EMSA	European Maritime Safety Agency
GMSA	Global Maritime Situational Awareness
LEO	Low Earth Orbit
GFW	Global Fishing Watching
LSTM	Long-Short Term Memory

RNN	Recurrent neural network
COG	Course over ground
SOG	Speed over ground
MMSI	Maritime Mobile Service Identity
VHF	Very high frequency
NM	Nautical miles
FPSO	Floating Production Storage and Offloading
OTH	Over-the-Horizon
SAR	Synthetic Aperture Radar
ft	Fishing trajectory
loi	Loitering
enc	Encounter
ts	Territorial sea
