

UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO

Alex Koji Misumi

**Mapeamento do Processo Decisório da Operação do
Porto de Santos utilizando análise de dados**

São Paulo

2023

ALEX KOJI MISUMI

Mapeamento do Processo Decisório da Operação do Porto de Santos utilizando análise de dados

— Versão original —

Trabalho de formatura do Programa de Graduação do Departamento de Engenharia de Produção da Escola Politécnica da Universidade de São Paulo.

Orientador: Daniel de Oliveira Mota

São Paulo
2023

Este trabalho é todo dedicado aos meus pais, pois é graças aos seus esforços que hoje posso concluir o meu curso e sem eles nada disso teria sido possível.

Agradecimentos

Aos meus pais (Lucia e Kenji) e irmão (Kenzo), que me incentivaram nos momentos difíceis e compreenderam a minha ausência enquanto eu me dedicava à realização deste trabalho.

Aos amigos, que sempre estiveram ao meu lado, pela amizade incondicional e pelo apoio demonstrado ao longo de todo o período de tempo em que me dediquei a este trabalho. Em especial os parceiros do time de tênis (Fabrizzio, Gabriel, Eduardo, Stefani, Helena, Rafael e Paulo) e a minha namorada (Wendy).

Ao professor Daniel de Oliveira Mota, por ter sido meu orientador e ter desempenhado tal função com muita dedicação e amizade.

Aos professores, pelas correções e ensinamentos que me permitiram apresentar um melhor desempenho no meu processo de formação profissional ao longo do curso.

Às pessoas com quem convivi ao longo desses anos de curso (Ricardo, Yan e Lincoln), que me incentivaram e que certamente tiveram impacto na minha formação acadêmica.

Aos meus colegas de turma (Bruno, Lucas, Gabriel, Estevan e Tulio) por compartilharem comigo tantos momentos de descobertas e aprendizado e por todo o companheirismo ao longo deste percurso.

À instituição de ensino Escola Politécnica da Universidade de São Paulo, essencial no meu processo de formação profissional, pela dedicação, e por tudo o que aprendi ao longo dos anos do curso.

Já tentei.
Já falhei.
Não importa.
Tente outra vez.
Falhe outra vez.
Falhe melhor.

— *Samuel Beckett*

Lista de Figuras

1	Movimentações Portuárias (Janeiro a Agosto de 2023)	15
2	Perfil de carga do Porto de Santos	16
3	Acesso Local	16
4	Fundeadouros do Porto de Santos	19
5	Trechos do Canal de Santos	21
6	Figura de uma lancha de Praticagem	22
7	Sistema para planejamento de portos e agendamento	24
8	Fluxo de trabalho do sistema de agentes	25
9	Fluxograma do algoritmo	28
10	Análise da Distribuição do Tempo de Estadia	30
11	Exemplo de um <i>grid</i> espaço-temporal de 4×4	32
12	Comparação do real x previsto pelo modelo com <i>Xgboost</i> em células de <i>grid</i> ocupadas para: a) base de treino e b) base de validação e teste . .	34
13	Comparação dos modelos no problema de VTFF	35
14	Exemplos de classificação comportamental para embarcações atracando	38
15	Gráfico de acurácia do Treino x Teste	39
16	Análise univariada das <i>features</i> notáveis para o modelo	40
17	Captura da Tela do site <i>marinetraffic</i> de acompanhamento global de embarcações	44
18	Resultado da função <i>.info()</i> na base filtrada	51
19	Polígonos formados pelas coordenadas das áreas de fundeio	55
20	Polígonos formados pelas coordenadas das áreas de dragagem	56
21	Comparativo entre Figura gerada pelo código e a Figura do <i>Google Maps</i> no site do Porto de Santos	57
22	Movimentação da embarcação 'ADAM I' e o local 'Trecho_1Csoleira' . .	60
23	Mapeamento do trajeto realizado pela embarcação 'ADVENTURE' . . .	62
24	Tipos de Embarcações mais Presentes na Base de Dados	63
25	Origem com maior representatividade na Base de Dados	64
26	Origem com maior representatividade na Base de Dados	65
27	Histograma da Distribuição da Velocidade ('SPEED') das embarcações	67
28	Distribuição da quantidade de embarcações pelo 'Estado_Navio' e 'Janela'	68
29	Distribuição da quantidade de embarcações distintas realizando operação de movimentação pelas Janelas	69
30	Análise da Característica de Atributos por Tipo de Embarcação	70
31	Gráficos <i>Boxplot</i> da métrica de tempo de operação de entrada	72
32	Funil do tamanho da base de dados pelo desenvolvimento do modelo .	73

33	Relação entre inteligência artificial, aprendizado de máquina, aprendizado profundo e ciência de dados	74
----	--	----

Lista de Tabelas

1	Tabela dos berços de atracação	20
2	Comparação dos modelos utilizados	31
3	Resultados do modelo baseado em VRF com diferentes parâmetros . .	35
4	Estrutura da base de dados	48
5	Trecho da Tabela Final	61
6	Tabela com dados descritivos sobre os campos calculados de intervalos de tempo	71
7	Tabela com dados descritivos sobre os campos calculados com trata- mento de <i>outliers</i> de intervalos de tempo	72
8	Pontos geográficos do Fundeadouro nº 1	87
9	Pontos geográficos do Fundeadouro nº 2	87
10	Pontos geográficos do Fundeadouro nº 3	87
11	Pontos geográficos do Fundeadouro nº 4	88
12	Pontos geográficos do Fundeadouro nº 5	88
13	Pontos geográficos do Fundeadouro nº 6	88

Lista de Códigos

1	<i>Webscrapping</i> dos dados da AIS no site da <i>Marinetraffic</i>	79
2	Criação das coluna 'Estado_Navio' e 'Janela'	79
3	Criação das colunas 'Flag_Mudanca_Estado' e 'Flag_3Mov'	79
4	Criação da coluna de 'Mudanca_Estado'	80
5	Aplicação da função 'verificar_fundeio'	81
6	Conversão dos dados geográficos de UTM para graus decimais	81
7	Aplicação da função 'verificar_dragagem'	82
8	Categorização do tipo de movimentação para as Áreas de Dragagem .	82
9	Coleta dos horários relacionados ao fundeio das embarcações	83
10	Marcação das linhas críticas para a documentação	84
11	Criação das colunas de horários dos momentos críticos da movimentação da embarcação	84
12	Agrupamento da movimentação a nível embarcação e área de dragagem	85
13	Tratamento dos dados de horários e geração da tabela descritiva	85
14	Função para tratar os <i>outliers</i>	86

Sumário

Resumo	12
Abstract	13
1 Introdução	14
1.1 Porto de Santos	14
1.1.1 Contextualização histórica	14
1.1.2 Panorama Operacional	14
1.1.3 Acesso e Conexão com o Porto	15
1.1.4 Operações Portuárias	17
1.2 Problemática	22
1.3 Objetivos	23
2 Revisão da Literatura	24
2.1 Sistemas de agentes distribuídos para planejamento de portos e agendamento	24
2.2 Modelo de agendamento de navios em canais de mão única	27
2.3 Classificação do período de estadia no porto	29
2.4 Modelos de <i>Machine Learning</i> para Previsão do Fluxo de Tráfego de Embarcações	31
2.4.1 VTFF baseado em VRF (<i>Vessel Route Forecasting</i>)	32
2.4.2 VTFF baseado em fluxo sequencial	33
2.5 Análises de entregas e operações do porto usando <i>Big Data</i>	37
2.6 Agendamento eficiente online para vias aquáticas restritas	41
3 Metodologia	43
3.1 Coleta de Dados	43
3.2 Desenvolvimento do Modelo	44
3.3 Dados e Códigos	48
3.3.1 Validação dos Dados	48
3.3.2 Filtragem Inicial dos Dados	50
3.3.3 Tratamento Inicial dos Dados	50
3.3.4 Criação de colunas de estado da embarcação	52
3.3.5 Criação de colunas de manipulação intermediárias	53
3.3.6 Criação de colunas de análise	53
3.3.7 Identificação de Localização das Embarcações nas Áreas de Fundeio	54

3.3.8	Identificação de Localização das Embarcações nas Áreas de Dragagem	55
3.3.9	Categorização do procedimento de movimentação	58
3.3.10	Documentação do procedimento de movimentação	58
3.3.11	Agrupamento dos procedimentos por embarcação e área de dragagem	61
3.4	Análise Exploratória da Base de Dados	62
3.4.1	Análise Primária da Base de Dados - Dados Categóricos	62
3.4.2	Análise Primária da Base de Dados - Dados Numéricos	66
3.4.3	Análise Secundária da Base de Dados	69
4	Discussão	73
5	Conclusões	76
	Apêndices	79
	Anexos	87

Resumo

MISUMI, A. *Mapeamento do Processo Decisório da Operação do Porto de Santos utilizando análise de dados 2023*. Trabalho de Formatura – Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção. Universidade de São Paulo, São Paulo, 2023.

No atual cenário global, é possível observar a crescente utilização de dados para melhorar múltiplos processos (ex: previsão e análise de dados) em diversas instituições. Desta maneira, instituições com grandes volumes de dados apresentam um grande potencial de adotarem este tipo de estratégia. A indústria naval portuária, apesar de ser rica em dados, ainda carece de sua utilização e, também, de um aprofundamento de uma mentalidade *data-driven*. Nos portos, existem diversos processos que podem ser aprimorados a partir de um melhor mapeamento e utilização de dados como as movimentações das embarcações (e subsequentes operações internas). Este trabalho se propõe a explorar e aprofundar o mapeamento do processo decisório das operações do Porto de Santos, uma vez identificada, nesta indústria, a necessidade da utilização e aplicação de análise de dados. A operação portuária, que está contida em um ambiente com múltiplos processos interligados, oferece oportunidades significativas para melhorias por meio da análise de dados, especialmente no que diz respeito ao sequenciamento das filas para entrada e atracação nos berços e terminais. O presente projeto fez uso de dados provenientes do sistema AIS (Sistema de Identificação Automática) e da Autoridade Portuária de Santos (SPA) para a construção de um mapeamento preciso e abrangente do fluxo de movimentação das embarcações, com ênfase na ordenação e priorização das movimentações de navios. Ao abordar essa lacuna na indústria, este trabalho visa não apenas aprimorar a eficiência operacional do Porto de Santos mas, também, contribuir para o desenvolvimento de abordagens estratégicas e baseada em dados no contexto portuário. Este projeto realizou diversos processos da área de Ciência de Dados como coleta e processamento de dados e a realização de análises e o levantamento do fluxo de movimentação das embarcações. O estudo proposto tem como objetivo, assim, promover a integração eficaz entre os dados disponíveis e a tomada de decisões, buscando uma operação mais eficiente, segura e adaptativa neste importante hub logístico nacional.

Palavras-chave: Análise de Dados, Ciência de Dados, Sequenciamento de navios, Porto de Santos

Abstract

MISUMI, A. *Mapping the decision-making process in the Port of Santos using data analysis* 2023. Trabalho de Formatura – Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção. University of São Paulo, São Paulo, Brazil. 2023.

In the current global scenario, it is possible to observe the growing use of data to improve multiple processes (e.g. forecasting and data analysis) in various institutions. Institutions with large volumes of data therefore have great potential for this strategy. However, one industry with this characteristic that still lacks both the use of data and the deepening of a more *data-driven* mentality is the shipping and ports industry. In ports, there are several processes that could be improved through better mapping and use of data, such as port operations. This work sets out to explore and deepen the mapping of the decision-making process of operations at the Port of Santos, given the need for a more data-driven approach in the port industry. The port operation, which is contained in an environment with multiple interconnected processes, offers significant opportunities for improvement through data analysis, especially with regard to the sequencing of queues for entry and berthing at berths and terminals. The project made use of data from the AIS (Automatic Identification System) and the Santos Port Authority (SPA) to build an accurate and comprehensive mapping of the decision-making process, with an emphasis on the ordering and prioritization of ship movements. By addressing this gap in the industry, this work aims not only to improve the operational efficiency of the Port of Santos, but also to contribute to the development of strategic and data-based approaches in the context of ports in general. This project carried out various processes in the area of Data Science, such as collecting and processing data, performing analysis and surveying the flow of ship movements. The proposed study thus seeks to promote effective integration between available data and decision-making, for more efficient, safe and adaptive operations in this important national logistics hub.

Keywords: Data Analysis, Data Science, Vessel Sequencing, Santos Port

1 Introdução

1.1 Porto de Santos

1.1.1 Contextualização histórica

O Porto de Santos, cujas raízes remontam ao início do século XV, desempenhou um papel de extraordinária relevância econômica ao longo dos anos, exercendo influência crucial tanto a nível nacional quanto regional. Originando-se no contexto da produção açucareira no final do século XVIII, o porto logo se tornou um centro estratégico para as exportações de café a partir de 1795, marcando o início de uma trajetória de destaque no cenário portuário brasileiro.

Ao longo das décadas subsequentes, o Porto de Santos testemunhou significativas expansões, modernizando suas instalações e adotando tecnologias de ponta. Esta evolução culminou na criação de terminais especializados para contêineres, carga geral, graneis sólidos e líquidos, conferindo-lhe status como o maior complexo portuário do país. Dono de recordes consecutivos de movimentação de carga, o Porto de Santos não apenas se estabeleceu como um protagonista incontestável no panorama nacional, mas também passou a figurar entre os maiores portos do mundo, consolidando sua posição de destaque no cenário global.

Estrategicamente localizado no litoral de São Paulo, a 70 km da capital, o Porto de Santos se destaca como um ponto crucial de conexão para o comércio nacional, firmando-se como um catalisador essencial para a economia brasileira. O estuário natural que delimita as cidades de Santos e Guarujá, nas ilhas de São Vicente e Santo Amaro, configura-se como o canal de navegação do Porto, tornando-o acessível por meio de infraestrutura aquaviária, dutoviária e ferroviária.

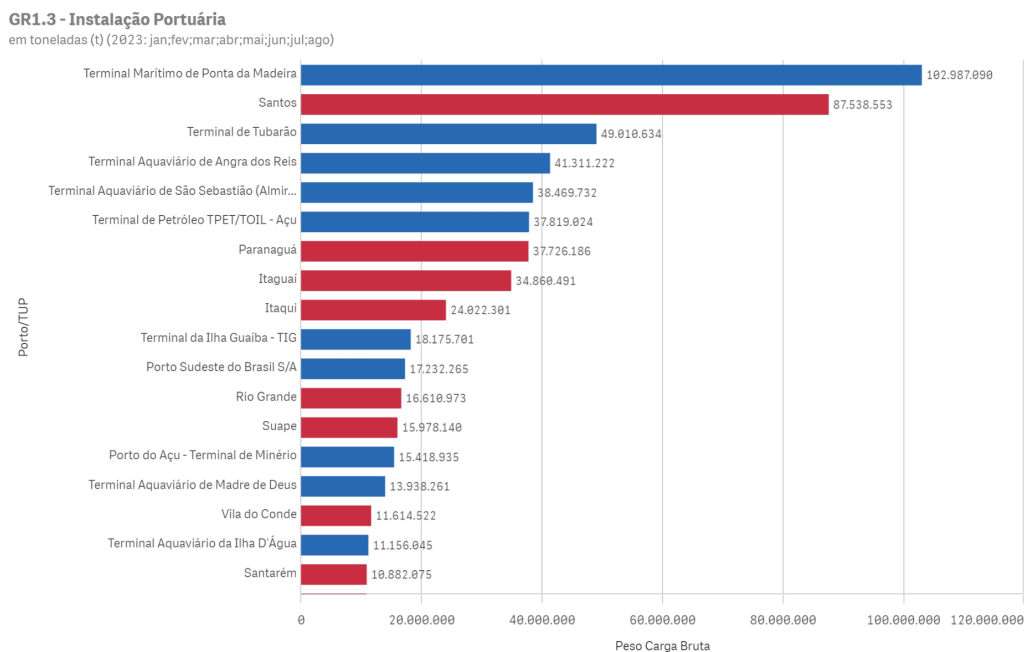
Desse modo, o Porto de Santos não apenas atende às demandas e necessidades de uma extensa porção do território nacional, mas também se destaca como um elemento vital da infraestrutura logística do Brasil, desempenhando um papel preponderante na integração econômica e no desenvolvimento sustentável do país. Sua trajetória histórica e evolução tecnológica não apenas refletem a capacidade de adaptação a diferentes contextos econômicos mas, também, evidenciam a relevância contínua desse porto icônico para a prosperidade e progresso do Brasil.

1.1.2 Panorama Operacional

Santos é uma das principais portas de entrada e saída de mercadorias do país, movimentando cargas de todos os estados brasileiros e de mais de 200 países. Comparando com os portos brasileiros, o Porto de Santos é o segundo maior porto no geral (portos públicos e privados) em relação a movimentações portuárias e o maior porto

público, como mostra a Figura 1.

Figura 1: Movimentações Portuárias (Janeiro a Agosto de 2023)



Fonte: de Transportes Aquaviários, s.d..

O Porto de Santos é um porto de destaque com desenvolvimento sólido e sustentável (crescimento médio de 5% ao ano entre 2011 e 2020 em toneladas movimentadas). Possui uma média de 4904 navios navegados por ano e seu perfil de carga pode ser conferido na Figura 2.

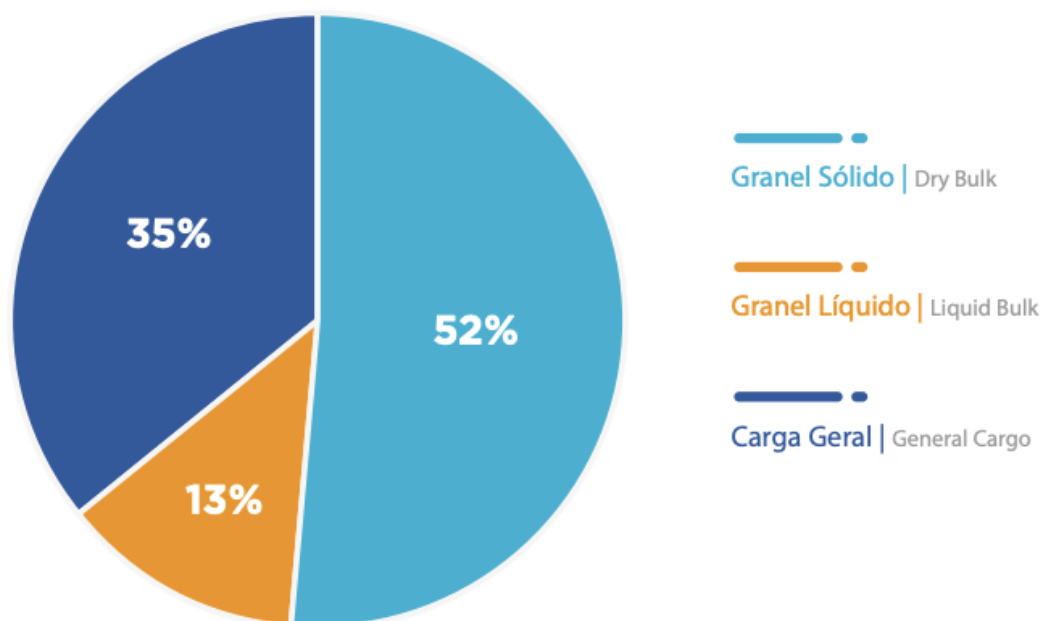
O Porto de Santos é servido por uma rede de transportes completa e integrada de modais rodoviário (55%), ferroviário (33%) e dutoviário (5%), oferecendo a solução mais eficiente do transporte brasileiro.

1.1.3 Acesso e Conexão com o Porto

Como mencionado anteriormente, a complexidade logística do Porto de Santos se destaca pela sua abrangência em diversos modais de transporte, desempenhando um papel essencial na eficiência e agilidade das operações comerciais. Dentre os sistemas de acesso direto ao porto, destacam-se os modais rodoviário e ferroviário, que formam a espinha dorsal desse gigantesco *hub* logístico.

O acesso rodoviário ao Porto de Santos é viabilizado principalmente pelo Sistema Anchieta-Imigrantes (SAI), composto pelas rodovias SP-160 (Imigrantes) e SP-150 (Anchieta). Além destas, as rodovias SP-41 (Interligação Planalto), SP-55 (Padre Manuel da Nóbrega), SP-248 (Cônego Domenico Rangoni) e SP-59 (Interligação Baixada) desempenham papéis de exímia importância na conectividade rodoviária que

Figura 2: Perfil de carga do Porto de Santos



Fonte: Heylin, 1972.

Figura 3: Acesso Local

	10.000 caminhões por dia trucks per day	20 km acessos internos inner accesses
	1.200 vagões por dia wagons per day	100 km acessos internos inner accesses
	7,6 mi toneladas/ano tonnes/year	55 km acessos internos inner accesses

Fonte: Heylin, 1972.

permeia a região. Das citadas, as duas primeiras assumem um papel vital na ligação entre a planície e o planalto, cruzando, de forma paralela, a imponente Serra do Mar. A Interligação Planalto, por sua vez, conecta essas duas rodovias ainda no planalto, enquanto as três últimas se mostram pilares estratégicos na distribuição eficiente de veículos até as cidades circunvizinhas e nas conexões com outras importantes rodovias.

Esse intrincado sistema rodoviário, conhecido como um dos maiores corredores de escoamento do comércio exterior brasileiro, não apenas facilita, mas também otimiza o transporte de cargas entre São Paulo, o centro econômico do país, e o Porto de Santos. A conexão eficaz do sistema citado anteriormente em menos de uma hora representa um feito logístico notável, proporcionando uma integração ágil e eficiente que contribui significativamente para o dinamismo do comércio internacional brasileiro. Essa infraestrutura robusta e interligada é fundamental não apenas para a competitividade do Porto de Santos, mas também para o progresso econômico sustentável do Brasil.

Já o Sistema Ferroviário é composto por principalmente 3 vias, sendo eles:

- **Sistema Ferroviário da Rumo:** A Rumo possui três malhas ferroviárias no Estado de São Paulo – a Malha Paulista, a Malha Norte e a Malha Oeste. As três malhas se interligam e acessam o Porto de Santos por meio da Malha Sul.
- **Corredor Centro-Sudeste da VLI:** a empresa opera a Ferrovia Centro Atlântica, que faz a conexão com a Malha Paulista da Rumo na cidade de Estrela d'Oeste; ou seja, a partir dessa conexão as composições da VLI pagam o direito de passagem à Rumo pelo uso da ferrovia.
- **Estrada de Ferro Santos-Jundiá – MRS:** a companhia opera potentes locomotivas Stadler no sistema de cremalheira, movimentando as cargas entre Santos e São Paulo – estação Brás.

O modal ferroviário responde por aproximadamente 30% do transporte das cargas movimentadas no Porto de Santos. A cada ano, essa participação tem aumentado progressivamente, a partir da unificação de procedimentos operacionais e de melhorias no atendimento aos clientes.

1.1.4 Operações Portuárias

O funcionamento integralmente contínuo do Porto de Santos é uma manifestação do comprometimento da Autoridade Portuária de Santos (SPA) em atender plenamente às demandas da logística naval, garantindo uma operação organizada e eficiente. A SPA desempenha um papel muito importante na gestão do porto, supervisionando desde a definição da programação de atracações e desatracações até o monitoramento constante das condições de navegabilidade do canal de navegação, abrangendo as condições marítimas e climáticas, assim como a profundidade dos berços de atracação.

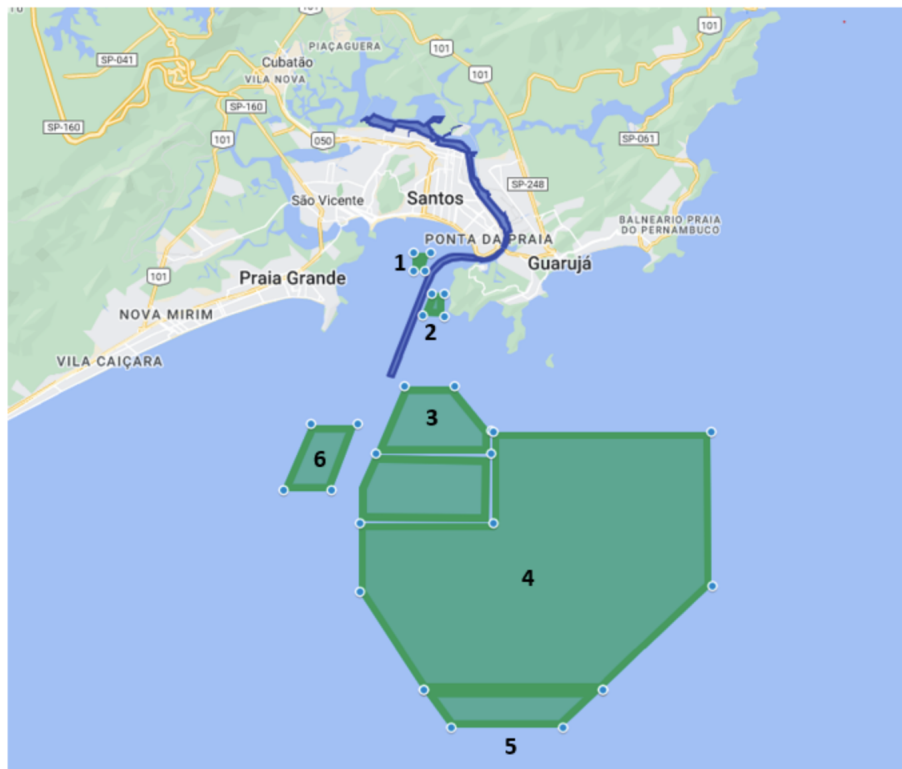
A fiscalização constante das atividades portuárias exercida pela SPA evidencia a complexidade das operações realizadas, uma vez que coordena e organiza diversas

manobras simultâneas. A garantia da fluidez dessas operações é essencial para que o Porto de Santos mantenha sua reputação em eficiência e produtividade. Os técnicos da SPA são essenciais nesse processo, assegurando que as operações ocorram de maneira regular, eficiente, segura e em conformidade com os padrões ambientais estabelecidos.

