

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
INSTITUTO DE COMPUTAÇÃO  
INSTITUTO TÉRCIO PACITTI DE APLICAÇÕES E PESQUISAS COMPUTACIONAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

**VALQUIRE DA SILVA DE JESUS**

AUTONOMIA DE MEIOS OPERATIVOS PARA APOIO À DECISÃO EM SISTEMAS  
DE COMANDO E CONTROLE: UMA ABORDAGEM PREDITIVA E CONCEITUAL  
UTILIZANDO APRENDIZADO DE MÁQUINA

RIO DE JANEIRO

2025

Valquire da Silva de Jesus

AUTONOMIA DE MEIOS OPERATIVOS PARA APOIO À DECISÃO EM SISTEMAS  
DE COMANDO E CONTROLE: UMA ABORDAGEM PREDITIVA E CONCEITUAL  
UTILIZANDO APRENDIZADO DE MÁQUINA

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática, Instituto de Computação e Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Informática.

Orientadora: Prof.<sup>a</sup> Giseli Rabello Lopes, D.Sc.

Coorientadora: Prof.<sup>a</sup> Kelli de Faria Cordeiro, D.Sc.

RIO DE JANEIRO

2025

## CIP - Catalogação na Publicação

J58a Jesus, Valquire da Silva de  
Autonomia de Meios Operativos para Apoio à  
Decisão em Sistemas de Comando e Controle: uma  
Abordagem Preditiva e Conceitual Utilizando  
Aprendizado de Máquina / Valquire da Silva de  
Jesus. -- Rio de Janeiro, 2025.  
133 f.

Orientadora: Giseli Rabello Lopes.  
Coorientadora: Kelli de Faria Cordeiro.  
Dissertação (mestrado) - Universidade Federal do  
Rio de Janeiro, Instituto de Computação e Instituto  
Tércio Pacitti de Aplicações e Pesquisas  
Computacionais, Programa de Pós-Graduação em  
informática, 2025.

1. ontologia bem fundamentada. 2. aprendizado de  
máquina. 3. logística militar. 4. comando e controle.  
5. autonomia. I. Lopes, Giseli Rabello, orient. II.  
Cordeiro, Kelli de Faria, coorient. III. Título.

VALQUIRE DA SILVA DE JESUS

AUTONOMIA DE MEIOS OPERATIVOS PARA APOIO À DECISÃO EM SISTEMAS  
DE COMANDO E CONTROLE: UMA ABORDAGEM PREDITIVA E CONCEITUAL  
UTILIZANDO APRENDIZADO DE MÁQUINA

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática, Instituto de Computação e Instituto Tércio Pacitti de Aplicações e Pesquisas Computacionais, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Informática.

Aprovado em 29 de maio de 2025:

---

Prof.<sup>a</sup> Giseli Rabello Lopes, D.Sc.,  
PPGI-UFRJ (Presidente)

---

Prof.<sup>a</sup> Kelli de Faria Cordeiro, D.Sc.,  
Subchefia de Comando e Controle do Ministério da Defesa

---

Prof. Sérgio Manuel Serra da Cruz, D.Sc.,  
PPGI-UFRJ

---

Prof.<sup>a</sup> Maria Cláudia Reis Cavalcanti, D.Sc.,  
PGSC-IME

---

Prof. João Luiz Rebelo Moreira, Ph.D.,  
Universidade de Twente

---

Cel Eng Mil Junier Caminha Amorim, D.Sc.,  
Subchefia de Comando e Controle do Ministério da Defesa

*Dedico à minha esposa, Vanessa Karam, que caminhou comigo em cada passo desta jornada, e à minha filha, Maria Luísa, que me ensina, mesmo sem saber, a importância de seguir em frente com leveza e curiosidade.*

## AGRADECIMENTOS

A realização deste trabalho foi possível graças ao apoio e incentivo de diversas pessoas e instituições, às quais expresso minha mais profunda gratidão.

À minha orientadora, Prof.<sup>a</sup> Giseli Rabello Lopes, por sua orientação clara, paciência e incentivo ao longo de toda a pesquisa. Seu conhecimento e dedicação foram essenciais para a conclusão deste trabalho. À minha coorientadora, Prof.<sup>a</sup> Kelli de Faria Cordeiro, pelo suporte técnico, pelas valiosas contribuições e pelo direcionamento assertivo que enriqueceram este estudo.

Aos professores e pesquisadores do Programa de Pós-Graduação em Informática da Universidade Federal do Rio de Janeiro (PPGI/UFRJ), pelo ensino de excelência e pelo ambiente acadêmico inspirador.

Aos professores membros da banca examinadora, Sérgio Manuel Serra da Cruz, Maria Cláudia Reis Cavalcanti, João Luiz Rebelo Moreira e Junier Caminha Amorim, por aceitarem avaliar este trabalho e por suas contribuições para o aprimoramento da pesquisa.

Aos meus colegas de pesquisa e amigos do grupo GRECO-PPGI/UFRJ, que compartilharam conhecimentos, trocaram ideias e proporcionaram momentos de aprendizado e colaboração inestimáveis.

À Comandante Luciene Carvalho Corrêa de Souza, pelo valioso apoio à pesquisa e pela mediação junto à Marinha do Brasil no fornecimento dos dados necessários. Aos profissionais e especialistas que colaboraram com informações valiosas para a modelagem conceitual e para a realização dos experimentos.

Aos órgãos FINEP/DCT/FAPEB, expresso minha sincera gratidão pelo suporte concedido ao projeto S2C2 (ref.: 2904/20, convênio 01.20.0272.00). O financiamento proporcionado por esse convênio possibilitou a participação em atividades colaborativas com pesquisadores e instituições, cuja contribuição foi fundamental para o desenvolvimento e aprimoramento desta pesquisa.

Por fim, aos meus familiares, pelo amor, paciência e incentivo incondicional. Aos demais entes queridos, por compreenderem minha ausência em muitos momentos e sempre acreditarem em mim, e a todos que, de alguma forma, contribuíram para que esta dissertação se tornasse realidade, meu mais sincero agradecimento.

## RESUMO

Em cenários militares complexos, como aqueles enfrentados pela Marinha do Brasil (MB) em operações navais, o planejamento logístico exige exatidão, especialmente quanto à autonomia de meios operativos, entendida como a capacidade de operar sem reabastecimento externo. Essa autonomia depende de fatores como consumo de suprimentos, número de militares embarcados e condições de emprego das plataformas navais, operando em contextos apoiados por sistemas de Comando e Controle (C<sup>2</sup>). Embora técnicas de Aprendizado de Máquina (AM) sejam usadas para antecipar demandas logísticas, a heterogeneidade semântica dos dados e a falta de contextualização reduzem a eficácia de abordagens puramente quantitativas. Para enfrentar esses desafios, este trabalho propõe a abordagem *ATOp* (*Autonomia de Meios Operativos*), que integra *Ontology-Driven Conceptual Modeling* (ODCM) e AM para aprimorar a previsão de autonomia. Fundamentada na *Unified Foundational Ontology* (UFO), a *ATOp* explicita o conhecimento logístico do domínio e viabiliza a segmentação semântica dos dados por meio da metacategoria ontológica *Situation*, que representa contextos relevantes ao comportamento operacional. Essa segmentação sustenta um *pipeline* de regressão com comitê de modelos (*ensemble*), que coordena múltiplos algoritmos especializados. A pesquisa incluiu o desenvolvimento da *ATOp-NavalOntology*, uma ontologia bem fundamentada, a engenharia de dados a partir de fontes heterogêneas da MB e a realização de experimentos com otimização *bayesiana* de hiperparâmetros e validação cruzada (*k-fold*), voltados à construção do *ATOp-PredictiveModel*. A avaliação utilizou métricas ( $R^2$  e RMSE) e testes estatísticos (*Friedman* e *Durbin-Conover*) para garantir comparações confiáveis. Os resultados mostram que a combinação entre ontologias bem fundamentadas e AM estrutura os dados de forma a evidenciar padrões relevantes, melhora a capacidade preditiva e supera modelos isolados, como o *Gradient Boosting*, quando aplicados diretamente a dados brutos. O *ATOp-PredictiveModel* obteve o melhor poder de previsão e menor custo computacional, apresentando diferenças estatisticamente significativas. Como contribuição, a *ATOp* oferece um modelo replicável que integra conhecimento semântico ao apoio à decisão em domínios operacionais sustentados por sistemas de C<sup>2</sup>.

**Palavras-chave:** ontologia bem fundamentada; aprendizado de máquina; UFO; OntoUML; comando e controle; autonomia;

## ABSTRACT

In complex military scenarios, such as those faced by the Brazilian Navy (MB) in naval operations, logistical planning demands rigor, especially regarding the autonomy of operational elements, understood as the ability to operate without external resupply. This autonomy depends on factors such as supply consumption, the number of embarked personnel, and the operational conditions of naval platforms, acting in contexts supported by Command and Control (C<sup>2</sup>) systems. Although Machine Learning (ML) techniques are used to anticipate logistical demands, the semantic data heterogeneity and the lack of contextual domain understanding limit the effectiveness of purely quantitative approaches. To address these challenges, this work proposes the *ATOp* (*Autonomia de Meios Operativos*) approach, which integrates *Ontology-Driven Conceptual Modeling* (ODCM) and ML to enhance autonomy prediction. Grounded in the *Unified Foundational Ontology* (UFO), *ATOp* makes domain knowledge explicit and enables semantic data segmentation through the ontological meta-category *Situation*, which represents contexts relevant to operational behavior. This segmentation supports a regression pipeline with an ensemble of models, coordinating multiple specialized algorithms. The research included the development of the *ATOp-NavalOntology*, a well-founded ontology, data engineering from heterogeneous MB sources, and experiments involving Bayesian hyperparameter optimization and k-fold cross-validation, aimed at constructing the *ATOp-PredictiveModel*. The evaluation used metrics (R<sup>2</sup> and RMSE) and statistical tests (Friedman and Durbin-Conover) to ensure reliable comparisons. The results show that combining well-founded ontologies and ML structures the data in a way that reveals relevant patterns, improves the quality of predictions, and outperforms individual models such as *Gradient Boosting* when applied directly to raw data. The *ATOp-PredictiveModel* achieved the best performance, with higher accuracy and lower computational cost, presenting statistically significant differences. As a contribution, *ATOp* offers a replicable model that integrates semantic knowledge into decision support in operational domains sustained by C<sup>2</sup> systems.

