



**PREVISÃO DA DURAÇÃO E PRODUTIVIDADE DE OPERAÇÕES DE
TRANSFERÊNCIA DE CARGA ENTRE EMBARCAÇÕES E UNIDADES
MARÍTIMAS CONSIDERANDO A INFLUÊNCIA DE CONDIÇÕES DE MAR**

Thonny Santos Jardim

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Produção, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Produção.

Orientador: Virgílio José Martins Ferreira Filho

Rio de Janeiro
Março de 2024

**PREVISÃO DA DURAÇÃO E PRODUTIVIDADE DE OPERAÇÕES DE
TRANSFERÊNCIA DE CARGA ENTRE EMBARCAÇÕES E UNIDADES
MARÍTIMAS CONSIDERANDO A INFLUÊNCIA DE CONDIÇÕES DE MAR**

Thonny Santos Jardim

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA
(COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE
DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ENGENHARIA DE PRODUÇÃO.

Examinada por:

Prof. Virgílio José Martins Ferreira Filho, D.Sc.

Prof^a. Juliana Souza Baioco, D.Sc.

Prof. Paulo Cesar Ribas, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2024

Jardim, Thonny Santos

Previsão da Duração e Produtividade de Operações de Transferência de Carga entre Embarcações e Unidades Marítimas Considerando a Influência de Condições de Mar / Thonny Santos Jardim. – Rio de Janeiro: UFRJ/COPPE, 2024.

X, 74 p.: il.; 29,7 cm.

Orientador: Virgílio José Martins Ferreira Filho
Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia de Produção, 2024.

Referências Bibliográficas: p. 69-72.

1. Movimentação de Carga. 2. Métodos de Regressão.
3. Logística Offshore. I. Ferreira Filho, Virgílio José Martins. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Produção. III. Título.

AGRADECIMENTOS

Agradeço a todos os que estiveram ao meu lado ao longo dessa grande jornada.

A Deus, por tudo o que tem proporcionado até aqui.

À minha família, por ser minha base.

Ao meu companheiro, Maykell, pela força e incentivo.

Aos amigos, pelo ânimo e companhia.

Ao LORDE/SAGE por ter sido uma segunda casa por tanto tempo.

Ao professor Virgílio por todo o suporte e presença marcante em minha caminhada acadêmica.

Aos professores do Programa de Engenharia de Produção – PEP/COPPE pela excelência.

Aos colegas Petrobras que muito contribuíram com o desenvolvimento da pesquisa, em especial ao Ricardo Leite.

À CAPES e à Petrobras pelo apoio financeiro.

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

**PREVISÃO DA DURAÇÃO E PRODUTIVIDADE DE OPERAÇÕES DE
TRANSFERÊNCIA DE CARGA ENTRE EMBARCAÇÕES E UNIDADES
MARÍTIMAS CONSIDERANDO A INFLUÊNCIA DE CONDIÇÕES DE MAR**

Thonny Santos Jardim

Março/2024

Orientador: Virgílio José Martins Ferreira Filho

Programa: Engenharia de Produção

Na indústria de exploração e produção de óleo e gás offshore, o abastecimento das unidades marítimas desempenha um papel crucial na manutenção da continuidade operacional. O transporte de carga, portanto, emerge como um elemento-chave para a sustentabilidade da indústria. Contudo, a transferência de carga entre embarcações está intrinsecamente condicionada a uma série de fatores, incluindo aqueles de natureza meteoceanográfica. A otimização do planejamento logístico, antecipando-se a esses fatores externos, torna-se de fundamental importância. Nesse contexto, o principal objetivo deste trabalho é desenvolver um modelo preditivo que estime a duração e a produtividade das operações de transferência de carga entre embarcações. Para alcançar esse propósito, propõe-se uma metodologia que abrange a análise exploratória de dados para a seleção e desenvolvimento de variáveis preditoras, incluindo informações sobre as condições do mar, e a avaliação de diversos modelos de previsão. A metodologia proposta foi aplicada a um estudo de caso onde dados operacionais reais e dados registros simulados de condição de mar são considerados. O modelo final, fundamentado em Support Vector Regression (SVR) para a duração e Random Forest Regressor (RFR) para eficiência, embora admita margem para aprimoramentos, oferece uma ferramenta útil para o planejamento logístico de suprimento.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

PREDICTION OF DURATION AND PRODUCTIVITY OF CARGO TRANSFER OPERATIONS BETWEEN VESSELS AND MARITIME UNITS CONSIDERING THE INFLUENCE OF SEA CONDITIONS

Thonny Santos Jardim

March/2024

Advisor: Virgílio José Martins Ferreira Filho

Department: Industrial Engineering

In the offshore oil and gas industry, maintaining operational continuity relies heavily on the timely supply of maritime units, making cargo transportation a critical component. However, the efficiency of cargo transfer operations between vessels is intricately tied to various factors, including meteoceanographic conditions. This study aims to develop a predictive model for estimating the duration and productivity of cargo transfers in offshore oil and gas operations. The methodology includes exploratory data analysis, incorporating sea condition data, and evaluating models such as Support Vector Regression (SVR) and Random Forest Regressor (RFR). While the model allows room for improvement, it serves as a useful tool for supply logistics planning. This underscores the importance of adaptable approaches and continuous refinement in predictive modeling for the complexities of offshore operations. The dynamic and ever-evolving nature of this approach is crucial for achieving efficiency and operational excellence in challenging maritime environments.

Sumário

1	INTRODUÇÃO	1
1.1	Objetivos	2
1.2	Estrutura do Trabalho.....	2
2	CONTEXTO DA APLICAÇÃO.....	4
2.1	O contexto logístico offshore	4
2.2	Embarcações de apoio logístico	5
2.3	Logística de cargas	7
2.3.1	Movimentação de Cargas	8
2.4	Trabalhos relacionados.....	13
3	MODELAGEM.....	20
3.1	Coleta de Dados	20
3.2	Processamento de Dados.....	20
3.2.1	Desenvolvimento de Variáveis.....	21
3.2.2	Remoção de Outliers	22
3.2.3	Seleção de Variáveis	24
3.2.4	Padronização	25
3.3	Modelos de Regressão.....	26
3.3.1	Modelos Paramétricos e não paramétricas	27
3.3.2	Regressão linear e regressão linear múltipla	28
3.3.3	Regressão Parcial de Mínimos Quadrados.....	29
3.3.4	SVM	29
3.3.5	Árvore de Decisão	30
3.3.6	Florestas Aleatórias	31
3.3.7	Redes neurais.....	31
3.4	Otimização e Análise de Modelos.....	33
3.4.1	Métricas de Avaliação	35
4	EXPERIMENTO.....	37
4.1	Definição do problema	37
4.2	Definição do Conjunto de Dados	39
4.3	Processamento dos Dados	45
4.3.1	Desenvolvimento de Variáveis.....	45
4.3.2	Remoção de Outliers	47
4.3.3	Seleção de Variáveis	50
4.4	Otimização dos Modelos de Previsão	54
4.5	Resultados e Avaliação da Previsão.....	58
5	CONCLUSÃO	66
5.1	Considerações Finais.....	66
5.2	Sugestão para Trabalhos Futuros	67
6	REFERÊNCIAS BIBLIOGRÁFICAS	69
	APÊNDICE I – MÉTRICAS DE AVALIAÇÃO	73

LISTA DE FIGURAS

Figura 1: Logística de apoio offshore	5
Figura 2: Fluxo esquemático da metodologia proposta.....	20
Figura 3: <i>Boxplot</i> para identificação de outliers: Método Tukey	23
Figura 4: Estrutura a árvore do <i>Isolation Forest</i> , com destaque para pontos outlier.	23
Figura 5: Validação cruzada de K ciclos: Tr são os subconjuntos de treinamento e Te são os subconjuntos de teste. Fonte: Evsukoff (2020)	34
Figura 6: Fluxo de otimização e avaliação dos modelos	35
Figura 7: Os movimentos de <i>roll</i> , <i>pitch</i> , <i>yaw</i> , <i>heave</i> , <i>sway</i> , <i>surge</i> de uma embarcação. Fonte: Winter (2018)	43
Figura 8: Histograma da distribuição das variáveis <i>wspd</i> e <i>wavhs</i>	47
Figura 9: Resultados do tratamento de outliers pelos métodos propostos.....	49
Figura 10: Histogramas de distribuição da Eficiência e da Duração.....	50
Figura 11: Matriz de correlação.....	51
Figura 12: Importância das variáveis para previsão da Duração.....	53
Figura 13: Importância as variáveis para previsão da Eficiência	53
Figura 14: Eficiência média por intervalo de valores de <i>max_engine</i>	54
Figura 15: Distribuição da métrica R^2 obtida por combinação de hiperparâmetros na validação cruzada para previsão da Eficiência no Cenário 1	56
Figura 16: Distribuição da métrica R^2 obtida por combinação de hiperparâmetros na validação cruzada para previsão da Eficiência no Cenário 2	57
Figura 17: Distribuição da métrica R^2 obtida por combinação de hiperparâmetros na validação cruzada para previsão da Eficiência no Cenário 3.	57
Figura 18: Distribuição da métrica R^2 obtida por combinação de hiperparâmetros na validação cruzada para previsão da Duração no Cenário 2.....	58
Figura 19: Distribuição das métricas de avaliação do conjunto de validação por cenário adotado.	62
Figura 20: Previsões de Duração pelo modelo SVR	63
Figura 21: Previsões de Duração pelo modelo SVR: Treino e Validação	63
Figura 22: Eficiência calculada a partir da duração prevista.....	63
Figura 23: Previsões de Eficiência pelo modelo RFR	64
Figura 24: Previsões de Duração pelo modelo RFR: Treino e Validação.....	64

Figura 25: Previsões de Eficiência pelo modelo SVR.....	64
Figura 26: Previsões de Duração pelo modelo RFR: Treino e Validação.....	64
Figura 27: RMSE por intervalo real da duração.....	65

LISTA DE TABELAS

Tabela 1: Resumo do material bibliográfico	18
Tabela 2: Cenários de condição de mar segundo a altura de onda, adaptado de Norlund & Gribkovskaia (2017).....	22
Tabela 3: Variáveis de condições meteoceanográficas	41
Tabela 4: Parâmetros operacionais relacionados ao PSV.....	41
Tabela 5: Variáveis de impacto das condições de mar sobre as embarcações	42
Tabela 6: Critérios aplicados para definição dos valores das variáveis por registro de operação.....	44
Tabela 7: Contagem de registros por variável categórica.....	45
Tabela 8: Criação de variáveis categóricas.....	46
Tabela 9: Relação de hiperparâmetros otimizados por modelo de previsão e seus respectivos conjuntos de valores testados.	55
Tabela 10: Métricas de avaliação dos modelos de previsão da duração	60
Tabela 11: Métricas de avaliação dos modelos de previsão da eficiência.....	61

1 INTRODUÇÃO

A logística, como conceituada por Ballou (2001), vai além da mera movimentação física de mercadorias, tornando-se um componente estratégico que agrega valor às operações. Dentro do contexto de Exploração e Produção (E&P), a logística offshore emerge como um ator central, coordenando o abastecimento de unidades marítimas com os materiais, produtos e recursos humanos necessários para manter a continuidade operacional.

Dentre as atividades fundamentais que compõem a logística de E&P, destacam-se o transporte marítimo de carga e o transporte aéreo de passageiros offshore. A logística de cargas é responsável por movimentar todo tipo de carga necessária para a operação de perfuração, produção e manutenção nas unidades marítimas. Os suprimentos, divididos em cargas gerais, granéis líquidos e granéis sólidos, são movimentados por embarcações especializadas, como as Platform Supply Vessels (PSVs). A eficácia das operações de abastecimento offshore depende significativamente da duração da movimentação desses suprimentos, influenciando o planejamento de viagens e a distribuição de cargas.

Porém, é notável que a movimentação de cargas entre embarcações, especialmente a operação de içamento por guindaste, é complexa e sujeita a rigorosos controles de segurança e regulamentações. Além disso, fatores operacionais e externos, como experiência da tripulação, condições meteoceanográficas, dimensões da embarcação e fatores ambientais, desempenham papéis essenciais na duração e viabilidade dessas operações. A análise desses fatores é desafiada pela ausência de registros históricos detalhados, dificultando a compreensão completa do impacto das condições meteoceanográficas na eficiência operacional.

Nesse contexto, a previsão da duração das operações de abastecimento em unidades marítimas permitiria otimizar recursos, reduzir custos e melhorar a segurança operacional. Esta pesquisa busca contribuir para o preenchimento dessa lacuna, utilizando dados operacionais e históricos para prever a duração e eficiência das operações de movimentação de carga, oferecendo *insights* para a otimização do planejamento logístico na indústria offshore de exploração e produção de óleo e gás.

O desafio central é desenvolver modelos de previsão robustos que considerem não apenas fatores operacionais, mas também as condições meteoceanográficas, uma vez que essas influenciam diretamente o êxito das operações. Contudo, a complexidade

inerente à interação de variáveis operacionais e condições do ambiente marinho demanda, além da coleta e disponibilização de dados confiáveis, uma abordagem analítica avançada para alcançar previsões confiáveis.

Portanto, o desenvolvimento deste trabalho se justifica pela importância de se antecipar a duração operacional de carga e descarga em unidades marítimas, conferindo uma base sólida para o planejamento logístico resiliente e otimizado. Ademais, a escassez de estudos abrangentes que integrem variáveis operacionais e meteoceanográficas para prever a duração e eficiência dessas operações específicas ressalta a necessidade de avanços nesse campo. Ao abordar essa lacuna, esta pesquisa visa contribuir para o conhecimento na indústria offshore de exploração e produção de óleo e gás, e para o aprimoramento da tomada de decisões e gestão estratégica nesse cenário.

1.1 Objetivos

O objetivo principal desse trabalho é obter um modelo para previsão da duração e eficiência da operação de transferência de carga entre embarcações e avaliar a influência das condições de mar sobre a operação.

Para ser possível alcançar o objetivo principal, alguns objetivos específicos foram postos como etapas a serem cumpridas:

- I. Realizar análise exploratória de dados, visando identificar as variáveis mais importantes para a previsão da variável alvo;
- II. Elaborar estratégias de *feature engineering* com base no conhecimento de especialistas;
- III. Avaliar diferentes modelos de previsão com diferentes abordagens, visando não só a previsão da duração das operações, mas também as influências das condições de mar sobre a duração;
- IV. Adotar uma estratégia de previsão com base nos modelos avaliados.

1.2 Estrutura do Trabalho

Esta dissertação está estruturada em seis capítulos, cada um desempenhando um papel específico.

O primeiro capítulo, Introdução, apresenta o tema da pesquisa e suas justificativas.

O segundo capítulo, Contexto da Aplicação, discorre sobre o ambiente no qual o problema abordado está inserido.

No terceiro capítulo, Modelagem, é detalhada a metodologia proposta para resolver o problema, delineando os passos e abordagens adotados, além de contextualizar os elementos logísticos das unidades marítimas.

No quarto capítulo, é descrito um experimento realizado com dados reais, aplicando a metodologia proposta.

No quinto capítulo, são apresentadas as considerações finais, resumindo os principais resultados e indicando possíveis direções para futuras pesquisas.

Finalmente, o sexto capítulo consiste na lista de referências bibliográficas.

2 CONTEXTO DA APLICAÇÃO

Neste capítulo será realizada uma apresentação acerca da logística offshore, ambiente que contextualiza o estudo deste trabalho.

2.1 O contexto logístico offshore

A ideia da logística como um componente estratégico que vai além da simples movimentação de mercadorias se aplica no contexto E&P, onde essa disciplina se destaca como um ator de extrema importância para a geração de valor da cadeia de produção. A logística offshore representa todo um segmento especializado da indústria que visa coordenar eficientemente o abastecimento das unidades de exploração e produção com os materiais, produtos e recursos humanos necessários para garantir a continuidade operacional, considerando a natureza desafiadora, remota e complexa das instalações de atendimento e operações realizadas.

Tal complexidade, inerente à logística de E&P, pode ser atribuída a fatores cruciais que definem esse contexto, como o elevado valor agregado à produção de petróleo, o alto custo das operações, demandando uma logística que otimize custos sem comprometer qualidade e eficiência, e ainda a urgência por entregas rápidas e eficazes, demandas variáveis, operações sujeitas a condições ambientais adversas e a necessidade de lidar com mudanças nos cronogramas dos clientes.

Dentre as atividades que caracterizam a logística de E&P, destacam-se o transporte marítimo de carga e o transporte aéreo de passageiros offshore. O transporte marítimo realizado por embarcações é essencial para movimentar grandes volumes de equipamentos, materiais e produtos entre as plataformas de produção e os centros de abastecimento em terra firme. Paralelamente, o transporte aéreo, predominantemente realizado por helicópteros, desempenha um papel crucial no deslocamento de pessoal para as plataformas. Essas operações dependem fortemente de uma infraestrutura robusta, incluindo portos, aeroportos e armazéns especializados. A gestão eficaz dessas instalações é determinante para garantir a eficiência do fluxo de materiais e pessoas.

Dessa maneira, a logística de E&P não apenas incorpora os princípios logísticos gerais, mas também adapta e especializa esses conceitos para atender às demandas específicas da indústria de exploração e produção de petróleo e gás. A Figura 1 representa de maneira integrada e dinâmica o movimento de produtos, materiais e

recursos humanos ao longo do sistema logístico de E&P, evidenciando a complexidade e a interdependência dos diversos elementos envolvidos.

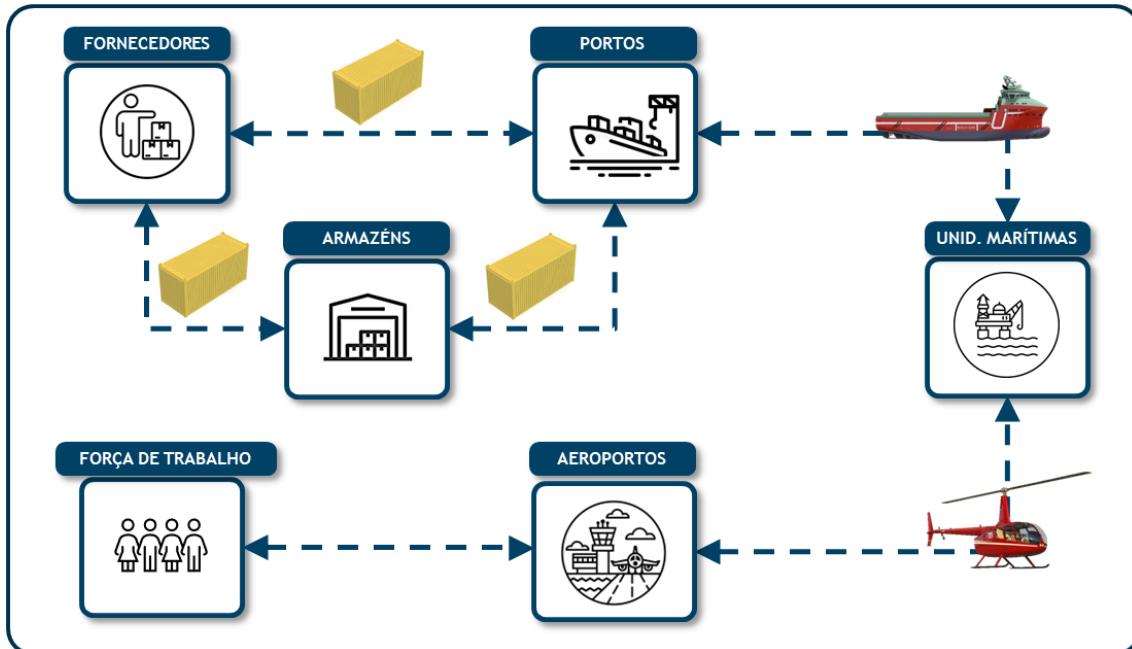


Figura 1: Logística de apoio offshore

O sistema logístico E&P na indústria de petróleo e gás atende a diversos clientes, incluindo navios sonda, unidades de produção, unidades de manutenção e segurança (UMS) e embarcações PLSV (Pipe-Laying Support Vessel). Cada cliente característico tem necessidades específicas decorrentes de suas funções na cadeia de produção, como abastecimento de materiais, suporte para operações de perfuração, fornecimento contínuo de insumos, remoção de produtos acabados e suporte para manutenção e instalação de dutos submarinos.

A movimentação de cargas por guindastes é uma atividade essencial nas operações de apoio offshore, desempenhando um papel crucial na transferência de cargas gerais unitizadas, itens volumosos, tubos, produtos químicos e até mesmo na execução de transbordo de passageiros entre embarcações e unidades marítimas.

2.2 Embarcações de apoio logístico

Cada tipo de embarcação utilizada no transporte de carga e apoio marítimo possui funções específicas e é projetada para atender a diferentes necessidades na indústria offshore. Em ABEAM (2023) faz-se também um grande apanhado dos tipos

de embarcações de apoio. Abaixo, são detalhados alguns dos principais tipos de embarcação:

- *PSV (Platform Supply Vessel)*: Fornece e abastece plataformas de exploração e produção offshore com uma variedade de cargas essenciais, como suprimentos, equipamentos, fluidos e produtos químicos. Possui capacidade para transportar uma ampla gama de cargas e oferecer suporte logístico às operações offshore. Suas características principais incluem:
 - *Operações de Carga e Descarga no Mar*: Os PSVs são especialmente projetados para realizar operações eficientes de carga e descarga em plataformas offshore. Essas operações abrangem o transporte de uma variedade de cargas, incluindo carga de convés, fluidos e granéis sólidos, fornecendo suprimentos essenciais às instalações offshore.
 - *Versatilidade de Carga*: Uma das principais vantagens dos PSVs é sua capacidade de transportar uma ampla gama de produtos, abrangendo desde equipamentos pesados até fluidos e granéis sólidos. Isso os torna peças fundamentais na cadeia logística das operações offshore.
 - *Classificação de Porte Bruto*: Os PSVs são classificados de acordo com o porte bruto, o que reflete sua capacidade de transporte e as especificações de carga. Os diferentes modelos, como PSV 1500, PSV 3000 e PSV 4500, indicam as capacidades específicas de cada embarcação.
 - *Restrições de Carga*: Ao contrário de restrições baseadas apenas no peso, nos PSVs, a limitação muitas vezes está relacionada à área do convés e ao volume dos tanques. Essas restrições são cruciais para garantir uma distribuição de carga segura e eficiente.
 - *Velocidade de Serviço*: Os PSVs geralmente operam a uma velocidade de serviço em torno de 10 nós. Essa velocidade é projetada para otimizar a eficiência das operações, garantindo a entrega rápida e segura de suprimentos às instalações offshore.
- *AHTS (Anchor Handling Tug Supply)*: Projetado para manusear âncoras e fornecer apoio durante as operações de ancoragem de plataformas e embarcações offshore. Além disso, oferece serviços de reboque e transporte de carga, sendo equipado com guinchos de grande capacidade.