Ao lidar com as embarcações que chegam ao porto, a SPA gerencia duas áreas distintas para a estabilização das mesmas: a área de fundeio e as regiões de dragagem, também conhecidas como berços de atracação. A área de fundeio (ou fundeadouro) funciona como um atracadouro pré-estabelecido e regulamentado e é destinada às embarcações que aguardam o momento adequado para entrar nos berços de atracação. A área de fundeio desempenha um papel estratégico na gestão eficiente do fluxo de navios que chegam ao porto, estejam eles "adiantados" em relação ao horário reservado para atracação, ou em fila de atracação. O planejamento e a coordenação cuidadosa demonstram o compromisso da SPA em manter o Porto de Santos como um ponto de referência em termos de operações portuárias no cenário nacional e internacional. A área de fundeio é composta por 6 fundeadouros, onde cada um possui as suas particularidades:

- **Fundeadouro nº 1:** Destinado a navios de guerra
- **Fundeadouro nº 2:** Destinado a navios que necessitem efetuar inspeção sanitária ou desembarço (desembarque e embarque de tripulantes, oficinas e materiais), com tempo de permanência não superior a 3 horas
- **Fundeadouro nº 3:** Destinado a navios com programação definida de atracação para as próximas 24 horas
- **Fundeadouro nº 4:** Destinado a navios com ou sem programação de atracação
- **Fundeadouro nº 5:** Destinado a navios com cargas inflamáveis ou explosivas
- **Fundeadouro nº 6:** Destinado a navios de quarentena e embarcações com suspeita de avaria no embalado e/ou vazamento de material radioativo

Figura 4: Fundeadouros do Porto de Santos



Fonte: Santos, s.d..

O Porto de Santos possui diversos berços de atracação com diferentes finalidades, principalmente no que se diz respeito ao perfil de carga dos produtos movimentados. O Porto possui em torno de 60 berços, estes sendo públicos ou privados. A Tabela 1 retrata a distribuição dos berços pelo perfil de carga e suas quantidades, sendo a categorização do berço, em determinado perfil, caracterizada pela movimentação predominante da embarcação.

Essa diversidade de berços se relaciona com os diferentes tipos de embarcação em movimento na região do porto durante todo o funcionamento dele. As principais embarcações são: Navios contêineres, Embarcações de pesca, Navios cargueiros de diversos tipos (*Dry Bulk*, *Dry Breakbulk*, *Wet Bulk*,...), Navios de passageiros e Embarcações de suporte.

A localização tática do Porto de Santos, embora proporcione uma vantagem logística inigualável, também impõe desafios significativos relacionados à sedimentação. A região recebe um volume substancial de sedimentos, provenientes tanto da Serra do Mar, no fundo do canal, quanto das marés e ressacas, na entrada da barra. Esses sedimentos, se acumulados, têm o potencial de causar assoreamento, comprometendo as cotas estabelecidas e, conseqüentemente, afetando a capacidade de navegação e operação do complexo portuário.

Tabela 1: Tabela dos berços de atracação

Perfil de Carga/Produtos movimentados	Quantidade
Granéis líquidos – Químicos/Combustíveis	8
Carga Geral – Contêineres	10
Granéis sólidos – Granéis sólidos vegetais (soja, milho, farelo, trigo)	14
Granéis sólidos – Granéis minerais de desembarque (fertilizantes, enxofre, sal)	4
Granéis líquidos – Sucos cítricos	3
Passageiros	1
Carga geral – Celulose	9
Carga geral – Produtos siderúrgicos	5
Carga Geral – Veículos	2
Berços públicos multipropósito	3
Carga offshore	1
Total	60

Fonte: Santos, s.d..

Diante desse cenário, a prática contínua de dragagem torna-se imperativa para a manutenção da profundidade adequada no complexo portuário. A dragagem não é apenas uma necessidade, mas um processo permanente, essencial para garantir a acessibilidade segura e eficiente aos navios que chegam ao Porto de Santos. Esse procedimento envolve não apenas a remoção dos sedimentos mas também a realização de levantamentos hidrográficos detalhados, visando obter informações precisas sobre as características físicas do leito marinho.

A combinação de dragagem e levantamentos hidrográficos é fundamental para calcular de forma acurada o volume de material a ser dragado, permitindo um planejamento eficiente e otimizado. A busca constante por maior eficiência nesses serviços é uma prioridade, não apenas para atender às demandas de navegação e operação do porto mas para mitigar riscos associados ao assoreamento, reduzir custos operacionais e proporcionar maior agilidade na execução das atividades.

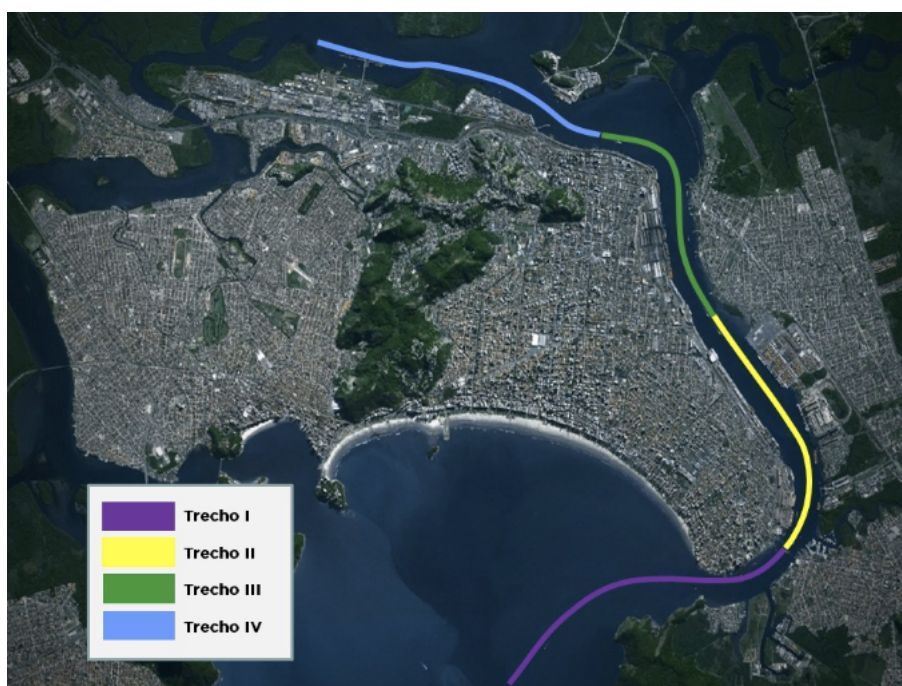
Assim, a gestão proativa da sedimentação no Porto de Santos reforça a importância da dragagem como uma prática vital e destaca a necessidade contínua de inovação e aprimoramento nos processos desse importante complexo portuário brasileiro, visando garantir sustentabilidade e eficiência a longo prazo.

O canal de navegação é subdividido em 4 trechos para efeitos de planejamento de

dragagem, ilustrados na Figura 5, sendo eles:

- **Trecho 1:** Barra até Entrepasto de Pesca
- **Trecho 2:** Entrepasto de Pesca à Torre Grande
- **Trecho 3:** Torre Grande até Armazém 06
- **Trecho 4:** Armazém 06 até o Terminal Alemoa e início do Canal de Piaçaguera

Figura 5: Trechos do Canal de Santos



Fonte: Santos, s.d..

Por fim, outra operação portuária de grande importância é a praticagem. A praticagem é o serviço responsável em levar as grande embarcações do mar aberto ou amplos espaços para as águas restritas como o canal de Santos. Um canal ou trecho de águas restritas são definidos por espaços confinados com limitações variadas para uma embarcação de grande porte se locomover como: pouca profundidade, perigos submersos, tráfego local, severas restrições de dimensão, correntezas e marés e perigos submersos. Para uma embarcação de grande dimensão e reação lenta (atraso para desviar de um barco durante uma manobra por exemplo), a praticagem é essencial uma vez que, o trajeto inteiro dentro de um canal é de alto risco de acidentes e, nos piores cenários, potencial risco de calamidade. Este serviço é regulado pela Autoridade Marítima e consiste, então, em assessorar o Comandante de uma embarcação no gerenciamento dos riscos inerentes à navegação em águas restritas. A praticagem

tem como objetivo evitar os acidentes ou, quando não for possível evitá-los, ao menos, minimizar suas consequências. Contribuindo, desta forma, para a segurança do tráfego aquaviário, a salvaguarda humana e a prevenção da poluição hídrica.

Figura 6: Figura de uma lancha de Praticagem



1.2 Problemática

Dada a magnitude das operações realizadas no Porto de Santos, torna-se indispensável implementar um mapeamento abrangente do fluxo de movimentações de navios e outras embarcações. Esse mapeamento, detalhando a sequência precisa da fila de atracções e saídas dos berços, oferece uma visão holística e valiosa das dinâmicas operacionais do porto. Esta informação é de extrema relevância, não apenas para monitorar o fluxo constante de embarcações, mas também para extrair *insights* profundos, realizar análises detalhadas e diagnosticar a eficiência global da operação portuária.

O conhecimento preciso da ordem das movimentações de navios permite a obtenção de um panorama abrangente das atividades portuárias, facilitando a identificação de padrões, o entendimento de variações de desempenho e a análise de possíveis gargalos operacionais. Além disso, essa abordagem possibilita distinguir entre situações mais ou menos críticas, contribuindo para a implementação de estratégias específicas de aprimoramento operacional e estrutural.

No entanto, é necessário destacar que, até o momento, não existe uma metodologia amplamente difundida para o processo decisório da operação do Porto de Santos, o que significa que não há uma maneira padronizada de prever, com precisão, quais serão as próximas movimentações de navios. A ausência desse protocolo centralizado pode representar um desafio na otimização efetiva das operações portuárias.

Dessa forma, a implementação de uma metodologia mais robusta e difundida para o processo decisório, que envolva um mapeamento detalhado das movimentações de

navios, se torna urgentemente necessária. Isso não só contribuirá para a eficiência operacional, como abrirá caminho para uma gestão mais estratégica e informada do Porto de Santos, promovendo avanços significativos na performance e na adaptação às demandas dinâmicas do comércio internacional. Assim, este projeto pretende estudar sobre o mapeamento do processo decisório das operações portuárias (tomando o caso do Porto de Santos) por meio da análise de dados a fim de avançar na pesquisa de formas de otimização da eficiência de Portos.

1.3 Objetivos

Tendo em vista o problema caracterizado na seção anterior, o presente projeto tem como motivação promover o mapeamento do processo decisório da operação do Porto de Santos utilizando análise de dados. No contexto atual em que as organizações precisam se adequar a "onda de digitalização", a utilização de dados históricos em decisões estratégicas são de grande valor (ex: modelos de previsão com *Machine Learning* e *Deep Learning*), podendo impactar de forma decisiva em resultados financeiros e econômicos. Assim, este projeto tem como motivação entender padrões e retirar conclusões e informações a partir dos dados do Porto de Santos, de forma a tornar possível a compreensão de quais são as características mais importantes na decisão da ordenação e priorização das movimentações dos navios.

Este projeto se fundamenta na motivação de compreender os padrões subjacentes e extrair conclusões valiosas tomando como estudo de caso o Porto de Santos. A partir de análises utilizando diversas informações e bases de dados, buscou-se um entendimento mais profundo das distintas operações que ocorrem dentro de um ambiente portuário. Nesse contexto, a análise de dados deixa de ser apenas uma ferramenta técnica; e passa a representar uma abordagem estratégica fundamental para promover a eficiência operacional e tomada de decisões assertivas e embasadas.

Dentro desse panorama, a busca por *insights* a partir dos dados do Porto de Santos não se limita apenas à otimização de processos, mas também abre portas para uma visão mais abrangente e estratégica do funcionamento do complexo portuário. Compreender as particularidades e correlações presentes nos dados contribui para uma gestão mais ágil, adaptativa e alinhada com as demandas dinâmicas do comércio internacional. Este projeto, assim, se configura como um passo na direção da modernização e aprimoramento contínuo de operações portuárias.

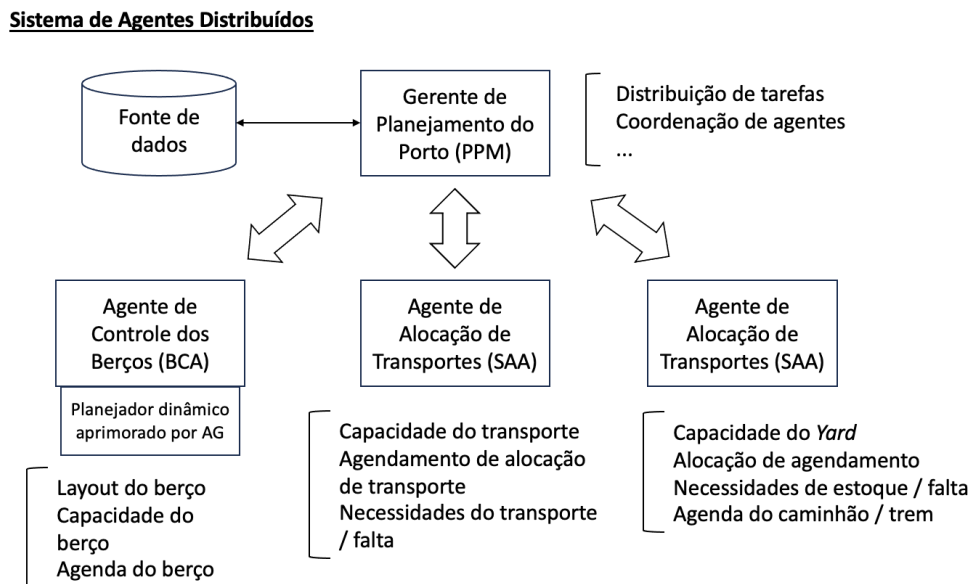
2 Revisão da Literatura

2.1 Sistemas de agentes distribuídos para planejamento de portos e agendamento

Yin et al., 2011 abordou o problema de planejamento do porto e agendamento por meio da utilização de agentes de software. Esses agentes são programas de computador que interagem com o ambiente de software como, por exemplo, sistemas operacionais, bases de dados, sites da internet, entre outros. Agentes podem automatizar tarefas repetitivas, resumir dados complexos de forma inteligente, aprender com outros agentes, fazer recomendações, entre outros.

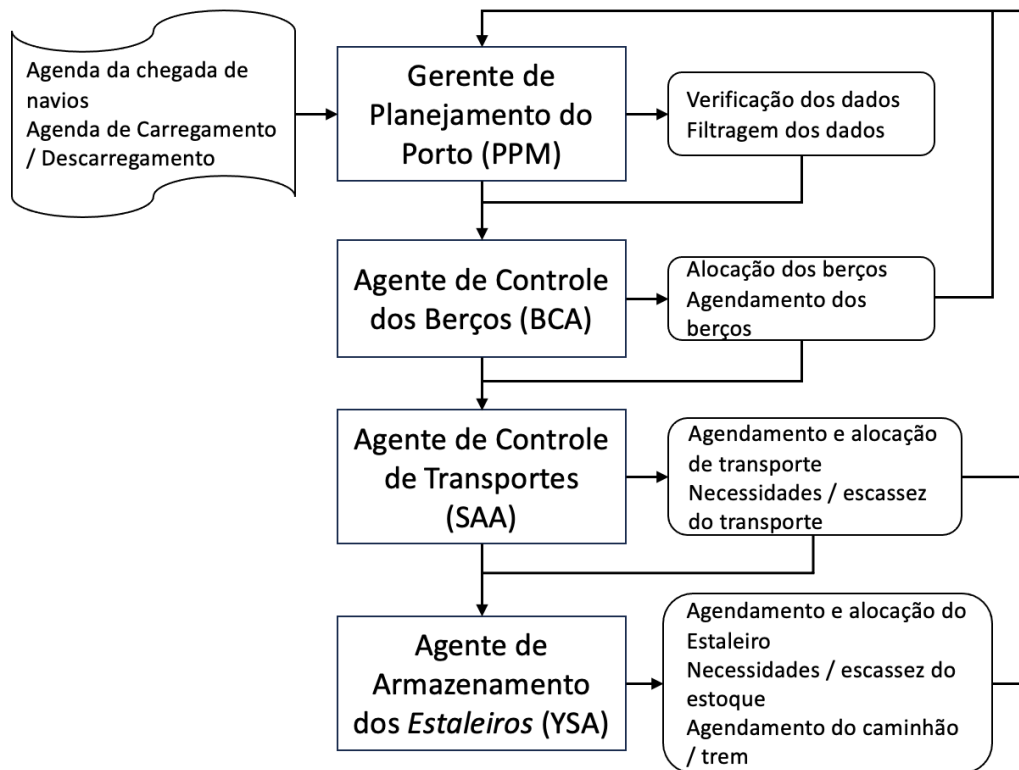
Neste artigo, o método proposto consiste em um sistema de agentes distribuídos para planejamento de portos e agendamento. Este modelo conta com 4 agentes, sendo eles: o Gerente de Planejamento do Porto (PPM), o Agente de Controle dos Berços (BCA), o Agente de Alocação de Transportes (SAA) e o Agente de Armazenamento dos Estaleiros (YSA). A estrutura e o fluxo do trabalho deste sistema está ilustrada nas Figuras 7 e 8 respectivamente.

Figura 7: Sistema para planejamento de portos e agendamento



Fonte: Adaptado de Yin et al., 2011.

Figura 8: Fluxo de trabalho do sistema de agentes



Fonte: Adaptado de Yin et al., 2011.

O Gerente de Planejamento do Porto (PPM) é responsável por uma série de tarefas, sendo elas: receber o agendamento tanto da chegada dos navios quanto dos carregamentos e descarregamentos, verificar e filtrar os dados (depois transfere as informações de chegada de navio válidas para o BCA, desencadeando o reagendamento do BCA, SAA e YSA), recuperar todos os agendamentos gerados pelos sistemas BCA, SAA e YSA, realizar uma comparação pareada utilizando os esquemas recuperados na etapa anterior para identificar possíveis conflitos, instruir o BCA a realizar reagendamentos, impondo restrições conforme necessário para harmonizar as operações portuárias e, por fim, repetir as etapas (3 a 5) até que uma programação livre de conflitos para toda a porta seja estabelecida. A persistência nesse processo é de extrema relevância para garantir uma operação portuária eficiente e sem entraves, demonstrando o papel central do Gerente de Planejamento na gestão estratégica do Porto.

Já as atividades do Agente de Controle dos Berços (BCA) se concentram em: aguardar a notificação do PPM, determinar se as informações canalizadas acionam um reagendamento na alocação do berço caso necessário, calcular o tempo de processamento para carregar ou descarregar um navio e, então, o planejador dinâmico

aprimorado por AG (Algoritmo Genético) é invocado para reprogramar o plano de alocação de berços (considerando restrições como a capacidade do berço e o plano de alocação atual) e, finalmente, o BCA atualiza o PPM e o SAA sobre qualquer alteração no cronograma resultante do reagendamento. Esse processo integrado demonstra a interdependência entre BCA, PPM e SAA para otimizar a alocação de berços e manter a eficiência operacional do porto.

No caso do Agente de Alocação de Transportes (SAA), suas principais tarefas são: aguardar notificações tanto do PPM quanto do BCA, determinar se as informações do BCA justificam um novo cálculo na alocação de transporte, em caso afirmativo, a tarefa seguinte é instanciada, envolvendo o cálculo das necessidades e carências, caso contrário a tarefa anterior se reinicia (aguardar pela notificação do PPM e BCA) para, em seguida, fornecer *feedback* sobre a falta de transportadores para o PPM e BCA, se aplicável, e, por último, atualizar o PPM e o YSA com o cronograma final de alocação de transportadores. Essa sincronização é essencial, pois o YSA, Agente de Sincronização de Atividades, ajustará seu plano de alocação de acordo com as informações fornecidas pelo SAA.

O Agente de Armazenamento dos Estaleiros (YSA) monitora constantemente as mudanças nas agendas extraídas pelas saídas do BCA e SAA. Ajusta a alocação de armazenamento e agenda além de gerar as necessidades ou carências do armazenamento e alerta o PPM de acordo. O YSA cria uma programação preliminar para o caminhão ou trem utilizado transportar os contêineres (ex: Essas informações podem ser usadas como um cronograma de entrega provisório para os clientes).

Através da metodologia de estudo de caso, o estudo chega à conclusão de que os agentes foram capazes de se comunicar e cooperar uns com os outros para elaborar o cronograma de alocação de berços e os requisitos para transportador. O PPM é capaz de manter todas as informações necessárias na fonte de dados de agendamento para o agendamento de instalações portuárias e fornecer um local de negociação e comunicação com o BCA, SAA e YSA, compartilhando as informações e resolvendo seus conflitos sob sua supervisão. O programador dinâmico aprimorado por GA do BCA demonstrou ser capaz de operar com eficácia na alocação de berços quase ideal e no cronograma de carregamento e descarregamento do navio. Serão realizados trabalhos adicionais visando resolver o problema de escassez de transportador, permitindo comunicações e negociações mais frequentes entre o SAA e o BCA. Além disso, uma melhoria para o sistema de agentes distribuídos proposto seria aprimorar ainda mais o escalonador dinâmico baseado em GA, para que pudesse ser adaptado e usado como mecanismo de otimização para SAA e YSA.