**Keywords:** well-founded ontology; machine learning; UFO; OntoUML; command and control; autonomy of operational elements.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Fluxograma do método empregado neste estudo. . . . .	20
Figura 2 – Fluxos de informações e ordens no SISMC <sup>2</sup> . . . . .	30
Figura 3 – Representação esquemática da estrutura do SISMC <sup>2</sup> . . . . .	31
Figura 4 – Tipos de ontologias. . . . .	34
Figura 5 – Fragmento da taxonomia da UFO-A. . . . .	39
Figura 6 – Fragmento da taxonomia da UFO-B. . . . .	43
Figura 7 – Hierarquia clássica de aprendizado. . . . .	48
Figura 8 – Uma árvore de decisão e a representação da divisão do espaço de entrada das variáveis. . . . .	51
Figura 9 – Relacionamentos entre modelos mentais, conceituais e de AM. . . . .	60
Figura 10 – Macroprocessos da abordagem <i>ATOp</i> . . . . .	67
Figura 11 – Fases que estruturam o Macroprocesso 1 da abordagem <i>ATOp</i> . . . . .	69
Figura 12 – Estrutura ontológica baseada em <i>Event</i> e <i>Situation</i> da UFO . . . . .	71
Figura 13 – Representação da regra <i>ATOp-GroupEntityRule</i> . . . . .	73
Figura 14 – Representação do mapeamento realizado na <i>ATOp-MapGroupEntity</i> . . .	81
Figura 15 – Representação do mapeamento com sobreposição de atributos correspondentes. . . . .	82
Figura 16 – Etapas que compõem o Macroprocesso 2 da abordagem <i>ATOp</i> . . . . .	84
Figura 17 – <i>ATOp-NavalOntology</i> baseada em UFO. . . . .	88
Figura 18 – Visualização do consumo Combustível, Lubrificante e Graxa. . . . .	94
Figura 19 – Visualização do consumo de Alimentos. . . . .	95
Figura 20 – Amostra de vinte e cinco registros extraídos do Conjunto Derivado. . . .	97
Figura 21 – Diagrama de Venn da aplicação da <i>ATOp-GroupEntityRule</i> sobre a <i>ATOp-NavalOntology</i> . . . . .	99
Figura 22 – Visualização dos resultados do <i>Treinamento de Modelos Individuais</i> . . . .	105
Figura 23 – Visualização dos resultados do <i>Treinamento de Comitês de Modelos sem Direcionamento Semântico</i> com divisão em dois subconjuntos. . . . .	110
Figura 24 – Visualização dos resultados do <i>Treinamento do ATOp-PredictiveModel</i> . .	114

Figura 25 – Distribuição das métricas $R^2$ e RMSE para os modelos avaliados no conjunto real. . . . .	117
Figura 26 – Tela do protótipo de modelo preditivo da autonomia. . . . .	122

## LISTA DE TABELAS

Tabela 1 – Contribuições da MC para o AM, com base nas fases do CRISP-DM. . . . .	59
Tabela 2 – Benefícios do uso de ontologias de fundamentação nas etapas do ciclo de vida de projetos de Ciência de Dados. . . . .	61
Tabela 3 – Comparação entre os trabalhos relacionados e a abordagem <i>ATOp</i> . . . . .	66
Tabela 4 – Esquema dos dados de abastecimento da plataforma naval militar. . . . .	92
Tabela 5 – Conversões de unidades de fornecimento de materiais. . . . .	93
Tabela 6 – Esquema dos dados operacionais da plataforma naval militar. . . . .	95
Tabela 7 – Correspondência entre Entidades e Atributos. . . . .	100
Tabela 8 – Hiperparâmetros empregados em cada modelo. . . . .	103
Tabela 9 – Resultado do <i>Treinamento de Modelos Individuais</i> . . . . .	104
Tabela 10 – Resultado do <i>Treinamento de Comitês de Modelos sem Direcionamento Semântico</i> com divisão em dois subconjuntos. . . . .	107
Tabela 11 – Resultado do <i>Treinamento de Comitês de Modelos sem Direcionamento Semântico</i> com formações completamente aleatórias. . . . .	109
Tabela 12 – Resultado do <i>Treinamento do ATOp-PredictiveModel</i> . . . . .	112
Tabela 13 – Resultado do teste <i>post-hoc Durbin-Conover</i> para $R^2$ e RMSE . . . . .	118

## LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
BPMN	<i>Business Process Model and Notation</i>
C <sup>2</sup>	Comando e Controle
CLG	Combustível, Lubrificante e Graxa
CRISP-DM	<i>Cross-Industry Standard Process for Data Mining</i>
EB	Exército Brasileiro
FAB	Força Aérea Brasileira
IA	Inteligência Artificial
MB	Marinha do Brasil
MC	Modelagem Conceitual
MD	Ministério da Defesa
ODCM	<i>Ontology-Driven Conceptual Modeling</i>
OMG	<i>Object Management Group</i>
RFC	Relatórios de Fim de Comissão
SINGRA	Sistema de Informações Gerenciais de Abastecimento
SIPLOM	Sistema de Planejamento Operacional Militar
SISMC <sup>2</sup>	Sistema Militar de Comando e Controle
UFO	<i>Unified Foundational Ontology</i>
UML	<i>Unified Modeling Language</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>14</b>
1.1	MOTIVAÇÃO . . . . .	14
1.2	CARACTERIZAÇÃO DO PROBLEMA . . . . .	15
1.3	QUESTÃO DE PESQUISA E HIPÓTESE . . . . .	15
1.4	OBJETIVO . . . . .	16
1.5	PROPOSTA DE SOLUÇÃO . . . . .	16
1.6	JUSTIFICATIVA . . . . .	17
1.7	ORGANIZAÇÃO DO TEXTO . . . . .	18
<b>2</b>	<b>METODOLOGIA</b> . . . . .	<b>19</b>
2.1	OBSERVAÇÃO E DEFINIÇÃO DO PROBLEMA . . . . .	20
2.2	FORMULAÇÃO DA HIPÓTESE . . . . .	21
2.3	DEDUÇÃO DE CONSEQUÊNCIAS . . . . .	21
2.4	PLANEJAMENTO E EXECUÇÃO DOS EXPERIMENTOS . . . . .	22
2.4.1	Estudo de Caso e Aplicação da Abordagem <i>ATOp</i> . . . . .	22
2.4.2	Aspectos Éticos e Confidencialidade . . . . .	22
2.4.3	Artefatos da Solução <i>ATOp</i> . . . . .	23
2.4.4	Estratégia Experimental . . . . .	24
2.4.5	Validação Estatística dos Resultados . . . . .	24
2.4.6	Codificação do Protótipo Funcional . . . . .	24
2.5	ANÁLISE DOS RESULTADOS . . . . .	25
2.6	CONCLUSÃO E VERIFICAÇÃO DA HIPÓTESE . . . . .	25
<b>3</b>	<b>REFERENCIAL TEÓRICO</b> . . . . .	<b>26</b>
3.1	OPERAÇÃO MILITAR, LOGÍSTICA MILITAR E COMANDO E CON- TROLE (C <sup>2</sup> ) . . . . .	26
3.1.1	Operação Militar . . . . .	26
3.1.2	Logística Militar . . . . .	27
3.1.3	Comando e Controle (C <sup>2</sup> ) . . . . .	28
3.2	MODELAGEM CONCEITUAL . . . . .	31

3.2.1	Ontologias . . . . .	33
3.2.2	Systematic Approach for Building Ontologies (SABiO) . . . . .	35
3.2.3	Unified Foundational Ontology (UFO) . . . . .	37
3.2.4	OntoUML . . . . .	43
3.3	APRENDIZADO DE MÁQUINA . . . . .	44
3.3.1	Paradigmas de aprendizado . . . . .	47
3.3.2	Modelos de Regressão . . . . .	49
3.3.2.1	Regressão Linear . . . . .	49
3.3.2.2	<i>K-Nearest Neighbors</i> . . . . .	49
3.3.2.3	Árvore de Decisão . . . . .	50
3.3.2.4	<i>Random Forest</i> . . . . .	51
3.3.2.5	<i>Gradient Boosting</i> . . . . .	52
3.3.2.6	Aprendizado em Comitê ( <i>Ensemble Learning</i> ) . . . . .	52
3.3.3	Avaliação de Modelos de Regressão e Testes Estatísticos . . . . .	53
3.3.3.1	Coefficiente de Determinação ( $R^2$ ) . . . . .	54
3.3.3.2	Erro Quadrático Médio ( <i>Mean Squared Error</i> - MSE) . . . . .	55
3.3.3.3	Raiz do Erro Quadrático Médio ( <i>Root Mean Squared Error</i> - RMSE) . . . . .	55
3.3.3.4	Testes de <i>Friedman</i> e <i>Durbin-Conover</i> . . . . .	55
4	TRABALHOS RELACIONADOS . . . . .	58
4.1	ANÁLISE DE PROCESSOS DE CIÊNCIA DE DADOS . . . . .	58
4.2	SIMILARIDADE SEMÂNTICA E REPRESENTAÇÕES DE DADOS . . . . .	62
4.3	QUALIDADE DOS DADOS . . . . .	63
4.4	CONSIDERAÇÕES FINAIS . . . . .	65
5	ABORDAGEM ATOP . . . . .	67
5.1	MACROPROCESSO 1: CONSTRUIR MODELO PREDITIVO DA AUTONOMIA DE MEIO OPERATIVO ( <i>ATOP-PREDICTIVEMODEL</i> ) . . . . .	68
5.1.1	Desenvolver Ontologia da Autonomia de Meios Operativos ( <i>ATOp-Ontology</i> ) . . . . .	70
5.1.2	Integrar Dados Históricos do Meio Operativo . . . . .	72
5.1.3	Agrupar Entidades da <i>ATOp-Ontology</i> . . . . .	73