- *OSRV (Oil Spill Response Vessel)*: Especializado em resposta a derramamentos de óleo e outros poluentes no ambiente marinho. Possui equipamentos de limpeza e contenção, sendo crucial para a mitigação dos impactos ambientais decorrentes de vazamentos de óleo.
- *LH (Line Handling)*: Desempenha um papel fundamental no manuseio de linhas e cabos durante operações offshore. Essas embarcações são frequentemente utilizadas para auxiliar em tarefas como atracação e desatracação de embarcações, assim como no manuseio de linhas durante as operações.
- *P (Passenger)*: Projetadas para transportar passageiros de e para instalações offshore. Essas embarcações focam na segurança e no conforto dos passageiros durante as viagens para as plataformas e outras estruturas marítimas.
- *UT (Utility Boat)*: Designação ampla para embarcações de uso geral que podem ser adaptadas para diversas funções, como transporte de pessoal, entrega de carga leve, e suporte em operações diversas. Sua versatilidade permite atender a várias demandas operacionais.

Essa variedade de embarcações reflete a complexidade e a diversidade de atividades na indústria. Cada tipo de embarcação desempenha um papel específico, contribuindo para a eficiência e segurança das operações em diferentes contextos.

2.3 Logística de cargas

O planejamento logístico para atender às Unidades Marítimas e outras demandas adota uma abordagem estratégica e adaptável, visando garantir eficiência. Para o atendimento periódico às UMs, são programadas operações regulares, ocorrendo de 1 a 2 vezes por semana ou a cada 5 a 10 dias, com o intuito de manter um fluxo contínuo de suprimentos e cargas gerais por meio de PSVs, otimizando a eficiência operacional. Em situações emergenciais, o cronograma prevê um atendimento dedicado com embarcações menores e mais ágeis, proporcionando uma resposta imediata a demandas críticas, sem uma periodicidade pré-estabelecida. Além disso, estratégias específicas são adotadas para o abastecimento de óleo diesel, atendimento de fluido, cimento e cascalho, transbordo entre UMs, buscando atender de maneira direcionada às diferentes necessidades, garantindo uma gestão eficiente dos recursos e insumos.

A logística de cargas é responsável por movimentar todo tipo de carga necessária para a operação de perfuração e produção das unidades marítimas. Os diversos tipos de carga aqui abordados englobam uma ampla gama de materiais, cada um com características específicas que requerem abordagens logísticas distintas. No que se refere à carga geral, que compreende itens como rancho, tubos, equipamentos e ferramentas, são necessárias estratégias logísticas focadas no armazenamento adequado, controle de estoque e manuseio especializado. Na categoria de granéis líquidos, que inclui fluidos de perfuração/completação, salmoura, olefina, água e diesel, é imperativo implementar logísticas específicas para garantir o transporte seguro e oportuno, assim como a integridade dessas substâncias. Já os granéis sólidos, como cimento, baritina e bentonita, demandam abordagens logísticas que considerem a proteção contra umidade, o manuseio adequado e a conformidade com padrões ambientais e regulamentações específicas, devido às suas aplicações em operações essenciais como a cimentação de poços e lamas de perfuração.

2.3.1 Movimentação de Cargas

A movimentação entre embarcações constitui uma operação de extrema importância e elevado grau de complexidade e risco. Essa atividade é conduzida sob rigoroso controle ativo, seguindo planejamentos minuciosos que se fundamentam em normativas internacionais e nacionais, bem como em diretrizes internas estabelecidas tanto pelas empresas fornecedoras do serviço de entrega quanto pelas empresas clientes. A eficiência da operação está intrinsecamente ligada a diversos fatores, sendo a natureza da carga um dos elementos preponderantes.

Içamento de Cargas por Guindaste

O içamento por guindaste assume papel crucial, sendo empregado na movimentação de cargas gerais unitizadas, itens volumosos, tubos, produtos químicos e até mesmo na execução de transbordo de passageiros. Essa operação, conhecida como lingada, comprehende distintas fases, cada uma desempenhando um papel fundamental no processo:

- *Elevação (ou Carga):* Nesta fase, a carga é erguida por meio de guindastes, iniciando o ciclo de movimentação.

- *Manobra Livre (ou Movimentação)*: Durante esta etapa, a carga é movida livremente para a posição desejada. A movimentação requer habilidade e precisão, pois a carga pode assumir diferentes posições e ângulos, exigindo controle preciso do guindaste.
- *Assentamento (ou Descarga)*: A fase final envolve o assentamento ou descarga da carga no local designado. A precisão e cuidado são essenciais para garantir a segurança da operação e evitar danos à carga ou à infraestrutura circundante.

Os guindastes desempenham um papel vital no sucesso dessas operações, sendo sua capacidade determinada pela soma do peso líquido estático, carga bruta estática (soma de todos os pesos reais aplicados no guindaste enquanto parado) e carga bruta dinâmica (incluindo os efeitos do movimento da peça). É crucial observar que o acréscimo na carga bruta estática, causado por inércia e movimento, deve ser minimizado. Portanto, a aceleração, frenagem e rotação do guindaste devem ocorrer de maneira controlada e lenta para evitar impactos adversos.

Os guindastes empregados nas operações de içamento em plataformas offshore, tanto fixas quanto flutuantes, frequentemente pertencem à categoria de "guindastes sobre pedestal" ou "*pedestal crane*". Nesse tipo de guindaste, a superestrutura é permanentemente fixada sobre um tubulão de aço, o qual é soldado ao casco ou à estrutura da plataforma. Essa configuração proporciona uma base robusta e estável, permitindo operações de içamento seguras e eficientes mesmo em ambientes marinhos desafiadores. A fixação do guindaste ao tubulão assegura a integridade estrutural necessária para lidar com cargas variadas, sendo uma escolha comum na indústria offshore devido à sua confiabilidade e capacidade de resistir às condições adversas encontradas no ambiente marítimo.

De acordo com a Norma Regulamentadora NR 37 (CTPP, 2020), que trata da segurança e saúde em plataformas de petróleo, a movimentação de cargas com guindastes está sujeita a restrições e proibições em situações específicas, visando a prevenção de riscos. Os casos nos quais é proibida a movimentação de cargas com guindaste incluem:

- *Iluminação Deficiente*: A movimentação de cargas com guindaste é proibida quando a iluminação no local é insuficiente. A adequada iluminação é essencial para garantir a segurança durante as operações de içamento.

- *Condições Climáticas Adversas ou Desfavoráveis:* Em situações de condições climáticas adversas ou desfavoráveis que exponham os trabalhadores a riscos, a movimentação de cargas com guindaste é proibida. Essas condições podem incluir ventos fortes, chuvas intensas, entre outras.
- *Inobservância das Limitações do Equipamento:* A movimentação de cargas com guindaste é proibida quando não são observadas as limitações do equipamento, conforme especificado no manual do fabricante ou fornecedor. O não cumprimento dessas limitações pode comprometer a segurança da operação.

Quanto às condições climáticas, a NR 37 estabelece critérios específicos para a movimentação de cargas com guindaste, considerando a velocidade do vento. Esses critérios são os seguintes:

- *Velocidade do Vento 0 a 38 km/h:* Permitidas todas as operações de movimentação de cargas.
- *Velocidade do Vento 39 a 49 km/h:* Acionamento de alarme sonoro a partir de 39 km/h; operações ordinárias devem ser interrompidas, e são permitidas apenas operações assistidas com observação contínua das condições climáticas.
- *Velocidade do Vento 50 a 61 km/h:* Permitidas apenas operações assistidas realizadas dentro da própria plataforma, com observação contínua das condições climáticas.
- *Velocidade do Vento Acima de 61 km/h:* Todas as operações devem ser interrompidas.

Essas medidas visam garantir a segurança dos trabalhadores envolvidos nas operações de movimentação de cargas, levando em consideração os potenciais riscos associados às condições ambientais.

LIEBHERR (2017), contudo, ressalta que as medidas padrão fornecida pelas normas podem não ser conservadoras o suficiente a depender da carga içada, e que um controle durante a operação precisa ser realizado. Isso porque a resistência exercida pela força do vento depende da superfície à qual o vento tem contato, neste caso, a área do material içado.

Outros parâmetros também interferem nas cargas de vento durante a operação. Alguns deles são dados que podem ser coletados no momento a operação, como o peso

da carga içada, a área máxima projetada da superfície, o coeficiente de resistência (C_w), e a velocidade do vento. Outros parâmetros devem ser calculados, como a área da superfície exposta, a pressão dinâmica e a força do vento.

A área projetada (A_P) refere-se à área da “sombra” projetada pela carga ao ser confrontada com o vento. A área projetada pode ser maior ou menor dependendo da direção do vento. O fabricante da carga pode fornecer a área projetada máxima. O coeficiente de resistência C_w é uma medida que descreve a obstrução que um corpo representa ao fluxo de ar. Esse coeficiente é utilizado para quantificar como a forma e o perfil de um objeto afetam a resistência ao movimento do ar ao redor dele. Também deve ser fornecido pelo fabricante.

Quanto aos parâmetros calculados, a área de superfície exposta ao vento (A_W) é dada pela Equação 1.

$$A_P = A_W \times C_w \quad \text{Equação 1}$$

A pressão dinâmica (p) é a pressão resultante do fluxo de ar ao redor de um objeto. Se o vento atinge uma placa fixada por uma mola, parte do vento é represado contra a superfície da placa, gerando um aumento de pressão conhecido como pressão dinâmica. A relação entre a velocidade do vento (v) e a pressão dinâmica é direta, quadruplicando quando a velocidade do vento dobra. Na Equação 2, ρ refere-se à densidade do ar, dada por $1,25 \text{ kg/m}^3$, e F_W refere-se à força do vento exercida pelo movimento do ar em um determinado ambiente. Essa força é gerada quando o ar se move em uma determinada direção e interage com objetos em seu caminho.

$$p = \frac{F_W}{A_W} = 0,5 \times \rho \times v^2 \quad \text{Equação 2}$$

LIEBHERR (2017) aborda outros elementos cruciais, incluindo a estrutura do guindaste e a velocidade máxima de vento permitida. A qualidade e a tecnologia dos guindastes, respaldadas por uma extensa experiência profissional e uma sólida formação dos operadores, juntamente com um planejamento prévio profissional das operações, contribuem significativamente para a redução do risco de acidentes durante a movimentação de cargas. Contudo, é essencial reconhecer que fatores imprevisíveis, como súbitas rajadas de vento, apresentam desafios difíceis e, muitas vezes, impossíveis de serem calculados com precisão antecipadamente.

Movimentação de Cargas Líquidas e Granéis

A movimentação de cargas líquidas e granéis entre embarcações é comumente realizada por meio de mangotes. Os mangotes, que são mangueiras industriais mais curtas, desempenham um papel crucial nesse processo, oferecendo flexibilidade, resistência à abrasão, altas temperaturas e produtos químicos.

Em sua maioria, os mangotes são fabricados a partir de borrachas de polietileno, garantindo propriedades específicas de acordo com a aplicação desejada. Essas características fazem dos mangotes uma escolha comum para diversas operações de transferência de cargas líquidas e granéis entre embarcações. Sua composição específica é adaptada para lidar com diferentes substâncias, proporcionando um excelente custo-benefício.

Os modelos de mangotes disponíveis no mercado são categorizados de acordo com suas aplicações específicas, abrangendo uma variedade de setores. Essas categorias incluem mangotes para ar e água, gases, solventes, vapor, óleo, areia e minério, alimentos e transparência. Cada tipo de mangote é projetado para atender aos requisitos específicos da carga a ser transferida, garantindo segurança e eficiência no processo.

2.4 Trabalhos relacionados

O trabalho proposto aborda a aplicação de métodos de aprendizado de máquina para previsão da duração e, assim, da eficiência, ou nível de produtividade de operações logísticas offshore considerando condições de mar. Deste modo, uma consulta bibliográfica foi realizada com o intuito de se observar como o problema tem sido tratado. Como poderá ser visto nesta seção, o tema é pouco discutido, e as condições de mar são mais comumente utilizadas para previsão de variáveis como velocidade de embarcações e consumo de combustível.

Embora não seja o tópico deste trabalho, a previsão ou simulação das condições ambientais do ambiente offshore é um tema recorrente na literatura. (Ali et al., 2021), por exemplo, aborda a previsão de duas características marítimas cruciais, a temperatura da superfície do mar e a altura significativa das ondas, aproveitando o grande volume e dimensionalidade dos dados marinhos disponíveis atualmente. O trabalho propõe a implementação de modelos estatísticos, de aprendizado profundo e de aprendizado de máquina para prever essas variáveis, utilizando dados reais da Agência Hidrográfica e Oceanográfica da Coreia. Os resultados experimentais revelam que o modelo de aprendizado profundo superou ligeiramente os modelos de aprendizado de máquina em desempenho geral, destacando uma lacuna significativa em relação ao modelo estatístico preditivo. A previsão de parâmetros marítimos é crucial para otimizar rotas de navegação, engenharia costeira, proteção ambiental e planejamento de atividades relacionadas ao mar, sendo essencial para decisões em tempo real e previsões de curto prazo. O estudo contribui ao introduzir uma arquitetura inovadora de modelo de redes neurais, além de oferecer uma comparação abrangente entre abordagens estatísticas, de aprendizado de máquina e de aprendizado profundo em um conjunto de dados real.

As condições ambientais são abordadas em diversos estudos para otimização de operações logísticas. O foco na determinação da velocidade das embarcações ou na viabilidade da realização de determinadas operações é destacado nos estudos a seguir.

Kisialiou et al. (2021), por exemplo, aborda um desafio crítico na logística offshore de produção de petróleo e gás, concentrando-se em resolver um problema de planejamento periódico de navios de abastecimento diante de incertezas na demanda e condições climáticas. O método combina uma metaheurística com um modelo de simulação de eventos discretos, permitindo a geração de cronogramas de navios com custo esperado mínimo sob condições de demanda e clima incertos. O estudo destaca a

importância de considerar tanto a incerteza na demanda quanto nas condições meteorológicas para um planejamento logístico mais eficaz. Os experimentos computacionais demonstram a eficiência da metodologia, oferecendo uma solução inovadora para o desafiador problema de planejamento de navios de abastecimento em ambientes offshore altamente incertos.

Norlund & Gribkovskaia (2017) abordam a influência das condições climáticas adversas nas operações de embarcações de abastecimento, com foco especial nas ondas como fator crucial para os planejadores operacionais. O estudo considera o clima como uma variável estocástica univariada, especificamente a altura significativa das ondas, modelada como uma cadeia de Markov para capturar dependências temporais. As condições climáticas variam sazonalmente e entre diferentes localidades, sendo abordadas por meio de um modelo estatístico que leva em consideração a variação mensal. O impacto do clima na duração e consumo de combustível de uma viagem é analisado, destacando que estratégias de otimização de velocidade podem resultar em aumento na duração da viagem em condições climáticas adversas. Para avaliar esses impactos, os autores desenvolvem uma ferramenta de simulação-otimização, demonstrando sua eficácia por meio de experimentos em instâncias reais, onde condições climáticas de inverno aumentam significativamente o consumo semanal de combustível e o número de embarcações necessárias. A aplicação de estratégias de otimização de velocidade, considerando o impacto do clima, permite reduções significativas no consumo de combustível, mas também destaca a necessidade de equilibrar objetivos ambientais e custos operacionais na geração de cronogramas de embarcações.

Gilbert et al. (2019) aborda a previsão de acesso a locais de operação em parques eólicos offshore, concentrando-se na previsão das condições durante a transferência de pessoal entre estruturas offshore e embarcações. Segundo os autores, as decisões de despacho/agendamento são baseadas em previsões univvaloradas da altura significativa das ondas de um modelo de previsão numérica do tempo. O estudo propõe avançar além da métrica tradicional da altura significativa das ondas, utilizando uma metodologia orientada por dados para estimar o movimento da embarcação durante a transferência. A relação entre o movimento vertical medido da embarcação e as condições das ondas, considerando altura significativa, período e direção de pico, é mapeada por meio de modelos aditivos generalizados para localização, escala e forma (GAMLSS). O estudo

demonstra a eficácia desse modelo em comparação com diferentes formulações, oferecendo uma abordagem operacional baseada em previsões de cenários temporais das medições de boias de ondas.

Pérez-Canosa et al. (2022) aborda a análise do movimento de rolamento de navios, com foco nos efeitos das condições do mar no comportamento da embarcação. O problema central reside na necessidade de compreender o comportamento do navio durante o movimento de rolamento em diversas condições de ondas, visando melhorar a segurança geral, especialmente no que diz respeito ao projeto de amarração de carga não padronizada. O estudo utiliza modelos matemáticos para avaliar o movimento rotacional do navio em torno do eixo longitudinal, abordando tanto o rolamento em águas calmas quanto o induzido por ondas regulares em condições específicas de mar.

Ogura et al. (2021) aborda a eficiência da cadeia de suprimentos associada ao transporte marítimo internacional, por meio de um método que visa atuar sobre o problema da influência de fatores como as condições meteorológicas no trajeto marítimo, frequentemente causador de atrasos nas chegadas programadas aos portos de destino. Para superar a limitação de estudos anteriores, que não consideravam o impacto das condições meteorológicas futuras, o estudo propõe um método de previsão que utiliza aprendizado bayesiano para calcular a velocidade da viagem levando em consideração tais condições. O método proposto demonstra uma precisão de previsão 28% superior em comparação com estudos anteriores, com base em dados operacionais históricos de embarcações que transportam cargas das indústrias de eletrodomésticos e automóveis.

Vários estudos recentes exploraram métodos para estimar o Consumo de Óleo Combustível e otimizar a eficiência energética de navios. Gkerekos et al. (2019) traz uma abordagem mais voltada para a previsão de variáveis. O estudo focaliza na previsão do Consumo de Óleo Combustível (FOC) de motores principais de navios, crucial para a sustentabilidade e rentabilidade das operações marítimas. Comparando algoritmos de regressão, como Support Vector Machines (SVMs), Random Forest Regressors (RFRs), Extra Trees Regressors (ETRs) e Artificial Neural Networks (ANNs), o estudo também considera diferentes estratégias de aquisição de dados a bordo: relatórios e sistemas de Registro e Monitoramento Automático de Dados (ADLM). Os modelos ETR e RFR mostraram o melhor desempenho em ambos os casos, sendo a existência de um sistema ADLM associada a uma precisão 7% maior e uma redução de até 90% no período

necessário para coleta de dados. Esses modelos são capazes de prever com precisão o FOC de navios sob diversas condições, como carga, clima, velocidade, distância e calados.

Bal Beşikçi et al. (2016) emprega uma metodologia pautada no uso de ANNs para prever o consumo de combustível em várias condições operacionais, utilizando dados operacionais diários conhecidos como “*Noon Data*”. No estudo, uma comparação do método com a análise de regressão múltipla confirma a superioridade da ANN. Além disso, é desenvolvido um Sistema de Apoio à Decisão baseado na ANN para operações eficientes em tempo real. O estudo destaca a importância de considerar condições ambientais, como vento e mar, na previsão de consumo de combustível e destaca a viabilidade do DSS como uma abordagem estratégica para tomada de decisões operacionais, considerando aspectos econômicos e ambientais.

Krata et al. (2021) aborda o desafio de prever com precisão a velocidade e o consumo de combustível de uma embarcação, considerando as condições de carga, previsões meteorológicas e outros parâmetros. O problema é abordado por meio de modelagem determinística e da aplicação de ANNs. A sensibilidade da ANN a diferentes configurações é analisada, incluindo a função de ativação, com destaque para o desempenho superior da função sigmoide. Os resultados indicam alta precisão na reprodução dos dados de velocidade e consumo de combustível, com notável agilidade na execução.

Lu et al. (2015) aborda o problema da otimização de viagens marítimas como uma prática para aumentar a eficiência energética e reduzir as emissões de gases de efeito estufa na indústria naval. O foco está na previsão do desempenho operacional do navio. Para isso, o estudo apresenta um método para prever a resistência adicional causada por ondas e vento específico para um tipo de navio. Além disso, é proposto um modelo semiempírico para prever o desempenho operacional do navio, permitindo a investigação da relação entre o consumo de combustível e as diversas condições do mar durante a viagem. O estudo utiliza o conceito de Eficiência Energética de Operação como indicador, demonstrando a relação entre o consumo de combustível e o trabalho de transporte.

Diversos outros estudos abordam métodos de estimativa da FOC, considerando ou não condições ambientais. Bialystocki & Konovessis (2016) realizaram uma análise estatística de relatórios de *Noon Data* de um navio Ro-Ro, identificando influências de

fatores como calado, deslocamento e condições meteorológicas e desenvolveram um algoritmo para aproximar o FOC baseado em curvas ajustadas para diferentes estados do mar. Meng et al. (2016) propuseram uma metodologia de pré-processamento de dados e dois modelos de regressão para vincular dados disponíveis ao FOC, considerando velocidade e condições meteorológicas. Simonsen et al. (2018) sugeriram o uso de dados do Sistema de Identificação Automática para estimar o FOC de navios de cruzeiro em águas norueguesas, destacando a utilidade para estimar emissões de gases de efeito estufa. Lundh et al. (2016) desenvolveram um método para estimar o FOC de navios com sistemas diesel-elétricos, otimizando o uso de geradores. Tsitsilis & Theotokatos (2018) criaram uma metodologia sistemática para o gerenciamento de energia de motores principais, combinando análise estatística e análises de energia. Wang et al. (2018) propuseram um modelo Lasso para estimar o FOC, mostrando desempenho otimizado em comparação com outros modelos.

A Tabela 1 resume alguns dos materiais bibliográficos expostos neste capítulo.