2.2 Modelo de agendamento de navios em canais de mão única

Neste artigo, o processo a ser otimizado consiste em reduzir ou até eliminar as inconsistências presentes no cronograma de coordenação do canal e do cais do porto. A capacidade do canal é um dos gargalos para o desenvolvimento do porto quando o tráfego no porto se aproxima de sua capacidade máxima. Assim, este estudo tem como objetivo endereçar o problema da coordenação da agenda do berço e do canal no porto. Zhang et al., 2017 analisa o problema pelo aspecto de serviço de tráfego de embarcações (VTS - *Vessel Traffic Service*) ao conceber um modelo de tempo total de programação e de tempo de espera, um modelo de inicialização, um modelo de atribuição de faixas horárias e um modelo de conversão do fluxo de tráfego. O modelo criado utiliza as seguintes considerações: Canal de mão única (uma embarcação é permitida a passar pela via em apenas uma direção por vez), Alocação de *slots* de tempo (Intervalo de tempo entre uma embarcação que entra ou sai e a seguinte) e a Conversão do Fluxo (Interconversão do fluxo de chegada e saída de embarcações). Para a abordagem deste problema de otimização o agendamento de navios, Zhang et al., 2017 utilizou um algoritmo genético de multiobjetivo (MOGA) a partir dos modelos matemáticos a seguir.

$$\min[F, W] \quad (1)$$

$$F = \sum_{i=1}^n (f_i - f_{i-1}) + f_1 - b_0 - h_0 \quad (2)$$

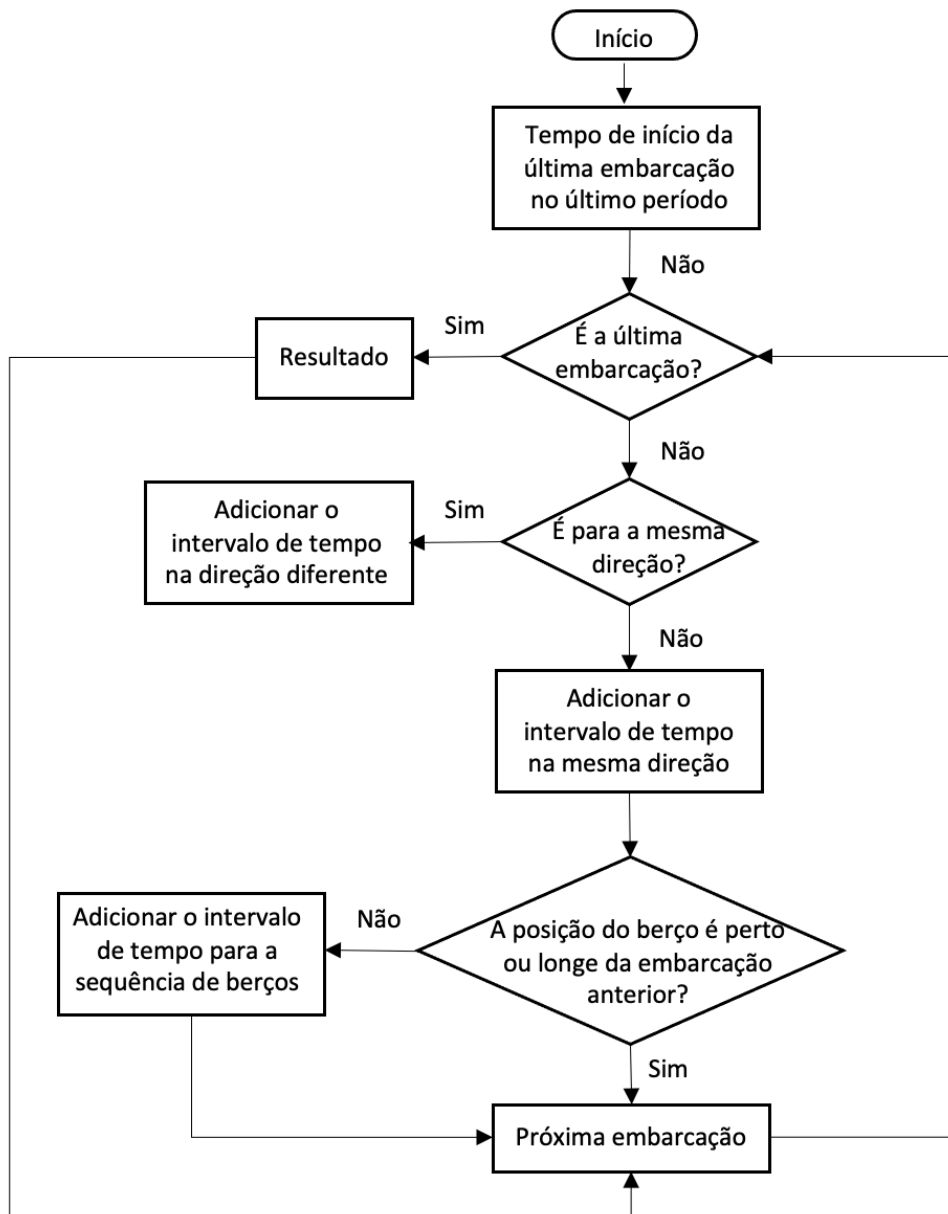
$$W = \sum_{i=1}^n (b_i - t0_i - \frac{s + s_i}{v_i}) + \max(b_i - t0_i) - \min(b_i - t0_i) \quad (3)$$

Onde i é o índice do navio, n é o total de navios, o tempo inicial e final do navio entrando ou saindo é b_i e f_i respectivamente, relacionado ao último estágio, b_0 é o tempo inicial do último navio no último estágio, $t0_i$ é o tempo que o navio i chega na *report line*, s é a distância entre a *report line* e a entrada da via navegável, s_i é a distância entre a entrada da via navegável até o berço alocado para o navio i , v_i é a velocidade média no navio i entre a entrada da via até o berço e h_0 é a folga de tempo segura entre o último navio do último estágio com o primeiro navio do estágio atual.

Além disso, o artigo considera diversas restrições que cujos conceitos serão citados brevemente. As restrições consideradas são: o valor numérico do tempo agendado do navio deve ser depois do que o valor numérico do tempo em que este atinge a entrada da via, o controle do intervalo de tempo de segurança associado para cada navio, o tempo atual não é menor que o tempo teórico para qualquer navio, a limitação do intervalo de tempo de segurança entre o horário de partida para dois navios, de

modo que não seja menor que o intervalo de tempo teórico em direções diferentes e navios se movimentando na mesma direção precisam ter alocações de espaço e tempo.

Figura 9: Fluxograma do algoritmo



Fonte: Adaptado de Zhang et al., 2017.

A partir disso, para obter uma solução adequada, calcula-se uma solução para o Algoritmo Genético de Multi Objetivos (MOGA) utilizando *MATLAB*. Os resultados calculados foram comparados com o método FCFS (*First Come, First Served*), isto é, utilizam a ordem de chegada ao porto como ordem de passagem no canal. Os

resultados foram de que os tempos de conversão de capacidade das embarcações entrando e saindo do porto podem ser reduzidos em nove vezes e o tempo total de agendamento pode ser reduzido em mais de 40%. Desta maneira, o método proposto garantiu que as embarcações entrassem e saíssem do porto de forma eficiente. Porém, por se tratar de um modelo simplificado, um modelo mais preciso e mais adequado precisa ser pesquisado no futuro. Além disso, também podem ser consideradas algumas situações específicas em que as embarcações estão sujeitas a esperar que a maré satisfaça o calado para passar pelo canal e a alocação de rebocadores pode ser levada em consideração para a construção de um modelo mais completo.

2.3 Classificação do período de estadia no porto

Para a melhoria da eficiência da operação de Portos, Abreu et al., 2023 desenvolveu uma metodologia baseada na classificação do período de estadia dos navios por meio dos dados. Este artigo realiza a previsão do tempo de estadia nos portos brasileiros. Assim, para o problema deste artigo é considerado o cálculo do tempo de estadia como:

$$ST = T_1 + T_2 + T_3 + T_4 \quad (4)$$

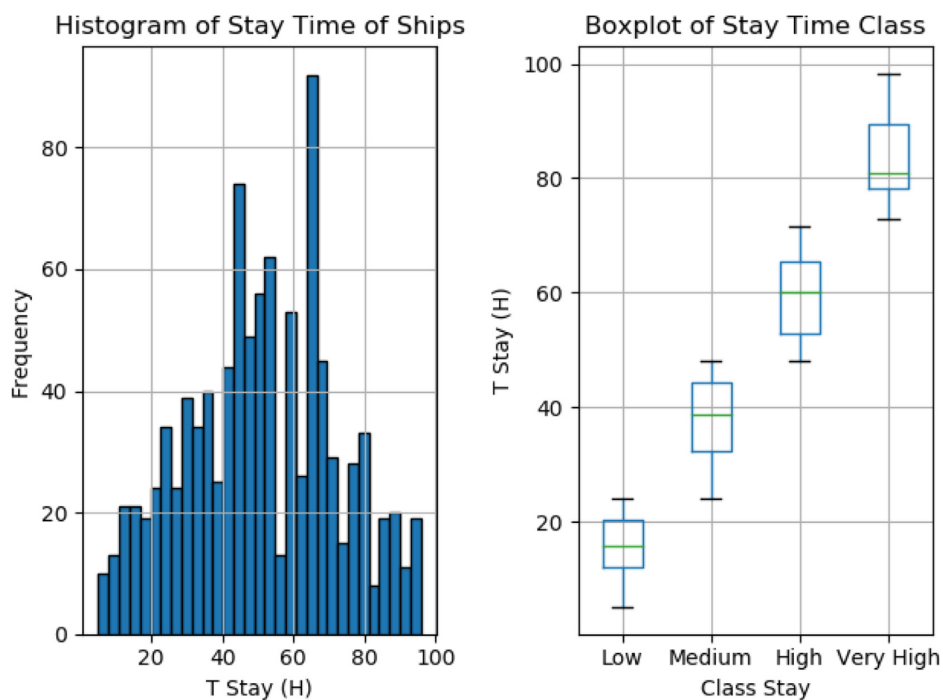
Onde T_1 é o tempo de movimentação pelo canal de acesso e qualquer tempo de espera para atracação, T_2 é o tempo quando o navio atracado espera para a operação carregamento/descarregamento começar, T_3 é o tempo de carregamento/descarregamento (operação) e T_4 é o tempo do navio espera no berço até a desatracação.

A metodologia delineada por Abreu et al., 2023, abrange cinco fases: Coleta de Dados, Limpeza dos Dados, Análise Exploratória dos Dados, Criação e Teste do Modelo e Análise dos Resultados e Impactos na Logística. Inicialmente, a etapa de coleta de dados foi conduzida através do acesso às informações disponíveis no site da ANTAQ (Agência Nacional de Transportes Aquaviários). Essa abordagem resultou em cinco bases de dados abrangendo informações cruciais sobre Entregas (como origem, destino e tipo de navegação), Contêineres (o carregamento em si e o seu peso), Berços (tipo de berço, datas e operação do porto), Atracadouros (relata as características cartográficas dos atracadouros, a região hidrográfica e a quantidade de carga comercializada) e Tempos (os períodos de estadias de cada movimento de carga na base de dados). Estas bases proporcionaram uma nova base robusta para a análise subsequente.

Posteriormente, a etapa de limpeza dos dados foi conduzida para otimizar a qualidade do modelo preditivo. Colunas com características indesejadas, como valores constantes, *outliers*, dados faltantes, alta cardinalidade, multicolinearidade e distorções excessivas, foram removidas para assegurar a integridade e eficácia do modelo.

Com base nas análises, foram realizados o desenvolvimento e teste do modelo. As *features* mais relevantes foram identificadas para a construção de modelos de previsão de tempo de estadia de navios no Brasil. Os modelos *Random Forest Classifier* e *Decision Tree Classifier* foram implementados e comparados com modelos frequentemente utilizados na literatura (como *Gradient Boosting Classifier*, *K Neighbors Classifier* e outros como mostra a Tabela 2), com avaliação baseada em métricas de desempenho (Acurácia, Revocação, Precisão, *F1-Score*) e tempo de execução.

Figura 10: Análise da Distribuição do Tempo de Estadia



Fonte: Abreu et al., 2023.

A etapa de análise exploratória dos dados foi conduzida de forma abrangente, permitindo a compreensão aprofundada das operações portuárias. Além disso, foi realizada a classificação dos tempos de estadia em categorias específicas - Baixo, Médio, Alto e Muito Alto - por meio de intervalo de quartis. Esta transformação pode ser observada pela Figura 10, na qual o gráfico de histograma da esquerda mostra a distribuição do tempo de estadia pela frequência ocorrida na base e o gráfico de *BoxPlot* da direita mostra a distribuição entre as categorias segmentadas do tempo de estadia.

Finalmente, através da análise comparativa dos resultados no qual foram utilizados 15 modelos diferentes indicado na Tabela 2, destacou-se o desempenho prático dos modelos *Random Forest Classifier* e *Decision Tree Classifier* no contexto do planejamento portuário. Ambos obtiveram um bom desempenho, com a liderança do primeiro, entretanto o segundo destaca-se pelo tempo de execução. Para estudos futuros, o au-

Tabela 2: Comparação dos modelos utilizados

Model	Accuracy	Recall	Prec.	F1	TT (s)
Random Forest Classifier	0.739	0.736	0.739	0.739	14.147
Decision Tree Classifier	0.728	0.726	0.727	0.727	0.667
Stochastic Gradient Descent	0.678	0.306	0.180	0.191	44.49
Gradient Boosting Classifier	0.675	0.642	0.680	0.667	73.481
K Neighbors Classifier	0.673	0.655	0.674	0.672	4.965
Linear Discriminant Analysis	0.612	0.581	0.615	0.601	1.792
Ridge Classifier	0.599	0.541	0.602	0.563	0.167
Logistic Regression	0.593	0.530	0.581	0.543	45.806
Ada Boost Classifier	0.545	0.514	0.535	0.534	4.710
Cat Boost Classifier	0.528	0.504	0.505	0.505	18.710
Artificial Neural Network	0.525	0.503	0.525	0.512	12.710
SVM	0.491	0.455	0.536	0.446	13.070
Naive Bayes	0.460	0.510	0.494	0.415	0.178
Quadratic Discriminant Analysis	0.386	0.428	0.458	0.374	0.712
ZeroR	0.364	0.250	0.132	0.194	0.077

Fonte: Abreu et al., 2023.

tor Abreu et al., 2023 recomenda a testagem de outros modelos de previsão como modelos de *Deep Learning* na tentativa de melhorar as métricas de desempenho. Além disso, o modelo pode ser testado com dados mais recentes e em um ambiente real para que a qualidade do modelo seja verificada. E, por último, é recomendado o estudo de outros fatores que influenciem no tempo de estadia do navio assim como a utilização um modelo de regressão ao invés de se limitar a técnicas de classificação.

2.4 Modelos de *Machine Learning* para Previsão do Fluxo de Tráfego de Embarcações

Esse artigo abordou o problema de previsão do fluxo de tráfego de embarcações (*VTFF - Vessel Traffic Flow Forecasting*) por duas abordagens: indireta - por meio da aplicação da previsão da rota da embarcação ao utilizar as localizações das embarcações no futuro, e direta - como um problema de previsão de fluxo sequencial. Diferente da literatura vista no artigo, que foca no comportamento das embarcações em lugares específicos de interesse, Mandalis et al., 2022 tem como objetivo o propósito geral, cobrindo, também, o mar aberto.

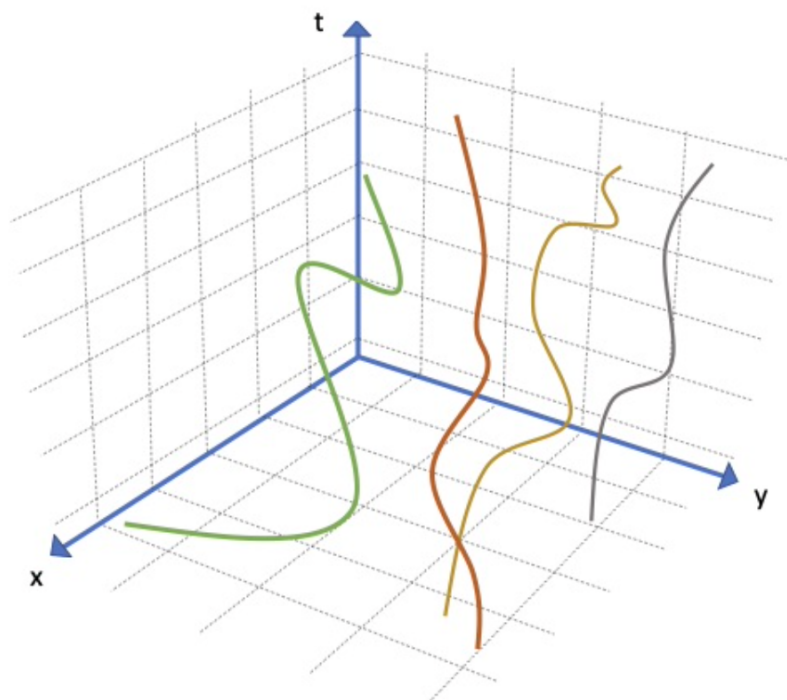
O problema VTFF proposto no artigo foi formulado como:

- **Dado:**

- um conjunto de trajetória de embarcações D abrangendo em D_s (conjunto delimitador mínimo de locais) no espaço e D_t no tempo

- o tempo de duração (horizonte de predição) Δt
- o número de transições temporais r
- o *grid* espaço-temporal (em 3D) que divide a) D_s em células do *grid* de resolução $G \times G$ e b) $D_t \cup \Delta t$ em trechos de tempo r
- **Prediz:** o número esperado de embarcações em cada célula do *grid* relacionado a Δt

Figura 11: Exemplo de um *grid* espaço-temporal de 4×4



Fonte: Mandalis et al., 2022.

2.4.1 VTFF baseado em VRF (*Vessel Route Forecasting*)

Nessa abordagem, a acurácia do modelo VTFF depende da performance do modelo de VRF (predição da rota das embarcações). Assim, foi aplicado o modelo de LSTM *Long Short Term Memory* que, de acordo com a literatura baseada, obteve o melhor resultado no problema de VRF ao ser comparado a modelos como: Regressão Linear, Métodos baseados em árvore, *Support Vector Machine* para regressão e Redes Neurais. Para a aplicação deste modelo, alguns processos de preparação foram realizados, sendo eles: simplificação estacionária (remover histórico correspondente a velocidade menor do que um nó), eliminação de trajetórias insignificantes (remover

trajetórias compostas por um número baixo de pontos; menos de 10 pontos) e a segmentação da trajetória em um número de sub-trajetórias quando o intervalo de tempo entre dois pontos consecutivos passar um limite inferior proposto (igual a 30 minutos). Dessa maneira, foi possível prever os locais resultado do modelo VRF alocados no *grid* espaço-temporal e determinar o número de embarcações localizadas em cada célula do *grid* obtendo, assim, o fluxo do tráfego numa dada área e período de tempo.

Para a avaliação do modelo, utilizou-se a proporção de 50% – 25% – 25% para as bases de treino, validação e teste, respectivamente. A partir disso, os resultados foram avaliados por duas métricas: Erro Percentual Absoluto da Média Simétrica (*SMAPE* - *Symmetric Mean Absolute Percentage Error*, que é uma métrica de acurácia independente de escala, onde quanto menor o valor, melhor a acurácia do modelo e o coeficiente de similaridade de Jaccard utilizado para medir a similaridade entre o conjunto de embarcações reais e o conjunto previsto em uma célula do *grid* específica. As equações a seguir mostram o cálculo para *SMAPE* e para *Jaccard* onde B e F são o número total de células e de espaços de tempo no *grid*, respectivamente; y e \hat{y} representam o número de embarcações atual e previsto; Y e \hat{Y} são os conjuntos incluindo as embarcações atuais e as previstas, respectivamente.

$$SMAPE = \frac{1}{B} \sum_{b=1}^B \frac{1}{F} \sum_{t=1}^F 2 \frac{|y_{b,t} - \hat{y}_{b,t}|}{|y_{b,t}| + |\hat{y}_{b,t}|} \quad (5)$$

$$Jaccard = \frac{1}{B} \sum_{b=1}^B \frac{1}{F} \sum_{t=1}^F 2 \frac{|Y_{b,t} \cap \hat{Y}_{b,t}|}{|Y_{b,t} \cup \hat{Y}_{b,t}|} \quad (6)$$

2.4.2 VTFF baseado em fluxo sequencial

Nessa abordagem, as trajetórias das embarcações são alocadas no *grid* espaço-temporal para formular a sequência do fluxo de tráfego em cada célula. Este modelo tem como entrada N que é composto pelo número de embarcações n em cada trecho de tempo t no b -ésimo conjunto dentro do *grid* de l conjuntos totais, podendo ser descritos como:

$$N^b = [n_{t-l}^b, \dots, n_{t-1}^b, n_t^b] \quad (7)$$

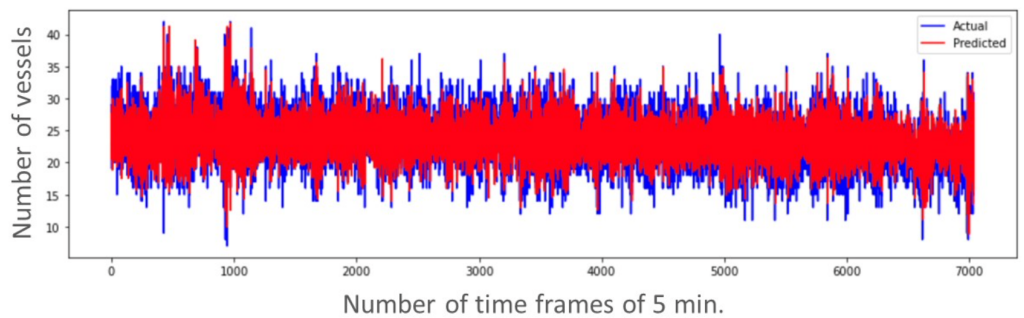
A saída do modelo (número de embarcações previsto) no b -ésimo conjunto do *grid* no trecho de tempo futuro $t + 1$ é \hat{n}_{t+1}^b .

Técnicas de *Machine Learning* diferentes possibilitam o entendimento de diferentes representações de dados. Neste estudo, foi aplicado os modelos *XgBoost* e o *ARIMA*. Não foi preciso realizar processos preparatórios pois, estes modelos podem ser aplicados diretamente na base de dados, já que simplesmente mapeiam a entrada

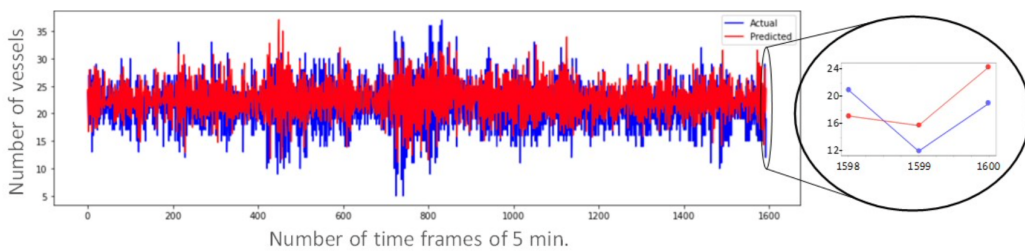
e a saída.

Na avaliação do modelo, para cada célula do *grid*, 75% do fluxo sequencial de tráfego inicial foi usado para treino enquanto os 25% restantes do fluxo foram utilizados para a validação, com exceção das últimas 3 observações. Para a avaliação, foi utilizada a métrica de *SMAPE*, como fora demonstrado o cálculo anteriormente.

Figura 12: Comparação do real x previsto pelo modelo com *Xgboost* em células de *grid* ocupadas para: a) base de treino e b) base de validação e teste



(a)



(b)

Fonte: Mandalis et al., 2022.

Tendo isso em vista, os modelos foram aplicados na base da AIS a fim de comparar os modelos existentes nas 20 células do *grid* mais ocupadas com $\Delta t = 15$ minutos. O *grid* espaço-temporal foi aplicado com $G = 10$ km e a janela de tempo com trechos de tempo de 5 minutos, obtendo os resultados a seguir:

Figura 13: Comparação dos modelos no problema de VTFF

PREDICTION RESULTS (SMAPE) IN THE TESTING SET (20 BUSIEST GRID CELLS), $G = 10\text{KM}$.

VTFF strategy	Method	Time prediction horizon (min)		
		5	10	15
Flow sequence-based	XgBoost	17.72	30.41	27.43
	ARIMA	46.94	37.75	48.73
VRF-based	LSTM	6.35	16.76	28.71

Fonte: Mandalis et al., 2022.

A partir da Figura 13, podemos complementar a análise comparativa ressaltando que o modelo baseado em VRF com LSTM não leva em conta os históricos de embarcações com velocidade baixa ou com poucos dados, ao passo que a estratégia baseada em fluxo sequencial é capaz de considerar todo o histórico. Por outro lado, dada a natureza do algoritmo, este apenas consegue prever o volume do tráfego em regiões ocupadas. Também foram analisados os resultados do modelo baseado em VRF com diferentes tamanhos de *grid* e janelas de tempo. Como o modelo é afetado pela granularidade, o melhor resultado foi obtido com o maior *grid* (no caso, de 15 km) como mostra a Figura abaixo.

Tabela 3: Resultados do modelo baseado em VRF com diferentes parâmetros

PREDICTION RESULTS (SMAPE, JACCARD) FOR THE VRF-BASED VTFF STRATEGY IN THE TESTING SET (ALL GRID CELLS) .

Grid cell (km)	Time frame (min)	SMAPE	Jaccard
5	5	9.57	0.95
	10	26.20	0.87
	15	44.00	0.78
10	5	4.97	0.97
	10	14.23	0.93
	15	24.90	0.87
15	5	3.52	0.98
	10	10.08	0.95
	15	18.04	0.91

Fonte: Mandalis et al., 2022.

Como observações finais, o autor comenta sobre a eficiência de ambas as abordagens, com seus respectivos modelos de predição do fluxo do tráfego na área marítima para tempos de horizonte curtos até $\Delta t = 15$ minutos. Porém, como próximos passos, menciona a investigação de outros dados como informações climáticas e mais parâmetros de tráfego para relacionar com o problema de VTFF. Além disso, Mandalis et al., 2022 planeja estudar a acurácia dos modelos propostos em um horizonte de tempo mais amplo e/ou com células de *grid* menores.

2.5 Análises de entregas e operações do porto usando *Big Data*

Nesse artigo, Bonham et al., 2018 realizou análises para explorar a operação, uso e relacionamentos entre portos no Reino Unido a nível macro, e as características comportamentais e operacionais a nível micro (especialmente relacionamentos nacionais e internacionais, tráfico em portos e fatores relacionados, atrasos na entrada e capacidade de utilização. Os dados utilizados vieram a partir de duas fontes: *Automatic Identification System* (AIS) para dados mais específicos de cada embarcação (como velocidade, posição, destino, entre outros) e *Consolidated European Reporting System* (CERS) para dados mais de nível macro (como tempo esperado de chegada para cada embarcação e porto de destino).

