



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA NAVAL

Predição do Consumo de Combustível em
Embarcações da Amazônia via *Machine Learning*

RHUAN FRACALLOSSI MELO

Orientação: Prof. Dr. Nélio Moura de Figueiredo

Belém

2023

RHUAN FRACALOSSI MELO

**Predição do Consumo de Combustível em
Embarcações da Amazônia via *Machine Learning***

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Naval – PPGENAV, do Instituto de Tecnologia – ITEC, da Universidade Federal do Pará – UFPA, como requisito para obtenção do título de Mestre em Engenharia Naval.

Linha de Pesquisa: Sistemas de Transporte Aquaviário.

Orientador: Prof. Dr. Nelio Moura de Figueiredo.

Orientação: **Prof. Dr. Nélio Moura de Figueiredo**

Belém

2023

RHUAN FRACALOSSI MELO

**Predição do Consumo de Combustível em Embarcações da Amazônia via
*Machine Learning***

Esta defesa final foi julgada e aprovada no Programa de Pós-Graduação em Engenharia Naval da Universidade Federal do Pará.

Belém, 30 de Junho de 2023

Prof. Dr. Said

(Coordenador do Programa de Pós-Graduação em Engenharia Naval – UFPA)

Banca Examinadora

Prof. Dr. Nélio Moura de Figueiredo
PPGNAV/UFPA

Orientador

Prof. Dr. Valcir João Farias da Cunha
PPGNAV/UFPA

Examinador Interno

Prof. Dr. Maisa Sales Gama Tobias
PPGENAV/UFPA

Examinador Interno

Prof. Dr. Lucélia Marques Lima da Rocha
SEDUC/PA

Examinador Externo

Dedico esse trabalho ao amor mais puro que
posso ter aqui nessa terra: minha família.
Mais do que para mim, é por vocês!

Agradecimentos

O filho de Paragominas cresceu. O encerramento de um ciclo significa, antes de tudo, que outro está para começar. Como um planejador nato, é inevitável olhar para os próximos passos e desejar um futuro de oportunidades. Entretanto, quando lembro da caminhada, é ainda mais difícil prosseguir sem antes ser grato a todos aqueles que estiveram comigo, alguns temporariamente, outros esporadicamente, mas, com toda certeza, cada um teve parte nas lições que foram lançadas e, com o passar dos anos, a gente descobre que, em verdade, sábio é aquele que faz das suas lutas um aprendizado diário e nunca deixa o sorriso no rosto se esvaecer.

Começo agradecendo ao Deus da minh'alma, sem O qual nada disso seria possível. Antes mesmo que eu viesse a viver o primeiro dos meus dias, todos eles já estavam escritos em Seu livro e eram caminhos de amor, paz e muita bondade. Nunca foi pela força do meu braço porque instrumento não tem importância se não houver quem saiba usá-lo, mas pelo Seu amor vivemos. Em segundo lugar, meu coração é de eterna gratidão por eles: José Aylton Melo e Cláudia Vitali, meus progenitores de quem tanto tenho orgulho e de quem faço minha motivação de todos os dias porque eu sei do sacrifício e esforço que tiveram durante todos esses anos da minha graduação. Mais do que por mim mesmo, saibam que faço isso por vocês.

Não menos importante, ao meu irmão Ramon Fracalossi, meu espelho, meu referencial que fez do nosso bem-estar seu maior sonho, provendo sempre tudo, não deixando faltar o que era essencial e segurando o mundo quando necessário era. A ajuda veio e, por isso, agradeço também à Mylla Christie por toda receptividade e carinho quando volto para casa e, principalmente, pela dedicação em cuidar daquele que se tornou nosso maior tesouro, meu sobrinho Lorenzo Fracalossi, minha maior saudade e exemplo de amor puro que Deus dá.

À minha parceira de vida, Emmanuelen Costa, que nunca me deixou desistir, mas sempre esteve me motivando com palavras de encorajamento e atos de dedicação em prol da realização desse sonho. Obrigado por ser meu exemplo de companheirismo e lealdade.

Aos meus mentores, a começar pelos professores e amigos Emmanuel Sant'Thiago, Pedro Lameira, Hito Braga e Nelio Figueiredo. A primeira oportunidade é aquela impossível de esquecer e foram de suas mãos que a minha veio. Obrigado por cada ensinamento, por toda instrução nos momentos delicados. Sem sombra de dúvidas, vocês

são parte importante de cada vitória que eu venha a conquistar a partir daqui.

Aos meus demais professores ao longo destes anos que, com tanta destreza, sempre souberam ensinar o técnico e, acima de tudo, a crescer independente das circunstâncias. Estendo esse agradecimento ao meu mestre de vida, Rubenson Santos, que sempre foi proativo em ouvir e tão complacente em ensinar.

À posteriori, sou grato ao Walter Júnior e a Andreza Matos, a quem considero pai e mãe de coração, que sempre souberam expor sua preocupação em nosso desenvolvimento de vida e bem-estar, assegurando que todo suporte fosse nos dado. Ainda, ao Walmir Matos e Walter Neto por serem irmãos de coração, cujo entusiasmo e cuidado para/comigo nunca deixaram faltar.

Obrigado, Naval, você é a realização dos planos de Deus pra minha vida, um sonho que cheguei a duvidar que se realizaria, mas que, no tempo oportuno, aconteceu. Vamos avante!

“Pede, e Eu te darei as nações como herança,
os confins da terra como tua propriedade.”

Salmos 2:8

Lista de Tabelas

1	Registros de embarcações do banco de dados utilizado.	40
2	Registros de embarcações do banco de dados utilizado.	41
3	Registros de embarcações do banco de dados utilizado.	43
4	Aplicação da técnica One-Hot para criação de novos parâmetros numéricos.	45
5	Distribuição dos registros após fase de processamento.	73
6	Métricas de desempenho dos seis modelos treinados por meio do Método de Validação Cruzada com 10 rodadas.	84
7	Hiperparâmetros do modelo após fase de otimização.	85
8	Métricas de desempenho do modelo vencedor, CatBoost.	86
9	Registros de embarcações do banco de dados utilizado.	88
10	Métricas de desempenho do modelo vencedor, CatBoost.	89

Lista de Figuras

1	Carga transportada (em TKU) por modo de transporte.	9
2	Emissão de dióxido de carbono (CO ₂) a partir do consumo de combustível pelos diferentes setores em 2014.	18
3	Fatores que influenciam o consumo de combustível das embarcações.	20
4	Pipeline do processo de construção do modelo.	28
5	Fluxograma de construção do modelo proposto.	36
6	Mapa do fluxo de passageiros total na região amazônica em 2017.	38
7	Evolução da estimativa do transporte de passageiros e cargas nos levantamentos realizados pela UFPA – 2011 a 2017.	39
8	Distribuição dos dados para treinamento, teste e validação hold-out.	51
9	Distribuição das embarcações de acordo com o regime de navegação.	65
10	Distribuição das embarcações conforme material de construção do casco.	66
11	Relação entre o material de construção do casco e a categoria de navegação da embarcação.	68
12	Distribuição das embarcações amazônicas em função do consumo de combustível.	69
13	Relação consumo de combustível versus duração da viagem e potência do motor.	70
14	Representação em box-plot dos parâmetros de entrada do modelo.	74
15	Matriz de Correlação de Pearson entre os parâmetros do modelo.	75
16	Ranking de importância dos parâmetros de entrada utilizados no modelo.	78
17	Representação de importância dos preditores através do método SHAP.	81
18	Curva de residuais entre os valores preditos e observados.	90
19	Curva de performance aderente à condição ótima do modelo.	91

Lista de Siglas

AD	Árvore de Decisão
AE	Árvores Extras
AEXG	Aumento Extremo de Gradiente
AG	Aumento de Gradiente
ANTAQ	Agência Nacional de Transportes Aquaviários
CB	Catboost
CFD	Computational Fluid Dynamics
CNT	Confederação Nacional de Transporte
DCBD	Método de Descoberta de Conhecimento em Banco de Dados
EAM	Erro Absoluto Médio
ECTFA	Estudo de Caracterização da Oferta e da Demanda do Transporte Fluvial de Passageiros e Cargas na Região Amazônica
EQM	Erro Quadrático Médio
FA	Florestas Aleatórias
OMI	Organização Marítima Internacional
KDD	<i>Knowledge Discovery in Databases</i>
LASSO	<i>Least Absolute Shrinkage and Selection Operator</i>
ML	<i>Machine Learning</i>
PNLT	Plano Nacional de Logística e Transportes
QME	Quadrado Médio do Erro
R^2	Coefficiente de Determinação
RNA	Redes Neurais Artificiais

SAD	Sistema de Apoio à Decisão
SIA	Sistema de Identificação Automática
SOBENA	Sociedade Brasileira de Engenharia Naval
TKU	Toneladas por Quilômetro Útil
TPB	Tonelagem Porte Bruto
UFPA	Universidade Federal do Pará

Sumário

1	INTRODUÇÃO	1
1.1	Justificativa	5
1.2	Objetivos	6
1.2.1	Objetivo geral	6
1.2.2	Objetivos específicos	6
1.3	Estrutura do Trabalho	7
2	REFERENCIAL TEÓRICO	8
2.1	A Matriz de Transporte Nacional e a Realidade do Modal Hidroviário para a Região Amazônica	8
2.2	Consumo de Combustível de Embarcações	16
2.2.1	Parâmetros de Projeto da Embarcação	19
2.2.2	Desempenho Operacional	23
2.2.3	Condições Ambientais	25
2.3	Método de Descoberta de Conhecimento em Banco de Dados (DCBC)	27
2.3.1	<i>Machine learning</i> Aplicado à Estimativa e Otimização de Consumo de Combustível na Navegação	29
3	METODOLOGIA	34
3.1	Aquisição de Dados	37
3.1.1	Coleta de Dados	37
3.1.2	Banco de Dados	39
3.2	Tratamento de Dados	44
3.2.1	Tratamento de Valores Desconhecidos	44
3.2.2	Transformação de Variáveis Categóricas	45
3.2.3	Remoção de Outliers	45
3.2.4	Normalização	46
3.3	Seleção de Atributos	47
3.3.1	Método de Correlação de Pearson	47
3.3.2	Método de Classificação de Atributos via F-Score	49
3.3.3	Método SHAP	50

3.4	Seleção de Algoritmos e Construção dos Modelos	51
3.4.1	Árvore de Decisão de Regressão (ACR)	52
3.4.2	Método de Conjunto	54
3.5	Otimização de Hiperparâmetros	59
3.6	Aplicação de Métricas de Desempenho	60
3.6.1	Erro Médio Absoluto (EMA)	61
3.6.2	Quadrado Médio do Erro (QME)	61
3.6.3	Erro Quadrático Médio (EQM)	62
3.6.4	Coefficiente de Determinação (R^2)	62
3.7	Validação do Modelo	64
4	RESULTADOS	64
4.1	Categorização dos Padrões	64
4.1.1	Estratificação do Universo de Embarcações Proposto	71
4.2	Seleção de Atributos	74
4.2.1	Matriz de Correlação via Método de Coeficiente de Pearson	74
4.2.2	Ranking de Importância dos Parâmetros via Aplicação de Método F-Score	77
4.2.3	Correlação de Atributos de Entrada em Função do Consumo de Combustível via Método SHAP	80
4.3	Treinamento e Avaliação dos Modelos	83
4.4	Otimização de Hiperparâmetros do Modelo Escolhido	85
4.5	Validação Holdout do Modelo Final	87
5	CONCLUSÃO	92
	REFERÊNCIAS	96
	APÊNDICE A - REGISTROS UTILIZADOS PARA VALIDAÇÃO	115

Resumo

Estima-se que cerca de dois terços dos custos operacionais estão relacionados ao consumo de combustível das embarcações no modal aquaviário, uma parcela significativa do ponto de vista econômico, uma vez que impacta a viabilidade das operações e eleva o custo dos serviços prestados por empresas e organizações que atuam neste setor. Sob tal motivação, este trabalho tem como principal objetivo desenvolver um modelo de predição analítica do consumo de combustível de embarcações operantes na região amazônica via machine learning, enfatizando os atributos que mais impactam a variável-alvo, se físicos, construtivos ou operacionais e a relação existente entre eles. Primeiramente, ocorreu o levantamento do banco de dados e pré-processamento dos registros, estabelecendo um padrão geral ao qual devem estar ajustados, de modo a garantir o desempenho ótimo da análise dos dados. O uso dos Algoritmos de Importância dos Atributos, teve, como finalidade, elencar os parâmetros de maior influência sobre o consumo de combustível, proporcionando também uma análise local dos registros acerca de cada atributo e a correlação das variáveis entre si. À posteriori, foram aplicados seis algoritmos de machine learning para extração dos padrões e desenvolvimento do modelo vencedor, avaliado de acordo com métricas de desempenho como Erro Absoluto Médio, Quadrado Médio do Erro, Erro Quadrático Médio e Coeficiente de Determinação. Por fim, foi possível investigar as principais características físicas, construtivas e operacionais das embarcações na região amazônica, como comprimento, boca, pontal, entre outras, e comprovar a eficácia dos modelos de aprendizado de máquina. O modelo preditivo desenvolvido com o algoritmo CatBoost obteve um coeficiente de determinação de 91%, fornecendo previsões satisfatórias para o consumo de combustível em embarcações conhecidas.

PALAVRAS-CHAVE: Amazônia, modelo de seleção de atributos, *machine learning*, embarcação, consumo de combustível, predição.

Abstract

It is estimated that approximately two-thirds of operational costs are related to the fuel consumption of vessels in the waterway transportation mode, a significant portion from an economic standpoint. This impacts the viability of operations and increases the cost of services provided by companies and organizations operating in this sector. With this motivation, this study's main objective is to develop an analytical predictive model for fuel consumption in vessels operating in the Amazon region using machine learning. It emphasizes the attributes that most impact the target variable, whether they are physical, structural, or operational, and the existing relationships between them. The process began with data collection and preprocessing, establishing a general standard to ensure optimal data analysis performance. Attribute Importance Algorithms were used to identify the parameters with the greatest influence on fuel consumption, also providing a local analysis of the records for each attribute and the correlation between variables. Subsequently, six machine learning algorithms were applied to extract patterns and develop the winning model, evaluated according to performance metrics such as Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, and Coefficient of Determination. Finally, it was possible to investigate the main physical, structural, and operational characteristics of vessels in the Amazon region, such as length, width, draft, among others, and confirm the effectiveness of machine learning models. The predictive model developed using the CatBoost algorithm achieved a coefficient of determination of 91%, providing satisfactory predictions for fuel consumption in known vessels.

KEYWORDS: Amazon, attribute selection model, machine learning, vessel, fuel consumption, prediction.

1 INTRODUÇÃO

Para a região amazônica, mesmo com a falta de incentivos e ineficiência do apoio público frente aos subsídios injetados para uma melhor infraestrutura aquaviária, seja sob a ótica da melhor qualidade de serviço prestado ao passageiro ou até mesmo quanto aos aportes financeiros aos pequenos e grandes armadores, segundo a Agência Nacional de Transportes Aquaviários (ANTAQ) por meio de registros mais recentes levantados no ano de 2017, são em torno de 9,8 milhões de passageiros e 3,4 milhões de toneladas de carga movimentados, representando uma injeção de recursos à economia de quase 1 bilhão de reais. A tendência é que esse número cresça ao passo que medidas de inclusão social também sejam implantadas, uma vez que o aumento do bem-estar da sociedade e a contribuição para a geração de renda e de emprego provoca também o maior trânsito de pessoas e trocas comerciais entre as regiões (JUNIOR *et al.*, 2018).

Vale ressaltar ainda que o transporte hidroviário prestado na Amazônia tem a seu favor o fato de ser o único meio possível para desembarque em algumas cidades no interior da região onde algumas viagens chegam a ter duração entre sete a dez dias, como é o caso do trecho Belém-Manaus (FIGUEIREDO *et al.*, 2018). Entretanto, mesmo sob a perspectiva de maior movimentação para os anos vindouros, assim como da necessidade de concepção de novas embarcações, a construção naval e hidroviária regional ainda segue a passos lentos e fundamentada no empirismo, cujos pesos de qualidade e concepção têm sido minimizados, ao ponto de serem, via de regra, consideradas padronizadas, “de prateleira” (PADOVEZI, 2003).

Como consequência, não há sequer um conhecimento acerca dos parâmetros que verdadeiramente influenciam o custo operacional da embarcação na região amazônica, realidade que, caso fosse alterada, certamente conduziria a novas perspectivas de redução de tais gastos motivadas por iniciativas inovadoras em que há a inserção de tecnologias para mapeamento e análise destes dados.

Para Avgouleas & Sclavounos (2014), a indústria utiliza 5% do consumo global de petróleo bruto como combustível no transporte marítimo. Kim e Roh (2020) corroboram que este montante equivale a quatro milhões de barris de combustível por dia, motivando as empresas de navegação a otimizar as operações do navio para alcançar uma eficiência energética eficaz e estratégica de navegação. Como alternativa, estudos têm sido

desenvolvidos para investigação da relação existente entre os fatores interferentes no consumo de combustível de um navio, estejam estes relacionados às características de projeto da embarcação e/ou ao desempenho operacional e/ou às condições ambientais do trajeto. Aliado ao conhecimento acerca dos atributos de maior impacto, o desenvolvimento de um modelo preditivo capaz de determinar analiticamente como dá-se o comportamento da curva de consumo de combustível de uma embarcação sob tais características e condições contribui para a otimização e respaldo dos projetos construtivos até então empíricos propiciados por grande parte dos armadores nortistas.

Wickramanayake e Bandara (2016) propuseram um comparativo entre três algoritmos distintos - Florestas Aleatórias, Redes Neurais Artificiais e Aumento de Gradiente - para predição do consumo de combustível em ônibus público em Sri Lanka para dados parâmetros de entrada como longitude, latitude, elevação, distância percorrida, velocidade, entre outros. Baseado nesta análise, a pesquisa concluiu que a técnica de Florestas Aleatórias gerou predições mais precisas com valores de QME, EAM e EQM como métricas de desempenho do modelo.

Karagiannidis *et al.* (2019) desenvolveram um modelo preditivo de consumo de combustível de embarcações por meio de aplicação de Redes Neurais Artificiais (RNA) utilizando dados coletados durante 1,5 anos de um navio de transporte de contêiner. O modelo final foi avaliado por meio de métricas de desempenho como Erro Quadrático Médio (EQM) e Coeficiente de Determinação (R^2), cuja avaliação definiu a boa precisão do algoritmo em prever o consumo de combustível da embarcação estudada.

Gkerekos *et al.* (2019) elaboraram um comparativo entre dez algoritmos de múltipla regressão projetados para prever o consumo de combustível de motores de combustão principal de navios através de parâmetros operacionais, como distância percorrida, calado, velocidade da embarcação, entre outros dados constantes no relatório de meio-dia (noon data report) enviados pela tripulação diariamente. O modelo de Árvores Extras obteve o melhor desempenho avaliado por meio de métricas de desempenho como Erro Médio Absoluto (EAM), Quadrado Médio do Erro (QME) e Coeficiente de Determinação (R^2). O estudo comprovou que a metodologia empregada pode ser aplicada a qualquer outra embarcação com perfil operacional semelhante.

Uyanik *et al.* (2020) promoveram vários modelos de predição, como Regressão Linear Múltipla, Regressão de Ridge e LASSO, Regressão de Vetor de Suporte, Algoritmos

Baseados em Árvores de Decisão, Algoritmos de Aumento, considerando dados coletados de um navio porta-contêineres, cujos parâmetros estavam relacionados à potência e consumo do motor de combustão principal, temperatura dos rolamentos, fluxo de massa de combustível, entre outros. A estimativa mais próxima dos dados reais de consumo de combustível foi obtida por regressão linear múltipla com Erro Quadrático Médio (EQM), Erro Absoluto Médio (EAM) e Coeficiente de Determinação (R^2) como métricas de avaliação do modelo.

Stepec *et al.* (2020) sugeriram um sistema automatizado para estimação do tempo de estadia de um navio em porto aplicando uma técnica de *machine learning* denominada Catboost, que utiliza o aumento de gradiente como fundamento. Os dados de entrada utilizados foram extraídos de relatórios de chamada gerados no Porto de Bordeaux na França ao longo de um período histórico de 11 anos, entre 2008 e 2018, contendo dados operacionais como dimensões principais dos navios, quantidade e tipo de carga carregada e descarregada no porto, berços de atracação e horas de chegada e saída do navio. Erro Absoluto Médio (EAM) e Erro Quadrático Médio (EQM) foram adotados como métricas de desempenho.

Abebe *et al.* (2020) pretenderam prever a velocidade de cruzeiro de navios por meio de uma abordagem baseada em algoritmos de *machine learning*, como Árvore de Decisão, Florestas Aleatórias, Árvores Extras, Aumento de Gradiente e Aumento Extremo de Gradiente. Foram utilizados dados provenientes do Sistema de Identificação Automática (SIA) e dos relatórios marítimos climáticos de meio-dia, capazes de prover as variáveis de entrada, como calado máximo, latitude, longitude, velocidade de cruzeiro, altura e direção de onda, altura e direção de vento, velocidade de corrente, temperatura ambiente, entre outras. Quanto às métricas de desempenho, foram avaliados o Coeficiente de Determinação (R^2) e o Erro Quadrático Médio (EQM), apontando um melhor desempenho do modelo de Árvores Extras em relação aos demais.

Hu *et al.* (2021) elaboraram um modelo híbrido de previsão de consumo de combustível baseado em dados de sensores coletados de um navio porta-contêineres oceânico. Um modelo de previsão de consumo de combustível híbrido foi estabelecido integrando métodos de Floresta Aleatória (RF), Aumento Extremo de Gradiente (AEG) e Regressão Linear Múltipla (RLM). O modelo proposto desempenhou um papel significativo no monitoramento em tempo real do consumo de combustível do navio, economia de energia

e redução de emissões, etc.

Okumus *et al.* (2021) propuseram o desenvolvimento de um modelo para predição de potência do motor principal e auxiliar de novos navios por meio de técnicas de *machine learning* como Regressão Linear Simples (RLS), Regressão Polinomial (RP) e Aumento Extremo de Gradiente (AEG), cujas métricas de desempenho foram Quadrado Médio do Erro (QME) e Erro Quadrático Médio (EQM).

A despeito da variedade de projetos construtivos em execução na região amazônica, persiste uma lacuna no atual cenário científico, caracterizada pela ausência de estudos que possam informar o desenvolvimento de um modelo capaz de estabelecer correlações entre atributos cruciais das embarcações operando na região, como comprimento, boca, pontal, entre outros, e seu consumo de combustível por viagem. É notável que, de acordo com a literatura, o consumo de combustível representa mais de um quarto dos custos operacionais totais de uma embarcação, um componente significativo quando considerado na perspectiva da matriz de custos e do balanço econômico-financeiro das empresas inseridas na cadeia de transporte e serviços.

O desenvolvimento de um modelo possibilita um planejamento mais eficiente por parte dos armadores, contribuindo para a mitigação das emissões de gases poluentes na atmosfera por meio da otimização dos parâmetros operacionais das embarcações. Além disso, ofereceria uma base sólida para a tomada de decisões na construção e operação de embarcações, beneficiando tanto os habitantes locais quanto as empresas de navegação, especialmente no que tange aos aspectos econômicos, resultando em uma maior congruência com os objetivos de planejamento estabelecidos.

Este projeto tem por propósito, portanto, desenvolver um modelo preditivo de consumo de combustível através da análise do banco de dados de embarcações operantes na Amazônia por meio de seis algoritmos de *machine learning* aplicados a partir da biblioteca *scikit-learn* via linguagem de programação *Python 3.6*, sendo: Árvore de Decisão (AD), Florestas Aleatórias (FA), Árvores Extras (AE), Aumento de Gradiente (AG), Aumento Extremo de Gradiente (AEXG), CatBoost (CB) - de modo a apontar quais são os parâmetros de maior importância relacionados ao consumo de combustível, predizendo numericamente o volume consumido, dadas as condições físicas, construtivas e operacionais da embarcação. Através do estudo dos padrões apontados acerca pelo modelo, é possível determinar se uma embarcação proposta, esteja em construção ou até

mesmo ainda em fase de projeto, está em concordância e operará sob regime satisfatório e comum às demais, evitando desempenhos inesperados e custos desnecessários ao armador posteriormente. Como consequência, respaldo técnico aprimorado gerando meios mais econômicos para a operação do transporte aquaviário na região, principalmente, quanto ao consumo de combustível.

1.1 Justificativa

A maioria das linhas de transporte aquaviário da Amazônia ainda são servidas por embarcações de tecnologia ultrapassada, construídas em madeira ou aço pelo simples processo de construção, pelo custo do material relativamente baixo e pelo modo artesanal como elas são construídas e recuperadas. Em geral, o empirismo é o principal fundamento na concepção destas embarcações, no qual as empresas de navegação fazem uso de projetos semelhantes já desenvolvidos, sem que haja uma análise prévia de compatibilidade ou busca por melhorias e inovações construtivas ou operacionais.

Explica-se que, apesar da alta gama de estudos desenvolvidos na região, ainda assim não há, por exemplo, um modelo ou sistema qualquer que seja capaz de expor a correlação existente entre as principais características dos navios operantes, como comprimento, boca, pontal, entre outras, em relação ao volume de combustível necessário por viagem, visando seu uso de forma racional, sustentável e econômica. Algumas literaturas, inclusive, apontam que o consumo de combustível representa cerca de mais de um quarto do total dos gastos operacionais de um navio. Por esta razão, busca-se uma comprovação científica do comportamento referente aos custos desempenhados pelas embarcações operantes na região amazônica fundamentada através da aplicação de um método de seleção de parâmetros renomado, tendo em vista sua ampla e eficiente utilização frente aos mais diversos segmentos, como: mercado financeiro, medicina, biologia, etc.

Por conseguinte, entendida a favorável extensão de malha hidroviária somada à significativa movimentação regional de passageiros e cargas, conhecer os parâmetros de maior relevância relacionados ao consumo de combustível das embarcações contribuiria para um planejamento mais centrado na economicidade por parte dos armadores, o que certamente traria impacto também quanto à redução da emissão de gases poluentes na atmosfera devido à adoção dos parâmetros ótimos para operação da embarcação. Aliás, o próprio nativo e as empresas de navegação estariam melhor respaldados na construção

e uso das embarcações quanto aos quesitos econômicos, gerando expectativas melhor enquadradas no planejamento.

Em vista da carência exposta, julga-se de extrema importância um estudo capaz de determinar quais são, de fato, os parâmetros físicos e operacionais de maior impacto sobre o volume de combustível consumido segundo o espaço amostral de embarcações presentes na região amazônica, capaz ainda de prever numericamente, a partir destes parâmetros físicos, construtivos e operacionais iniciais, o consumo gerado pela embarcação, contribuindo, dessa maneira, para a elaboração de novos projetos que atendam não só os anseios financeiros dos pequenos e grandes armadores, mas também a lacuna científica sobre a relação de cada propriedade com o consumo de combustível sob o cerne regional, além de apontar padrões construtivos e operacionais intrínsecos e permitir o ajuste desses parâmetros quando em uma análise preliminar de projeto.

1.2 Objetivos

1.2.1 Objetivo geral

Desenvolver um modelo para predição do consumo de combustível dada a representatividade do mesmo na estrutura de custos operacionais a partir das características físicas, propulsivas e operacionais de embarcações navegantes na região amazônica via *Machine Learning*.

1.2.2 Objetivos específicos

1. Levantar as propriedades físicas, propulsivas e operacionais de embarcações que navegam na região amazônica para o criação do banco de dados.
2. Desenvolver um modelo para tratamento dos dados levantados e aplicação de algoritmos de *machine learning* como Árvore de Decisão (AD), Florestas Aleatórias (FA), Árvores Extras (AE), Aumento de Gradiente (AG), Aumento Extremo de Gradiente (AEXG), CatBoost (CB), sendo seus desempenhos avaliados por meio de métricas como EMA, QME, EMQ e R^2 , para obtenção do modelo vencedor frente às predições referentes à variável-alvo, o consumo de combustível.
3. Validar o modelo vencedor por meio de novos registros de embarcações levantadas.

4. Estratificar os grupos de embarcações existentes na região através de análises inerentes aos atributos correlacionados e de maior influência para o modelo, clarificando padrões empregados na construção destas embarcações analisadas..

1.3 Estrutura do Trabalho

A seguinte seção evidencia a estrutura da dissertação, com destaque para a divisão dos capítulos, esboçando-a abaixo.

Já exposto, o Capítulo 1 faz a apresentação do trabalho através da introdução, objetivos geral e específicos, assim como sua estrutura e fluxograma. Retratando ao leitor a realidade a qual o transporte aquaviário está inserido na Amazônia atual, fundamentando sua importância para os nativos da região, além de propor um modelo inovador que contribua com um conhecimento científico cujas consequências trazem consigo impactos sociais, econômicos e ambientais tanto para prestadores do serviço, ribeirinhos e empresas de navegação, quanto para os passageiros.

O Capítulo 2 destaca-se pela revisão bibliográfica, buscando contextualizar as temáticas norteadoras do presente trabalho, como a atual matriz logística de transporte brasileira, enfatizando o panorama ao qual está enquadrado o modal hidroviário nacional, bem como fazendo alusão à maneira como o transporte misto tem sido pautado sob a ótica regional, principalmente no que diz respeito à problemática do consumo de combustível, a partir da qual foram estudados os parâmetros de maior influência, sejam físicos, operacionais ou ambientais, além de possíveis soluções voltadas à sua estimativa e otimização através da adoção de modelos de análise de dados e *machine learning* pelas empresas de navegação mundiais, metodologia que será aplicada ao banco de dados no projeto proposto.

O Capítulo 3 apresenta o detalhamento da metodologia utilizada na dissertação, com foco em sete aspectos principais: aquisição do banco de dados para desenvolvimento do modelo; a preparação e limpeza dos registros, mantendo um padrão geral dos mesmos; a seleção e extração dos atributos de maior relevância através da matriz de correlação de Pearson, método de Importância de Atributo via Aumento Extremo de Gradiente (AEXG) e método de Importância de Atributo via SHAP; aplicação de seis algoritmos de *machine learning* para desenvolvimento do modelo: Árvore de Decisão (AD), Florestas Aleatórias (FA), Árvores Extras (AE), Aumento de Gradiente (AG), Aumento Extremo de Gradiente

(AEXG), CatBoost (CB); otimização dos hiperparâmetros; avaliação do desempenho dos modelos através de métricas de desempenho: EAM, QME, EQM e R^2 ; validação hold-out do modelo escolhido.

O Capítulo 4 apresenta os resultados, com a compilação das informações, dados e parâmetros obtidos nos tópicos anteriores alcançados através da aplicação do modelo proposto e desenvolvido na metodologia tomando como base as informações e características das embarcações operantes na Amazônia conforme apresentado pela Agência Nacional de Transporte Aquaviário (ANTAQ).

O Capítulo 5 apresenta as considerações finais do trabalho, assim como identifica e recomenda o aprofundamento desse estudo, visando o aperfeiçoamento do conhecimento gerado até o presente momento.

O Capítulo 6 informa as referências bibliográficas.

2 REFERENCIAL TEÓRICO

2.1 A Matriz de Transporte Nacional e a Realidade do Modal Hidroviário para a Região Amazônica

A atividade de transporte baseia-se no movimento de bens e pessoas entre as diferentes origens e destinos, onde sua economia tem por propósito a alocação de recursos para viabilizar a movimentação de cargas e passageiros de um lugar para outro (SENNA, 2014). Inevitavelmente, de acordo com Ferreira (2013), todos os setores econômicos apresentarão dependência, seja mais ou menos direta, da funcionalidade e eficiência geradas pelo setor de transportes.

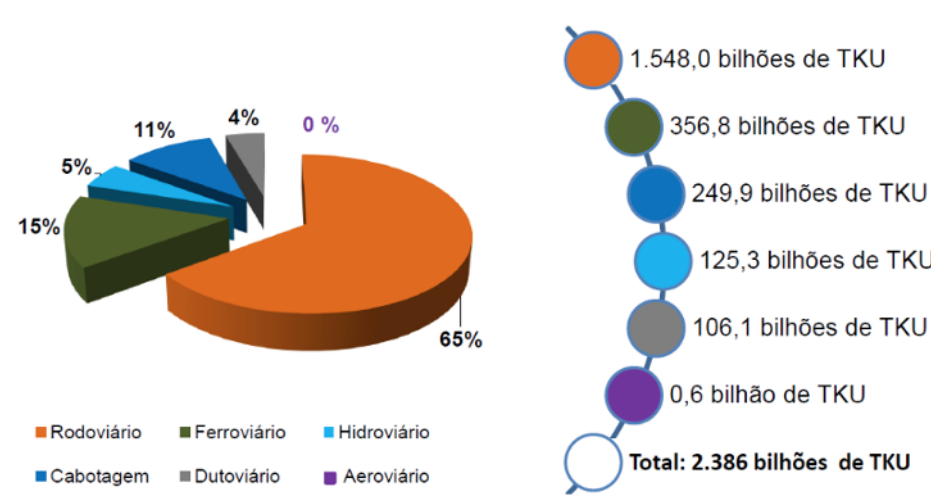
Sendo assim, é justificável afirmar que o crescimento econômico de uma nação é proporcional “às facilidades de mobilidade e acessibilidade de sua população em termos de deslocamento urbano, entre suas regiões e países, bem como o escoamento de sua produção de mercadorias de maneira que cheguem aos pontos de consumo” (SOARES, 2019).

Entendida a importância do transporte para o avanço econômico e também para a estabilidade financeira de um país, julga-se, à luz da atual matriz de transporte brasileira, haver um desequilíbrio entre os diversos modais, considerando o alto grau de dependência

em relação ao transporte rodoviário, fato que pode ser considerado preocupante tendo em vista as características negativas deste modal que, segundo a Confederação Nacional de Transporte (2013), são: a baixa eficiência energética, o elevado custo para longa distância e, por fim, os altos índices de acidentes e de roubo/avaria de cargas.

Segundo a Empresa de Planejamento e Logística (EPL), em 2015, sob a ótica da Tonelada por Quilômetro Útil (TKU), está representado na Figura 1 a participação de cada modal de transporte na matriz brasileira, estando o rodoviário em primeiro lugar com mais de 1500 bilhões de TKU, algo em torno de 64,88% de toda a movimentação provida em 2015, seguido pelos modais ferroviário (356,8 bilhões de TKU, 14,95%), cabotagem (249,9 bilhões de TKU, 10,47%), hidroviário (125,3 bilhões de TKU, 5,25%), dutoviário (106,1 bilhões de TKU, 4,42%) e aeroviário (0,6 bilhão de TKU, 0,03%). Corroborase, portanto, que, mesmo frente às deficiências julgadas anteriormente para este tipo de transporte, ocorre a predominância errônea de uma logística rodoviarista no país.

Figura 1: Carga transportada (em TKU) por modo de transporte.



Fonte: Autor (2023).

Outro ponto que o gráfico pretérito expõe diz respeito à baixa utilização do modal aquaviário para transporte de carga, reforçando o que foi pronunciado por Oliveira (2020) que coloca o Brasil partícipe de um grupo de países com o menor índice de aproveitamento de suas vias potencialmente navegáveis. Atualmente, segundo a Agência Nacional de Transportes Aquaviários (ANTAQ, 2015), o Brasil conta com uma rede hidroviária economicamente navegada de aproximadamente 22.037 quilômetros, tendo

como suas principais hidrovias a bacia Amazônica (17.651 km), Tocantins-Araguaia (1.360 km), Paraná-Tietê (1.359 km), Paraguai (591 km), São Francisco (576 km), Sul (500 km). Embora corresponda a um número expressivo, vale ressaltar que, de acordo com a CNT (2013), se somados todos os trechos de vias navegáveis e os potencialmente navegáveis durante os períodos das cheias, a rede hidroviária do país pode ultrapassar 40.000 quilômetros de extensão.

Com isso, conclui-se que apenas 52% do potencial navegável do país é utilizado para o transporte de cargas ou passageiros, enfatizando ainda que 80% das hidrovias estão na região amazônica. Embora tais características espaciais e geográficas proporcionem uma situação favorável para o desenvolvimento do modal aquaviário, este meio de transporte, frente aos dados apresentados, não tem sido bem aproveitado (CNT, 2018). O Brasil, nação de grande potencial hídrico, está desenvolvendo suas hidrovias mais tarde do que aqueles países europeus e norte-americanos que utilizam, significativamente, a navegação fluvial como meio de transporte efetivo de cargas como granéis e contêineres (PADOVEZI, 2003).

A potencialidade de navegação característica de cada hidrovia advém de diversos fatores, cuja combinação pode limitar ou até mesmo ampliar as condições de navegabilidade, bem como determinar as embarcações-tipo operantes do trecho, no que tange à sua dimensão e à sua capacidade de transporte (MORGADO *et al.*, 2013). A geografia do território brasileiro, por exemplo, é uma condição limitante das rotas hidroviárias, onde, conseqüentemente, é essencial uma maior flexibilidade das rotas através da integração com outros modais terrestres, diante disso, o rodoviário para distâncias curtas e o ferroviário para média distância (NEIVA, 2019).