Tabela 1: Resumo do material bibliográfico

Artigo	Problema Abordado	Método Aplicado
Ali et al., 2021	Previsão de temperatura do mar e altura das ondas	Modelos estatísticos, aprendizado profundo, aprendizado de máquina
Kisialiou et al., 2021	Planejamento periódico de navios de abastecimento	Metaheurística, simulação de eventos discretos
Norlund & Gribkovskaia, 2017	Influência de condições climáticas adversas nas operações	Modelo estatístico com cadeia de Markov para altura significativa das ondas
Gkerekos et al., 2019	Previsão do Consumo de Óleo Combustível (FOC)	Support Vector Machines (SVMs), Random Forest Regressors (RFRs), Extra Trees Regressors (ETRs), Artificial Neural Networks (ANNs)
Gilbert et al., 2019	Previsão de acesso a locais de operação em parques eólicos	Modelos aditivos generalizados para localização, escala e forma (GAMLSS)
Pérez-Canosa et al., 2022	Análise do movimento de rolamento de navios	Modelos matemáticos para avaliar o movimento rotacional do navio em diferentes condições de ondas
Krata et al., 2021	Previsão de velocidade e consumo de combustível	Modelagem determinística, Redes Neurais Artificiais (ANN)
Lu et al., 2015	Otimização de viagens marítimas para eficiência energética	Método para prever a resistência adicional causada por ondas e vento, modelo semiempírico para prever o desempenho operacional do navio
Bal Beşikçi et al., 2016	Previsão de consumo de combustível em várias condições	Redes Neurais Artificiais (ANN), Sistema de Apoio à Decisão (DSS)
Ogura et al., 2021	Melhoria da eficiência na cadeia de suprimentos do transporte marítimo	Aprendizado bayesiano para calcular a velocidade da viagem considerando condições meteorológicas futuras

Em suma, a literatura revisada aborda de maneira abrangente a importância de parâmetros ambientais para otimizar operações logísticas offshore. Os estudos discutidos variam desde a previsão de condições climáticas, como temperatura do mar e altura das ondas, até a previsão de variáveis específicas, como consumo de combustível e eficiência operacional de embarcações considerando parâmetros ambientais. A aplicação de métodos de aprendizado de máquina e modelos estatísticos emerge como uma abordagem eficaz para lidar com a complexidade dos dados marítimos disponíveis. A literatura destaca a relevância dessas previsões para otimizar rotas de navegação, planejamento logístico e tomada de decisões em tempo real. Além disso, ressalta a necessidade de considerar tanto a incerteza nas condições ambientais quanto na demanda para um planejamento logístico mais eficaz. As contribuições dos estudos incluem a introdução de modelos inovadores, como o modelo GRU, e a avaliação comparativa entre abordagens estatísticas, de aprendizado de máquina e de aprendizado profundo. Portanto, há espaço significativo na literatura para a continuidade dessas pesquisas, visando aprimorar a eficiência e sustentabilidade das operações logísticas offshore.

Muitos outros trabalhos podem ser encontrados na literatura. A grande maioria deles, que consideram condições de meteoceanográficas como parâmetro de modelos de previsão, debruçam-se sobre o problema da otimização da velocidade dos navios ou do consumo de combustível. Não há, contudo, um aprofundamento acerca da previsão de parâmetros de eficiência das operações logísticas em si, nem do impacto das condições de mar sobre essas operações, o que tem muita importância dado que a incerteza dessas operações é capaz de alterar todo um planejamento de viagem para suprimento de unidades marítimas.

3 MODELAGEM

Este capítulo aborda o procedimento metodológico adotado para modelagem do problema proposto: prever a eficiência de operações de movimentação de carga entre embarcações, a partir da duração, que é uma variável desconhecida. Para isso, foi determinado um fluxo simples de etapas, conforme Figura 2.



Figura 2: Fluxo esquemático da metodologia proposta

A seguir, é realizado um detalhamento de cada etapa e as tarefas compreendidas em cada uma delas

3.1 Coleta de Dados

Esta etapa consiste no levantamento das bases de dados disponíveis e suas respectivas características. O objetivo principal é estabelecer as premissas necessárias para a obtenção de uma base de dados consolidada e robusta, que será fundamental na fase subsequente de modelagem do problema.

3.2 Processamento de Dados

Durante esta fase, é conduzida uma análise exploratória abrangente com o intuito de conhecer a estrutura e a natureza dos dados disponíveis. Esta análise proporciona *insights* cruciais sobre padrões, distribuições e possíveis peculiaridades nos dados. Tais informações são essenciais para a tomada de decisões informadas sobre a abordagem mais adequada na construção do modelo.

Além disso, a análise exploratória envolve a identificação e tratamento de dados ausentes, outliers e a compreensão da relação entre diferentes variáveis. Isso contribui para garantir a qualidade e confiabilidade dos dados que serão utilizados na modelagem posterior. A seguir, as tarefas adicionais complementares à análise exploratória dos dados.

3.2.1 Desenvolvimento de Variáveis

A engenharia de atributos, também conhecida como *feature engineering* em inglês, é um processo fundamental no desenvolvimento de variáveis para modelos preditivos. Este processo envolve a seleção, manipulação e transformação de dados brutos, resultando na criação de novas variáveis que não estão originalmente presentes no conjunto de treinamento. Tais transformações podem ser realizadas por meio da aplicação de conhecimento especializado ou de práticas estabelecidas, como a seleção de variáveis correlacionadas, entre outras técnicas.

No contexto específico deste estudo, além da aplicação de práticas convencionais, o objetivo é enriquecer o modelo com informações adicionais, gerando novas variáveis categóricas a partir de fontes externas. Por exemplo, ao considerar as diretrizes da Norma Regulamentadora NR 37 (CTPP, 2020) para a movimentação de cargas com guindastes, há restrições específicas relacionadas a condições climáticas adversas, iluminação deficiente e desrespeito às limitações do equipamento.

Quanto às condições climáticas, conforme já apresentado, a NR 37 estabelece critérios específicos para a movimentação de cargas com guindaste, considerando a velocidade do vento de 0 a 38 km/h, entre 39 e 49 km/h, 50 e 61 km/h e acima de 61 km/h.

Além disso, outros critérios, propostos por Norlund & Gribkovskaia (2017) e Halvorsen-Weare & Fagerholt (2011b), classificam as condições do mar com base na altura das ondas. Esses critérios, originalmente desenvolvidos para o Mar do Norte, são adaptados ao contexto deste estudo para classificar dados de condições marítimas, embora sejam reconhecidas as diferenças ambientais entre o cenário base desses estudos e o deste, além da operação marítima considerada. Os cenários estabelecidos encontram-se descritos na Tabela 2.

Tabela 2: Cenários de condição de mar segundo a altura de onda, adaptado de Norlund & Gribkovskaia (2017)

Perfil climático	Altura da onda (m)	Aumento do tempo de serviço (%)
1	< 2.5	0
2	[2.5, 3.5)	20
3	[3.5, 4.5)	30
4	≥ 4.5	Aguardando condição climática

Assim, três variáveis categóricas são adicionadas ao modelo, representando a luminosidade ambiente (dia ou noite, com base no horário de início da operação), a classificação da velocidade do vento e a classificação da altura das ondas. Essas variáveis enriquecem a análise, fornecendo ao modelo informações valiosas sobre o comportamento das embarcações em diferentes cenários operacionais.

Além disso, uma vez disponíveis dados o suficiente para a análise, variáveis adicionais devem ser implementadas para tentar capturar os efeitos do vento descritos na Seção 2.3.1 deste estudo, definindo, por exemplo, a força exercida pelo vento sobre a carga içada, dentre outros fatores detalhados em LIEBHERR (2017).

3.2.2 Remoção de Outliers

A identificação e eliminação de valores atípicos (*outliers*) constituem um procedimento crucial, uma vez que tais valores têm o potencial de comprometer o desempenho dos modelos (Evsukoff, 2020). Neste trabalho, propõe-se a utilização de três métodos distintos para a detecção de outliers, visando selecionar a abordagem mais adequada ao conjunto de dados em questão.

Na abordagem inicial, conhecida como método Tukey (Tukey, 1977) baseia-se na utilização do gráfico conhecido como *boxplot*, ilustrado na Figura 3. Este gráfico é composto por caixas que são calculadas para cada variável, representando a chamada distância interquartil (IQR). A IQR é a diferença entre o valor do 75º percentil (terceiro quartil, $Q3$) e o valor do 25º percentil (primeiro quartil, $Q1$). Essa medida representa o intervalo que contém 50% dos registros da variável. A linha no interior da caixa representa a mediana (50º percentil) e auxilia na identificação de distribuições assimétricas caso não esteja centralizada na caixa.

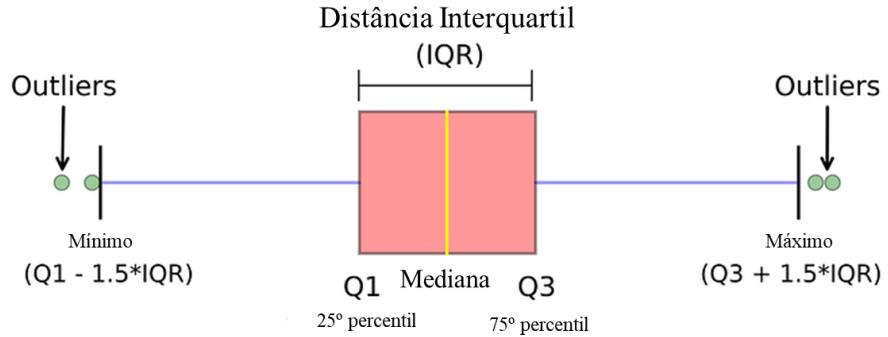


Figura 3: *Boxplot* para identificação de outliers: Método Tukey

As linhas à direita e à esquerda da caixa delimitam intervalos $[Q3, Q3 + 1,5IQR]$ e $[Q1 - 1,5IQR, Q1]$, respectivamente. Valores que se encontram fora desses intervalos são considerados outliers para a variável correspondente.

Outra abordagem é o algoritmo *Isolation Forest* (Liu et al., 2008) para a detecção de outliers. O *Isolation Forest* fundamenta-se na ideia de que pontos de dados anômalos são mais facilmente separados do restante da amostra. Para isolar um ponto, o algoritmo realiza partições recursivas na amostra, escolhendo aleatoriamente atributos e valores de divisão. A estrutura do *Isolation Forest* representa esse particionamento, onde o comprimento do caminho até um nó terminal, a partir da raiz, indica a probabilidade de anomalia. Os pontos mais fáceis de isolar, portanto, têm menores comprimentos de caminho.

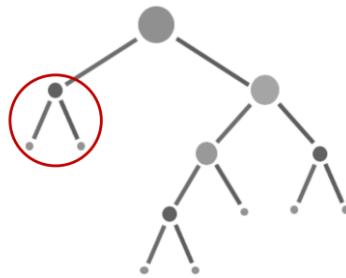


Figura 4: Estrutura a árvore do *Isolation Forest*, com destaque para pontos outlier.

Finalmente, incorpora-se o método baseado nas distâncias euclidianas entre todos os registros da base de dados. Este método, ao considerar simultaneamente todas as variáveis, utiliza a média ordenada das distâncias de um registro para todos os outros

registros. Um registro é considerado um outliers quando a média ordenada das distâncias desses registros é maior que um parâmetro definido.

Essas estratégias de detecção de outliers, juntamente com a análise da distância euclidiana entre registros, proporcionam uma abordagem abrangente para lidar com valores atípicos e melhorar a robustez dos modelos preditivos.

3.2.3 Seleção de Variáveis

A seleção de variáveis desempenha um papel crucial na construção de modelos preditivos robustos e interpretações eficazes. Neste trabalho, estuda-se a aplicação de duas abordagens. A primeira, prioriza a análise de correlação entre as variáveis preditoras. A segunda, ordena as variáveis mais importantes para a predição da variável alvo.

Inicialmente, realiza-se uma análise das relações de correlação entre as variáveis, visando identificar padrões de associação. A correlação é uma medida estatística que descreve o grau de relação linear entre duas variáveis, não implicando em causalidade. Usualmente, a correlação entre duas variáveis é calculada pelo coeficiente de Pearson, que varia de -1 a 1, sendo +1 uma correlação positiva perfeita, -1 uma correlação negativa perfeita, e 0 indica nenhuma correlação linear.

Pares de variáveis altamente correlacionadas são submetidos a uma etapa de filtragem, com o objetivo de mitigar multicolinearidade e otimizar a estabilidade do modelo. O critério adotado para a escolha das variáveis a serem mantidas no modelo não se limita-se apenas à força da correlação, mas incorpora a importância relativa das variáveis como critério de escolha. A importância das variáveis, ou *feature importance*, fornece *insights* sobre quais características são mais relevantes para o modelo. O algoritmo Random Forest calcula a importância das variáveis de duas maneiras principais: a importância Gini (ou redução média da impureza) e a redução da variância.

A importância Gini é derivada da estrutura do Random Forest que, como já apresentado, consiste em um conjunto de Árvores de Decisão. Cada árvore é composta por nós internos e folhas. Nos nós internos, as decisões de divisão do conjunto de dados são baseadas em critérios como impureza de Gini, ganho de informação (para tarefas de classificação) ou redução de variância (para regressão). A importância de uma variável é então medida pela média de como ela reduz a impureza das divisões nos nós internos. Essa média sobre todas as árvores na floresta compõe a medida de importância da variável.

Este valor também é a referência utilizada na segunda abordagem de seleção de variáveis utilizada neste estudo, onde as variáveis são ordenadas de acordo com a sua importância para o modelo Random Forest. As variáveis abaixo de um limiar específico são consideradas irrelevantes e podem ser removidas.

Ao empregar esses procedimentos de seleção de variáveis, busca-se não apenas otimizar o desempenho do modelo, mas também obter uma interpretação mais aprofundada do problema em questão. A primeira abordagem visa mitigar potenciais vieses decorrentes da multicolinearidade, ao mesmo tempo em que aproveita a capacidade do Random Forest em destacar as variáveis mais relevantes para a tarefa em questão. Contudo, ao ter como parâmetro a correlação linear entre variáveis, se houver uma relação não linear significativa, a correlação pode não refletir adequadamente a associação entre as variáveis. Além disso, a correlação entre variáveis não implica necessariamente que uma delas seja redundante para a modelagem. Duas variáveis podem ter uma correlação alta, mas ainda assim contribuir independentemente para a explicação do modelo.

Por isso, a técnica baseada na importância das variáveis também é implantada, de modo a garantir uma seleção de variáveis mais robusta e adequada às características específicas do conjunto de dados e do problema em questão.

3.2.4 Padronização

O conjunto de dados em sua forma bruta frequentemente apresenta variáveis expressas em escalas distintas, uma disparidade que pode impactar os resultados de algoritmos analíticos. A padronização, uma técnica de transformação, é empregada com o propósito de equalizar as variáveis para uma mesma escala (Evsukoff, 2020). Essa prática é essencial para garantir uma análise mais consistente e precisa, evitando que disparidades nas escalas das variáveis influenciem indevidamente os resultados obtidos por algoritmos analíticos.

Todas as variáveis numéricas no conjunto de dados, então, são padronizadas pela estimativa do *z-score*, dada por:

$$\hat{x}_l = \frac{x_i - \bar{x}_l}{\hat{\sigma}_i} \quad \text{Equação 3}$$

onde \bar{x}_l é o valor médio de todas as observações pertencentes a essa variável e $\hat{\sigma}_i$ é o desvio padrão. Todas as variáveis são padronizadas dessa maneira para assegurar

que todas elas possam contribuir igualmente para a função objetivo utilizada no treinamento do modelo. Essa prática visa garantir que as diferenças nas escalas das variáveis não influenciem indevidamente o desempenho do algoritmo.

Com o intuito de explorar outras abordagens de normalização, propõe-se também a implementação do *Robust Scaler*, um método que utiliza estatísticas robustas para lidar com valores discrepantes. Esse método consiste em remover a mediana e redimensionar os dados no intervalo compreendido entre o 1º quartil e o 3º quartil, ou seja, entre o 25º e o 75º percentil. Caso valores discrepantes estejam presentes no conjunto de dados, a utilização da mediana e do intervalo interquartil proporciona resultados mais robustos, superando a eficácia da média e da variância amostral. A transformação *Robust Scaler* é dada por:

$$\hat{x}_i = \frac{x_i - Q_1(X)}{Q_3(X) - Q_1(X)} \quad \text{Equação 4}$$

onde $Q_1(X)$ e $Q_3(X)$ correspondem ao primeiro e terceiro quartil, respectivamente, da variável X .

3.3 Modelos de Regressão

Como foi apresentado entre os objetivos deste trabalho, pretende-se obter um modelo que preveja um valor numérico contínuo, a saber, a duração ou a eficiência de uma operação. Para isso, algoritmos supervisionados de aprendizado de máquinas foram as principais ferramentas de modelagem de dados empregadas. O aprendizado de máquinas estuda como computadores podem “aprender” a reconhecer padrões complexos e tomar decisões inteligentes, baseando-se no conjunto de dados que analisaram. De acordo com o tipo de problema que aborda, o aprendizado de máquinas pode ser dividido em: supervisionado, não-supervisionado, semi-supervisionado e ativo (Han et al., 2011).

O aprendizado supervisionado é caracterizado por possibilitar a automatização de processos de classificação ou previsão generalizando a partir de exemplos conhecidos. Um subconjunto dos dados rotulados é utilizado para “treinar” um modelo, ou seja, buscar encontrar parâmetros que melhor ajustem o modelo ao conjunto de dados. Utilizando outro subconjunto dos dados, o conjunto de teste, o modelo é avaliado de acordo com sua capacidade de prever os rótulos desconhecidos deste conjunto.

Quanto aos rótulos, estes classificam o problema resolvido pelo modelo em um problema de classificação, caso o rótulo seja uma variável discreta, regressão, quando se tem uma variável contínua.

No aprendizado não-supervisionado, os algoritmos exploram padrões intrínsecos nos dados sem a orientação de rótulos específicos. O objetivo é identificar estruturas ou agrupamentos naturais nos dados, como clusters ou associações, sem a necessidade de exemplos rotulados. Por outro lado, o aprendizado semi-supervisionado combina elementos do aprendizado supervisionado e não-supervisionado, utilizando um conjunto limitado de dados rotulados e uma grande quantidade de dados não rotulados. Isso permite ao modelo aprender com as informações disponíveis e generalizar para dados não rotulados. Já o aprendizado ativo, ou por reforço, envolve um processo interativo, no qual o modelo solicita, estratégicamente, rótulos de instâncias não rotuladas, otimizando assim seu desempenho com o menor esforço de rotulagem possível. Essas abordagens oferecem flexibilidade e adaptabilidade em diferentes cenários, explorando diferentes formas de extrair conhecimento de conjuntos de dados complexos e variados.

Neste capítulo é realizada uma exposição acerca dos métodos de aprendizado de máquina utilizados em problemas de regressão. Existe uma grande e variada gama de metodologias possíveis em abordagens envolvendo regressões, de modo que diferentes níveis de complexidade de implementação e precisão de resultados podem ser obtidos. (Bishop, 2006). Portanto, os métodos aqui considerados abrangem desde modelos lineares até redes neurais profundas. Além disso, aborda-se também métodos de preparação de conjunto de dados (seleção de variáveis, remoção de *outliers*, etc.) e de avaliação dos modelos.

3.3.1 Modelos Paramétricos e não paramétricas

Existem duas categorias principais de abordagens de modelagem de dados: paramétricas e não paramétricas. Modelos paramétricos buscam otimizar um conjunto específico de parâmetros θ , ajustando-o ao conjunto de treinamento durante a fase de aprendizado. Após essa fase, quaisquer previsões futuras, realizadas pelo modelo, x são independentes do conjunto de dados observado D , de modo que:

$$P(x|\theta, D) = P(x|\theta) \quad \text{Equação 5}$$

Ou seja, θ é assumido como capturando toda a variância contida no conjunto de dados D (Clarke et al., 2009). Modelos como Regressão Linear, Redes Neurais Artificiais e Regressores de Vetor de Suporte (com uma função linear) são exemplos de modelos paramétricos.

Já os modelos não paramétricos operam sob a ideia de que a distribuição do conjunto de dados não pode ser descrita por um número finito de parâmetros. Esses modelos utilizam os dados de treinamento ou uma amostra deles para realizar previsões, proporcionando flexibilidade para adaptar-se a estruturas mais complexas e não lineares nos dados. Ao contrário dos modelos paramétricos, não fazem suposições rígidas sobre a forma da distribuição subjacente, sendo especialmente úteis em situações em que a relação entre variáveis é intrinsecamente complexa. Exemplos incluem Regressores de Árvores de Decisão, Floresta Aleatória, SVRs com uma Função de Base Radial (RBF), já que o número de parâmetros aumenta com o tamanho do conjunto de dados de treinamento (Bishop, 2006).

Dessa forma, a estratégia de modelagem não paramétrica tem a possibilidade de oferecer modelos com desempenho superior devido à diminuição das suposições em relação aos parâmetros. Entretanto, essa abordagem requer uma quantidade mais significativa de dados de treinamento e resulta em um aumento do custo computacional. Determinar a metodologia ideal para a derivação do modelo não é uma tarefa trivial, uma vez que é influenciada por diversos fatores, incluindo a quantidade e qualidade dos dados disponíveis, além da natureza (e complexidade) do problema em questão.

3.3.2 Regressão linear e regressão linear múltipla

A regressão linear (LR) é uma técnica estatística amplamente utilizada para modelar a relação entre uma variável dependente e uma ou mais variáveis independentes. O objetivo principal da regressão linear é entender e quantificar a relação linear entre essas variáveis, permitindo fazer previsões ou inferências sobre a variável dependente com base nas variáveis independentes. É um modelo paramétrico conhecido como o algoritmo de regressão mais simples e são frequentemente utilizados como uma linha de base comparação e avaliação do desempenho de outros modelos (Bishop, 2006).

Hastie et al. (2008) apresenta a equação básica da regressão linear simples da seguinte forma:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Equação 6

onde Y é a variável dependente, X é a variável independente, β_0 é o intercepto, β_1 é o coeficiente de inclinação e ε é o erro. O objetivo é encontrar os valores ótimos de β_0 e β_1 que minimizam a soma dos quadrados dos resíduos, ou seja, a diferença entre os valores observados e os valores preditos pelo modelo. Essa estimativa é geralmente feita usando uma abordagem de Mínimos Quadrados.

A regressão linear pode ser estendida para casos de múltiplas variáveis independentes, resultando na regressão linear múltipla. Nesse cenário, a equação é expandida para

$$Y = \beta_0 + \sum_{i=1}^n \beta_i X_i + \varepsilon$$

Equação 7

3.3.3 Regressão Parcial de Mínimos Quadrados

A Regressão Parcial de Mínimos Quadrados (PLS) (Wold et al., 2001) é uma técnica estatística que reduz os preditores a um conjunto menor de componentes não correlacionados, efetuando regressão de mínimos quadrados para esses componentes no lugar dos dados originais. Essa abordagem é particularmente valiosa em situações em que os preditores são altamente colineares ou quando há mais preditores do que observações, desafiando a regressão de mínimos quadrados ordinários. Sua aplicação visa o desenvolvimento de modelos preditivos robustos, sendo menos adequada para filtrar variáveis desnecessárias à explicação da resposta.