5.1.4	Mapear Grupos de Entidades nos Dados Integrados do Meio Operativo ( <i>ATOp-MapGroupEntity</i> ) . . . . .	76
5.1.5	Treinar Modelo de Comitê Preditivo da Autonomia do Meio Operativo ( <i>ATOp-PredictiveModel</i> ) . . . . .	83
5.2	MACROPROCESSO 2: APLICAR <i>ATOP-PREDICTIVEMODEL</i> EM CENÁRIO DE C <sup>2</sup> . . . . .	83
5.2.1	Aplicar <i>ATOp-PredictiveModel</i> . . . . .	84
5.2.2	Apoiar a Tomada de Decisão para o C <sup>2</sup> Conjunto . . . . .	85
6	APLICAÇÃO DA ABORDAGEM <i>ATOP</i> NO CENÁRIO DE PLATAFORMAS NAVAIS MILITARES . . . . .	86
6.1	CONCEPÇÃO . . . . .	86
6.1.1	Desenvolvimento da <i>ATOp-NavalOntology</i> . . . . .	86
6.1.2	Integração Dados Históricos da <i>MilitaryNavalPlatform</i> . . . . .	91
6.1.3	Agrupamento das Entidades da <i>ATOp-NavalOntology</i> . . . . .	96
6.1.4	Mapeamento dos Grupos de Entidades nos Dados Integrados . . . . .	98
6.2	RESULTADOS DO TREINAMENTO . . . . .	101
6.2.1	Treinamento de Modelos Individuais . . . . .	102
6.2.2	Treinamento de Comitês de Modelos sem Direcionamento Semântico . . . . .	106
6.2.3	Treinamento do <i>ATOp-PredictiveModel</i> . . . . .	111
6.2.4	Comparação dos Modelos . . . . .	115
6.3	CASO DE APLICAÇÃO . . . . .	119
6.3.1	Aplicação do <i>ATOp-PredictiveModel</i> . . . . .	120
6.3.2	Apoio à Tomada de Decisão na Operação Conjunta <i>TYPHOON</i> . . . . .	121
6.4	AMEAÇAS À VALIDADE . . . . .	123
7	CONCLUSÃO . . . . .	125
7.1	CONTRIBUIÇÕES . . . . .	126
7.2	CONSIDERAÇÕES E LIMITAÇÕES . . . . .	126
7.3	TRABALHOS FUTUROS . . . . .	127
	REFERÊNCIAS . . . . .	128

# 1 INTRODUÇÃO

A logística militar, historicamente reconhecida como fator decisivo para o sucesso de operações, segue sendo essencial para a prontidão das Forças Armadas, especialmente em cenários hostis ou com infraestrutura limitada. A capacidade de sustentar operações com recursos próprios reduz vulnerabilidades e aumenta a resiliência. Nesse contexto, o avanço da tecnologia e a disponibilidade de dados operacionais ampliam as possibilidades do planejamento logístico, com destaque para o uso de Inteligência Artificial (IA) (SIMON, 2001; BRASIL, 2016; KRESS et al., 2002). A IA tem ganhado relevância no meio militar, sendo considerada estratégica por várias nações. Estudiosos como Heller (2019) destacam seu impacto potencial nas capacidades econômicas e militares, o que tem levado países, inclusive o Brasil, a investir em pesquisa e desenvolvimento na área (HELLER, 2019; ARAYA; KING, 2022; GRAND-CLÉMENT, 2023; KONDRATENKO et al., 2023).

Apesar dos avanços, persistem desafios nas atividades de Comando e Controle (C<sup>2</sup>), que exigem modelos capazes de integrar dados diversos com conhecimento especializado. A combinação entre Modelagem Conceitual (MC) e Aprendizado de Máquina (AM) surge como alternativa promissora, ao estruturar conceitos militares e lidar com grandes volumes de dados de forma coerente (BRASIL, 2020; BORK; ALI; ROELENS, 2023). Diante disso, este estudo propõe uma abordagem para aprimorar a predição da autonomia de plataformas navais da Marinha do Brasil, utilizando ontologias bem fundamentadas para explicitar a semântica dos dados do domínio. A proposta baseia-se na *Unified Foundational Ontology* (UFO) e na linguagem OntoUML, promovendo a integração semântica de dados heterogêneos e fortalecendo a tomada de decisão em sistemas de C<sup>2</sup>.

## 1.1 MOTIVAÇÃO

A crescente complexidade dos cenários operacionais modernos tem evidenciado a necessidade de maior autonomia de meios por parte das Forças Armadas. Em conflitos assimétricos, por exemplo, adversários não convencionais frequentemente adotam táticas voltadas à interrupção das linhas logísticas, o que exige das forças regulares a capacidade de sustentar suas operações de forma independente. De modo semelhante, em missões humanitárias ou de manutenção da paz, essa autonomia é essencial para garantir apoio contínuo às populações

locais sem sobrecarregar recursos externos (BRASIL, 2016; BRASIL, 2015c).

Diante desse cenário, a adoção de soluções tecnológicas capazes de fortalecer o planejamento logístico torna-se uma prioridade. O avanço de métodos analíticos e computacionais permite explorar dados históricos estruturados e não estruturados oriundos de operações militares, viabilizando abordagens mais adequadas para a predição de desempenho logístico, inclusive em contextos com restrições de sigilo (KRESS et al., 2002). Assim, explorar essas tecnologias oferece não apenas ganhos operacionais, mas também maior capacidade de resposta em ambientes incertos e dinâmicos.

## 1.2 CARACTERIZAÇÃO DO PROBLEMA

Os processos de tomada de decisão em logística militar enfrentam desafios significativos, especialmente em atividades de  $C^2$ , onde a predição da autonomia dos meios operativos é essencial para o sucesso das operações. Situações como deslocamentos prolongados, posicionamento de meios em regiões de difícil acesso, sustentação em áreas remotas e reabastecimento sob condições adversas exigem análises rápidas e exatas. Contudo, nesses contextos, os dados disponíveis costumam ser heterogêneos, incompletos ou ambíguos, o que compromete a confiabilidade das análises e dificulta a formulação de estratégias logísticas eficazes.

Modelos preditivos surgem como ferramentas promissoras para apoiar a logística militar, pois permitem antecipar necessidades, otimizar a alocação de recursos e melhorar a tomada de decisão. No entanto, sua efetividade depende da superação de limitações relacionadas à interpretação semântica dos dados, à integração de fontes diversas e à incorporação de conhecimento especializado. A combinação entre MC e AM tem sido apontada como uma abordagem promissora nesse cenário, ao permitir a estruturação explícita do conhecimento do domínio e o uso eficiente de grandes volumes de dados operacionais (BORK; ALI; ROELEN, 2023). Ainda assim, é necessário investigar sua aplicação prática em ambientes militares, onde decisões críticas devem ser tomadas com agilidade, exatidão e respaldo técnico. Essa integração pode representar um avanço significativo na gestão logística, promovendo soluções mais resilientes, adaptáveis e alinhadas às exigências contemporâneas das operações militares.

## 1.3 QUESTÃO DE PESQUISA E HIPÓTESE

Após os esclarecimentos realizados sobre o tema de estudo, chegou-se à seguinte questão de pesquisa:

**Questão de Pesquisa:** Como prever a autonomia dos meios operativos com base em dados históricos e semanticamente heterogêneos?

Para abordar o problema previamente apresentado, formulou-se a seguinte hipótese:

**Hipótese:** A utilização de uma abordagem orientada ao aprimoramento da aplicação de métodos preditivos da autonomia de plataformas navais militares, por meio da explicitação semântica dos dados do domínio com o uso de ontologias bem fundamentadas, eleva a qualidade das previsões quando comparada a modelos preditivos que não empregam tal abordagem.

#### 1.4 OBJETIVO

Como forma de comprovar a hipótese levantada foi estabelecido o seguinte objetivo geral:

- Desenvolver e avaliar uma abordagem para um modelo preditivo baseado em técnicas de AM, que incorpore a explicitação semântica dos dados do domínio por meio do uso de ontologias bem fundamentadas, com o intuito de viabilizar sua aplicação em conjuntos de dados heterogêneos e apoiar a tomada de decisão quanto à autonomia de meios operativos em cenários de  $C^2$ .

Para se alcançar este objetivo foram delineados os seguintes objetivos específicos:

1. Desenvolver uma ontologia bem fundamentada que capture os conceitos relacionados à autonomia de meios operativos no contexto militar;
2. Definir uma trajetória de ciência de dados que utilize um algoritmo preditivo fundamentado na base ontológica desenvolvida, com o objetivo de aprimorar a qualidade das previsões relacionadas à autonomia de meios operativos.

#### 1.5 PROPOSTA DE SOLUÇÃO

Este trabalho propõe a abordagem denominada *Autonomia de Meios Operativos (ATOp)*, que visa aprimorar a previsão da autonomia de plataformas navais militares. A *ATOp* converge técnicas de MC com AM, utilizando ontologias bem fundamentadas para estruturar semanticamente dados operacionais heterogêneos.