Ainda assim, a ampliação do modal hidroviário brasileiro tenderá a contribuir com uma maior competitividade às indústrias e aos produtores, além de um menor desperdício da produção, haja vista ser mais seguro e menos poluente, trazendo eficiência ao transporte e contribuindo para um meio ambiente sustentável (OLIVEIRA, 2020).

Denomina-se transporte hidroviário um tipo de transporte aquaviário realizado por meio de hidrovias cujo intuito permanece sendo o transporte de pessoas e mercadorias por meio de rios, lagos, lagoas navegáveis e quaisquer outros, entretanto, que sejam naturalmente navegáveis ou tenham adquirido a condição de navegabilidade mediante intervenções de dragagem, sinalização, balizamento, implantação de sistemas de controle

operacional, etc (CUNHA & INSPER JR., 2020).

Segundo a ANTAQ (2013), no que tange ao transporte aquaviário de cargas no Brasil, estão presentes algumas particularidades como: grande capacidade de carga, sendo possível o deslocamento de produtos de maior tamanho ou em maior quantidade por distâncias até mesmo intercontinentais; baixo custo de manutenção; menor poluição por tonelada de mercadoria transportada em comparação aos demais modais; custo inferior da tonelada por quilômetro. Além disso, não há perdas relacionadas ao desgaste da via, assim como uma menor utilização de combustíveis, contribuindo para um panorama menos nocivo para o meio ambiente tendo em vista a redução dos índices de emissão de poluentes (CUNHA E INSPER JR., 2020).

Não obstante as características apresentadas anteriormente relacionadas ao transporte aquaviário, ressalta-se que o território brasileiro possui condições geográficas extremamente favoráveis para a prática deste modal, o que maximiza ainda mais as vantagens de seu uso. Entretanto, para que isto aconteça de fato, é necessário renunciar a priorização do rodoviário em detrimento aos demais modais. A hidrovia, por exemplo, é o modelo menos oneroso se comparado a qualquer outro modal, sendo cerca de 20 vezes mais barato que o rodoviário (OLIVEIRA, 2020). Aliás, segundo levantamentos do Ministério dos Transportes, através do PNLT (2011), os fretes hidroviários podem ser 62% mais baratos que os rodoviários.

Em contrapartida, quanto às desvantagens, apresenta pouca flexibilidade da carga, baixa velocidade de transporte, maior tempo gasto para carregamento e descarregamento, distância entre os portos e os centros de fabricação, sendo sua utilização adequada para grandes volumes de carga a serem transportados por longas distâncias (CUNHA e INSPER JR., 2020).

Dentre os tipos de navegação relacionados ao transporte aquaviário, vale destacar que, além da participação no transporte de carga, a navegação interior é aquela de importante papel no transporte de passageiros, principalmente, na Amazônia, cuja movimentação estimada, em 2017, ronda um pouco menos de dez milhões de passageiros, transportados seja pelas linhas estaduais ou interestaduais nas vias interiores dessa região (ANTAQ, 2017). O principal motivo dá-se devido os rios viabilizarem os deslocamentos e o abastecimento de grande parte da população, sendo, para muitas localidades, o único meio de conexão com os demais lugares. A nível de comércio internacional, o Brasil

tem nesta modalidade do transporte aquaviário o responsável por 98,7% das exportações e 91,3% das importações totais do país, sendo números bem expressivos (CUNHA & INSPER JR., 2020).

A Amazônia brasileira possui cerca de 24 mil quilômetros de vias fluviais, das quais aproximadamente 15 mil quilômetros são navegáveis, com profundidades superiores a um metro durante 90% do ano (PEIXOTO, 2006). Portanto, há uma grande rede natural apta ao transporte fluvial, que se estende por toda a região, e correspondente ao tamanho expressivo de mais de 50.000 quilômetros de trechos navegáveis, sendo extremamente bem drenada (BRASIL, 2006).

Elucida-se que a colonização no Brasil teve início no litoral e, com o tempo, se desenvolveu para o interior do país. Tal fenômeno, chamado de interiorização, trouxe a necessidade de utilizar meios de transporte que deslocassem mercadorias e pessoas aos portos e centros urbanos até então localizados no litoral. Conforme aponta Neto *et al.* (2014), durante esse período, as hidrovias sofreram concorrência do transporte ferroviário, mas, em muitas ocasiões, foi o modal escolhido em diversas rotas, desencadeando o início da reorientação das políticas de transportes no Brasil baseada no transporte rodoviário.

Pertinente ao cenário, o setor de transportes, baseado fundamentalmente na modalidade hidroviária, assume um papel de relevo, seja com o propósito de atender ao fluxo de pessoas e cargas ou de tornar sua infraestrutura capaz de influenciar o desenvolvimento econômico e social da região (MORGADO *et al.*, 2013). Ademais, essas condições naturais, especialmente a grande extensão de vias navegáveis, fazem com que o transporte hidroviário seja tomado como ponto de referência básico, no qual as demais modalidades de transporte são chamadas a integrar-se e a complementá-lo.

Hodiernamente, os rios continuam a ser o principal meio de mobilidade na região (ANTAQ, 2017), seja por meio de pequenos barcos de madeira ou até mesmo para a atuação de embarcações de grande porte ligadas ao abastecimento do mercado regional ou a conglomerados comerciais. Classifica-se o transporte fluvial realizado na Amazônia em três segmentos: navegação de transporte exclusivo de cargas; transporte de derivados de petróleo; e o transporte misto de passageiros e de cargas (COUTO, 2015).

Sob a ótica dos dados mais recentes, nota-se que o transporte de passageiros é o mais relevante na região da Amazônia devido às suas características físicas, sociais e econômicas, visto que há pouca competição entre modais, sendo as hidrovias, muitas

vezes, meio de transporte único para determinadas populações. Ferreira (2016) corrobora com tal assertiva ao afirmar que na maior parte dos casos, foram motivações econômicas e políticas, e não propriamente uma vocação geográfica do Brasil, que determinaram o predomínio do rodoviarismo.

Atualmente, sabe-se que a navegação na Amazônia caracteriza-se da seguinte maneira, segundo Ferreira (2016):

- Uso de balsas, cuja finalidade é o transporte de cargas diversas, como veículos, madeiras, maquinário e mercadorias;
- Viagens por meio das lanchas, utilizadas, basicamente, no transporte de passageiros com volume pequeno de bagagem;
- Para longos cursos, existem transatlânticos, oriundos de outros países, que movimentam o turismo; e
- Movimentação de embarcações mistas que visam atender passageiros e cargas, sendo esta a predominante.

O transporte misto é caracterizado pela movimentação concomitante de passageiros e cargas em geral para pequenos e grandes centros comerciais. Para Morgado *et al.* (2013), quando viabilizado por embarcações rústicas, há uma tendência no uso de canoas a remo ou até mesmo pequenas embarcações do tipo voadeira operadas por seus proprietários, consistindo, inclusive, no único meio de acesso às comunidades ou casas isoladas localizadas na beira de igapós ou rios pequenos. Pode também acontecer de embarcações de linha transportarem passageiros e distribuírem cargas ao longo de toda a malha hidroviária da região, com barcos maiores, conhecidos por ‘batelão’ ou ‘recreio’ (MORGADO *et al.*, 2013).

Entretanto, através de uma análise mais profunda frente à realidade apontada na Amazônia, o transporte fluvial por estas embarcações ditas mistas não apresenta um equilíbrio em termos sociais (segurança, conforto e higiene), econômicos (custos com terminais e custos com a viagem) e ambientais (poluição da água e poluição sonora). Com isso, o princípio defendido por Ferreira (2016) no que cerne ao papel da sustentabilidade em assegurar o bem-estar do homem e do ambiente considerando em equilíbrio seus três aspectos clássicos, ambiental, econômico e social, acaba por ser ferido.

Nos portos de Manaus, por exemplo, as irregularidades são inúmeras, a citar: embarcações cuja estrutura muito se assemelha com a de barcos contruídos no século passado, porém, que chegam a trafegar com excesso de passageiros e carga, onde as mercadorias, em geral, ocupam todo o espaço do porão e do convés inferior destinado a passageiros (DAVID, 2019). Acerca da perspectiva arcaica de construção aplicada à região, explana-se que na equação da navegação fluvial, as concepções e as construções das embarcações fluviais são, via de regra, consideradas padronizadas, “de prateleira”.

A maioria dos barcos amazônicos utilizados para o transporte de carga e passageiros é produzida de modo artesanal, sendo sua tecnologia transmitida através da tradição oral, onde os desenhos de projeto são feitos somente para que a embarcação possa ser legalizada após o término da construção (SOARES e DE BRITTO, 2014).

Na prática, verifica-se que, em vários casos, ao optar-se pelo uso de padrões sem sequer atribuir a devida importância à qualidade final da embarcação, são gerados altos riscos de perdas econômicas ou até mesmo ambientais que, a longo prazo, podem significar penalizações de eficiência. Além disso, há também os erros de concepção, de projeto ou de construção que, seja em conjunto ou não, tendem a elevar os riscos de uma embarcação sofrer um determinado tipo de acidente (PADOVEZI, 2003). Através do relatório técnico da Sociedade Brasileira de Engenharia Naval (SOBENA, 2010), corrobora-se a precariedade do transporte em embarcações de casco de madeira devido à falta de controle de qualidade na construção das mesmas, o que, posteriormente, coloca em risco a conservação das condições estruturais com o passar dos anos.

Seguindo o mesmo relatório técnico, relata-se que em boa parte das embarcações construídas de casco de madeira, constatou-se não haver acomodações satisfatórias aos passageiros, nem tampouco compartimentação adequada que possibilitasse seu transporte simultâneo ao de carga. Como consequência, de acordo com Frota (2008), os passageiros, na sua maioria, estão insatisfeitos em relação à qualidade dos serviços oferecidos e, normalmente, queixam-se da falta de uma boa infraestrutura portuária, da desordem e demora no atendimento, da despreocupação com a higiene e limpeza das embarcações, da falta de cortesia dos funcionários. Conforme o mesmo autor, as embarcações são de pequeno e médio porte e podem representar alto risco para os passageiros, devido à má qualidade da manutenção, à superlotação, do desconforto das acomodações, do carregamento desordenado e da baixa qualidade dos alimentos servidos.

Na visão do armador, várias são as adversidades que ainda reverberam no cenário da navegação amazônica. Longos trechos, preço do combustível, embarcações clandestinas, ausência do Estado, risco de acidentes pela condição natural da região, como troncos submersos nos rios, fenômenos atmosféricos como temporais e ventanias, especialmente na época do verão amazonense (FERREIRA, 2016).

Apesar, desses e de outros fatores, aliada à falta de regulamentação para o transporte fluvial, os armadores decidem permanecer com seu trabalho motivados por diversos fatores, como o afeto à profissão, por meio das gerações, a rentabilidade da atividade, bem como da sua representatividade dentro do município. Por conseguinte, ocorre a falta de programas de governo no sentido de legislar e regulamentar o transporte intermunicipal de passageiros. Embora já sejam observadas algumas melhorias recentes em algumas linhas, Cunha e Insper Jr. (2020) afirmam que há uma carência de condições mínimas de eficiência, conforto, regularidade e segurança inerentes à prestação de todo serviço público.

Portanto, expostas as falhas no planejamento do setor de transportes brasileiro, seja no passado ou presente, ressalta-se a necessidade do desenvolvimento de novas ferramentas e metodologias que haverá de contribuir para que decisões sejam tomadas quanto a novos investimentos. De acordo com Duarte *et al.* (2009), é latente a falta de conhecimento sobre as verdadeiras condicionantes regionais.

Diferentemente do modal rodoviário, privilegiado ao longo dos anos com vultosos investimentos mitigados pela política nacional, o transporte fluvial regional sempre esteve à mercê dos poucos investimentos locais, fato que tornava ainda mais difícil a ocorrência de melhorias físicas nas embarcações ou até mesmo modernização da frota, tendo em vista que o capital privado, próprio dos armadores, era o único subsídio disponível e, muitas vezes, não suficiente para atender às necessidades. Para Morgado *et al.* (2013), o planejamento apropriado do sistema de transportes torna-se importante ao atender as demandas existentes através de projetos de investimento em infraestrutura, aumentando as chances de inserção da população ribeirinha na dinâmica econômica vigente.

Contudo, a despeito de todos os reveses do transporte fluvial na região amazônica, este modal é um dos mais viáveis no que diz respeito à sustentabilidade, por ser considerado como o modo que provoca os menores impactos ambientais, além de mais econômico, por conta do seu baixo custo de manutenção (CUNHA & INSPER JR., 2020), quando

comparado ao transporte rodoviário, ferroviário e aéreo.

2.2 Consumo de Combustível de Embarcações

Boa parte dos órgãos nacionais e internacionais relacionados à navegação tem, nos últimos anos, trazido à tona como tema principal das discussões a eficiência quanto ao consumo de combustível dos navios, recebendo atenção especial devido, principalmente, segundo Kee *et al.* (2018), a dois fatores: o aumento do custo do combustível e à deterioração ambiental, em particular, a poluição do ar. Entre os anos de 1970 e 1980, o preço do combustível aumentou de forma considerável, cerca de dez vezes, obrigando muitos navios de alto consumo a permanecerem inoperantes. A situação só foi revertida em 1985, impulsionada pela queda no preço do óleo, motivo que levou à estagnação da pesquisa e do desenvolvimento voltado à eficiência energética no campo da indústria marítima. Entretanto, a partir de 2000, o custo do petróleo bruto voltou a subir, levando fabricantes de motores, estaleiros e projetistas a investir outra vez em soluções de projeto e operacionais de modo a reduzir o consumo de combustível e elevar a eficiência energética (BIALYSTOCKI & KONOVESSIS, 2016).

Embora o preço do petróleo tenha caído por um breve período, sobretudo, após a recessão de 2008, atualmente encontra-se em níveis recordes ao ponto dos operadores de navios, segundo Bialystocki e Konovessis (2016), não podem mais ignorar essa despesa como outrora aconteceu ou apenas incorporá-la ao preço das mercadorias transportadas, mas há a necessidade de projetar e operar navios mais eficientes, consumindo menos combustível por capacidade de carga.

O aumento e a volatilidade dos preços dos combustíveis constituem um grande problema para as companhias de navegação, pois o custo do combustível representa cerca de 60% do custo operacional de um navio (GOLIAS *et al.*, 2009). De modo mais distintivo, Stopford (2009) afirma que o consumo de combustível constitui aproximadamente dois terços dos custos de viagem de um navio e mais de um quarto dos custos gerais de funcionamento de um navio. Por esta razão, uma possível subida no preço do petróleo é capaz de afetar negativamente a rentabilidade das companhias (SOURTZI, 2019).

Tendo em vista que o princípio econômico visa o aumento da margem de lucro por meio da maximização da receita e da minimização do custo operacional (MAK *et al.*, 2014), a eficiência de combustível, portanto, é assumida como uma das características

mais importantes para que as empresas de navegação se mantenham competitivas no mercado (EIDE *et al.*, 2011; HASSELAAR, 2011). Como consequência, procedimentos e operações eficientes estão sendo adotadas para reduzir o consumo de energia, a fim de reduzir seus custos de gestão e, assim, manter sua posição competitiva no mercado (BEŞIKÇI *et al.*, 2019).

A maioria das previsões sugere que os volumes de transporte devem aumentar em um futuro previsível (SMITH *et al.*, 2014). Com isso, o desenvolvimento de novas estratégias para melhorar a eficiência energética, de fato, é uma direção a ser tomada por uma série de partes interessadas na indústria naval na condição de desempenhar um papel fundamental em relação ao planejamento futuro sustentável (BEŞIKÇI *et al.*, 2019). A operação eficiente dos navios pode levar à redução dos custos operacionais e consequente aumento da lucratividade (GKEREKOS *et al.*, 2019) em razão da redução do consumo de combustível e dos custos de manutenção (REHMATULLA *et al.*, 2017).

Em relação ao intenso foco na proteção ambiental, principalmente devido ao aquecimento global impulsionado pelo recorrente aumento dos gases de efeito estufa na atmosfera, estima-se, segundo Tzannatos & Stournaras (2014), que a indústria naval contribuiu com cerca de 3,3% das emissões de dióxido de carbono (CO₂) gerados a partir da queima de combustíveis fósseis entre os anos de 2007 e 2012, sendo 2,7% referente à navegação internacional e 0,6% para a navegação doméstica e de pesca, como também apresenta o Figura 2. Smith (2014) afirma que para o dado período, foram cerca de 1016 milhões de toneladas de dióxido de carbono por ano em média, ao passo que esta quota pode crescer entre 50 a 250% até 2050, a depender do crescimento do comércio internacional e dos desenvolvimentos energéticos.

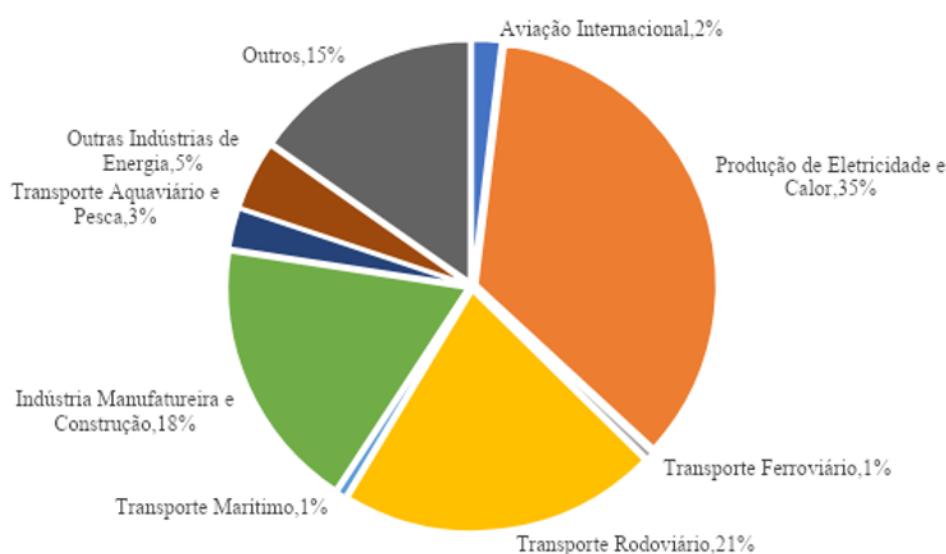
As emissões de CO₂ do setor marítimo em comparação com outras indústrias estão presentes no Figura 2 a partir do qual observa-se que o setor marítimo é o segundo maior emissor de dióxido de carbono quando comparado aos meios de transporte: rodoviário, aquaviário, aviação, etc.

Tal questão tem sido cada vez mais considerada fundamental em diferentes áreas, sejam elas industrial, civil ou doméstica. O uso do combustível fóssil leva, além da produção de gases do efeito estufa, a geração de óxidos de nitrogênio (NO_x) e óxidos de enxofre (SO_x), relacionados às fatalidades humanas e degradação ambiental (MARINE INSIGHT, 2017).

A crescente demanda de bens e transporte de pessoas levará a um aumento da frota

mundial, resultando não apenas em um crescimento nas emissões globais do transporte aquático, mas também mais combustível fóssil será necessário para a operação dos navios (NWAHOA *et al.*, 2016). Sabe-se que as atividades marítimas, por exemplo, dependem do consumo de combustível, impactando não somente o meio ambiente, mas também a saúde pública (SANABRA *et al.*, 2013).

Figura 2: Emissão de dióxido de carbono (CO₂) a partir do consumo de combustível pelos diferentes setores em 2014.



Fonte: Autor (2023).

Todavia, apenas nos últimos anos é que a legislação no setor de transportes, em especial o do transporte marítimo, foi reorganizada, tendo em vista que sua contribuição frente ao lançamento de gases de efeito estufa não pode ser negligenciada (ANCONA *et al.*, 2018). Além disso, consideráveis resultados de pesquisas levaram a Organização Marítima Internacional (OMI) a tomar medidas concertadas nesse sentido, de modo a limitar significativamente o impacto ambiental dos navios (ANDERSON, 2012; LINDSTAD, 2016). Com isso, sendo a emissão de gases poluentes intimamente relacionada ao consumo de combustível, os esforços voltados à melhoria da eficiência energética empregada à indústria naval tornaram-se primordiais para redução da taxa de CO₂ lançados na atmosfera (CIPOLLINI *et al.*, 2018).

Para Sarkis (2006) e Epstein (2008), quanto melhor for o desempenho social e ambiental de uma empresa, maior valor será agregado ao negócio. Sendo assim,

o aprimoramento da eficiência energética dos navios impactará tanto em relação a sustentabilidade ambiental quanto para o sucesso econômico das cadeias de abastecimento marítimo (MANSOURI *et al.*, 2015; QIU *et al.*, 2018).

Ademais, o objetivo de reduzir as emissões de CO₂ também está associado ao aumento do preço do combustível, estimulando, assim, a adoção de inovações tecnológicas e operacionais para diminuição de seu consumo (BOUMAN *et al.*, 2017). Dentre essas novas práticas e tecnologias, cujo enfoque é melhorar a eficiência energética no setor de transporte marítimo, normalmente dividem-se entre características relacionadas ao projeto da embarcação, desempenho operacional e condições ambientais.

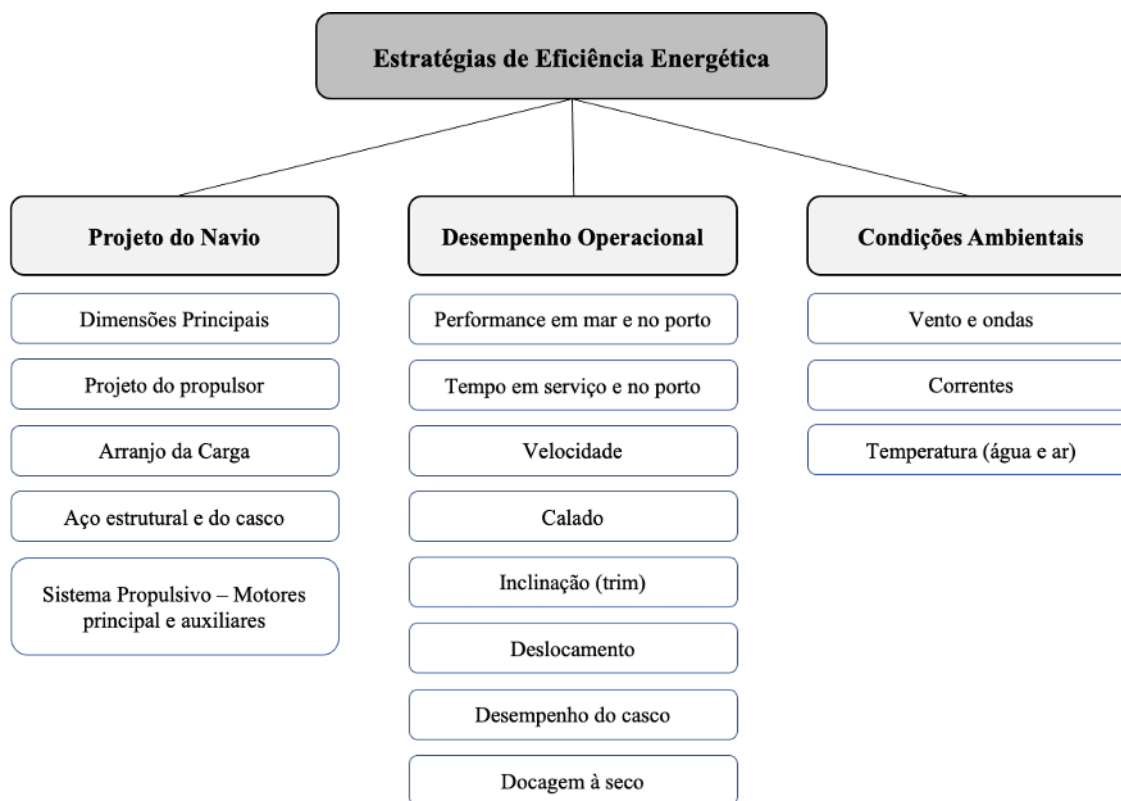
Os parâmetros levados em consideração cada uma destas frentes apontadas, segundo Sourtzi (2019), se fazem presentes na Figura 3. Embora não seja foco deste estudo, é importante ressaltar que além das medidas apresentadas anteriormente, alguns autores consideram também as condições climáticas influenciadas por vento e ondas, correntes, temperatura.

2.2.1 Parâmetros de Projeto da Embarcação

Diversas investigações têm sido promovidas de modo a tentar estimar e calcular o consumo de combustível levando em consideração o projeto do navio. Conforme Lu *et al.* (2013), a maioria das pesquisas baseia-se nas questões relacionadas à potência propulsiva, resistências do casco, especificações dos motores em concordância com vários outros fatores operacionais e ambientais, cujo intuito é obter uma previsão mais precisa do consumo de combustível.

Quando um método investigativo está relacionado ao projeto do navio, em boa parte dos estudos, o enfoque se dará em otimizar a forma do casco, a configuração do hélice, uso de materiais leves, etc., fatores que, por muitos, são considerados como tendo maiores potenciais do que se comparado, por exemplo, aos métodos relacionados ao desempenho operacional (TILLIG *et al.*, 2017). Na comunidade marinha, inclusive, a otimização da forma do casco continua a ser reconhecida como um campo em crescimento de modo a melhorar a eficiência energética dos navios.

Figura 3: Fatores que influenciam o consumo de combustível das embarcações.



Fonte: Autor (2023).

Tratam-se, portanto, de medidas referentes a soluções técnicas físicas. Em outras palavras, o impacto requerido está na evolução do desempenho de partes individuais dos sistemas, como o motor, o hélice e o casco (ANCONA *et al.*, 2018). Fontes de energia adicionais podem ser usadas para propulsão (por exemplo, velas e rotores) e para geração de energia auxiliar (por exemplo, células de combustível), proposto por Coraddu *et al.* (2017).

Outra estratégia consiste na regulação da velocidade de navegação: através da determinação da velocidade ótima do motor, é possível reduzir em até 19% o consumo da embarcação se comparado ao valor típico de combustível por unidade de distância (BALDI *et al.*, 2016). Todavia, as principais medidas para aprimoramento das condições de navegação dos navios, de fato, dependem do tipo da embarcação, cargas, rotas, entre outros fatores. Segundo Nwaoha *et al.* (2016), a dificuldade maior enfrentada atualmente neste campo está em determinar quais são os mais adequados parâmetros para um determinado navio e serviço.

Sobretudo, entende-se ser importante elencar alguns fatores de projeto preponderantes para a redução do consumo de combustível e conseqüente amortecimento dos impactos ambientais e econômicos:

Dimensões principais Conforme Lu *et al.* (2013), o desempenho de um navio em condições operacionais muda sob fatores climáticos reais, os quais é importante citar: altura de onda, período de onda, velocidade do vento, direção do navio, velocidade do navio e aspectos de projeto do navio (dimensões principais e forma do casco, formato da superestrutura, calado, etc.).

Uma parte significativa do consumo de combustível do navio, como discutido anteriormente, pode ser influenciada pelas dimensões do navio (comprimento, boca e calado). De acordo com Brynolf *et al.* (2016), tais parâmetros são, de forma frequente, limitados por fatores externos, como a profundidade da hidrovia, o comprimento do cais nos portos e a largura dos canais. Em particular, a resistência à formação de ondas é fortemente influenciada pela esbeltez do casco. Ainda segundo Brynolf *et al.* (2016), como consequência, navios rápidos tendem a ter cascos mais longos e delgados, enquanto que navios mais lentos, para os quais o componente de atrito da resistência domina, têm cascos mais delgados e possivelmente desempenho reduzido.

Por fins de exemplificação, os navios empregados em viagens marítimas atendem à lógica da influência acerca da relação existente entre as dimensões principais, como abordado anteriormente. Segundo Sharifi *et al.* (2017), o aumento da razão comprimento – feixe ou apenas do comprimento, bem como redução do coeficiente de bloco, acarreta uma diminuição no consumo de combustível de até 3 a 5%. Em contrapartida, quando em comparação com o aumento de outras propriedades como boca ou calado, o comprimento é a dimensão mais cara. Tomando por base um navio-tanque do tipo Aframax, quando acrescida a relação comprimento/boca de 5,5 para 5,75, mantendo a velocidade do navio e volume de carga constantes, o custo de construção da embarcação aumenta de 0,25 a 1%. Outras implicações dizem respeito ao fato de que para a mesma relação apresentada, segundo Molland (2011), a resistência de onda é diminuída, enquanto que para a razão boca/calado a resistência friccional decai em função da redução da superfície molhada.

Isto posto, o projeto do casco do navio é um processo iterativo que deve atender a requisitos que, na maioria dos casos, são tidos como conflitantes, tendo em vista que

devem ser determinadas variáveis de projeto sujeitas às relações e restrições envolvendo umas às outras (SONER *et al.*, 2019). Em geral, muitos fatores devem ser considerados e nem todos eles são de natureza hidrodinâmica.

Velocidade de Cruzeiro A redução da velocidade do navio é o método mais eficiente em termos de economia de combustível, ao passo que, conforme proposto por Wijnolst e Wergeland (2011), Stopford (1999) e Chang e Chang (2013), quando adotada uma velocidade de cruzeiro cerca de 2/3 nós abaixo da velocidade de projeto, ocorre um impacto considerável sobre o consumo diário de combustível e, portanto, pode reduzir pela metade o custo operacional das companhias de navegação.

O motivo principal dá-se por conta da relação não-linear entre a velocidade do navio e o consumo de combustível, no qual, o primeiro tem um impacto maior sobre o segundo devido à sua função de terceira ordem com a potência consumida pelo motor principal (HARVALD, 1977). Em outras palavras, se a velocidade da embarcação dobra, a potência utilizada do motor aumentará, pelo menos, oito vezes. Sendo assim, em caso de uma velocidade do navio reduzir em 10%, o volume de combustível necessário diminuirá em 27% (FABER *et al.*, 2012). Portanto, mais especificamente, de acordo com Oleksiy *et al.* (2013), observa-se que para altas velocidades, as perdas mecânicas aumentam se comparadas aquelas quando para um desempenho a baixa velocidade, o que acarreta um consumo de combustível maior.

À luz desse princípio, muito tem sido discutido acerca da redução de velocidade, adotando o chamado *slow steaming*, cujo enfoque é operar o navio em velocidade mais devagar do que a sua própria de projeto. Para Psaraftis e Kontovas (2016), consiste em uma técnica que envolve apenas desacelerar ou mesmo reduzir o motor de um navio, ou seja, reconfigurar o motor para que uma produção de energia menor seja alcançada e, conseqüentemente, velocidades mais lentas sejam adotadas. Embora seja, em termos de consumo de combustível, mais lucrativo, é uma prática que deve estar em equilíbrio quanto as outras necessidades comerciais e operacionais, tendo em vista que reduzir a velocidade tem como efeito a expansão do tempo de viagem (MANDER, 2017), resultando em uma frota maior para manter o mesmo nível de serviço (BEŞIKÇI *et al.*, 2019).

Projetar para a velocidade certa, ou faixa certa de velocidades, tem outros benefícios também: ao estimar a velocidade ideal e os perfis de rota com base em dados empíricos

e modelos estatísticos, por exemplo, é possível obter economias de até 10% (BEŞIKÇI *et al.*, 2019). Além disso, uma forma de casco otimizada para a velocidade mais lenta geralmente provém uma forma mais cheia e maior porte de carga. O motor principal e o hélice também podem ser aprimorados em torno de uma velocidade mais lenta para o máximo benefício (SHARIFI *et al.*, 2017).

Capacidade de Carga Muitos modelos relacionados à investigação dos fatores que influenciam a demanda de combustível pelas embarcações consideram não haver dependência alguma quanto à carga útil do navio. Em contrapartida, segundo Psaraftis e Kontovas (2016), essa suposição é razoável apenas para quando a carga útil do navio é constante ou que não tenha muitas variações consideráveis, e caso não seja válida, uma séria sub ou superestimação dos custos de combustível pode ocorrer. Sendo assim, a resistência do navio e, portanto, o consumo de combustível de uma dada velocidade podem ser drasticamente diferentes, a depender da condição de carregamento: cheio, vazio ou intermediário. Assim, caso essa prática seja aplicada em embarcações que não estejam totalmente carregadas, a redução no uso de combustível é estimada em até 5% (REICHEL *et al.*, 2014).

A recíproca é verdadeira ao passo que, à medida com que o calado do navio difere dependendo do volume de carga transportada, maior será a resistência ao avanço especialmente na região da proa (TILLIG *et al.*, 2018). Por esta razão, o aumento da quantidade de carga contribui para um calado e deslocamento maiores, produzindo maior resistência e, portanto, maior consumo de combustível (BEŞIKÇI *et al.*, 2019).

Para Ventikos *et al.* (2018), o consumo de combustível está estritamente relacionado à carga do motor e também com as horas de operação do mesmo, ao passo que, caso opere continuamente além de um limite predefinido, pode resultar em maior consumo de combustível.

2.2.2 Desempenho Operacional

Embora muitas sejam as vantagens referentes aos métodos de projeto, para as empresas de navegação, torna-se mais atraente aplicar métodos operacionais, tendo em vista requererem um investimento de capital menor, cujo potencial de economia é significativo (CORADDU *et al.* 2017; TILLIG *et al.*, 2017).

No mesmo grau, as empresas de navegação devem, no entanto, aplicar métodos operacionais que requerem menos investimento de capital e ainda fornecem potenciais de economia significativos (CORADDU *et al.*, 2017; TILLIG *et al.*, 2017). As medidas operacionais incluem esforços que não requerem a instalação de novos equipamentos a bordo (BALDI *et al.*, 2016). Visando a eficiência energética voltado às questões operacionais, a OMI (2009) traz como principais abordagens: a otimização da velocidade de uma embarcação; controle do consumo de combustível na base atual; o aperfeiçoamento do plano de viagens; a otimização do calado e trim, dentre outros.

Existe uma associação estreita entre projeto de navios, avaliação de desempenho, operação, gerenciamento de regulamentações e competitividade, onde o consumo de combustível tem grande influência. Dentre tais, Coraddu *et al.* (2017) determinou, como aspecto chave, a eficiência operacional do navio, no qual, o setor de transporte marítimo desenvolveu novas abordagens operacionais para atingir os objetivos anteriores e, por decorrência, reduzir o consumo de combustível:

- Adaptar as rotas a fim de evitar condições de mau tempo, reduzindo assim o impacto negativo de ondas altas e vento forte no consumo de combustível do navio;
- Maximizar a carga transportada, reduzindo o volume do lastro; e
- Melhorando as configurações de trim e calado, juntamente com a otimização dos horários e as práticas para o polimento do casco e da hélice, levando à redução da resistência do navio para uma dada velocidade, etc.

Normalmente, a otimização da viagem baseia-se em objetivos múltiplos que, muitas vezes, são considerados conflitantes e, por isso, são ponderados por grau de importância a depender dos armadores operantes (LU *et al.*, 2015). Por exemplo, algumas empresas de navegação priorizam a chegada no horário e tempos de trânsito mais curtos ao invés do consumo reduzido de combustível; para outras, fornecer um serviço “sustentável” deve ser inegociável. É importante enfatizar que melhorar um objetivo pode reduzir a eficiência de outro e é por isso que cada setor está procurando maneiras de aumentar a eficiência do combustível por meio de produtos de alto teor de inovação (UYANIK *et al.*, 2019).

O planejamento ideal da viagem consiste, portanto, em transportar a carga máxima disponível para as condições operacionais da embarcação, implicando em um volume de lastro reduzido (NORSTAD, 2011). Por outro lado, a adaptação das rotas contribui para

evitar condições de mau tempo e, conseqüentemente, diminuir os impactos negativos causados pelas ondas e ventos fortes quando relacionados ao consumo de combustível do navio (NISHIDA *et al.*, 2011; SHAO, 2006); melhorar a configuração do trim e de calado, bem como adotar práticas acertadas de polimento do casco e da hélice, são considerações importantes para redução da resistência do navio para uma determinada velocidade (LEE *et al.*, 2013; MOUSTAFA *et al.*, 2015); o conceito de slow steaming ajuda na diminuição, de maneira drástica, dos custos com combustível: entende-se que a quantidade de carga transportada é proporcional à velocidade (decrece linearmente com a velocidade), enquanto que a demanda de potência dos motores depende aproximadamente do cubo da velocidade, logo, a vantagem em relação ao uso de velocidades reduzidas torna-se notável (CARIOU (2011); GUAN *et al.* (2014)).

Outros autores corroboraram com o uso destes métodos operacionais bem conhecidos e maduros com o objetivo de aumentar a eficiência energética das embarcações através do slow steaming (MALONI *et al.*, 2013), otimização da velocidade de cruzeiro (FAGERHOLT *et al.*, 2010) e planejamento de viagem (LU *et al.*, 2015). Os modelos de previsão de consumo de combustível formam a base para as ferramentas de apoio à decisão usadas em terra/a bordo para avaliar o impacto do combustível de instruções de viagem alternativas, opções de roteamento e de manobra da embarcação no mar e devem, de preferência, ser precisos e confiáveis (BAN *et al.*, 2017; KANELLOS, 2014).