3.3.4 SVM

As Máquinas de Vetores de Suporte (SVMs), em sua forma mais simples, constituem um classificador de duas classes em casos em que as duas classes são linearmente separáveis. As SVMs funcionam derivando o hiperplano ótimo, ou seja, o hiperplano que oferece a margem mais ampla possível entre instâncias das duas classes (Gkerekos et al., 2019). Sua funcionalidade pode ser estendida pela introdução de um kernel não linear, permitindo que aprendam mapeamentos não lineares, classificando elementos de classes não linearmente separáveis. Dependendo das propriedades do kernel selecionado, as SVMs podem ser modelos paramétricos ou não paramétricos.

As SVMs também podem ser construídas como regressores (Evsukoff, 2020). Regressores de Vetor de Suporte (SVRs) funcionam de maneira semelhante, desta vez tentando ajustar um hiperplano que prevê com precisão os valores-alvo das amostras de treinamento dentro de uma margem de tolerância ε .

3.3.5 Árvore de Decisão

Existem dois tipos principais de Árvores de Decisão: aquelas destinadas à classificação e aquelas voltadas para a regressão. Nas Árvores de Decisão de Classificação, o objetivo é atribuir uma classe ou categoria a uma instância de dados, enquanto nas Árvores de Decisão de Regressão, o foco é prever um valor numérico. As Árvores de Decisão para Regressão (DTRs) representam uma abordagem não paramétrica para modelagem de problemas de regressão. Esses modelos dividem o espaço de características em retângulos e aprendem modelos simples, como constantes, em cada uma dessas regiões (Hastie et al., 2008). Esses modelos são especialmente valiosos para compreender o processo de tomada de decisões em um conjunto de dados complexo.

O conceito fundamental por trás das Árvores de Decisão é a divisão recursiva do espaço de características em regiões distintas, cada uma associada a uma decisão ou previsão. Cada nó na árvore representa uma condição sobre uma característica específica, e os ramos indicam as possíveis saídas ou divisões com base nessa condição. Ao percorrer a árvore, seguindo as condições nos nós, chegamos a uma folha que fornece a previsão ou classificação final (Hastie et al., 2008).

O processo de construção de uma Árvore de Decisão envolve a seleção da melhor característica para dividir o conjunto de dados em cada nível da árvore. Uma medida de impureza, como o índice Gini para classificação ou o erro quadrático médio para regressão, é utilizada para avaliar a qualidade das divisões. A árvore cresce de forma recursiva, sendo “podada” posteriormente para evitar overfitting.

Uma característica notável das Árvores de Decisão é a capacidade de lidar com dados categóricos e numéricos sem a necessidade de pré-processamento complexo. Além disso, esses modelos oferecem uma interpretabilidade natural, permitindo que usuários e especialistas entendam facilmente as decisões tomadas pelo algoritmo. Contudo, elas também podem ser suscetíveis a overfitting, especialmente quando não são adequadamente podadas. Estratégias como Random Forests e Gradient Boosting

Trees foram desenvolvidas para melhorar essas limitações, utilizando múltiplas árvores para realizar previsões mais robustas.

3.3.6 Florestas Aleatórias

Florestas Aleatórias para Regressão (RFR, inglês Random Forest Regressor) é um conjunto de árvores de decisão, treinadas através do método *Bagging*. *Bagging*, ou agregação *bootstrap*, é uma estratégia para redução da variância de uma função de previsão estimada que combina diferentes modelos de aprendizagem para melhorar a acurácia da previsão do modelo. Essa abordagem demonstra eficácia especialmente em procedimentos com alta variância e baixo viés, como é o caso das árvores de decisão (Hastie et al., 2008). No contexto de regressão, aplica-se repetidamente a mesma árvore de regressão a versões amostradas por *bootstrap* dos dados de treinamento, para posteriormente calcular a média dos resultados.

Florestas aleatórias representam uma modificação substancial do *bagging*, construindo uma extensa coleção de árvores não correlacionadas que são posteriormente combinadas por meio de uma média.

Uma variação da floresta aleatória é o método Árvores Extras para Regressão (ETR, do inglês *Extra Trees Regressor*), onde o conjunto de dados completo é utilizado em cada instância, e as divisões da árvore são escolhidas completamente ao acaso. As árvores extras têm se mostrado eficazes em muitos problemas de aprendizado de máquina, especialmente em conjuntos de dados com alta dimensionalidade e quando há a presença de características irrelevantes ou redundantes. Elas ajudam a mitigar o *overfitting* e melhoram a generalização do modelo.

3.3.7 Redes neurais

As Redes Neurais (NNs) são sistemas computacionais inspirados na forma como os sistemas nervosos biológicos funcionam. Existem várias arquiteturas de ANN que oferecem desempenho superior em muitas tarefas de aprendizado de máquina, incluindo classificação e regressão. As ANNs são extremamente versáteis, pois podem modelar com precisão comportamentos não lineares complexos (Gkerekos et al., 2019).

As ANNs são baseadas em um grupo interconectado de unidades conectadas (neurônios), onde cada conexão entre essas unidades transmite um sinal de uma para outra quando a combinação linear de suas entradas excede algum limiar (Russell & Norvig, 2010). A unidade receptora pode processar esse sinal e então transmiti-lo para a

próxima unidade. Devido à arquitetura descrita, as ANNs são consideradas modelos paramétricos.

Dois parâmetros importantes das ANNs são o número de camadas ocultas (entre entrada e saída) e o número de neurônios por camada. Excluindo as camadas de entrada e saída que sempre existem, diferentes arquiteturas pedem diferentes números de camadas ocultas e neurônios. Consequentemente, diferentes funções de ativação podem ser implementadas, alterando a complexidade que o modelo pode aprender.

À medida que o número de camadas aumenta, o modelo pode aprender comportamentos não lineares mais complexos. Ao mesmo tempo, o treinamento se torna computacionalmente mais caro e o risco de sobreajuste ao conjunto de dados de treinamento também aumenta.

As funções de ativação são usadas para introduzir uma não linearidade em cada neurônio. A derivada dessas funções não deve ser muito complexa computacionalmente, já que são utilizadas no processo de retropropagação, para otimização da rede. Algumas das ativações amplamente utilizadas são ReLU, Sigmoid e tangente hiperbólica.

As funções de custo, ou funções de perda, são utilizadas para medir a discrepância entre as previsões da rede neural e os valores reais nos dados de treinamento. A escolha da função de custo depende da natureza da tarefa (classificação, regressão etc.) e do tipo de saída da rede. O Erro Quadrático Médio, do inglês Mean Squared Error (*MSE*) é um exemplo de função de custo amplamente usada em problemas de regressão, que penaliza fortemente erros maiores.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{Equação 8}$$

3.4 Otimização e Análise de Modelos

Entende-se por “parâmetros” valores determinantes que os algoritmos de aprendizado de máquina ajustam de acordo com o conjunto de dados fornecido. Os hiperparâmetros são os parâmetros definidos manualmente antes do início do treinamento e que se baseiam em propriedades como as características dos dados e a capacidade de aprendizado do algoritmo.

A otimização dos hiperparâmetros representa a fase de calibração dos modelos, envolvendo a exploração de diversas combinações de hiperparâmetros específicos de cada modelo de aprendizado. Esse processo visa aprimorar o desempenho dos modelos, levando em conta o estimador em questão, o espaço de hiperparâmetros disponíveis, a estratégia de busca ou candidatos de amostras, e a métrica de erro utilizada para avaliar a qualidade dos resultados. A busca pelos melhores parâmetros é essencial para maximizar a eficácia do modelo e adaptá-lo de forma mais precisa às características dos dados em análise.

Neste estudo, a abordagem de otimização dos hiperparâmetros consiste em se estabelecer um conjunto de valores para cada hiperparâmetro pré-definido de cada modelo avaliado. O modelo é, então, treinado e avaliado para todas as combinações possíveis desses valores. É uma técnica simples, mas pode se tornar computacionalmente cara, especialmente com um grande número de hiperparâmetros e valores.

Normalmente, quando se avalia um modelo a partir do conjunto de treinamento, pode-se não obter uma boa escolha de melhor modelo, pois um modelo com sobreajustado tem bom desempenho no conjunto de treinamento. O método mais simples para avaliar o desempenho de um modelo em um conjunto de dados diferente do utilizado no ajuste envolve a divisão do conjunto C em dois subconjuntos: C_{treino} para treinamento e C_{teste} para teste, geralmente com tamanhos proporcionais ($N_{treino} = 2N / 3$ e $N_{teste} = N / 3$). Por essa abordagem pressupor a representatividade das amostras, sua eficácia pode ser comprometida quando se tem conjuntos pequenos de dados. Além disso, estatísticas de validação ficam muito dependentes da partição (Evsukoff, 2020).

A validação cruzada apresenta-se, então, como um procedimento essencial para avaliar o desempenho de um modelo em dados distintos do conjunto de treinamento. Na validação cruzada k -fold, o conjunto de dados é dividido em K subconjuntos, e a

validação é conduzida em K iterações. O modelo é ajustado em cada ciclo utilizando $K - 1$ subconjuntos e avaliado no subconjunto restante, gerando estatísticas de validação a cada ciclo. A Figura 5 ilustra o procedimento.

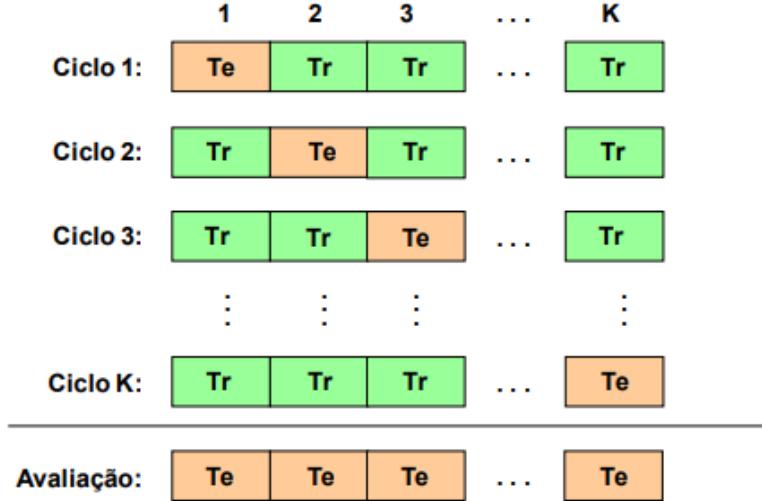


Figura 5: Validação cruzada de K ciclos: Tr são os subconjuntos de treinamento e Te são os subconjuntos de teste. Fonte: Evsukoff (2020)

A escolha do número de ciclos K é crucial, sendo comum optar por $K = 5$ ou $K = 10$ para equilibrar eficácia e custo computacional. Apesar do alto custo computacional, o k -fold fornece a melhor estimativa de desempenho do modelo em dados não utilizados no treinamento. Contudo, é sensível à amostra e pode variar entre diferentes conjuntos de dados.

Para assegurar de maneira razoável que os valores dos hiperparâmetros selecionados estejam realmente próximos do ótimo e não resultem apenas em um ajuste excessivo do modelo, a validação cruzada k -fold é implementada. Assim, para cada combinação de hiperparâmetros, vários resultados são obtidos e, em seguida, são calculadas médias desses resultados. Ao aplicar essa técnica a todos os modelos, é possível identificar aquele que apresenta o melhor desempenho, garantindo simultaneamente boas capacidades de generalização.

A Figura 6 ilustra o fluxo de otimização e avaliação de modelos adotado neste estudo. Os dados são, inicialmente, divididos aleatoriamente em um Conjunto de Treino e Teste (80% do conjunto de dados), e em um Conjunto de Validação (20%). O Conjunto de Treino e Teste é utilizado na calibração dos modelos via validação cruzada

k -fold

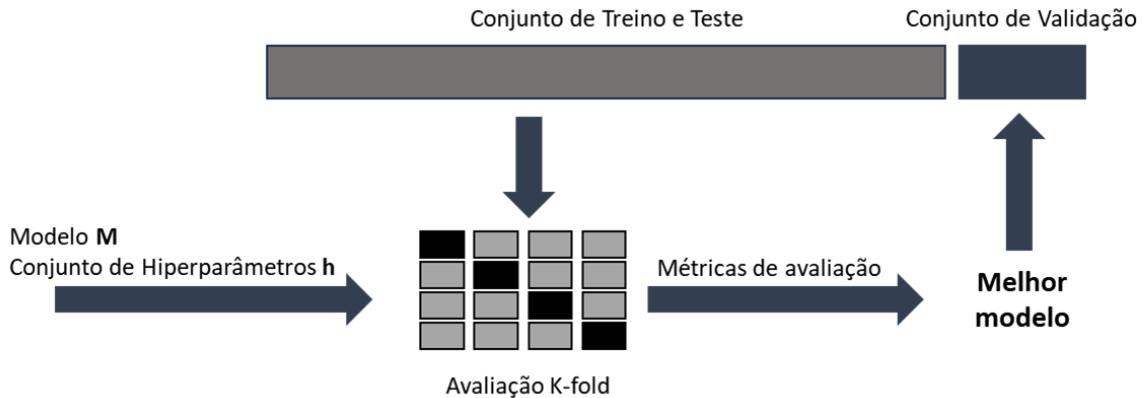
, processo que é realizado para cada combinação de hiperparâmetros h para cada modelo M . A melhor combinação de hiperparâmetros para cada modelo é estabelecida a partir da média das métricas de avaliação geradas em cada ciclo da validação cruzada. O modelo otimizado é então treinado em todo o conjunto de treino e teste e avaliado junto ao conjunto de validação.

Figura 6: Fluxo de otimização e avaliação dos modelos

3.4.1 Métricas de Avaliação

Nesta sessão são apresentadas as métricas estatísticas utilizadas para avaliação dos modelos. As estatísticas de validação medem o desempenho do modelo num determinado conjunto de dados. Várias métricas podem ser utilizadas para avaliação de modelos de regressão. Neste trabalho, foram utilizadas as métricas R^2 , o Erro Percentual Absoluto Médio (*MAPE*, do inglês *Mean Absolute Percentual Error*), e a Raiz do Erro Quadrático Médio (*RMSE*, do inglês *Root Mean Squared Error*).

O coeficiente de determinação, representado por R^2 , é uma métrica estatística amplamente empregada na avaliação de modelos de regressão. Variando de 0 a 1, o R^2 quantifica a proporção da variabilidade total na variável de saída que é explicada pelo modelo. Um R^2 próximo de 1 sugere que o modelo é capaz de explicar de forma significativa a variabilidade nos dados, indicando um desempenho adequado. Intuitivamente, o R^2 compara a eficácia do modelo com um modelo simples que se baseia na média da variável de saída para fazer previsões. Um R^2 próximo de 0 indica que o modelo não oferece melhorias substanciais em relação à previsão da média. Vale notar que, em alguns casos, o R^2 pode ser inferior a 0, indicando que o modelo é menos

eficaz do que simplesmente prever a média. Além disso, o coeficiente de determinação é sensível a valores extremos na variável de saída, uma vez que os erros são elevados ao quadrado, destacando a importância de considerar sua aplicabilidade em cenários com dados atípicos.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad \text{Equação 9}$$

O *MAPE* é uma variação do *MAE* (Erro Médio Absoluto, do inglês *Mean Absolute Error*). O *MAE* possui a mesma escala que a variável alvo e pode ser analisada no contexto do problema. O *MAPE*, contudo, oferece uma visão por vezes mais intuitiva, ao apresentar os erros de forma relativa.

$$MAPE(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad \text{Equação 10}$$

Por fim, o *RMSE* é uma métrica de avaliação semelhante ao Erro Quadrático Médio (*MSE*, Equação 8), mas inclui a raiz quadrática para lidar com diferenças de escala entre unidades. Essa modificação preserva a interpretabilidade da avaliação, mantendo a unidade do *RMSE* na mesma escala que os dados originais. Embora o valor do *RMSE* não necessariamente se assemelhe ao do *MAE*, sua interpretação segue a lógica de indicar a magnitude média dos erros do modelo. O *RMSE* é valioso quando se busca uma avaliação criteriosa das previsões do modelo, considerando sua sensibilidade a desvios significativos e outliers.

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad \text{Equação 11}$$

4 EXPERIMENTO

4.1 Definição do problema

A indústria de exploração e produção de óleo e gás offshore representa uma parte vital da economia global, sendo fundamental para o abastecimento de energia. Como já foi argumentado, para operar de forma eficaz nesta indústria, é crucial o transporte de suprimentos, a partir de bases de apoio offshore, para abastecimento de unidades marítimas, tais como plataformas de produção, sondas de perfuração e embarcações especiais.

Como destacado no anteriormente, esses suprimentos são comumente categorizados em três grupos principais: cargas gerais, granéis líquidos e granéis sólidos. Para o transporte desses itens, são utilizadas embarcações especializadas, como, dentre outras, as *Platform Supply Vessels* (PSVs), já apresentadas anteriormente. A duração da movimentação desses suprimentos entre um PSV e uma UM é um elemento crítico para o sucesso das operações offshore, afinal, a partir desse dado, todo o planejamento de viagens e distribuição de cargas pode ser otimizado.

A produtividade, ou eficiência, de uma operação de transferência de carga é avaliada de maneira diversificada, dependendo do tipo de carga e da operação, sempre atrelada à duração da operação. No transporte de cargas gerais no convés de PSVs, a produtividade é geralmente medida pelo tempo necessário para movimentar um item entre o PSV e a UM. Em contrapartida, no transporte de granéis em tanques e silos sob o convés, a produtividade é avaliada em termos de vazão, com a transferência realizada por tubulações e mangotes. Além disso, os tempos de preparação antes e após a movimentação também afetam a eficiência operacional, apesar de serem raramente registrados.

No entanto, tanto a duração quanto a viabilidade de uma operação marítima estão condicionadas a uma série de fatores, tanto intrínsecos à infraestrutura operacional quanto externos. No âmbito operacional, esses fatores incluem desde a experiência da tripulação, passando pela posição dos guindastes, o aproamento das plataformas, as condições dos equipamentos da embarcação, o sistema de posicionamento dinâmico e as dimensões da embarcação, até o bordo em que a embarcação operou ou tentou operar. Quanto aos fatores externos, as condições meteoceanográficas no local da plataforma no

momento da operação desempenham um papel crucial na segurança, eficiência e viabilidade da operação, impactando parâmetros operacionais, como a potência dos motores. Entre esses fatores ambientais, destacam-se a intensidade e direção da corrente marítima, velocidade e direção do vento, altura e direção das ondas, bem como o período de pico, que representa o intervalo de tempo entre duas cristas sucessivas (ou vales) de uma onda (períodos mais curtos podem resultar em mares agitados, enquanto períodos mais longos indicam ondas mais espaçadas).

O maior desafio encontrado ao realizar uma análise de maior abrangência, contudo, é a ausência de registros históricos desses fatores, seja pela não coleta de informações cruciais, como tempos de preparação ou posição de guindastes, o não detalhamento das operações, ou não medição.

A literatura já estabeleceu correlações entre as condições climáticas ambientais e parâmetros críticos das operações marítimas, como eficiência no consumo de combustível e velocidade de deslocamento de embarcações. Segundo Norlund & Gribkovskaia (2017), condições climáticas adversas, notadamente ondas e vento, prolongam os tempos de navegação e as operações de serviço em embarcações de abastecimento, sendo as ondas apontadas como o fator mais significativo por Halvorsen-Weare & Fagerholt (2011a). Além disso, em operações de movimentação mecânica de cargas, a NR 37 (CTPP, 2020) estabelece limites em relação a condições de vento e luminosidade. Contudo, a influência do vento nas operações de carga e descarga em unidades marítimas é mencionada por Norlund & Gribkovskaia (2017) sem uma análise aprofundada desse impacto.

Apesar de estudos prévios terem examinado a capacidade de prever a viabilidade operacional com base em previsões meteoceanográficas, persiste uma lacuna significativa na compreensão de como essas condições afetam a produtividade da movimentação de suprimentos além do impacto previsto pelas normas reguladoras. Assim, permanece incerto se, mesmo quando as operações são consideradas viáveis, as condições meteoceanográficas podem influenciar os tempos de ligação, as vazões e os tempos de preparação, impactando a eficiência global do processo.

Portanto, de modo a fornecer uma ferramenta de apoio à otimização do planejamento logístico de E&P, o principal foco deste trabalho é prever, a partir de dados operacionais e meteoceanográficos históricos, a duração ou produtividade de operações de movimentação de carga entre embarcações. A partir disso, pretende-se é

investigar como as condições meteoceanográficas impactam a produtividade das operações de movimentação de suprimentos entre PSVs e UMs na indústria offshore de exploração e produção de óleo e gás. Busca-se preencher essa lacuna de conhecimento ao examinar de que forma fatores meteorológicos e oceanográficos específicos influenciam os tempos de ligada e as vazões. Dessa maneira, espera-se contribuir com insights valiosos para aprimorar a eficiência e segurança das operações offshore, promovendo o progresso desta crucial indústria.

Neste capítulo, o procedimento metodológico mostrado no Capítulo 3 é experimentado e avaliado na resolução de um problema real de determinação da duração e produtividade da tarefa de transferência de cargas entre embarcações. Para isso, são considerados registros operacionais e climáticos de operações realizadas em unidades marítimas dispostas ao longo da Bacia de Santos e Bacia de Campos, em um período de 2 anos.

4.2 Definição do Conjunto de Dados

A base de dados inicial consiste em um conjunto de registros de rotas de viagem. Cada registro corresponde a um tipo de entrega realizada por uma embarcação específica a uma determinada unidade marítima em uma viagem. Um registro é composto pelas informações:

- início da operação (data e horário do início da transferência de cargas),
- fim da operação (data e horário da conclusão da operação),
- um identificador da viagem,
- tipo de embarcação realizando a operação,
- identificador da embarcação,
- tipo da unidade marítima,
- identificador da unidade marítima;
- tipo de operação (movimentação de carga geral em convés ou granéis líquidos e sólidos);
- o peso total transferido (que inclui carga e descarga das unidades marítimas, e também contabiliza o peso das embalagens de unitização);
- o volume total transferido e

- a área total ocupada no convés (assumindo o não empilhamento das cargas).

Neste trabalho, a eficiência total da operação é calculada como a razão entre o peso total transferido e a duração total da operação. Isso se deve ao fato de que, para cada operação, tem-se apenas a duração total da operação que, no caso da movimentação de carga geral, compreende todas as lingadas realizadas, além de outras operações de preparação que são realizadas. Não é possível, portanto, especificar a duração de cada lingada, o que seria interessante, dado que o conjunto de materiais unitizados para içamento em uma lingada possui seu próprio conjunto de características, como quantidade, peso, volume etc. que, sob efeito de condições climáticas adversas, podem interferir na duração da tarefa de modo mais perceptível do que quando analisada a operação como um todo.

Foram analisadas informações sobre as condições marítimas durante a operação, levando em consideração as premissas necessárias devido à disponibilidade limitada de dados. É importante destacar que essas premissas podem influenciar o desempenho do modelo utilizado nesta pesquisa.