A abordagem se fundamenta na *Unified Foundational Ontology (UFO)* e utiliza a linguagem OntoUML para representar os conceitos do domínio logístico de forma adequada. Essa

representação semântica viabiliza o agrupamento coerente de entidades em subconjuntos de dados, por meio da metacategoria *Situation*, com base em situações típicas extraídas do domínio. Esses subconjuntos são, então, utilizados para alimentar um *pipeline* de regressão baseado em comitês de modelos (*ensemble*), otimizados por meta-aprendizagem.

Ao explicitar o conhecimento do domínio antes da etapa de modelagem, a *ATOp* busca mitigar os efeitos da heterogeneidade semântica e melhorar a capacidade preditiva dos modelos. Com isso, a abordagem contribui para decisões logísticas mais exatas e resilientes em ambientes operacionais mediados por sistemas de C<sup>2</sup>.

## 1.6 JUSTIFICATIVA

A obtenção de informações exatas sobre a autonomia dos meios empregados é um fator determinante para o êxito das operações militares. Essa necessidade justifica o investimento em métodos computacionais que apoiem a tomada de decisão logística, especialmente em cenários complexos e dinâmicos. Nesse sentido, a abordagem *ATOp*, ao combinar técnicas de AM com a explicitação semântica dos dados por meio de ontologias, apresenta potencial para aprimorar significativamente a eficácia dos processos decisórios no âmbito das Forças Armadas.

A adoção da *ATOp* se mostra particularmente relevante no contexto do Sistema Militar de Comando e Controle (SISMC<sup>2</sup>) (BRASIL, 2015a), que concentra os principais recursos de apoio à decisão operacional do Ministério da Defesa. Dentro dessa estrutura, destaca-se o Sistema de Planejamento Operacional Militar (SIPLM) (OURIQUES et al., 2012), plataforma dedicada ao planejamento de operações militares conjuntas. Por suas características funcionais, o SIPLM se apresenta como um candidato promissor para incorporar a abordagem *ATOp* (detalhada no Capítulo 5), contribuindo para tornar as análises logísticas mais ágeis, exatas e robustas (BRASIL, 2020).

Além disso, esta pesquisa dialoga diretamente com iniciativas estratégicas em curso, como o projeto Sistemas de Comando e Controle (S2C2), instituído por meio de convênio entre a FINEP e a FAPEB. O S2C2 tem como foco a modernização, interoperabilidade e inovação nos sistemas de C<sup>2</sup> das Forças Armadas, e a presente proposta busca contribuir com soluções tecnológicas alinhadas a esses objetivos, em especial no aprimoramento da capacidade preditiva da logística militar.

O avanço das tecnologias emergentes e a disponibilidade crescente de dados oriundos

de operações e treinamentos criam um ambiente propício ao desenvolvimento de modelos baseados em IA. A integração entre dados históricos e conhecimento semântico, viabilizada por técnicas de AM e MC, representa uma oportunidade concreta de evolução na forma como a autonomia de meios é estimada e utilizada no planejamento operacional.

Por fim, esta pesquisa busca fornecer uma base tecnológica sólida para a aplicação de soluções baseadas em IA no contexto da Marinha do Brasil, com foco na predição da autonomia de plataformas navais durante as atividades de C<sup>2</sup>. Ao mesmo tempo, espera-se que seus resultados contribuam para o amadurecimento científico em áreas como ciência de dados aplicada à defesa, ontologias e sistemas inteligentes, setores com grande potencial de impacto para a inovação em segurança e soberania nacionais.

## 1.7 ORGANIZAÇÃO DO TEXTO

O presente texto está organizado em sete capítulos. No Capítulo 1, dedicado à introdução, são apresentados o tema, os objetivos e a justificativa do estudo. No Capítulo 2, descreve-se em detalhes a metodologia adotada nesta pesquisa. No Capítulo 3, expõem-se as teorias fundamentais que compõem o arcabouço conceitual necessário para a compreensão do problema investigado. No Capítulo 4, realiza-se a análise dos trabalhos relacionados, por meio de uma revisão da literatura que posiciona o presente estudo no estado da arte. No Capítulo 5, detalha-se a abordagem metodológica empregada, incluindo as técnicas e ferramentas utilizadas para a resolução do problema de pesquisa. No Capítulo 6, apresenta-se a aplicação da abordagem proposta, acompanhada de uma análise detalhada dos resultados obtidos. No Capítulo 7, conclui-se o texto com uma discussão sobre os achados e suas implicações, destacando as contribuições da pesquisa e apontando possíveis direções para trabalhos futuros.

## 2 METODOLOGIA

Este estudo adota o método hipotético-dedutivo (POPPER, 2002) como abordagem principal para avaliar a adequabilidade de uma solução baseada em MC para o aprimoramento de modelos preditivos em um projeto de ciência de dados. A pesquisa parte da formulação da hipótese de que a introdução de uma etapa semântica, fundamentada na Modelagem Conceitual Orientada por Ontologias (*Ontology-Driven Conceptual Modeling* – ODCM) (GUIZZARDI, 2012), contribui para a melhoria da capacidade preditiva de modelos de AM. A ODCM é empregada na coleta de requisitos, concepção e desenvolvimento da solução, devido à sua capacidade de capturar conhecimentos intuitivos e de senso comum do domínio. A partir dessa hipótese, foram deduzidas consequências específicas sobre o impacto da abordagem *ATOp*, posteriormente testadas de forma empírica.

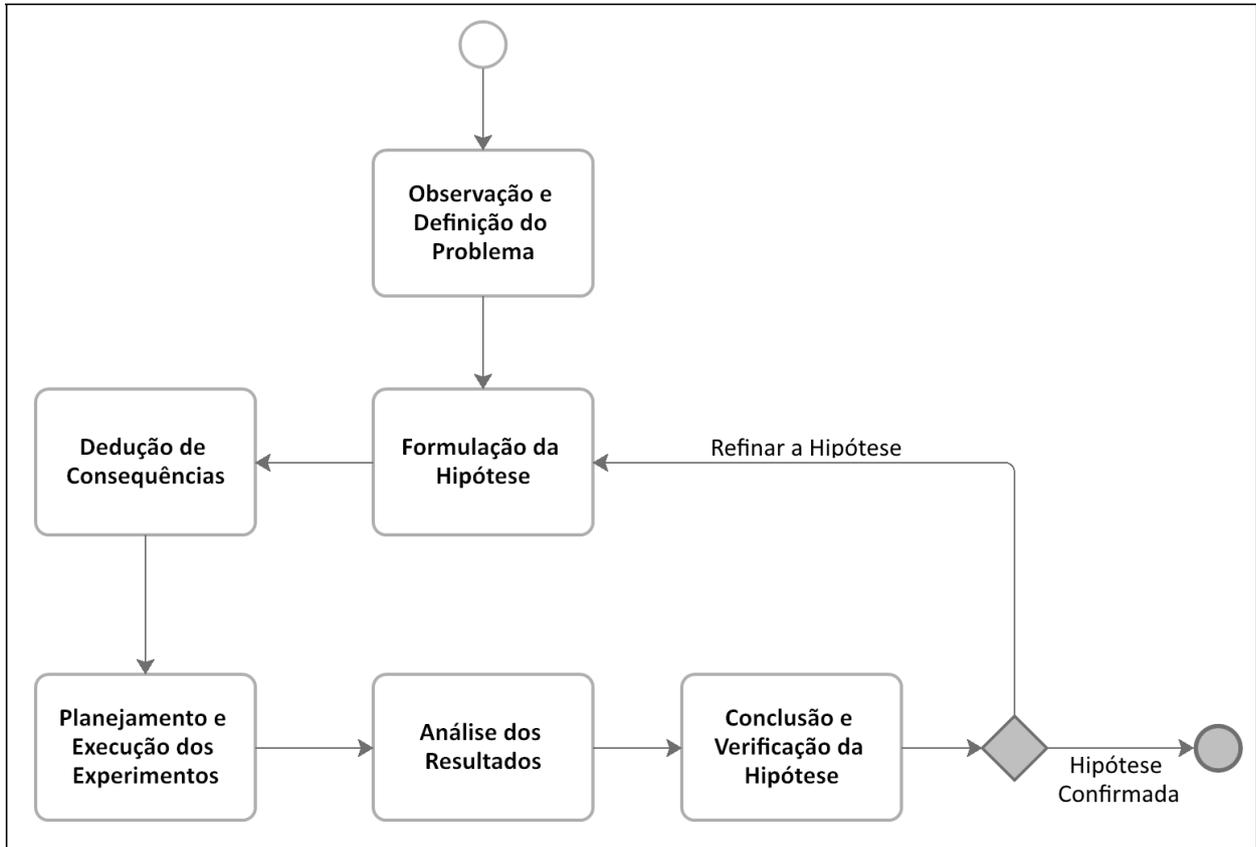
Como método de procedimento, foi conduzido um estudo de caso (YIN, 2009) no qual a abordagem *ATOp* (descrita no Capítulo 5) foi aplicada a dados históricos da Marinha do Brasil (MB). Esse estudo permitiu avaliar, de maneira sistemática, a influência da segmentação semântica sobre a capacidade preditiva dos modelos, utilizando métricas como o coeficiente de determinação ( $R^2$ ) e a raiz do erro quadrático médio (RMSE), bem como a execução de testes estatísticos apropriados (descritos na Seção 3.3.3).

A metodologia desta pesquisa é composta pelas etapas de: observação e definição do problema; formulação da hipótese; dedução de consequências; planejamento e execução dos experimentos; análise dos resultados; e verificação da hipótese. Nesse ciclo, a concepção da abordagem *ATOp* ocorre de forma central na etapa de dedução de consequências. É nesse momento que a hipótese proposta é operacionalizada, por meio da formulação de uma ontologia bem fundamentada, estruturada com base na UFO e na linguagem OntoUML, além da proposta de segmentação semântica dos dados como estratégia de apoio ao treinamento preditivo.