2.2.3 Condições Ambientais

O desempenho operacional do navio não está apenas relacionado às condições do navio (ou seja, casco e maquinário), mas também a recursos adicionais, como condições climáticas e de carregamento (ORIHARA & TSUJIMOTO, 2017). Dessa forma, é importante frisar que as condições climáticas preponderantes ao longo da viagem certamente são parâmetros que contribuem para o consumo de combustível, haja vista ainda a possibilidade de ocorrerem mudanças de acordo com o tempo transcorrido e a localização apresentada (PSARAFTIS & KONTOVAS, 2016). Ressalta-se ainda que, durante o trajeto, a velocidade do navio tende a ser impactada, influenciando diretamente no tempo de navegação. Ocorre que, para contornar tal situação, muitos operadores optam por adotar diferentes velocidades para cada condição climática, exigindo um alto desempenho do motor e, portanto, um consumo maior é necessário (ZHANG *et al.*, 2019).

Estima-se que, devido às incertezas proporcionadas pelas circunstâncias ambientais, quando em consideração nos cálculos de projeto da embarcação, uma margem de 10 a 15% do consumo de combustível previsto é definida, justamente, levando em consideração as influências do vento e das ondas predominantes (BIALYSTOCKI & KONOVESSIS, 2016). Cada navio em particular, quando sob várias condições de viagem (velocidade, incrustação e degradação do sistema de propulsão e calado) apresentará desempenho bastante diferente, especialmente ao lidar com condições climáticas severas.

Magnussen (2017) estabelece que existem fatores ambientais constituintes da chamada resistência adicional, parcela que contribui para o aumento da resistência total do navio. Entende-se que o termo resistência adicional é utilizado para explicar a ocorrência de perda de energia devido a fatores externos, descrita como a soma das resistências geradas por: ondas (MOLLAND, 2011); vento (MOLLAND, 2011); composição e temperatura da água do mar (MAGNUSSEN, 2017).

Destarte, a resistência provocada pelas ondas e pelo vento não podem ser negligenciadas, haja vista terem impacto significativo na velocidade do navio e, conseqüentemente, no consumo de combustível. Quando em referência à resistência gerada pelas ondas, refere-se aos efeitos das ondas sobre o casco da embarcação (MOLLAND, 2011). Segundo De Grado *et al.* (2016), a diferença no que tange à resistência de uma embarcação quando operando em condições reais se comparado à sua operação em águas calmas pode ser de mais de 30%.

Quanto à resistência adicionada devido ao vento, distingue-se como os efeitos da força do vento não apenas sobre o casco imerso, mas também na seção emersa da embarcação, acima da água. Molland (2011) define que a resistência do ar é responsável por 4 a 8% da resistência total do navio, associando-a à velocidade e direção do vento, além da dependência inerente às particularidades da superestrutura (isto é, dimensões e forma) e, por fim, da velocidade da embarcação.

Outro parâmetro, que tem impacto na resistência do casco, é devido a água do mar, fato que pode ser justificado por conta das propriedades da água doce serem diferentes daquelas da água do mar (MAGNUSSEN, 2017). Mais precisamente, esta composição pode ser expressa em função da temperatura, densidade, viscosidade e salinidade (NPL, 2018). Portanto, um aumento potencial na densidade da água causa não apenas um aumento na resistência do casco, mas também um aumento na potência efetiva, fato que

tem um efeito negativo na eficiência de combustível da embarcação (FESTUS & SAMSON, 2015).

A temperatura da água do mar tem assumido papel crucial no impacto à resistência do navio e, conseqüentemente, influenciado no consumo de combustível, tendo em vista que a viscosidade e a densidade são atributos dependentes daquele. Salienta-se que essas propriedades são relativas à localidade e estação do ano, proporcionando variações mínimas na temperatura, viscosidade e densidade (MAGNUSSEN, 2017).

A partir de alguns levantamentos baseados em dados operacionais reais, é possível reduzir o consumo de combustível em até 3% com a adoção de planos meteorológicos de rotas (weather routing plans). Segundo Beşikçi *et al.* (2019), trata-se de um planejamento voltado à determinação da rota ideal, levando em consideração as condições climáticas (vento, corrente e efeitos do mar) na viagem proposta. A rota ótima designada para a viagem é aquela que permite a conclusão mais rápida do trajeto, na qual estão garantidos o conforto e a segurança, além da máxima eficiência energética sob quaisquer condições climáticas. O objetivo principal é minimizar a resistência total.

2.3 Método de Descoberta de Conhecimento em Banco de Dados (DCBC)

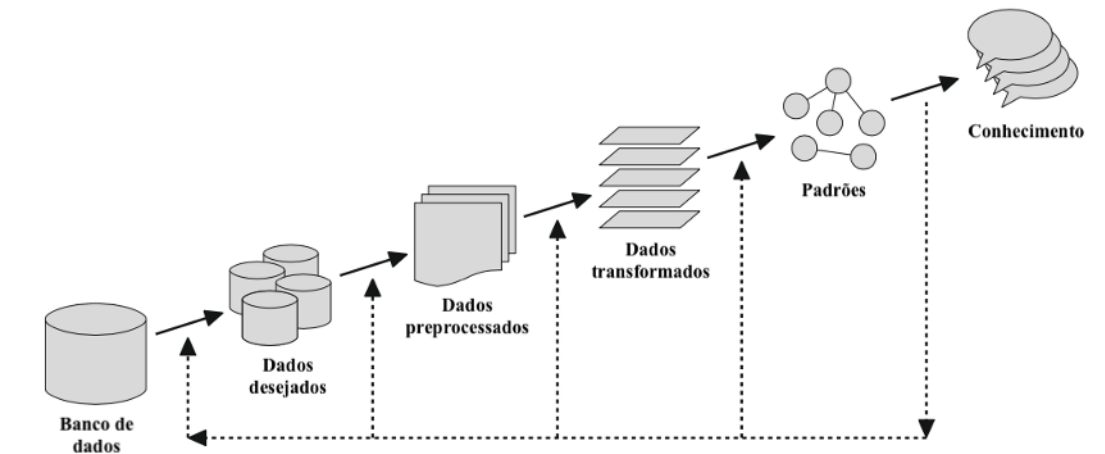
A abordagem de um conjunto de dados, normalmente, requer planejamento prévio (STOROPOLI, 2016). Por esta razão, tem-se como intuito primordial acerca do planejamento a priorização da escolha do universo de dados a serem explorados, bem como dos resultados que pretende-se alcançar. De modo a cumprir tal intento, a estrutura básica escolhida é chamada Descoberta de Conhecimento em Banco de Dados ou, em inglês, Knowledge Discovery in Database, comumente utilizada nas áreas de ciência da informação e engenharia da computação para devidos fins (LAVALLE *et al.*, 2011; MANYIKA *et al.*, 2011).

O DCBC consiste no conjunto de ações de exploração e análise de dados, envolvendo a seleção, pré-processamento, limpeza, transformação, data mining e avaliação (FAYYAD *et al.*, 1996; TSAI *et al.*, 2014; ADHIKARI & ADHIKARI, 2015). Constatou-se que, quanto ao referido método, há um espectro amplo devido ao fato de sua natureza genérica (STOROPOLI, 2016). Ademais, trata-se de uma estrutura de processo organizada para

encontrar conhecimento útil de um banco de dados (YAMASAKI et. al, 2019).

O fluxo de construção do modelo baseado na metodologia apontada está apresentado na Figura 4.

Figura 4: Pipeline do processo de construção do modelo.



Fonte: Autor (2023).

A fase de processamento do banco de dados tem início com coleta de inúmeros dados, a partir dos quais são selecionados aqueles cujas atribuições são desejadas. Posteriormente, estes mesmos registros escolhidos, chamados dados de destino, são pré-processados para corresponder a um formato adequado, removendo informações desnecessárias ou interpolando dados ausentes.

Os registros de dados pré-processados são transformados em um formulário de redução de dimensionalidade (dados transformados) e padrões úteis são extraídos usando algoritmos de mineração de dados. Isso significa que, após ter acesso aos dados, algoritmos específicos serão usados para a remoção de duplicidades, enriquecimento dos dados pela complementação de partes ausentes e ainda realização da transformação dos formatos (YAMASAKI *et al.*, 2019). Um exemplo a ser citado a respeito da transformação que ocorre é o preenchimento de campos que estejam vazios. Por fim, os padrões extraídos são interpretados como conhecimento.

É uma atividade de descoberta automática de conhecimentos, modelos, padrões e anomalias em grandes repositórios de dados. Conhecimento é a compreensão do fenômeno que produz os dados observados; modelos são conjuntos matemáticos e lógicos de

funções que descrevem os dados e explicam seu comportamento; padrões e anomalias são respectivamente regularidades esperadas e observações anormais nos dados (GUARASCIO *et al.*, 2018).

2.3.1 *Machine learning* Aplicado à Estimativa e Otimização de Consumo de Combustível na Navegação

Conforme visto anteriormente, há uma relação forte entre o consumo de combustível e a quantidade de gases poluentes lançados à atmosfera, quando sob a ótica ambiental, e também relacionado ao aumento do custo operacional, financeiramente falando. Além disso, outros vários métodos e formas foram discutidos, nos quais as empresas de navegação têm investido na tentativa de proporcionar um consumo de combustível mais eficiente nas embarcações, a serem, de acordo com Uyanik *et al.* (2019): limpeza do casco, trim, energia eólica, energia solar, energia das ondas e célula de combustível.

Mesmo com o uso desses sistemas eficazes, uma outra solução muito usada envolve a criação de um modelo de estimativa que possa revelar as variáveis (e fatores externos) que afetam diretamente o consumo de combustível, tornando este método mais eficaz para monitoramento e consequente aumento da eficiência do combustível (UYANIK *et al.*, 2020). Outro benefício diz respeito à possibilidade de reconhecimento de desvios de desempenho considerados padrões, acusando possíveis degradações ou prevenindo problemas dos sistemas e subsistemas da embarcação (CIPOLLINI *et al.*, 2018; RAPTODIMOS & LAZAKIS, 2018; LAZAKIS *et al.*, 2018; LAZAKIS *et al.*, 2019). Dessa maneira, torna-se mais barato para as empresas implementar o modelo proposto do que adaptar novos sistemas à sua frota para uma gestão eficiente.

Partindo desse pressuposto, segundo Gkerekos *et al.* (2019), várias tentativas de modelagem têm sido levantadas na literatura, variando desde abordagens com foco em fórmulas semi-empíricas de modelos de testes até modelos orientados a dados, chamado *machine learning*, juntamente com sistemas instalados de registro e monitoramento automatizado de dados que extraem informações de uma grande quantidade de pontos levantados.

De acordo com Wilson e Keil (2001), *machine learning* é definido como um campo científico que permite o aprendizado a partir de dados de medição reais baseados em dados históricos de um sistema. Em um outro conceito abordado por Alpaydin (2014),

trata-se de um processo de programação de computadores cujo propósito é otimizar um determinado critério de desempenho com base em dados de exemplo (também chamados dados de treinamento) ou experiência anterior (dados pretéritos). Por conseguinte, ao analisar o desempenho passado do sistema, é determinada sua condição de saída e melhorado o desempenho, quando possível (WITTEN *et al.*, 2011).

Embora sejam mais caros computacionalmente, os modelos de *machine learning*, portanto, são especialmente adotados como os mais promissores na estimativa e otimização sobre o consumo de combustível por meio de diferentes maneiras, como supervisionado, não supervisionado, etc. Julga-se importante explicitar que, no caso da aprendizagem não supervisionada, os dados de treinamento são apenas valores de entrada e o objetivo do algoritmo é fornecer alguns insights sobre esse conjunto de dados (BISHOP, 2006).

Concomitantemente, para a aprendizagem supervisionada, o foco do algoritmo é utilizar dados de treinamento provindos do conjunto de amostras de entradas e do alvo (saída) para conseguir prever um valor alvo relevante quando para novas observações (LAZAKIS *et al.*, 2019). Mais a fundo, para problemas em que é existente um número finito de categorias (classificações) de saída do modelo, chama-se de classificação, enquanto que para variáveis de saída contínuas, é determinado como de regressão (BARUA *et al.*, 2020).

Atrelado o uso de métodos estatísticos ao desenvolvimento de modelos com *machine learning*, a estimativa de consumo de combustível de uma embarcação é processada com uma baixa taxa de erro, proporcionando à empresa e aos seus engenheiros uma visão importante voltada à gestão da eficiência energética (UYANIK *et al.*, 2020). Com isso, a avaliação e otimização do desempenho do navio tem sido incentivada, de modo primordial a explicar a relação entre o consumo de combustível e a velocidade do navio, potência do motor e as condições climáticas diversas de operação as quais são submetidas a embarcação. Sabe-se, conforme defendido por Haranen *et al.* (2016), que a análise relacional entre tais fatores é expressa através de um modelo com parâmetros que são físicos ou estatísticos, tendo em vista a otimização das operações dos navios de modo a minimizar os custos operacionais e maximizar as receitas (LAZAKIS *et al.*, 2019).

As abordagens de aprendizagem estatística utilizadas para o monitoramento do desempenho operacional do navio têm oferecido previsões interpretáveis e robustas quando comparadas aos modelos teóricos (CORADDU *et al.*, 2017; EL NOSHOKATY, 2013).

Através da adoção do aprendizado de máquina, ocorre o tratamento do conjunto de dados operacionais de alta dimensão referentes aos navios, possibilitando a extração das informações, até então, “ocultas”. Existem vários recursos que afetam o desempenho operacional intrínseco a cada navio, como velocidade e ângulo do vento, altura da onda, pressão do ar, peso da carga, entre outros.

Segundo Soner *et al.* (2019), alguns destes parâmetros são altamente correlacionados, pois aumentam a variância e o viés, causando um modelo complexo e uma taxa de precisão baixa. Embora os modelos lineares forneçam uma alta taxa de precisão de predição, as relações entre as características não são bem explicadas, não sendo possível interferir no funcionamento do navio e apontar uma solução que atinja o objetivo da eficiência energética.

Para James *et al.* (2013), os modelos estatísticos, cujo principal objetivo é extrair tais informações ocultas do conjunto de dados disponíveis, são utilizados sob a premissa da previsão e da interpretação. Em outras palavras, para alguns casos, a análise leva a revelar a relação entre a resposta e a relevância de cada preditor. O estudo abrangente acerca da relação entre os fatores de influência e o consumo de energia do navio, por exemplo, torna os metamodelos, aproximações estatísticas de modelos físicos originais ou modelos de simulação, a saída mais indicada para aprender a relação entre os fatores de entrada e saídas de modelos de conjunto de dados de treinamento (YUAN & NIAN, 2018).

Diferentemente de uma série de sistemas operados em terra, os navios atuam em muitas condições diferentes e, portanto, com grandes variações na demanda de energia, como acontece com os navios de cruzeiro, por exemplo, onde podem ser observadas demandas de energia em diferentes formas (mecânica, elétrica, térmica) (BALDI *et al.*, 2016). Por esta razão, investimentos vultosos têm sido empregados no desenvolvimento dos Sistemas de Apoio à Decisão (SAD) tipos de sistemas de informação baseados em computador que podem ajudar os tomadores de decisão a utilizar dados, modelos e outros conhecimentos para resolver problemas semi-estruturais e alguns não estruturais, ou seja, que não podem ser medidos ou modelados analiticamente (BEŞIKÇI *et al.*, 2019).

A necessidade de maior precisão do que os modelos empíricos ou teóricos e menor custo computacional do que as simulações CFD, tem levado a comunidade científica à experimentação com métodos baseados em dados e específicos para navios. A ideia básica por trás desses modelos é explorar os dados coletados da operação de um determinado

navio e usá-los para criar um modelo estatístico que pudesse estimar suas necessidades de energia ou prever seu consumo e monitorar seu desempenho (KARAGIANNIDIS *et al.*, 2019).

Segundo Barua *et al.* (2020), as abordagens tradicionais, portanto, se enquadram amplamente em duas categorias: modelagem estatística, como visto anteriormente, e pesquisa operacional. Entretanto, em comparação com tais abordagens tradicionais, a modelagem baseada em *machine learning* sustenta-se em três vantagens:

- a. As abordagens tradicionais baseiam-se em pressupostos definidos inicialmente sobre a relação entre as entradas e saídas do sistema que, outrora, tendem a ser menos de 100% precisas e frequentemente simplificadas, levando à uma descrição e previsão tendenciosas do sistema e influenciando a tomada de decisão abaixo do ideal. A modelagem de *machine learning*, segundo Barua *et al.* (2020), evita tal ocorrência tendo em vista que nenhuma suposição a priori é necessária, o que acontece é o treinamento e aprendizado por parte da máquina.
- b. O ambiente operacional, grande e complexo, traduz-se em dados de grandes dimensões. Consequentemente, a modelagem estatística pode encontrar multicolinearidade com a qual há de ser causado um possível comprometimento da validade dos modelos. Para o caso de uma multicolinearidade presente na regressão linear, pequenas mudanças nos dados causarão grandes oscilações nas estimativas dos coeficientes (BARUA *et al.*, 2020). Muitos métodos de ML, entretanto, são bastante robustos à multicolinearidade (CARAVIELLO *et al.*, 2018; KANG *et al.*, 2015).
- c. Ainda sob a premissa do tamanho e complexidade dos problemas envolvendo a abordagem das embarcações, o tratamento por meio de pesquisa operacional requer modelos de grande escala que são computacionalmente caros para resolver. Adotando uma filosofia diferente e livre de modelos, a otimização baseada em ML precisa de um tempo de solução ínfimo, se comparado ao nível de alta qualidade com o qual produz as soluções (NAZARI *et al.*, 2018). A maior parte do tempo de computação é gasto no treinamento do modelo antes da otimização.

Bialystocki e Konovessis (2016) realizaram uma análise estatística acerca de *noon reports* (dados de meio-dia) referentes à uma embarcação roll-on/roll-off (ro-ro) com o

intuito de identificar a influência de fatores como calado do navio, deslocamento, condições climáticas e rugosidade do casco e do hélice. Após a aplicação de várias correções sugeridas aos dados obtidos juntamente a uma filtragem relevante, as curvas para cada estado de mar observado durante o estudo foram ajustadas, obtendo como resultado um algoritmo simples que se aproxima do consumo de óleo combustível. Lu *et al.* (2015) desenvolveram um método semi-empírico para a previsão do desempenho operacional de navios. Este método é baseado na modelagem em águas tranquilas e parâmetros adicionais de resistência, além de levar em consideração as condições climáticas e estado do mar. A partir deste modelo gerado, acontece a otimização da rota do navio.

Besikçi *et al.* (2016) sugeriram o uso de Redes Neurais Artificiais (RNA) para a previsão do consumo de óleo combustível do navio para diversas condições operacionais. Isto posto, também um Sistema de Apoio à Decisão (SAD), elaborado para operações em tempo real, contribuindo para a eficiência energética. A metodologia sugerida compara-se à análise de Regressão Múltipla, entretanto, exibindo resultados superiores. Petersen *et al.* (2012), por sua vez, avaliaram abordagens de modelagem de consumo de combustível do motor principal de balsas também utilizando RNAs, no qual a saída dos modelos foi usada para fins de otimização do trim da embarcação.

Meng *et al.* (2016) sugerem uma metodologia de pré-processamento de dados com base na pontuação dos outliers presentes no conjunto. À posteriori, dois modelos de regressão são desenvolvidos a fim de vincular as informações disponíveis com a variável-alvo, o consumo de óleo combustível. O primeiro modelo conecta o consumo de combustível do navio com a sua velocidade e deslocamento. Tomando como parâmetro os resultados gerados pelo primeiro modelo, são incluídos os dados a respeito das condições meteorológicas. O trabalho foi validado utilizando dados de meio-dia de navios petroleiros. Simonsen *et al.* (2018) propôs outro método de utilização de dados do Sistema de Identificação Automática (SIA) para estimar o consumo de combustível de navios de cruzeiro navegando em águas norueguesas, no qual pode ser também usado para estimar as emissões de gases de efeito estufa. Yao *et al.* (2012) investigaram a correlação entre o consumo de combustível e a velocidade de navios porta-contêineres de diferentes tamanhos.

Tsitsilonis e Theotokatos (2018) desenvolveram uma metodologia sistemática para gerenciamento de energia de motores primários de navios. Uma análise estatística é

combinada com análises de energia para identificar as principais áreas onde a economia de energia pode ser obtida. Wang *et al.* (2018) propôs um modelo de regressão LASSO para a estimativa do consumo de combustível de uma embarcação. Este modelo mostrou ter um desempenho ótimo quando comparado aos modelos de Redes Neurais Artificiais, Support Vector Regressor e Processos Gaussianos em um estudo de caso utilizando dados de baixa frequência obtidos de uma frota de navios porta-contêineres.

Abebe *et al.* (2020) propôs uma abordagem de aprendizado de máquina para prever a velocidade do navio usando o SIA e dados meteorológicos marítimos ao meio-dia. Para treinar e validar o modelo desenvolvido, foram utilizados os dados do SIA e meteorológicos marítimos das setenta e seis embarcações para o período de um ano.

Do exposto, pode-se deduzir que a modelagem de consumo de combustível dos navios é um campo de pesquisa ativo com várias abordagens diferentes sendo realizadas simultaneamente. Corroborar-se que, até o presente, a maioria dos estudos utiliza diferentes conjuntos de dados, com diferentes aquisições e particularidades de modelagem e outras suposições inerentes, como dados filtrados para clima adverso. Devido a essas inconsistências na literatura atual, é impossível identificar as abordagens de modelagem que produzem resultados ótimos para o problema comum de modelagem do consumo de combustível de uma embarcação.

3 METODOLOGIA

Em vista da proposta de construção de um modelo preditivo de consumo de combustível de embarcações que navegam na região amazônica, foi aplicada a metodologia de Descoberta de Conhecimento em Banco de Dados (DCBD) via *machine learning* (LAVALLE *et al.*, 2011; MANYIKA *et al.*, 2011; PANI, 2014).

Para a operacionalização do modelo através do *machine learning*, foi utilizado um banco de dados constituído por variáveis, a seguir detalhadas, obtido com base no Estudo de Caracterização da Oferta e da Demanda do Transporte Fluvial de Passageiros e Cargas na Região Amazônica, elaborado por docentes da Faculdade de Engenharia Naval da Universidade Federal do Pará para a Agência Nacional de Transporte Aquaviário no ano de 2017 (FIGUEIREDO *et al.*, 2018).

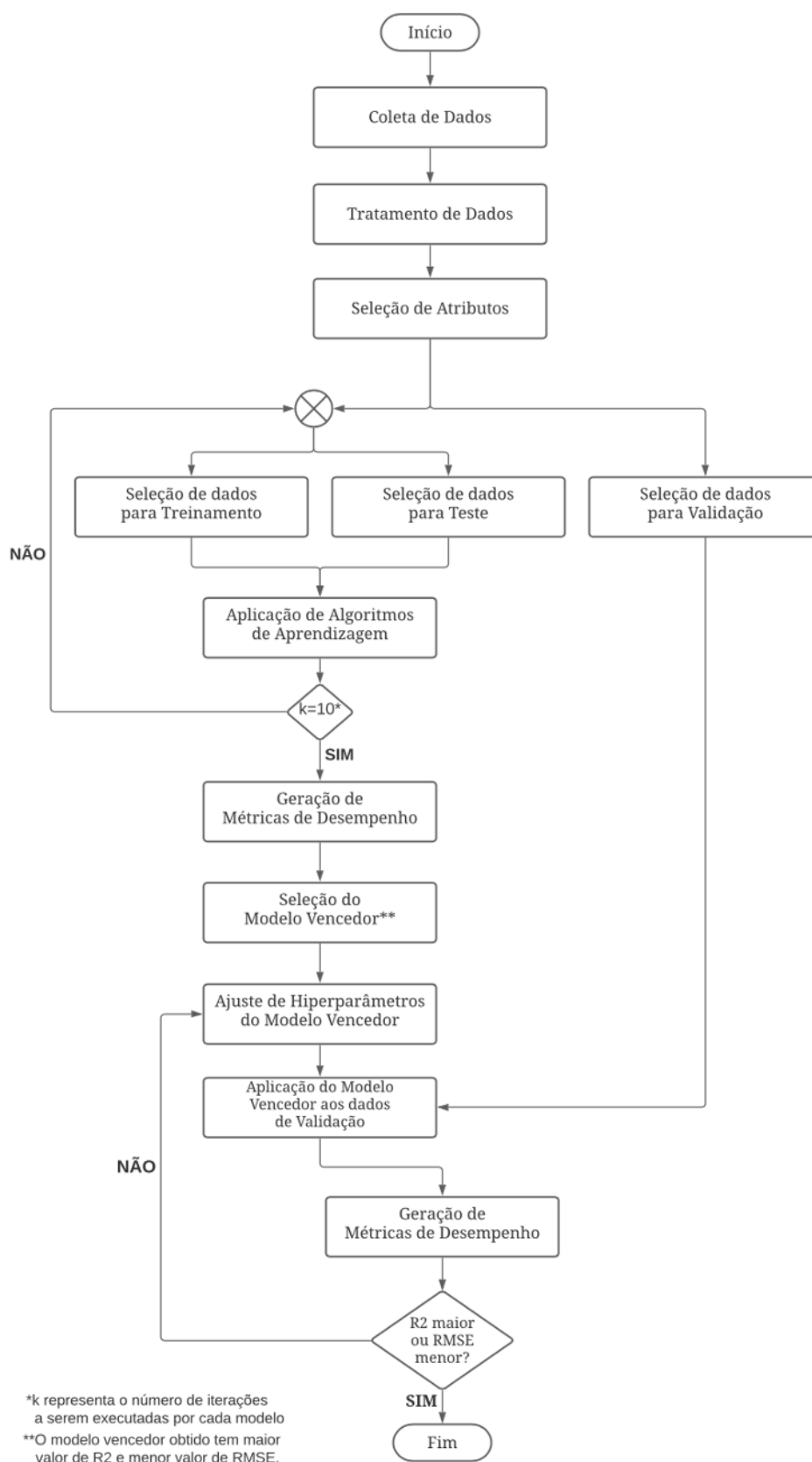
O modelo de predição desenvolvido, o qual apresentou como variável de saída uma

estratigrafia regional do consumo de combustível das embarcações que navegam na região amazônica, teve por base a técnica de *machine learning*, a qual teve sua estruturação pautada em características de projeto da embarcação e de desempenho operacional. A metodologia aplicada por Petersen *et al.* (2015), Wickramanayake e Bandara (2016), Gkerekos *et al.* (2019) e, principalmente, Abebe *et al.* (2020) proporcionou a fundamentação necessária para construção do modelo predição do consumo de combustível em embarcações a partir das variáveis contidas no banco de dados.

A construção do modelo de predição de combustível compreendeu as etapas de aquisição dos dados, aplicação de técnicas de pré-processamento das informações e seleção dos principais atributos do banco de dados. Foi abordado também o desenvolvimento e a implementação de seis modelos construídos a partir de algoritmos de *machine learning*, sendo avaliados através de métricas de desempenho que sustentaram na escolha do modelo de maior destaque quanto à capacidade preditiva. Para este, aplicou-se o ajuste dos hiperparâmetros visando um modelo cujas predições fossem mais assertivas.

Nota-se, através da Figura 5, o processo de desenvolvimento proposto, bem como as seções subsequentes trazem de maneira mais aprofundada cada etapa abordada.

Figura 5: Fluxograma de construção do modelo proposto.



Fonte: Autor (2023).

3.1 Aquisição de Dados

A aquisição de dados foi importante em termos de reconhecimento e interpretação do sistema (HU *et al.*, 2021). Na pesquisa, a etapa de aquisição de dados teve por base a coleta de dados necessários e suficientes à caracterização das variáveis no desenvolvimento do modelo preditor proposto, tais quais calado, boca, comprimento, etc. O “Estudo de Caracterização da Oferta e da Demanda do Transporte Fluvial de Passageiros e Cargas na Região Amazônica - ECTFA” (FIGUEIREDO *et al.*, 2019) serviu como fonte de coleta de informações e dados necessários ao desenvolvimento do modelo.

O ECTFA (FIGUEIREDO *et al.*, 2018) foi elaborado com o intuito conhecer e dimensionar a demanda de passageiros e misto (carga e passageiros) transportados pelos rios da Amazônia. Buscou-se ainda atualizar e verificar a evolução dos dados do transporte longitudinal de passageiros e cargas em embarcações mistas que navegam na região desde 2013, ano em que havia sido realizado o último estudo, além de aprofundar o diagnóstico das condições operacionais dos terminais e portos.

3.1.1 Coleta de Dados

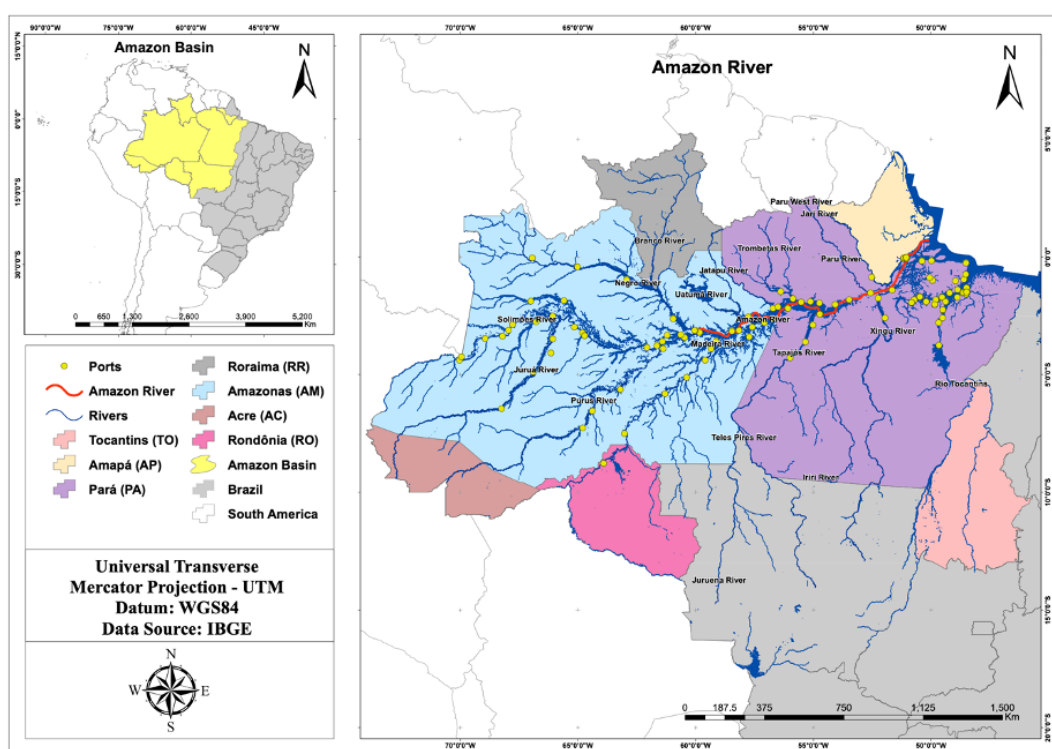
A base de dados foi composta de 1343 registros, para os quais foram identificados dezessete principais parâmetros de entrada para o modelo, originados dos vinte e nove parâmetros particulares iniciais referentes aos aspectos físico-construtivos, propulsivos, operacionais, funcionais e de sustentabilidade das embarcações coletados por meio do ECTFA (FIGUEIREDO *et al.*, 2019):

- a. Características físicas construtivas - ano de construção, tipo de embarcação, material do casco, comprimento, boca, calado;
- b. Parâmetros propulsivos: número de propulsores e geradores, potência do motor;
- c. Valores operacionais - tempo de viagem, tonelagem porte bruto, capacidade de passageiros, capacidade de carga, custo para transporte;
- d. Características funcionais das embarcações - tripulação, acomodação e salvaguarda;
- e. Ambientes das embarcações - quantidades de bares, cozinhas, porões, conveses; e
- f. Práticas sustentáveis das embarcações - existência de coleta seletiva de lixo, destinação adequada dos resíduos recicláveis e não-recicláveis, etc.

A categoria que apresentou o maior número de atributos foi a de valores operacionais com onze, sendo: tempo de viagem, deslocamento leve, velocidade de cruzeiro, distância entre origem e destino, tonelagem porte bruto, capacidade de passageiros, capacidade de carga, preço da passagem, periodicidade da manutenção, número de viagens, consumo de combustível. Em seguida, a categoria física-constructiva com oito atributos: ano de construção, ano da última reforma, tipo de embarcação, material do casco, comprimento, boca, calado e pontal. Por fim, ambientes das embarcações, parâmetros propulsivos e características funcionais.

A área de abrangência do ECTFA, conforme apresenta a Figura 6, compõe-se das principais Unidades da Federação potenciais de fluxo fluvial na Região Hidrográfica Amazônica, sendo: Pará (PA), Amapá (AP), Amazonas (AM) e Rondônia (RO). Foram coletadas informações sobre as linhas de transporte, a quantidade e as características das embarcações e dos terminais hidroviários que integram o sistema de transporte de passageiros e misto quanto à prestação dos serviços de transporte estadual, interestadual e de travessias na região amazônica.

Figura 6: Mapa do fluxo de passageiros total na região amazônica em 2017.

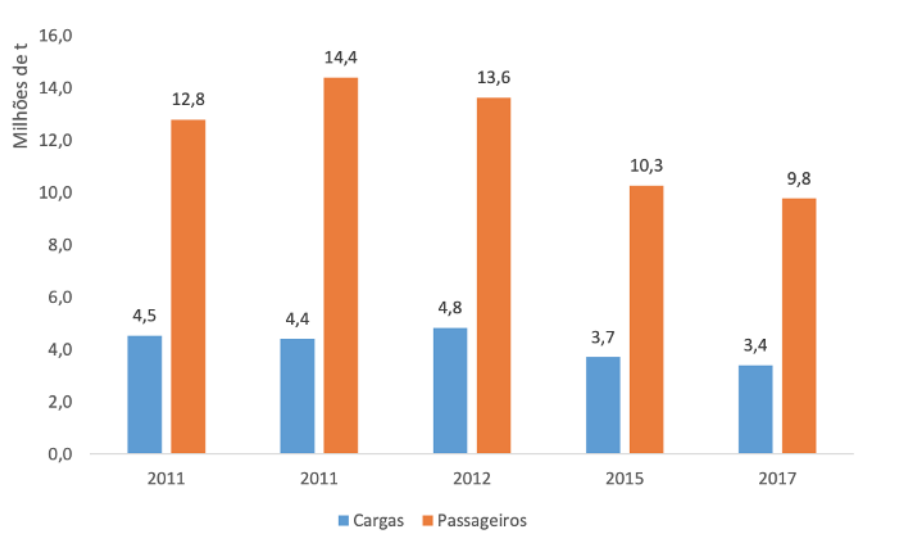


Fonte: Autor (2023).

Os dados levantados no estudo apontaram uma movimentação estimada em torno de 9,8 milhões de passageiros para o ano de 2017, abrangendo as linhas estaduais, interestaduais e de travessia. Em relação ao levantamento realizado pela UFPA em 2015, constatou-se uma redução da ordem de 6,6%. Quando comparadas as estimativas de 2017 com o histórico dos levantamentos pretéritos realizados pela Universidade, sendo dois em 2011 e um outro em 2012, tendo sido aplicados a mesma metodologia e público alvo, a redução registrada é ainda maior, cerca de 32%.

Acerca da demanda de cargas, o estudo estimou que em 2017 foram transportadas 3,4 milhões de toneladas pelas embarcações de passageiros e misto, uma redução de 7,6% em relação ao levantamento promovido pela UFPA em 2015. Para uma comparação entre os resultados obtidos em 2011 e 2017, houve uma diminuição de 30% na movimentação de cargas. Constata-se, por meio da Figura 7, a evolução das estimativas de movimentação de passageiros e cargas no transporte fluvial da Região Amazônica.

Figura 7: Evolução da estimativa do transporte de passageiros e cargas nos levantamentos realizados pela UFPA – 2011 a 2017.



Fonte: Autor (2023).

3.1.2 Banco de Dados

As características levantadas pelo ECTFA, principal instrumento do processo de coleta de dados, foram utilizadas para compor as entradas do modelo proposto. Para cada registro obtido, que representa uma embarcação levantada pelo Estudo,

adotou-se 29 parâmetros particulares, conforme a Tabela 1, referentes às características das embarcações de acordo com aspectos físico-construtivos, propulsivos, operacionais, funcionais e de sustentabilidade.

Tabela 1: Registros de embarcações do banco de dados utilizado.