A partir de análises dos dados, o autor concluiu que as visualizações derivadas dos dados da AIS poderiam suportar um entendimento a alto nível sobre a embarcação, como identificação da linha de entrega, berços do porto, carregamento do porto, entre outros. Sendo assim, era considerado de uso limitado para classificação quantitativa comportamental em uma base individual de embarcação para embarcação. Posto isso, foi aplicado um modelo de segmentação não supervisionada (*K-means classifier*) utilizando um recorte da base da AIS durante o período entre Agosto de 2016 até Julho de 2017 (janela de período de 1 ano) com diversas *features* incluindo: Velocidade territorial em nós (*Speed over Ground* - SOG), Aceleração (derivada da velocidade), taxa de rotação em graus por segundo (*Rate of Turn* - ROT) entre outros). O algoritmo foi aplicado em um conjunto de dados de treino contendo pares de velocidade territorial e taxa de rotação para cada momento de cada embarcação. Além disso, para remover ruído do modelo causado por embarcações menores, como Navios rebocadores e balsas, assim, os dados extraídos continham os seguintes tipos de embarcação: navios de contêineres, navios de cargas em geral, navios-tanque de produtos químicos e petrolíferos, navios de cargo e navios graneleiros.

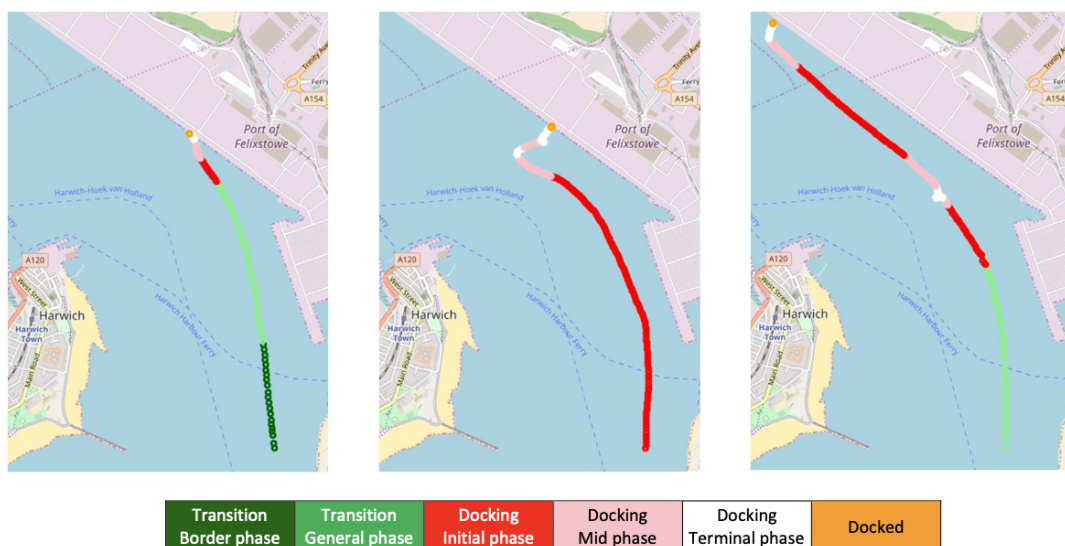
O algoritmo foi aplicado nos dados de treino e o número de segmentos gerados (variável k) foi ajustado para o valor arbitrário de 8 mas a partir da junção de segmentos com centróides próximos e tratativas de interpretação, os segmentos agrupados resultantes foram (6 centróides):

- **Embarcações com comportamento de Transição:** embarcações com maior velocidade média territorial (SOG) e menor taxa de rotação (ROT). Essa categorização englobou 2 grupos (centróides) sendo eles: *Border Phase* (corresponde a 4% da base de treino) e o *General Phase* (corresponde a 57% da base de treino).
- **Embarcações com comportamento de Atracação (processo para ser atracado):** embarcações de velocidade média territorial mediana (SOG) e taxa de rotação mediana (ROT). Essa categorização englobou 3 grupos (centróides) sendo

eles: *Initial Phase* (corresponde a 10% da base de treino), o *Mid Phase* (corresponde a 5% da base de treino) e o *Terminal Phase* (corresponde a 13% da base de treino).

- **Embarcações com comportamento de Repouso (já estão atracados):** embarcações com baixíssima velocidade média territorial (SOG), embarcações em repouso e maior taxa de rotação (ROT). Essa categorização englobou é de um único centroide que corresponde a 12% da base de treino.

Figura 14: Exemplos de classificação comportamental para embarcações atracando



Fonte: Bonham et al., 2018.

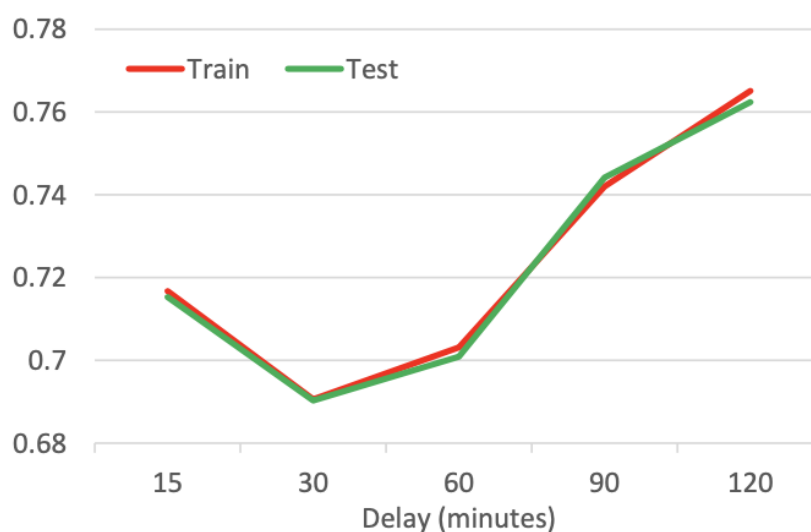
Além da segmentação das embarcações pelo modelo não supervisionado, o autor investigou a probabilidade de uma dada embarcação atrasar no tempo de chegada ao destino. Para isso, os atrasos de cada embarcação ao subtrair o campo de Tempo Estimado de Chegada (*Estimated Time of Arrival* - ETA) e o campo de Tempo Real de Chegada (*Actual Time of Arrival* - ATA). Como o tempo de atraso impacta de forma crítica para diferentes situações, foram estabelecidos 5 campos binários de *target*, cada um relacionado a diferentes limites de atraso, sendo eles de: 15, 30, 60, 90 e 120 minutos. Com isso, o modelo foi criado com diversas *features* que estão agrupadas em:

- **Horário e Sazonalidade (mês, dia, hora,...)**
- **Tipo de embarcação (Peso bruto, *flag* de atraso anterior, tipo do navio,...)**
- **Dinâmica da embarcação (SOG, ROT, aceleração,...)**

- **Classificação da embarcação (resultado do modelo de segmentação das embarcações)**
- **Situação Local (distância até a embarcação mais próxima e número de embarcações próximas)**
- **Separação de Portos (distância entre o porto anterior e o de destino)**
- **Lotação do Porto (número de embarcações nos limites do porto por tipo, por classificação de segmentação e absoluto)**
- **Clima no porto (temperatura, pressão do nível do mar, visibilidade, velocidade do vento, temperatura máxima e mínima,...)**

A partir disso, realizou-se a preparação dos dados para incorporação no modelo. Realizou-se o balanceamento de dados com embarcações atrasadas e não atrasadas e dividiu-se uma parte da base (20 %) aleatoriamente como base para teste (para avaliar a qualidade do modelo). Assim, foram utilizados diversos modelos de *Machine Learning* como: *Random Forest*, *AdaBoost*, *Gradient Boosted Trees* e o algoritmo *XG-Boost*, o qual obteve o melhor resultado dentre eles. Para a avaliação da performance do modelo pelos limites de tempo, primeiro foi utilizada a métrica de acurácia, que é o cálculo da divisão do número de classificações corretas pelo total de casos, obtendo o gráfico da Figura 9.

Figura 15: Gráfico de acurácia do Treino x Teste



Fonte: Bonham et al., 2018.

Focando no modelo de predição de atraso de 15 minutos, outras métricas de avaliação foram calculadas: precisão de 69 % (ou seja, o modelo prediz atraso corretamente

em 69 % das vezes), *recall* de 80 % (o que indica que dos atrasos que aconteceram, o modelo acertou 80 % das vezes). As 5 *features* mais importantes para o modelo foram: distância entre o porto anterior e o de destino, hora do dia, temperatura máxima diária, pressão do nível do mar e peso bruto da embarcação. Além disso, Bonham et al., 2018 realizou análise de *features* notáveis para o modelo a fim de avaliar o impacto individual no aumento ou redução da probabilidade de atraso da embarcação, como mostrado na Figura 10.

Figura 16: Análise univariada das *features* notáveis para o modelo



Fonte: Bonham et al., 2018.

Por fim, o artigo finaliza apontando o trabalho futuro, em quais áreas esses modelos de *Machine Learning* poderiam ser utilizados para obter bons resultados, sendo eles: indicadores iniciais ao PIB (Produto Interno Bruto) para ser possível analisar e apoiar decisões de alocação das operações dos portos de tal região ou país para re-

duzir um possível estresse operacional futuro, simulação de lotação dos portos por meio da previsão de atrasos que ao serem conectados com os dados de porto de destino, concluem a quantidade de embarcações em horários futuros, Planejamento de cenários futuros no geral (não só a lotação do porto, mas para calcular a mão de obra operacional em determinados dias, recursos que precisam ser realocados, preços de mercadorias, realocação de embarcações para outros portos, priorização para embarcações com produtos perigosos entre outros) e Otimização das operações do porto (minimizar o tempo no berço, minimizar o número de embarcações atrasadas, maximizar o número de embarcações processadas em um dia, minimizar o impacto ambiental entre outros).

2.6 Agendamento eficiente online para vias aquáticas restritas

Para melhorar a eficiência do agendamento e sequenciamento de navios, Gan et al., 2021 atacou este problema com uma abordagem diferente de soluções anteriores. Apesar da ASS (Aircraft Sequencing and Scheduling - pegar a referência desta solução) minimizar o atraso do tempo de aeronaves no ar utilizando simulações de Monte Carlo, esta metodologia não pode ser usada de forma direta também para o problema de navios em portos. Isto por conta de 3 pontos principais:

1. Os navios passarão pela via aquática restrita pelos dois sentidos (montante e jusante) enquanto no problema ASS, apenas as aeronaves de pouso eram consideradas.
2. Para o problema ASS o tempo de pouso é ignorado por ser bem rápido em aeronaves. Já para os navios, o tempo de movimentação pode variar entre 10 a 60 minutos, isto não pode ser ignorado por impactar significativamente no resultado do agendamento.
3. Na gestão do tráfego aquático restrito, navios a jusante possuem prioridade para passar por meio da via aquática restrita. Esta é uma restrição pra o problema na otimização do sequenciamento de navios. No problema ASS, todas as aeronaves são tratadas igualmente (sem prioridade) para o pouso no aeroporto.

Gan et al., 2021 propõe um algoritmo de Agendamento e Sequenciamento Online (OSS) para alcançar a sequência de navios e de tempo de passagem ótima quando navios passam em vias aquáticas restritas. Assim, o objetivo é minimizar o tempo total de espera T_{tw} como mostrado na Equação 1.

$$T_{tw} = \sum_{i=1}^N [t_{ATT}(i) - t_{PAT}(i) + t_{ACT}(i) - t_{PCT}(i)] \quad (8)$$

Onde t_{ATT} é o tempo permitido de movimentação, t_{PAT} é o tempo previsto de chegada, t_{ACT} é o tempo permitido para cruzar e t_{PCT} é o tempo previsto para cruzar.

O processo, então, do algoritmo OSS é descrito como:

1. Seja $S_d = a_d^1, a_d^2, \dots, a_d^n$ e $S_u = a_u^1, a_u^2, \dots, a_u^n$ denotando a sequência original a jusante e a montante respectivamente, a_d^i e a_u^j são i-ésimo navio na sequência a jusante e o j-ésimo da sequência a montante. Seus t_{PAT} estão dentro do período de interesse, isso precisa ser considerado no subproblema.
2. Ordenar os navios a jusante de acordo com o t_{PAT} em ordem ascendente. Ajustar seus tempos de movimentação t_{ACT} pela Equação 2 para prevenir ultrapassagens em vias aquáticas restritas.

$$t_{ACT}(i) = \max\{t_{ATT}(i-1) + t_{ACT}(i-1) + t_{PAT}(i) + \delta, t_{PCT}(i)\} \quad (9)$$

3. Se existir tempos desocupados entre navios a jusante, o que significa que o navio anterior deixou a via antes que o seguinte chega, i.e. $t_{PAT}(i-1) + t_{PCT}(i-1) < t_{PAT}(i)$. Atravesse por todos os navios a montante em S_u e ache navios a_u^j que podem passar pela via aquática restrita durante essa folga. Insira a_u^j na folga. O t_{ATT} não precisa ser recalculado já que a_u^j não impacta na sequência dos navios a jusante.
4. Coloque os navios a jusante não agendados no fim da sequência. Organize o primeiro navio a montante com o devido t_{ATT} pela Equação 3 para prevenir colisões.

$$t_{ATT}(i) = \max\{t_{ATT}(i-1) + t_{ACT}(i-1) + \delta, t_{PAT}(i)\} \quad (10)$$

5. Organize todos os navios a montante no final da sequência com t_{ACT} factíveis pela Equação 2 para prevenir ultrapassagens nas vias aquáticas restritas.

A janela deslizante no processo acima é um fator crucial para a redução da onerosidade computacional. Ao analisar apenas uma janela ao invés do período inteiro, este método consegue obter uma solução de forma eficiente. Após alcançar a solução ótima de uma determinada janela, prossegue-se para a janela seguinte, repetindo o processo. Este mecanismo gera um algoritmo com forte robustez dentro de um ambiente dinâmico e incerto devido aos dados *online*, já que mudanças e atualizações que geram variações dos tempos previstos impactam de forma significativa e, normalmente, geram um custo computacional muito grande com outras abordagens.

3 Metodologia

O desenvolvimento do modelo proposto demandou a aplicação de códigos de programação para o tratamento e análise dos dados obtidos da plataforma *marinetraffic*. Através desse processo, buscou-se a obtenção de um histórico detalhado das embarcações selecionadas para realizar a atividade da praticagem, que consiste nas manobras necessárias para atracar nos berços do Porto de Santos. A análise dos dados, portanto, teve como objetivo principal compreender, a partir de uma perspectiva analítica sobre os processos decisórios para o fluxo de sequenciamento das embarcações e características da base de dados portuários tomando o caso do Porto de Santos.

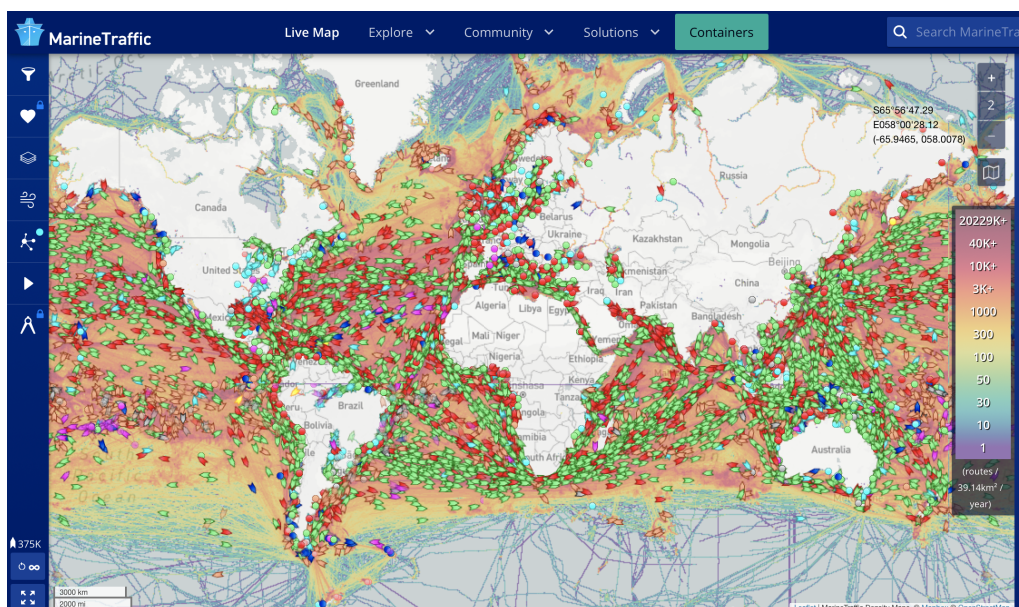
A utilização de códigos de programação permitiu a manipulação eficiente e aprofundada dos dados coletados presentes no site da *marinetraffic* (MarineTraffic, s.d.), fornecendo uma base sólida e praticidade para a análise subsequente. O histórico das embarcações escolhidas para a atividade de praticagem foi analisado, proporcionando *insights* sobre padrões, tendências e características distintivas que influenciaram as decisões de escolha. Essa abordagem, embasada em dados, busca entender os critérios e fatores que desempenham um papel essencial na seleção dos navios, contribuindo, assim, para a compreensão mais aprofundada dos processos decisórios envolvidos na operação portuária. Ao obter o histórico passado de movimentações das embarcações, abre-se a possibilidade não apenas de otimizar o processo de ordenação e priorização mas também de aprimorar a eficiência operacional como um todo.

3.1 Coleta de Dados

A coleta de dados foi feita a partir do site da *marinetraffic*, empresa líder na disponibilização de dados de monitoramento de embarcações e inteligência marítima. A partir de dados recolhidos de uma rede de estações costeiras de recepção de AIS e complementadas por receptores de satélite, são aplicados algoritmos e integradas fontes de dados complementares para fornecer às indústrias de transporte marítimo, comércio e logística informações práticas sobre a atividade marítima.

O site também possui as informações das embarcações em tabelas que é de onde o presente projeto concentra os esforços de coleta de informações. Esta etapa de obtenção dos dados utilizou como base as tabelas fornecidas pelo site realizando o filtro para a região do Porto de Santos (escopo do projeto) e o período de coleta foi dos meses de Julho a Outubro de 2023. Para este propósito, foi elaborado um sistema de captura por meio de código de *webscrapping*, isto é, o processo de coletar dados de sites ou outras plataformas *online*. No contexto deste projeto, a coleta de dados é realizada a cada intervalo de 5 minutos, uma vez que o site constantemente atualiza

Figura 17: Captura da Tela do site *marinetraffic* de acompanhamento global de embarcações



Fonte: MarineTraffic, s.d..

as informações em tempo real. Essa frequência de coleta foi escolhida com o objetivo de capturar as possíveis variações de rotas, velocidades, localizações e estados das embarcações ocorridas neste intervalo de tempo. Dessa forma, a tabela utilizada para armazenar os dados reflete de maneira precisa e dinâmica as mudanças em curso, proporcionando uma visão atualizada e detalhada da situação das embarcações no Porto de Santos. Cada coleta de dados refletiu em um arquivo que era salvo em formato de planilhas csv.

O código utilizado foi escrito na linguagem *Python* e esteve ativo o tempo todo. Para fazer a coleta de dados, foi utilizado o código 1 na página 79.

3.2 Desenvolvimento do Modelo

Com a coleta dos dados em diversos arquivos, a primeira etapa consistiu na integração desses dados, reunindo-os em uma única tabela para viabilizar a análise consolidada do período todo. Este processo de junção estabeleceu a base para todas as etapas subsequentes do projeto.

A partir dessa tabela, os dados coletados foram submetidos a diversos tratamentos, de forma a melhorar a manipulação, facilitar análises e construir o fluxo do modelo do atual projeto (identificação da rota de navegação de cada embarcação). Para isso, o desenvolvimento do modelo seguiu as etapas descritas a seguir.

Primeiramente, realizou-se a etapa inicial de Filtragem da base de dados. Assim,

foram excluídas da base de dados, as embarcações que se enquadram no tipo *Tug*, especificamente as embarcações de rebocadores de porto (*Harbour Tug*). Assim, foram removidas do modelo devido à sua natureza de suporte, que não faz parte do escopo de análise do projeto por não agregarem valor ao processo decisório uma vez que adentram ao canal apenas para oferecer assistência a outras embarcações. Ainda foram eliminadas as embarcações que não continham o código IMO, ou seja, embarcações envolvidas exclusivamente na pesca, navios sem meios mecânicos de propulsão, iates pessoais, navios afetos a serviços especiais (como navios-farol e embarcações de busca e salvamento), barcas de funil, hidrofólios, veículos de almofada de ar, docas flutuantes, vasos e embarcações de guerra e transporte de tropas, barcos de madeira também não estão no escopo deste projeto.

Na etapa subsequente que compreende o Tratamento Inicial, foi realizada a transformação e alteração no tipo de dados na tabela. Campos como 'LAT', 'LON', 'SPEED', 'LENGTH' e 'WIDTH', embora contenham números, eram interpretados como textos, necessitando de uma manipulação para possibilitar a interpretação numérica. Além disso, o formato da data e horário de coleta dos dados não estava adequado para manipulação e foi alterado para garantir a precisão da análise.

Em seguida, a etapa de Criação de Colunas de Estado da Embarcação adicionou informações relevantes à tabela inicial, proporcionando uma visão mais abrangente sobre o estado dos navios. Foram introduzidas colunas específicas relacionadas ao estado da embarcação, sendo elas: Estado do Navio (indicando a condição de movimento da embarcação contendo três categorias distintas: Parado para embarcações com velocidade igual a 0, Manobrando para embarcações com velocidade até 0.5 e Movimentando para embarcações com velocidade maior que 0.5) e Janela (coluna que determina a janela de horário a partir do horário da coleta dos dados, dividindo-se em quatro categorias: Janela 1 para o período entre 1 da manhã e 7 da manhã, Janela 2 para o período entre 7 da manhã e 1 da tarde, Janela 3 para o período entre 1 da tarde e 7 da noite e Janela 4 para o período entre 7 da noite e 1 da manhã).

Posteriormente, realizou-se a etapa de Criação de Colunas de Manipulação Intermediárias, na qual foram introduzidas colunas adicionais para simplificar a identificação de mudanças ou comportamentos fundamentais para o modelo, através da inclusão de *flags* indicativas de elementos específicos presentes em determinados dados. Nesse contexto, foram criadas duas colunas distintas: 'Flag_Mudança_Estado' (Esta coluna atua como uma indicação de se a embarcação alterou seu 'Estado_Navio', facilitando a identificação de momentos em que ocorrem transições de movimento - Parado, Manobrando ou Movimentando) e 'Flag_3Mov' (indica se a embarcação em questão realizou os três tipos de estados, ou seja, se passou pelos estados de Parado, Manobrando e Movimentando durante a análise; caso a embarcação esteja marcada neste campo, pode indicar que a embarcação foi bem ativa durante a coleta de dados).

Essas colunas intermediárias de manipulação proporcionam uma camada adicional de informações, permitindo uma identificação mais eficiente de padrões e comportamentos significativos para o modelo, contribuindo, assim, para uma análise mais detalhada e precisa das atividades das embarcações no Porto de Santos.

Após a etapa anterior, seguiu-se para a Criação de Colunas de Análise, onde foram desenvolvidas colunas mais detalhadas para indicar elementos específicos de interesse para o modelo. A principal adição foi a coluna de Mudança de Estado, que oferece uma visão mais aprofundada das transições ocorridas. Essa coluna categórica apresenta diversos valores possíveis, fornecendo informações detalhadas sobre a classificação da mudança de estado.

- **Parou:** A embarcação estava se movimentando e parou (Estado Movimentando -> Parado)
- **Moveu:** A embarcação estava parada e começou a se movimentar (Estado Parado -> Movimentando)
- **Iniciou manobra de parada:** A embarcação estava se movimentando e começou a manobrar (Estado Movimentando -> Manobrando)
- **Parou manobra e moveu:** A embarcação estava manobrando e parou (Estado Manobrando -> Parado)
- **Iniciou manobra de movimentação:** A embarcação estava parada e começou a manobrar (Estado Parado -> Manobrando)
- **Parou manobra e estacionou:** A embarcação estava manobrando e parou (Estado Manobrando -> Parado)

Na etapa de Identificação de Localização das Embarcações nas Áreas de Fundeio, foi implementado um código para determinar a posição das embarcações nos 6 fundeadouros do Porto de Santos. Essas áreas são estrategicamente designadas para atender diferentes tipos de navios, seguindo as regulamentações portuárias. Essas áreas são determinantes para a operação portuária, funcionando tanto como uma 'fila' (onde as embarcações aguardam sua vez de atracação) quanto como um local de espera para aquelas que ainda não desejam atracar, mas precisam ou preferem aguardar nestes locais. O código, então, identificou em qual dos 6 fundeadouros a embarcação em questão se localiza em um determinado horário, com base em sua posição registrada pelas colunas de 'LAT' (latitude) e 'LON' (longitude), caso estivesse em uma delas.