As características meteoceanográficas foram especificadas com base na localização da Unidade Marítima onde a operação foi realizada, bem como no horário correspondente. Ressalta-se que tais dados foram obtidos por meio de simulações, cujos procedimentos detalhados não estão incluídos no escopo deste estudo. A Tabela 2 Tabela 3 fornece uma visão detalhada de cada variável considerada.

Tabela 3: Variáveis de condições meteoceanográficas

Campo	Nome	Conceito
Vento	<i>wspd</i>	Velocidade do vento (m/s)
	<i>wdir</i>	Direção do vento (graus)
Corrente	<i>cspd</i>	Velocidade da corrente (m/s)
	<i>cdir</i>	Direção da corrente (graus)
Onda (mar local)	<i>wavhs</i>	Altura significativa (m)
	<i>wavtp</i>	Período (s)
	<i>wavdir</i>	Direção da onda (graus)
Swell (marulho)	<i>swhs</i>	Altura significativa (m)
	<i>swtp</i>	Período (s)
	<i>swdir</i>	Direção da onda (graus)

As informações meteoceanográficas, juntamente com os parâmetros operacionais delineados na Tabela 3, representam a base de parâmetros utilizados para calcular outras variáveis relevantes relacionadas à resposta do PSV às condições climáticas momentâneas.

Tabela 4: Parâmetros operacionais relacionados ao PSV

Campo	Nome	Conceito
PSV	<i>PSVgroup</i>	Classificação PSV
	<i>relHead</i>	Aproamento relativo do PSV para com a plataforma (graus)
	<i>point</i>	Ponto de operação
	<i>side</i>	Bordo de operação

A variável *PSVgroup* categoriza os PSVs com base na relação entre o comprimento total (LOA) e a largura máxima (Boca) da embarcação, dividindo-os em quatro distintos. As categorias foram estabelecidas para proporcionar uma classificação mais precisa. Em relação às variáveis *relHead* e *point*, que dizem respeito ao aproamento relativo do PSV em relação à Unidade Marítima e à distância entre o PSV e a UM, respectivamente, vale salientar que essas informações não foram diretamente registradas. Para preencher as lacunas e gerar os dados referentes às variáveis presentes na Tabela 5, tomou-se a decisão de estabelecer valores específicos com base no conhecimento especializado dos operadores envolvidos na operação. Assim, foi atribuído um *relHead* de 185 graus e uma distância de 10 metros entre o PSV e a UM. Além disso, em todas as operações, o bordo de operação foi definido como bombordo.

Tabela 5: Variáveis de impacto das condições de mar sobre as embarcações

Variável	Descrição
<i>engine1</i>	Porcentagem de uso no propulsor 1
<i>engine2</i>	Porcentagem de uso no propulsor 2
<i>engine3</i>	Porcentagem de uso no propulsor 3
<i>engine4</i>	Porcentagem de uso no propulsor 4
<i>engine5</i>	Porcentagem de uso no propulsor 5
<i>engine6</i>	Porcentagem de uso no propulsor 6
<i>SigRoll</i>	Ângulo significativo de roll ($^{\circ}$)
<i>MaxRoll</i>	Máximo Roll ($^{\circ}$)
<i>SigPitch</i>	Ângulo significativo de pitch ($^{\circ}$)
<i>MaxPitch</i>	Máximo Pitch($^{\circ}$)
<i>SigHeave</i>	Altura significativa de heave (m)
<i>MaxHeave</i>	Máximo Heave
<i>ForceX</i>	Força de propulsão resultante na direção longitudinal do navio (positivo em direção à proa)
<i>ForceY</i>	Força de propulsão resultante na direção transversal do navio (positivo a bombordo)
<i>MomentZ</i>	Momento de propulsão resultante no plano horizontal (positivo sentido anti-horário de rotação)

Roll (rolagem) refere-se ao movimento de inclinação lateral de uma embarcação em torno do eixo longitudinal. Quando uma embarcação rola, ela se inclina para um lado ou para o outro, como um balanço lateral. *Pitch* (arfagem) refere-se ao movimento de inclinação para frente ou para trás de uma embarcação em torno do eixo transversal. *Heave* (arremesso) refere-se ao movimento vertical para cima e para baixo de uma embarcação ou aeronave e ocorre ao longo do eixo vertical. Esses termos são fundamentais para descrever e compreender os diferentes movimentos que uma embarcação pode experimentar durante sua operação. A Figura 7 demonstra esses movimentos.

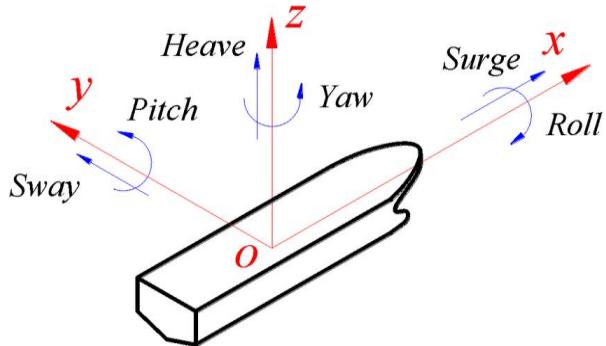


Figura 7: Os movimentos de *roll*, *pitch*, *yaw*, *heave*, *sway*, *surge* de uma embarcação.

Fonte: Winter (2018)

Assim, as informações são organizadas com base na localização da Unidade Marítima e nos horários das operações. Por exemplo, se uma operação teve início às 17 horas de um dia qualquer e durou 3 horas, serão gerados 3 registros correspondentes às horas específicas: 17h, 18h e 19h. No entanto, neste estudo, é essencial consolidar as informações de cada horário para associá-las aos registros operacionais. Dessa forma, para cada cenário em cada horário durante uma operação, foram aplicadas ponderações específicas, conforme detalhado na Tabela 6.

Tabela 6: Critérios aplicados para definição dos valores das variáveis por registro de operação

Nome	Descrição	Valor considerado
<i>wspd</i>	Velocidade do vento (m/s)	Máximo
<i>wdir</i>	Direção do vento (graus)	Médio
<i>cspd</i>	Velocidade da corrente (m/s)	Máximo
<i>cdir</i>	Direção da corrente (graus)	Médio
<i>wavhs</i>	Altura significativa (m)	Máximo
<i>wavtp</i>	Período (s)	Mínimo
<i>wavdir</i>	Direção da onda (graus)	Máximo
<i>swhs</i>	Altura significativa (m)	Máximo
<i>swtp</i>	Período (s)	Mínimo
<i>swdir</i>	Direção da onda (graus)	Médio
<i>engine1</i>	Porcentagem de uso no propulsor 1	Máximo
<i>engine2</i>	Porcentagem de uso no propulsor 2	Máximo
<i>engine3</i>	Porcentagem de uso no propulsor 3	Máximo
<i>engine4</i>	Porcentagem de uso no propulsor 4	Máximo
<i>engine5</i>	Porcentagem de uso no propulsor 5	Máximo
<i>engine6</i>	Porcentagem de uso no propulsor 6	Máximo
<i>SigRoll</i>	Ângulo significativo de roll (°)	Máximo
<i>MaxRoll</i>	Máximo Roll (°)	Máximo
<i>SigPitch</i>	Ângulo significativo de pitch (°)	Máximo
<i>MaxPitch</i>	Máximo Pitch(°)	Máximo
<i>SigHeave</i>	Altura significativa de heave (m)	Máximo
<i>MaxHeave</i>	Máximo Heave	Máximo
<i>ForceX</i>	Força de propulsão resultante na direção longitudinal do navio (positivo em direção à proa)	Máximo
<i>ForceY</i>	Força de propulsão resultante na direção transversal do navio (positivo a bombordo)	Máximo
<i>MomentZ</i>	Momento de propulsão resultante no plano horizontal (positivo sentido anti-horário de rotação)	Máximo

Os PSVs foram também categorizados com base no arranjo de seus propulsores, sendo identificados diferentes padrões de configuração. Cada propulsor recebe uma alocação de potência, sendo representado pelas variáveis *engine1* a *engine6*. No entanto, é importante ressaltar que o propulsor designado, por exemplo, pela variável *engine1* em um PSV com um determinado arranjo não é equivalente ao propulsor *engine1* de outro PSV com uma classificação diferente.

Diante dessa distinção, optou-se por simplificar a representação, substituindo as variáveis *engine1* a *engine6* por uma única variável, denominada *maxEngine*. Esta variável reflete o valor máximo de porcentagem de uso entre todos os propulsores, considerando as diferenças nos arranjos de propulsores dos diversos PSVs.

Portanto, inicialmente, foram consideradas as seguintes variáveis neste estudo:

- **Variáveis contínuas:** *nLingada* (número de lingadas realizadas em uma operação), *Peso* (peso total transferido na operação), *wspd*, *wdir*, *cspd*, *cdir*, *wavhs*, *wavtp*, *wavdir*, *swhs*, *swtp*, *swdir*, *max_engine*, *sigRoll*, *sigPitch*, *sigHeave*, *maxRoll*, *maxPitch*, *maxHeave*, *forceX*, *forceY*, *momentZ*, *volumeTotal*, *areaTotal*, *PesoLingada* (peso médio de carga por lingada), *Duração* e *Eficiencia*.
- **Variáveis categóricas:** *unitType*, *vesselGroup*.

Devido à disponibilidade de dados, neste trabalho foram consideradas apenas operações de movimentação de carga geral, desconsiderando as movimentações de fluidos via mangotes. Deste modo, são considerados 1730 registros distribuídos entre as variáveis categóricas conforme a Tabela 7.

Tabela 7: Contagem de registros por variável categórica

Variável	Valor	Número de registros
vesselGroup	1	1547
	2	126
	3	57
unitType	1	1037
	2	429
	3	264

4.3 Processamento dos Dados

Nesta etapa, foram realizados os processos de desenvolvimento de variáveis, remoção de outliers e seleção de variáveis, descritos na Seção 3.2 deste estudo.

4.3.1 Desenvolvimento de Variáveis

Foram estabelecidos intervalos para as variáveis *wspd* e *wavhs*, referentes à velocidade do vento e altura de onda, respectivamente, conforme os cenários

estabelecidos por CTPP (2020) e Halvorsen-Weare & Fagerholt (2011b). A partir desses intervalos, foram criadas variáveis categóricas. Além disso, operações com horário de início superiores às 18h e inferiores às 6h foram considerados eventos noturnos. A Tabela 8 descreve as variáveis criadas.

Tabela 8: Criação de variáveis categóricas

Variável	Intervalo	Valor Categórico	Variável Categórica
<i>wavhs</i>	< 2.5 m	1	<i>classeOnda</i>
	[2.5 m, 3.5 m)	2	
	[3.5 m, 4.5 m)	3	
<i>wspd</i>	≥ 4.5 m	4	<i>classeVento</i>
	< 38 km/h	1	
	[39 a 49 km/h)	2	
	[50 a 61 km/h)	3	
Horário Inicial	≥ 61 km/h	4	<i>classeNoturno</i>
	[18h, 23h] + [0h, 6h]	1	

Entretanto, ao examinar o conjunto de dados por meio dessas segmentações, observou-se que a grande maioria dos registros está concentrada nos intervalos considerados mais brandos, tanto em termos de altura de onda quanto de velocidade de vento. Diante disso, em virtude da natureza limitada dos dados disponíveis para esta pesquisa, caracterizada pela ausência de registros de eventos ocorridos em condições mais adversas, a análise da influência das condições de mar se restringe apenas a cenários mais amenos. A distribuição das variáveis *wspd* e *wavhs* pode ser observada na Figura 8.

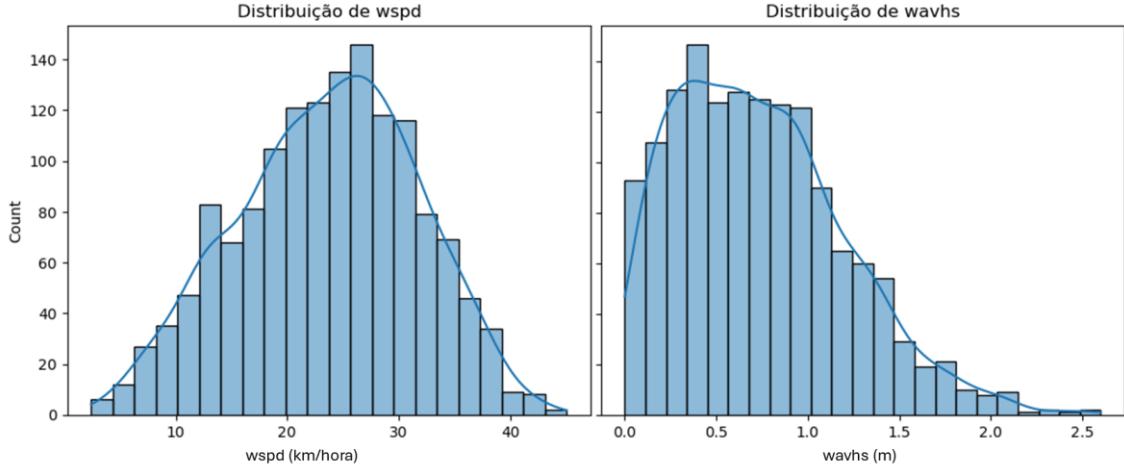


Figura 8: Histograma da distribuição das variáveis *wspd* e *wavhs*

Por fim, foram implementadas algumas variáveis adicionais para estimar a força exercida pelo vento sobre as cargas içadas, seguindo as diretrizes da LIEBHERR (2017). No entanto, algumas premissas foram definidas devido às limitações do conjunto de dados.

A força do vento, é uma função de vários fatores como o formato da carga içada, a direção do vento em relação à carga, a velocidade resultante do vento e a área da carga exposta. Devido à falta de informações detalhadas por lingada, foi adotada uma abordagem que assume uma área média da carga içada por lingada, determinada pela razão entre a medida de área disponível e o número de lingadas por operação. Além disso, a velocidade do vento máxima prevista para todo o período foi utilizada, sem calcular a resultante, devido à falta de dados que permitiriam esse aprofundamento. A pressão dinâmica e a força do vento foram então calculadas e adicionadas ao conjunto de dados.

É válido ressaltar que, devido às premissas e simplificações adotadas, as estimativas de força e pressão aqui utilizadas não têm credibilidade. Por esse motivo, foram estabelecidos cenários, apresentados adiante, nos quais essas variáveis não são consideradas.

4.3.2 Remoção de Outliers

Durante esta fase, foram empregados os métodos Tukey, Isolation Forest e a Média das Distâncias para detecção de registros anômalos. A Figura 9 apresenta a distribuição de algumas variáveis antes e depois da remoção os outliers. Tomando como

referência a variável *Duracao* antes da identificação de anomalias, fica evidente que parte delas são provenientes da coleta imprecisa das datas de realização das operações, resultando em durações extremamente discrepantes. As variáveis *areaTotal*, *volumeTotal* e *pesoTotal* também apresentam valores atípicos. Essas variáveis são preenchidas por operadores, logo, são registros susceptíveis a falhas de digitação, por exemplo.

Sendo assim, método Tukey foi aplicado individualmente a cada variável, focalizando a remoção de outliers nas dimensões de *areaTotal*, *volumeTotal*, *pesoTotal*, *Duracao* e *Eficiencia*. É possível observar na Figura 9 que o método consegue identificar os registros mais discrepantes, de modo que 260 pontos foram removidos da base de dados. Nesta configuração, a duração máxima das operações estudadas neste trabalho ficou em torno de 10 horas.

Durante a aplicação do método das Médias das Distâncias, foi assumido que os maiores valores de distância que constituíssem 15% dos registros seriam considerados outliers. Esse valor foi estabelecido conforme a taxa de outliers detectada pelo método Tukey. Neste método, foi identificada uma maior tolerância em relação às variáveis *volumeTotal* e *areaTotal*, como pode ser observado na Figura 9. Além disso a duração máxima das operações ficou em torno de 20 horas.

O método de *Isolation Forest* incorpora o hiperparâmetro de contaminação. Esse parâmetro representa a quantidade de contaminação no conjunto de dados, ou seja, a proporção de outliers presentes no conjunto de dados. Ao ajustar o modelo, esse parâmetro é utilizado para definir o limiar nos escores das amostras. Embora existam outros hiperparâmetros que podem ser ajustados, neste estudo, optou-se por especificar apenas a taxa de contaminação. Um valor de 0,15 foi adotado para este parâmetro, indicando que aproximadamente 15% dos dados são considerados como outliers durante a aplicação do método de Isolation Forest. Esta proporção, como no método das Médias das Distâncias, foi escolhida levando-se em consideração a proporção de outliers identificada pelo método Tukey.

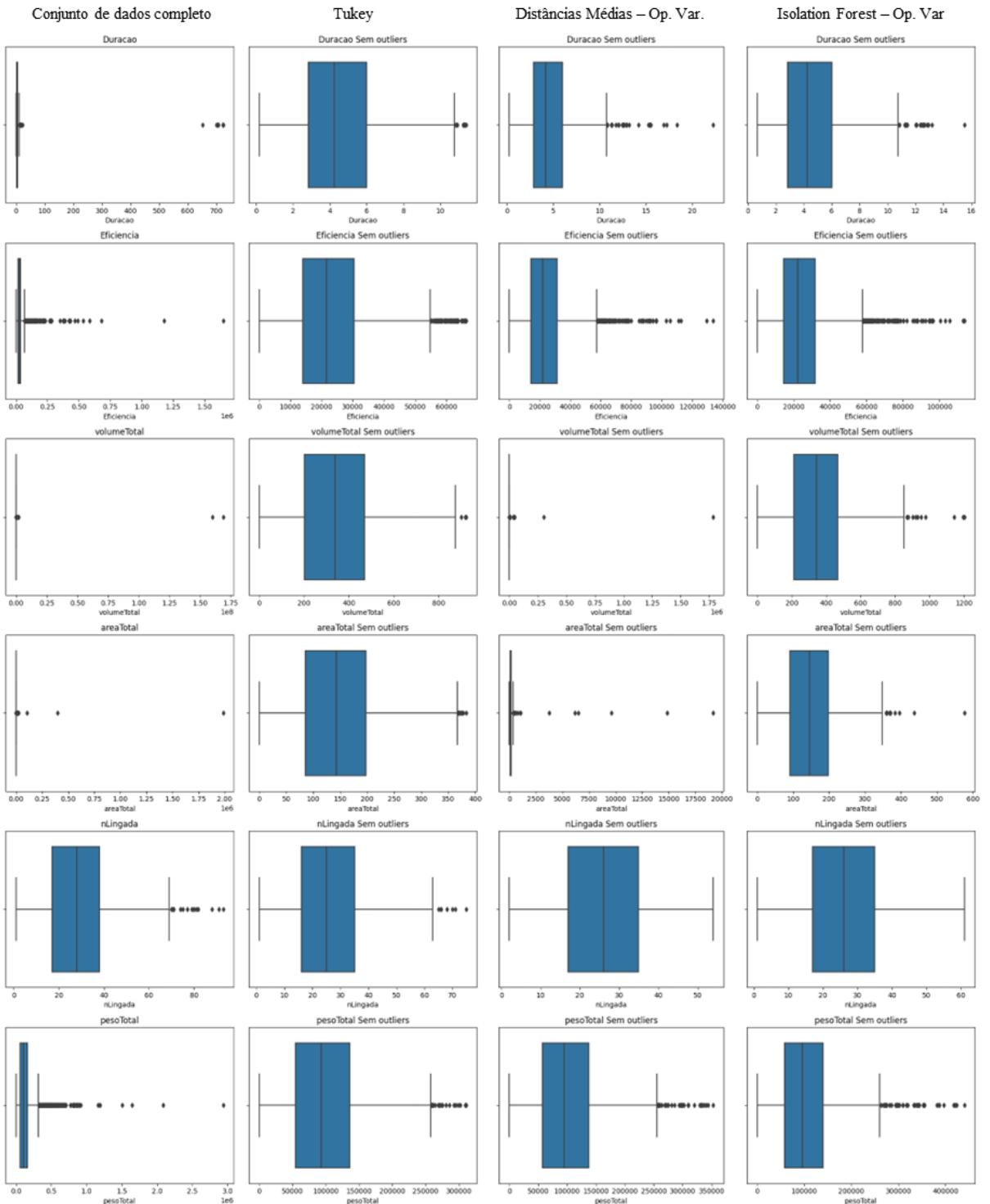


Figura 9: Resultados do tratamento de outliers pelos métodos propostos

O Isolation Forest revelou-se menos tolerante do que o método da Média das Distâncias em relação às variáveis *volumeTotal*, *areaTotal*, *Eficiência* e *Duração*, contudo, mais flexível que o método Tukey. Devido a essa natureza intermediária, optou-se pelo Isolation Forest, uma vez que não restringe excessivamente o modelo

quanto a valores mais elevados, como no caso da duração de operação. A previsão desses valores, se não forem considerados outliers gerados por erro de registro, por exemplo, é essencial, dada a sua influência significativa em estratégias de planejamento de rotas para navios de apoio. É um fator crítico, contudo, a existência de informações na base de dados que expliquem tais ocorrências. Caso não existam, espera-se que o modelo tenha maior dificuldade em prever esses valores.

Uma vez removidos os *outliers*, é possível observar melhor a distribuição das variáveis dependentes duração e eficiência, bem como descrever características gerais da operação. A eficiência média registrada é de 25833,6 Kg/hora enquanto observa-se uma duração média de operação de 4,6 horas. A Figura 10 apresenta a distribuição de valores de Eficiência e Duração na base de dados.