	Sigla	Parâmetro	Unidade	Categoria
1	AC	Ano de Construção	Ano	
2	AUR	Ano da Última Reforma	Ano	
3	TE	Tipo de Embarcação	Passageiros, Carga Geral, Misto	
4	MC	Material do Casco	Madeira, Aço Naval, Fibra e Alumínio	Físico-
5	CT	Comprimento	Metro	Construtivo
6	BO	Boca	Metros	
7	CD	Calado	Metros	
8	PL	Pontal	Metros	
9	NP	Número de Propulsores	Metros	Parâmetros
10	NG	Número de Geradores	Metros	Propulsivos
11	PM	Potência do Motor	HP	
12	TV	Tempo de Viagem	Dias	
13	DL	Deslocamento Leve	Metros Cúbicos	
14	VC	Velocidade de Cruzeiro	Nós	
15	DOD	Distância entre Origem e Destino	Quilômetros	
16	TPB	Tonelagem Porte Bruto	Toneladas	Valores
17	CP	Capacidade de Passageiros	Passageiros	Operacionais
18	CG	Capacidade de Carga	Toneladas	
19	PP	Preço da Passagem	Real	
20	PMT	Periodicidade de Manutenção	Mensal, Trimestral, Semestral e Anual	
21	CC	Consumo de Combustível	Litros	
22	NV	Número de Viagens	Viagens/Mês	
23	NT	Número de Tripulantes	Tripulantes	Variáveis
24	AC	Acomodações	Unidades	Funcionais
25	SL	Salvaguarda	Unidades	
26	NB	Número de Bares	Unidades	
27	NC	Número de Cozinhas	Unidades	Ambientes da
28	NR	Número de Porões	Unidades	Embarcação
29	NV	Número de Conveses	Unidades	

Fonte: Autor (2023).

Segundo defendido por Dietterich (1997) & Batista (2003), a presença de muitos atributos relevantes prejudica o desempenho do modelo. Por esta razão, foi necessária uma análise prévia para seleção dos atributos relevantes disponíveis, culminando, como

consequência, na remoção dos preditores considerados duplicados ou redundantes, ou seja, quando uma informação essencialmente idêntica é armazenada em mais de um atributo (BATISTA, 2003).

Esta análise inicial dos dados foi proporcionada por meio da aplicação do método de seleção de atributos denominado “filtro” (OSHIRO *et al.*, 2012). Dessa maneira, todos os atributos levantados foram analisados e mantidos apenas aqueles relacionados às características de projeto da embarcação, desempenho operacional e condições ambientais, considerados parâmetros de influência para o consumo de combustível de uma embarcação, conforme abordado por Padovezi (2003), Gainza & Brinati (2010) e Schiller (2016).

Os dezessete parâmetros considerados preditores de entrada para construção do modelo podem ser vistos na Tabela 2.

Tabela 2: Registros de embarcações do banco de dados utilizado.

Entrada	Sigla	Parâmetro	Unidade	
1	TE	Tipo de Embarcação	Passageiros, Carga Geral, Misto	
2	MC	Material do Casco	Madeira, Aço Naval, Fibra e Alumínio	
3	CT	Comprimento	Metro	
4	BO	Boca	Metros	
5	CD	Calado	Metros	
6	PL	Pontal	Metros	
7	NP	Número de Propulsores	Metros	
8	NG	Número de Geradores	Metros	
9	PM	Potência do Motor	HP	
10	TV	Tempo de Viagem	Dias	
11	DL	Deslocamento Leve	Metros Cúbicos	
12	VC	Velocidade de Cruzeiro	Nós	
13	TPB	Tonelagem Porte Bruto	Toneladas	
14	CP	Capacidade de Passageiros	Passageiros	
15	CG	Capacidade de Carga	Toneladas	
16	NT	Número de Tripulantes	Tripulantes	
Variável-Alvo	17	CC	Consumo de Combustível	Litros

Fonte: Autor (2023).

Os dados ambientais inerentes às condições climáticas e estados de viagem não foram

considerados para compor as entradas do modelo proposto, haja vista não terem sido levantadas pelo ECTFA.

Os 17 parâmetros selecionados em meio às 29 características coletadas para cada uma das 1343 embarcações através do ECTFA subsidiaram a construção de uma matriz de registros, conforme indicado, por exemplo, pelos dez registros elencados na Tabela 3. Seguindo este padrão, o algoritmo identifica que cada linha representa uma embarcação, ou seja, um registro, e que cada coluna evidencia uma das características particulares apresentadas anteriormente pela Tabela 3, sejam elas de caráter numérico ou categórico, necessárias às fases posteriores de desenvolvimento do projeto.

A variável-alvo adotada para o modelo foi o Consumo de Combustível (CC) tendo em vista ser o objetivo do projeto proposto a predição do consumo de combustível em embarcações de transporte de passageiros e misto da região amazônica. Por esta razão, conforme argumentado por Padovezi (2003), Gainza & Brinati (2010) e Schiller (2016), os demais atributos, expostos na Tabela 3, estariam fortemente relacionados com o consumo de combustível apresentado pela embarcação, sendo assim, definidos como variáveis preditoras.

A identificação e remoção de atributos redundantes que, porventura, ainda permaneceram no sistema após a etapa de aquisição de dados, foi realizada por meio de métodos de seleção de atributos (UYANIK *et al.*, 2020).

Tabela 3: Registros de embarcações do banco de dados utilizado.

	TE	MC	CT	BC	CD	PL	PM	VC	NP	TPB	DL	NG	CP	CG	NT	TV	CC
1	Lancha	Alumínio	17	3,45	1,2	1,9	215	22	1	35	7	0	100	0	3	0,45	70
2	Ferryboat	Aço Naval	32	8	5	6	450	19	1	528	98	2	99	670	8	36	3000
3	Lancha	Aço Naval	25,6	6,3	1,8	2,2	315	18	1	21	6	2	90	0	2	24	2500
4	Lancha	Aço Naval	24,89	6,3	1,4	2,3	550	18	1	21	6	2	90	0	2	24	2500
5	Passageiro/Carga Geral	Madeira	19	5,3	1,8	2,8	200	10	2	109	21	2	65	25	3	7	300
6	Passageiro/Carga Geral	Madeira	24	6	2,15	2,8	315	15	3	109	21	2	88	34	3	8	400
7	Passageiro/Carga Geral	Madeira	32	8	2,4	3	400	20	3	109	21	1	126	288	7	40	2300
8	Passageiro/Carga Geral	Madeira	28	7,4	1,5	1,75	367	14	2	109	21	2	130	90	5	144	5000
9	Passageiro/Carga Geral	Madeira	17,6	4,2	1,8	1,8	612	15	1	109	21	1	60	19	2	6	180
10	Lancha	Alumínio	12	2,4	1	1,8	481	20	1	35	7	0	35	0	1	2	87

TE: Tipo de Embarcação; MC: Material do Casco; CT: Comprimento; BC: Boca; CD: Calado; PL: Pontal; PM: Potência do Motor; VC: Velocidade de Cruzeiro; NP: Número de Propulsores; TPB: Tonelagem Porte Bruto; DL: Deslocamento Leve; NG: Número de Geradores; CP: Capacidade de Passageiros; CG: Capacidade de Carga; NG: Número de Geradores; TV: Tempo de Viagem; CC: Consumo de Combustível.

Fonte: Autor (2023).

3.2 Tratamento de Dados

A etapa de tratamento dos dados da matriz de registros constantes na Tabela 3 compreendeu a adoção das seguintes etapas:

- a. Preenchimento dos dados faltantes;
- b. Transformação das variáveis categóricas;
- c. Remoção dos outliers; e
- d. Normalização.

O objetivo foi garantir que todas as informações estivessem organizadas dentro de um padrão geral, ou seja, sem a presença de valores nulos ou anormais, alcançando uma melhor interpretação por parte do modelo que, conseqüentemente, teria uma probabilidade maior de alcançar um alto desempenho.

3.2.1 Tratamento de Valores Desconhecidos

Para tratamento dos registros com valores nulos, ou seja, ausentes no conjunto de dados, mas existentes no contexto em que a medida foi realizada (UYANIK *et al.*, 2020), a saída mais viável para obtenção de dados persistentes foi a aplicação de duas técnicas: a) substituição pela média do atributo (PETERSEN *et al.*, 2015); LIANG *et al.*, 2019; e b) remoção do registro KHUN & JOHNSON, 2013; KWAK & KIN, 2017).

De acordo com a aplicação de Petersen *et al.* (2015), a média de cada atributo foi calculada a partir do somatório dos valores de cada registro dividido pelo número total de registros, conforme apresenta a Equação 1:

$$Média = \frac{1}{M} \sum_{n=0}^M X_n \quad (1)$$

Em que M é o número de registros utilizado e x_n é o valor apresentado pela amostra dado o atributo analisado.

Baseado em Khun & Johnson (2013), tendo em vista ser pequena a parcela de valores desconhecidos quando comparada à grande quantidade total de registros do banco de dados empregado no desenvolvimento do modelo, os valores desconhecidos, que não puderam ser preenchidos com a média, foram removidos do banco de dados. A aplicação

deste método condiciona-se a uma quantidade de valores desconhecidos inferior a 10% do valor total de registros contidos no banco de dados. Kwak & Kin (2017) sugerem a adoção do método de Análise de Caso Completa em que todos os valores desconhecidos são removidos.

3.2.2 Transformação de Variáveis Categóricas

A transformação das variáveis categóricas em variáveis numéricas se fez necessária para que pudessem ser aplicados os algoritmos de aprendizado (JIAN *et al.*, 2017). Por meio da aplicação do método de codificação 1-de-n ou one-hot (TAN *et al.*, 2013), os valores nominais de cada atributo categórico (Tipo de embarcação e Material do casco) foram transformados em novas colunas de atributos. Para cada nova coluna gerada a partir dos valores categóricos existentes para os atributos da Tabela 4, foi inserido o valor 1 apenas na coluna a qual o valor do atributo fosse referente ao registro, enquanto todas as demais foram preenchidas com 0, representando a característica nominal do atributo de maneira numérica (POTDAR *et al.*, 2017; YIN, 2019).

Tabela 4: Aplicação da técnica One-Hot para criação de novos parâmetros numéricos.

Nome da Embarcação	Madeira	Aço Naval	Alumínio	Fibra
Embarcação A	1	0	0	0
Embarcação B	0	0	1	0

Fonte: Autor (2023).

Dessa maneira, conforme exposto pela Tabela 4, nas etapas iniciais de análise dos dados, havia cerca de 17 atributos referentes aos registros elencados na Tabela 2. Por meio do método de transformação das variáveis categóricas em numéricas, este número aumentou para 25 atributos, ditos necessários para construção e desenvolvimento do modelo.

3.2.3 Remoção de Outliers

A detecção paramétrica dos pontos fora da curva também conhecidos como outliers foi executada por meio da Distância de Mahalanobis (DE MAESSCHALCK *et al.* 2000;

LEYS *et al.* 2018), método que se baseia na detecção de valores muito distantes do centróide da “nuvem” formada pela maioria dos dados, algo em torno de 99% das amostras. Conforme a definição original estabelecida por Tukey (1977) e reproduzida por Kwak & Kim (2017), um diagrama de caixa (ou box-plot) representa graficamente um lote de dados gerado a partir de algumas sínteses estatísticas que seriam a mediana e os quartis. A linha central no interior da caixa representa a mediana, ou seja, 50% do lote de dados.

A utilização do método da Distância de Mahalanobis e a plotagem dos registros, concernentes à cada um dos preditores, no formato de diagrama de caixa (*box-plot*) deu-se por conta da natureza não-linear a qual a distribuição dos dados estava disposta. Devido a este comportamento, impossibilitou-se o uso de histogramas para cumprir com detecção de pontos extremos, tendo em vista que seu uso é restrito para distribuições normais ditas gaussianas (JONES, 1994).

Por meio da mediana e do intervalo de quartis, adicionalmente à utilização de diagramas de caixa (box plot), os outliers puderam ser identificados, haja vista que quaisquer dados que encontravam-se fora das linhas das cercas superior ou inferior, calculadas a partir da mediana e do intervalo de quartil, eram considerados outliers (KWAK & KIM, 2017; LEYS *et al.*, 2018). Sendo assim, quando na ocorrência de um ponto extremo, conforme aplicado por Kwak & Kim (2017), optou-se pela remoção do registro para quaisquer que fossem os atributos.

3.2.4 Normalização

O processo de normalização foi empregado devido às diferenças de escala existentes entre os preditores de entrada considerados no projeto. Por esta razão, foram alterados os valores das colunas numéricas no conjunto de dados para usar uma escala comum, sem distorcer as diferenças nos intervalos ou ocasionar perda de informações (SINGH & SINGH, 2020).

Com isso, evitou-se que a variável-alvo, o consumo de combustível, pudesse ser enviesada negativamente, o que tornaria o estimador incapaz de aprender de maneira correta também com os demais parâmetros, além de realçar os dados considerados inconsistentes que, porventura, tenham permanecido, mesmo após a fase pretérita de remoção dos outliers.

Com esta finalidade, foi aplicada a técnica de normalização Z-score (PANDEY &

JAIN, 2017; GKEREKOS *et al.*, 2019; HAN *et al.*, 2022), na qual o valor normalizado de uma amostra x foi obtido através do quociente entre a diferença do seu valor original e a média dos registros considerados do atributo pelo desvio padrão de todas as amostras, conforme exposto na Equação 2:

$$Z_i = \frac{X_i - \mu}{\sigma}; \quad (2)$$

Em que μ é a média dos registros considerados; σ é o desvio padrão de todas as amostras; X_i é o valor de cada registro observado.

As pontuações Z resultantes foram definidas em um intervalo ilimitado de números negativos e positivos. Ao contrário dos valores normalizados, não houve mínimo e máximo predefinidos.

3.3 Seleção de Atributos

Para a seleção do conjunto mínimo de atributos, originados do banco de dados extraído do ECTFA, necessária para alcançar um critério de desempenho desejado, utilizou-se o Método de Seleção de Atributos (TARADEH *et al.*, 2019). Atributos que não apresentaram relação com a variável-classe foram considerados irrelevantes, ao passo que atributos com alta correlação em função às demais variáveis foram rotulados como redundantes, haja vista contribuir para uma diminuição de desempenho do modelo (MANJU *et al.*, 2019). Dessa maneira, a aplicação da técnica de Seleção de Atributos provocou uma redução de dimensionalidade, através da qual, atributos relevantes foram selecionados, enquanto que atributos irrelevantes e redundantes foram descartados (GUYON & ELISSEEFF, 2006).

Para desenvolvimento do sistema de predição do consumo de combustível das embarcações, na etapa de seleção dos atributos, adotou-se: elaboração da matriz de correlação; seleção dos parâmetros via Frequência de Pontuação; e definição da importância dos parâmetros via SHAP.

3.3.1 Método de Correlação de Pearson

Para investigação da correlação existente entre os parâmetros de entrada do modelo, foi aplicado o Método de Correlação de Pearson, de tal modo a apontar se forte ou fraca, bem como se os registros evoluíam, em função uns dos outros, de forma direta ou inversamente

proporcional. Sendo assim, quando na ocorrência de uma alta correlação entre diferentes variáveis, escolheu-se um único parâmetro para prosseguir à etapa de modelagem, processo também conhecido como engenharia de atributos.

Para determinação dessa relação entre preditores, calculou-se o coeficiente de correlação de Pearson (r), definido através da covariância entre duas variáveis dividida pelo produto de seus desvios-padrão, conforme apresenta a Equação 3, tendo sido aplicado por Pani (2014) e Brillante *et al.* (2015) e Schober, Boer & Schwarte (2018):

$$\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}; \quad (3)$$

Em que n é o tamanho do banco de amostras; x_i e y_i são os valores individuais dos registros indexados com i ; $\bar{x} = \frac{1}{n} \sum_{i=1}^n (x_i)$ é a medida da amostragem; e analogamente para y .

A aplicação do coeficiente de correlação permitiu retratar a força da associação linear entre dois parâmetros distintos característicos das embarcações por meio de valores que variam entre -1 e +1, haja vista ser um índice adimensional (UYANIK *et al.*, 2020). Consequentemente, convencionou-se que:

- A força do relacionamento entre as variáveis poderia admitir qualquer valor entre -1 e +1 e quanto mais próximo de um dos extremos, mais forte a correlação seria; assumiu-se ainda que, para a ocorrência de uma relação linear perfeita, a correlação apresentada deveria ser igual a -1 ou +1;
- Quando igual a zero, intentou-se não haver uma relação linear entre as duas variáveis em análise;
- Um valor positivo de coeficiente de correlação determinaria a existência de uma relação diretamente proporcional entre duas variáveis, ou seja, à medida com que houvesse o crescimento de uma destas, o comportamento da curva da outra variável também aumentaria; enquanto que, por outro lado, para um coeficiente negativo, os atributos seriam considerados inversamente proporcionais, o que, em outras palavras, quer dizer que, concomitantemente ao aumento de um atributo, o outro decresce).

Obtidos os coeficientes pertinentes às relações lineares entre os pares de atributos, para melhor identificação da intensidade de cada uma das correlações, foi gerado um

mapa de calor, combinando cores quentes para correlações próximas a +1 e cores frias ao à medida com que a correlação representava um valor perto de -1. O coeficiente de correlação de Pearson também foi representado na intersecção entre as colunas e linhas dos eixos horizontal e vertical, onde estão expostos os preditores, assim como aplicado nos estudos de Singh *et al.* (2021) e Xu e Deng (2017).

Embora a seleção dos atributos através do uso de coeficientes de correlação de Pearson tenha mostrado ser eficaz e de fundamental importância para conhecimento do tipo, se direta ou inversamente proporcional, bem como da significância do relacionamento existente entre os parâmetros em si quando em pares, é importante frisar que, quando referida em termos estatísticos, a correlação é um método para avaliação de uma possível associação linear bidirecional entre duas variáveis contínuas (LEE RODGERS, 1988).

Em síntese, tendo em vista que os registros deste projeto não se encaixam na definição de distribuição normal, mas sim, apresentam distribuição não-linear, a seleção de atributos por meio dos coeficientes de correlação de Pearson não pôde ser tomada como única e exclusiva técnica para o devido fim (XU & DENG, 2017).

3.3.2 Método de Classificação de Atributos via F-Score

A alternativa mais viável encontrada para contornar este problema, então, foi empregar, adicionalmente à correlação de Pearson, a técnica de Classificação de Atributos (Feature Ranking), para a qual cada atributo recebe uma classificação de acordo com o peso associado a ele (MANJU *et al.*, 2019).

Segundo Manju *et al.* (2019), o peso do atributo é equivalente ao número de vezes que o mesmo aparece quando na construção da estrutura da árvore de decisão. Sendo assim, os atributos foram ranqueados em ordem decrescente, definida pelo número de vezes que um dado atributo era selecionado para divisão da árvore durante o treinamento do modelo (TARADEH *et al.*, 2019). Em outras palavras, quanto mais o parâmetro era empregado na tomada de decisão para “divisão” do nó raiz, maior era a sua relevância para explicar o comportamento dos padrões dos dados utilizados neste projeto e, por esta razão, ocupava os lugares mais altos do ranking.

Por meio da função *feature_importances_* disponível pela biblioteca *scikit-learn* via linguagem de programação *Python 3.6*, as pontuações de importância relativa à cada um dos preditores puderam ser calculados e apresentados em um gráfico de barra, cujas

colocações foram organizadas através do método Pontuação por Frequência (ou F-Score), aplicado por Dao *et al.* (2018) e Nguyen *et al.* (2019), pelo qual elencou-se de maneira crescente os atributos em função da porcentagem da frequência com que os atributos foram usados para divisão das árvores, baseando-se na definição de separação pela impureza quando aplicada a técnica de índice de Gini, definida pela Equação 4 (TAYEFI *et al.*, 2017; KIM, 2016).

$$Gini(D) = 1 - \sum_{i=1}^m P_i^2 \quad (4)$$

Para o qual, D representa as amostras do banco de dados e P_i é a proporção do número de amostras relacionadas ao atributo i em D.

3.3.3 Método SHAP

A aplicação do Método SHapley Additive exPlanations (SHAP), proposto para interpretação da importância das variáveis de entrada do modelo (LUNDBERG & LEE, 2017) determinou a ordenação dos atributos de maneira crescente por grau de importância, exibindo ainda a influência de cada amostra do banco de dados em relação à variável-alvo (o consumo de combustível) conforme defendido Ribeiro *et al.* (2016) e Marcílio-Jr & Eler (2021) em relação às interpretações locais de cada atributo. A contribuição de cada atributo (Φ_i é a contribuição do atributo i) na saída do modelo $E(F)$ foi alocada com base em sua contribuição marginal (SHAPLEY, 1953). Baseado em vários axiomas de modo a alocar de forma justa a contribuição de cada recurso, os valores de SHAP são determinados por meio da Equação 5 a seguir:

$$\Phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \{E[f(X)|X_{S \cup i} = x_{S \cup i}] - E[f(X)|X_S = x_S]\} \quad (5)$$

Em que F é o conjunto de atributos de entrada do modelo; S denota um subconjunto; $S \cup i$ é a união entre o subconjunto S e o atributo i ; $E[f(X)|X_S = x_S]$ é a expectativa condicional do modelo $f(\cdot)$ quando um subconjunto S de atributos é fixado no ponto local x .

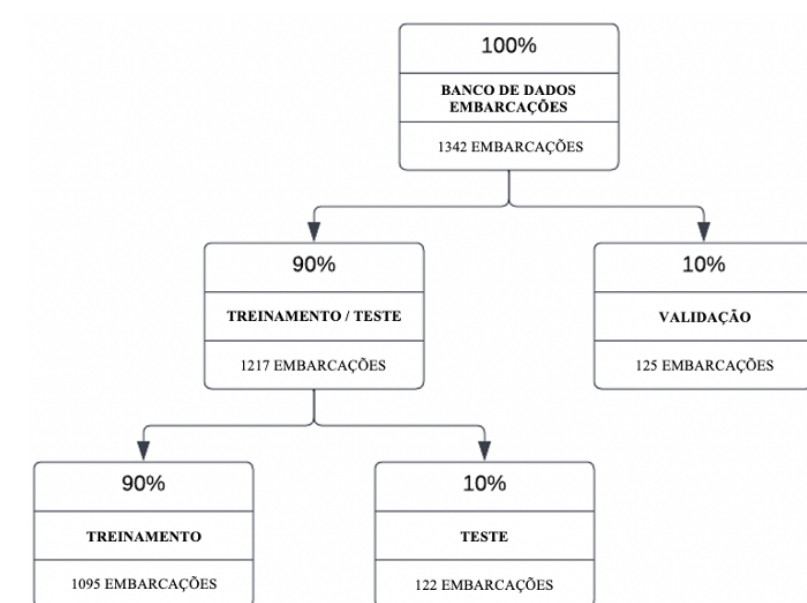
A partir dos valores de SHAP gerados, foram empregadas cores frias para representar valores pequenos, ao passo que cores quentes para registro dos valores grandes. Consecutivamente, a representação destes dados sobre o eixo x, ocorreu do seguinte modo: à esquerda do valor médio de SHAP, registros de fraca correlação com a variável-alvo, enquanto que à direita, aqueles que tivessem uma forte correlação.

3.4 Seleção de Algoritmos e Construção dos Modelos

Através de uma abordagem clássica de *machine learning*, empregada por Izbicki & Santos (2019), o banco de dados foi dividido em dois conjuntos: treinamento e validação. No total, 1342 registros caracterizados por 29 parâmetros foram selecionados para compor a entrada do modelo, a partir dos quais criou-se duas partições: uma para treinamento, contendo 90% do conjunto de dados, e outra, com os outros 10% das amostras restantes, reservada exclusivamente para validação do modelo vencedor por meio da técnica de hold-out, etapa adicional na qual o modelo escolhido deve ser capaz de prever amostragens aleatórias, até então, não vistas anteriormente (BREIMAN, 2001).

Inicialmente, a técnica de validação cruzada k-fold foi aplicada sobre o conjunto de dados de treinamento, tendo sido dividido em dez subconjuntos de amostras, a partir dos quais, sete conjuntos foram destinados para treinamento, enquanto que o restante para teste, alternando entre si até que todos os subconjuntos tenham sido usados tanto para teste quanto para treinamento (HASTIE; TIBSHIRANI; FRIEDMAN, 2009), ou seja, o modelo será executado k vezes iterativamente. Durante cada iteração de uma configuração de hiperparâmetros combinado, vários resultados de precisão do modelo foram obtidos e calculados. A Figura 8 demonstra a distribuição proposta dos registros.

Figura 8: Distribuição dos dados para treinamento, teste e validação hold-out.



Fonte: Autor (2023).

Conforme apresentado na Figura 8, 1342 amostras foram utilizadas para treinamento, sendo, mais especificamente, 1095 embarcações para treinamento do modelo e outras 122 para teste. Restantes para a etapa adicional de validação via Método de Resistência (hold-out), exclusiva ao modelo vencedor, foram apontadas outras 125 embarcações. É importante ainda ressaltar que os registros pertencentes à cada partição foram escolhidos de maneira aleatória em que o banco de dados foi dividido sem que houvesse um critério avaliativo para tal, tendo em vista a capacidade de generalização esperada acerca do modelo escolhido (HUANG *et al.*, 2019).

O desenvolvimento do modelo, de fato, teve início ao passo que foram treinados, por meio da programação em Python da biblioteca *scikit-learn*, os seis algoritmos do tipo CART escolhidos: Árvore de Decisão (AD), Florestas Aleatórias (FA), Árvores Extras (AE), Gradient Boosting Regressor (GBR), Extreme Gradient Boosting Regressor (XGBoost), CatBoost Regressor.

3.4.1 Árvore de Decisão de Regressão (ACR)

No projeto proposto, utilizou-se o método de Árvores de Classificação e Regressão (ACR), um dos métodos mais comuns para métodos de regressão baseados em árvore de decisão. Em uma ACR, após a escolha do ponto de divisão ideal para obter o melhor ajuste do modelo definido através do Critério de Gini, o espaço do parâmetro é dividido em dois caminhos, processo executado recursivamente até que as regras de parada sejam alcançadas e obtida a predição final.

As árvores de decisão apresentam-se como um algoritmo de *machine learning* em que, a primeiro momento, são utilizados os valores conhecidos das observações para que, em seguida, o modelo seja capaz de prever os dados presentes no banco de teste, onde os valores das observações são desconhecidos (BREIMAN, 2001).

Trata-se de um método de aprendizagem supervisionada não paramétrico de regressão na forma de uma estrutura de árvore com nós e ramos. Para o desenvolvimento do ACR, os parâmetros foram particionados, ao passo que um modelo simples (árvore) foi treinado para cada um, utilizando um conjunto de dados de treinamento em uma faixa contínua. Como resultado do processo, obteve-se o valor médio das observações dos conjuntos de treinamento que estão localizados no mesmo nó (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Ainda segundo os autores, para desenvolvimento do modelo, propõe-se um número n de amostras do banco de dados e d diferentes parâmetros, conforme Equação 6:

$$D\{(x_i; y_i)\}(|D| = n, x_i \in R)^d, y_i \in R \quad (6)$$

Em que D é o conjunto do número de amostras, ao passo que sua magnitude tende a ser o próprio valor de n ; x_i representa a variável de entrada dos recursos disponíveis e deve estar contido no intervalo dos valores reais de d ; quanto ao valor de y_i , variável de saída, representa o resultado obtido pelo modelo e que deve ser melhorado com base em cada iteração realizada pelo algoritmo.

O conjunto de parâmetros disponível foi, então, dividido em um número de regiões K , chamado de R_K , e o valor de predição do modelo foi obtido por meio do valor médio da observação encontrada na região de valor K -ésima, ou seja, na última região. Em outras palavras, o valor médio da região K foi encontrado ao longo do desenvolvimento dos valores nas regiões anteriores. A combinação não-linear envolvendo o vetor de regressão é dada pela Equação 7:

$$\hat{y}_i = \text{media}(y_i | x_i \in R_k); \quad (7)$$

O valor correspondente a uma predição mais concisa com base nos dados observados, ou seja, \hat{y}_i ideal, foi alcançado em virtude da minimização do erro quadrático mínimo, σ , acarretando em uma performance de maior rendimento do modelo. Por esta razão, a minimização de σ indica a melhoria de \hat{y}_i , definida através da Equação 8:

$$\sigma = \sum (y_i - \bar{y}_i)^2; \quad (8)$$

Os valores σ podem ser simplesmente calculados, no entanto, é alto o esforço para divisão da região. Para superar isso, foi aplicado um algoritmo recursivo para determinar os nós de divisão ideais até que o ponto de parada fosse alcançado. Normalmente, isso depende dos hiperparâmetros e da dificuldade do problema fundamental. Os hiperparâmetros selecionáveis do modelo em estudo baseados em Abebe *et al.* (2020) são:

- A profundidade máxima da árvore (*max_depth*) que indica a profundidade da construção da árvore. Salienta-se que quanto mais profunda a árvore, maior número de divisões (*splits*) haverá, capturando mais informações acerca dos dados;

como consequência, aumentando a profundidade, o tempo computacional para processamento dos dados e treinamento do modelo também crescerão;

- *min_samples_split* representa o número mínimo de amostras necessárias para a divisão de um nó interno. Este item varia entre considerar apenas uma amostra em cada nó e considerar todas as amostras em cada nó. Quando este parâmetro é aumentado, a árvore é “forçada” a considerar mais dados em cada nó.
- O número mínimo de amostras requerido em cada nó folha (*min_samples_leaf*). Muito semelhante ao *min_samples_splits*, de qualquer modo, para descrever o número mínimo de amostras presentes nos nós folhas.
- Adicionalmente, o número de recursos (*max_features*) a serem considerados ao pesquisar a melhor divisão deve ser especificado.

3.4.2 Método de Conjunto

O conceito básico no uso do Método de Conjunto é o desenvolvimento de um modelo preditivo formado a partir da integração de modelos simples para obter um modelo de melhor desempenho (ABEBE *et al.*, 2020). Duas técnicas de Método de Conjunto foram utilizadas no projeto proposto, denominadas: Aumento (*boosting*) e Ensacamento (*bagging*). Ensacamento (*bagging*) é um método que visa a formação de um único e inclusivo modelo de conjunto a partir da média de diferentes regressões geradas por diferentes modelos, a fim de reduzir a variância. Dessa forma, o ensacamento pode ser aplicado para um modelo com alta variância e baixo viés.

Ao contrário do ensacamento, o Aumento (*boosting*) é capaz de gerar um modelo de conjunto a partir de um único modelo, como ocorre com o algoritmo de Árvores de Decisão. Define-se como uma técnica sequencial que integra um conjunto de classificadores fracos, definidos como árvores de decisão simples que, se usadas por conta própria, preveriam a classe de observações com baixa precisão (HOLLOWAY *et al.*, 2019). O objetivo é tornar mais precisa a estimativa resultante do modelo tendo em vista que os novos resultados do modelo desenvolvido têm pesos baseados nas saídas anteriores do modelo. Se as saídas forem previstas corretamente, um peso menor será atribuído; caso contrário, o peso atribuído será maior (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Florestas Aleatórias (FA) O algoritmo de floresta aleatória, proposto por Breiman (2001), é um método de árvore de decisão que usa a técnica de ensacamento (*bagging*) para reduzir a variância, calculando a média de muitas árvores de decisão imprecisas, mas aproximadamente imparciais (HASTIE; TIBSHIRANI; FRIEDMAN, 2008).

Para a construção de um modelo de Florestas Aleatórias, uma série de Árvores de Decisão Regressoras (ACR) independentes ($n_estimators$) são geradas utilizando a base de dados de treinamento disposta. A resposta do modelo é determinada pela média dos resultados individuais das árvores de decisão, conforme exposto pela Equação 9:

$$\hat{y}_i(x) = \frac{1}{M} \sum_{m,i=1}^M f_m(x_i); \quad (9)$$

Em que M é o número de árvores de decisão do modelo $n_estimators$.

Para construir uma árvore de decisão, o método utiliza dados escolhidos a partir de uma amostragem aleatória do conjunto de dados de treinamento com reposição, ou seja, mesmo que um registro já tenha sido escolhido para treinamento do modelo, é permitido que possa ser escolhido de novo para treinamento de outra árvore (BREIMAN, 2001).

Além disso, em cada ponto de divisão, ao invés de considerar todas as variáveis no conjunto de dados de treinamento, cada árvore cresce pesquisando a melhor divisão a partir das variáveis disponíveis, escolhidas aleatoriamente pelo modelo. Assim como o conjunto de dados de treinamento, para cada nova árvore construída, são escolhidos aleatoriamente diferentes atributos. Com isso, espera-se que, como consequência, o crescimento das árvores aconteça independente e diferente umas das outras.

Na implementação do *scikit-learn*, semelhante aos ACRs, o número mínimo de amostras necessárias para dividir um nó interno é controlado pelo parâmetro *min_samples_split*.

Árvores Extras (AE) Semelhantemente ao algoritmo de Florestas Aleatórias (FA), o método de Árvores Extras (AE) tem como princípio o desenvolvimento de um conjunto de árvores de decisão do tipo regressão, partindo de uma estrutura padrão contendo sequencialmente nós raiz, filho e folha, construídos de cima para baixo da árvore. Entretanto, para um modelo de Árvores Extras, os pontos de corte selecionados para divisão dos nós das árvores são extremamente aleatórios, além de não aplicar uma subamostragem do banco de dados de treinamento: todas as amostras de treinamento

são usadas (GEURTS *et al.*, 2006).

Quanto às variáveis numéricas, o procedimento de divisão dos nós pelo método de Árvores Extras possui dois parâmetros básicos, sendo o número de atributos a serem escolhidos aleatoriamente em cada nó e o tamanho mínimo da amostra necessário para divisão de um nó. Consecutivamente, são gerados os modelos preditivos das árvores individuais, e o resultado final adotado é igual à média das regressões de todas as árvores construídas, assim como ocorre no desenvolvimento de uma FA.

Os hiperparâmetros básicos adotados por este algoritmo são o número de atributos, para garantir a eficácia do procedimento de seleção dos atributos; o tamanho mínimo da amostra, para suavizar a influência do ruído (dados incorretos) resultante; e o número de árvores, visando a redução da variância da combinação do modelo de conjunto.

Em uma implementação do *scikit-learn*, os hiperparâmetros são semelhantes àqueles utilizados em uma Árvore de Decisão, incluindo também o parâmetro referente ao número de árvores *n_estimators* na floresta. Normalmente, um número maior de árvores treina melhor os dados. No entanto, adicionar muitas árvores pode retardar consideravelmente o processo de treinamento, portanto, uma pesquisa paramétrica para encontrar a configuração ideal é necessária.

Aumento de Gradiente (AG) O modelo de Aumento de Gradiente (AG) tem como fundamento o método de Aumento (*boosting*), que produz um modelo preditivo eficaz a partir de um conjunto de modelos de regressão fracos, geralmente usando árvores de decisão (FRIEDMAN, 2002). Um novo regressor é construído adicionalmente ao modelo em diferentes estágios, permitindo a otimização das funções de perda (*loss function*). No desenvolvimento de um modelo de Aumento de Gradiente, um conjunto de árvores usa M funções aditivas para estimar a saída, de acordo com a Equação 10:

$$\hat{y}_i(x) = \sum_{m=1}^M f_m(x_i), f_m \in \zeta. \quad (10)$$

Conforme visto na Equação 11, ζ denota o domínio da função que inclui todas as árvores de regressão:

$$\zeta = \{f(x) = w_q(x), w \in \mathbb{R}^T, q : \mathbb{R}^d \rightarrow T\} \quad (11)$$

Em que q denota a estrutura de cada árvore que mapeia o correspondente índice da folha; T denota o número de folhas em uma árvore. Cada f_m corresponde à uma estrutura

de árvore independente, q , e a um peso de folha, w . Diferente das ADRs, cada árvore de regressão contém uma pontuação contínua em cada folha, sendo w_j a pontuação da respectiva folha j th. O peso da folha é calculado pela minimização da função de perda, vide Equação 12:

$$\Omega = \sum_i l(\hat{y}_i, y_i) + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (12)$$

Para o qual, l representa a função de perda diferenciável que mede a diferença entre o valor predito, \hat{y}_i , e a amostra, y_i . λ denota um valor constante de regularização para penalizar a complexidade do modelo, e o w_j ótimo pode ser obtido por meio de uma aproximação da série de Taylor de segunda ordem da Equação 13.

$$w_j = \frac{\sum_{i \in I_j} \frac{\partial l(y_i, \hat{y}_i)}{\partial (\hat{y}_i = 0)}}{\sum_{i \in I_j} \left(\frac{\partial^2 l(y_i, \hat{y}_i)}{\partial (\hat{y}_i = 0)^2} \right) + \lambda} \quad (13)$$

Em que I_j é o banco de dados contido em uma folha j .

Na implementação do *scikit-learn*, um AGR também tem os mesmos hiperparâmetros principais que um ACR com a adição de $n_estimators$ e a taxa de aprendizagem que podem ajudar o modelo a diminuir a contribuição de cada árvore.

Aumento Extremo de Gradiente (AEXG) O Aumento Extremo de Gradiente (AEXG) fornece hiperparâmetros de regularização adicionais, conforme apresenta a Equação 14, o que contribui para a redução de ocorrência do sobreajuste (*overfitting*) do modelo, além de diminuir a variabilidade das predições, aumentando a sua precisão. A saída predita y_i é obtida minimizando a função de regulação Φ :

$$\Phi = \sum_i l(\hat{y}_i, y_i) + \sum_m \Omega(f_m) \therefore \Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 + \alpha |w| \quad (14)$$

Para os quais Ω representa o parâmetro de regularização que penaliza a complexidade do modelo como funções de árvore de regressão e suaviza os pesos finais aprendidos de modo a evitar o sobreajuste (*overfitting*). T representa o número de nós folha e w é a pontuação do nó folha. γ , λ e α são parâmetros que definem o grau de regularização.