Já na etapa de Identificação de Localização das Embarcações nas Áreas de Dragagem, o processo é semelhante à etapa anterior, mas concentra-se na identificação dos

locais no canal de dragagem, representando as áreas de atracação das embarcações no Porto de Santos. Dada a complexidade do porto, que abrange diversos terminais e pontos de atracação, esta análise visa determinar precisamente a localização das embarcações utilizando os dados de latitude (LAT) e longitude (LON) presentes na base de dados.

A etapa seguinte foi a de Categorização do Procedimento de Movimentação que, utilizando as colunas de Fundeio e Terminal, desenvolveu processos para categorizar o tipo de movimento - entrada ou saída - de uma embarcação em uma área de fundeio ou área específica de dragagem. Com base nessas informações, foram criadas as colunas de Estado de Entrada para Fundeio e Área de Dragagem, assim como as colunas de Estado de Saída para Fundeio e Área de Dragagem. Essas colunas seguem a lógica a seguir:

- **Estado Entrada Fundeio/Área de Dragagem:** coluna de *flag* que indica que a embarcação não estava em nenhum Fundeio/Área de Dragagem e logo em seguida entrou em um determinado Fundeio ou Área de Dragagem
- **Estado Saída Fundeio/Área de Dragagem:** coluna de *flag* que indica que a embarcação estava em um determinado Fundeio ou Área de Dragagem e logo em seguida não estava mais em nenhum Fundeio/Área de Dragagem

Após essas etapas, foi realizada, então, a etapa de Documentação do Procedimento de Movimentação que, baseando-se nas colunas de categorização dos procedimentos de movimentação, gerando uma documentação detalhada contendo o horário no qual cada movimento estava sendo executado. Para isso, identificou-se o momento de início de cada procedimento, desde o início da operação de entrada até o término da operação de saída. Com esse objetivo, foram criadas quatro colunas de *flag* para identificar cada momento crucial para o modelo, sendo elas:

- **flag_operacao_entrada_inicio:** A embarcação é marcada neste campo quando a coluna de 'Mudanca_Estado' está como "Iniciou manobra de parada" e a embarcação já está na região de algum terminal de dragagem
- **flag_operacao_entrada_fim:** A embarcação é marcada neste campo quando a coluna de 'Mudanca_Estado' está como 'Parou manobra e estacionou' e a embarcação está na região de algum terminal de dragagem
- **flag_operacao_saida_inicio:** A embarcação é marcada neste campo quando sua coluna de 'Mudanca Estado' está como 'Iniciou manobra de movimentação' e a embarcação está na região de algum terminal de dragagem

- **flag_operacao_saida_fim:** A embarcação é marcada neste campo quando sua coluna de Terminais está com um terminal associado diferente do anterior (movimentando para outro terminal, seja para apenas passar ou para estacionar) ou já saiu de regiões de terminais

E, desta forma, foi possível criar as colunas com os horários marcados:

- **Hora Entrada Início:** horário que a embarcação iniciou a operação de entrada
- **Hora Entrada Fim:** horário que a embarcação finalizou a operação de entrada
- **Hora Saída Início:** horário que a embarcação iniciou a operação de saída
- **Hora Saída Fim:** horário que a embarcação finalizou a operação de saída

Por fim, na etapa de Agrupamento dos Procedimentos por Embarcação os dados foram consolidados por embarcação e área de dragagem. Essa etapa resultou em uma tabela que apresenta, para cada embarcação e área de dragagem, a ordem das movimentações e os períodos correspondentes a cada processo (entrada, repouso e saída). Esse agrupamento facilita a visualização e análise do histórico operacional de cada embarcação em relação aos terminais no Porto de Santos. A tabela resultante proporciona uma representação organizada e estruturada das operações, permitindo uma compreensão mais detalhada da sequência de movimentos de cada embarcação nos diferentes locais do porto (desde a área de fundeio ao local de início da praticagem, área de dragagem para qual se movimentou e a sua saída).

3.3 Dados e Códigos

3.3.1 Validação dos Dados

O código foi realizado dentro do sistema *Jupyter Notebook* que roda em linguagem de programação *Python*. A coleta de dados resultou em diversos arquivos (todos compilados em um mesmo arquivo) em formato de tabela (*Pandas Dataframe* como mostra a Tabela 4).

Tabela 4: Estrutura da base de dados

Nome da coluna	Descrição do dado
SHIP_ID	Código ID do navio
Continua na próxima página	

Tabela 4 – Continuação da página anterior

Nome da coluna	Descrição do dado
IMO	Código IMO (<i>International Maritime Organization</i>) de uma embarcação, o qual é um código vinculado ao casco independentemente de mudança de bandeira, nome ou proprietário
CALLSIGN	Indicativo de Chamada Internacional - código de identificação de estação radiotransmissora em navios e aeronaves
SHIPNAME	Nome da embarcação
TYPE_COLOR	Tipo de cor da embarcação
LAST_POS	Última posição da embarcação
TIMEZONE	Fuso horário local
CODE2	Código que representa o nome do país
COUNTRY	País da embarcação
NEXT_PORT_NAME	Nome do porto de destino
NEXT_PORT_COUNTRY	País do porto de destino
NEXT_PORT_ID	ID do porto de destino
ETA	<i>Estimated Time of Arrival</i> - Tempo Estimado de Chegada no destino
DESTINATION	Destino da embarcação
CURRENT_PORT_COUNTRY	País do porto atual
TYPE_SUMMARY	Tipo da embarcação
COURSE	Direção cardinal da embarcação que a embarcação está direcionada
LON	Ponto longitudinal (em graus decimais)
LAT	Ponto latitudinal (em graus decimais)
AREA_CODE	Código da área do local
ETA_OFFSET	<i>Estimated Time of Arrival</i> - Tempo Estimado de Chegada com o atraso
SPEED	Velocidade de movimentação da embarcação (em nós)
Continua na próxima página	

Tabela 4 – Continuação da página anterior

Nome da coluna	Descrição do dado
LENGTH	Comprimento da embarcação (em metros)
WIDTH	Largura da embarcação (em metros)
ETA_UPDATED	Tempo Estimado de Chegada no destino atualizado
DISTANCE_TO_GO	Distância restante para o destino
PORT_ID	ID do Porto atual
CURRENT_PORT	Porto atual em que está localizado
CTA_ROUTE_FORECAST	Indica se há uma previsão de rota disponível para o canal ou rota marítima
current_time	Tempo de captura dos dados

A tabela inicial possuía 6679042 linhas e 32 colunas.

3.3.2 Filtragem Inicial dos Dados

A partir da primeira análise dos dados, foram descobertas algumas manipulações de filtragem de dados com pouca relevância para o modelo utilizado, precisando ser retiradas da base de dados onde o mesmo será construído.

Essa filtragem concentrou em 2 processos:

- **Retirada de Dados Estranhos:** Como o código concatenou diversos arquivos que foram coletados pelo processo de *webscrapping*, neste processo o título do arquivo concatenado era interpretado como um dado comum e assim resultou em diversas linhas com dados idênticos aos nomes das colunas.
- **Filtragem de Dados fora do escopo do projeto:** O nosso modelo desconsidera alguns tipos de embarcações, em especial o tipo 'Tug' (*Harbour Tug*) já que são embarcações de suporte e embarcações sem um código IMO (embarcações menores de diversos tipos como mencionado anteriormente) e ambos casos não estão dentro do escopo de análise do projeto.

O resultado da tabela filtrada reduziu a tabela inicial para 1386101 linhas, mantendo o mesmo número de colunas (32).

3.3.3 Tratamento Inicial dos Dados

Realizada a filtragem inicial dos dados, o desenvolvimento avançou para a etapa de tratamento inicial dos dados. Para isso, os tipos de dados foram investigados e

analisados para averiguar se estavam adequados, a fim de melhorar sua manipulação nas etapas posteriores do modelo. Por se tratar de uma tabela de tipo *Pandas DataFrame*, foi utilizada a função *.info()* para observar os tipos dos dados da base filtrada demonstrada na Figura 18.