Na etapa seguinte, de planejamento e execução dos experimentos, essa concepção é concretizada: a *ATOp* é aplicada, com desenvolvimento da ontologia, integração e preparação dos dados, aplicação dos algoritmos de regressão e condução dos experimentos com rigor estatístico. Assim, a abordagem *ATOp* emerge como resultado direto da dedução das consequências esperadas da hipótese, posicionando-se metodologicamente como o elo entre a

teoria formulada e a validação empírica subsequente. A metodologia geral seguida é ilustrada na Figura 1 e detalhada nas seções a seguir.

Figura 1 – Fluxograma do método empregado neste estudo.



## 2.1 OBSERVAÇÃO E DEFINIÇÃO DO PROBLEMA

Esta seção apresenta a identificação do problema de pesquisa, com foco nas limitações do planejamento logístico de plataformas navais em atividades de C<sup>2</sup>. A partir do mapeamento de dificuldades operacionais, lacunas conceituais e limitações de abordagens quantitativas de AM, estabelecem-se as bases para uma proposta integrando ontologias e AM.

- **Levantamento de dificuldades no planejamento logístico:** mapeamento preliminar das dificuldades enfrentadas na gestão da autonomia de plataformas navais militares, destacando a importância da antecipação de demandas em cenários de C<sup>2</sup>.
- **Identificação de lacunas em MC e AM:** verificação de limitações tanto na representação semântica (etapa conceitual) quanto no poder preditivo de métodos tradicionais de AM aplicados de forma puramente quantitativa por meio da revisão bibliográfica.

- **Revisão bibliográfica:** análise de estudos que abordam a combinação entre MC e AM em bases como *ScienceDirect*, *IEEE Xplore*, *Web of Science*, *ACM Digital Library* e *Scopus*, e revisão de doutrinas militares do Ministério da Defesa<sup>1</sup>.

## 2.2 FORMULAÇÃO DA HIPÓTESE

Conforme exposto na Seção 1.3, parte-se da premissa de que a modelagem semântica contribui significativamente para o aumento da capacidade preditiva de modelos de AM. Espera-se, com isso, elevar as previsões da autonomia em atividades de C<sup>2</sup>.

## 2.3 DEDUÇÃO DE CONSEQUÊNCIAS

É nesta etapa do ciclo hipotético-dedutivo que se dá a concepção da abordagem *ATOp*, momento em que a hipótese inicialmente formulada é transformada em uma proposta concreta de solução. Essa concepção envolve, de forma articulada, o desenvolvimento teórico da abordagem, a definição dos seus macroprocessos e a estruturação ontológica do domínio com base na UFO e na linguagem OntoUML. Complementarmente, inclui-se a formulação de um pipeline de regressão, que incorpora a segmentação semântica dos dados como elemento central.

Como parte desta etapa, foi produzido um artefato computacional funcional que materializa a proposta metodológica. Esse artefato consiste em um protótipo implementado em ambiente *Python*, integrando os mecanismos de segmentação semântica e o modelo preditivo baseado em comitês de algoritmos. O protótipo viabiliza a operacionalização da hipótese formulada e serve como base para a execução dos experimentos e para a validação empírica da abordagem nas etapas subsequentes do ciclo metodológico.

Desse modo, a *ATOp* emerge como uma consequência lógica da hipótese de que a explicitação semântica pode aprimorar a capacidade preditiva de modelos de AM aplicados a contextos logísticos militares. Essa hipótese orienta a estruturação conceitual da proposta, a qual será posteriormente validada por meio de experimentação empírica. Nesse contexto, destacam-se duas premissas fundamentais associadas à hipótese:

- **Melhoria na capacidade preditiva:** a convergência entre ontologias bem fundamentadas e AM deve proporcionar desempenho superior às abordagens puramente

<sup>1</sup> <https://www.gov.br/defesa/pt-br/assuntos/estado-maior-conjunto-das-forcas-armadas/doutrina-militar/publicacoes-1>

quantitativas, refletido em métricas como  $R^2$  e RMSE;

- **Facilitação do treinamento:** a explicitação semântica tende a favorecer um ajuste mais eficiente dos modelos, contribuindo para um processo de aprendizagem mais eficaz.

## 2.4 PLANEJAMENTO E EXECUÇÃO DOS EXPERIMENTOS

As atividades de preparação dos dados, incluindo a extração de informações de sistemas administrativos e de relatórios operacionais, transformação de formatos, padronização de unidades e segmentação semântica, integram o escopo da engenharia de dados no contexto da ciência de dados. Essas tarefas foram fundamentais para garantir a qualidade e coerência dos dados utilizados na construção do *ATOp-PredictiveModel*.

Implementação prática da *ATOp*, desenvolvimento da ontologia, integração de dados, segmentação, treinamento e avaliação do modelo *ATOp-PredictiveModel*. Para esse fim, foram desenvolvidas atividades que abrangem desde a definição do estudo de caso até a preparação dos dados e a execução dos modelos de AM. Ao longo desta seção, descreve-se a aplicação da abordagem *ATOp* no contexto investigado.

### 2.4.1 Estudo de Caso e Aplicação da Abordagem *ATOp*

A abordagem *ATOp* combina a MC ontológica e técnicas de AM, aplicadas ao domínio logístico de plataformas navais da Marinha do Brasil. O estudo de caso focou-se em dados históricos do período de 2013 a 2017, abrangendo fatores como consumo de combustível, alimentos, água e composição da tripulação, que constituem elementos críticos para a avaliação da autonomia em operações navais. Esses fatores distinguem o meio operativo naval de outras forças, a exemplo do meio operativo aéreo da Força Aérea Brasileira, cuja autonomia costuma ser mensurada em horas de voo, ou do meio operativo terrestre do Exército Brasileiro, que depende quase exclusivamente de combustível. No contexto da Marinha, é comum o afastamento das redes de suprimento por períodos prolongados, exigindo, assim, um planejamento logístico apurado.

### 2.4.2 Aspectos Éticos e Confidencialidade

O acesso aos dados utilizados nesta pesquisa foi condicionado à obtenção de autorização formal da MB, em conformidade com os protocolos de segurança institucional e as normas

éticas aplicáveis à pesquisa científica. Os dados em questão são classificados como sigilosos, com prazo de restrição de acesso estabelecido em cinco anos, conforme as diretrizes internas da instituição. Após esse período, os dados poderão ser reclassificados ou desclassificados, a depender de avaliação da autoridade competente.

Durante a realização da pesquisa, foi assegurada a completa anonimização de quaisquer informações sensíveis, de forma a preservar a integridade institucional e a confidencialidade de operações, sistemas ou agentes envolvidos. Eventuais solicitações de acesso a esses dados, uma vez desclassificados, deverão ser encaminhadas formalmente à MB, observando os trâmites legais e administrativos previstos. Todo o processo respeitou os princípios da ética em pesquisa, com especial atenção à proteção de informações estratégicas e ao cumprimento das normas de sigilo vigentes.

### 2.4.3 Artefatos da Solução *ATOp*

As tarefas de preparação e organização dos dados, que envolvem desde a unificação de formatos até a segmentação semântica orientada por ontologias, configuram uma estratégia de engenharia de dados voltada à ciência de dados. A execução dessas atividades resulta na geração de artefatos distintos, que estruturam a solução *ATOp* e são essenciais para assegurar a qualidade das informações utilizadas na modelagem preditiva.

1. **Desenvolvimento da *ATOp-NavalOntology*:** modelagem do domínio com UFO e OntoUML.
2. **Coleta e Integração de Dados:** atividades de engenharia de dados voltadas à integração de dados estruturados (Sistema de Informações Gerenciais do Abastecimento – SINGRA) (BRASIL, 2020) e não estruturados (Relatórios de Fim de Comissão – RFC), incluindo etapas como extração, padronização, transformação e unificação das informações.
3. **Agrupamento de Entidades e Atributos:** definição de grupos semânticos com base em situações operacionais (como *em movimento*, *atracado na sede*, etc.), baseados em metacategorias da UFO, especialmente o uso da metacategoria *situation*.
4. **Mapeamento Ontológico dos Atributos:** separação dos atributos preditivos em subconjuntos de dados semanticamente coerentes.

5. **Treinamento e Avaliação dos Modelos:** construção do *ATOp-PredictiveModel* com *ensemble* otimizado via *Optuna* (AKIBA et al., 2019) e validação cruzada *k-fold*.

#### 2.4.4 Estratégia Experimental

Três blocos experimentais foram conduzidos:

- **Treinamento de Modelos Individuais:** aplicação de modelos individuais sobre os dados completos, sem segmentação semântica.
- **Treinamento de Comitês de Modelos sem Direcionamento Semântico:** formação de comitês de com combinações de modelos individuais, de forma parcialmente direcionada e aleatória.
- **Treinamento do ATOp-PredictiveModel:** aplicação do *ATOp-PredictiveModel* com divisão semântica baseada na ontologia.

#### 2.4.5 Validação Estatística dos Resultados

Após a coleta dos resultados:

- **Teste de *Friedman*:** utilizado para identificar diferenças estatisticamente significativas entre os modelos com base nas métricas  $R^2$  e RMSE.
- **Teste *post-hoc* de *Durbin-Conover*:** empregado para determinar quais pares de modelos apresentaram diferenças significativas, assegurando rigor estatístico às comparações.

#### 2.4.6 Codificação do Protótipo Funcional

Para materializar a abordagem *ATOp*, foi desenvolvido um protótipo funcional utilizando a linguagem *Python* e a biblioteca *Streamlit*<sup>2</sup>, que viabilizou a criação de uma interface interativa para emprego em sistemas de C<sup>2</sup>. O protótipo implementa todas as etapas da solução proposta, desde a preparação e segmentação semântica dos dados até o treinamento e aplicação do *ATOp-PredictiveModel*. As bibliotecas *pandas*, *scikit-learn*, *optuna* e

<sup>2</sup> <https://streamlit.io/>

`imbalanced-learn` foram empregadas para manipulação dos dados, modelagem e otimização. O código-fonte está disponível no repositório oficial do projeto<sup>3</sup>, permitindo sua replicação e extensão por outros pesquisadores.