Em relação aos parâmetros λ e α , também definidos para regularização L_1 e L_2 , respectivamente, têm diferentes influências no peso; α induz o peso aprendido a ser zero, enquanto que λ intervém para que o peso seja pequeno. λ é um hiperparâmetro de pseudo-regularização, denominado multiplicador de Lagrange, que controla a complexidade de

uma determinada árvore e especifica a redução de perda mínima necessária para fazer mais partições em um nó folha, o que significa que, quanto mais alto for seu valor, menor será o número de divisões geradas. Além do uso de um termo de regularização, a subamostragem do preditor foi usada para evitar o sobreajuste (CHEN & GUESTRIN, 2016).

Durante o processo de predição, os resultados de cada árvore são somados para que sejam obtidos os resultados finais do modelo. Os parâmetros de cada árvore (f_t), que inclui a estrutura da árvore e as pontuações obtidas por cada nó folha, deve ser determinado. Por meio do método de treinamento aditivo, o resultado de uma árvore é adicionado ao modelo em determinado momento. O valor predito ($\hat{y}_i^{(t)}$) obtido na etapa t pode ser utilizado no processo de desenvolvimento do algoritmo, conforme Equação 15:

$$\Phi \hat{y}_i^{(t)} = \sum_{m=1}^M f_m(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (15)$$

Quando na configuração dos parâmetros iniciais do modelo, cada observação foi prevista como a média de todas as variáveis de resposta observadas, determinando um peso igual para todos. O ajuste dos pesos ocorreu ao longo de cada iteração, onde quanto mais uma previsão falhava, maior o peso desse dado observado naquela árvore, haja vista que o algoritmo foi forçado a focar nessa observação.

Conseqüentemente, as árvores adicionadas em sequência foram treinadas a partir das variações do conjunto de registros não “reconhecidos” adequadamente pelo modelo. O critério de parada quanto à criação de novas árvores foi acionado quando no número máximo de árvores ou quando as predições não estavam mais a contribuir com avanços consideráveis, ou seja, a capacidade do modelo em predizer novos registros não apresentava mais avanços significativos. Em uma implementação de *scikit-learn*, os parâmetros adicionais em AEXGs são γ , λ e α , como mencionado acima. Esses parâmetros de regularização limitam o quão extremos os pesos (ou influência) das folhas em uma árvore podem se tornar.

Catboost CatBoost é uma biblioteca de software de código aberto que possui capacidade de tratamento de dados categóricos. Na fase de treinamento do modelo, diferente dos demais algoritmos, não há a substituição binária de valores categóricos, ocorrendo, em contrapartida, a execução de uma permutação aleatória do conjunto de dados e cálculo do valor médio da predição resultante que deve ser comparado ao

mesmo valor da categoria quando colocado antes do dado de permutação. Com essa forma de utilização das variáveis categóricas, CatBoost reduz o sobreajuste (overfitting) (PROKHORENKOVA *et al.*, 2018).

Para lidar de maneira eficaz com a presença de parâmetros categóricos, o algoritmo CatBoost utiliza o método de Estatísticas Alvo (EA), semelhante à codificação média, em que acontece a substituição do atributo y_k^i referente à iteração K por um equivalente numérico igual à estatística da variável-alvo. A estimação do valor esperado é condicionado pelo atributo, de acordo com Prokhorenkova *et al.* (2018) e está apresentado pela Equação 16:

$$\hat{y}_k^i \approx E(y|y^i = y_k^i) \quad (16)$$

Para transformação da variável categórica em numérica, de acordo com a Equação 17, o valor alvo médio é, então, determinado através de um estimador baseado nas amostras de mesma categoria que y^i , correspondente à amostra de valor K. Dessa maneira, o residual é suavizado em virtude de determinado p anterior através do aprendizado, com $a > 0$.

$$\hat{y}_k^i = \frac{\sum_{j=1}^n l_{y_j^i=y_k^i} \cdot y_j + ap}{\sum_{j=1}^n l_{y_j^i=y_k^i} + a} \quad (17)$$

Em que o valor p é definido pela média do valor alvo sobre a amostra, a identificado como um parâmetro que desempenha função do grau de regularização. Quando na construção de uma nova divisão para a árvore, o algoritmo Catboost utilizou-se da estratégia de não usar uma variável formada através da combinação de atributos como parâmetro de divisão da árvore. Em contrapartida, para o próximo corte, ocorreu a combinação de todos os atributos categóricos que, posteriormente, foram dinamicamente convertidos em números por meio da técnica de Estatísticas Alvo (EA), também aplicado por Fei, Fang e Ling (2020).

3.5 Otimização de Hiperparâmetros

Existem vários hiperparâmetros em um modelo e a alteração nos valores dos hiperparâmetros pode afetar o desempenho do modelo construído (ABEBE *et al.*, 2020). Uma vez que os valores ideais do hiperparâmetro não são identificados inicialmente, a otimização deve ser realizada para selecionar os valores adequados para cada modelo.

Para o projeto proposto, referente à Bergstra *et al.* (2011), aplicou-se a Busca Aleatória (Random Search) como técnica de otimização de hiperparâmetros, na qual os

valores dos hiperparâmetros são escolhidos aleatoriamente, a partir de uma distribuição normal, dado um específico determinado espaço de hiperparâmetro (PROBST; WRIGHT; BOULESTEIX, 2019). Inicialmente, é definida manualmente uma faixa de limites dos parâmetros possíveis e o algoritmo faz uma pesquisa sobre eles para o número de iterações pré-determinadas.

Conforme aplicado por Bergstra & Bengio (2012), quando em um problema de otimização hiperparamétrica, o método de Busca Aleatória gera, como amostra, valores candidatos para cada hiperparâmetro (i) do algoritmo de aprendizado, para este caso de regressão, por meio da definição de uma função de densidades. Dessa forma, consiste em uma amostragem seguida pelo treinamento e avaliação do modelo utilizando o vetor amostrado. Estas duas tarefas acontecem em ciclo, de maneira recursiva, até que uma condição de parada ocorra, seja esta pelo número máximo de treinamentos ou pelo tempo de processamento.

3.6 Aplicação de Métricas de Desempenho

A partir da aplicação da técnica de validação cruzada para particionamento do espaço amostral em treinamento e teste, na qual alternou-se, a cada rodada, a porção a ser utilizada, foi necessária a avaliação do desempenho para estes cenários, tendo em vista a diversidade dos registros contidos no banco de dados, mensurando a capacidade de extração dos padrões e predição da variável-alvo pelos algoritmos propostos (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Para tal, tornou-se mais vantajoso examinar diferentes métricas de desempenho para que fossem proporcionadas múltiplas perspectivas acerca dos aspectos qualitativos de cada modelo. O objetivo, dessa maneira, era de contribuir para a tomada de decisão quanto ao modelo que mais se adequa à aplicação exigida, seja no treinamento, teste ou até mesmo na validação, este último para o caso do algoritmo vencedor (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

Dessarte, as métricas de desempenho escolhidas com base nos critérios de maior utilização e nos objetivos de cada métrica foram: Erro Médio Absoluto (EMA), Quadrado Médio do Erro (QME), Erro Quadrático Médio (EQM) e Coeficiente de Determinação (R^2).

3.6.1 Erro Médio Absoluto (EMA)

O Erro Médio Absoluto (EMA) foi calculado a partir da média dos erros obtidos em um intervalo de valores de previsões. Em outras palavras, foi empregada a média aritmética das diferenças absolutas entre o valor predito e o valor real, para que, dessa forma, fosse possível a mensuração da proximidade entre a previsão e os resultados reais (TOQUÉ *et al.*, 2017; PANI, 2014).

O principal objetivo em seu uso deu-se em estabelecer, através de uma magnitude, comparações entre previsões e resultados finais obtidos. Vale enfatizar que cada erro cometido ao longo do processo de previsão contribuiu para o EMA de forma proporcional.

Sendo assim, relacionou-se os valores dos erros gerados pela diferença entre o valor predito e o valor real através de uma média aritmética, em função do tamanho do conjunto de previsão, conforme segue na Equação 18, explicitada por Pani (2014), Kitsikoudis (2014), Toqué *et al.* (2017) e Fan *et al.* (2018).

$$EMA = \frac{1}{n} \sum |y_j - \hat{y}_j| \quad (18)$$

Em que \hat{y} é definido como o valor predito; y é determinado como o valor real da amostra; e n , descrito como o número de vezes em que a diferença entre os termos foi calculada.

3.6.2 Quadrado Médio do Erro (QME)

De forma semelhante ao Erro Médio Absoluto (EMA), foi considerada uma medida de tendência central pela qual fosse mensurada a qualidade da previsão através de um valor calculado. Determinou-se, por meio do Quadrado Médio do Erro (QME), uma medida da qualidade da previsão através da comparação do valor da diferença média quadrática entre o valor real e o valor predito, conforme abordado pela Equação 19 (CORADDU *et al.*, 2017; DAWOOD, 2019):

$$QME = \frac{1}{n} \sum (y_j - \hat{y}_j)^2 \quad (19)$$

Em que y é o valor real da amostra; \hat{y} é o valor predito pelo modelo; e n é o número de vezes em que a diferença entre os termos foi calculada. O principal objetivo em sua utilização deu-se em virtude da busca pelo melhor resultado do preditor, caracterizado por um valor mais próximo de zero possível, indício de uma previsão eficaz, além de nunca poder assumir um valor negativo (JEON *et al.*, 2018).

3.6.3 Erro Quadrático Médio (EQM)

Para determinação da acurácia, visando a comparação de erros existentes nas previsões de diferentes modelos aplicados em um conjunto de dados em particular, o Erro Quadrático Médio (EQM) foi empregado no projeto proposto (ALAWADI *et al.*, 2019; LIANG *et al.*, 2019).

O principal propósito no uso deste indicador de desempenho deu-se pela necessidade em indicar quão próximas as estimativas do modelo estavam em relação aos valores das amostras originais. Quanto menor seu valor, mais ajustado estava o modelo, demonstrando uma ótima performance do mesmo (IZBICKI & SANTOS, 2019).

Definido como a raiz quadrada da relação entre o desvio padrão, obtido através do quadrado da diferença entre o valor predito e o valor original, e o número de amostras do conjunto de dados, conforme apresentado pela Equação 20. Tais discrepâncias entre previsões e as amostras, quando ocorridas na fase de treinamento, foram determinadas como residuais, enquanto que, quando ocorridas na fase de teste, reconhecidas como erros de previsão.

$$EQM = \sqrt{\frac{\sum_{i=1}^n (y_j - \bar{y}_j)^2}{n}} \quad (20)$$

Em que X_0 representa os valores das amostras; X_m representa os valores preditos pelo modelo.

Devido ao uso do quadrado dos residuais, diferenças muito grandes foram penalizadas de maneira severa, o influenciando no treinamento do modelo ao atribuir mais peso a este tipo de atributo de maior magnitude em detrimento aqueles de menores proporções.

3.6.4 Coeficiente de Determinação (R^2)

O coeficiente de determinação (R^2) teve por função quantificar a correlação entre as probabilidades obtidas diretamente das amostras de dados e as probabilidades estimadas derivadas da distribuição teórica treinada. Em outras palavras, trata-se de uma medida comum de quão bem o modelo se ajusta à variável dependente no conjunto de treinamento (HAREL, 2009); expressa como a proporção da variância do parâmetro dependente predito em relação aos valores observados de uma variável aleatória independente (SCHOBBER; BOER; SCHWARTE, 2018).

O coeficiente de determinação baseou-se em dois principais parâmetros (ABEBE *et*

al., 2020):

- SS_{res} tido como o somatório dos quadrados da diferença entre os valores observado e predito, conforme Equação 21, responsável por quantificar quão distante as previsões geradas pelo modelo estavam dos registros originais;

$$SS_{res} = \sum (y_i - \hat{y}_i)^2 \quad (21)$$

Para o qual y_i representa cada registro original e \hat{y}_i é o valor predito pelo modelo.

- SS_{tot} , definido como o somatório dos quadrados da diferença entre os valores observados e a média das amostras de entrada de acordo com a Equação 22, cujo resultado apresentou o afastamento dos dados originais quanto ao valor médio de todos os registros.

$$SS_{tot} = \sum (y_i - \bar{y})^2 \quad (22)$$

Em que y_i representa cada registro original e \bar{y} é a média dos registros originais.

A diferença entre os dois fatores demonstrou quão afim estavam os dados preditos pelo modelo em relação ao modelo médio, justamente porque quando dividida tal diferença pelo próprio SS_{tot} , foi possível encontrar o coeficiente de distribuição R^2 , dito indicador da representatividade da boa adequação do modelo, conforme apresenta a Equação 23.

$$R^2 = \frac{SS_{tot} - SS_{res}}{SS_{tot}} \therefore R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (23)$$

Os valores para R^2 variam entre 0 e 1 (0 e 100%); quando igual a 0, trouxe como significado que o modelo proposto não obteve boa performance quando na tentativa de prever os dados originais; em contrapartida, quando alcançado o valor 1, constatou-se como predição perfeita (PANAPAKIDIS; SOURTZI; DAGOUMAS, 2020). Assim, quanto maior o R^2 , mais explicativo tornou-se o modelo linear para o projeto proposto, ou seja, melhor ele se ajusta à amostra (FAN *et al.*, 2018).

A determinação do modelo vencedor, ou seja, aquele que há de apresentar a melhor performance por conseguir prever de maneira mais satisfatória os valores da variável-alvo dados determinados preditores, avaliou-se o coeficiente de determinação, R^2 , alcançado por cada modelo em todas as rodadas da validação cruzada. Um R^2 baixo significaria que

as predições proporcionadas estavam distantes dos valores reais e, portanto, o modelo desenvolvido não seria suficiente para explicar o comportamento da variável-alvo.

Como exemplo, um coeficiente de determinação igual a 0,50 significa que o modelo desenvolvido foi capaz de “aprender” cerca de 50% da variância da variável dependente a partir do regressores (variáveis independentes) aplicados no projeto.

3.7 Validação do Modelo

Para mitigar o problema de sobreajuste (overfitting) gerado durante as fases de treinamento e de teste do modelo, adotou-se a Validação de Retenção (MOHR *et al.* 2018). Os dados foram separados inicialmente, em três partes não sobrepostas para treinamento, teste (retenção) e validação final (YADAV & SHUKLA, 2016). Foram selecionadas 852 embarcações para treinamento e 365 para a etapa de teste. Outras 125 embarcações restaram para a etapa adicional de validação via Método de Retenção (Hold-out), exclusiva ao modelo de melhor desempenho ainda na fase de construção e treinamento do modelo.

A validação de retenção pode ter diferentes porcentagens de dados sendo retidos para teste, desde 20% de validação de retenção ou até 10% de validação de retenção (YADAV & SHUKLA, 2016). Os registros pertencentes à cada partição foram escolhidos de maneira aleatória em que o banco de dados foi dividido sem que houvesse um critério avaliativo para tal, tendo em vista a capacidade de generalização esperada acerca do modelo escolhido (ZHANG *et al.*, 2019). Diferentes aproximações do desempenho de generalização podem levar a diferentes desempenhos de teste (ZENG & LUO, 2017).

4 RESULTADOS

4.1 Categorização dos Padrões

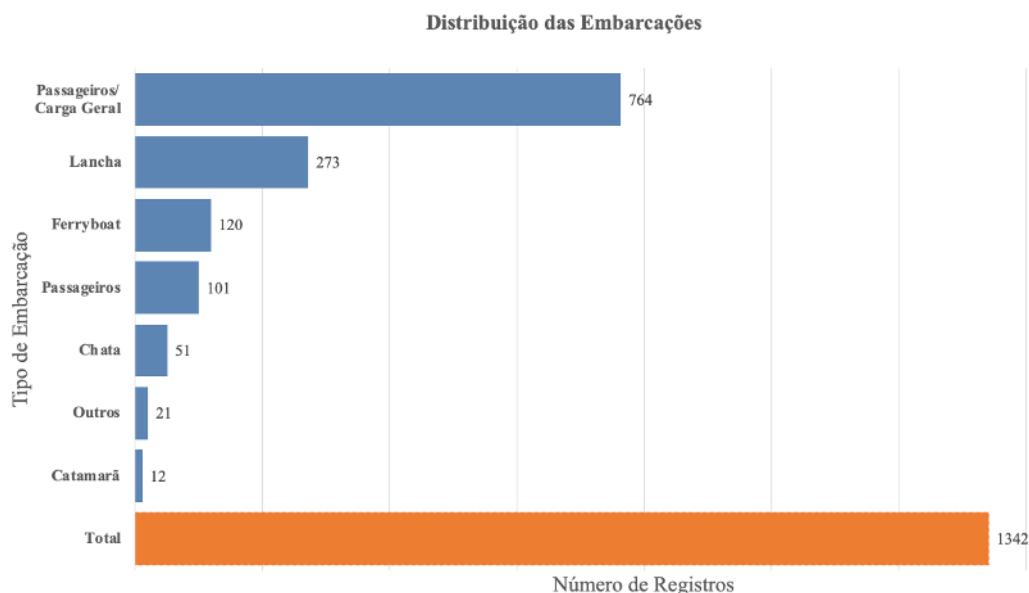
Cerca de 1342 embarcações contribuíram com a movimentação de passageiros e de carga na região amazônica por meio do transporte aquaviário considerando o período entre os anos 2015 e 2017: foram catalogadas 692 e 651 embarcações, respectivamente, em cada ano.

Conforme exposto no Figura 9, o transporte misto, concentrado na movimentação

de passageiros e carga, tem maior participação no modal aquaviário da região, contando com 764 registros, cerca de 56,9% das embarcações, uma diferença de 496 embarcações em relação à quantidade de lanchas, segundo colocado, que detém 268 embarcações e representa 20,3% da amostragem das embarcações. Os número de ferryboats apresentado foi de 120 unidades, contabilizando 8,9% do montante, enquanto que para embarcações que transportam apenas passageiros foram 101 registros, cerca de 7,5% das embarcações. Além disso, foram 51 chatas, 3,8% do número total de embarcações, e 21 embarcações cujas características não coincidem com as demais categorias e, por isso, categorizadas apenas como Outras, representando 1,5% das embarcações. Por fim, 12 catamarãs, quantidade equivalente à 0,9% do montante.

Estas categorias de transporte hidroviário foram anteriormente definidas pela Agência Nacional de Transporte Aquaviário (ANTAQ) face ao perfil de navegação interior empregado na Amazônia em que muitas pessoas têm no modal aquaviário a única alternativa para deslocamento próprio, bem como envio de mercadorias.

Figura 9: Distribuição das embarcações de acordo com o regime de navegação.



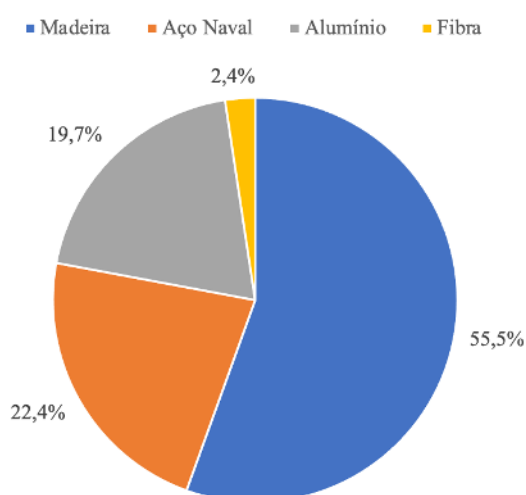
Fonte: Autor (2023).

O tipo de navegação ao qual a embarcação está dedicada influencia na concepção do material a ser utilizado em sua construção. De acordo com que apresenta-se na Figura 10, há uma expressiva utilização de embarcações construídas em madeira, motivada pela

identidade regional em que muitos armadores tendem a optar por um material que seja resistente, leve e, ao mesmo tempo, mais facilmente disponível para aquisição e com baixo custo. Dessa forma, são alcançadas as necessidades construtivas e operacionais suficientes para um ótimo desempenho dessas embarcações construídas em madeira que apresentam um projeto mais robusto de maneira a comportar o arranjo das cargas em conjunto com os passageiros.

A madeira é destaque, inclusive, em detrimento ao aço naval e ao alumínio, estes que concentram uma quantidade similar de embarcações: foram 745 embarcações de madeira, 55,5% do total; outras 301 embarcações em aço naval, 22,4% do montante; 264 embarcações com casco em alumínio, cerca de 19,6% das embarcações; e 32 embarcações em fibra, um percentual de apenas 2,4% do geral.

Figura 10: Distribuição das embarcações conforme material de construção do casco.



Fonte: Autor (2023).

Uma visão mais centrada nos materiais utilizados de acordo com os tipos de embarcação está evidenciada no Figura 11. Cada quadrante corresponde à quantidade de embarcações edificadas por tipo de material (alumínio, fibra, aço naval e madeira) de acordo com os seguintes tipos de embarcação: ferryboat, lancha, passageiro ou misto – parcela que representa 93,7% dos registros.

É possível inferir a partir do Figura 11 que, de 753 embarcações destinadas ao transporte misto, cerca de 588 embarcações tiveram o casco construído em madeira,

enquanto que o restante, 176 barcos, em aço naval. Como parâmetro comparativo, embarcações do tipo ferryboat e lancha têm sua representatividade baseada no alumínio e aço naval, respectivamente, quando se tratando sobre o material do casco. Nota-se, dessa maneira, que a madeira tem uma participação significativa quando para transporte de passageiros e cargas, fato que não se repete para lanchas e ferryboats.

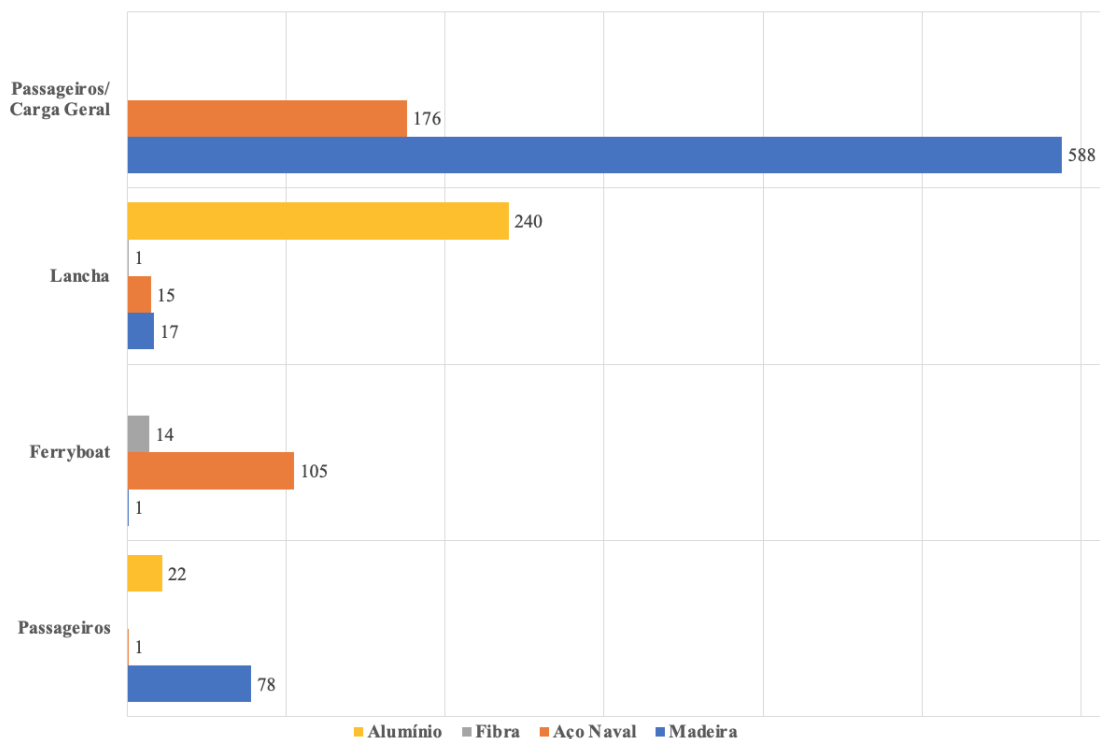
As lanchas apresentaram, em sua maioria, o alumínio como material construtivo mais utilizado, sendo 240 embarcações, cerca de 87,9% de um total de 273 lanchas. Caracterizam-se como embarcações de velocidade que, embora transportem cargas, tem no deslocamento rápido de passageiros o seu principal propósito, sendo o alumínio o material mais indicado para sua construção trazendo vantagens como: menor peso, resistência satisfatória e boa conformação construtiva.

No caso dos ferryboats, 105 embarcações foram construídas em aço naval, 87,5% de 120 embarcações levantadas. Como são embarcações empregadas, principalmente, no transporte de cargas mais pesadas, como veículos e mercadorias, além de passageiros, é imprescindível que, frente ao peso requisitado, seja aplicado um material de grande resistência e, por esta razão, há uma preferência maior no uso do aço naval.

Embarcações dedicadas exclusivamente ao transporte de passageiros, cuja alta capacidade de acomodação de pessoas é prioridade, inclusive, em detrimento à velocidade de cruzeiro e capacidade de transporte de carga, semelhante às embarcações de transporte misto, há uma tendência na utilização de casco em madeira: são 78 embarcações, representando 77,2% do montante.

Conhecer o comportamento dos dados a partir do tipo de embarcação e material utilizado para sua construção possibilita uma análise mais clara da distribuição do consumo de combustível das embarcações operantes na região amazônica. O fracionamento dos dados em função do consumo de combustível está demonstrado na Figura 12, trazendo a quantidade de embarcações em referência à quantidade de combustível utilizado por viagem. Nota-se que grande parte das embarcações consome um volume de combustível entre 500 e 1000 litros, bem como uma parcela significativamente menor chega a um valor acima dos 4000 litros, por exemplo. Importante dizer que as constatações apresentadas fazem menção a um consumo de combustível absoluto por viagem.

Figura 11: Relação entre o material de construção do casco e a categoria de navegação da embarcação.



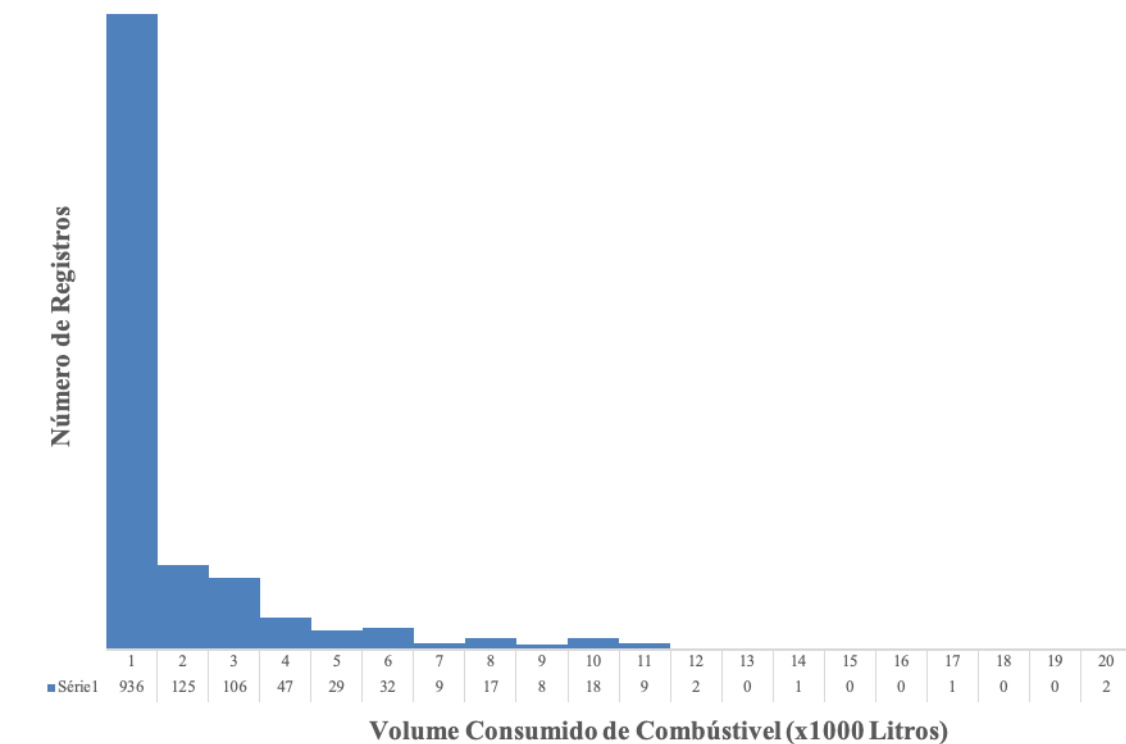
Fonte: Autor (2023).

Salienta-se, através deste embasamento, duas principais constatações:

- A distribuição não é normal (ou gaussiana) e os atributos apresentam relações não-lineares com a variável-alvo;
- Existem valores extremos, também chamados raros, presentes na variável dependente do estudo, caracterizada pela “cauda” deslocada à direita da figura, sendo necessário uma etapa de pré-processamento para tratamento dos dados, ao passo que outliers indesejados devem ser encontrados e removidos.

Acontece que há um número maior de embarcações que operam com até mil litros por viagem do que a quantidade das mesmas que consomem um volume superior a este limite. Embarcações que registram um consumo acima de quatro mil litros, por sua vez, tendem a ser consideradas “eventos” raros.

Figura 12: Distribuição das embarcações amazônicas em função do consumo de combustível.



Fonte: Autor (2023).

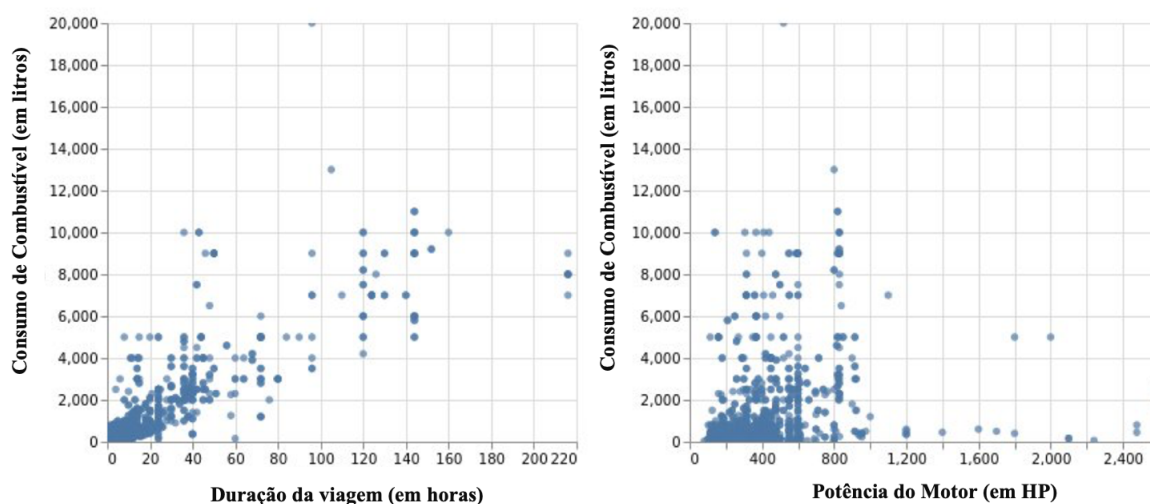
De modo a corroborar com a assertiva anterior, a Figura 13 esboça a relação da variável-alvo, consumo de combustível, com dois outros preditores: duração da viagem e potência do motor. Ambos os gráficos trazem a dispersão dos registros concernentes ao banco de dados da ANTAQ explicitando, neste caso, qual o consumo de combustível das embarcações quando a partir da duração de sua viagem e potência do seu motor.

Observa-se que há uma concentração de embarcações operantes apresentando um volume de óleo consumido em torno de 2000 litros com duração de viagem até vinte horas. Quanto à potência do motor, nota-se um acúmulo de embarcações na faixa entre 100 HP e 400 HP para um mesmo consumo de até 2000 litros. Os pontos fora da curva apresentam-se mais distantes da origem, em que quanto maior a duração da viagem, maior o consumo de combustível. A mesma prerrogativa, no entanto, não pode ser estabelecida para o padrão de potência do motor tendo em vista que o comportamento da distribuição apresenta-se predominantemente não-linear.

Explica-se, portanto, que a concentração característica das embarcações quanto ao

consumo de combustível menor do que 1000 litros ocorre devido à boa parte das viagens serem rápidas, com duração inferior à dez horas, e também pelas potências propulsivas utilizadas não serem altas, algo em torno de até 350 HP.

Figura 13: Relação consumo de combustível versus duração da viagem e potência do motor.



Fonte: Autor (2023).

A relação potência do motor versus consumo de combustível revela ainda a existência de embarcações cujas potências são as mesmas, porém, apresentando volumes consumidos diferentes. Por esta razão, evidencia-se a infalível necessidade em estratificar tanto parâmetros físicos quanto operacionais, bem como propulsivos.

O levantamento anterior é explicado devido à diferença de regime operacional sob os quais estão submetidas as embarcações, seja regime de planeio ou deslocamento. Fica manifesto, dessa maneira, que para um transporte de altas cargas é fundamental um motor de potência proporcional, gerando, conseqüentemente, um consumo de combustível menor. Os ferry-boats, exemplo de regime em deslocamento, se utilizam de uma potência propulsiva maior do que embarcações para transporte misto de porte médio, até 300 passageiros. Entretanto, a mesma assertiva não é cabível quando em alusão às lanchas, uma vez que, embora transportem um valor de carga menor do que os ferry-boats, caracterizam-se por motores de alta potência devido à necessidade de atingir grandes velocidades, ou seja, um regime de planeio.

Imprescindível conhecer o regime de operação das embarcações, haja vista que devem ser identificados os valores que, de fato, são outliers, isto é, dados que se diferenciam

drasticamente dos outros: são pontos fora da curva que podem enviesar negativamente o projeto proposto. À exemplo disto, percebe-se que existem embarcações de mesmo regime operacional cuja potência é similar, porém, apresentando diferença discrepante quanto ao consumo de combustível. O mesmo parecer é válido para a duração da viagem.

Por esta razão, os gráficos no formato diagrama de caixa apresentados, à posteriori, contribuem para a constatação acerca da presença de valores que fogem à média ou padrão dos valores alcançados quando nas tratativas que dizem respeito às embarcações operantes na região, sendo assim, considerados valores extremos. Importante citar ainda que, devido ao acesso restrito das localidades pesquisadas e ambiente desconfortável onde os questionários eram aplicados, erros de origem humana possivelmente aconteceram quando no momento do levantamento dos dados por parte da equipe e que não distinguem-se em relação aos demais registros, evidenciando uma coleta não-otimizada.

4.1.1 Estratificação do Universo de Embarcações Proposto

Após aquisição de dados, cuja base para coleta dos registros foi o ECTFA, houve tratamento dos registros e aplicação inicial de método de seleção de atributos do tipo filtro, reduzindo o número de atributos escolhidos e contidos na etapa de pré-processamento para dezessete, das quais quinze eram numéricas e outras duas categóricas. Parâmetros referentes às condições ambientais (onda, vento e correnteza) não foram consideradas devido não estarem disponíveis para análise pelo estudo-base.

Os quinze parâmetros numéricos adotados estão apresentados na Tabela 5, gerados após fase de processamento e seleção de atributos, juntamente às métricas quantitativas, como média, mediana, mínimo, máximo e limites característicos aos primeiro e terceiro quartil de cada atributo. Sob a ótica de comportamento dos dados apresentados, entende-se que:

- Cerca de metade das viagens performadas tem duração de até 9 horas;
- 75% do total de embarcações tem comprimento até 31,3 metros;
- Embarcações que navegam com cinco propulsores não apresentam ganhos significativos de velocidade ou desempenho para os trajetos da região;
- A maior embarcação para transporte de passageiros contida no banco de dados é capaz de transportar 1400 pessoas;

- A embarcação de maior capacidade de carga utilizada no transporte misto é capaz de movimentar até 1600 toneladas;
- O calado de grande parte das embarcações amazônicas restringe-se a 1,80 metros devido às restrições inerentes às vias navegáveis;
- A média de consumo de combustível das embarcações operantes na Amazônia está em torno de 1270,88 litros.

Os diagramas de caixa apresentados na Figura 14 correspondem aos valores anteriormente apresentados por meio da Tabela 4, assim apresentados com o intuito de identificar os pontos extremos (outliers) ainda contidos na base de dados mesmo após a fase de processamento.

Quanto às características físicas, ocorre que, para os registros levantados na região amazônica, o comprimento médio das embarcações apresentado foi igual a 26 metros, enquanto que para os demais parâmetros como boca, calado e pontal foram obtidos valores médios iguais a 5,92 metros, 1,53 metros e 2,10 metros, respectivamente. Grande parcela dos pontos extremos para os quesitos citados, tornam evidente a preferência na construção de embarcações de pequeno a médio porte para durações de viagem e carga transportada menores.