Figura 18: Resultado da função *.info()* na base filtrada

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 949184 entries, 10 to 4549421
Data columns (total 32 columns):
#   Column                               Non-Null Count  Dtype
---  -
0   SHIP_ID                             949184 non-null object
1   IMO                                  949184 non-null object
2   MMSI                                 949184 non-null object
3   CALLSIGN                             949104 non-null object
4   SHIPNAME                             949184 non-null object
5   TYPE_COLOR                           949176 non-null object
6   LAST_POS                             949184 non-null object
7   TIMEZONE                             949184 non-null object
8   CODE2                                928659 non-null object
9   COUNTRY                              928659 non-null object
10  COUNT_PHOTOS                         949184 non-null object
11  NEXT_PORT_NAME                       826307 non-null object
12  NEXT_PORT_COUNTRY                   825786 non-null object
13  NEXT_PORT_ID                        826307 non-null object
14  ETA                                  883739 non-null object
15  DESTINATION                         924450 non-null object
16  CURRENT_PORT_COUNTRY               948686 non-null object
17  TYPE_SUMMARY                       949176 non-null object
18  COURSE                              946331 non-null object
19  LON                                  949184 non-null object
20  LAT                                  949184 non-null object
21  AREA_CODE                           949184 non-null object
22  ETA_OFFSET                          762773 non-null object
23  SPEED                               949089 non-null object
24  LENGTH                              949184 non-null object
25  WIDTH                               949184 non-null object
26  ETA_UPDATED                         66830 non-null object
27  DISTANCE_TO_GO                     949184 non-null object
28  PORT_ID                             949184 non-null object
29  CURRENT_PORT                       948686 non-null object
30  CTA_ROUTE_FORECAST                 949184 non-null object
31  current_time                       949184 non-null object
dtypes: object(32)
memory usage: 239.0+ MB
```

Pela Figura, é possível observar que todos os dados vem no formato *object*, formato esse de quando o modelo interpreta os dados como texto (em programação chamamos de *string*). Nesse formato, dados numéricos e de data são difíceis de manipular e, desta forma, realizamos a tratativa para transformá-los em dados numéricos e do tipo data (*datetime*). Essa tratativa foi realizada em dois processos, sendo eles:

- **Transformação dos dados para numéricos:** A base de dados possui diversas colunas numéricas mas restringimos a tratativa para as colunas 'LAT', 'LON',

'SPEED', 'LENGTH' e 'WIDTH'. Todas elas estavam com dados de texto e para realizar a transformação, utilizou-se a função *to_numeric()* da biblioteca *Pandas*

- **Transformação dos dados para o formato de data:** Para essa tratativa, criou-se uma coluna nova chamada 'data_tratada' e utilizando a função *strptime()* da biblioteca *datetime*, foi possível transformar os dados para o formato desejado

3.3.4 Criação de colunas de estado da embarcação

Em adição a manipulação da base de dados, foram construídas algumas colunas para melhorar análises posteriores. Pensando nisso, foram geradas colunas de identificação do estado no qual a embarcação se encontra, além da categorização do horário através da 'Janela'. Esse processo foi realizado da seguinte forma:

- **'Estado_Navio':** Coluna que indica a condição de movimento da embarcação, tendo 3 categorias.
 - **Parado:** Estado que indica quando a embarcação está com velocidade (coluna 'SPEED') igual a 0
 - **Manobrando:** Estado que indica quando a embarcação está com velocidade (coluna 'SPEED') maior que 0 e menor ou igual a 0.5 (velocidade de manobra imposta)
 - **Movimentando:** Estado que indica quando a embarcação está com velocidade (coluna 'SPEED') acima de 0.5
- **Janela:** A partir do horário da coleta do dado, foi determinada a janela de horário que são divididas em 4 categorias.
 - **Janela 1:** Esta janela corresponde ao período de tempo entre 1 da manhã e 7 da manhã (filtro de horas entre hora 1 e 7 inclusivo)
 - **Janela 2:** Esta janela corresponde ao período de tempo entre 7 da manhã e 1 da tarde (filtro de horas entre hora 7 e 13 inclusivo)
 - **Janela 3:** Esta janela corresponde ao período de tempo entre 1 da tarde e 7 da noite (filtro de horas entre hora 13 e 19 inclusivo)
 - **Janela 4:** Esta janela corresponde ao período de tempo entre 7 da noite e 1 da manhã (filtro de horas entre hora 19 e 1 inclusivo)

O código foi construído para percorrer (por meio da função *apply* a coluna 'SPEED' (coluna de velocidade) e a coluna 'data_tratada' (coluna com a data no formato devido) visando categorizar nos 3 Estados (Parado, Manobrando e Movimentando) e nas 4 Janelas de período. O código para a criação destas colunas utilizado foi o 2 e está na página 79.

3.3.5 Criação de colunas de manipulação intermediárias

Para facilitar a observação de certos comportamentos da base para o modelo, criaram-se duas colunas intermediárias indicando características específicas das embarcações. Estas colunas foram:

- **'Flag_3Mov'**: Coluna indicando a característica geral de movimentação da embarcação na base de dados. Essa *flag* aponta quando a embarcação em questão possui pelo menos um dado de todas as condições de Estado possíveis (Parado, Manobrando e Movimentando). Para isso, o agrupamento se deu por embarcação e foi realizado a contagem distinta de 'Estado_Navio'. Quando a embarcação possuía 3 como contagem distinta, a marcação deste tipo de caso era realizada na base de dados original na coluna em questão.
- **'Flag_Mudanca_Estado'**: Coluna que indica o momento de mudança de um estado para outro de uma dada embarcação. Por exemplo, quando a embarcação muda do 'Estado_Navio' Parado para o 'Estado_Navio' Manobrando, a *flag* indica 1.

O código utilizado foi o 3 e está na página 79.

A partir dessas colunas, foi possível analisar embarcações específicas, ao filtrar embarcações que possuíam a 'Flag_3Mov', ou seja, um indicativo de uma embarcação bem ativa (realizou os 3 tipos de 'Estado_Navio') durante o período da base e o momento exato da mudança desses estados.

3.3.6 Criação de colunas de análise

Com as colunas intermediárias de estado das embarcações criadas, foi possível criar colunas com mais detalhes sobre os movimentos realizados por cada embarcação em uma determinada mudança de estado. Essa coluna indica mais elementos de interesse ao modelo. A coluna criada foi a de 'Mudanca_Estado' que indica com maior precisão qual foi a mudança em questão (em qual estado que estava para qual estado que mudou). Para a construção desta coluna, foram utilizadas duas colunas: a coluna de 'Flag_Mudanca_Estado' (para filtrarmos os momentos nos quais uma embarcação realiza uma mudança de 'Estado_Navio') e a coluna de 'Estado_Navio' (para identificarmos com exatidão qual foi a mudança observada). Se trata de uma coluna categórica com valores em *string* com as possíveis variações listadas a seguir:

- **Parou**: A embarcação estava se movimentando e parou (Estado Movimentando -> Parado)
- **Moveu**: A embarcação estava parada e começou a se movimentar (Estado Parado -> Movimentando)

- **Iniciou manobra de parada:** A embarcação estava se movimentando e começou a manobrar (Estado Movimentando -> Manobrando)
- **Parou manobra e moveu:** A embarcação estava manobrando e parou (Estado Manobrando -> Parado)
- **Iniciou manobra de movimentação:** A embarcação estava parada e começou a manobrar (Estado Parado -> Manobrando)
- **Parou manobra e estacionou:** A embarcação estava manobrando e parou (Estado Manobrando -> Parado)

O código utilizado foi o 4 e está na página 80.

3.3.7 Identificação de Localização das Embarcações nas Áreas de Fundeio

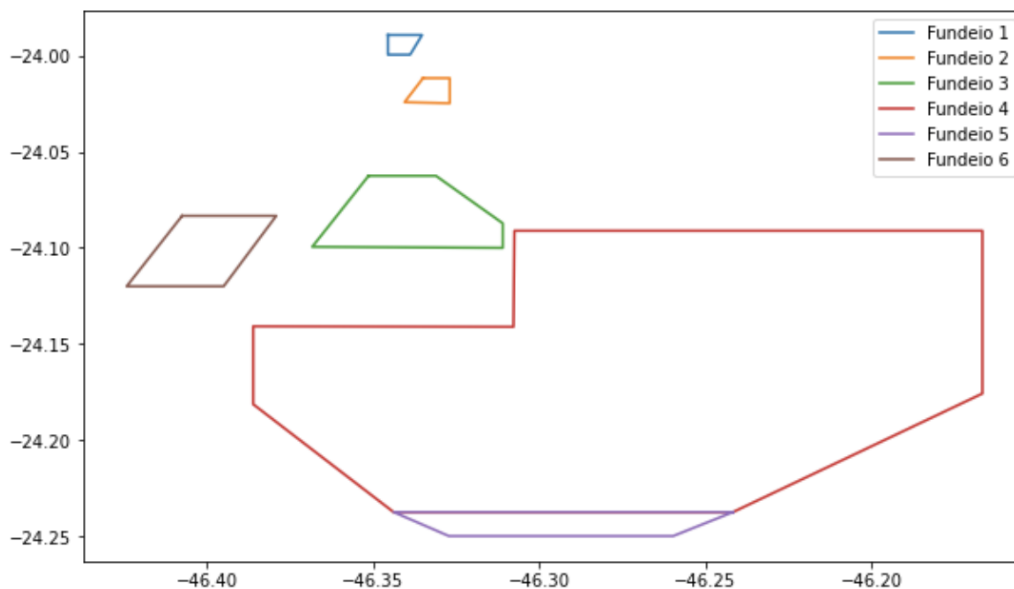
O porto de Santos possui 6 áreas de fundeio que atendem diferentes tipos de navio de acordo com regras do porto como apresentado na Figura 4. Como explicado anteriormente, essas áreas são muito importantes para a operação portuária já que funcionam tanto como uma "fila"(embarcações aguardam nesses locais por sua vez de atracação) assim como local de espera (para embarcações que ainda não desejam atracar mas precisam/desejam aguardar ali). Nesta etapa, é realizada a identificação de qual dos 6 fundeadouros (caso a embarcação esteja em um deles) a embarcação em questão se localiza naquele determinado horário pela sua posição (a partir das colunas de LAT e LON).

Para isso, é preciso primeiro construir um código que consiga determinar a área geográfica com a dimensão e formato de cada fundeadouro. Tendo isso em vista, a partir dos dados presentes no site do Porto de Santos (Santos, s.d.), geramos tabelas com pontos geográficos que descrevem o polígono relacionado a cada área de fundeio (dados presentes nas páginas 87 a 88).

As tabelas, com as coordenadas por ponto, estão em graus, minutos e segundos (ou em graus, minutos decimais) enquanto que a base de dados está com latitude e longitude com graus decimais. Dessa forma, foi realizada a conversão dos pontos de fundeadouro para graus decimais buscando realizar a posterior identificação do local das embarcações nas áreas de fundeio. Assim, criou-se uma função ('verificar_fundeio') que verificava se uma embarcação estava em uma das áreas de fundeio adicionando uma margem de erro (margem de erro arbitrária de 0.01). Para isso, foram utilizadas funções da biblioteca *shapely* que tornam possível a construção de polígonos a partir de pontos em coordenadas, além de realizar a checagem para averiguar-se se um polígono contém um ponto pela função *contain()* como mostra o código 5 que está na página 81.

Dessa forma, o código realiza a identificação, retornando o fundeadouro referido. Primeiro, criamos os polígonos que representam os fundeadouros pelas coordenadas dadas como mostra a Figura 19.

Figura 19: Polígonos formados pelas coordenadas das áreas de fundeio



Fonte: Elaboração própria.

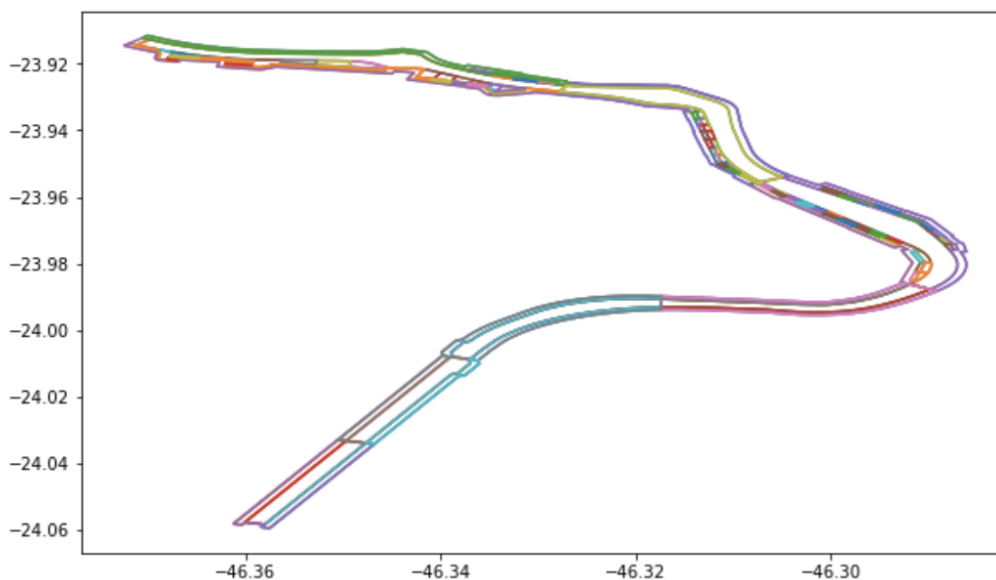
3.3.8 Identificação de Localização das Embarcações nas Áreas de Dragagem

De forma análoga, foi realizado o processo de identificação da localização das embarcações pelas áreas de dragagem. O porto de Santos possui em torno de 120 áreas de dragagem conforme bases recebidas pelo Porto de Santos (existem partições de uma mesma área de dragagem porém o código as interpreta como áreas distintas). Nesta etapa, é realizada a identificação de qual das áreas de dragagem (caso a embarcação esteja em um deles) a embarcação em questão se localiza naquele determinado horário pela sua posição (a partir das colunas de LAT e LON). Os dados geográficos das áreas de dragagem foram disponibilizados pela Autoridade Portuária de Santos e estavam contidos em arquivos *shapefile* em dimensão UTM, enquanto os dados da base da AIS estavam em graus decimais. Assim, foi necessário a conversão dos dados de UTM em graus decimais visando adequar a base como mostra o código abaixo e, assim, tornar possível a realização da identificação das áreas de dragagem.

Com a função descrita no código 6 - página 81, os dados foram convertidos em graus decimais. A partir disso, foi realizada uma função análoga a de 'verificar_fundeio' (código 5 - página 81), desta vez para as áreas de dragagem. Para este fim, o código analisou se cada ponto (par de dado de latitude e longitude - coluna 'LAT' e 'LON')

estava contido nas geometrias de dragagem presentes na lista *dragagem_geometria* e, caso estivesse presente, retornava o nome da área que está contida nas lista *dragagem_areas* na mesma ordem (este código é o código 7 da página 82).

Figura 20: Polígonos formados pelas coordenadas das áreas de dragagem



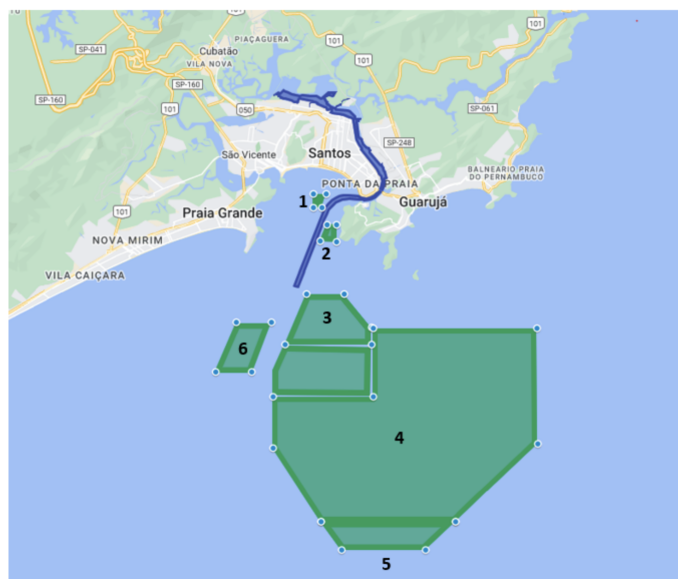
Fonte: Elaboração própria.

A partir da Figura 20, é possível observar que as áreas de dragagem estão espalhadas por todo o canal do Porto de Santos e se assemelha muito a Figura 4 mostrada na introdução do projeto. A caráter ilustrativo, criou-se o gráfico com as áreas de dragagem e áreas de fundeio juntos para comparar com a Figura 4, mostrando elas na Figura 21.

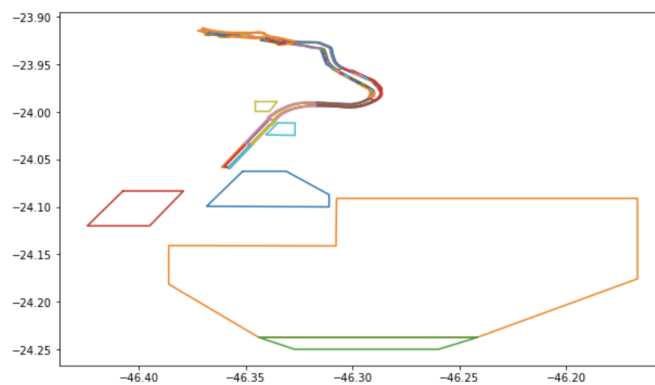
Através desta comparação, foi possível observar a semelhança de ambas as figuras com a maior diferença sendo a escala e a ausência da área em verde na Figura 4 entre os fundeadouros 3 e 4 a qual aponta para uma área de despejo do Porto de Santos. Na Figura 4, o canal em azul apontando para o canal de acesso ao Porto e as áreas entre os Trechos 1 e 4 se relacionam muito bem com as áreas de dragagem identificadas pelo código.

Figura 21: Comparativo entre Figura gerada pelo código e a Figura do *Google Maps* no site do Porto de Santos

(a) Figura 4



(b) Figura gerada pelo código com áreas de dragagem e fundeio



3.3.9 Categorização do procedimento de movimentação

Os passos anteriores identificaram em quais das áreas de dragagem ou fundeadouros as embarcações estavam. Feito as colunas de 'Fundeiio' e 'Area_Dragagem', foi realizado um processo de categorização do procedimento de movimentação realizado para identificar se uma dada embarcação está entrando ou saindo de um determinado Fundeio ou Terminal (se estiver realizando tal procedimento). Primeiramente, foi criada a coluna de 'Estado_Entrada_Area_Dragagem' (o processo é semelhante para o Fundeio). As condições utilizadas para identificar a entrada de uma dada embarcação na área de dragagem são: a partir de uma tabela ordenada pela data, a linha atual e a seguinte são referentes ao mesmo navio e, a linha seguinte possui uma área de dragagem identificada diferente da anterior. Dessa forma, uma embarcação que estava anteriormente em alguma Área de Dragagem ou não estava em nenhuma e passou a adentrar uma área diferente, será marcada pelo código. De maneira análoga, foi feito o processo para marcar a saída que, sob a mesma condição, marcava a saída da Área de Dragagem ou Fundeio que estava anteriormente.

Ao entrar em uma Área de Dragagem, o resultado do código indicará "Entrou na (área de dragagem referida)" na linha que indicava a entrada da embarcação, ou seja, na linha seguinte sob a referência da condição, assim como ocorre no processo de Fundeio. O processo da saída de área de dragagem acontece quase da mesma forma, com a diferenciação de que a marcação de saída acontece na linha de saída, ou seja, na linha atual sob a referência da condição, preenchendo, assim, com "Saiu da (área de dragagem referida)" (código 8 - página 82).

3.3.10 Documentação do procedimento de movimentação

O mapeamento do fluxo do navio foi realizado para capturar 8 momentos: o momento que a embarcação está pronta para entrar (NOR - *Notice of Readiness*), o início da movimentação de entrada (que se dá pela saída da área de fundeio), o acoplamento a praticagem (momento chamado de POB, *Pilot On Board*), início da atracação (momento que entra em uma área de dragagem e começa a operação de manobra), início da pré-operação (parou a manobra e estacionou, se torna apto a realizar manobra de descarregamento ou carregamento de mercadorias), início da desatracação (inicia a operação de manobra de saída), início da navegação de saída (sai da área de dragagem) e, por fim, o fim da praticagem (o segundo momento de POB, determinando o fim da operação de praticagem para a saída). A base trabalhada não possuía todos os dados como os momentos de POB (tanto para entrar no canal como para sair) e o NOR. Assim, calcularemos esses momentos como o primeiro momento de entrada na área 'Trecho1B_soleira' como o primeiro POB, e o último momento nesta mesma área como o segundo POB, uma vez que é a área que identificamos como a

área mais próxima do trecho onde a praticagem é iniciada. Já para o NOR pegaremos o primeiro horário que a embarcação foi identificada dentro do Fundeadouro.

Podemos então, separar a documentação dos horários em 3 grupos:

- **Horários relacionados ao fundeio:** NOR (primeiro momento mapeado com a embarcação dentro do Fundeio) e Horário de Saída do Fundeio
- **Horários críticos da movimentação:** Horário de Início da Entrada, Horário do Fim da Entrada, Horário do Início da Saída e Horário do Fim da Saída de uma Área de Dragagem
- **Horários POB:** Horários relacionados a Praticagem, manobra realizada para reduzir os riscos das movimentações de grandes embarcações dentro do canal do Porto

Os horários relacionados ao Fundeio foram capturados a partir do momento em que uma embarcação é identificada em uma área de fundeio. Para mapearmos a embarcação que passa por uma área de fundeio, criamos 3 colunas: 'flag_fundeio' (indicando 1 quando a embarcação está em uma área de fundeio e 0 caso contrário), 'hora_fundeio' (coluna com horário caso a embarcação esteja em uma área de fundeio) e 'hora_saida_fundeio' (coluna com o horário quando a coluna de 'Estado_Saida_Fundeio' seja diferente de 0, indicando que a embarcação está saindo de uma área de fundeio). Com isso, agrupamos por embarcação e trazemos estes campos da seguinte forma: 'flag_fundeio' como máximo, uma vez que a embarcação tenha estado em um fundeadouro é o suficiente para indicar a passagem da embarcação pelo fundeadouro (nomeando este campo como 'flag_navio_fundeio'), 'hora_fundeio' como mínimo para calcular o NOR da embarcação (dando este nome) e 'hora_saida_fundeio' como o primeiro valor já que queremos pegar a saída para as áreas de dragagem (nomeando este campo como 'hora_saida_fundeio_navio'). Após este passo, unimos os dados obtidos à base inicial para fins de registro como o código 9 da página 83 indica.

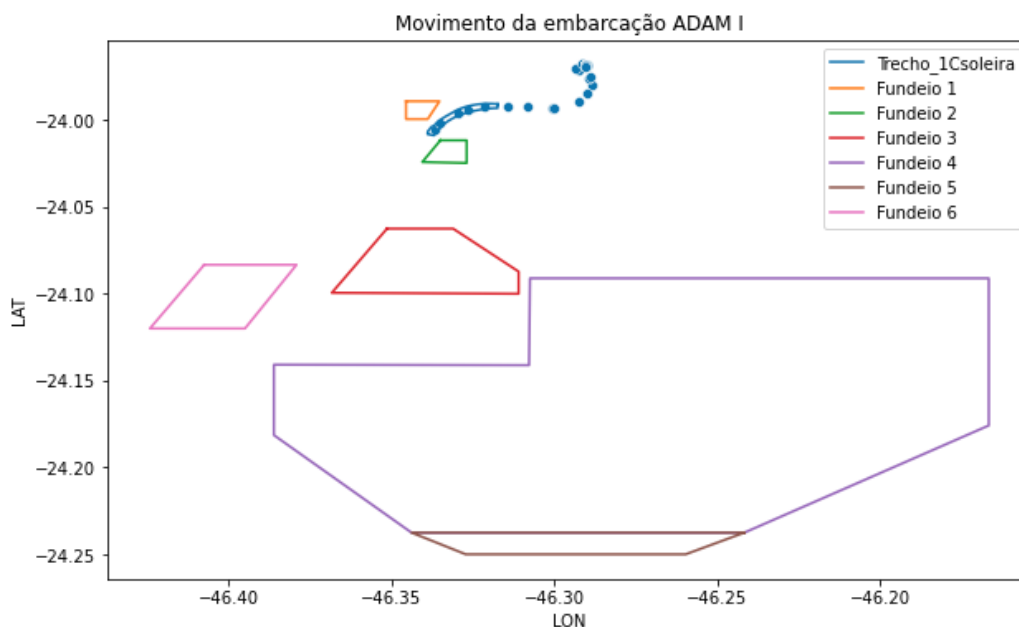
Com a categorização da movimentação realizada anteriormente, a qual indica quando e para onde estava entrando (ou saindo se fosse esse o caso), foi realizada a marcação deste procedimento. Esta marcação se dá pela indicação do intervalo de horas em que a movimentação ocorreu, ou seja, o horário de início e fim para entrada e saída. Então, primeiramente foi realizada a identificação de 4 momentos críticos para esta documentação, sendo eles: o início da entrada, o fim da entrada, o início da saída e o fim da saída. Para este fim, criaram-se colunas que realizaram o *flag* destes momentos a partir da coluna de 'Mudanca_Estado' e se a Área de Dragagem associada a embarcação for válida (ou seja, não é um "Local não identificado"). O código utilizado para este passo foi o 10 que está na página 84.

A partir da marcação das linhas críticas para a documentação (a indicação pelas colunas 'fl_operacao_entrada_inicio', 'fl_operacao_entrada_fim', 'fl_operacao_saida_inicio' e 'fl_operacao_saida_fim'), foi necessário obter os horários de cada um desses momentos. Para que o código colete o horário presente na coluna de "data_tratada", a coluna *flag* deverá estar marcada. O horário, então, é associado a uma nova coluna para cada um dos 4 cenários supracitados. (início da entrada, fim da entrada, início da saída e fim da saída), como mostrado no código 11 que está na página 84.

Desta forma, foi possível obter o fluxo de operações realizadas por uma embarcação em uma dada área de dragagem. Este fluxo de operações de movimentação são dados pelos movimentos de entrada e saída (início e fim de cada operação).

Por fim, obtiveram-se os horários relacionados a operação de praticagem (POB 1 - horário de início da praticagem de entrada e POB 2 - horário de fim da praticagem de saída). Como apontado anteriormente, a base não continha este dado e não foi possível obtê-lo de outras bases. Visto isso, para obter um dado próximo a estes horários, observou-se que a Área de Dragagem correspondente a 'Trecho_1Csoleira' corresponde a um dos trechos iniciais da movimentação em direção ao canal do Porto, como mostrado pela Figura 22, na qual é retratado o movimento da embarcação 'ADAM I', e os pontos iniciais e finais de sua movimentação estão localizados nesta área de dragagem ('Trecho_1Csoleira').

Figura 22: Movimentação da embarcação 'ADAM I' e o local 'Trecho_1Csoleira'



Fonte: Elaboração própria.

Desta forma, bastou apenas tomar o horário mínimo e máximo de cada embarcação no 'Trecho_1Csoleira' como os momentos POB 1 e POB 2 (POB 1: momento de

entrada sendo um dos primeiros momentos da entrada no canal e, POB 2: momento de saída).

3.3.11 Agrupamento dos procedimentos por embarcação e área de dragagem

Levantados os 8 horários para cada embarcação (NOR, horário de saída da área de fundeio, POB 1, horário de entrada de início, horário de entrada final, horário de saída início, horário de saída final e POB 2), foi realizado o agrupamento destes horários para cada embarcação e área de dragagem. Os respectivos horários identificam os momentos críticos de uma movimentação completa de uma embarcação no Porto de Santos, desde a área de fundeio até a saída para o próximo destino. Assim, já que os horários dos momentos críticos das movimentações estão relacionados à Área de dragagem, era preciso apenas agrupar por embarcação e área de dragagem e puxar os horários utilizando uma função de operação de agrupamento (como a função *max()*).

A partir do código 12 da página 85, foi possível gerar a tabela final de horários dos pontos críticos da movimentação por embarcação e área de dragagem, sendo mostrada na Tabela 5.

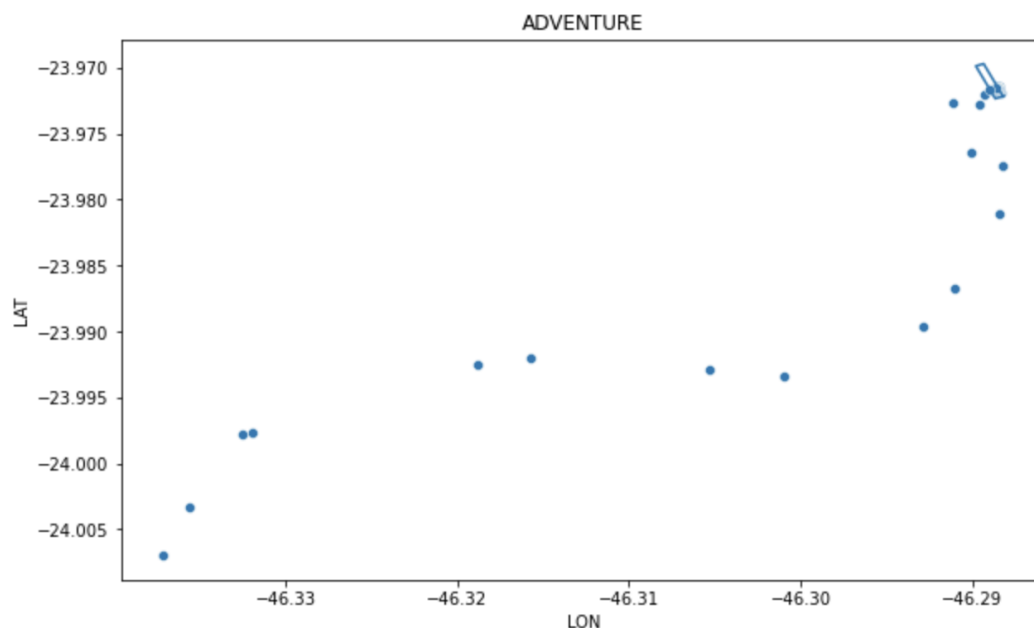
Tabela 5: Trecho da Tabela Final

SHIPNAME	Area_Dragagem	NOR	hora_saida_fundeio_navio	POB_1	hora_entrada_inicio	hora_entrada_fim	hora_saida_inicio	hora_saida_fim	POB_2
ADAM I	Box_TGG	2023-07-25 22:31:47	2023-07-25 22:41:54	2023-07-25 22:31:47	2023-07-25 23:27:27	2023-07-25 23:32:30	2023-07-27 11:50:20	2023-07-27 11:50:20	2023-07-27 12:30:57
ADVENTURE	Box_TERMAG	2023-08-02 19:05:51	2023-08-02 19:15:59	2023-08-02 19:05:51	2023-08-02 19:51:25	2023-08-04 22:01:10	2023-08-04 21:46:00	2023-08-05 02:39:21	2023-08-05 03:19:50
ADVENTURER	Box_Arm30	2023-07-09 06:29:02	2023-07-09 06:39:10	2023-07-09 06:29:02	2023-07-09 07:24:46	2023-07-09 07:29:50	2023-07-12 03:19:34	2023-07-12 03:19:34	2023-07-12 04:00:11
AEOLIAN	Box_Arm38	2023-09-13 18:39:21	2023-09-13 18:44:26	2023-09-13 18:39:21	2023-09-13 19:19:48	2023-09-17 03:45:11	2023-09-17 06:17:14	2023-09-17 06:22:18	2023-09-17 06:52:43
AEOLIAN VICTORY	Box_Arm39	2023-08-01 07:55:22	2023-08-01 08:05:32	2023-08-01 07:55:22	2023-08-01 08:36:02	2023-08-03 06:10:02	2023-08-03 05:54:47	2023-08-03 09:18:01	2023-08-03 09:43:22
...
YU HONG	Box_Arm20_21	2023-08-08 01:17:13	2023-08-08 01:27:20	2023-08-08 01:17:13	2023-08-08 02:23:05	2023-08-10 16:34:54	2023-08-12 02:56:50	2023-08-12 02:56:50	2023-08-12 03:47:38
YU TONG	Box_Arm38	2023-07-25 08:40:09	2023-07-25 08:45:14	2023-07-25 08:40:09	2023-07-25 09:15:39	2023-07-27 16:44:59	2023-07-27 16:39:56	2023-07-27 21:23:41	2023-07-27 21:59:09
ZEPHYROS	Box_TGG	2023-08-25 18:42:06	2023-08-25 18:47:09	2023-08-25 18:42:06	2023-08-25 19:17:30	2023-08-30 10:29:47	2023-08-30 10:24:43	2023-08-30 13:02:02	2023-08-25 18:47:09
ZHONG XIN PEARL	Talude_T2_Norte	2023-09-17 14:43:48	2023-09-17 14:53:56	2023-09-17 14:43:48	2023-09-17 15:24:24	2023-09-19 22:52:54	2023-09-19 23:23:16	2023-09-19 23:28:19	2023-09-20 00:13:51
ZHONG YUAN HAI YUN JIN QU	Box_TEV	2023-08-28 01:57:02	NaT	2023-08-28 01:51:58	2023-08-27 15:53:27	2023-08-27 18:56:11	2023-08-27 18:51:07	2023-08-28 01:06:13	2023-08-28 02:02:06

Fonte: Elaboração própria.

Para ilustrar o mapeamento de uma embarcação realizada pelo modelo do projeto, tomou-se a embarcação de nome 'ADVENTURE' cujo destino foi a área de dragagem chamada 'Box_TERMAG'. Assim, foi gerado um gráfico de dispersão utilizando os pontos de latitude e longitude dos locais que esta embarcação passou, além de plotar o polígono referente a área de dragagem em questão. O resultado consta na Figura 23.

Figura 23: Mapeamento do trajeto realizado pela embarcação 'ADVENTURE'



Fonte: Elaboração própria.

3.4 Análise Exploratória da Base de Dados

Por fim, foi realizada a Análise Exploratória da Base de Dados, abrangendo desde análises dos dados primários (da base da AIS coletada sem as manipulações realizadas), para entender algumas características da base e também dos dados secundários, para mostrar detalhes mais profundos e de uma análise e estudo mais avançado. A Análise Primária da Base de Dados percorreu dados categóricos e numéricos que existiam na base além de análises preliminares com campos criados pelo presente projeto

3.4.1 Análise Primária da Base de Dados - Dados Categóricos

A Análise Exploratória se iniciou pelo estudo dos tipos de embarcações presentes na base. Vale ressaltar que para o nosso estudo, foram filtradas as embarcações do tipo *Tug* (rebocadores) e embarcações que não possuíam código IMO (tipicamente as embarcações de menor porte). Como mostra a Figura 24, mais de 70% das embarcações são do tipo navios cargueiros seguidos pelos *Tankers* (embarcações que fazem o transporte de Gás Liquefeito de Petróleo (GLP), Gás Natural Liquefeito (GNL), produtos líquidos, produtos químicos industriais além de cargas sensíveis) e navios específicos (como navios de exploração da marinha entre outros).

Figura 24: Tipos de Embarcações mais Presentes na Base de Dados

Tipo_Embarcacao	Quantidade	Porcentagem	
0	Cargo	983331	70.9428
1	Tanker	264343	19.0711
2	Special Craft	129101	9.3140
3	Unspecified	6833	0.4930
4	Sailing Vessel	1830	0.1320
5	Passenger	303	0.0219
6	High Speed Craft	176	0.0127
7	Pleasure Craft	84	0.0061
8	Fishing	70	0.0051
9	Other	15	0.0011
10	Navigation Aid	3	0.0002

Fonte: Elaboração própria.