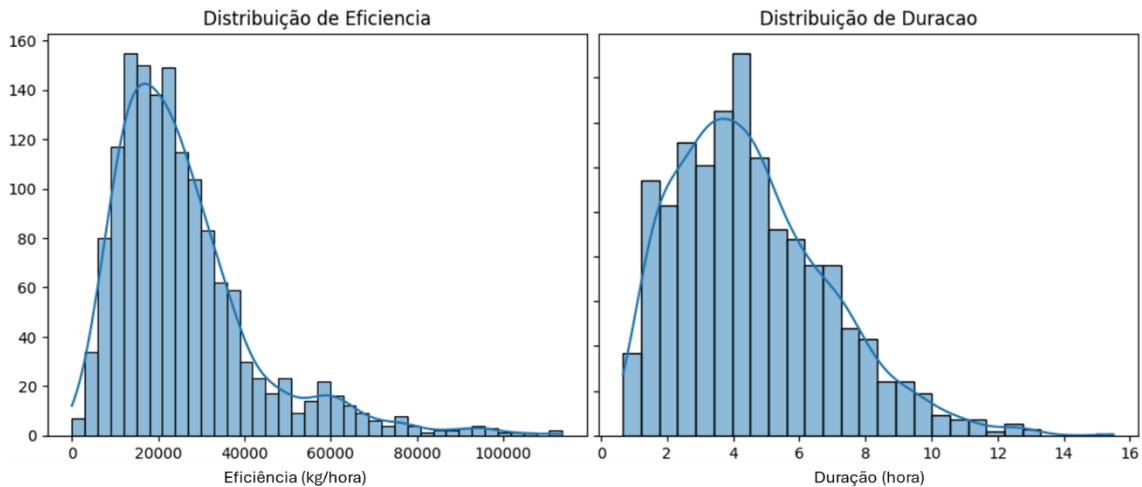


Figura 10: Histogramas de distribuição da Eficiência e da Duração

4.3.3 Seleção de Variáveis

Nesta etapa, inicialmente, observou-se que apenas uma das categorias da variável *classeOnda* possuía cerca de 99% dos registros, isto é, apenas 2 registros ocorreram em cenários de altura de onda entre 2,5 e 3,5m. Todos os demais apresentam altura de onda menor que 2,5. Deste modo, optou-se por remover a variável *classeOnda* do conjunto de dados.

Em seguida, foi realizada a análise de correlação entre as variáveis, que pode ser observada graficamente na Figura 11. Chama a atenção a baixa correlação linear entre as variáveis meteocanográficas e as variáveis alvo *Duracao* e *Eficiencia*, com valores próximos de zero. Já as variáveis de natureza mais operacional, que fazem parte do

planejamento da operação, como *nLingada*, *pesoTotal*, *volumeTotal* e *areaTotal*, possuem maior correlação com *Duracao* e *Eficiencia*.

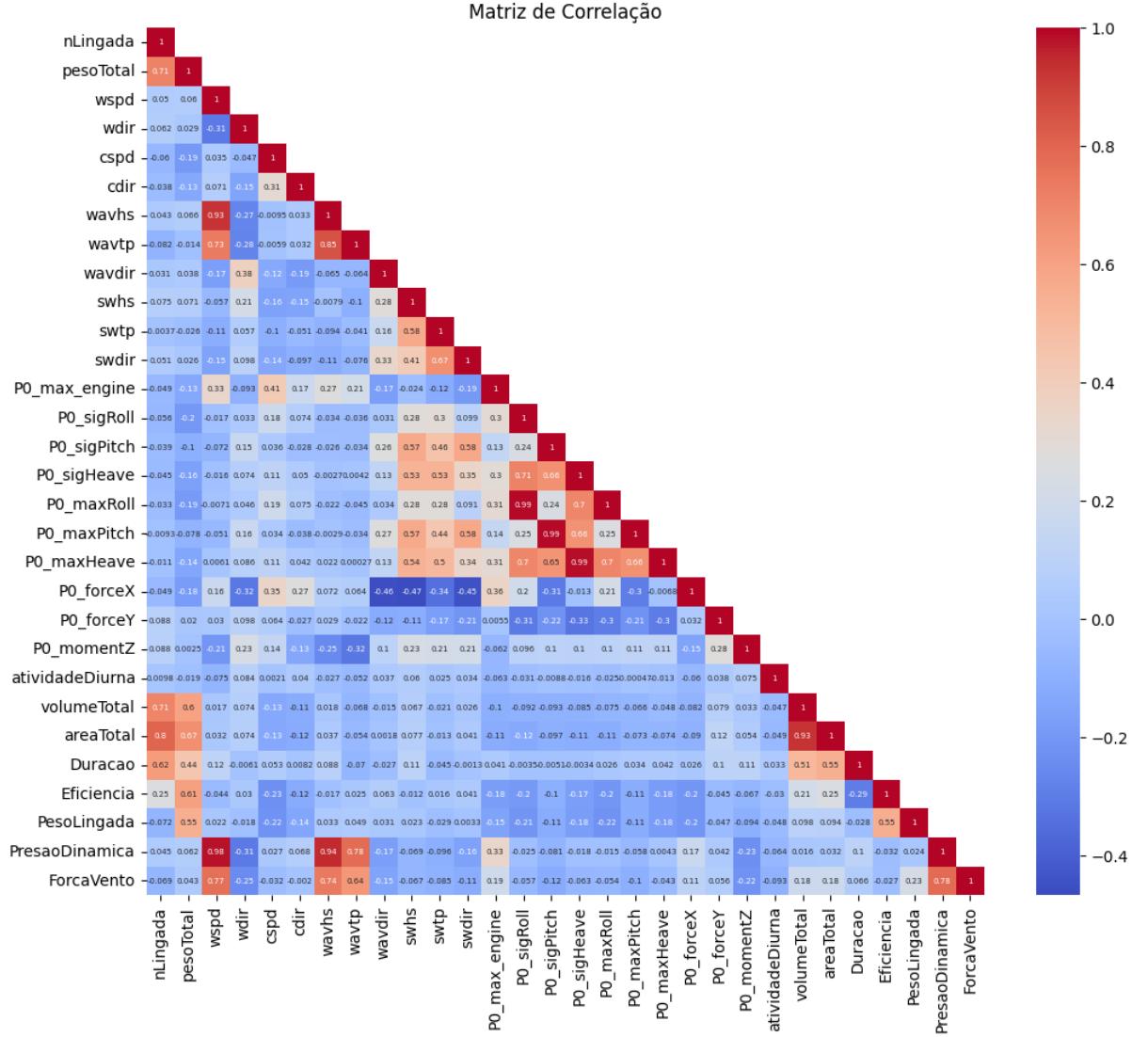


Figura 11: Matriz de correlação

Dentre as variáveis preditoras, estabeleceu-se um intervalo de correlação entre -0,85 e 0,85. Deste modo tem-se os seguintes pares de variáveis altamente correlacionadas linearmente:

- *volumeTotal* e *areaTotal*: 0,93
- *sigRoll* e *maxRoll*: 0,99
- *wavhs* e *wavtp*: 0,85
- *sigPitch* e *maxPitch*: 0,99
- *sigHeave* e *maxHeave*: 0,99

- *PressaoDinamica e wspd*: 0,98
- *PressaoDinamica e wavhs*: 0,94

A avaliação da relevância das variáveis por meio do algoritmo Random Forest foi, então, conduzida com o objetivo de estabelecer critérios para a seleção das mesmas. Os resultados dessa análise estão representados na Figura 12 e na Figura 13, que exibem a importância relativa de cada variável para a previsão da Duração e da Eficiência, respectivamente.

Com base nesses critérios, observou-se que a variável *areaTotal* desempenha um papel mais preponderante na definição da Duração em comparação com a Eficiência. Consequentemente, conforme as diretrizes estabelecidas, ela seria selecionada em um modelo, enquanto o *volumeTotal* seria empregado no outro. Contudo, é crucial destacar que a elevada correlação linear entre essas variáveis não implica necessariamente em redundância, conforme evidenciado pela análise da importância das variáveis. Além disso, tanto *volumeTotal* quanto *areaTotal* apresentam correlações promissoras com as variáveis alvo, indicando que podem fornecer informações cruciais para sua previsão. Por esses motivos, optou-se por retê-las na modelagem.

As variáveis *maxRoll*, *maxPitch* e *maxHeave* foram mantidas com base no critério de desempate segundo a importância. A variável *PressaoDinamica* foi excluída devido à sua baixa relevância e à presença de outras variáveis capazes de compensar sua ausência, conforme indicado pela análise de correlação. Por fim, embora a variável *wavhs* tenha sido mantida em detrimento de *wavtp* e a variável *wspd* tenha sido utilizada apenas quando a força do vento não era considerada no modelo, foi necessário um discernimento cuidadoso para tomar essas decisões.

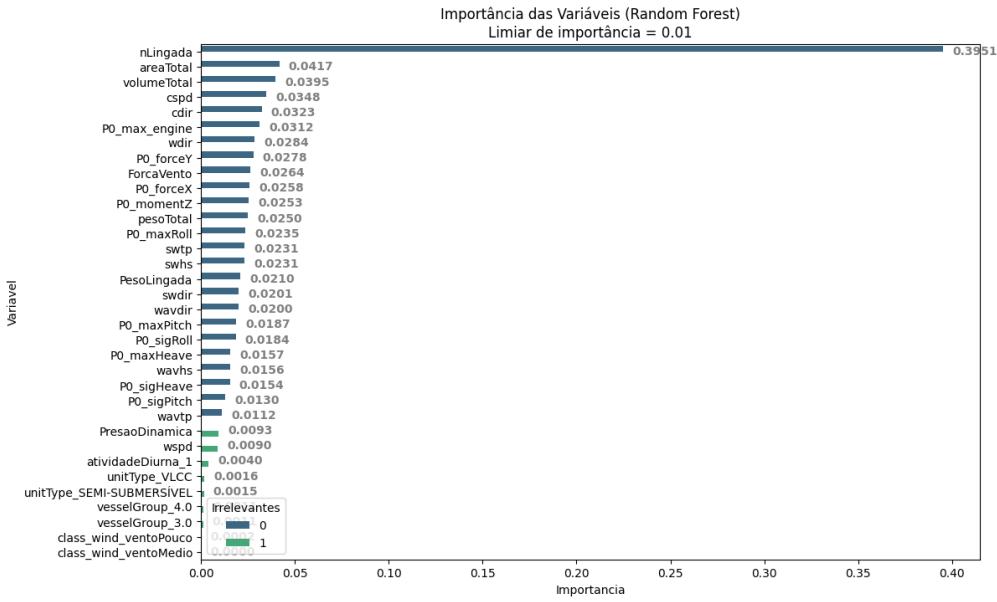


Figura 12: Importância das variáveis para previsão da Duração

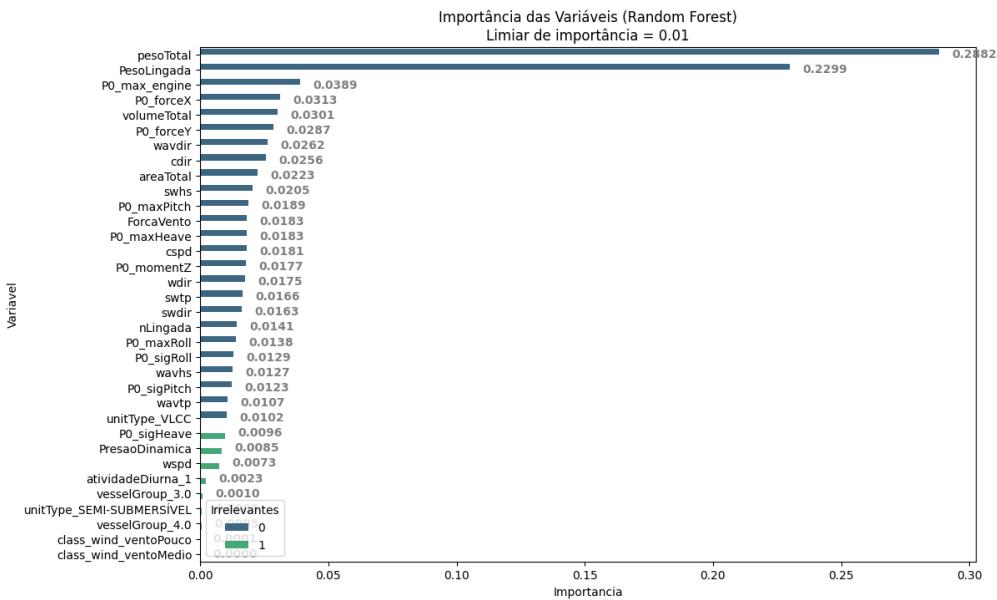


Figura 13: Importância as variáveis para previsão da Eficiência

Outro aspecto importante revelado pela análise de importância das variáveis é que as variáveis categóricas, quando estabelecemos uma importância relativa mínima de 1%, foram consideradas irrelevantes para o problema em questão. Portanto, *unitType*, *vesselGroup*, *atividadeDiurna* e *classeVento* foram desconsideradas no processo de modelagem.

Ainda observando a Figura 12 e na Figura 13, é interessante destacar a importância de *max_engine* (maior taxa de uso de propulsor, nos gráficos

$P0_max_engine$) e sua correlação com a velocidade do vento, e a importância de $cspd$ (velocidade da corrente) para previsão da eficiência e da duração. De fato, como pode ser mostrado na Figura 14, a eficiência média por intervalo de valores de max_engine cai conforme os limites dos intervalos aumentam. Contudo, os valores médios de eficiência observados por intervalo são muito menores que os valores máximos de eficiência apontados na distribuição da Figura 10.



Figura 14: Eficiência média por intervalo de valores de max_engine .

4.4 Otimização dos Modelos de Previsão

Antes da análise dos modelos de regressão, foram estabelecidos cenários para avaliar a inclusão de variáveis relacionadas às condições de mar. Assim, três cenários foram criados segundo as variáveis a serem mantidas na modelagem de dados. No primeiro cenário (Cenário 01), apenas as variáveis operacionais, que podem ser coletadas ainda no planejamento da operação, foram consideradas. No segundo (Cenário 02) além das condições operacionais, foram consideradas também as condições de mar obtidas por simulação. O último cenário (Cenário 03), abrange todas as variáveis anteriores, além da estimativa da força média do vento sobre as cargas içadas.

Além disso, foram implementadas, nos três cenários, as duas normalizações citadas na Seção 3.2.4. Desta forma, o procedimento proposto na seção 3.3 foi executado para cada cenário de modelo de dados e para cada técnica de padronização de dados implementada.

Inicialmente, o conjunto de dados foi particionado em um conjunto de treino (80% do conjunto total) e um conjunto de validação (20%). Essa divisão foi uniformemente aplicada em todos os cenários, garantindo que os conjuntos de treino e teste de cada cenário compartilhassem os mesmos registros, possibilitando uma comparação consistente dos resultados.

Neste estudo, foram implementados, utilizando a biblioteca *scikit learning* do *python*, os seguintes modelos de regressão para previsão da duração e eficiência das operações de transferência de carga entre as embarcações: Regressão Linear, Regressores de Vetor de Suporte, Árvore de Decisão, Florestas Aleatórias, Árvores Extras, e Regressão Parcial de Mínimos Quadrados. Os modelos foram ajustados, a partir do conjunto de teste, usando o método *GridSearchCV* do pacote *scikit learning*. Nesse método, cada combinação de valores de hiperparâmetros são avaliados via validação cruzada *k-fold*, tendo sido estabelecido um número de 5 partições no conjunto de treino. A métrica R^2 foi escolhida como critério para a determinação dos hiperparâmetros ótimos, visto que é uma medida que, teoricamente, penaliza de forma mais significativa os erros mais substanciais. A Tabela 9 apresenta os hiperparâmetros otimizados e os valores considerados. O modelo de Regressão Linear foi estabelecido como um modelo base, por ser o mais simples, e nenhum hiperparâmetro foi alterado, tenho sido mantidas as configurações padrão da biblioteca.

Tabela 9: Relação de hiperparâmetros otimizados por modelo de previsão e seus respectivos conjuntos de valores testados.

Modelo	Hiperparâmetros Otimizados	Valores Considerados
Support Vector Regressor	C	[0.1, 1, 10]
	kernel	['poly', 'rbf']
	max_depth	[None, 10, 20, 30]
Decision Tree	min_samples_split	[2, 5, 10]
	min_samples_leaf	[1, 2, 4]
Random Forest Regressor	n_estimators	[200, 800, 1400]
	max_features	[1.0, 'sqrt']
	max_depth	[10, 60, 110]
Partial Least Squares	n_components	[1, 2, 3, 4, 5]
Extra Trees Regressor	n_estimators	[50, 100, 200]
	max_features	['auto', 'sqrt', 'log2']
	min_samples_split	[2, 5, 10]

Os gráficos apresentados na Figura 15 fornecem uma representação da distribuição dos valores da métrica R^2 para cada modelo obtida no Cenário 1 para a modelagem da Eficiência da operação. Essa distribuição é constituída pelos valores médios do R^2 , os quais são calculados por meio de validação cruzada para todas as combinações de hiperparâmetros. Em outras palavras, cada valor é a média dos R^2 médios obtidos após o treinamento do modelo com o mesmo conjunto de hiperparâmetros em todas as cinco rodadas da validação cruzada. Os gráficos aqui representados visam ilustrar observações feitas a partir desta etapa.

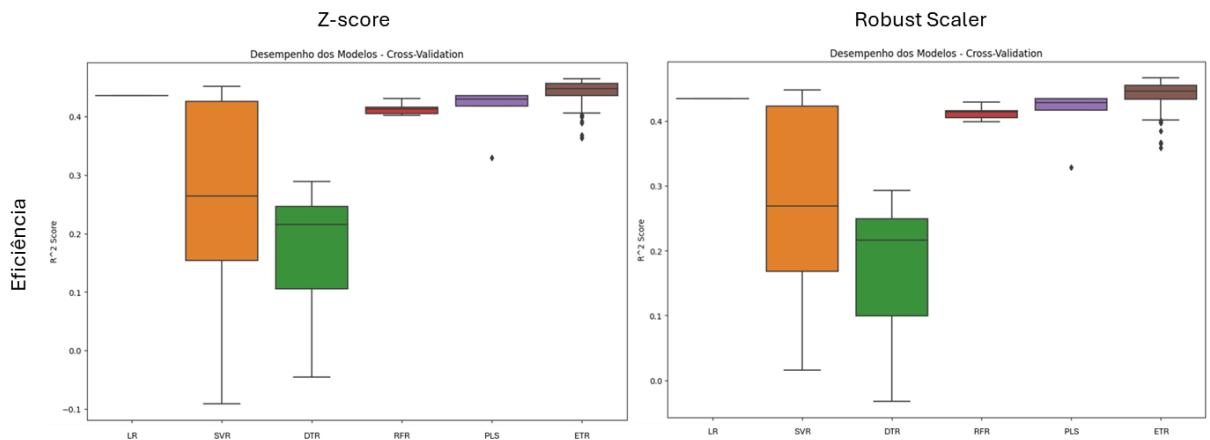


Figura 15: Distribuição da métrica R^2 obtida por combinação de hiperparâmetros na validação cruzada para previsão da Eficiência no Cenário 1

De forma geral, observa-se que os modelos DTR e SVR exibem métricas mais voláteis, dependendo dos hiperparâmetros utilizados. Essa instabilidade pode ser observada na Figura 15, onde percebe-se que os valores de R^2 do modelo SVR variam de, aproximadamente -0,1 a 0,4. Contudo, com a adição das variáveis de condição de tempo (Cenário 2), percebe-se uma diminuição na variação dos resultados obtidos por esse modelo, enquanto o modelo DTR mantém-se com as métricas menos favoráveis, vide Figura 16.

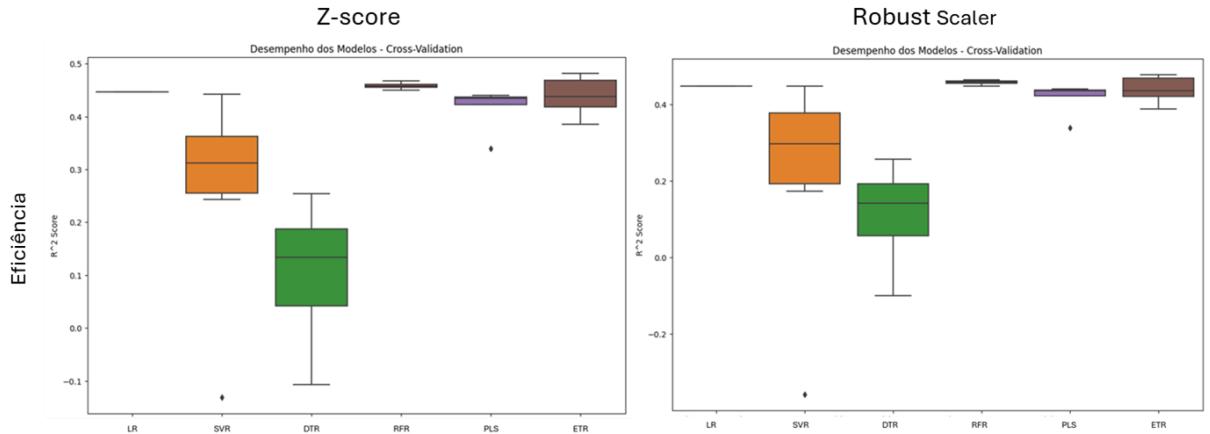


Figura 16: Distribuição da métrica R^2 obtida por combinação de hiperparâmetros na validação cruzada para previsão da Eficiência no Cenário 2

Em contrapartida, os modelos ETR, PLS e RFR demonstram um comportamento mais estável em todos os cenários, apresentando métricas bastante semelhantes, que forma um platô na métrica de avaliação entre 0,45 e 0,5. Nessa faixa de valores, no entanto, encontra-se também o resultado gerado pelo método RL, não observando-se muita vantagem no uso de um desses modelos em detrimento do modelo mais básico. Além disso, observando-se a Figura 17, não é possível observar nenhuma melhora significativa em decorrência do uso da estimativa de força do vento na modelagem.

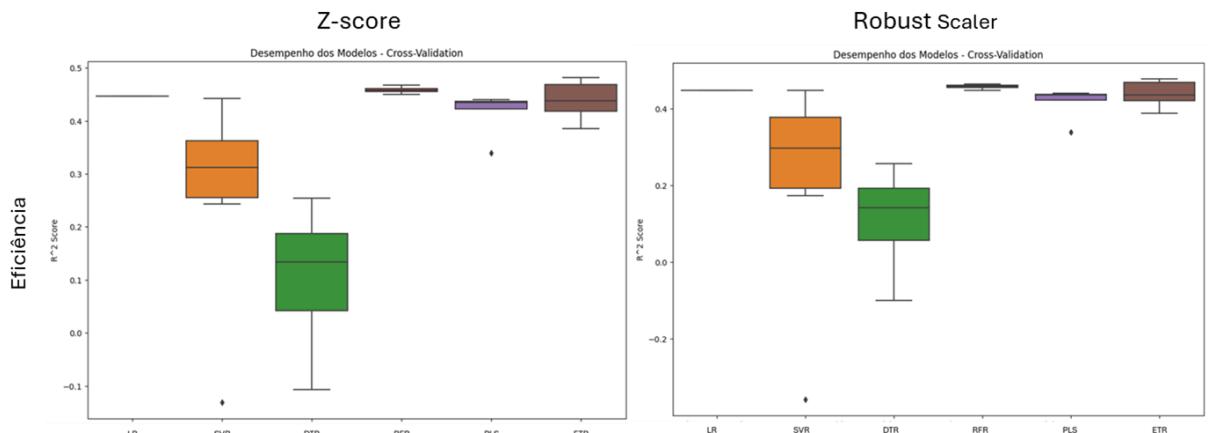


Figura 17: Distribuição da métrica R^2 obtida por combinação de hiperparâmetros na validação cruzada para previsão da Eficiência no Cenário 3.