Para as variáveis propulsivas, a citar o número de propulsores, potência média do motor e velocidade de cruzeiro, o valor médio apontado foi de 1,38 propulsor, 393,18 HP e 16,28 nós, respectivamente. O número de propulsores médio encontrado demonstra uma tendência maior no uso de um propulsor do que dois propulsores, haja vista ser uma variável “binária”, não passível de fracionamento. A velocidade média apontada revela um regime de operação adotado por boa parte das embarcações como sendo de planeio em que é priorizada a velocidade de cruzeiro em detrimento ao volume de passageiros e carga transportados pela embarcação, uma vez que quanto mais peso, maior resistência contra o deslocamento e menor a velocidade do barco.

Tabela 5: Distribuição dos registros após fase de processamento.

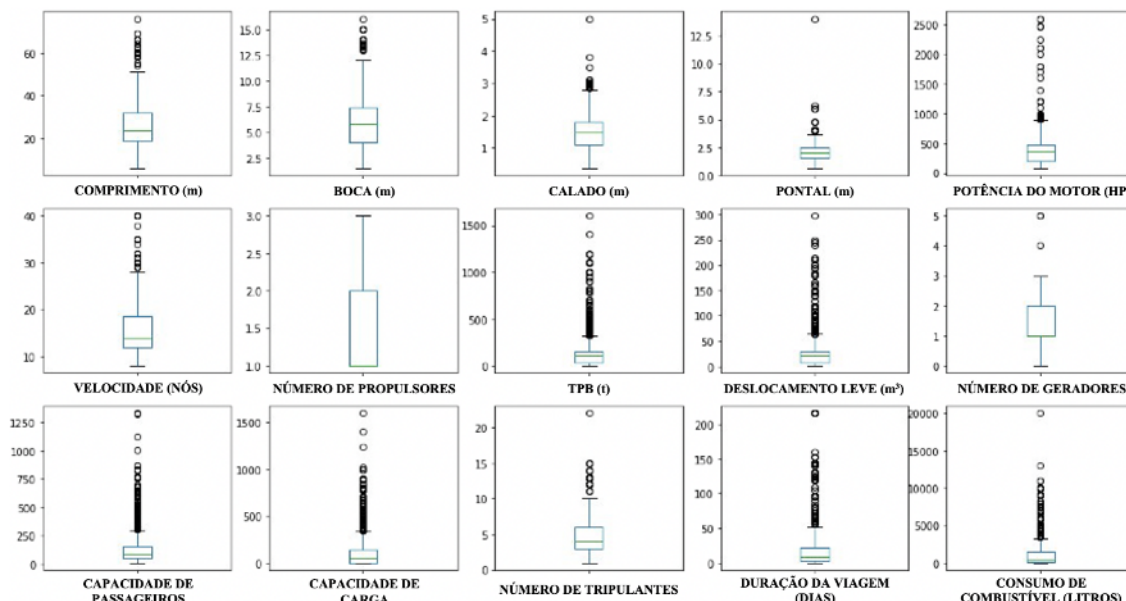
Parâmetro	QTD	M	MD	MN	25%	50%	75%	MX
Comprimento		26,0	10,9	6,0	19,0	23,3	31,3	76,0
Boca		5,9	2,6	1,5	4,0	5,7	7,4	21,4
Calado		1,5	0,58	0,3	1,1	1,5	1,8	5,0
Pontal		2,1	0,79	0,6	1,6	2,0	2,5	14,0
Potência do Motor		393,1	281,1	80,0	200,0	350,0	480,0	2750,0
Velocidade de Cruzeiro		16,2	6,34	8,0	12,0	15,0	19,7	40,0
Número de Propulsores		1,3	0,59	1,0	1,0	1,0	2,0	5,0
TPB	1342	150,6	187,3	5,0	35,0	109,0	141,5	1600,0
Deslocamento Leve		29,0	35,5	2,0	7,0	21,0	28,0	297,0
Número de Geradores		1,2	0,7	0,0	1,0	1,0	2,0	5,0
Capacidade de Passageiros		140,0	159,3	10,0	52,0	88,0	145,0	1400,0
Capacidade de Carga		113,3	182,1	0,0	5,0	46,0	130,0	1600,0
Número de Tripulantes		4,4	2,42	1,0	3,0	4,0	6,0	22,0
Duração de Viagem		20,4	37,5	0,2	3,0	9,0	21,7	768,0
Consumo de Combustível		1270,8	2136,6	5,0	150,0	400,0	1300,0	20000,0

QTD: Quantidade de registros; M: Média; MD: Mediana; MN: Mínimo; MX: Máximo.

Fonte: Autor (2023).

Os preditores concernentes à operação das embarcações, como tonelagem porte bruto (capacidade máxima, em toneladas, a ser transportada por uma embarcação sem considerar seu peso próprio), deslocamento leve (apenas o peso próprio da embarcação sem quaisquer outras cargas como consumíveis ou tripulação), capacidade de passageiros e de carga, número de tripulantes e duração da viagem apresentaram média de 150,64 toneladas, 29,07 m³, 140 passageiros, 113,37 toneladas e 20,40 horas, consecutivamente.

Figura 14: Representação em box-plot dos parâmetros de entrada do modelo.



Fonte: Autor (2023).

4.2 Seleção de Atributos

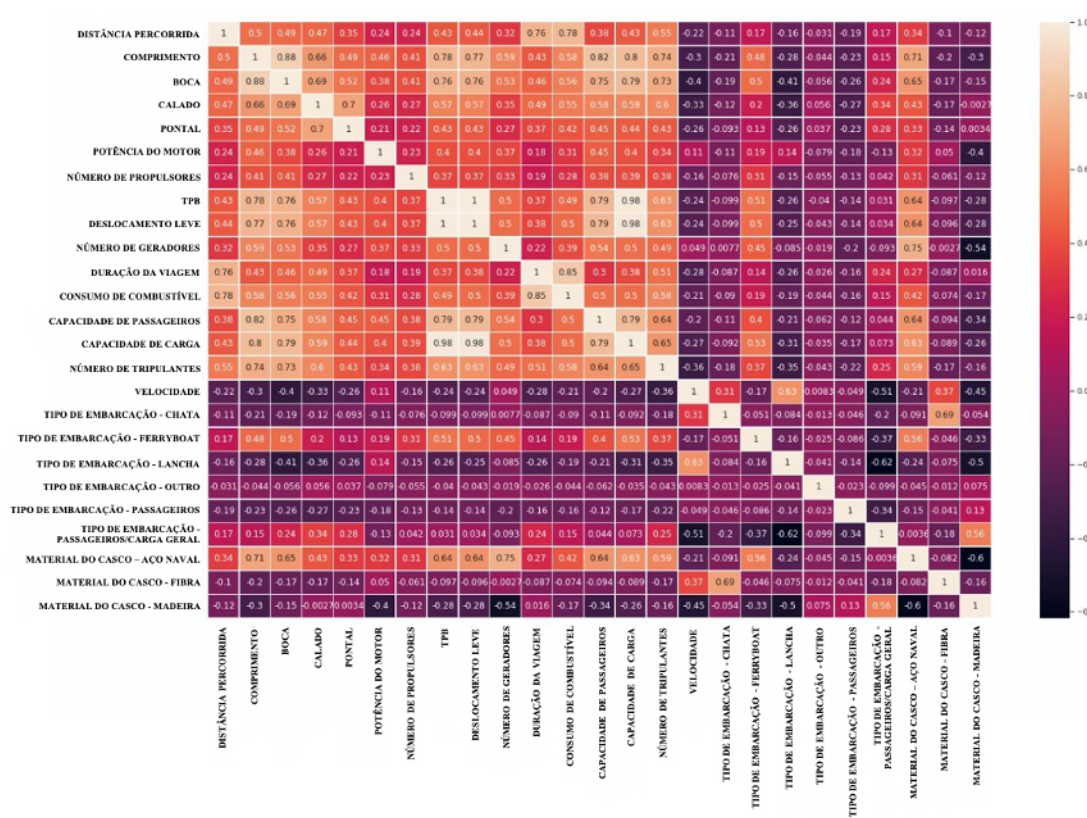
4.2.1 Matriz de Correlação via Método de Coeficiente de Pearson

Para validação dos algoritmos propostos - Árvore de Decisão, Florestas Aleatórias, Árvores Extras, Aumento de Gradiente, Aumento Extremo de Gradiente e Catboost - uma análise de regressão foi desenvolvida por meio da repartição do banco de dados das embarcações levantadas em parcelas para treinamento, teste e validação na proporção de 70%, 20% e 10%, respectivamente, do contingente total dos registros.

Para a etapa de seleção de atributos, tal como defendido pelos trabalhos desenvolvidos por Abebe *et al.* (2020) e Uyanik *et al.* (2020), em função da necessidade em conhecer a relação existente entre as variáveis de entrada e alvo do modelo, optou-se pela aplicação do Método de Correlação de Pearson. A matriz de correlação entre os parâmetros está apresentada na Figura 15, cujos valores variam de -1 e +1 entre tons claros e escuros, demonstrando a afinidade entre tais variáveis: quanto maior a dependência entre duas variáveis, mais próximo de +1 será o termo contido na interseção entre ambas, além de assumir uma cor de tonalidade mais clara, fator que pode, inclusive, representar redundância entre variáveis. Os parâmetros de entrada do modelo estão identificados nos

eixos x e y, enquanto que a diagonal principal apresenta um espalhamento das relações entre os parâmetros em que, seja no triângulo inferior ou superior à diagonal principal, as variáveis e as relações não se alteram (LIU *et al.*, 2017).

Figura 15: Matriz de Correlação de Pearson entre os parâmetros do modelo.



Fonte: Autor (2023).

Para os parâmetros operacionais tempo de viagem, distância e velocidade, os coeficientes de correlação obtidos foram de +0,85, +0,78 e -0,21. Quanto mais tempo uma embarcação estiver navegando, conseqüentemente, uma quantidade maior de óleo será demandada para combustão, impactando no volume de combustível consumido. A mesma premissa aplicou-se para viagens com grandes distâncias: viagens com itinerários mais longos tendem a consumir mais do que se comparado à viagens rápidas. A velocidade, por sua vez, apresentou a correlação negativa mais forte de todos os preditores levantados. Quanto mais alta a velocidade de cruzeiro da embarcação, menor será o consumo de combustível desempenhada pela mesma, uma vez que o tempo de viagem será menor.

Levando em consideração apenas os atributos dimensionais dos registros coletados, o comprimento (+0,58) destacou-se com a maior correlação com o consumo de combustível,

demonstrando que quanto maior for a embarcação em seu sentido longitudinal, maior será a tendência da embarcação consumir um volume considerável de combustível. Os demais atributos dimensionais e físicos como boca (+0,56), calado (+0,55) e pontal (+0,42) apresentaram o mesmo comportamento, cuja relação é diretamente proporcional em função do consumo de combustível, porém, com intensidade correlativa menor quando comparado ao comprimento. Para o projetista, é de extrema importância conhecer a dimensão física de construção que contribuirá para um alto consumo de combustível, uma vez que, optando por otimização de custos, o comprimento deve ser o primeiro parâmetro a ser revisado.

Tendo em vista que o coeficiente de arrasto e a resistência friccional são funções dependentes do tamanho da embarcação, quanto maior for o comprimento do navio, maior será a área de contato em relação ao fluido, sendo esta uma constatação intuitiva e comprovada pela correlação intrínseca às embarcações operantes na região amazônica. A baixa correlação entre as embarcações de madeira (-0,17) com o consumo de combustível torna-se evidente na Figura 4. Em um cenário onde duas embarcações apresentam dimensões semelhantes e condições operacionais idênticas, porém, tendo sido uma construída em aço naval e outra em madeira, a primeira apresentará um consumo de combustível maior do que esta segunda. Em outras palavras, barcos construídos em aço naval (+0,42) tendem a um volume consumido de combustível maior se comparado às embarcações de alumínio (+0,19), fibra (-0,074) e madeira (-0,17). Em contrapartida, do ponto de vista econômico, embarcações projetadas com casco de madeira apresentaram-se mais vantajosas por conta do seu menor consumo de combustível.

Cascos construídos em aço naval apresentaram a mais expressiva correlação com o consumo de combustível, cerca de +0,42. A rugosidade do casco mais acentuada impacta diretamente no coeficiente de fricção. Além disso, boa parte destas embarcações transportam um alto volume de carga, solicitando materiais com resistência e durabilidade maiores. Como consequência, motores de combustão cada vez maiores são necessários para sustentar a potência propulsiva necessária para vencer o peso e a dificuldade de deslocamento da embarcação.

Sob a ótica dos parâmetros propulsivos, o número de geradores apresentou a maior correlação (+0,39), seguido pela potência do motor (+0,31) e pelo número de propulsores (+0,28). Entende-se que embarcações maiores adotam uma quantidade redundante de

geradores por fator de segurança: na parada ou mau funcionamento de um gerador, um segundo deve imediatamente ser colocado em barra. Uma vez que os motores auxiliares (geradores) são responsáveis por fornecer energia elétrica para funcionamento dos equipamentos de comando e navegação, o barco ficaria à deriva ou extremamente suscetível a uma colisão com outro navio em caso de ocorrência de um apagão.

A potência do motor determina a capacidade de transporte e/ou velocidade desenvolvida pela embarcação. Quanto maiores o peso e as dimensões ou a velocidade performada por uma embarcação, maior será a potência requerida para deslocá-la. Considerando a curva de desempenho da mesma, o consumo de combustível tenderá a aumentar também. O número de propulsores, por sua vez, apresentou comportamento semelhante: para alcançar a força necessária para deslocamento ou planeio da embarcação e preservar a vida útil dos motores, uma quantidade maior de propulsores é solicitada.

Embarcações do tipo ferryboat, intrínsecas à região amazônica por conta das suas grandes dimensões e capacidade de carga utilizadas, inclusive, para transporte de carros junto aos passageiros, apresentou a correlação mais alta, +0,19, quanto à premissa de tipos de embarcação, seguido pelas embarcações de transporte de passageiros e carga geral (+0,15), chata (-0,09), transporte de passageiros (-0,16) e lancha (-0,19). Comparando os dois extremos, entende-se que quanto maior a capacidade de carga requerida, por exemplo, para um ferryboat, maior será o consumo de combustível, enquanto que quanto maior a velocidade desempenhada por uma lancha, menor será o consumo de combustível, uma vez que a duração da viagem será reduzida.

É possível inferir, dessa maneira, que a duração da viagem é o atributo que mais influencia no consumo de combustível, seguido pelo uso de madeira como casco e, em terceiro, o comprimento da embarcação. Com isso, se uma viagem de longa distância ou uma embarcação construída de madeira ou de grande comprimento são fatores que certamente ocasionarão um consumo de combustível maior por parte da embarcação, enquanto que o inverso também é verdadeiro.

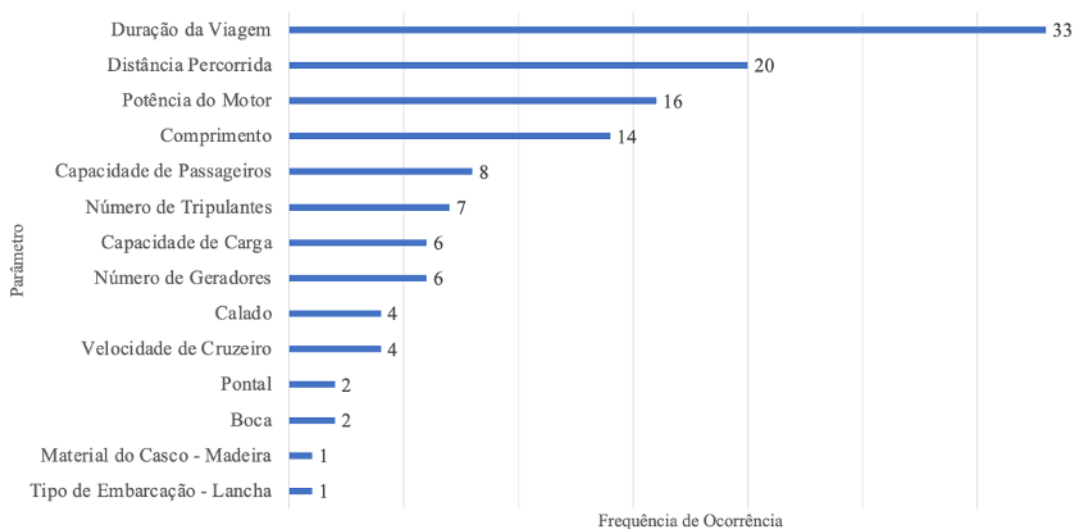
4.2.2 Ranking de Importância dos Parâmetros via Aplicação de Método F-Score

A técnica de Importância de Atributo, desenvolvida especificamente para uso do algoritmo do tipo Classification and Regression Trees (CART), foi aplicada por meio

da função `feature_importances_` disponível pela biblioteca `scikit-learn` via linguagem de programação Python 3.6 em um ambiente Jupyter®.

As pontuações de importância relativa à cada um dos preditores puderam ser calculados e apresentados em um gráfico de barra, cujas colocações foram ordenadas levando em consideração a pontuação definida pelo método de Pontuação por Frequência (ou F-Score), aplicado por Dao *et al.* (2018) e Nguyen *et al.* (2019), conforme pode ser visto no Figura 16. A porcentagem da frequência com que os atributos foram usados para divisão das árvores determinou a importância de cada um destes e sua influência em relação ao consumo de combustível.

Figura 16: Ranking de importância dos parâmetros de entrada utilizados no modelo.



Fonte: Autor (2023).

Os índices localizados ao lado das barras indicam a frequência do resultado (frequency score), ou seja, a parcela de vezes em que o preditor foi escolhido como nó raiz, o qual, geralmente, apresenta o menor ganho e, portanto, é escolhido como o primeiro divisor do modelo, sendo considerado também o de maior relevância na determinação do valor final da variável-alvo, neste caso, o consumo de combustível (SNAPIR *et al.*, 2019). Em síntese, foram geradas cerca de 124 árvores de decisão durante a aplicação do método, ao passo que apenas 13 parâmetros foram selecionados como nó raiz durante as rodadas de criação destas árvores, demonstrando a influência destes atributos em referência ao consumo de combustível.

O tempo de viagem foi escolhido em 33 árvores de decisão, uma representatividade de 26,61% do total, sendo o atributo mais relevante em relação ao consumo de combustível. Em segundo lugar, a distância percorrida obteve a liderança em 20 árvores de decisão, cerca de 16,12% da totalidade. A potência do motor, logo em seguida, apresentou índice igual a 16, parcela equivalente a 12,90% do montante.

Quanto maior a distância entre as cidades de origem e destino, maior será o tempo necessário para cumprir a viagem, o que conseqüentemente implicará em um consumo de combustível proporcionalmente maior. A potência do motor é o único parâmetro capaz de alterar esta sinergia, haja vista que embarcações cujo sistema propulsivo apresenta uma potência maior, certamente alcançarão um tempo de viagem reduzido, porém, aumentando ainda mais o consumo de combustível.

O comprimento obteve 14 árvores de decisão, cerca de 11,29% do todo, seguido pela capacidade de passageiros com 8 árvores de decisão, representando 6,45% das árvores geradas, pelo número de tripulantes com 7 árvores de decisão, 5,64% do montante, e pela capacidade de carga com 6 árvores de decisão, em torno de 4,83% do total. Ao passo com que o comprimento das embarcações tornam-se maiores, por norma de salvaguarda e segurança do meio ambiente e da vida humana, fica estabelecido um número maior de tripulantes, bem como de patentes mais altas. Intuitivamente, isto acontece porque cresce a capacidade de transporte de passageiros e de carga da embarcação. Em síntese, se elevado o peso a ser transportado, infere-se haver uma tendência de consumo de combustível também maior.

Para o número de geradores, foram 6 árvores de decisão, 4,83% do total de árvores criadas, enquanto que calado e velocidade obtiveram igualmente 4 árvores de decisão para cada, cerca de 3,22%.

O número de geradores é proporcional ao tamanho da embarcação e, conseqüentemente, necessário para uma boa acomodação de tripulantes e passageiros, além de bom armazenamento das cargas, em especial, se perecíveis congeladas, influenciando no consumo de combustível. Acontece ainda que é empregado um quesito de redundância em que há a consideração de um gerador reserva em caso de o primário sofrer algum dano.

Embarcações cujos calados operacionais são elevados por conta da alta capacidade de carga transportada exigem um alto rendimento do motor, acarretando em um alto

consumo de combustível. Sob a mesma ótica de desempenho, para embarcações de rápida velocidade, a potência exigida tende a ser maior, consumindo um volume de combustível também maior.

Os quatro últimos listados – Tipo de Embarcação Lancha, Material do Casco Madeira, Boca e Pontal - apresentaram índices iguais a 1 para os dois primeiros, e 2 para os últimos, respectivamente, demonstrando uma participação em grupo de apenas 4,8% em relação ao número de árvores geradas. Ainda assim, o tipo de embarcação (condicionado a uma lancha), do material do casco (condicionado à madeira) e as dimensões de boca e/ou de pontal tem uma influência ligeiramente relevante quanto ao consumo de combustível.

Mesmo apresentando uma baixa pontuação, considerando que não foram selecionados outros tipos de embarcação, nem tampouco outros materiais de casco, é possível inferir que o uso de uma embarcação do tipo lancha, bem como se o material de casco da embarcação for madeira, contribui para um maior consumo de combustível se comparado aos demais grupos.

Em outras palavras, uma relação de correlação destes com a variável-alvo é exposta: por exemplo, quanto maior for o tempo de viagem e/ou o comprimento da embarcação e/ou a quantidade de tripulantes, maior será o consumo de combustível predito para a dada embarcação.

4.2.3 Correlação de Atributos de Entrada em Função do Consumo de Combustível via Método SHAP

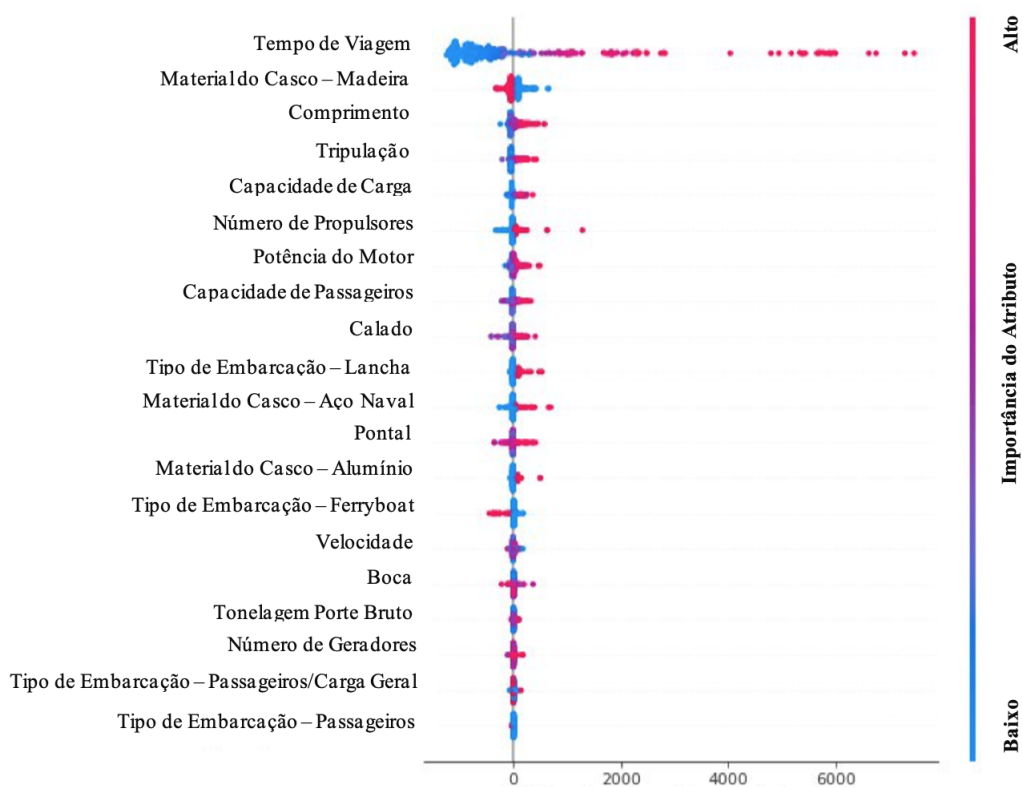
O método SHapley Additive exPlanation (SHAP) foi aplicado visando uma análise local de contribuição de cada registro frente à variável de saída, o consumo de combustível, podendo ser positiva ou negativa. Como resultado, foram expostas a hierarquia de importância de cada atributo e a distribuição dos registros e seus impactos relacionados à variável-alvo, o consumo de combustível, conforme apresentado na Figura 17.

A representação denota no eixo x os valores SHAP de cada registro para um determinado atributo, evidenciando o impacto que cada dado gera em relação ao consumo de combustível. Quando positivo, a característica predominante da distribuição dos dados está intimamente ligada com o comportamento da variável-alvo, ao passo que crescem/decrecem sob a mesma proporção, enquanto que para valores negativos, entende-se como parâmetros cuja relação é inversamente proporcional. Para o eixo y, as variáveis

utilizadas foram apresentadas em ordem crescente em função do grau de correlação existente com o consumo de combustível: de baixo para cima, foram demonstradas as variáveis detentoras da menor até a maior correlação.

A paleta de cores representa o valor normalizado do registro tomando como referência a média do valor de SHAP: quando acima, cor mais quente significando valor alto; quando abaixo, cor mais fria cujo valor é baixo. Cada ponto configura uma embarcação e assume cor equivalente ao seu valor, ao passo que quanto mais afastado o ponto em relação à origem, maior o peso em determinar a relação do atributo com a variável-alvo.

Figura 17: Representação de importância dos preditores através do método SHAP.



Fonte: Autor (2023).

A duração da viagem ocupa a primeira posição no gráfico de SHAP, representando a maior importância e influência na determinação do consumo de combustível. Acontece que os dados estiveram mais dispersos e distantes da origem para a direita e apresentando uma cor mais avermelhada, enquanto que, para a esquerda, há uma concentração maior dos registros e uma distância menor da origem, além de uma cor mais azulada. Intuitivamente, afirma-se, através deste comportamento, que valores altos para a duração da viagem

implicam em um consumo de combustível maior, enquanto que, para durações mais curtas, há uma redução no consumo de combustível.

Outro parâmetro representado pelo gráfico SHAP diz respeito ao material do casco, mais especificamente, embarcações construídas em madeira. Por se tratar de uma variável binária, existem apenas duas cores distintas categorizando os registros. Ao contrário do comportamento ocorrido com a duração da viagem, o conjunto dos dados localizado à direita está predominantemente avermelhado, enquanto que, ao lado oposto, encontra-se azulado. Ressalta-se, dessa maneira, que embarcações construídas em madeira tendem a gastar um volume menor de combustível quando comparado às embarcações edificadas com quaisquer outros tipos de material. Os registros em azul tem peso contribuinte ao consumo de combustível e estão relacionados aos cascos que não são de madeira.

A lancha destacou-se como o tipo de embarcação melhor colocado, cuja distribuição aponta uma relação de proporcionalidade direta: lanchas consomem um volume maior de combustível do que demais categorias. Em contrapartida, para embarcações cujo intuito é exclusivamente o transporte de passageiros, ocorre a redução do consumo de combustível.

Embarcações equipadas com um número maior de propulsores, tendem a gastar um volume de combustível elevado, haja vista a posição agrupada de dados de maior valor de SHAP e, por isso, avermelhados, localizados à direita do gráfico. A mesma analogia pode ser considerada para a potência do motor.

Para a velocidade, a presença de uma cor cuja tonalidade tende ao roxo evidencia ser este um parâmetro inconclusivo quanto para um projeto propulsivo de embarcações operantes na região. Sabe-se que boa parte dos barcos tem utilizado motores de potências altas, mas navegado com a mesma velocidade daquelas outras que detém motores de baixa potência. Devido a isso, o método proposto determina ser inconclusiva a constatação de que a velocidade seja um agente influenciador do consumo de combustível, tendo em vista que, de pequena a grande potência, as velocidades têm sido similares, porém, com um consumo divergente.

A duração da viagem, além de atributo de maior importância, tem os dados concentrados à direita, se em relação à coordenada de origem, e coloridos em escala mais extrema da cor vermelha. Deduz-se, portanto, que a duração da viagem tem uma relação diretamente proporcional com o consumo de velocidade, ou seja, quanto mais distante os portos de origem e destino da viagem, maior será o consumo de combustível,

uma constatação dita intuitiva por confirmar o que outrora fora apresentado nas seções anteriores.

4.3 Treinamento e Avaliação dos Modelos

Visando a otimização dos hiperparâmetros do modelo, o Método de Busca Aleatória (Grid Search Method) foi adotado, conforme defendido pelos trabalhos de Stepec *et al.* (2020) e Zhang *et al.* (2022). Os valores que melhor se ajustam ao modelo e proporcionam uma maior eficiência são atingidos a partir de incremento passo-a-passo (step-by-step), ou seja, por meio de recurso iterativo em que a máquina altera os hiperparâmetros até que seja alcançado o critério de parada, quando o modelo não mais é capaz de apresentar predições mais assertivas (ABEBE *et al.*, 2020).

Para definição do modelo vencedor, que obteve o melhor desempenho em prever de maneira mais satisfatória o consumo de combustível das embarcações durante a etapa de treinamento, foram avaliadas as métricas de desempenho Erro Absoluto Médio (EAM), Quadrado Médio do Erro (QME), Erro Quadrático Médio (EQM), Coeficiente de Determinação (R^2) e Tempo Computacional de Processamento (TCP), segundo proposto por Abebe *et al.* (2020) em seu estudo.

Para avaliação de desempenho de cada modelo gerado, aplicou-se o Método de Validação Cruzada, também aplicado por Gkerekos *et al.* (2019), que consiste em dividir o conjunto de dados em k partes iguais, tendo sido o k definido como 10, denominadas dobras. O processo repetiu-se por dez vezes, em que, a cada iteração, uma dobra diferente foi selecionada como conjunto de teste, enquanto as demais ($k-1$) serviram como conjunto de treinamento. A pontuação final do modelo foi, dessa maneira, calculada como a média das pontuações obtidas em cada iteração.

Os desempenhos de cada modelo estão presentes na Tabela 6.

O modelo de maior acurácia foi o CatBoost com 85,2%, justamente pela técnica exclusiva de tratamento às variáveis categóricas presentes no banco de dados, sendo robusto a outliers e o mais vantajoso quando lidando com levantamentos tabulares, seguindo o direcionamento aplicado por Stepec *et al.* 2020. Em segundo, terceiro e quarto lugares, Aumento Extremo de Gradiente (HU *et al.*, 2021), Florestas Aleatórias (WICKRAMANAYAKE & BANDARA, 2016) e Aumento de Gradiente (UYANIK *et al.*, 2020) cujos valores dos coeficientes de determinação foram, respectivamente, 79,7%,

Tabela 6: Métricas de desempenho dos seis modelos treinados por meio do Método de Validação Cruzada com 10 rodadas.

	Modelo	EAM	QME	EQM	R^2	TCP
1	Catboost	5,9	2,6	1,5	4,0	5,7
2	Aumento de Gradiente Extremo	1,5	0,58	0,3	1,1	1,5
3	Florestas Aleatórias	2,1	0,79	0,6	1,6	2,0
4	Aumento de Gradiente	393,1	281,1	80,0	200,0	350,0
5	Árvores Extras	16,2	6,34	8,0	12,0	15,0
6	Árvore de Decisão	1,3	0,59	1,0	1,0	1,0

Fonte: Autor (2023).

79,0% e 78,7%. Nas últimas posições, apresentaram-se Árvores Extras (GKEREKOS *et al.*, 2019) com 75,1% e Árvore de Decisão (ABEBE *et al.*, 2020) com 66,95%, este último com o menor coeficiente de determinação dentre os seis modelos considerados.

Para os parâmetros de desempenho QME e EQM, empregados por Ahmad *et al.* (2021), permanece a mesma ordenação de desempenho dos modelos gerada em função do coeficiente de determinação (R^2) em que as predições feitas pelo CatBoost foram as mais assertivas e, por isso, apresentaram QME e EQM menores quando comparados aos demais modelos. Em contrapartida, a avaliação do EAM, demonstra que o Catboost apresentou o menor erro (348,48), seguido por Florestas Aleatórias (352,67), Florestas Aleatórias (387,08), Aumento de Gradiente (406,33), Aumento Extremo de Gradiente (409,23) e Árvore de Decisão (433,02). Os valores EAM e EQM têm as mesmas unidades das variáveis medidas (AFRIFA-YAMOAHA *et al.*, 2020).

Sob a ótica do ponto de vista relacionado ao tempo computacional, como feito por Abebe *et al.* (2020), o DTR destacou-se como aquele que mais rápido conseguiu processar os dados e chegar a um resultado, mesmo que, como visto, apresentando um erro mais alto e, conseqüentemente, predições divergentes dos dados observados. O CatBoost, por sua vez, por conta da sua complexidade em trabalhar com registros categóricos, apresentou o maior tempo para processamento, cerca de 3,80 segundos. Nas posições subsequentes,

ênfase para os modelos de Aumento Extremo de Gradiente e de Florestas Aleatórias que obtiveram um ótimo rendimento nas predições, embora tenham gastado um tempo de processamento também maior.

O CatBoost foi escolhido como modelo vencedor devido às avaliações de desempenho satisfatórias para processamento dos dados e predição da variável-alvo. Foi executado mais uma vez o ajuste dos hiperparâmetros com o objetivo de tornar o modelo ainda mais eficaz, refinando seu tratamento dos residuais e, conseqüentemente, atingindo uma acurácia de maior magnitude, conforme Okumuş *et al.* (2021).

4.4 Otimização de Hiperparâmetros do Modelo Escolhido

Os hiperparâmetros encontrados via aplicação do Método de Busca Aleatória (STEPEC *et al.*, 2020; ZHANG *et al.*, 2019) geraram valores finais e que podem ser checados na Tabela 7, os quais asseguram ao modelo gerado um desempenho preditivo mais eficaz dentre os conjuntos de hiperparâmetros testados.

Tabela 7: Hiperparâmetros do modelo após fase de otimização.

Hiperparâmetros	Valores
base_estimator_iterations	1000
base_estimator_learning_rate	0,2
base_estimator_depth	5
base_estimator_l2_leaf_reg	10
base_estimator_loss_function	RMSE
base_estimator_border_count	32
base_estimator_random_state	955
base_estimator	CatBoost Regressor
n_estimators	10

Fonte: Autor (2023).

Por meio da Tabela 8, é mostrado o Erro Absoluto Médio (EAM), o Quadro Médio do Erro (QME), o Erro Quadrático Médio (EQM) e o Coeficiente de Determinação (R^2),

porém, tendo por base apenas os resultados apresentados pelo CatBoost após a etapa final de otimização de hiperparâmetros. Por meio da aplicação do Método de Validação Cruzada (GKEREKOS *et al.*, 2019), dada linha expõe os resultados obtidos em uma rodada, na qual uma parte dos dados foi retirada do conjunto e recolocada na rodada posterior. Almejou-se reduzir o viés do modelo e evitar o sobreajuste do modelo em relação aos dados de treinamento e teste.

Tabela 8: Métricas de desempenho do modelo vencedor, CatBoost.

Rodada	EAM	QME	EQM	R^2
1	294,056	261096,97	510,976	0,931
2	325,923	381541,46	617,690	0,908
3	569,990	2726221,30	1651,127	0,693
4	237,294	465902,40	682,570	0,877
5	382,079	1078487,04	1038,502	0,788
6	429,818	796458,25	892,445	0,791
7	271,196	219112,81	468,094	0,919
8	297,043	378350,54	615,102	0,932
9	209,607	124925,29	353,447	0,937
10	261,891	339094,75	582,318	0,870
Média	327,890	677119,09	741,227	0,865

Fonte: Autor (2023).

A otimização dos hiperparâmetros executada provocou a melhora das métricas de desempenho do modelo Catboost, aumentando o Coeficiente de Determinação (R^2) em +1,5% e variando os demais: EAM em -5,91%, QME em -7,01% e EQM em -4,97%.

Enfatizando o Coeficiente de Determinação (R^2), sob direcionamento de Chicco *et al.* 2021, a rodada 9 apresentou o melhor desempenho, algo em torno de 93,7%. Por conseguinte, as rodadas 8 e 1, respectivamente, 93,2% e 93,1%, ocupam segundo e terceiro lugares, considerados valores altos de acurácia. Os menores resultados, por sua vez, ocorreram nas rodadas 3 e 5 em que os R^2 atingidos foram de 0,693 e 0,788.

4.5 Validação Holdout do Modelo Final

A validação final do modelo desenvolvido por meio do Método de Validação Hold-out: aleatoriamente escolhidos registros e separados do banco de dados inicial para aferir o desempenho do modelo frente aos novos dados de entrada. Foram separadas 125 embarcações, das quais 5 estão elencadas na Tabela 9. As numerações presentes na primeira linha indicam os preditores de entrada do modelo, bem como a variável-alvo.