## 2.5 ANÁLISE DOS RESULTADOS

- **Comparação da capacidade preditiva:** análise dos resultados de  $R^2$ , RMSE e tempo de treinamento entre os três blocos experimentais.
- **Discussão, limitações e implicações práticas:** identificação de desafios na integração ontológica e impacto prático da segmentação semântica em contextos de  $C^2$  e limitações de escopo ao domínio naval.

## 2.6 CONCLUSÃO E VERIFICAÇÃO DA HIPÓTESE

- **Confirmação ou refinamento da hipótese:** os resultados dos testes estatísticos foram utilizados para confirmar a hipótese inicial ou sugerir ajustes no modelo teórico, conforme evidências empíricas.
- **Contribuições:** reforço da importância da etapa semântica no aprimoramento da capacidade preditiva e na adequabilidade de modelos de AM aplicados a domínios críticos como o militar naval.

---

<sup>3</sup> <https://github.com/Valquire/ATOp>

### 3 REFERENCIAL TEÓRICO

Ao explorar em detalhes os fundamentos de ODCM, destacam-se suas vantagens e contribuições em diferentes aspectos. Entre eles, estão a compreensão aprofundada do domínio, a representação do conhecimento e a melhoria na comunicação entre as partes interessadas. Esses fatores ressaltam o papel crucial de ODCM no desenvolvimento de sistemas de informação eficientes, capazes de atender aos requisitos de negócios propostos.

O AM tem ganhado destaque como uma técnica fundamental para a resolução de problemas em diferentes áreas. Isso se deve à sua capacidade de permitir que os computadores aprendam diretamente a partir dos dados, sem depender de regras explicitamente programadas. Ao identificar padrões e regularidades nos dados, o AM possibilita a obtenção de *insights* valiosos, além de promover soluções mais eficientes e inovadoras (FACELI et al., 2021).

Enquanto os humanos podem levar horas ou até dias para analisar um conjunto de dados, as máquinas podem processá-los em questão de minutos ou segundos. Fato que ajuda a consolidar o discurso em prol da Tomada de Decisão Orientada a Dados (*Data-Driven Decision-Making* - DDDM) (PROVOST; FAWCETT, 2013). A DDDM pode contribuir para que as empresas tomem decisões mais rápidas e confiáveis com base em informações mais completas.

#### 3.1 OPERAÇÃO MILITAR, LOGÍSTICA MILITAR E COMANDO E CONTROLE (C<sup>2</sup>)

Esta seção aborda os principais conceitos de logística militar, operações militares, e comando e controle, com foco em suas definições, importância e inter-relações no contexto militar. O objetivo é fornecer uma compreensão de como esses elementos se conectam para assegurar o suporte eficiente, tanto em situações de conflito quanto em tempos de paz.

##### 3.1.1 Operação Militar

A operação militar é caracterizada como processo produtivo, constituído de atividades planejadas e executadas sob a responsabilidade de autoridade militar legalmente constituída, com o propósito de cumprir missões relacionadas à defesa nacional, à manutenção da ordem interna ou à realização de exercícios e manobras estratégicas. De forma geral, esse tipo

de operação visa proteger territórios, eliminar ameaças ou fornecer suporte em cenários adversos, com o uso de estratégias ofensivas ou defensivas conforme a necessidade da situação (BRASIL, 2015c; KRESS et al., 2002).

As operações militares podem ser classificadas quanto às forças empregadas em singulares, conjuntas e combinadas. Operações singulares são conduzidas por apenas uma das Forças Armadas (Exército, Marinha ou Força Aérea) de forma independente. Operações conjuntas envolvem a atuação coordenada de mais de uma força sob um comando unificado, garantindo sinergia operacional. Já as operações combinadas ocorrem com a participação de Forças Armadas de diferentes países, geralmente no contexto de alianças estratégicas ou missões de paz. Essa classificação é essencial para o planejamento eficaz das missões, permitindo melhor aproveitamento dos recursos e interoperabilidade entre as forças (BRASIL, 2020).

As operações militares podem ser realizadas não só em tempos de guerra quanto também em tempos de paz, como, por exemplo: as operações em apoio ao Estado ou em resposta a desastres naturais. Nesse contexto, as Forças Armadas são mobilizadas para auxiliar no socorro e recuperação de áreas atingidas, prestando assistência à população afetada. Esse tipo de operação é acionado quando os mecanismos civis, como os órgãos de segurança pública ou defesa civil, tornam-se insuficientes para atender à demanda gerada pela situação de calamidade, exigindo a atuação das Forças Armadas como instrumento de ampliação da capacidade operativa do Estado. (BRASIL, 2020).

Durante essas operações, os militares podem ser responsáveis por realizar tarefas como distribuição de alimentos e medicamentos, resgate de vítimas, reconstrução de infraestrutura básica e manutenção da ordem pública. Além de proteger as pessoas, eles também garantem a segurança das equipes de ajuda humanitária que atuam no local, criando um ambiente mais seguro para a execução dos trabalhos de recuperação.

### **3.1.2 Logística Militar**

A logística militar é uma função crítica dentro das operações militares, responsável por garantir que as Forças Armadas tenham os suprimentos necessários, como comida, água, munição e combustível, para sustentar sua capacidade de combate e sobrevivência. A logística militar inclui não apenas o transporte de materiais para as tropas em campo, mas também envolve um planejamento complexo da cadeia de suprimentos, adaptando-se às necessidades variáveis de cada missão e à evolução tecnológica (SIMON, 2001).

No contexto do planejamento da logística militar, autonomia diz respeito à capacidade de um veículo ou sistema, como navios, viaturas ou aeronaves, de funcionar de maneira contínua sem a necessidade de reabastecimento ou apoio externo. Esse conceito é essencial no planejamento de missões, pois garante que as operações possam ser conduzidas sem interrupções, maximizando a eficácia e a independência das forças no campo. Dependendo do tipo de veículo, a autonomia é definida de formas específicas, como por exemplo, é o tempo que um navio consegue permanecer no mar sem necessitar de reabastecimento de combustível, água potável ou suprimentos (BRASIL, 2015c).

Segundo Kress et al. (2002), as previsões relacionadas à autonomia de equipamentos militares podem ser elaboradas a partir de três métodos genéricos: abordagem intuitiva, métodos extrapolativos e modelos causais. A abordagem intuitiva baseia-se na experiência e no conhecimento tácito de especialistas do domínio. Os métodos extrapolativos, por sua vez, utilizam dados históricos com o objetivo de identificar padrões recorrentes e projetá-los no futuro, assumindo que as condições subjacentes permanecerão relativamente estáveis. Já os modelos causais incorporam o conhecimento situacional e as relações de causa e efeito entre variáveis, sendo particularmente úteis em contextos onde há mudanças frequentes ou fatores contextuais relevantes que influenciam os resultados.

Neste trabalho, optou-se por integrar técnicas extrapolativas e causais, visando combinar o poder preditivo dos dados históricos com a capacidade de adaptação às variáveis contextuais específicas do domínio militar. Essa integração permite capturar tanto regularidades estatísticas quanto nuances operacionais, resultando em previsões mais robustas, explicáveis e condizentes com a complexidade do cenário analisado.

### **3.1.3 Comando e Controle (C<sup>2</sup>)**

Comando e Controle (C<sup>2</sup>) constitui uma função crítica para a condução bem-sucedida das operações militares, compreendendo o conjunto de atividades voltadas ao planejamento, direção e coordenação de ações no ambiente operacional, com vistas ao alcance de objetivos estratégicos, operacionais e táticos. De acordo com a doutrina do Sistema Militar de Comando e Controle (SISMC<sup>2</sup>) (BRASIL, 2015a), o C<sup>2</sup> é simultaneamente uma ciência, por seu embasamento em doutrinas, processos e métodos sistematizados, e uma arte, ao demandar julgamento, intuição e experiência por parte dos tomadores de decisão. Essa função é sustentada por três componentes fundamentais e interdependentes: autoridade, processo