Outro ponto observado é que o método de padronização dos dados não parece interferir muito nas métricas de avaliação. Contudo, na previsão da Duração, não só o modelo ETR mostrou-se mais estável quando os dados foram normalizados pelo Robust

Scaler, como também o SVR apresentou métricas mais próximas do platô atingido nesta fase da execução da metodologia proposta.

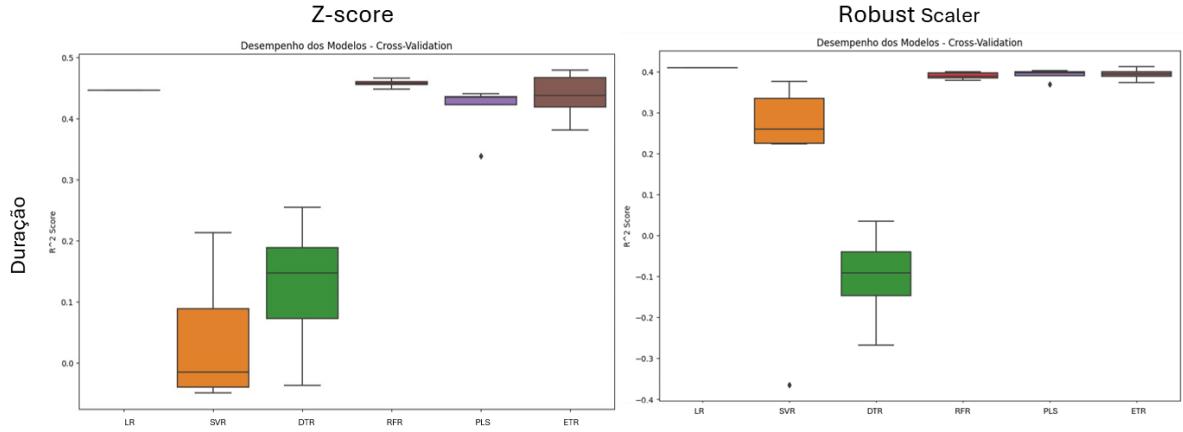


Figura 18: Distribuição da métrica R^2 obtida por combinação de hiperparâmetros na validação cruzada para previsão da Duração no Cenário 2.

Os valores de R^2 médio por modelo obtidos nesta fase, tanto para os modelos de previsão da Eficiência quanto da Duração, estão dispostos no Apêndice I.

4.5 Resultados e Avaliação da Previsão

Após definir os conjuntos ideais de hiperparâmetros para cada método de previsão e cenário, os modelos foram treinados novamente utilizando o conjunto de treino completo, ajustando assim os parâmetros. A avaliação dos resultados foi realizada com base nas previsões feitas para o conjunto de validação.

A Tabela 10 mostra os valores das métricas de avaliação obtidas para o conjunto de validação na previsão da duração. Em destaque, estão os melhores resultados para a variável alvo. O modelo SVR aplicado à modelagem do Cenário 02 normalizado pelo método z-score resultou nos menores valores de erro MAPE e RMSE, e no maior valor de R^2 . Quantifica-se uma melhora de 6,5% de melhoria da métrica MAPE em relação à melhor métrica obtida pelo método de Regressão Linear, 3% na RMSE e 8,6% no R^2 . Segundo a métrica RMSE, o método é capaz de prever a duração da operação com um erro médio de 1,68 horas.

A Tabela 11 mostra os valores das métricas de avaliação obtidas para o conjunto de validação na previsão da eficiência. No destaque, percebe-se que modelo RFR aplicado à modelagem do Cenário 03 normalizado pelo método z-score resultou nos menores valores de RMSE, e no maior valor de R^2 , enquanto o modelo SVR do cenário 03 com *Robust Scaler* apresentou o menor erro MAPE. Em relação ao melhor modelo LR, o RFR teve uma melhora de 3% na métrica RMSE, e 7% no R^2 , porém uma piora de 1% na MAPE. O SVR, no entanto, obteve uma melhora de 15% no MAPE, 2% no RMSE e 5% no R.

Na Tabela 10, também são apresentadas métricas de avaliação calculadas com base nos valores de eficiência derivados da duração prevista. Nota-se que uma previsão mais precisa da duração não necessariamente se traduz em uma previsão mais precisa da eficiência. Assim, a previsão da duração pode ser mais confiável dependendo de sua usabilidade. Além disso, conforme evidenciado na Tabela 11, as métricas de MAPE indicam que o modelo RFR oferece a melhor previsão para a eficiência, enquanto as métricas R^2 e RMSE mostram desempenho superior para o modelo SVR. Dada a natureza da variável eficiência, que abrange uma ampla escala de valores, esse padrão pode sugerir que o SVR é mais eficaz na manipulação de valores mais elevados de eficiência, que são responsáveis pelos erros mais pronunciados nos modelos, enquanto as previsões do RFR tendem a se aproximar mais dos valores médios reais.

Tabela 10: Métricas de avaliação dos modelos de previsão da duração

Cenário	Modelo	Normalização	Duração - Validação			Eficiência Calculada		
			MAPE	RMSE (hora)	R ²	MAPE	RMSE (Kg/hora)	R ²
Cenário 01	DTR	Z-score	0,461	2,225	0,027	0,385	14113	0,253
	DTR	Robust Scaler	0,462	2,235	0,018	0,393	14194	0,245
	ETR	Z-score	0,411	1,804	0,360	0,318	12876	0,378
	ETR	Robust Scaler	0,408	1,784	0,375	0,315	12918	0,374
	LR	Z-score	0,406	1,805	0,360	0,314	12977	0,369
	LR	Robust Scaler	0,406	1,805	0,360	0,314	12977	0,369
	PLS	Z-score	0,406	1,805	0,360	0,314	12977	0,369
	PLS	Robust Scaler	0,406	1,805	0,360	0,314	12977	0,369
	RFR	Z-score	0,403	1,834	0,339	0,319	12823	0,384
	RFR	Robust Scaler	0,403	1,837	0,337	0,320	12858	0,380
Cenário 02	SVR	Z-score	0,375	1,822	0,347	0,334	12481	0,416
	SVR	Robust Scaler	0,376	1,820	0,349	0,331	12481	0,416
	DTR	Z-score	0,488	2,240	0,014	0,397	14873	0,171
	DTR	Robust Scaler	0,486	2,247	0,008	0,395	14800	0,179
	ETR	Z-score	0,402	1,752	0,397	0,305	12786	0,387
	ETR	Robust Scaler	0,403	1,732	0,410	0,303	12705	0,395
	LR	Z-score	0,391	1,731	0,411	0,308	12509	0,413
	PLS	Z-score	0,398	1,732	0,411	0,316	12745	0,391
	PLS	Robust Scaler	0,398	1,732	0,411	0,316	12745	0,391
	RFR	Z-score	0,412	1,733	0,410	0,304	12602	0,405
Cenário 03	RFR	Robust Scaler	0,413	1,727	0,414	0,303	12606	0,404
	SVR	Z-score	0,365	1,679	0,446	0,305	11744	0,483
	SVR	Robust Scaler	0,367	1,683	0,443	0,305	11706	0,486
	DTR	Z-score	0,486	2,233	0,020	0,394	14878	0,170
	DTR	Robust Scaler	0,489	2,251	0,004	0,396	14943	0,163
	ETR	Z-score	0,410	1,767	0,387	0,306	12896	0,377
	ETR	Robust Scaler	0,401	1,743	0,403	0,304	12687	0,397
	LR	Z-score	0,392	1,733	0,409	0,304	12563	0,408
	LR	Robust Scaler	0,391	1,731	0,411	0,308	12509	0,413
	PLS	Z-score	0,397	1,732	0,410	0,312	12630	0,402
	PLS	Robust Scaler	0,397	1,732	0,410	0,312	12630	0,402
	RFR	Z-score	0,415	1,736	0,408	0,304	12637	0,401
	RFR	Robust Scaler	0,413	1,720	0,419	0,303	12644	0,401
	SVR	Z-score	0,368	1,679	0,446	0,307	11740	0,483
	SVR	Robust Scaler	0,368	1,679	0,446	0,304	11667	0,490

Tabela 11: Métricas de avaliação dos modelos de previsão da eficiência

Cenário	Modelo	Normalização	MAPE	RMSE (Kg/hora)	R ²
Cenário 01	ETR	Z-score	0,384	12213,53	0,441
	ETR	Robust Scaler	0,378	12265,94	0,436
	PLS	Z-score	0,386	12453,60	0,419
	PLS	Robust Scaler	0,386	12453,60	0,419
	LR	Robust Scaler	0,387	12463,21	0,418
	LR	Z-score	0,387	12463,21	0,418
	SVR	Z-score	0,338	12480,87	0,416
	RFR	Robust Scaler	0,380	12485,58	0,416
	RFR	Z-score	0,380	12487,33	0,415
	SVR	Robust Scaler	0,339	12576,39	0,407
Cenário 02	DTR	Robust Scaler	0,397	13967,55	0,269
	DTR	Z-score	0,400	13969,80	0,268
	SVR	Robust Scaler	0,316	11803,77	0,478
	RFR	Z-score	0,373	11833,45	0,475
	RFR	Robust Scaler	0,373	11847,84	0,474
	SVR	Z-score	0,322	11856,70	0,473
	ETR	Robust Scaler	0,372	11906,48	0,469
	ETR	Z-score	0,372	11927,89	0,467
	LR	Z-score	0,369	12034,85	0,457
	LR	Robust Scaler	0,369	12034,85	0,457
Cenário 03	PLS	Z-score	0,372	12057,49	0,455
	PLS	Robust Scaler	0,372	12057,49	0,455
	DTR	Robust Scaler	0,401	14181,65	0,246
	DTR	Z-score	0,403	14321,26	0,231
	RFR	Z-score	0,371	11668,37	0,490
	RFR	Robust Scaler	0,370	11784,02	0,479
	SVR	Robust Scaler	0,314	11787,00	0,479
	SVR	Z-score	0,318	11832,56	0,475
	ETR	Robust Scaler	0,370	11884,97	0,470
	ETR	Z-score	0,371	11920,58	0,467

Entretanto, percebem-se diferenças pouco expressivas entre as métricas dos modelos obtidos. A grande maioria dos valores para a métrica R², por exemplo, situa-se em um platô entre 0,45 e 0,5. Isso sugere uma dificuldade dos estimadores em definir o tempo de duração ou a eficiência com base nas variáveis disponíveis, seja devido à

ausência de variáveis que expliquem melhor a duração das operações, ou pela imprecisão das variáveis meteoceanográficas. No entanto, é notável que, de maneira geral, os cenários nos quais as informações sobre as condições do mar foram incorporadas na modelagem (cenários 02 e 03) apresentaram uma clara melhoria em comparação com o Cenário 01, conforme registrado na Figura 19. As métricas MAPE, RMSE e R² tiveram uma melhoria aproximada de 2,5%, 4% e 14% respectivamente.

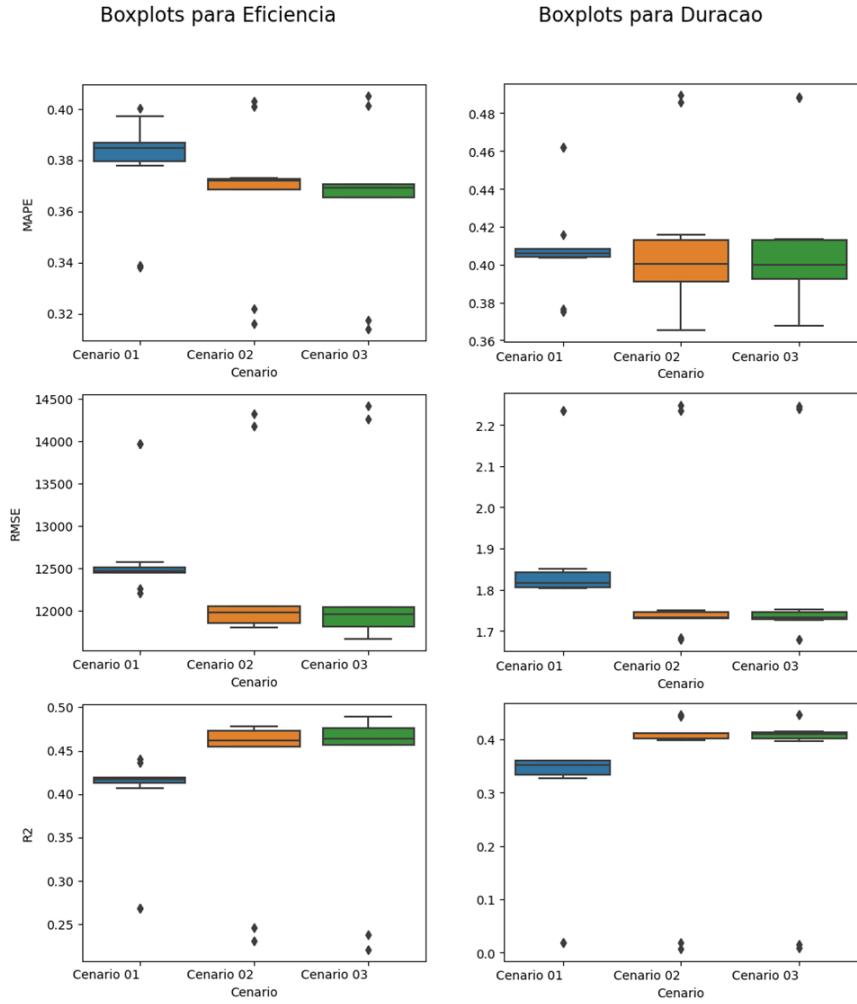


Figura 19: Distribuição das métricas de avaliação do conjunto de validação por cenário adotado.

Gkerekos et al. (2019) ao abordar o desafio de definir a FOC, alcançaram valores de R² superiores a 0,95. Embora as naturezas distintas dos problemas inviabilizem uma comparação direta, a métrica R² obtida neste trabalho pode ser considerada moderada, uma vez que os modelos conseguem identificar alguns cenários e reproduzir padrões dos dados reais, porém geram previsões dispersas. Esse fenômeno de subajuste é observado consistentemente em todos os modelos, evidenciado da Figura

20 à Figura 26, que ilustram a comparação entre os valores reais e as previsões dos modelos considerados mais eficazes.

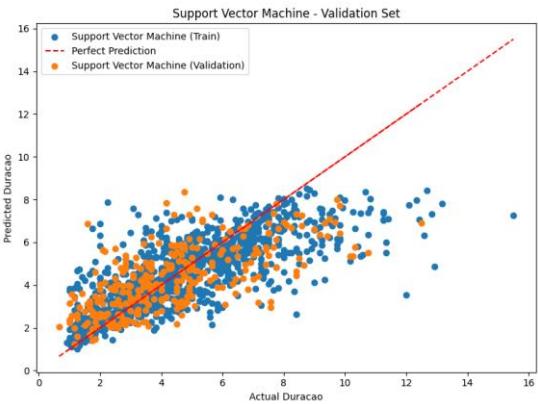


Figura 20: Previsões de Duração pelo modelo SVR

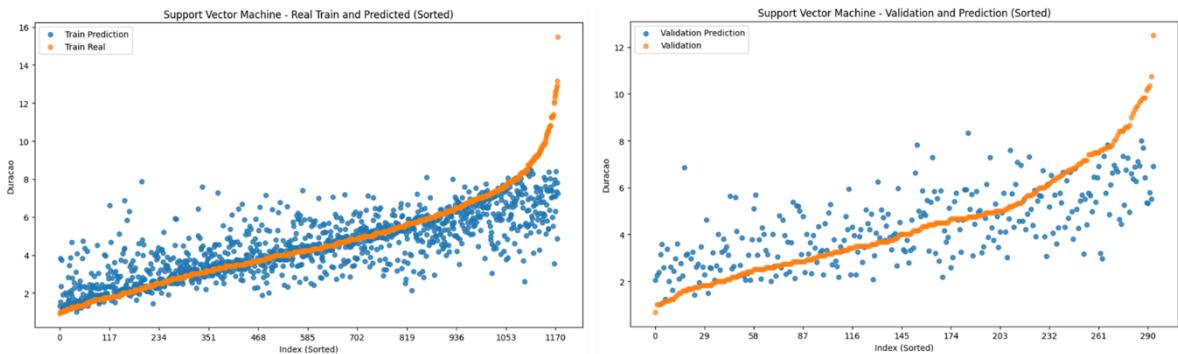


Figura 21: Previsões de Duração pelo modelo SVR: Treino e Validação

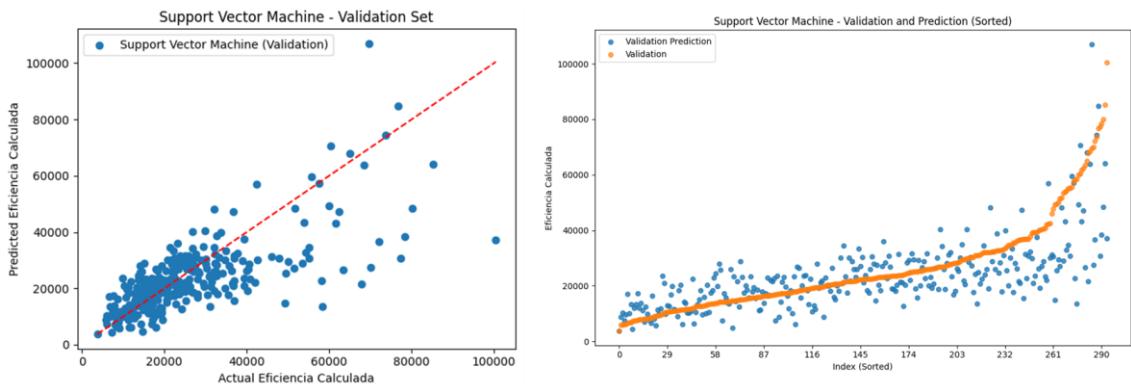


Figura 22: Eficiência calculada a partir da duração prevista

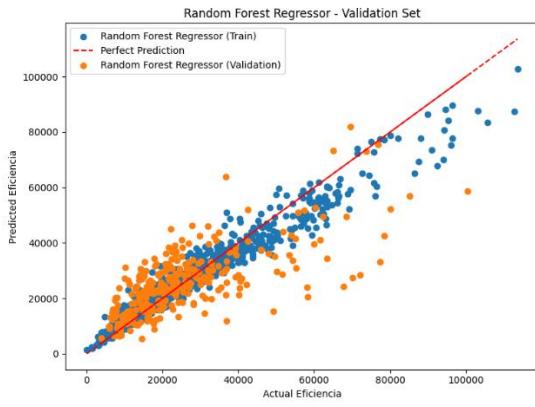


Figura 23: Previsões de Eficiência pelo modelo RFR

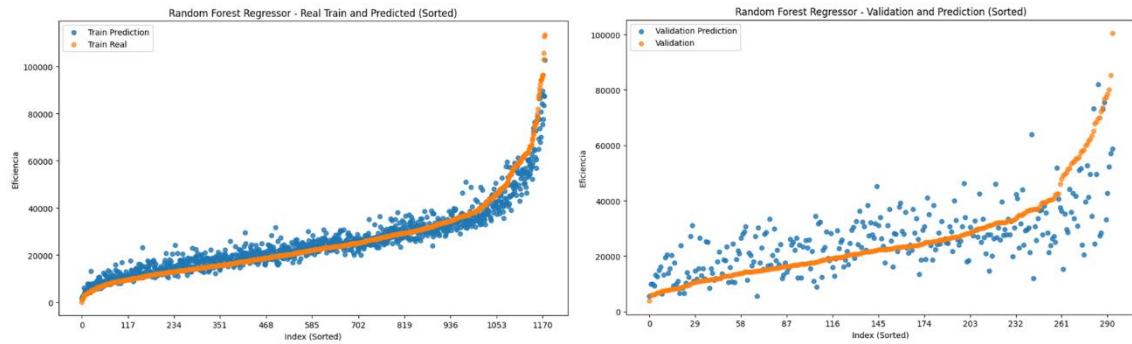


Figura 24: Previsões de Duração pelo modelo RFR: Treino e Validação

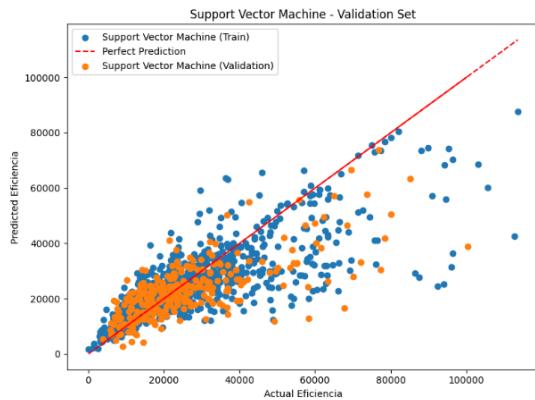


Figura 25: Previsões de Eficiência pelo modelo SVR

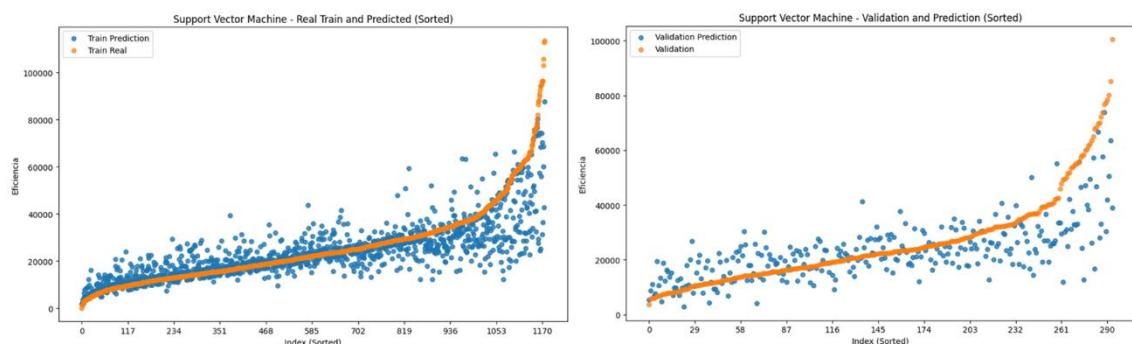


Figura 26: Previsões de Duração pelo modelo RFR: Treino e Validação

A análise da métrica RMSE do modelo SVR, representada na Figura 27, proporciona insights sobre o desempenho do modelo em diferentes intervalos de valores reais de duração. Destaca-se a presença de erros mais significativos nos extremos da escala de duração real, revelando uma tendência de subajuste nesses cenários específicos. Especificamente, quando a duração real ultrapassa 6 horas, observa-se um aumento substancial no erro médio, chegando a superar 2,5 horas (cerca do dobro do RMSE observado nos intervalos médios). Essa tendência ressalta a complexidade do modelo em prever com precisão eventos de longa duração, evidenciando a necessidade de aprimoramentos para explicar de forma eficaz situações mais prolongadas de transferência de carga entre embarcações.