Frente aos novos dados, a Tabela 10 traz os principais parâmetros de desempenho. O coeficiente de determinação atingido, por sua vez, foi de 0,910, demonstrando, assim, a sua alta performance mesmo quando para registros nunca antes vistos.

A análise residual do modelo teve papel crucial na verificação de desempenho do modelo de regressão de forma generalizada, conforme defendido por Okumus *et al.* (2021), tendo em vista que os residuais representam a diferença entre os valores estimados pelo modelo e os reais observados. Considerando as características adversas contidas no banco de dados, é notório, pela Figura 18, que o modelo apresentou predições divergentes aos dados reais observados ao longo das fases de treinamento e de teste, gerando residuais que permitiram ao modelo ajustar as próximas predições a partir destes residuais, aumentando seu desempenho final.

Tabela 9: Registros de embarcações do banco de dados utilizado.

	TE	MC	CT	BC	CD	PL	PM	VC	NP	TPB	DL	NG	CP	CG	NT	TV	CC
1	Lancha	Alumínio	17	3,45	1,2	1,9	215	22	1	35	7	0	100	0	3	0,45	70
2	Ferryboat	Aço Naval	32	8	5	6	450	19	1	528	98	2	99	670	8	36	3000
3	Lancha	Aço Naval	25,6	6,3	1,8	2,2	315	18	1	21	6	2	90	0	2	24	2500
4	Lancha	Aço Naval	24,89	6,3	1,4	2,3	550	18	1	21	6	2	90	0	2	24	2500
5	Passageiro/Carga Geral	Madeira	19	5,3	1,8	2,8	200	10	2	109	21	2	65	25	3	7	300

TE: Tipo de Embarcação; MC: Material do Casco; CT: Comprimento; BC: Boca; CD: Calado; PL: Pontal; PM: Potência do Motor; VC: Velocidade de Cruzeiro; NP: Número de Propulsores; TPB: Tonelagem Porte Bruto; DL: Deslocamento Leve; NG: Número de Geradores; CP: Capacidade de Passageiros; CG: Capacidade de Carga; NG: Número de Geradores; TV: Tempo de Viagem; CC: Consumo de Combustível.

Fonte: Autor (2023).

Tabela 10: Métricas de desempenho do modelo vencedor, CatBoost.

Modelo	EAM	QME	EQM	R^2
Catboost	274,5	346143,13	588,34	0,91

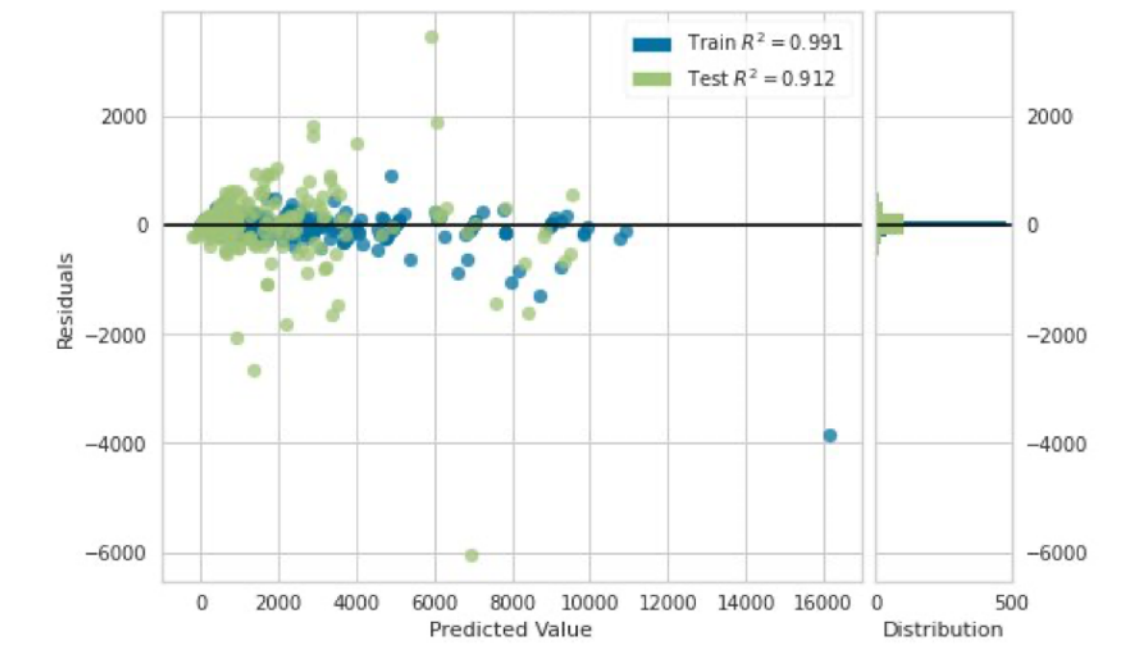
Fonte: Autor (2023).

A distribuição dos registros conforme Figura 18 caracteriza-se como função do valor da predição, no eixo das abscissas, pelo erro residual, no eixo das ordenadas. Quanto menor o erro residual, mais precisa a estimativa gerada pelo modelo. Os pontos em azul referem-se à fase de treinamento, enquanto que os pontos em verde, à fase de teste.

Inferre-se que o modelo performou muito bem durante a fase de treinamento, haja vista a concentração dos registros em torno da linha posicionada na origem e ótimo Coeficiente de Determinação (R^2) igual a 99,1%. Entretanto, é possível notar alguns outliers, predições distantes do valor real observado e que geraram altos residuais. O mesmo acontece quando considerando apenas as predições durante a fase de teste em que há um número maior de residuais, impactando em um menor Coeficiente de Determinação (R^2), 91,2%, indicando uma perda de desempenho do modelo, porém, dentro da normalidade, haja vista serem dados nunca antes vistos pelo modelo e provocados por meio do Método de Validação Cruzada, evitado enviesamento do modelo (GKEREKOS *et al.*, 2019).

Observa-se, pela Figura 10, que o modelo apresentou dificuldade na predição de registros a partir da faixa de 4000 litros, tendo gerado, inclusive, na fase de teste, três dos maiores residuais. Como consequência, este dado pode, certamente, influenciar no aumento de métricas como EQM, que penaliza grandes erros entre predição e real observado (CHICCO *et al.*, 2021).

Figura 18: Curva de residuais entre os valores preditos e observados.



Fonte: Autor (2023).

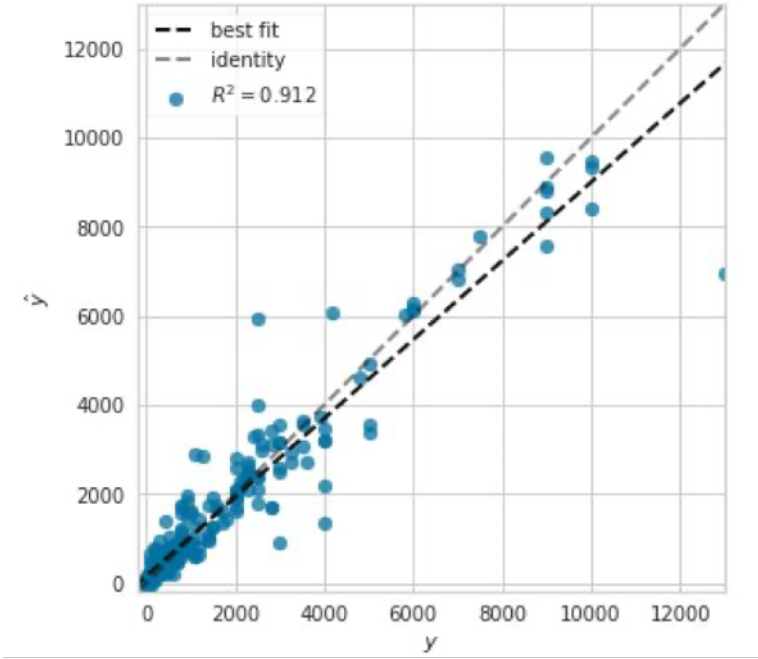
A distribuição dos registros na Figura 19 fez-se dependente de cada valor real do registro observado (eixo x) e das estimações promovidas pelo modelo durante a fase de treino (eixo y). A linha tracejada em preto representa uma regressão linear que melhor se adequa em estimar a correlação entre o valor previsto e medido da variável-alvo. A linha tracejada em cinza, denominada identidade, esboça uma função em que $y=x$, formando 45 graus, apenas para evidenciar se o modelo está super ou subestimado para os valores analisados (OKUMUS *et al.*, 2021).

O acúmulo de predições ao longo das linhas tracejadas para a faixa próxima à origem determina que o modelo tem uma capacidade assertiva para embarcações que consomem acima de 2000 litros. É notório, pela Figura 11, que a partir desta faixa, ocorre um número maior de outliers, ou seja, predições com altos residuais. Isto é causado porque:

- A parcela de embarcações que consomem mais do que 2000 litros é pequena se comparada ao restante do banco de dados (cerca de 29,3% do montante), deixando o modelo mais suscetível ao erro em caso de embarcações com dimensões discrepantes àquelas contidas em grande quantidade;
- Estes outliers podem ser um indicativo de um consumo de combustível além do

normal para embarcações com estas dimensões, evidenciando um comportamento operacional fraudulento.

Figura 19: Curva de performance aderente à condição ótima do modelo.



Fonte: Autor (2023).

5 CONCLUSÃO

Por meio desta pesquisa, constatou-se de maneira clara e fundamentada a importância vital do modal aquaviário na região amazônica. Através da análise minuciosa dos dados coletados, foi possível confirmar que muitas localidades na Amazônia dependem exclusivamente das embarcações como a única alternativa viável de transporte para passageiros e mercadorias. O fato de aproximadamente 56,9% das embarcações analisadas estarem dedicadas ao transporte misto, com as lanchas (20,3%) e ferry-boats (8,9%) também desempenhando papéis significativos, ressalta a centralidade desse meio de transporte na região.

Essa constatação não apenas corrobora a relevância do modal aquaviário na Amazônia, mas também destaca a necessidade de otimização do desempenho dessas embarcações para garantir a eficiência econômica e reduzir o impacto ambiental. O conhecimento científico adquirido em virtude desta investigação, bem como os resultados apresentados pelo modelo de predição desenvolvido, oferecem valiosas ferramentas para o planejamento e operação dessas embarcações.

Através da análise dos parâmetros relacionados ao consumo de combustível, foram identificados fatores cruciais que influenciam diretamente nesse consumo. Primeiramente, observa-se que o tempo de viagem, a distância e a velocidade das embarcações estão fortemente correlacionados com o volume de combustível consumido. Conforme o tempo de viagem aumenta, o consumo de óleo combustível tende a crescer, visto que uma maior duração da viagem demanda uma quantidade superior de combustível para a combustão, impactando diretamente nos custos operacionais. Dessa maneira, ressalta-se que a necessidade de planejamento preciso das rotas e operações das embarcações, de modo a cooperar com a economia significativa de combustível.

Viagens com itinerários mais longos, por sua vez, tendem a consumir mais combustível do que viagens de curta duração, o que reflete em gastos significativos para os armadores. No entanto, é importante destacar que a velocidade das embarcações apresenta uma correlação negativa, ou seja, quanto maior a velocidade de cruzeiro, menor é o consumo de combustível. Isso se deve ao fato de que viagens mais rápidas reduzem o tempo de viagem, compensando o aumento da velocidade com uma menor quantidade de combustível consumido.

Outro aspecto relevante é o tamanho das embarcações, expresso pelo comprimento, boca, calado e pontal. O comprimento da embarcação mostra-se como o atributo com a correlação mais forte em relação ao consumo de combustível, indicando que embarcações mais longas tendem a consumir uma quantidade considerável de combustível. Os outros atributos dimensionais e físicos, como boca, calado e pontal, também apresentam correlações positivas, sugerindo que embarcações maiores tendem a ter um consumo de combustível mais elevado.

Comprova-se, com isso, a importância de um projeto básico e detalhado para a construção adequada das embarcações, especialmente em uma região onde muitas delas são fabricadas com base apenas em projetos empíricos e sem validação técnica.

No que diz respeito ao material do casco das embarcações, observa-se que aquelas construídas em aço naval apresentam uma correlação positiva significativa com o consumo de combustível. Isso se deve à rugosidade do casco, que impacta diretamente no coeficiente de fricção, aumentando a resistência ao deslocamento na água. No entanto, embarcações de madeira mostram uma correlação negativa, indicando que seu consumo de combustível tende a ser menor. Do ponto de vista econômico, isso sugere que a construção de embarcações de madeira pode ser mais vantajosa devido ao menor consumo de combustível, sendo uma constatação valiosa para armadores e construtores de embarcações na escolha de materiais e designs apropriados.

Quanto aos parâmetros propulsivos, como o número de geradores, potência do motor e número de propulsores, também desempenham um papel importante no consumo de combustível. Embarcações maiores geralmente adotam um número maior de geradores como medida de segurança. A potência do motor determina a capacidade de transporte e velocidade da embarcação, influenciando diretamente o consumo de combustível. Além disso, um maior número de propulsores é necessário para atender à força de deslocamento ou planeio da embarcação, o que também contribui para um aumento no consumo de combustível.

A pesquisa identificou que a aplicação de técnicas de machine learning é uma abordagem altamente eficaz para prever o consumo de combustível das embarcações na região, superando abordagens empíricas e reduzindo os riscos associados à falta de conhecimento técnico. O uso desses modelos não apenas contribui para o planejamento econômico eficiente das viagens, mas também oferece uma base sólida para a tomada

de decisões relacionadas à gestão de combustível, substituindo métodos empíricos e tradicionais.

O modelo de maior acurácia foi o CatBoost, com uma precisão de 85,2%. Essa alta precisão se deve à técnica exclusiva de tratamento de variáveis categóricas, tornando-o robusto a outliers e adequado para dados tabulares. Em segundo, terceiro e quarto lugares, encontram-se o Gradient Boosting, Random Forest e Gradient Boosting, com coeficientes de determinação de 79,7%, 79% e 78,7%, respectivamente. Nas últimas posições, estão Árvores Extras, com 75,1%, e Árvore de Decisão, com 66,95%, este último apresentando o menor coeficiente de determinação entre os seis modelos considerados.

Quanto aos parâmetros de desempenho QME e EQM, seguindo a mesma ordem de desempenho dos modelos em relação ao coeficiente de determinação (R^2), o CatBoost apresentou as melhores previsões, resultando em QME e EQM mais baixos em comparação com os outros modelos. Em relação ao EAM, que mensura o erro absoluto médio, o CatBoost também se destacou com o menor erro, seguido por Florestas Aleatórias, Aumento Extremo de Gradiente e demais.

Quando se analisa o tempo computacional, a Árvore de Decisão foi o mais rápido em processar os dados e gerar resultados, embora tenha apresentado um erro mais elevado e previsões menos precisas. Em contraste, o CatBoost exigiu o maior tempo de processamento. Os modelos Aumento de Gradiente e Florestas Aleatórias obtiveram bons desempenhos nas previsões, mas também demandaram tempo de processamento maior.

O CatBoost foi escolhido como o modelo vencedor devido ao seu desempenho satisfatório no processamento de dados e na previsão da variável-alvo. A otimização de hiperparâmetros e a validação hold-out resultaram em um coeficiente de determinação de aproximadamente 91%, o que comprova a eficiência do modelo proposto na previsão do consumo de combustível.

Portanto, sabidos os atributos de maior importância, bem como desenvolvido um sistema de previsão do consumo de combustível, estudo que certamente contribui para um melhor planejamento acerca das viagens, bem como ajuda na evolução do conhecimento científico acerca do comportamento das embarcações intrínsecas à Amazônia, vislumbra-se a aplicação da metodologia em demais regiões e bancos de dados de embarcações, de modo que seja possível uma análise comparativa entre os métodos construtivos e operacionais

adotados por cada região, além da inserção dos parâmetros relacionados às condições ambientais, não abordados no desenvolvimento do modelo.

REFERÊNCIAS

ABEBE, Misganaw; SHIN, Yongwoo; NOH, Yoojeong; LEE, Sangbong; LEE, Inwon. **Machine Learning Approaches for Ship Speed Prediction towards Energy Efficient Shipping**. Applied Sciences, 10(7), 2325, (2020).

ADHIKARI, Animesh; ADHIKARI, Jhimli. **Advances in knowledge discovery in databases**. Springer. 2015.

AFRIFA-YAMOA, E.; MUELLER, UA.; TAYLOR, SM.; FISHER, AJ. **Missing data imputation of high-resolution temporal climate time series data**. Meteorological Applications, v. 27, n. 1, p. e1873, 2020.

AHMAD, Ayaz; OSTROWSKI, Krzysztof Adam; MASLAK, Mariusz; FAROOQ, Furqan; MEHMOOD, Imran; NAFEES, Afnan. **Comparative study of supervised machine learning algorithms for predicting the compressive strength of concrete at high temperature**. Materials, v. 14, n. 15, p. 4222, 2021.

ALAWADI, Sadi; MERA, David; DELGADO, Manuel Fernández; ALKHABBAS, Fahed; OLSSON, Carl Magnus; DAVIDSSON, Paul. **A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings**. Energy Systems, p. 1-17, 2020.

ALPAYDIN, Ethem. **Introduction to Machine Learning**. MIT press, 2014.

ANCONA, Maria Alessandra; BALDI, Francesco; BIANCHI, Michele; BRANCHINI, Lisa; MELINO, Francesco; PERETTO, Antonio; ROSATI, Jessica. **Efficiency improvement on a cruise ship: Load allocation optimization**. Energy Conversion and Management, 164, 42-58, 2018.

ANDERSON, Kevin; BOWS, Alice. **Executing a Scharnow turn: reconciling shipping emissions with international commitments on climate change**. Carbon Manage. 3 (6), 615–628, 2012.

AVGOULEAS, Kyriakos; SCLAVOUNOS, Paul D. **Fuel-efficient ship routing**. Nausivios Chora, v. 5, p. 39-72, 2014.

BALDI, Francesco; AHLGREN, Fredrik; MELINO, Francesco; GABRIELII, Cecilia; ANDERSSON, Karin. **Optimal load allocation of complex ship power plants**. Energy Conversion and Management, 124, 344-356, 2016.

- BARUA, Limon; ZOU, Bo; ZHOU, Yan. **Machine learning for international freight transportation management: a comprehensive review**. Research in Transportation Business & Management, v. 34, p. 100453, 2020.
- BATISTA, Gustavo Enrique. **Pré-processamento de dados em aprendizado de máquina supervisionado**. 2003. Tese de Doutorado. Universidade de São Paulo.
- BAN, Bo; YANG, Junjie; CHEN, Pengguang; XIONG, Jianbin; WANG, Qinruo. **Ship track regression based on support vector machine**. IEEE, 2017.
- BERGSTRA, James; BARNEDET, Remi; BENGIO, Yoshua; KEGI, Balazs. **Algorithms for hyper-parameter optimization**. Advances in neural information processing systems, v. 24, 2011.
- BERGSTRA, James; BENGIO, Yoshua. **Random search for hyper-parameter optimization**. Journal of machine learning research, v. 13, n. 2, 2012.
- BESİKÇI, Bal; ARSLAN, O.; TURAN, O.; ÖLÇER, A. I. **An artificial neural network based decision support system for energy efficient ship operations**. Computers & Operations Research, 66, 393-401, 2019.
- BIALYSTOCKI, Nicolas; KONOVESSIS, Dimitris. **On the estimation of ship's fuel consumption and speed curve: A statistical approach**. Journal Ocean Engineering and Science, pp.157-166, 2016.
- BISHOP, Cristopher. **Pattern Recognition and Machine Learning (Information Science and Statistics)**. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- BOUMAN, Evert; LINDSTAD, Elizabeth; RIALLAND, Agathe; STRØMMAN, Anders. **State-of-the-art technologies, measures, and potential for reducing GHG emissions from shipping—a review**. Transportation Research Part D: Transport and Environment, 52, 408-421, 2017.
- BRASIL. Ministério do Meio Ambiente/Secretaria de Recursos Hídricos. **Caderno da Região Hidrográfica Amazônica**. Brasília: MMA, 2006.
- BRASIL. Presidência da República. Secretaria de Portos. Agência Nacional de Transportes Aquaviários. **Caracterização da oferta e da demanda do transporte fluvial de passageiros da região**. Brasília: ANTAQ, 2013.
- BRASIL. Presidência da República. Secretaria de Portos. Agência Nacional de

Transportes Aquaviários. **Caracterização da oferta e da demanda do transporte fluvial de passageiros da região**. Brasília: ANTAQ, 2015.

BRASIL. Presidência da República. Secretaria de Portos. Agência Nacional de Transportes Aquaviários. **Caracterização da oferta e da demanda do transporte fluvial de passageiros da região**. Brasília: ANTAQ, 2017.

BRASIL. Ministério dos Transportes. **Projeto de Reavaliação de Estimativas e Metas do PNLT**. Brasília: MMA, 2012.

BREIMAN, Leo. **Random forests**. Machine learning, v. 45, n. 1, p. 5-32, 2001.

BRILLANTE, Luca; GAIOTTI, Federica; LOVAT, Lorenzo; VINCENZI, Simone; GIACOSA, Simone; TORCHIO, Fabrizio; SEGADE, Susana; ROLLE, Luca; TOMASI, Diego. **Investigating the use of gradient boosting machine, random forest and their ensemble to predict skin flavonoid content from berry physical–mechanical characteristics in wine grapes**. Computers and Electronics in Agriculture, v. 117, p. 186-193, 2015.

BRYNOLF, Selma; BALDI, Francesco; JOHNSON, Hannes. **Energy efficiency and fuel changes to reduce environmental impacts**. Shipping and the Environment (pp. 295-339). Springer, Berlin, Heidelberg, 2016.

NETO, Campos; DE PAULA, Pepino; POMPERMAYER, Fabiano. **Hidrovias brasileiras: elevado potencial, mas por que não são implantadas?**. Gargalos e soluções na infraestrutura de transportes. 1. ed. Rio de Janeiro: Editora FGV, 2014a. p. 243–284, 2014.

CARAVIELLO, D. Z.; PATEL, R.; PAI, R. **U.S. Patent No. 10,102,476**. Washington, DC: U.S. Patent and Trademark Office, 2018.

CARIOU, Pierre. **Is slow steaming a sustainable means of reducing CO2 emissions from container shipping?**. Transp Res Part D Transp Environ, 2011.

CHANG, Ching-Ching; CHANG, Chia-Hong. **Energy conservation for international dry bulk carriers via vessel speed reduction**. Energy Policy, 59:710–5, 2013.

CHEN, Tianqi; GUESTRIN, Carlos. **Xgboost: A scalable tree boosting system**. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016. p. 785-794.

- CHICCO, Davide; WARRENS, Matthijs J.; JURMAN, Giuseppe. **The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation.** PeerJ Computer Science, v. 7, p. e623, 2021.
- CIPOLLINI, F.; ONETO, L.; CORADDU, A.; MURPHY, A.J.; ANGUIA, D. **Condition based maintenance of naval propulsion systems with supervised data analysis.** Ocean Eng. 149, 268–278, 2018.
- CNT. **Pesquisa CNT do transporte aquaviário – Cabotagem.** Brasília: Confederação Nacional do Transporte, 2013.
- CNT. **Plano o transporte que move o Brasil.** Brasília: Confederação Nacional de Transporte, 2018.
- CORADDU, A.; ONETO, L.; BALDI, F.; ANGUIA, D. **Vessels fuel consumption forecast and trim optimisation: A data analytics perspective.** Ocean Eng. 130, 351–370, 2017.
- COUTO, Mayra Hermínia. **Qualidade de vida dos usuários do transporte fluvial na Amazônia.** Tese de doutorado. Universidade Federal do Pará, Núcleo de Altos Estudos Amazônicos, Pós-graduação em Desenvolvimento Sustentável do Trópico Úmido, Belém, 2015.
- CUNHA, Evelinn Flores; INSPER JR, A. S. **O desafio na regulação do transporte hidroviário intermunicipal no Amazonas.** Brazilian Journal of Development, v. 6, n. 3, p. 15736-15752, 2020.
- DAO, Fu-Ying; LV, Hao; WANG, Fang; FENG, Chao-Qin; DING, Hui; CHEN, Wei; LIN, Hao. **Identify origin of replication in Saccharomyces cerevisiae using two-step feature selection technique.** Bioinformatics, v. 35, n. 12, p. 2075-2083, 2019.
- DAVID, Robert Carvalho. **Esse rio é minha rua: perspectivas para o transporte fluvial de passageiros no Amazonas.** Manaus: Universidade Federal do Amazonas (UFMA), 2019.
- DAWOOD, Edel Goreil. **Geo-locating UEs Using Multi-output Decision Tree Regressor.** 2019. Tese de Doutorado.
- DE MAESSCHALCK, Roy; JOUAN-RIMBAUD, Delphine; MASSART, Désiré L. **The**

mahalanobis distance. Chemometrics and intelligent laboratory systems, v. 50, n. 1, p. 1-18, 2000.

DIETTERICH, Thomas G. **Approximate statistical tests for comparing supervised classification learning algorithms.** Neural computation, v. 10, n. 7, p. 1895-1923, 1998.

DUARTE, R.; KUWAHARA, Nelson; ALENCAR, Larissa. **Perspectiva Ergonômica para Embarcações do Estado do Amazonas.** Vitória: 2009.

EIDE, Magnus; LONGVA, Tore; HOFFMANN, Peter; ENDRESEN, Øyvind; DALSRØREN, Stig. **Cost scenarios for reduction of ship CO2 & Management.** 38(1), 11-37, 2011.

EL NOSHOKATY, Said. **Shipping optimisation systems (SOS): liner optimisation perspective.** Int J Shipp Transport Logist, 2013.

EPL. Empresa de Planejamento e Logística. **Transporte interregional de cargas no Brasil: Panorama 2015.** Brasília: EPL, 2016.

EPSTEIN, Marc. **Making sustainability work: Best practices in managing and measuring corporate social environmental and economic impacts.** Londres: Routledge, 2008.

FABER, Jasper; NELISSEN, Dagmar; HON, Galen; WANG, Haifeng; TSIMPLIS, Mikis. **Regulated slow steaming in maritime transport: an assessment of options.** Costs Benefits; 7 442-1, 2012.

FAGERHOLT, Kjetil; LAPORTE, Gilbert; NORSTAD, Inge. **Reducing fuel emissions by optimizing speed on shipping routes.** J. Oper. Res. Soc., 61, 523–529, 2010.

FAN, Junliang; WANG, Xiukang; WU, Lifeng; ZHOU, Hanmi; ZHANG, Fucang; YU, Xiang; LU, Xianghui; XIANG, Youzhen. **Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China.** Energy conversion and management, 164, 102-111, 2018.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. **From data mining to knowledge discovery in databases.** AI Mag 17:37–54, 1996.

FEI, Xihong; FANG, Yi; LING, Qiang. **Discrimination of excessive exhaust**

emissions of vehicles based on Catboost algorithm. 2020 Chinese Control And Decision Conference (CCDC). IEEE, 2020. p. 4396-4401.

FERREIRA, Vitor Hugo. **O setor dos transportes de mercadorias em Portugal: a intermodalidade enquanto fator dinamizador das empresas exportadoras.** Porto: Faculdade de Economia, 2013.

FERREIRA, Márcio Antônio. **Transporte fluvial por embarcações mistas no Amazonas: uma análise do trecho Manaus-Coari e Manaus-Parintins.** Tese (Doutorado em Ciências Ambientais e Sustentabilidade na Amazônia) – Universidade Federal do Amazonas, Manaus, 2016.

FESTUS, Dick Ibitoru; SAMSON, Nitonye. **Effect of fluid density on ship hull resistance and powering.** International Journal of Engineering Research and General Science, [online] 3(1), p.623, 2015.

FRIEDMAN, Jerome H. **Stochastic gradient boosting.** Computational statistics & data analysis, v. 38, n. 4, p. 367-378, 2002.

FIGUEIREDO, Nelio; MORAES, Hito; LOUREIRO, Emannuel; LAMEIRA, Pedro. **Caracterização da Oferta e da Demanda do Transporte Fluvial de Passageiros e Cargas na Região Amazônica,** (2018).

FROTA, Cláudio Dantas. **Gestão da qualidade aplicada às empresas prestadoras do serviço de transporte hidroviário de passageiros na amazônia ocidental: Uma proposta prática.** Rio de Janeiro: UFRJ/ COPPE/ Programa de Engenharia de Transportes, 2008.

GAINZA, Jorge Alberto Nuñez; BRINATI, Hernani L. **Análise da operação de navios porta contêineres em velocidade reduzida.** Instituto Pan-Americano de Engenharia Naval, p. 1-15, 2010.

GEURTS, Pierre; ERNST, Damien; WEHENKEL, Louis. **Extremely randomized trees.** Machine learning, v. 63, n. 1, p. 3-42, 2006.

GKEREKOS, Christos; LAZAKIS, Iraklis; THEOTOKATOS, Gerasimos. **Machine learning models for predicting ship main engine Fuel Oil Consumption: A comparative study.** Ocean Engineering, v. 188, p. 106282, 2019.

GOLIAS, Mihalis; SAHARIDIS, Georgios; BOILE, Maria; THEOFANIS, Sotirios;

- IERAPETRITOU, Marianthi. **The berth allocation problem: optimizing vessel arrival time.** Maritime Economics Logistics, vol. 11, no. 4, pp. 358–377, 2009.
- GUAN, Cong; THEOTOKATOS, Gerasimos; ZHOU, Peilin; CHEN, Hui. **Computational investigation of a large containership propulsion engine operation at slow steaming conditions.** Appl Energy, 2014.
- GUARASCIO, Massimo; MANCO, Giuseppe; RITACCO, Ettore. **Knowledge discovery in databases.** Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, p. 336, 2018.
- GUYON, Isabelle; ELISSEEFF, André. **An introduction to feature extraction.** Springer, Berlin, Heidelberg, 2006. p. 1-25.
- HAN, Jiawei; PEI, Jian; TONG, Hanghang. **Data mining: concepts and techniques.** Morgan Kaufmann, 2022.
- HARANEN, Michael; PAKKANEN, Pekka; KARIRANTA, Risto; SALO, Jouni. **White, grey and black-box modeling in ship performance evaluation.** Hull Performance and Insight Conference, Pavone Canavese, Torino, Italy, 2016.
- HAREL, Ofer. **The estimation of R² and adjusted R² in incomplete data sets using multiple imputation.** Journal of Applied Statistics, v. 36, n. 10, p. 1109-1118, 2009.
- HARVALD, Svend Aage. **Prediction of Power of Ships.** Department of Ocean Engineering, Technical University of Denmark, Lyngby, Denmark, 1977.
- HASSELAAR, Thijs Willem. **An investigation into the development of an advanced ship performance monitoring and analysis system.** Tese de Doutorado - University of Newcastle. 2011.
- HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The elements of statistical learning: data mining, inference, and prediction.** Springer Science & Business Media, 2009.
- DE GRADO, Herradon; BETRAM, Volker. **Predicting added resistance in wind and waves employing artificial neural nets.** Proceedings of 1st Hull Performance & Insight Conference HullPIC'16, Padova, Italy, pp. 14e22, 2016.

HOLLOWAY, Jacinta; HELMSTEDT, Kate; MENGERSEN, Kerrie; SCHMIDT, Michael. **A decision tree approach for spatially interpolating missing land cover data and classifying satellite images**. *Remote Sensing*, v. 11, n. 15, p. 1796, 2019.

HU, Zhihui; ZHOU, Tianrui; OSMAN, Tarmizi Mohd; LI, Xiaohe; JIN, Yongxin; ZHEN, Rong. **A novel hybrid fuel consumption prediction model for ocean-going container ships based on sensor data**. *Journal of Marine Science and Engineering*, v. 9, n. 4, p. 449, 2021.

HUANG, Guomin; WU, Lifeng; ZHANG, Weiqiang; FAN, Junliang; YU, Xiang; ZENG, Wenzhi; ZHOU, Hanmi. **Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions**. *Journal of Hydrology*, v. 574, p. 1029-1041, 2019.

IZBICKI, Rafael; SANTOS, T. M. **Machine learning sob a ótica estatística: Uma abordagem preditivista para estatística com exemplos em R**. 2018.

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. **An introduction to statistical learning**. Vol 112. Springer, New York, 2013.

JEON, Miyeon; NOH, Yoojeong; SHIN, Yongwoo; LIM, O-Kaung; LEE, Inwon; CHO, Daeseung. **Prediction of ship fuel consumption by using an artificial neural network**. *Journal of Mechanical Science and Technology*, 32, 5785-5796, 2018.

JEON, Miyeon; NOH, Yoojeong; SHIN, Yongwoo; LIM, O-Kaung; LEE, Inwon; CHO, Daeseung. **Prediction of ship fuel consumption by using an artificial neural network**. *Journal of Mechanical Science and Technology*, v. 32, n. 12, p. 5785-5796, 2018.

JIAN, Songlei; CAO, Longbing; PANG, Guansong; LU, Kai; GAO, Hang. **Embedding-based representation of categorical data by hierarchical value coupling learning**. *IJCAI International Joint Conference on Artificial Intelligence*, 2017.

JONES, Chris. **Expectiles and M-quantiles are quantiles**. *Statistics & Probability Letters*, v. 20, n. 2, p. 149-153, 1994.

JUNIOR, Betarelli; AMBROSIO, Diogo; FARIA, Weslem; MONTENEGRO, Rosa. (2018). **Interdependências, influências e produtividade: O papel do transporte aquaviário nas economias mundiais**. *Revista Econômica do Nordeste*, 49(4), 53-73.

KANELLOS, Fotios. **Optimal power management with GHG emissions limitation in all-electric ship power systems comprising energy storage systems.** IEEE Trans Power Syst, 2014.

KANG, John; SCHWARTZ, Russell; FLICKINGER, John; BERIWAL, Sushil. **Machine learning approaches for predicting radiation therapy outcomes: A clinician's perspective.** International Journal of Radiation Oncology, Biology, Physics, 93(5), 1127–1135, 2015.

KARAGIANNIDIS, Pavlos; THEMELIS, Nikos; ZARAPHONITIS, George; SPANDONIDIS, Christos; GIORDAMLIS, Christos. **Ship Fuel Consumption Prediction using Artificial Neural Networks,** 2019.

KEE, Keh-Kim; SIMON, B.-Y. Lau; RENCO, K.-H. Yong. **Prediction of ship fuel consumption and speed curve by using statistical method.** J. Comput. Sci. Comput. Math, v. 8, n. 2, p. 19-24, 2018.

KIM, Ki-Su; ROH, Myung-Il. **ISO 15016: 2015-based method for estimating the fuel oil consumption of a ship.** Journal of Marine Science and Engineering, v. 8, n. 10, p. 791, 2020.

KITSIKOUDIS, Vasileios; SIDIROPOULOS, Epaminondas; HRISSANTHOU, Vlassios. **Machine learning utilization for bed load transport in gravel-bed rivers.** Water resources management, v. 28, n. 11, p. 3727-3743, 2014.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling.** New York: Springer, 2013.

KWAK, Sang Kyu; KIM, Jong Hae. **Statistical data preparation: management of missing values and outliers.** Korean journal of anesthesiology, v. 70, n. 4, p. 407-411, 2017.

LAVALLE, Steve; LESSER, Eric; SHOCKLEY, Rebecca; HOPKINS, Michael; KRUSCHWITZ, Nina. **Big data, analytics and the path from insights to value.** MIT Sloan Management Review, 52(2), 21. 2011.

LAZAKIS, Iraklis; RAPTODIMOS, Yannis; VARELAS, T. **Predicting ship machinery system condition through analytical reliability tools and artificial neural networks.** Ocean Eng., 2018.

- LAZAKIS, Iraklis; GKEREKOS, Christos; THEOTOKATOS, Gerasimos. **Investigating an SVM-driven, one-class approach to estimating ship systems condition.** Ships Offshore Struct, 2019.
- LEE, Jisun; SEONOH, Yoo; SANGkyu, Choi; HEETAEK, Kim; CHUNBEOM, Hong; JONGSOO, Seo. **Development and application of trim optimization and parametric techniques using an evaluation system (SoLuTion) based on the RANS for improvement of EEOI.** ASME, p. 784–91. vol. 2 CFD VIV, 2013.
- LEE RODGERS, Joseph; NICEWANDER, W. Alan. **Thirteen ways to look at the correlation coefficient.** The American Statistician, v. 42, n. 1, p. 59-66, 1988.
- LEYS, Christophe; KLEIN, Olivier; DOMINICY, Yves; LEY, Christophe. **Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance.** Journal of experimental social psychology, v. 74, p. 150-156, 2018.
- LIANG, Yunxin; WU, Jiyu; CAO, Yujun; ZHONG, Biliang; CHEN, Zhenkun; LI, Zhenzhang. **Product marketing prediction based on XGboost and LightGBM algorithm.** Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition. 2019. p. 150-153.
- LINDSTAD, Haakon; ESKELAND, Gunnar. **Environmental regulations in shipping: Policies leaning towards globalization of scrubbers deserve scrutiny.** Transport Res Part D: Transport Environ; 47:67–76, 2016.
- LIU, Yong; LIAO, Shizhong; JIANG, Shali; DING, Lizhong; LIN, Hailun; WANG, Weiping. **Fast cross-validation for kernel-based algorithms.** IEEE transactions on pattern analysis and machine intelligence, v. 42, n. 5, p. 1083-1096, 2019.
- LU, Ruihua; TURAN, Osman; BOULOUGOURIS, Evangelos. **Voyage optimization: Prediction of ship specific fuel consumption for energy efficient shipping.** Anais da 3ª Conferência Internacional sobre Tecnologias, Operações, Logística e Modelagem para Transporte de Baixo Carbono, Londres, UK, 9–11, 2013; pp. 1–11.
- LUNDBERG, Scott M.; LEE, Su-In. **A unified approach to interpreting model predictions.** Advances in neural information processing systems, v. 30, 2017.
- MAGNUSSEN, Anna Karina. **Rational calculation of sea margin.** Ph.D. Norwegian University of Science Application of AIS Data in Vessel Performance Analysis. Master

Thesis, 2017.