Posteriormente, foram analisados os portos de destino. A base de dados apresentavam diversos nomes para o porto de Santos então, após fazer um tratamento que agrupou todos os nomes contendo 'Santos' e, também, os casos que chamavam 'BR SSZ', outro nome para o Porto de Santos, obteve-se a Tabela 25 (dividida em duas visualizações: uma por local e outra por país de destino). Como esperado, a grande maioria das embarcações se dirige ao Porto de Santos, mas varias outras embarcações estavam se direcionando a outros lugares (em torno de 30% da base). Importante frisar que foram mapeados mais de 800 locais de destino, o que mostra uma grande variedade de embarcações em Santos. A visão a nível país de destino também foi explorada, mostrando uma grande maioria de embarcações com o próximo país definido como Brasil, já que muitas das embarcações estudadas estavam se movimentando em destino ao próprio Porto de Santos. Já a quantidade de países diferentes mapeados como destino foi de 65 países diferentes, mas com pouca representatividade para cada um deles (o 4º maior país de destino tem apenas 0.68% de representatividade).

Em contrapartida, os países de origem são mais bem distribuídos. Como mencionado na Introdução, o Porto de Santos recebe embarcações de diversos países e regiões do mundo. Pela base de dados deste trabalho, destacam-se a Libéria, Panamá e Ilhas Marshall somando quase 50%. Além disso, a base indicou 48 países diferentes presentes neste campo (incluindo Brasil) e, portanto, com grande represen-

Figura 25: Origem com maior representatividade na Base de Dados

(a) Local de Destino

	Destino	Quantidade	Porcentagem
0	SANTOS	959846	71.1523
1	KLAIPEDA	29285	2.1709
2	NIT PORT	29285	2.1709
3	AO SEBASTIAO	29285	2.1709
4	MATANZAS	28346	2.1013
5	BRRIO	7979	0.5915
6	STS AREA	6712	0.4976
7	BRPNG	6114	0.4532
8	SINGAPORE	5490	0.4070
9	SG SIN	5443	0.4035
10	GUARUJA	4616	0.3422

(b) País de Destino

	Pais_Destino	Quantidade	Porcentagem
0	BR	1110156	91.5956
1	PA	28474	2.3493
2	SG	25218	2.0807
3	AR	8222	0.6784
4	US	4225	0.3486
5	NG	3752	0.3096

tatividade internacional. Para uma visão com granularidade a nível Região do Mundo, por não haver um campo destinado a isso, foi coletada uma base de dados da *Kaggle* (plataforma que possui diversas bases de dados com foco principal para o uso em projetos de Ciência de Dados e Aprendizado de Máquina) que tivesse tal detalhamento. A partir disso, foi possível unir as bases e adicionar este novo campo que, obteve um pareamento bastante efetivo (de 1.36 M de casos, 1.31 M foram mapeados) e, assim, foi gerada a análise mostrada pela Figura 26. A base mostra novamente que há uma boa distribuição com exceção de pouquíssimos casos de embarcações da América do Norte, mostrando que provavelmente chegam ao Brasil por outros portos ou existe um local intermediário antes de chegar ao Brasil.

Figura 26: Origem com maior representatividade na Base de Dados

(a) País Origem

	Pais_Origem	Quantidade	Porcentagem
0	Liberia	250043	18.4322
1	Panama	224026	16.5143
2	Marshall Islands	200535	14.7826
3	Brazil	150166	11.0696
4	Hong Kong	87536	6.4528
5	Malta	85649	6.3137
6	Singapore	64872	4.7821
7	China	36402	2.6834
8	Norway	31799	2.3441
9	Bahamas	31661	2.3339
10	Denmark	29981	2.2101
11	Cyprus	28829	2.1252
12	Portugal	21047	1.5515
13	United Kingdom	15476	1.1408
14	Barbados	11673	0.8605
15	Greece	11568	0.8527
16	Netherlands	11118	0.8196
17	Italy	9522	0.7019
18	Cayman Islands	9256	0.6823
19	Korea	7759	0.5720

(b) Região Origem

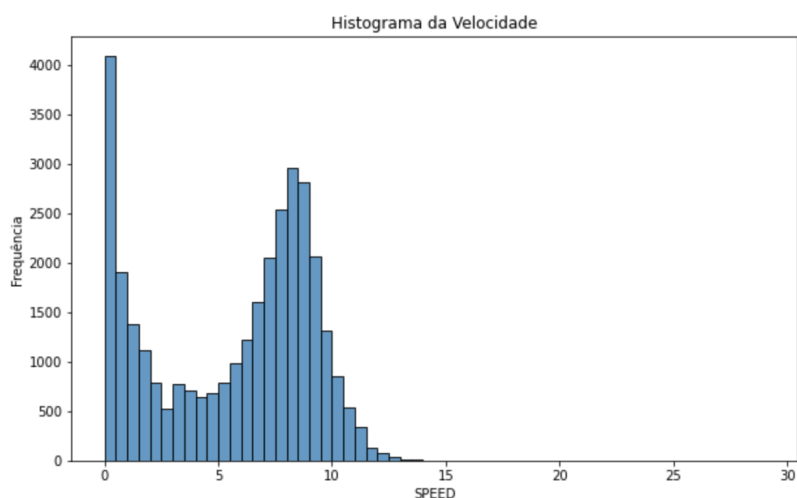
	Regiao_Origem	Quantidade	Porcentagem
0	LATIN AMER. & CARIB	396362	30.2032
1	SUB-SAHARAN AFRICA	252868	19.2688
2	WESTERN EUROPE	220696	16.8173
3	ASIA (EX. NEAR EAST)	201958	15.3894
4	OCEANIA	201677	15.3680
5	NEAR EAST	37095	2.8267
6	NORTHERN AMERICA	1663	0.1267

3.4.2 Análise Primária da Base de Dados - Dados Numéricos

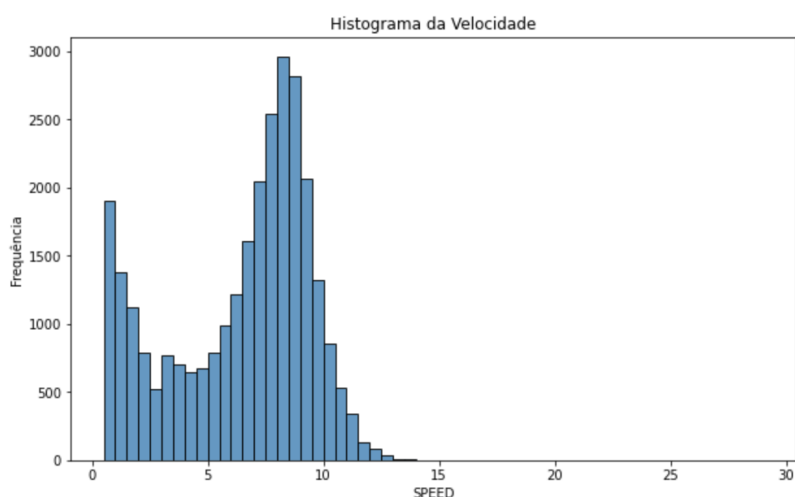
A Análise Primária dos Dados Numéricos da Base foi mais superficial pois grande parte da análise numérica focou em detalhes mais aprofundados, ficando para a parte seguinte deste relatório. Assim, neste momento foram analisados os campos de: velocidade ('SPEED'), dimensões da embarcação ('WIDTH' - Largura e 'LENGTH' - Comprimento), 'Estado_Navio', 'Janela' (período de operação do Porto) e 'Flag_3Mov' (indica se a embarcação esteve nos 3 tipos de 'Estado_Navio' possíveis). Para a velocidade, foram construídos dois histogramas que mostram a distribuição de velocidades das embarcações. Como a base de dados foi coletada de forma contínua, muitos dados são "repetidos" e, assim, para fazer uma melhor análise, tiramos as duplicidades de uma mesma embarcação com uma mesma velocidade, ou seja, para uma embarcação em uma velocidade específica, mapeamos apenas um registro. Destes dois histogramas, um foi construído com a aplicação de tratamento para apenas as velocidades de movimentação (acima de 0.5 nós) e a outra sem este tratamento e o resultado está na Figura 27. A distribuição sem tratamento aponta uma grande concentração de embarcações Paradas, o que pode significar uma grande quantidade de embarcações paradas nos terminais ou nos fundeadouros em comparação com a quantidade de embarcações em movimentação ou manobra. Já a distribuição com o tratamento permite uma melhor visualização da distribuição das velocidades de movimentação, onde é possível notar uma concentração no intervalo de 7-9 nós de velocidade.

Figura 27: Histograma da Distribuição da Velocidade ('SPEED') das embarcações

(a) Histograma Sem Tratamento



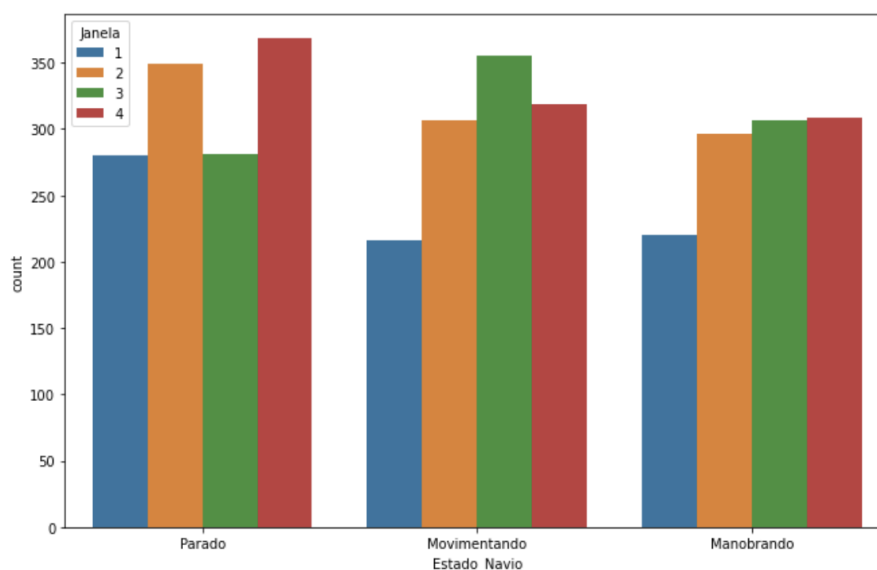
(b) Histograma Com Tratamento



Fonte: Elaboração própria.

Para a análise da 'Janela', foi realizada a distribuição das embarcações levando em consideração seu 'Estado_Navio'. Realizamos, também, o mesmo tratamento de retirada de duplicações para uma melhor identificação da distribuição das embarcações pelas categorias como mostra a Figura a seguir. O período da Janela 1 possui a menor quantidade de embarcações enquanto a 4 possui o maior número, outro destaque vai para a Janela 3 que conta com a maior quantidade de embarcações no estado de movimentação, além de uma quantia bem considerável manobrando. Esta análise pode indicar que a Janela 1 possui um indicador de dificuldade para a atracação das embarcações, seja burocrática ou climática (temperatura, chuva no período da madrugada - 1 a 7 da manhã) e que, as Janelas 3 e 4 possuem as melhores condições de operação, principalmente para a movimentação e atracação (manobra) das embarcações.

Figura 28: Distribuição da quantidade de embarcações pelo 'Estado_Navio' e 'Janela'



Fonte: Elaboração própria.

Ainda utilizando o campo Janela, foi realizada, também, a análise da distribuição de quantidade de embarcações realizando algum tipo de movimentação (entrada ou saída), a fim de identificar algum possível padrão presente na base de dados. Para isso, aplicou-se um filtro na base, e foram apresentadas apenas as linhas que capturavam momentos de entrada ou saída de embarcações e, a partir do resultado obtido, foi realizado um agrupamento para contar quantas embarcações distintas fizeram operações de movimentação pelas janelas. O resultado obtido está na Figura 29, onde é possível ver uma boa distribuição entre as Janelas com a maior concentração na Janela 3 (entre 1 da tarde e 7 da noite) e a menor concentração na Janela 1 (período entre 1 da manhã e 7 da manhã).

Figura 29: Distribuição da quantidade de embarcações distintas realizando operação de movimentação pelas Janelas

Janela	Embarcações	Porcentagem %	
0	1	685	22.33
1	2	776	25.29
2	3	835	27.22
3	4	772	25.16

Fonte: Elaboração própria.

Por último, foi realizada uma análise de forma a identificar as características de dimensões do navio, velocidade e comportamento ('Flag_3Mov') por tipo da embarcação ('TYPE_SUMMARY'). Para isso, novamente utilizamos a base tratada sem as duplicações e retiramos o comportamento médio destes atributos numéricos, resultando na análise da Figura a seguir. Pela Figura 30, é possível observar quais são as embarcações com maior interação dentro da base utilizando como critério o campo campo de 'Flag_3Mov' (quanto mais perto de 1, maior a proporção de embarcações que realizaram os 3 estados pelo total) sendo elas as embarcações do tipo: *High Speed Craft*, *Cargo* e *Tanker*. Já para os outros campos, é preciso tomar que quanto maior o valor, maior o atributo. Dito isso, a embarcação mais rápida foi a de *Passenger*, e a mais comprida e mais larga foi a embarcação do tipo *Cargo*. Vale lembrar que a métrica utilizada foi a média por agrupamento, e que é importante que mais análises sejam realizadas, uma vez que podem haver *outliers* e valores incorretos que possam interferir negativamente no processo de análise.

3.4.3 Análise Secundária da Base de Dados

A Análise Secundária da Base de Dados partiu da tabela resultante da modelagem realizada neste projeto. A tabela contém os eventos nos quais uma dada embarcação realiza durante sua passagem no Porto de Santos (além dos horários de movimentação da embarcação com relação as operações de entrada e saída da área de dragagem, também foram levantados horários da praticagem e relacionados ao fundeio), dando ênfase para uma área de dragagem específica. A partir disso, foram calculados os tempos relacionado a operações específicas, sendo elas:

- **Tempo da Operação de Entrada:** Tempo decorrido entre o horário de entrada

Figura 30: Análise da Característica de Atributos por Tipo de Embarcação

	TYPE_SUMMARY	SPEED	LENGTH	WIDTH	Flag_3Mov
0	Cargo	2.8302	223.3128	33.6197	0.9509
1	Fishing	1.0306	21.0556	6.8611	0.0000
2	High Speed Craft	0.0000	12.0000	4.0000	1.0000
3	Navigation Aid	NaN	30.0000	7.0000	0.0000
4	Other	0.0500	12.1667	3.6667	0.0000
5	Passenger	3.9667	91.6333	16.0000	0.0000
6	Pleasure Craft	1.3943	23.4286	11.2857	0.0000
7	Sailing Vessel	1.0862	20.4483	5.4759	0.1034
8	Special Craft	2.7963	60.6648	13.3407	0.4444
9	Tanker	2.9176	170.2408	28.3964	0.9271
10	Unspecified	0.8000	20.0769	8.4615	0.1538

Fonte: Elaboração própria.

final pelo horário de entrada de início

- **Tempo da Operação de Saída:** Tempo decorrido entre o horário de saída final pelo horário de saída de início
- **Tempo da Operação de Praticagem:** Tempo decorrido entre o momento identificado como fim da praticagem (POB 2) e o do início da praticagem (POB 1), calculados com referência a área de dragagem 'Trecho_1Csoleira'
- **Tempo da Operação da Embarcação Atracada:** Tempo decorrido entre o horário do início da saída com o horário final de entrada, ou seja, o tempo no qual a embarcação ficou atracada na área de dragagem em questão
- **Tempo Total:** Tempo decorrido entre o horário final da praticagem e o horário 'NOR', indicando o intervalo de tempo mapeado de cada embarcação dentro do percurso considerado no modelo

Realizamos a simples subtração entre os dados da coluna *datetime* e, com os resultados gerados a partir deste cálculo, foi realizado um tratamento de forma que o resultado final retornasse o período identificado em formato numérico para possibilitar a melhor manipulação para as análises posteriores destes dados, uma vez que

os dados sem tratamento se apresentavam no formato "1 dia e 3 horas", por exemplo. Assim, foi realizada a conversão deste período para minutos, permitindo que os intervalos de tempo calculados estivessem como um dado numérico. A aplicação do tratamento do tempo e a geração da saída descritiva dos dados é o código 13 e está na página 85.

Tabela 6: Tabela com dados descritivos sobre os campos calculados de intervalos de tempo

	tempo_operacao_entrada	tempo_operacao_saida	tempo_operacao_praticagem	tempo_operacao_atracado	tempo_total
count	799.0000	1155.0000	8328.0000	1011.0000	8309.0000
mean	5331.9150	3862.4422	34333.3619	-653.4852	34250.0531
std	21177.4981	17353.2572	50348.9500	12443.9814	50214.9919
min	-122383.0833	-86.0333	0.0000	-128665.5000	0.0000
25%	5.4917	0.0000	2361.1333	-5.0833	2330.7500
50%	847.8333	50.6000	6555.9667	-5.0500	6573.7333
75%	2848.8500	473.9667	42664.9667	177.4667	42664.9667
max	163815.5500	179270.1500	185090.1333	102114.9500	185085.0833

Fonte: Elaboração própria.

Pelas métricas ilustradas na Tabela 6, é possível notar uma grande variação dos dados, ou seja, os dados não representam uma distribuição bem equilibrada pela base de dados. Podemos notar uma diferença entre dados de máximo e/ou de mínimo com o resto da distribuição nos quartis, o que possivelmente mostra a presença de *outliers* (dados muito superiores ou inferiores a distribuição da base, podendo ser ou de casos muito específicos que não representam o comportamento usual dos dados estudados ou uma coleta/inserção de dados com casos errados pontuais). Tendo isso em vista, foi realizado um tratamento destes dados anulando estes casos (para evitar uma redução caso fosse optado a exclusão deles) pelas equações a seguir:

$$IQR = Q_3 - Q_1 \quad (11)$$

$$LL = Q_1 - 1.5 * IQR \quad (12)$$

$$UL = Q_3 + 1.5 * IQR \quad (13)$$

Nas equações *LL* aponta o limite inferior (*lower limit*) enquanto *UL* aponta o limite superior (*upper limit*), *IQR* é o intervalo interquartil e Q_3 e Q_1 são os dados respectivos aos quartis 3 e 1. A partir das equações é possível identificar os dados *outliers*, dados ou maiores que *UL* ou menores que *LL*, superiores e inferiores respectivamente.

Desta forma, o tratamento pela retirada dos dados *outliers* foi realizado pelo có-

digo 12 da página 85 e novamente foi levantada a Tabela com a descrição dos dados (Tabela 7).

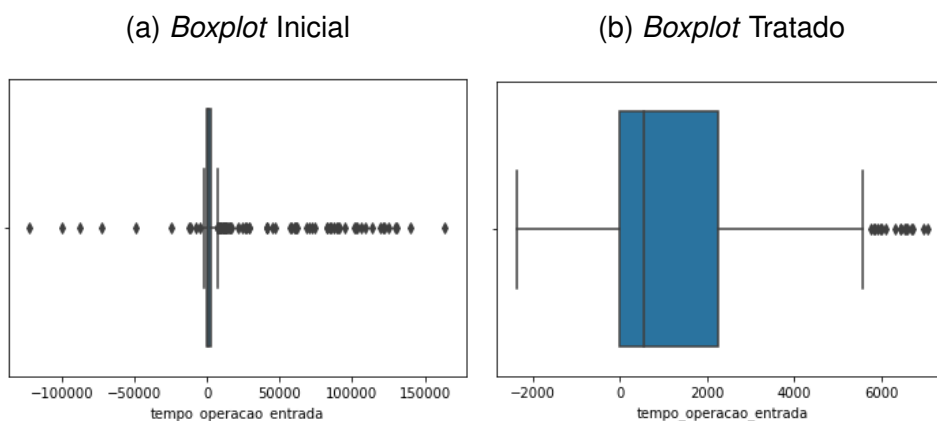
Tabela 7: Tabela com dados descritivos sobre os campos calculados com tratamento de *outliers* de intervalos de tempo

	tempo_operacao_entrada	tempo_operacao_saida	tempo_operacao_praticagem	tempo_operacao_atracado	tempo_total
count	713.0000	970.0000	7037.0000	771.0000	7039.0000
mean	1288.3595	144.7571	15481.2771	27.6601	15664.8231
std	1664.7365	237.6677	25001.4672	90.6886	25230.3465
min	-2366.5000	-86.0333	0.0000	-268.7667	0.0000
25%	5.1000	0.0000	1911.6333	-5.0833	1873.9833
50%	547.3667	10.1333	4525.7333	-5.0667	4575.9000
75%	2247.8000	213.1833	15997.8833	10.1833	16151.1333
max	7068.4500	1156.7667	102547.1667	431.4333	103152.4500

Fonte: Elaboração própria.

Utilizando a métrica do tempo de operação de entrada, conseguimos gerar uma visualização gráfica, correspondida por dois gráficos do tipo *Boxplot* que mostram, por meio de uma 'caixa', os dados presentes no interquartil (entre o Q_3 e o Q_1 , o *IQR*). Os dados dentro da régua são aqueles que estão dentro do limite inferior e superior do cálculo de *outliers*, enquanto os dados restantes são aqueles discrepantes (dados *outliers*).

Figura 31: Gráficos *Boxplot* da métrica de tempo de operação de entrada



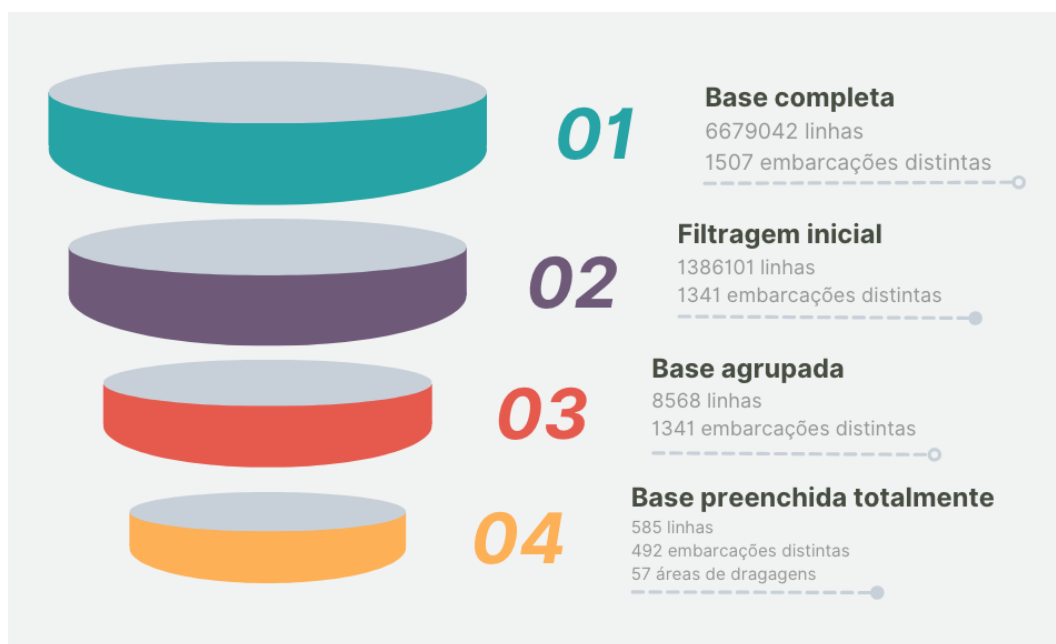
Fonte: Elaboração própria.

Pela Figura 31, é possível observar a importância da aplicação do tratamento de *outliers* e como a presença destes pode atrapalhar uma análise gráfica dos dados.

4 Discussão

O modelo desenvolvido neste projeto utilizou bases de dados do Porto de Santos e, utilizando técnicas de programação e análise de dados, passou por diversos tratamentos até o seu resultado final, gerando uma base capaz de mapear as operações das embarcações de acordo com locais designados dentro do Porto de Santos. Para ilustrar de forma enxuta a evolução das bases por meio do desenvolvimento do projeto, tem-se a Figura 32 que apresenta o tamanho da base (pela quantidade de linhas) e a quantidade de embarcações distintas depois da realização de diferentes processos do projeto. As etapas descritas são: a base completa obtida pela coleta de dados (1), base após aplicação da filtragem inicial (2), base agrupada por embarcação e área de dragagem (3), base agrupada filtrada quando todos os campos estão preenchidos (4). Desta forma, o presente projeto concentra os esforços no tema de Ciência de Dados ao processar e armazenar grandes volumes de dados e, a partir destes, realizar análises descritivas e exploratórias com o objetivo de gerar *insights* relevantes ao tema abordado.

Figura 32: Funil do tamanho da base de dados pelo desenvolvimento do modelo

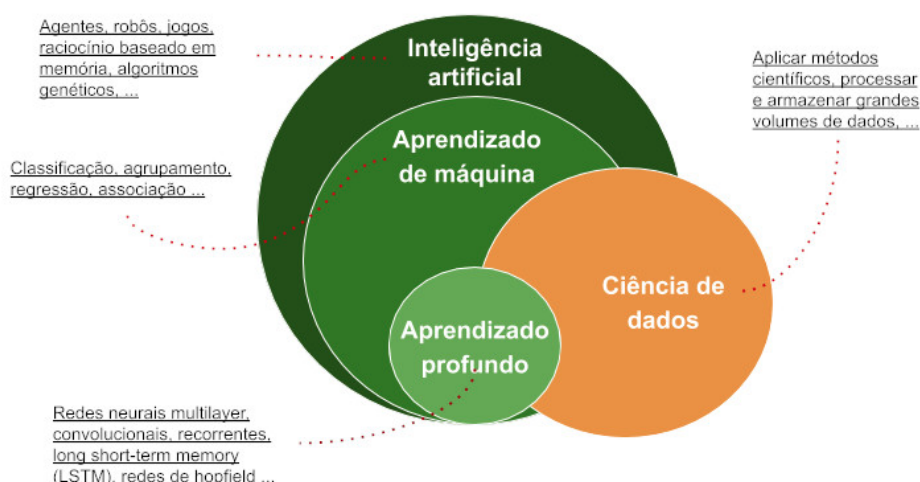


Fonte: Elaboração própria.

Ao comparar com as metodologias levantadas nas referências literárias, podemos segmentá-las em 3 tipos: modelos com análise de dados e uso de modelos de *Machine Learning* com Abreu et al., 2023, Mandalis et al., 2022 e Bonham et al., 2018, modelos de algoritmos de agendamento com Zhang et al., 2017 e Gan et al., 2021 e modelos de software com Yin et al., 2011. Ainda que todos estejam dentro do contexto portuário, os modelos mais próximos ao presente projeto abordam o tema de análise

de dados e modelos de *Machine Learning* como pode ser visto pela Figura 33, na qual Aprendizado de Máquina (ou *Machine Learning*) possui pontos de congruência com Ciência de Dados (tema atacado pelo projeto).

Figura 33: Relação entre inteligência artificial, aprendizado de máquina, aprendizado profundo e ciência de dados



Fonte: Dias, 2019.

O projeto de Abreu et al., 2023 mostra uma abordagem bem detalhada de um modelo de previsão do período de estadia no porto em 4 categorias: Baixa, Média, Alta e Muito Alta. Para isso, foram avaliados diversos modelos utilizados na literatura por métricas de desempenho, no qual o melhor avaliado foi o *Random Forest Classifier* com mais de 70% em Acurácia, *Recall* e Precisão. Já o modelo de Zhang et al., 2017 aborda outro objetivo para previsão, o Fluxo de Tráfego de Embarcações de um canal portuário. Desta maneira, construiu um modelo baseado em VRF (*Vehicle Route Forecasting*) capaz de prever a quantidade de embarcações numa determinada célula de grid em um tempo futuro. Para isso, foram utilizados 3 modelos para fazer uma análise comparativa sendo eles: *XgBoost*, ARIMA e LSTM, obtendo os melhores resultados com LSTM e *XgBoost* (a variar do horizonte de tempo previsto). Por fim, o projeto de Bonham et al., 2018 construiu um modelo de previsão de atrasos sendo agrupados em limites de atraso, sendo eles: 15, 30, 60, 90 e 120 minutos. Além do modelo, este estudo trouxe uma análise bem aprofundada dos dados tanto antes de aplicá-los no modelo (categorização das embarcações por um modelo de segmentação e *feature engineering*) como após (*feature importance*).

Estes 3 modelos apontados realizaram diferentes abordagens com o intuito de entender melhor as operações portuárias por meio da análise e construção de modelos de previsão de: tempo de estadia, fluxo de tráfego e atraso. Por outro lado, apenas o estudo de Bonham et al., 2018 realizou análises detalhadas sobre quais foram as mé-

tricas com maior relevância no resultado do modelo, o que permite um aprimoramento pela redução dos atrasos no Porto. Ao passo que, nos outros modelos, a melhoria se concentra em utilizar o resultado previsto como entrada para a tomada de decisão (como escolher horários com menos fluxo de tráfego para aumentar a quantidade de embarcações entrando no canal e melhorar a distribuição delas, para uniformizar o fluxo pelos períodos de tempo).

O presente projeto, por sua vez, não constrói um modelo de previsão de algum objetivo mas, foca no mapeamento das operações portuárias do Porto de Santos e análises da base concentrando as atividades na área de Ciência de Dados. Assim, o modelo desenvolvido obtém como resultado o fluxo de movimentação de embarcações associadas a uma área de dragagem em questão, por meio de uma tabela contendo os diversos momentos realizados pela embarcação durante a sua passagem no Porto de Santos. Com este resultado, é possível levantar a ordenação das embarcações ocorrida para determinada área de dragagem em um certo horário, atingindo o objetivo do projeto. Pode-se dizer, assim, que o resultado realiza um mapeamento das operações portuárias do Porto de Santos, principalmente no que se diz respeito a movimentação das embarcações que entram no canal.

5 Conclusões

A partir da metodologia descrita neste relatório, foi possível realizar o mapeamento das operações portuárias do Porto de Santos. Para isso, foram coletados dados da *marinetraffic* que utiliza a AIS como fonte de extração para mapear a situação de diversos portos no mundo todo. Depois de diversos processos e análises dos dados da base, o resultado final foi de uma base de 585 linhas com 492 embarcações distintas e 57 áreas de dragagem no formato de tabela. Esta tabela estava a nível de embarcação e área de dragagem contendo os horários de movimentação delas, sendo: NOR (*Notice of Readiness* - momento que aciona o Porto dizendo que a embarcação está pronta para entrar), horário de saída da área de fundeio, POB 1 (horário de início da praticagem de entrada), horário de início da entrada, horário de fim da entrada (horário de fim da praticagem de entrada), horário de início da saída (horário de início da praticagem de saída) e horário de fim da saída relacionados a área de dragagem em questão e, por fim, o POB 2 (horário de fim da praticagem de saída). Com essa tabela, foi possível calcular os tempos de algumas operações realizadas pelas embarcações, como tempo da operação de entrada, de saída, da praticagem e o tempo total de movimentação com o objetivo de possibilitar o estudo da eficiência dessas operações, além de identificar oportunidades de melhorias (como identificação de gargalos e momentos de subutilização). Além disso, com o objetivo de obter maior entendimento da base de dados trabalhada, foram realizadas diversas análises tanto descritivas como exploratórias dos dados onde foram obtidos *insights* relevantes. Alguns destes resultados foram: distribuição das velocidades das embarcações (geral e apenas da movimentação), métricas estatísticas dos tempos de operação e padrões de movimentação das embarcações. Dessa forma, o presente projeto possibilitou não apenas o mapeamento do fluxo de movimentação das embarcações por meio da análise de dados como uma maior compreensão das operações do Porto de Santos. Este projeto, portanto, centralizou a atenção na área de Ciência de Dados ao atuar desde a coleta e processamento dos dados até a identificação dos deslocamentos das embarcações e a realização de análises dos dados.

Como limitações do projeto, é importante ressaltar que a coleta de dados não continha todos os dados precisos para a indicação mais adequada dos horários mapeados, sendo necessárias algumas aproximações e considerações para atingir o resultado final (como adotar que o primeiro horário da embarcação no fundeadouro era o NOR e que os horários da praticagem estavam diretamente relacionados a área de dragagem 'Trecho_1Csoleira'). Além disso, os dados da base de dados não continham uma grande quantidade de histórico de embarcações nos fundeadouros, o que pode apontar uma limitação na forma como foi realizada a coleta de dados (ou a disponibilização destes no site utilizado - *marinetraffic*). Assim, para estudos futuros é relevante

e necessária a busca de outras fontes de dados, ou ainda, tentar identificar as limitações desta forma de coleta. Ademais, são necessários estudos de maior profundidade acerca dos critérios para a ordenação ocorrida, de forma a entender possíveis padrões nos dados e, assim, possivelmente, construir um modelo de previsão de qual a próxima embarcação a atracar em uma determinada área de dragagem, em um dado período de tempo.

Referências

- Abreu, L. R., Maciel, I. S., Alves, J. S., Braga, L. C., & Pontes, H. L. (2023). A decision tree model for the prediction of the stay time of ships in Brazilian ports. *Engineering Applications of Artificial Intelligence*, 117(June 2022), 105634. <https://doi.org/10.1016/j.engappai.2022.105634>
- Bonham, C., Noyvirt, A., Tsalamani, I., & Williams, S. (2018). Analysing port and shipping operations using big data. *ONS Data Science Campus*, (June). <https://datasciencecampus.ons.gov.uk/wp-content/uploads/sites/10/2018/06/here.pdf>
- de Transportes Aquaviários, A. -. A. N. (s.d.). *Desempenho Portuário - Estatístico Aquaviário*. Recuperado 28 outubro 2023, de <https://web3.antaq.gov.br/ea/sense/movport.html#pt>
- Dias, S. M. (2019). *Democratizando a Inteligência Artificial*. Recuperado 26 novembro 2023, de <https://www.serpro.gov.br/menu/noticias/noticias-2019/democratizando-a-inteligencia-artificial>
- Gan, S., Wang, Y., Li, K., & Liang, S. (2021). Efficient online one-way traffic scheduling for restricted waterways. *Ocean Engineering*, 237(August), 109515. <https://doi.org/10.1016/j.oceaneng.2021.109515>
- Heylin, M. (1972). Facts and figures. *Chemical and Engineering News*, 50(23), 3. <https://doi.org/10.1021/cen-v050n023.p003>
- Mandalis, P., Chondrodima, E., Kontoulis, Y., Pelekis, N., & Theodoridis, Y. (2022). Machine Learning Models for Vessel Traffic Flow Forecasting: An Experimental Comparison. *Proceedings - IEEE International Conference on Mobile Data Management, 2022-June(1500)*, 431–436. <https://doi.org/10.1109/MDM55031.2022.00094>
- MarineTraffic. (s.d.). *Live Map*. Recuperado 12 novembro 2023, de <https://www.marinetraffic.com/en/ais/home/centerx:-46.4/centery:-24.0/zoom:11>
- Santos, A. P. (s.d.). *Porto de Santos - Áreas de Fundeio*. Recuperado 4 outubro 2023, de <https://www.portodesantos.com.br/informacoes-operacionais/operacoes-portuarias/areas-de-fundeio/>
- Yin, X. F., Khoo, L. P., & Chen, C. H. (2011). A distributed agent system for port planning and scheduling. *Advanced Engineering Informatics*, 25(3), 403–412. <https://doi.org/10.1016/j.aei.2010.10.004>
- Zhang, X., Chen, X., Ji, M., & Yao, S. (2017). Vessel Scheduling Model of a One-Way Port Channel. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 143(5), 1–11. [https://doi.org/10.1061/\(asce\)ww.1943-5460.0000388](https://doi.org/10.1061/(asce)ww.1943-5460.0000388)

Apêndices

Códigos Implementados no Projeto