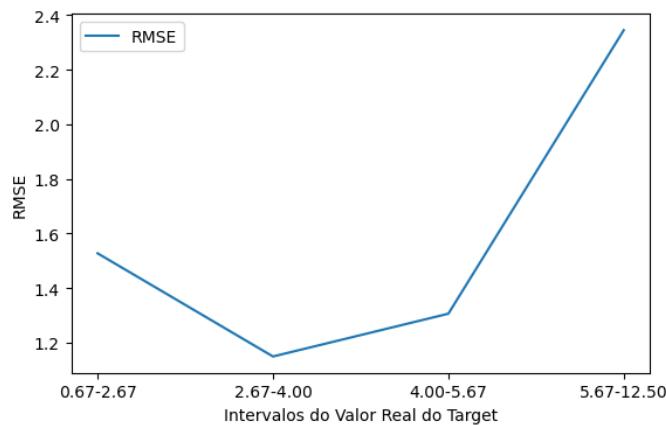


Figura 27: RMSE por intervalo real da duração.

5 CONCLUSÃO

5.1 Considerações Finais

O presente trabalho buscou desenvolver um modelo preditivo para estimar a duração e eficiência (ou produtividade) das operações de transferência de carga entre embarcações na indústria de exploração e produção de óleo e gás offshore. Para atingir esse objetivo, foram realizadas análises exploratórias de dados, avaliações de diferentes modelos de previsão e a incorporação de informações sobre as condições de mar, considerando variáveis meteoceanográficas.

No desenvolvimento do modelo considerou-se, inicialmente, variáveis contínuas e categóricas, estas tendo se mostrado menos importantes para a previsão. Foram criados cenários distintos, contemplando diferentes níveis de detalhamento das informações sobre as condições de mar, a fim de avaliar o impacto dessas informações no modelo. O processo de desenvolvimento de variáveis, remoção de outliers e seleção de variáveis foi conduzido com critérios rigorosos, considerando a relevância das informações para a previsão desejada.

A análise dos modelos revelou que a inclusão de variáveis relacionadas às condições do mar trouxe melhorias na precisão das previsões, especialmente para a duração das operações. O modelo Support Vector Regression (SVR) aplicado ao Cenário 02, se destacou na previsão da duração, enquanto o Random Forest Regressor (RFR) aplicado ao Cenário 03 apresentou desempenho superior na previsão da eficiência. Tais melhorias, contudo não são muito significativas. Afinal, em todos os casos as previsões atingem um nível máximo R^2 em torno de 0,5. Isso se deve à dificuldade de se explicar mais precisamente, com os dados disponíveis, a ocorrência de eventos de alta duração, levando os modelos a um subajuste especialmente nessa faixa de valores.

A ausência de registros históricos detalhados contribuiu para desafios adicionais na modelagem. Entende-se que as premissas adotadas para consolidação da base de dados podem não representar a realidade operacional e meteoceanográfica dos eventos registrados. Além disso, a ausência de registros de eventos ocasionados sob condições climáticas mais adversas impossibilita a análise desse impacto sobre as operações, uma vez que a quase totalidade dos eventos registrados ocorreram em situações normais ou

ideais meteoceanográficas, segundo as faixas de valores consideradas mais críticas por CTPP (2020), Norlund & Gribkovskaia (2017) e Halvorsen-Weare & Fagerholt (2011b).

Certa influência foi, contudo, observada, uma vez que os cenários que consideram as condições de mar apresentam melhorias em relação ao cenário onde tais condições são descartadas. A análise de importância das variáveis aponta a variável *max_engine*, que tem correlação com a velocidade das correntes marítimas, como uma das variáveis mais importantes para a previsão. Isso sugere que um aprofundamento do trabalho nessa direção pode retornar ainda melhores resultados.

Em síntese, este trabalho contribui para a compreensão da influência das condições de mar nas operações logísticas offshore, e fornece *insights* valiosos para otimização do planejamento logístico na indústria de exploração e produção de óleo e gás. O modelo desenvolvido, embora apresente espaço para melhorias, oferece um bom recurso para o planejamento logístico de suprimento, reforçando a importância de abordagens adaptáveis e contínuo aprimoramento na modelagem preditiva para cenários desafiadores como os encontrados na indústria offshore.

5.2 Sugestão para Trabalhos Futuros

Diversas sugestões para trabalhos futuros emergem da análise dos resultados obtidos e das complexidades identificadas durante o desenvolvimento deste estudo. Uma proposta promissora consiste na implementação de intervalos de previsão para a duração e eficiência das operações de transferência de carga, em vez de apresentar apenas valores pontuais. Essa abordagem permitiria uma avaliação mais abrangente da incerteza associada às estimativas, fornecendo insights valiosos para uma gestão estratégica e informada das operações offshore.

Outro aspecto crucial para aprimorar o modelo diz respeito à análise mais aprofundada das variáveis meteoceanográficas. Especificamente, explorar relações mais complexas entre essas variáveis, como a direção da onda e do vento, pode enriquecer significativamente a capacidade preditiva do modelo. A consideração de cálculos de resultantes desses elementos também pode contribuir para uma representação mais fiel das condições ambientais durante as operações.

É essencial realizar análises de sensibilidade em relação às premissas adotadas para a obtenção de variáveis meteoceanográficas. Embora a computação dessas

derivadas tenha sido limitada devido ao custo computacional, recomenda-se a coleta de informações adicionais para gerar dados mais confiáveis, proporcionando uma base sólida para a modelagem.

A exploração de técnicas avançadas, como redes neurais, para a previsão das operações logísticas offshore pode representar um avanço significativo. Esses modelos têm a capacidade de lidar com a complexidade inerente aos dados marítimos, oferecendo uma abordagem mais adaptável e flexível.

Uma revisão das abordagens utilizadas para considerar as variáveis meteoceanográficas também é uma linha de pesquisa interessante. A análise de alternativas que capturem informações mais detalhadas durante a duração da operação pode fornecer insights valiosos sobre a influência dessas condições nas operações logísticas.

Por fim, investigar a existência de outros aspectos operacionais que possam enriquecer o modelo é uma perspectiva relevante. A busca por variáveis adicionais, que possam contribuir para uma melhor explicabilidade do modelo, pode resultar em melhorias significativas na compreensão e previsão das operações logísticas offshore. Essas sugestões representam direções promissoras para pesquisas subsequentes, visando aprimorar a eficiência e sustentabilidade das operações na indústria de exploração e produção de óleo e gás offshore.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- ABEAM. (2023). *Tipos de embarcações*. <https://abeam.org.br/embarcacoes/>.
- Bal Beşikçi, E., Arslan, O., Turan, O., & Ölçer, A. I. (2016). An artificial neural network based decision support system for energy efficient ship operations. *Computers and Operations Research*, 66, 393–401. <https://doi.org/10.1016/j.cor.2015.04.004>
- Ballou, R. H. (2001). *Gerenciamento da cadeia de suprimentos: planejamento, organização e logística empresarial* (4º ed). Bookmann.
- Bialystocki, N., & Konovessis, D. (2016). On the estimation of ship's fuel consumption and speed curve: A statistical approach. *Journal of Ocean Engineering and Science*, 1(2), 157–166. <https://doi.org/10.1016/j.joes.2016.02.001>
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning* (M. Jordan, J. Kleinberg, & B. Schölkopf, Orgs.; 1º ed). Springer.
- Clarke, B., Fokoue, E., & Zhang, H. H. (2009). *Principles and theory for data mining and machine learning*. Springer Science & Business Media.
- CTPP. (2020). *Norma Regulamentadora No. 37 (NR-37)*. <https://www.gov.br/trabalho-e-emprego/pt-br/assuntos/inspecao-do-trabalho/seguranca-e-saude-no-trabalho/ctpp-nrs/norma-regulamentadora-no-37-nr-37>.
- Evsukoff, A. G. (2020). *INTELIGÊNCIA COMPUTACIONAL: Fundamentos e aplicações*. <http://www.e-papers.com.br>
- Gilbert, C., Browell, J., & McMillan, D. (2019). A Data-driven Vessel Motion Model for Offshore Access Forecasting. *OCEANS 2019 - Marseille*, 1–6. <https://doi.org/10.1109/OCEANSE.2019.8867176>
- Gkerekos, C., Lazakis, I., & Theotokatos, G. (2019). Machine learning models for predicting ship main engine Fuel Oil Consumption: A comparative study. *Ocean Engineering*, 188, 106282. <https://doi.org/10.1016/j.oceaneng.2019.106282>

Halvorsen-Weare, E. E., & Fagerholt, K. (2011a). *Robust Supply Vessel Planning* (p. 559–573). https://doi.org/10.1007/978-3-642-21527-8_62

Halvorsen-Weare, E. E., & Fagerholt, K. (2011b). *Robust Supply Vessel Planning* (p. 559–573). https://doi.org/10.1007/978-3-642-21527-8_62

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques* (3º ed.). Elsevier Inc.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction* (2º ed.). Springer.

Kisialiou, Y., Gribkovskaia, I., & Laporte, G. (2021). Periodic supply vessel planning under demand and weather uncertainty. *INFOR: Information Systems and Operational Research*, 59(3), 411–438. <https://doi.org/10.1080/03155986.2021.1907872>

Krata, P., Kniat, A., Vettor, R., Krata, H., & Guedes Soares, C. (2021). The Development of a Combined Method to Quickly Assess Ship Speed and Fuel Consumption at Different Powertrain Load and Sea Conditions. *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, 15(2), 437–444. <https://doi.org/10.12716/1001.15.02.23>

LIEBHERR. (2017). *Documento para treinamento: As influências do vento na operação do guindaste*. www.liebherr.com

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. *2008 Eighth IEEE International Conference on Data Mining*, 413–422. <https://doi.org/10.1109/ICDM.2008.17>

Lu, R., Turan, O., Boulogouris, E., Banks, C., & Incecik, A. (2015). A semi-empirical ship operational performance prediction model for voyage optimization towards energy efficient shipping. *Ocean Engineering*, 110, 18–28. <https://doi.org/10.1016/j.oceaneng.2015.07.042>

- Lundh, M., Garcia-Gabin, W., Tervo, K., & Lindkvist, R. (2016). Estimation and Optimization of Vessel Fuel Consumption. *IFAC-PapersOnLine*, 49(23), 394–399. <https://doi.org/10.1016/j.ifacol.2016.10.436>
- Meng, Q., Du, Y., & Wang, Y. (2016). Shipping log data based container ship fuel efficiency modeling. *Transportation Research Part B: Methodological*, 83, 207–229. <https://doi.org/10.1016/j.trb.2015.11.007>
- Norlund, E. K., & Gribkovskaia, I. (2017). Environmental performance of speed optimization strategies in offshore supply vessel planning under weather uncertainty. *Transportation Research Part D: Transport and Environment*, 57, 10–22. <https://doi.org/10.1016/j.trd.2017.08.002>
- Ogura, T., Inoue, T., & Uchihira, N. (2021). Prediction of arrival time of vessels considering future weather conditions. *Applied Sciences (Switzerland)*, 11(10). <https://doi.org/10.3390/app11104410>
- Pérez-Canosa, J. M., Orosa, J. A., Fraguera, F., & López-Varela, P. (2022). Proposal of Optimal Operation in Ship Rolling Motion Considering Sea State Conditions. *Journal of Marine Science and Engineering*, 10(5). <https://doi.org/10.3390/jmse10050669>
- Russell, S. J., & Norvig, P. (2010). *Artificial Intelligence A Modern Approach* (3th ed). Pearson Education, Inc.
- Simonsen, M., Walnum, H., & Gössling, S. (2018). Model for Estimation of Fuel Consumption of Cruise Ships. *Energies*, 11(5), 1059. <https://doi.org/10.3390/en11051059>
- Tsitsilis, K.-M., & Theotokatos, G. (2018). A novel systematic methodology for ship propulsion engines energy management. *Journal of Cleaner Production*, 204, 212–236. <https://doi.org/10.1016/j.jclepro.2018.08.154>
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, Mass. : Addison-Wesley Pub. Co.

Wang, S., Ji, B., Zhao, J., Liu, W., & Xu, T. (2018). Predicting ship fuel consumption based on LASSO regression. *Transportation Research Part D: Transport and Environment*, 65, 817–824. <https://doi.org/10.1016/j.trd.2017.09.014>

Winter, R. de. (2018). *Designing Ships using Constrained Multi-Objective Efficient Global Optimization* [MASTER'S THESIS]. Leiden University.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)

APÊNDICE I – MÉTRICAS DE AVALIAÇÃO

Cenario	Alvo	Modelo	Normalização	Validação cruzada			Treino			Validação		
				MAPE	RMSE	R2	MAPE	RMSE	R2	MAPE	RMSE	R2
1	Eficiencia	LR	Z-score	0,63	12499,41	0,45	0,62	12333,61	0,46	0,39	12463,21	0,42
1	Eficiencia	SVR	Z-score	0,44	12385,97	0,46	0,40	11832,95	0,50	0,34	12480,87	0,42
1	Eficiencia	DTR	Z-score	0,47	14047,07	0,30	0,31	8993,81	0,71	0,40	13969,80	0,27
1	Eficiencia	RFR	Z-score	0,47	12571,37	0,44	0,27	7191,01	0,82	0,38	12487,33	0,42
1	Eficiencia	PLS	Z-score	0,63	12498,72	0,45	0,62	12334,65	0,46	0,39	12453,60	0,42
1	Eficiencia	ETR	Z-score	0,49	12222,63	0,47	0,39	9989,93	0,65	0,38	12213,53	0,44
1	Eficiencia	LR	RobustScaler	0,63	12499,41	0,45	0,62	12333,61	0,46	0,39	12463,21	0,42
1	Eficiencia	SVR	RobustScaler	0,48	12412,42	0,45	0,42	11832,68	0,50	0,34	12576,39	0,41
1	Eficiencia	DTR	RobustScaler	0,47	13992,64	0,31	0,31	8993,81	0,71	0,40	13967,55	0,27
1	Eficiencia	RFR	RobustScaler	0,47	12564,59	0,44	0,27	7200,15	0,82	0,38	12485,58	0,42
1	Eficiencia	PLS	RobustScaler	0,63	12498,72	0,45	0,62	12334,65	0,46	0,39	12453,60	0,42
1	Eficiencia	ETR	RobustScaler	0,48	12230,53	0,47	0,40	10263,24	0,63	0,38	12265,94	0,44
1	Duracao	LR	Z-score	0,38	1,82	0,40	0,38	1,81	0,41	0,41	1,81	0,36
1	Duracao	SVR	Z-score	0,36	1,84	0,38	0,33	1,78	0,42	0,38	1,82	0,35
1	Duracao	DTR	Z-score	0,45	2,20	0,13	0,26	1,31	0,69	0,46	2,24	0,02
1	Duracao	RFR	Z-score	0,40	1,88	0,36	0,23	1,09	0,78	0,41	1,85	0,33
1	Duracao	PLS	Z-score	0,38	1,82	0,40	0,38	1,81	0,41	0,41	1,81	0,36
1	Duracao	ETR	Z-score	0,39	1,84	0,39	0,34	1,62	0,53	0,42	1,80	0,36
1	Duracao	LR	RobustScaler	0,38	1,82	0,40	0,38	1,81	0,41	0,41	1,81	0,36
1	Duracao	SVR	RobustScaler	0,36	1,84	0,38	0,33	1,78	0,42	0,38	1,82	0,35
1	Duracao	DTR	RobustScaler	0,45	2,19	0,13	0,26	1,31	0,69	0,46	2,24	0,02
1	Duracao	RFR	RobustScaler	0,40	1,88	0,36	0,23	1,08	0,79	0,40	1,84	0,34
1	Duracao	PLS	RobustScaler	0,38	1,82	0,40	0,38	1,81	0,41	0,41	1,81	0,36
1	Duracao	ETR	RobustScaler	0,39	1,85	0,38	0,33	1,58	0,55	0,40	1,81	0,35
2	Eficiencia	LR	Z-score	0,54	12363,58	0,46	0,51	12012,81	0,49	0,37	12034,85	0,46
2	Eficiencia	SVR	Z-score	0,52	12480,15	0,45	0,27	10261,10	0,63	0,32	11856,70	0,47
2	Eficiencia	DTR	Z-score	0,64	14284,89	0,28	0,29	8596,51	0,74	0,40	14321,26	0,23
2	Eficiencia	RFR	Z-score	0,46	12199,67	0,47	0,17	4486,89	0,93	0,37	11833,45	0,48
2	Eficiencia	PLS	Z-score	0,44	12418,17	0,45	0,42	12082,80	0,48	0,37	12057,49	0,46
2	Eficiencia	ETR	Z-score	0,45	12056,29	0,49	0,22	6373,26	0,86	0,37	11927,89	0,47
2	Eficiencia	LR	RobustScaler	0,54	12363,58	0,46	0,51	12012,81	0,49	0,37	12034,85	0,46
2	Eficiencia	SVR	RobustScaler	0,52	12406,63	0,46	0,28	10380,86	0,62	0,32	11803,77	0,48
2	Eficiencia	DTR	RobustScaler	0,64	14285,40	0,28	0,29	8596,51	0,74	0,40	14181,65	0,25
2	Eficiencia	RFR	RobustScaler	0,46	12177,48	0,48	0,17	4458,57	0,93	0,37	11847,84	0,47
2	Eficiencia	PLS	RobustScaler	0,44	12418,17	0,45	0,42	12082,80	0,48	0,37	12057,49	0,46
2	Eficiencia	ETR	RobustScaler	0,46	12053,94	0,49	0,22	6324,92	0,86	0,37	11906,48	0,47
2	Duracao	LR	Z-score	0,37	1,80	0,41	0,36	1,75	0,44	0,39	1,73	0,41
2	Duracao	SVR	Z-score	0,37	1,85	0,38	0,25	1,50	0,59	0,37	1,68	0,45
2	Duracao	DTR	Z-score	0,46	2,30	0,04	0,21	1,16	0,76	0,49	2,23	0,02
2	Duracao	RFR	Z-score	0,39	1,82	0,40	0,14	0,67	0,92	0,41	1,73	0,41
2	Duracao	PLS	Z-score	0,38	1,81	0,41	0,37	1,76	0,43	0,40	1,73	0,41

2	Duracao	ETR	Z-score	0,38	1,80	0,41	0,26	1,29	0,70	0,40	1,75	0,40
2	Duracao	LR	RobustScaler	0,37	1,80	0,41	0,36	1,75	0,44	0,39	1,73	0,41
2	Duracao	SVR	RobustScaler	0,37	1,85	0,38	0,25	1,48	0,60	0,37	1,68	0,44
2	Duracao	DTR	RobustScaler	0,46	2,30	0,04	0,21	1,16	0,76	0,49	2,25	0,01
2	Duracao	RFR	RobustScaler	0,39	1,82	0,40	0,14	0,67	0,92	0,42	1,74	0,41
2	Duracao	PLS	RobustScaler	0,38	1,81	0,41	0,37	1,76	0,43	0,40	1,73	0,41
2	Duracao	ETR	RobustScaler	0,38	1,80	0,41	0,26	1,30	0,70	0,40	1,74	0,40
3	Eficiencia	LR	Z-score	0,57	12344,61	0,46	0,54	11976,78	0,49	0,37	12041,67	0,46
3	Eficiencia	SVR	Z-score	0,55	12411,28	0,45	0,27	10236,98	0,63	0,32	11832,56	0,48
3	Eficiencia	DTR	Z-score	0,63	14100,86	0,30	0,29	8449,53	0,75	0,41	14414,59	0,22
3	Eficiencia	RFR	Z-score	0,53	12193,61	0,47	0,17	4409,17	0,93	0,37	11668,37	0,49
3	Eficiencia	PLS	Z-score	0,49	12374,68	0,46	0,46	12045,91	0,49	0,37	12008,50	0,46
3	Eficiencia	ETR	Z-score	0,46	12080,32	0,48	0,22	6395,28	0,86	0,37	11920,58	0,47
3	Eficiencia	LR	RobustScaler	0,57	12344,61	0,46	0,54	11976,78	0,49	0,37	12041,67	0,46
3	Eficiencia	SVR	RobustScaler	0,54	12364,08	0,46	0,28	10376,86	0,62	0,31	11787,00	0,48
3	Eficiencia	DTR	RobustScaler	0,63	14133,45	0,29	0,29	8449,53	0,75	0,40	14257,81	0,24
3	Eficiencia	RFR	RobustScaler	0,50	12181,74	0,47	0,17	4449,25	0,93	0,37	11784,02	0,48
3	Eficiencia	PLS	RobustScaler	0,49	12374,68	0,46	0,46	12045,91	0,49	0,37	12008,50	0,46
3	Eficiencia	ETR	RobustScaler	0,46	12013,17	0,49	0,26	7307,43	0,81	0,37	11884,97	0,47
3	Duracao	LR	Z-score	0,37	1,80	0,41	0,36	1,75	0,44	0,39	1,73	0,41
3	Duracao	SVR	Z-score	0,37	1,85	0,38	0,25	1,50	0,59	0,37	1,68	0,45
3	Duracao	DTR	Z-score	0,46	2,32	0,02	0,21	1,16	0,76	0,49	2,24	0,02
3	Duracao	RFR	Z-score	0,39	1,82	0,40	0,14	0,68	0,92	0,41	1,74	0,40
3	Duracao	PLS	Z-score	0,38	1,81	0,41	0,37	1,77	0,43	0,40	1,73	0,41
3	Duracao	ETR	Z-score	0,38	1,81	0,41	0,27	1,30	0,69	0,40	1,75	0,40
3	Duracao	LR	RobustScaler	0,37	1,80	0,41	0,36	1,75	0,44	0,39	1,73	0,41
3	Duracao	SVR	RobustScaler	0,37	1,85	0,38	0,25	1,48	0,60	0,37	1,68	0,45
3	Duracao	DTR	RobustScaler	0,46	2,32	0,02	0,21	1,16	0,76	0,49	2,25	0,01
3	Duracao	RFR	RobustScaler	0,39	1,82	0,40	0,14	0,67	0,92	0,41	1,73	0,41
3	Duracao	PLS	RobustScaler	0,38	1,81	0,41	0,37	1,77	0,43	0,40	1,73	0,41
3	Duracao	ETR	RobustScaler	0,39	1,81	0,41	0,00	0,00	1,00	0,41	1,73	0,41