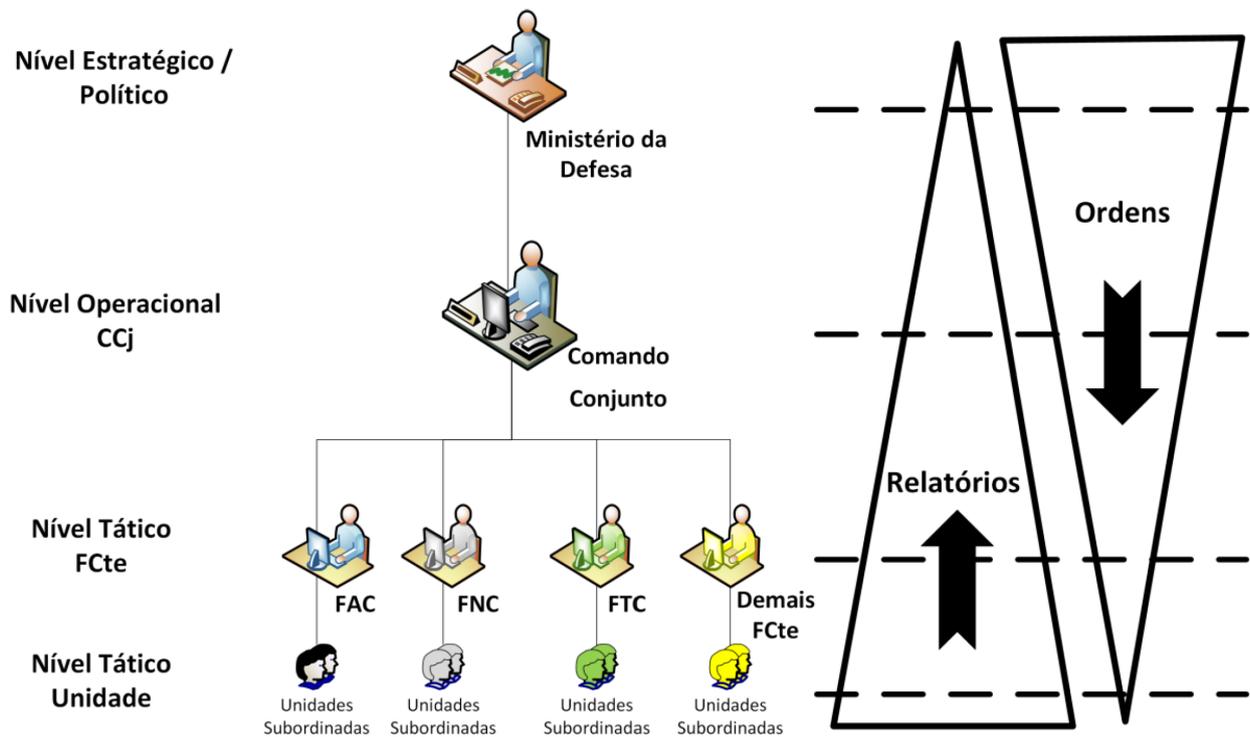
decisório e estrutura. A autoridade refere-se ao poder legítimo conferido ao comandante, que lhe permite emitir ordens e receber as informações necessárias à supervisão da execução. O processo decisório corresponde ao mecanismo por meio do qual o comandante formula decisões e orienta o fluxo de informações críticas à condução das operações, com base em doutrinas estabelecidas. Já a estrutura abrange os meios físicos e tecnológicos, como pessoal, sistemas, instalações e equipamentos, que viabilizam a materialização das atividades de comando e a efetivação do controle sobre a força empregada (BRASIL, 2015a).

O sistema de C<sup>2</sup> permite que o comandante tenha uma visão clara do ambiente operacional, desenvolva e execute planos com eficácia, e reaja rapidamente às mudanças na área de operação. Ele engloba também o uso de sistemas de informação e comunicação, que auxiliam na coleta, análise e disseminação de dados e informações para apoiar a tomada de decisão. Os sistemas de informação permitem flexibilidade e adaptabilidade ao ambiente dinâmico de C<sup>2</sup>, possibilitando respostas imediatas à medida que as situações operacionais mudam.

A Figura 2 apresenta o fluxo de informações e ordens dentro do SISMC<sup>2</sup>, destacando o papel da comunicação entre os níveis estratégico, operacional e tático. Esse fluxo ocorre de duas formas principais:

- **Descendente:** as ordens são transmitidas do nível estratégico para os níveis operacional e tático, assegurando a aplicação estruturada das diretrizes estabelecidas pelo alto comando no campo de operações. No nível tático, estão presentes as Forças Componentes (FCte), as quais podem ser classificadas em três categorias: Força Aérea Componente (FAC), Força Naval Componente (FNC) e Força Terrestre Componente (FTC).
- **Ascendente:** as informações coletadas nos níveis mais baixos retornam aos níveis superiores por meio de relatórios operacionais e atualizações situacionais, contribuindo para a construção da consciência situacional dos tomadores de decisão.

Complementando essa perspectiva, a Figura 3 evidencia o papel dos Centros de Comando e Controle (CC<sup>2</sup>), que são responsáveis pela recepção, análise e transmissão de informações dentro do sistema. A Rede Operacional de Defesa (ROD), o Sistema de Comunicações Militares por Satélite (SISCOMIS) e o Sistema Tático de Enlaces de Dados (SISTED) aparecem como elementos essenciais para garantir a conectividade segura entre os diferentes níveis de comando, assegurando que as informações sejam compartilhadas de forma rápida e confiável.

Figura 2 – Fluxos de informações e ordens no SISMC<sup>2</sup>.

Fonte: (BRASIL, 2015b).

Além disso, a Figura 3 apresenta um diagrama esquemático da organização do SISMC<sup>2</sup> dentro da Estrutura Militar de Defesa (EttaMiD). Nessa figura, são detalhados os diferentes sistemas de C<sup>2</sup> das Forças Armadas, incluindo:

- **SISNC<sup>2</sup>** (Sistema Naval de Comando e Controle) da Marinha do Brasil;
- **SisC<sup>2</sup>FTer** (Sistema de Comando e Controle da Força Terrestre) do Exército Brasileiro;
- **SisC<sup>2</sup>FAB** (Sistema de Comando e Controle da Força Aérea Brasileira).

A Figura 3 também destaca a presença de CC<sup>2</sup> permanentes e temporários, ativados conforme a necessidade operacional. Isso inclui os CC<sup>2</sup> do Ministério da Defesa (CC<sup>2</sup>MD), CC<sup>2</sup> dos Comandos Conjuntos (CC<sup>2</sup>CCj) e CC<sup>2</sup> de Forças de Paz (CC<sup>2</sup>FPaz), utilizados em operações internacionais.

Outro aspecto relevante dessa representação gráfica é a diferenciação entre ligações permanentes e temporárias. As ligações permanentes representam conexões estabelecidas continuamente entre os sistemas, garantindo o funcionamento regular da rede de comando e controle. Já as ligações temporárias são ativadas sob demanda, sendo empregadas em operações conjuntas.



A ciência da computação utiliza modelos conceituais para representar sistemas e domínios independente de tecnologia, permitindo que os modelos sejam aplicados a uma ampla variedade de contextos. Além disso, a MC é usada para apoiar a comunicação entre especialistas de domínio e desenvolvedores de software, ajudando a garantir que o software desenvolvido atenda às necessidades do usuário final, ou seja, possa representar as características físicas e sociais presentes no universo de discurso (MYLOPOULOS, 1992).

Segundo Guarino, Guizzardi e Mylopoulos (2020), os modelos conceituais inevitavelmente fazem compromissos ontológicos, pois estão conectados a uma conceituação capaz de fornecer sua semântica. O compromisso ontológico tem como cerne a premissa de que todo modelo conceitual se compromete com uma visão de mundo ou teoria do que existe em um domínio. Ainda segundo os autores, um modelo conceitual faz afirmações sobre a existência de entidades, suas qualidades e relações, bem como leis que restringem como esses elementos podem ser combinados. Dessa forma, todo artefato produzido com técnicas de representações visuais e linguagens empregadas em MC possui uma semântica conceitual, logo, faz um compromisso ontológico.

Um dos principais marcos na história da MC foi o trabalho de Chen (1976), que propôs o modelo Entidade-Relacionamento (ER). Esse modelo adere a uma visão do mundo real, constituída por entidades e relacionamentos. A visão de mundo real tem como base a linguagem natural, onde entidades, normalmente, são representadas por substantivos e relacionamentos por verbos, e está permeada por uma ontologia, provendo semântica aos dados. O modelo ER não se preocupa com a ontologia, ele somente é capaz de fornecer uma representação, tendo como suporte a linguagem natural. Devido à extensa pesquisa e exploração ao longo do tempo, alguns estudos publicados em anos subsequentes ampliaram o modelo ER, incorporando elementos de agregação e generalização.

Outro importante fato histórico para a MC foi a criação da Linguagem de Modelagem Unificada (*Unified Modeling Language - UML*) (BOOCH et al., 1996), que carrega o conceito de orientação a objeto. A UML tornou-se uma linguagem padrão para modelagem e é mantida pela *Object Management Group* (OMG). A UML emprega uma abordagem com identificadores, independentes de atributos, para tratamento da representação única dos objetos.

A utilização dos modelos ER e UML, dá ao modelador a liberdade para criar visões de mundo a partir do seu próprio senso, ou seja, um modelo poderia ser carregado com um

viés. Esse viés representa o conhecimento do modelador sobre o micromundo ou universo de discurso. As diferentes visões ou semânticas divergentes elaboradas por diversos modeladores contribuíram de maneira significativa para o surgimento de um dos principais desafios enfrentados pelos sistemas de informação atualmente: a falta de interoperabilidade.

A semântica desempenha um papel fundamental na interoperabilidade de sistemas de informação. Ela se refere ao significado e interpretação dos dados e informações trocados entre sistemas, o que é crucial para garantir que diferentes sistemas possam compreender e usar esses dados de maneira adequada e exata.

### 3.2.1 Ontologias

Guarino (1998) foi um dos pioneiros a reconhecer que alcançar a interoperabilidade semântica demandava a utilização de padrões semânticos, como ontologias. As ontologias definem, de forma concisa, os significados e relacionamentos entre os termos e conceitos usados nos dados e informações.

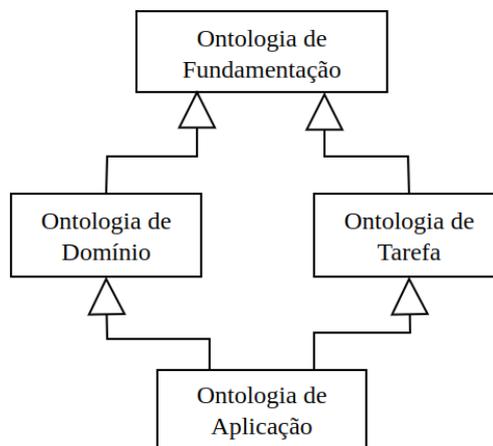
Portanto, a semântica desempenha um papel crítico na interoperabilidade de sistemas de informação, garantindo que a troca de dados e informações seja eficaz, clara e significativa, independentemente das diferenças tecnológicas e estruturais entre os sistemas envolvidos. As soluções baseadas em abordagens ontológicas têm trazido enormes contribuições nessa direção. O uso de ontologias na ciência da computação se concentra na representação e organização do conhecimento de um domínio específico. Elas são usadas para modelar e descrever conceitos, entidades e suas relações dentro de um determinado campo de conhecimento. Uma ontologia pode ser considerada como uma estrutura formal que captura o significado e a semântica dos objetos e conceitos em um domínio específico (GUARINO, 1998).