MAK, Lawrence; SULLIVAN, Michael; KUCZORA, Andrew; MILLAN, James. **Ship performance monitoring and analysis to improve fuel efficiency**. Oceans-St. John's. IEEE, p. 1-10, 2014.

MALONI, Michael; PAUL, Jomon Aliyas; GLIGOR, David. **Slow steaming impacts on ocean carriers and shippers**. Maritime Economics & Logistics, 15(2), 151–171, 2013.

MANDER, Sarah. **Slow steaming and a new dawn for wind propulsion: A multi-level analysis of two low carbon shipping transitions**. 18th International Conference on Ships and Shipping Research., 75, pp.210-216, 2017.

MANJU, N.; HARISH, B. S.; PRAJWAL, V. **Ensemble feature selection and classification of internet traffic using XGBoost classifier**. International Journal of Computer Network and Information Security, v. 10, n. 7, p. 37, 2019.

MANSOURI, Afshin; LEE, Habin; ALUKO, Oluwakayode. **Multi-objective decision support to enhance environmental sustainability in maritime shipping: A review and future directions**. Transportation Research Part E: Logistics and Transportation Review, 78(2015), 3–18, 2015.

MANYIKA, James; CHUI, Michael; BROWN, Brad; BUGHIN, Jacques; DOBBS, Richard; ROXBURGH, Charles; BYERS, Angela. **Big data: the next frontier for innovation, competition, and productivity**. McKinsey Global Institute, 2011.

MARCILIO-JR, Wilson; ELER, Danilo. **Explaining dimensionality reduction results using Shapley values**. Expert Systems with Applications, v. 178, p. 115020, 2021.

MARINE INSIGHT. **What is Nitrogen Oxides or NOx air pollution from Ships?**. 2017.

MENG, Qiang; DU, Yuquan; WANG, Yadong. **Shipping log data based container ship fuel efficiency modeling**. Transp. Res. B 83, 207–229, 2016.

MINISTÉRIO DOS TRANSPORTES, PORTOS E AVIAÇÃO CIVIL – MTPA. Secretaria de Portos da Presidência da República – SEP/PR. **Plano Nacional de Logística Portuária – PNLP, Relatório de Diagnostico**. Estudo (2015).

- MOHR, Felix; WEVER, Marcel; HÜLLERMEIER, Eyke. **ML-Plan: Automated machine learning via hierarchical planning**. Machine Learning, v. 107, n. 8, p. 1495-1515, 2018.
- MOLLAND, Anthony; TURNOCK, Stephen; HUDSON, Dominic. **Ship Resistance and Propulsion, Practical Estimation of Propulsive Power**. New York: Cambridge University Press, 2011.
- MORGADO, Andréa; PORTUGAL, Licínio; MELLO, Andréa. **Acessibilidade na Região Amazônica através do transporte hidroviário**. Journal of Transport Literature, v.7, n.2, p. 97-123. Rio de Janeiro: Universidade Federal do Rio de Janeiro (UFRJ), 2013.
- MOUSTAFA, Moustafa; YEHA, W.; HUSSEIN, Arwa. **Energy efficient operation of bulk carriers by trim optimization**. Proc 18th int conf ships shipp res, Lecco, Italy; 2015.
- NPL, National Physical Laboratory. **Physical properties of sea water**. 2018.
- NAZARI, Mohammadreza; OROOJLOOY, Afshin; SNYDER, Lawrence; TAKÁC, Martin. **Reinforcement learning for solving the vehicle routing problem**. In Advances in Neural Information Processing Systems, 9839–9849, 2018.
- NEIVA, Christian; MENDONÇA, Joaquin. **Estudo do setor de transporte do brasil e sua influência na competitividade das commodities nacionais**, 2019.
- NGUYEN, Xuan-Vinh; NGUYEN, Phu-Khoa; NGUYEN, Vu. **Clustering automation test faults**. 11th International Conference on Knowledge and Systems Engineering (KSE). IEEE, 2019. p. 1-7.
- NORSTAD, Inge; FAGERHOLT, Kjetil; LAPORTE, Gilbert. **Tramp ship routing and scheduling with speed optimization**. Transp Res Part C Emerg Technol;19:853–65, 2011.
- NISHIDA, Tomoya; KATORI, Masanobu; OHUCHI, Kazuyuki; WASEDA, Takuji. **Optimization of integrated weather routing systems for sailing cargo ships**. Proceeding 21st int offshore polar eng conf, Maui, USA, vol. 8; p. 283–9, 2011.
- NWAOHA, Thaddeus; OMBOR, Garrick; OKWU, Modestus. **A combined algorithm approach to fuel consumption rate analysis and prediction of sea-worthy diesel**

engine-powered marine vessels. Proc. Inst. Mech. Eng. Part M J. Eng. Marit. Environ. 231, 542–554, 2016.

OKUMUŞ, Fatih; EKMEKÇIOĞLU, Araks; KARA, Selin Soner. **Modelling ships main and auxiliary engine powers with regression-based machine learning algorithms.** Polish Maritime Research, 2021.

OLEKSIY, Bondarenko; FUKUDA, Tetsugo; YOO, Dong-Hoon; TANIZAWA, Katsuji. **Development of Diesel Engine Simulator for Use with Self-Propulsion Ship Model,** 48(5), pp.684-691, 2013.

OLIVEIRA, Ana Caroline Dorxas de. **Os benefícios da implantação da hidrovía Araguaia Tocantins para o transporte de mercadorias.** Araguaína: Universidade Federal do Tocantins, 2020.

ORIHARA, Hideo; TSUJIMOTO, Masaru. **Performance prediction of full-scale ship and analysis by means of Part 2: Validation of full-scale ship and analysis predictions in actual seas.** Journal of and Technology, 1-20, 2017.

OSHIRO, Thais Mayumi; PEREZ, Pedro Santoro; BARANAUSKAS, José Augusto. **How many trees in a random forest?.** Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 8. Springer Berlin Heidelberg, 2012. p. 154-168.

PADOVEZI, Carlos Daher. **Conceito de embarcações adaptadas à via aplicado à navegação fluvial no Brasil.** São Paulo: Escola Politécnica da Universidade de São Paulo Departamento de Engenharia Naval e Oceânica, 2003.

PANAPAKIDIS, Ioannis; SOURTZI, Vasiliki-Marianna; DAGOUMAS, Athanasios. **Forecasting the fuel consumption of passenger ships with a combination of shallow and deep learning.** Electronics, v. 9, n. 5, p. 776, 2020.

PANDEY, Amit; JAIN, Achin. **Comparative analysis of KNN algorithm using various normalization techniques.** International Journal of Computer Network and Information Security, v. 9, n. 11, p. 36, 2017.

PANI, Claudia. **Managing vessel arrival uncertainty in container terminals: A machine learning approach.** 2014.

PEIXOTO, T. F. de A. **Quadro Comparativo entre as Diferentes Modalidades**

de Transporte de Mercadorias na Amazônia. Estudos de transporte e logística na Amazônia. Manaus, AM: Novo Tempo, 2006.

PETERSEN, Joan; WINTHER, Ole; JACOBSEN, Daniel. **A machine-learning approach to predict main energy consumption under realistic operational conditions.** Ship Technol. Res. 59 (1), 64–72, 2012.

POTDAR, Kedar; PARDAWALA, Taher S.; PAI, Chinmay D. **A comparative study of categorical variable encoding techniques for neural network classifiers.** International journal of computer applications, v. 175, n. 4, p. 7-9, 2017.

PROBST, Philipp; WRIGHT, Marvin; BOULESTEIX, Anne-Laure. **Hyperparameters and tuning strategies for random forest.** Wiley Interdisciplinary Reviews: data mining and knowledge discovery, v. 9, n. 3, p. e1301, 2019.

PROKHORENKOVA, Liudmila; GUSEV, Gleb; VOROBEV, Aleksandr; DOROGUSH, Anna Veronika, GULIN, Andrey. **CatBoost: unbiased boosting with categorical features.** Advances in neural information processing systems, 31, 2018.

PSARAFTIS, Harilaos; KONTOVAS, Christos. **Green maritime transportation: Speed and route optimization.** Green transportation logistics. Springer, Cham, p. 299-349, 2016.

QIU, Xuan; WONG, Eugene YC; LAM, Jasmine Siu Lee. **Evaluating economic and environmental value of liner vessel sharing along the maritime silk road.** Maritime Policy & Management, 45(3), 336–350, 2018.

RAPTODIMOS, Yiannis; LAZAKIS, Iraklis. **Using artificial neural network-self-organising map for data clustering of marine engine condition monitoring applications.** Ships Offshore Struct, 2018.

REHMATULLA, Nishatabbas; CALLEYA, John; SMITH, Tristan. **The implementation of technical energy efficiency and CO2 emission reduction measures in shipping.** Ocean Engineering. 139, 184e197, 2017.

REICHEL, Maciej; MINCHEV, A.; LARSEN, NL. **Trim optimisation-theory and practice.** Int. J. Marine Navig. Saf. Sea Transp. 8(3), 387–392, 2014.

RIBEIRO, Marco Tulio; SINGH, Sameer; GUESTIN, Carlos. **“Why should I trust you?” Explaining the predictions of any classifier.** Proceedings of the 22nd ACM

- SIGKDD international conference on knowledge discovery and data mining. 2016. p. 1135-1144.
- SANABRA, Castells Sanabra; SANTAMARÍA, Juan Usabiaga; OSÉS, Martínez. **Manoeuvring and hotelling external costs: enough for alternative energy sources?**. 41(1), pp.42-60, 2013.
- SARKIS, Joseph. **Greening the supply chain**. Londres: Springer, 2006.
- SCHILLER, Rodrigo Achilles. **Análise da eficiência energética em navios mercantes e estudo de caso do consumo de combustível em navio aliviador do tipo Suezmax**. 2016. Universidade de São Paulo.
- SCHOBER, Patrick; BOER, Christa; SCHWARTE, Lothar A. **Correlation coefficients: appropriate use and interpretation**. Anesthesia & Analgesia, v. 126, n. 5, p. 1763-1768, 2018.
- SENN, Luiz Afonso. **Economia e planejamento dos transportes**. Rio de Janeiro: Elsevier, 2014.
- SHAO, Wei; ZHOU, Peilin; THONG, Sew Kait. **Development of a novel forward dynamic programming method for weather routing**. Journal Marine Science Technol. 11:239–51, 2006.
- SHAPLEY, Lloyd. **A value for n-person games**. Contributions to the Theory of Games (AM-28), Volume II. Princeton University Press, 2016. p. 307-318.
- SHARIFI, Yaser; GHASSEMI, Hassan; ZANGANEH, Hamid. **Various innovative technologic devices in shipping energy saving and diminish fuel consumption**. Int. J. Phys, v. 5, p. 21-29, 2017.
- SIMONSEN, Morten; WALNUM, Hans Jakob; GÖSSLING, Stefan. **Model for estimation of fuel consumption of cruise ships**. Energies 11 (5), 1059. 2018.
- SINGH, Dalwinder; SINGH, Birmohan. **Investigating the impact of data normalization on classification performance**. Applied Soft Computing, v. 97, p. 105524, 2020.
- SINGH, Koushlendra; KUMAR, Suraj; DIXIT, Prachi; BAJPAI, Manish. **Kalman filter based short term prediction model for COVID-19 spread**. Applied Intelligence, 51(5), 2714-2726, 2021.

- SMITH, Tristan; JALKANEN, J.P.; ANDERSON, B.A.; CORBETT, J.J.; FABER, J.; HANAYAMA, S.; **Third IMO GHG Study**. Londres, UK, 2014.
- SNAPIR, Boris; WAINE, Toby W.; BIERMANN, Lauren. **Maritime vessel classification to monitor fisheries with SAR: Demonstration in the North Sea**. Remote sensing, v. 11, n. 3, p. 353, 2019.
- SOARES, Jones Alexandre. **Cabotagem como complemento na matriz de transportes: uma análise dos desafios para seu crescimento no sistema de transportes brasileiro no século 21**. Lisboa: ISCTE-IUL, 2019.
- SOARES, Flávio José; DE BRITTO, Walter. **Caracterização Dinâmica de Embarcações Regionais do Amazonas**. UNOPAR: Científica Ciências Exatas e Tecnológicas, v. 13, n. 1, 2014.
- SOBENA, Sociedade Brasileira de Engenharia Naval. **Recomendações para o setor de transporte fluvial de passageiros na Amazônia**. Relatório técnico, Rio de Janeiro: Sociedade Brasileira de Engenharia Naval, 2010.
- SOURTZI, Vasiliki Marianna. **Forecasting the fuel consumption on passenger vessels**. 2019. Tese de Doutorado. University of Piraeus (Greece).
- SONER, Omer; AKYUZ, Emre; CELIK, Metin. **Statistical modelling of ship operational performance monitoring problem**. Journal of Marine Science and Technology, v. 24, n. 2, p. 543-552, 2019.
- ŠTEPEC, Dejan; MARTINCIC, Tomaz; KLEIN, Fabrice; VLADUSIC, Daniel; COSTA, João Pita. **Machine learning based system for vessel turnaround time prediction**. 21st IEEE International Conference on Mobile Data Management (MDM). IEEE, 2020. p. 258-263.
- STOPFORD, Martin. **Maritime economics**. 2nd ed. London: Routledge, 1999.
- STOPFORD, Martin. **The Organization of the Shipping Market**. Maritime Economics, 3rd ed., pp. 47-90; Routledge: London, UK, 2009.
- STOROPOLI, José Eduardo. **O uso do Knowledge Discovery in Database (KDD) de informações patentárias sobre ensino a distância: contribuições para instituições de ensino superior**. 2016.
- TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Data mining cluster**

analysis: basic concepts and algorithms. Introduction to data mining, v. 487, p. 533, 2013.

TARADEH, Mohammad; MAFARJA, Majdi; HEIDARI, Ali Asghar; FARIS, Hossam; ALJARAH, Ibrahim; MIRJALILI, Seyedalli; FUJITA, Hamido. **An evolutionary gravitational search-based feature selection.** Information Sciences, 497, 219-239, 2019.

TAYEFI, Maryam; ESMAEILI, Habibollah; KARIMIAN, Maryam; ZADEH, Alireza; EBRAHIMI, Mahmoud; SAFARIAN, Mohammad; NEMATY, Mohsen; PARIZADEH, Seyed; FERNS, Gordon; MOBARHAN, Majid. **The application of a decision tree to establish the parameters associated with hypertension.** Computer methods and programs in biomedicine, v. 139, p. 83-91, 2017.

TILLIG, Fabian; RINGSBERG, J.; MAO, W.; RAMNE, B. **A generic energy systems model for efficient ship design and operation.** Anais da Instituição de Engenheiros Mecânicos, Parte M: Journal of Engineering for the Maritime Environment, 231(2), 649–666, 2017.

TILLIG, Fabian; RINGSBERG, J.; MAO, W.; RAMNE, B. **Analysis of uncertainties in the prediction of ships' fuel consumption – from early design to operation conditions.** 13(sup1), pp.13-24, 2018.

TOQUÉ, Florian; KHOUADJIA, Mostepha; COME, Etienne; TREPANIER, Martin; OUKHELLOU, Latifa. **Short & long term forecasting of multimodal transport passenger flows with machine learning methods.** 20th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2017. p. 560-566.

TUKEY, John W. **Exploratory data analysis**, Vol. 2. Reading, MA, 1977.

TSAI, Chun-Wei; LAI, Chin-Feng; CHIANG, Ming-Chao; YANG, Laurence. **Data mining for internet of things: a survey.** IEEE Commun Surv Tutor 16(1):77–97, 2014.

TSITSILONIS, Konstantinos-Marios; THEOTOKATOS, Gerasimos. **A novel systematic methodology for ship propulsion engines energy management.** J. Cleaner Prod, 2018.

TZANNATOS, Ernestos; STOURNARAS, Lefteris. **EEDI analysis of Ro-Pax and passenger ships in Greece.** Maritime Policy Managment. 2014, 42, 305–316.

- UYANIK, Tayfun; KARATUĞ, Çağlar; ARSLANOĞLU, Yasin. **Machine learning approach to ship fuel consumption: A case of container vessel**. Transportation Research Part D: Transport and Environment, v. 84, p. 102389, 2020.
- UYANIK, Tayfun; ARSLANOGLU, Yasin; KALENDERLI, Ozcan. **Ship fuel consumption prediction with machine learning**. Proceedings of the 4th International Mediterranean Science and Engineering Congress, Antalya, Turkey. p. 25-27, 2019.
- VENTIKOS, N. P.; PAPANIKOLAOU, A. D.; LOUZIS, K.; KOIMTZOGLU, A. **Statistical analysis and critical review of navigational accidents in adverse weather conditions**. Ocean Engineering, 163, 502-517, 2018.
- WANG, Shengzheng; JI, Baoxian; ZHAO, Jiansen; LIU, Wei; XU, Tie. **Predicting ship fuel consumption based on LASSO regression**. Transp. Res, 2018.
- WICKRAMANAYAKE, Sandareka; BANDARA, HMN Dilum. **Fuel consumption prediction of fleet vehicles using machine learning: A comparative study**. 2016 Moratuwa Engineering Research Conference (MERCOn). IEEE, 2016. p. 90-95.
- WIJNOLST, Niko; WERGELAND, Tor. **Shipping Innovation**. Delft: Delft University Press, 2011.
- WILSON, Robert; KEIL, Frank. **The MIT encyclopedia of the cognitive sciences**. Choice Rev. 37-1902–37-1902, 2001.
- WITTEN, Ian; FRANK, Eibe; HALL, Mark; PAL, Christopher. **Data Mining: Practical Machine Learning Tools and Techniques**. 2011.
- XU, Honghui; DENG, Yong. **Dependent evidence combination based on shearman coefficient and pearson coefficient**. IEEE Access, v. 6, p. 11634-11640, 2017.
- YADAV, Sanjay; SHUKLA, Sanyam. **Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification**. 2016 IEEE 6th International conference on advanced computing (IACC). IEEE, 2016. p. 78-83.
- YAMASAKI, Shintaro; YAJI, Kentaro; FUJITA, Kikuo. **Knowledge discovery in databases for determining formulation in topology optimization**. Structural and Multidisciplinary Optimization, v. 59, n. 2, p. 595-611, 2019.
- YAO, Zhishuang; NG, Szu Hui; LEE, Loo Hay. **A study on bunker fuel management**

for the shipping liner services. 2012.

YIN, Wenjie. **Machine Learning for Adaptive Cruise Control Target Selection.** 2019.

YUAN, Jun; NIAN, Victor. **Ship energy consumption prediction with Gaussian process metamodel.** Energy Procedia, v. 152, p. 655-660, 2018.

ZENG, Xueqiang; LUO, Gang. **Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection.** Health information science and systems, v. 5, n. 1, p. 1-21, 2017.

ZHANG, Yanfei; WEN, Yiyang; HUANG, ZhenPing. **A Ship Main Engine Power Predictive Model based on Big Data Analytics and Machine Learning.** The 29th International Ocean and Polar Engineering Conference. International Society of Offshore and Polar Engineers, 2019.

APÊNDICE A - REGISTROS PARA VALIDAÇÃO

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	LANCHA	ALUMÍNIO	17	3,45	1,2	1,9	215	22	1	35	7	0	100	0	3	0,45	70
2	FERRY BOAT	AÇO NAVAL	32	8	5	6	450	19	1	528	98	2	99	670	8	36	3000
3	LANCHA	AÇO NAVAL	25,6	6,3	1,8	2,2	315	18	1	21	6	2	90	0	2	24	2500
4	LANCHA	AÇO NAVAL	24,89	6,3	1,4	2,3	550	18	1	21	6	2	90	0	2	24	2500
5	PASSAGEIRO/CARGA GERAL	MADEIRA	19	5,3	1,8	2,8	200	10	2	109	21	2	65	25	3	7	300
6	PASSAGEIRO/CARGA GERAL	MADEIRA	24	6	2,15	2,8	315	15	3	109	21	2	88	34	3	8	400
7	PASSAGEIRO/CARGA GERAL	MADEIRA	32	8	2,4	3	400	20	3	109	21	1	126	288	7	40	2300
8	PASSAGEIRO/CARGA GERAL	MADEIRA	28	7,4	1,5	1,75	367	14	2	109	21	2	130	90	5	144	5000
9	PASSAGEIRO/CARGA GERAL	MADEIRA	17,6	4,2	1,8	1,8	612	15	1	109	21	1	60	19	2	6	180
10	LANCHA	ALUMÍNIO	12	2,4	1	1,8	481	20	1	35	7	0	35	0	1	2	87
11	LANCHA	MADEIRA	22	5,1	1,5	1,9	375	20	1	28	5	2	81	55	4	6	300
12	PASSAGEIRO/CARGA GERAL	MADEIRA	13	4	1,5	2,5	280	15	2	109	21	1	98	138	4	6	350
13	LANCHA	ALUMÍNIO	20	3,75	0,6	1,25	448	20	2	35	7	0	105	0	3	5	350
14	PASSAGEIRO/CARGA GERAL	MADEIRA	19	4,5	1	1,8	240	14	1	109	21	2	60	12	3	12	220
15	PASSAGEIRO/CARGA GERAL	MADEIRA	27	6,8	1,2	2,1	315	15	2	109	21	2	140	120	5	14	600
16	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	30	5	1,3	1,8	470	12	1	330	64	1	120	80	4	14	300
17	FERRY BOAT	AÇO NAVAL	42	10,3	1615	1,9	450	12	2	528	98	2	451	200	6	10	1000
18	PASSAGEIRO/CARGA GERAL	MADEIRA	16,7	3,8	1	1,4	110	8	1	109	21	1	10	17	3	10	200
19	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	39,5	8,16	1,8	2,5	600	15	1	330	64	2	365	140	8	30	2700

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
20	PASSAGEIRO/CARGA GERAL	MADEIRA	25	6,6	1,8	2,7	367	13	1	109	21	1	100	143	7	30	2000
21	FERRY BOAT	AÇO NAVAL	39	10,2	1,57	1,85	570	15	3	528	98	2	340	300	8	36	2500
22	FERRY BOAT	AÇO NAVAL	30	7	1,2	1,6	600	14	2	528	98	2	181	80	9	42	1400
23	FERRY BOAT	AÇO NAVAL	41	10,3	1,6	1,6	400	18	2	528	98	2	249	350	8	42	2500
24	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	60	15	2,6	3,2	829	13	2	330	64	2	637	900	9	130	9000
25	FERRY BOAT	AÇO NAVAL	45,6	10	3	3,6	824	12	1	528	98	2	400	700	7	40	3000
26	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	46	9,6	2,8	3	590	12	2	330	64	2	430	380	8	120	9000
27	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	38	9	2,4	2,9	315	18	2	330	64	2	221	200	8	40	3500
28	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	41	8,2	2,2	3,2	820	12	1	330	64	2	265	400	8	144	11000
29	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	40,4	8,5	1,8	2,65	711	16	2	330	64	1	442	294	6	15	1500
30	PASSAGEIRO/CARGA GERAL	MADEIRA	25	6	2,5	2,8	350	12	1	109	21	2	80	40	3	20	400
31	LANCHA	ALUMÍNIO	24	7	1,4	1,8	400	20	1	35	7	1	72	0	5	3	300
32	PASSAGEIRO/CARGA GERAL	MADEIRA	26	7	1,6	2	290	18	1	109	21	1	90	120	4	18	600
33	FERRY BOAT	MADEIRA	38	4,3	1,1	1,5	190	10	1	528	98	1	30	30	5	40	2000
34	FERRY BOAT	AÇO NAVAL	46	11	2,2	3,7	450	8	2	528	98	1	95	300	7	45	4000
35	PASSAGEIRO/CARGA GERAL	MADEIRA	17	3,5	1,1	2	160	15	1	109	21	1	25	14	2	8	250
36	LANCHA	ALUMÍNIO	19	3,1	1	1,6	475	38	1	35	7	1	65	0	3	10	700
37	LANCHA	ALUMÍNIO	19	4,3	1	1,5	390	23	1	35	7	0	60	0	2	2	150
38	PASSAGEIRO/CARGA GERAL	MADEIRA	21	5,25	1,8	2,2	250	12	1	109	21	1	60	40	3	12	200
39	PASSAGEIRO/CARGA GERAL	MADEIRA	24	6,15	1	1,6	210	13	1	109	21	1	71	56	5	20	800
40	PASSAGEIRO/CARGA GERAL	MADEIRA	23,2	5	1	1,8	180	15	1	109	21	0	61	30	4	3	80

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
41	LANCHA	ALUMÍNIO	17,2	2,4	0,8	1	200	20	1	35	7	0	60	0	2	1,3	80
42	PASSAGEIRO/CARGA GERAL	MADEIRA	16,45	3,1	0,9	1,8	369	12	1	109	21	0	20	14	3	8	100
43	PASSAGEIRO/CARGA GERAL	MADEIRA	26	7,6	1	1,2	215	19	1	109	21	2	76	98	5	9	250
44	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	37	8	2,22	2,6	600	13	1	330	64	2	327	406	7	10	700
45	PASSAGEIRO/CARGA GERAL	MADEIRA	31	6,8	2,1	2,7	420	8	1	109	21	2	132	150	4	5	200
46	PASSAGEIRO/CARGA GERAL	MADEIRA	28,99	7,25	1,3	2,3	420	10	2	109	21	2	130	100	7	5	400
47	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	40	6	1,8	2,4	800	9	2	330	64	2	420	200	8	10	800
48	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	38,5	8	1,7	2,55	954	15	2	330	64	1	300	199	6	4	400
49	PASSAGEIRO/CARGA GERAL	MADEIRA	34	7	2,5	2,9	600	10	1	109	21	2	154	150	7	33	2000
50	PASSAGEIRO/CARGA GERAL	MADEIRA	30	8	2	2,3	477	10	2	109	21	2	159	240	6	34	1800
51	LANCHA	ALUMÍNIO	19,9	6	1,2	1,7	426	30	1	35	7	0	48	0	2	1,3	125
52	LANCHA	AÇO NAVAL	25	3,5	0,8	1,5	600	19	2	21	6	1	114	0	2	1,4	140
53	PASSAGEIRO/CARGA GERAL	MADEIRA	20,5	5,4	1,85	1,85	200	18	1	109	21	2	68	36	3	5	150
54	FERRY BOAT	FIBRA	7,2	1,9	0,7	1	200	40	1	528	98	1	16	8	1	0,3	50
55	FERRY BOAT	FIBRA	7,82	1,7	0,6	1	200	30	1	528	98	0	22	12	1	0,3	40
56	FERRY BOAT	FIBRA	8,6	2,1	0,5	0,65	200	30	1	528	98	0	22	12	1	0,3	50
57	FERRY BOAT	FIBRA	8	2,1	0,4	1	200	40	1	528	98	1	18	8	1	0,3	45
58	PASSAGEIRO	MADEIRA	14	13,1	1,2	1,35	114	18	1	29	7	0	57	0	2	0,45	40
59	PASSAGEIRO/CARGA GERAL	MADEIRA	20	4	1,3	1,5	114	10	1	109	21	1	36	36	4	3	50
60	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	51,2	10	2,75	3,2	915	15	1	330	64	2	836	440	8	44	5000
61	PASSAGEIRO/CARGA GERAL	MADEIRA	25	5	1,5	2	114	12	1	109	21	1	60	36	3	2	20

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
62	PASSAGEIRO/CARGA GERAL	MADEIRA	16	3,2	1,2	1,8	114	12	2	109	21	1	30	15	2	5	60
63	PASSAGEIRO/CARGA GERAL	MADEIRA	20	5,5	2	2,5	270	16	1	109	21	1	60	20	3	5	130
64	LANCHA	ALUMÍNIO	19,9	3,3	0,52	1	580	25	2	15	4	0	100	0	3	0,8	180
65	PASSAGEIRO/CARGA GERAL	MADEIRA	26,5	6,9	1,4	2,09	360	13	1	95	21	2	124	75	5	13	800
66	PASSAGEIRO/CARGA GERAL	MADEIRA	27	6,15	1,5	2,1	412	20	1	115	20	2	98	90	5	18	1100
67	PASSAGEIRO/CARGA GERAL	MADEIRA	17,6	4,2	1,8	2	612	15	1	22	4	1	60	19	2	6	180
68	FERRY BOAT	AÇO NAVAL	42,3	10	1,93	2,15	830	11	2	1100	210	2	95	840	5	126	8000
69	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	41,9	10	2,4	3,1	829	14	1	260	48	2	520	232	8	42	7500
70	PASSAGEIRO/CARGA GERAL	MADEIRA	25	6,3	1,8	2,9	190	10	1	50	8	1	80	40	3	5	350
71	PASSAGEIRO/CARGA GERAL	MADEIRA	21	4,4	1,44	1,65	240	13	2	25	4	1	45	25	3	4	300
72	FERRY BOAT	AÇO NAVAL	36	7,2	1,8	1,7	370	12	1	235	47	2	185	200	7	84	5000
73	PASSAGEIRO/CARGA GERAL	MADEIRA	39	6,8	1,8	2,8	210	13	2	90	20	2	92	70	6	144	5800
74	PASSAGEIRO/CARGA GERAL	MADEIRA	20	5,2	1,7	2,3	280	14	1	100	24	1	100	88	4	12	225
75	PASSAGEIRO/CARGA GERAL	MADEIRA	20	4	1,3	1,5	114	10	1	58	12	1	36	36	4	3	50
76	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	39	9,5	1,8	3	618	17	3	300	60	2	650	263	4	1	750
77	PASSAGEIRO	MADEIRA	18	3,7	1,2	1,6	270	12	1	29	7	1	100	0	3	0,8	30
78	LANCHA	ALUMÍNIO	16	4	0,9	1,2	650	15	1	35	7	1	30	0	2	1	50
79	LANCHA	ALUMÍNIO	8	2	1	1,4	150	18	1	30	8	1	20	0	2	2	80
80	LANCHA	ALUMÍNIO	21,1	3,1	1	1,2	552	15	2	28	7	0	64	0	2	3	150
81	LANCHA	ALUMÍNIO	39	5,3	1,6	2,2	1600	25	2	480	85	2	171	400	4	2	600
82	LANCHA	ALUMÍNIO	20	3	1,2	1,4	550	25	1	15	4	1	75	0	2	2	150

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
83	LANCHA	ALUMÍNIO	19	4,8	1,6	2,7	600	35	1	35	7	1	48	0	2	1	155
84	FERRY BOAT	AÇO NAVAL	40,46	9,3	1,6	2,15	500	13	3	116	20	2	180	100	8	72	6000
85	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	45	10	3	3,5	829	10	3	446	90	2	400	380	7	50	3500
86	PASSAGEIRO/CARGA GERAL	MADEIRA	24	6	1,9	2,4	190	16	1	72	12	1	80	56	3	2	200
87	FERRY BOAT	AÇO NAVAL	46	11	2,2	3,7	450	12	2	380	70	1	95	300	7	45	4000
88	PASSAGEIRO/CARGA GERAL	MADEIRA	24,7	6,3	1,92	2,4	313	12	1	190	34	0	100	162	5	4	350
89	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	46	9,6	2,8	3	590	12	2	474	90	2	430	380	8	120	9000
90	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	41	8,6	1,2	3	915	16	1	300	50	2	322	248	4	5	500
91	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	46,4	10,4	2,51	2,97	440	12	3	392	75	2	700	320	13	43	10000
92	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	37	8,5	2,5	2,9	815	12	2	440	90	2	340	340	12	56	4600
93	LANCHA	ALUMÍNIO	17	3,45	1,2	1,9	215	22	1	28	6	0	100	0	3	0,8	70
94	LANCHA	ALUMÍNIO	22	4,3	1,15	1,75	480	25	1	35	7	0	55	0	2	2	150
95	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	38	8	1,7	2,8	375	17	2	200	35	2	350	180	4	5	350
96	PASSAGEIRO/CARGA GERAL	MADEIRA	23	5,3	1,8	2,4	180	14	1	79	16	1	100	70	6	15	450
97	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	60	12	1,4	3	600	14	2	30	4	2	870	0	6	1,5	600
98	PASSAGEIRO/CARGA GERAL	MADEIRA	32	8	2,4	3	400	20	3	345	70	1	126	288	7	40	2300
99	PASSAGEIRO/CARGA GERAL	MADEIRA	32	8	2,4	2,8	480	15	1	205	40	1	60	180	7	72	1200
100	PASSAGEIRO	MADEIRA	16	4	0,8	1,1	180	14	1	33	8	1	65	0	3	0,8	40
101	PASSAGEIRO/CARGA GERAL	MADEIRA	31	6	1,2	1,5	420	15	1	60	15	2	135	60	5	5	250
102	PASSAGEIRO/CARGA GERAL	MADEIRA	15,8	4,3	1,2	1,2	200	15	1	20	4	1	30	17	2	10	350
103	PASSAGEIRO	MADEIRA	18,78	4,8	0,66	1	240	13	2	20	5	2	115	0	3	0,7	70

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
104	PASSAGEIRO/CARGA GERAL	MADEIRA	29,6	7,8	2,2	2,7	552	12	1	190	35	2	200	160	7	140	7000
105	PASSAGEIRO/CARGA GERAL	MADEIRA	22	5,18	1,2	1,7	175	12	1	55	10	1	90	42	6	58	2250
106	PASSAGEIRO	MADEIRA	20	4,5	0,99	1	114	12	1	29	7	1	30	15	3	10	400
107	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	38	7,3	1,65	2	412	20	1	169	30	2	97	150	5	9	400
108	PASSAGEIRO	MADEIRA	16,9	3,6	0,75	1	125	12	1	23	6	1	66	0	3	0,8	20
109	PASSAGEIRO/CARGA GERAL	MADEIRA	18	5	1,2	2	160	11	1	15	4	1	38	15	2	6	60
110	LANCHA	ALUMÍNIO	22	5,5	2,5	3,3	240	23	1	90	20	1	93	70	8	2,5	90
111	PASSAGEIRO/CARGA GERAL	MADEIRA	26	6,5	2	3	377	17	1	120	22	1	90	90	4	24	600
112	FERRY BOAT	AÇO NAVAL	41	9	1,8	2	420	12	2	360	72	2	315	298	4	5	500
113	PASSAGEIRO/CARGA GERAL	MADEIRA	20	4	1,2	1,4	280	15	1	20	4	1	45	20	3	8	160
114	PASSAGEIRO/CARGA GERAL	MADEIRA	26	6	1,5	2,8	373	10	2	700	145	2	105	615	6	10	800
115	PASSAGEIRO/CARGA GERAL	MADEIRA	16	4	1,4	1,9	125	12	1	146	30	1	40	120	2	12	800
116	FERRY BOAT	AÇO NAVAL	76	16	2,8	2,8	932	15	2	1410	250	2	1120	1240	5	3	400
117	FERRY BOAT	AÇO NAVAL	32	8,5	1,5	2,2	300	12	2	150	36	3	170	130	8	30	1600
118	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	35	8,4	2,1	2,8	600	17	2	184	35	2	200	150	12	28	2000
119	PASSAGEIRO/CARGA GERAL	MADEIRA	22,15	6,3	1,5	2,3	306	12	1	98	20	2	95	80	5	18	550
120	PASSAGEIRO/CARGA GERAL	MADEIRA	26	6,1	1,64	2	450	12	1	60	10	1	132	60	4	8	250
121	PASSAGEIRO/CARGA GERAL	AÇO NAVAL	27	6,5	1,6	2	170	16	1	110	25	1	140	83	6	22	600
122	LANCHA	ALUMÍNIO	12	3	0,9	1,3	350	20	1	25	5	1	40	0	2	0,4	30
123	PASSAGEIRO/CARGA GERAL	MADEIRA	30	6,4	1,6	2	600	14	2	87	18	2	161	60	6	15	750
124	PASSAGEIRO/CARGA GERAL	MADEIRA	22	5,8	1,9	2,4	180	12	2	60	15	1	50	50	4	4	250

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
125	PASSAGEIRO/CARGA GERAL	MADEIRA	27	6	2	2,4	315	10	2	96	14	1	107	80	5	2	160