```
from aisexplorer.AIS import AIS
import time
from datetime import datetime

# loop para manter o código sempre ativo
while(1==1):
try:
    now = datetime.now()
    localtime = time.asctime(time.localtime(time.time()))
    url = "https://www.marinetraffic.com/en/data/asset_type=vessels&columns=flag,ship_
↪ name,imo,ship_type,dwt&flag_in|in|Brazil|flag_in=BR&dwt_between|range|dwt_bet_
↪ ween=0,410000&ship_type_in|in|Tugs%20%26%20Special%20Craft|ship_type_in=3"

    base = AIS(return_df = True).get_data_by_url(url)
    base["current_time"] = localtime

    time.sleep(5*60)
```

Código 1: *Webscrapping* dos dados da AIS no site da *Marinetraffic*

```
base_AIS_f['Estado_Navio'] = base_AIS_f['SPEED'].apply
else ('Manobrando' if x>0 else 'Parado'))

base_AIS_f['Janela'] = base_AIS_f['data_tratada']
.apply(lambda x: 1 if x.hour>1 and x.hour<=7 else
(2 if x.hour>7 and x.hour<=13 else
(3 if x.hour>13 and x.hour<=19 else 4))
```

Código 2: Criação das coluna 'Estado_Navio' e 'Janela'

```
df_gb = base_AIS_f.groupby('SHIPNAME')['Estado_Navio'].nunique()
df_3 = df_gb[df_gb['Estado_Navio']==3]
df_3['Flag_3Mov'] = 1

df_sort_3['Flag_Mudanca_Estado'] = np.where((df_sort_3['SHIPNAME'] ==
↪ df_sort_3['SHIPNAME_anterior']) & (df_sort_3['Estado_Navio']
!= df_sort_3['Estado_Navio_anterior']),1,0)
```

Código 3: Criação das colunas 'Flag_Mudanca_Estado' e 'Flag_3Mov'

```
# condições para cada categorização da mudança de estado
```

```

conditions = [
    # condições para a mudança 'Parou'
    (df_sort_3['SHIPNAME'] ==
     ↪ df_sort_3['SHIPNAME_antesior']) & (df_sort_3['Estado_Navio_antesior']
     ↪ == 'Movimentando') & (df_sort_3['Estado_Navio'] == 'Parado'),
    # condições para a mudança 'Moveu'
    (df_sort_3['SHIPNAME'] ==
     ↪ df_sort_3['SHIPNAME_antesior']) & (df_sort_3['Estado_Navio_antesior'] ==
     ↪ 'Parado') & (df_sort_3['Estado_Navio'] == 'Movimentando'),
    # condições para a mudança 'Iniciou manobra de parada'
    (df_sort_3['SHIPNAME'] ==
     ↪ df_sort_3['SHIPNAME_antesior']) & (df_sort_3['Estado_Navio_antesior'] ==
     ↪ 'Movimentando') & (df_sort_3['Estado_Navio'] == 'Manobrando'),
    # condições para a mudança 'Parou manobra e moveu'
    (df_sort_3['SHIPNAME'] ==
     ↪ df_sort_3['SHIPNAME_antesior']) & (df_sort_3['Estado_Navio_antesior'] ==
     ↪ 'Manobrando') & (df_sort_3['Estado_Navio'] == 'Movimentando'),
    # condições para a mudança 'Iniciou manobra de movimentação'
    (df_sort_3['SHIPNAME'] ==
     ↪ df_sort_3['SHIPNAME_antesior']) & (df_sort_3['Estado_Navio_antesior'] ==
     ↪ 'Parado')
    & (df_sort_3['Estado_Navio'] == 'Manobrando'),
    # condições para a mudança 'Parou manobra e estacionou'
    (df_sort_3['SHIPNAME'] ==
     ↪ df_sort_3['SHIPNAME_antesior']) & (df_sort_3['Estado_Navio_antesior'] ==
     ↪ 'Manobrando') & (df_sort_3['Estado_Navio'] == 'Parado')
]

# valores a serem categorizados
choices = [
    'Parou',
    'Moveu',
    'Iniciou manobra de parada',
    'Parou manobra e moveu',
    'Iniciou manobra de movimentação',
    'Parou manobra e estacionou'
]

# Criar coluna 'Mudanca_Estado' com base nos valores (choices)
df_sort_3['Mudanca_Estado'] = np.select(conditions, choices, default='Sem mudanças ou
↪ sem relação')

```

Código 4: Criação da coluna de 'Mudanca_Estado'

```

from shapely.geometry import Polygon, Point

def verificar_fundeio(dado, locais, margem_de_erro=0.01):

```

```

latitude, longitude = dado['LAT'], dado['LON']
ponto = Point(longitude, latitude)

for local_info in locais:
    pontos = local_info['pontos']
    poligono = Polygon(pontos)
    poligono_buffer = poligono.buffer(margem_de_erro)

    if poligono_buffer.contains(ponto):
        return local_info['nome']
    else:
        return "Local não identificado"

# criação da coluna que identifica o local de fundeio que está associado a
df_sort_3['Fundeio'] = df_sort_3.apply(verificar_fundeio, locais=areas_fundeio,axis=1)

```

Código 5: Aplicação da função 'verificar_fundeio'

```

# Define o sistema de coordenadas de destino (WGS84 para graus decimais)
wgs84 = pyproj.Proj(proj="latlong", datum="WGS84")
utm_zone = pyproj.Proj(proj="utm", zone=23, south=True, ellps="WGS84")

# Função que converte os dados de localização de utm para graus decimais
def utm_to_decimal(latitude_utm, longitude_utm, zone, is_south=True):
    # Define o sistema de coordenadas de origem (UTM)
    utm = pyproj.Proj(proj="utm", zone=zone, south=is_south, ellps="WGS84")

    # Define o sistema de coordenadas de destino (WGS84 para graus decimais).
    wgs84 = pyproj.Proj(proj="latlong", datum="WGS84")

    # Realiza a conversão de UTM para graus decimais.
    longitude, latitude = pyproj.transform(utm, wgs84, longitude_utm, latitude_utm)

    return latitude, longitude

```

Código 6: Conversão dos dados geográficos de UTM para graus decimais

```

def verificar_dragagem(dado, dragagem_geometria, dragagem_areas):
    latitude, longitude = dado['LAT'], dado['LON']
    ponto = Point(longitude, latitude)

    for i in range(len(dragagem_geometria)):
        geometry = dragagem_geometria[i]

        if geometry.contains(ponto):
            return dragagem_areas[i]

```

```

    return "Local não identificado"

# criação da coluna que identifica o local de fundeio que está associado a
df_sort_3['Area_Dragagem'] =
↳ df_sort_3.apply(verificar_dragagem,dragagem_geometria=terminal_geometria,
↳ dragagem_areas=terminal_areas, axis=1)

```

Código 7: Aplicação da função 'verificar_dragagem'

```

# Crie uma coluna auxiliar para verificar se a condição foi atendida
df_sort_3['Condicao'] = (
    (df_sort_3['SHIPNAME'] == df_sort_3['SHIPNAME'].shift(1)) &
    (df_sort_3['Area_Dragagem'] != df_sort_3['Area_Dragagem'].shift(1)))

# Utilize o método 'apply' para criar a coluna 'Estado_Local' com a string formatada
df_sort_3['Estado_Entrada_Area_Dragagem'] = df_sort_3.apply(lambda row: f"Entrou na
↳ {row['Area_Dragagem']}" if row['Condicao'] else '0', axis=1)

# Remova a coluna auxiliar 'Condicao' se não for mais necessária
df_sort_3.drop('Condicao', axis=1, inplace=True)

# Crie uma coluna auxiliar para verificar se a condição foi atendida
condicao = (
    (df_sort_3['SHIPNAME'] == df_sort_3['SHIPNAME'].shift(1)) &
    (df_sort_3['Area_Dragagem'] != df_sort_3['Area_Dragagem'].shift(1))
)

# Encontre os índices das linhas que atendem à condição
indices_condicao = df_sort_3.index[condicao]

# Atualize a coluna 'Estado_Saida_Local' nas linhas que atendem à condição
for idx in indices_condicao:
    idx_anterior = idx - 1
    if idx_anterior >= 0: # Verifica se o índice anterior é válido (não negativo)
        df_sort_3.at[idx, 'Estado_Saida_Area_Dragagem'] = f"Saiu da
↳ {df_sort_3.at[idx_anterior, 'Area_Dragagem']}"

# tratamento da coluna
df_sort_3['Estado_Saida_Area_Dragagem'] =
↳ df_sort_3['Estado_Saida_Area_Dragagem'].fillna(0)

```

Código 8: Categorização do tipo de movimentação para as Áreas de Dragagem

```

# Criacao dos campos
df_sort_3['flag_fundeio'] = np.where(df_sort_3['Fundeio']!='Local não
↳ identificado',1,0)

```

```

df_sort_3['hora_fundeio'] = df_sort_3.apply(lambda row: row['data_tratada'] if
↳ row['flag_fundeio'] == 1 else None, axis = 1)
df_sort_3['hora_saida_fundeio'] = df_sort_3.apply(lambda row: row['data_tratada'] if
↳ row['Estado_Saida_Fundeio'] != 0 else None, axis = 1)

# agrupamento e calculo dos campos a nivel embarcacao
df_fundeio = df_sort_3.groupby('SHIPNAME').agg({'flag_fundeio': 'max', 'hora_fundeio': '
↳ min', 'hora_saida_fundeio': 'first'}).reset_index()
df_fundeio.columns =
↳ ['SHIPNAME', 'flag_navio_fundeio', 'NOR', 'hora_saida_fundeio_navio']

# junta a base inicial
df_sort_3_fundeio = df_sort_3.merge(df_fundeio, on = 'SHIPNAME', how = 'left')

```

Código 9: Coleta dos horários relacionados ao fundeio das embarcações

```

# Cria coluna com a area de dragagem da linha seguinte
df_sort_3['Area_Dragagem_seguinte'] = df_sort_3['Area_Dragagem'].shift(-1)

# Início Entrada Operação

# Crie uma coluna auxiliar para verificar se a condição foi atendida
df_sort_3['Condicao'] = (
    (df_sort_3['Mudanca_Estado'] == 'Iniciou manobra de parada') &
    (df_sort_3['Area_Dragagem'] != 'Local não identificado')
)

# Utilize o método 'apply' para criar a coluna 'Estado_Local' com a string formatada
df_sort_3['fl_operacao_entrada_inicio'] = df_sort_3.apply(lambda row: '1' if
↳ row['Condicao'] else '0', axis=1)

# Remova a coluna auxiliar 'Condicao' se não for mais necessária
df_sort_3.drop('Condicao', axis=1, inplace=True)

# Fim Entrada Operação

# Crie uma coluna auxiliar para verificar se a condição foi atendida
df_sort_3['Condicao'] = (
    (df_sort_3['Mudanca_Estado'] == 'Parou manobra e estacionou') &
    (df_sort_3['Area_Dragagem'] != 'Local não identificado')
)

# Utilize o método 'apply' para criar a coluna 'Estado_Local' com a string formatada
df_sort_3['fl_operacao_entrada_fim'] = df_sort_3.apply(lambda row: '1' if
↳ row['Condicao'] else '0', axis=1)

# Remova a coluna auxiliar 'Condicao' se não for mais necessária

```

```

df_sort_3.drop('Condicao', axis=1, inplace=True)

# Início Saída Operação

# Crie uma coluna auxiliar para verificar se a condição foi atendida
df_sort_3['Condicao'] = (
    (df_sort_3['Mudanca_Estado'] == 'Iniciou manobra de movimentação') &
    (df_sort_3['Area_Dragagem'] != 'Local não identificado')
)

# Utilize o método 'apply' para criar a coluna 'Estado_Local' com a string formatada
df_sort_3['fl_operacao_saida_inicio'] = df_sort_3.apply(lambda row: '1' if
    ↪ row['Condicao'] else '0', axis=1)

# Remova a coluna auxiliar 'Condicao' se não for mais necessária
df_sort_3.drop('Condicao', axis=1, inplace=True)

# Fim Saída Operação

# Crie uma coluna auxiliar para verificar se a condição foi atendida
df_sort_3['Condicao'] = (
    #(df_sort_3['Mudanca_Estado'] == 'Iniciou manobra de movimentação') &
    (df_sort_3['Area_Dragagem'] != 'Local não identificado')
    & (df_sort_3['Area_Dragagem'] != df_sort_3['Area_Dragagem_seguinte'])
    & (~df_sort_3['Area_Dragagem'].isnull())
)

# Utilize o método 'apply' para criar a coluna 'Estado_Local' com a string formatada
df_sort_3['fl_operacao_saida_fim'] = df_sort_3.apply(lambda row: '1' if
    ↪ row['Condicao'] else '0', axis=1)

# Remova a coluna auxiliar 'Condicao'
df_sort_3.drop('Condicao', axis=1, inplace=True)

```

Código 10: Marcação das linhas críticas para a documentação

```

df_sort_3['hora_entrada_inicio'] = df_sort_3.apply(lambda row: row['data_tratada'] if
    ↪ row['fl_operacao_entrada_inicio'] == '1' else None, axis = 1)
df_sort_3['hora_entrada_fim'] = df_sort_3.apply(lambda row: row['data_tratada'] if
    ↪ row['fl_operacao_entrada_fim'] == '1' else None, axis = 1)
df_sort_3['hora_saida_inicio'] = df_sort_3.apply(lambda row: row['data_tratada'] if
    ↪ row['fl_operacao_saida_inicio'] == '1' else None, axis = 1)
df_sort_3['hora_saida_fim'] = df_sort_3.apply(lambda row: row['data_tratada'] if
    ↪ row['fl_operacao_saida_fim'] == '1' else None, axis = 1)

```

Código 11: Criação das colunas de horários dos momentos críticos da movimentação da embarcação

```

df_sort_3_fundeio_pob_gb = df_sort_3_fundeio_pob.groupby(['SHIPNAME', 'Area_Dragagem' ]
↳ ).agg({'NOR': 'min', 'hora_saida_fundeio_navio': 'max', 'POB_1': 'min', 'hora_entrada_
↳ inicio': 'max', 'hora_entrada_fim': 'max', 'hora_saida_inicio': 'max', 'hora_saida_fim'
↳ : 'max', 'POB_2': 'max'}).reset_index()

```

Código 12: Agrupamento da movimentação a nível embarcação e área de dragagem

```

# cria as colunas de horários pela subtração do horário final pelo inicial de cada
↳ operação
df_final['tempo_operacao_entrada'] = df_final['hora_entrada_fim'] -
↳ df_final['hora_entrada_inicio']
df_final['tempo_operacao_saida'] = df_final['hora_saida_fim'] -
↳ df_final['hora_saida_inicio']
df_final['tempo_operacao_praticagem'] = df_final['POB_2'] - df_final['POB_1']
df_final['tempo_operacao_atracado'] = df_final['hora_saida_inicio'] -
↳ df_final['hora_entrada_fim']
df_final['tempo_total'] = df_final['POB_2'] - df_final['NOR']

# trata as colunas de horários
df_final['tempo_operacao_entrada'] = df_final['tempo_operacao_entrada'].apply(lambda
↳ x: x.days * 24 * 60 + x.seconds/60)
df_final['tempo_operacao_saida'] = df_final['tempo_operacao_saida'].apply(lambda x:
↳ x.days * 24 * 60 + x.seconds/60)
df_final['tempo_operacao_praticagem'] =
↳ df_final['tempo_operacao_praticagem'].apply(lambda x: x.days * 24 * 60 +
↳ x.seconds/60)
df_final['tempo_operacao_atracado'] =
↳ df_final['tempo_operacao_atracado'].apply(lambda x: x.days * 24 * 60 +
↳ x.seconds/60)
df_final['tempo_total'] = df_final['tempo_total'].apply(lambda x: x.days * 24 * 60 +
↳ x.seconds/60)

# gera tabela descritiva sem a coluna ('flag_navio_fundeio')
df_final.drop(columns='flag_navio_fundeio').dropna().describe()

```

Código 13: Tratamento dos dados de horários e geração da tabela descritiva

```

def remove_outliers(df, column_name):

    # Extrai os valores da coluna de interesse
    data = df[column_name].values

    # Extrai dados nao nulos para o calculo de outliers
    dados_nao_nulos = data[~np.isnan(data)]

    # Calcula Q1, Q3 e IQR para os dados nao nulos

```

```

q1 = np.percentile(dados_ao_nulos, 25)
q3 = np.percentile(dados_ao_nulos, 75)
iqr = q3 - q1

# Define os limites inferior e superior para os outliers
lower_bound = q1 - 1.5 * iqr
upper_bound = q3 + 1.5 * iqr

# Substitui os outliers com NaN em valores nao nulos
df[column_name] = np.where((~np.isnan(df[column_name])) & ((df[column_name] <
↳ lower_bound) | (df[column_name] > upper_bound)), np.nan, df[column_name])

return df

```

Código 14: Função para tratar os *outliers*

Anexos

Tabelas das áreas de fundeadouros com pontos de latitude e longitude

Latitude	Longitude
23° 59' 21"S	46° 20' 44"W
23° 59' 21"S	46° 20' 07"W
23° 59' 58"S	46° 20' 20"W
23° 59' 58"S	46° 20' 44"W

Tabela 8: Pontos geográficos do Fundeadouro nº 1

Latitude	Longitude
24° 00' 42"S	46° 20' 06"W
24° 00' 42"S	46° 19' 37"W
24° 01' 29"S	46° 19' 37"W
24° 01' 27"S	46° 20' 26"W

Tabela 9: Pontos geográficos do Fundeadouro nº 2

Latitude	Longitude
24° 03' 45"S	46° 21' 05"W
24° 03' 45"S	46° 19' 16"W
24° 05' 14"S	46° 17' 56"W
24° 06' 00"S	46° 17' 57"W
24° 05' 58"S	46° 22' 06"W

Tabela 10: Pontos geográficos do Fundeadouro nº 3

Latitude	Longitude
24° 13.75' S	46° 13.90' W
24° 13.75' S	46° 20.38' W
24° 10.53' S	46° 22.70' W
24° 08.27' S	46° 22.70' W
24° 08.28' S	46° 17.88' W
24° 05.28' S	46° 17.88' W
24° 05.28' S	46° 10.00' W
24° 10.33' S	46° 09.97' W

Tabela 11: Pontos geográficos do Fundeadouro nº 4

Latitude	Longitude
24° 13.75' S	46° 13.90' W
24° 13.75' S	46° 20.38' W
24° 15.00' S	46° 19.38' W
24° 15.00' S	46° 15.35' W

Tabela 12: Pontos geográficos do Fundeadouro nº 5

Latitude	Longitude
24° 05' 00"S	46° 24' 27"W
24° 05' 00"S	46° 22' 45"W
24° 07' 12"S	46° 23' 42"W
24° 07' 12"S	46° 25' 27"W

Tabela 13: Pontos geográficos do Fundeadouro nº 6