O conceito filosófico de ontologia refere-se ao estudo do que existe no mundo e como as coisas estão relacionadas entre si. Ele lida com questões como: Que tipos de coisas existem? Quais são suas propriedades e relações? Como podemos classificá-las? A ontologia está preocupada com a natureza da realidade e as categorias que usamos para descrevê-la. No contexto da MC, as ontologias são representações formais de uma visão de mundo ou teoria particular do que existe em um domínio. Elas fornecem um vocabulário compartilhado e um conjunto de restrições que pode ser usado para raciocinar sobre as entidades, suas qualidades e suas relações em um domínio (GUARINO; GUIZZARDI; MYLOPOULOS, 2020).

As ontologias desempenham um papel crucial na ciência da computação. Elas são uma ferramenta poderosa para a modelagem, organização e processamento de informações, permitindo que sistemas de computação entendam e usem o conhecimento de maneira aprimorada e inteligente (GUARINO; GUIZZARDI; MYLOPOULOS, 2020).

Uma forma de classificação de ontologias foi proposta por Guarino (1998), e tem sido largamente empregada. A Figura 4 ilustra que essa classificação está baseada no nível de generalização da ontologia.

Figura 4 – Tipos de ontologias.



Fonte: (GUARINO, 1998).

As ontologias de fundamentação ou nível superior exploram conceitos universais e amplamente abstratos, como espaço, tempo, matéria, objeto, evento, entre outros, que são independentes de um problema ou domínio específico. Elas representam os alicerces sobre os quais todo o conhecimento posterior é construído.

As ontologias de domínios são especializadas e adaptadas a um domínio específico. Elas são construídas com base em especializações de conceitos introduzidos em ontologias de fundamentação e se concentram em detalhes específicos do domínio. Por exemplo, uma ontologia de domínio para a segurança e defesa pode incluir informações sobre a atividade militar, as forças militares e as operações militares.

As ontologias de tarefa são direcionadas à representação de conceitos, propriedades e relações pertinentes à execução de uma tarefa específica. De modo análogo às ontologias de domínio, elas se baseiam em ontologias de fundamentação. Por exemplo, uma ontologia para suprimento logístico militar pode incluir informações sobre meios operativos, materiais de consumo e locais de apoio.

As ontologias de aplicação são desenvolvidas para atender a necessidades específicas de uma aplicação ou sistema. Elas podem ser construídas com base em ontologias de domínio e de tarefa, adaptando-se às demandas do contexto de uso. Um exemplo seria uma ontologia de aplicação para gerenciamento de materiais de consumo em um sistema de gestão administrativa.

As ontologias de fundamentação mais amplamente reconhecidas na literatura incluem a *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE) (MASOLO et al., 2002), a *Unified Foundational Ontology* (UFO) (GUIZZARDI, 2005) e a *Basic Formal Ontology* (BFO) (ARP; SMITH; SPEAR, 2015; SPEAR; CEUSTERS; SMITH, 2016). Embora compartilhem o objetivo comum de fornecer um arcabouço conceitual sólido para a modelagem do conhecimento, essas ontologias diferem em suas abordagens filosóficas e na forma como estruturam a realidade. A UFO apresenta uma das abordagens mais desenvolvidas para a representação de entidades e relações, proporcionando uma descrição mais explícita e estruturada do domínio. Além disso, oferece uma estrutura conceitual robusta e expressiva, essencial para uma modelagem ontológica detalhada (BERNASCONI et al., 2022).

### 3.2.2 Systematic Approach for Building Ontologies (SABiO)

O desenvolvimento de ontologias é uma atividade complexa que requer a adoção de processos sistemáticos, estruturados em diretrizes bem definidas, para garantir a consistência, a clareza e a utilidade do modelo resultante. Nesse contexto, a metodologia SABiO (FALBO, 2014) destaca-se como uma abordagem de grande relevância, pois incorpora recomendações fundamentadas em boas práticas de modelagem.

A metodologia SABiO (FALBO, 2014) é uma abordagem sistemática para o desenvolvimento de ontologias de domínio, que incorpora melhores práticas da Engenharia de Software e da Engenharia de Ontologias. O método distingue entre ontologias de referência e ontologias operacionais, aplicando processos específicos para a construção de cada tipo.

As ontologias de referência são concebidas com o propósito de representar, de forma concisa, clara e independente de solução, a estrutura conceitual de um domínio. Elas funcionam como modelos conceituais ricos e bem fundamentados, voltados para a compreensão, comunicação e análise do conhecimento sobre um determinado campo. Por isso, são altamente expressivas, podendo incluir axiomas complexos e distinções ontológicas refinadas, geralmente fundamentadas em uma ontologia de alto nível, como a UFO. Essas ontologias

são voltadas para uso humano e têm como foco a adequação da representação.

Já as ontologias operacionais derivam, geralmente, de ontologias de referência e têm como foco a eficiência computacional e a aplicabilidade prática em sistemas computacionais. Elas são implementadas em linguagens formais específicas, como OWL<sup>1</sup> ou RDF<sup>2</sup>, levando em conta restrições tecnológicas e requisitos não funcionais, como capacidade de inferência, interoperabilidade e usabilidade em ambientes reais. Para isso, muitas vezes é necessário simplificar ou adaptar a estrutura rica da ontologia de referência para que ela seja processável e utilizável por máquinas.

A metodologia SABiO é estruturada em cinco fases principais que compõem o seu processo de desenvolvimento, cada uma delas com objetivos e atividades específicas voltadas à construção sistemática de ontologias de domínio. Essas fases abrangem desde a identificação do propósito e requisitos da ontologia até a sua implementação e teste, promovendo um fluxo coerente e iterativo de trabalho., conforme descrito a seguir:

1. Identificação de propósito e elicitação de requisitos: Esta fase inicial foca na definição do propósito da ontologia, nos usos pretendidos e nos requisitos necessários, incluindo requisitos funcionais, também conhecidos como questões de competência, e não funcionais. Além disso, promove a modularização da ontologia em sub-ontologias.
2. Captura e formalização da ontologia: Baseando-se nas questões de competência, os conceitos e relações são identificados e organizados em modelos conceituais. A análise ontológica, fundamentada em uma ontologia de base, é aplicada para garantir rigor semântico e consistência. Nesta fase, axiomas são definidos em linguagem natural e, posteriormente, formalizados em linguagens expressivas.
3. Projeto da ontologia: Para ontologias operacionais, esta fase traduz a especificação conceitual em uma especificação de design, considerando requisitos tecnológicos não funcionais e o ambiente de implementação. Isso inclui a definição da arquitetura da ontologia e ajustes para atender às limitações das linguagens operacionais.
4. Implementação: Nesta etapa, a ontologia é codificada em uma linguagem operacional apropriada, como OWL. A *W3C Web Ontology Language* (OWL) é uma linguagem de

---

<sup>1</sup> <https://www.w3.org/OWL/>

<sup>2</sup> <https://www.w3.org/RDF/>

representação do conhecimento que compõem a pilha da Web Semântica, considerando as especificações de *design* previamente elaboradas.

5. Teste: A ontologia é testada de forma dinâmica, verificando sua capacidade de responder às questões de competência e avaliando requisitos não funcionais, como desempenho e robustez. O teste ocorre em sub-ontologias, integrações e no modelo completo, podendo incluir validação por meio de aplicações práticas.

Além das etapas do desenvolvimento, a metodologia SABiO inclui processos de suporte, como aquisição de conhecimento, reuso, documentação, gestão de configuração e avaliação. Esses processos asseguram a qualidade e a continuidade do projeto, alinhando o desenvolvimento com os requisitos do domínio e promovendo a reutilização de recursos existentes.

Essa metodologia não apenas organiza o processo de construção de ontologias, mas também promove a reutilização de recursos, a análise rigorosa baseada em ontologias de base e a formalização clara dos conceitos, relações e restrições que compõem o domínio em questão. Assim, a SABiO tem contribuído de maneira significativa para a sistematização e evolução das práticas na engenharia de ontologias, consolidando-se como um referencial amplamente reconhecido na área.

Neste trabalho, optou-se pela utilização exclusiva de uma ontologia de referência. Tal escolha justifica-se pelo objetivo principal de fornecer uma representação conceitual clara e devidamente fundamentada do domínio da autonomia de meios operativos, sem a exigência imediata de uma implementação computacional.

### 3.2.3 Unified Foundational Ontology (UFO)

A UFO é uma ontologia de fundamentação proposta por Guizzardi (2005). A UFO tem como base alguns princípios de áreas multidisciplinares, como a Ontologia Formal na Filosofia, a Psicologia Cognitiva e a Linguística.

O desenvolvimento da UFO teve como elemento substancial a distinção entre os conceitos ontológicos de Tipos (*Types*) e de Indivíduos (*Individuals*). Tipos são coisas abstratas que são criadas para tarefas de classificação de tudo que existe no mundo. Nesse sentido, essas coisas carregam diversas características que podem ser encontradas em outras coisas mais específicas, chamadas de indivíduos ou instâncias dos tipos. Por exemplo, Marruá 9009 é um

indivíduo que é instância do tipo Viatura Blindada, de forma semelhante, o soldado Lucas instancia Militar e Pessoa.

Outro conceito ontológico relevante para a UFO é o Princípio de Identidade, usado para distinguir dois indivíduos. Esse princípio é derivado da filosofia da identidade e é crucial para a criação de modelos ontológicos precisos. O Princípio de Identidade estabelece que duas entidades são idênticas, ou seja, representam a mesma coisa no mundo real, se e somente se elas possuem exatamente as mesmas características essenciais e são indistinguíveis em todos os aspectos relevantes.

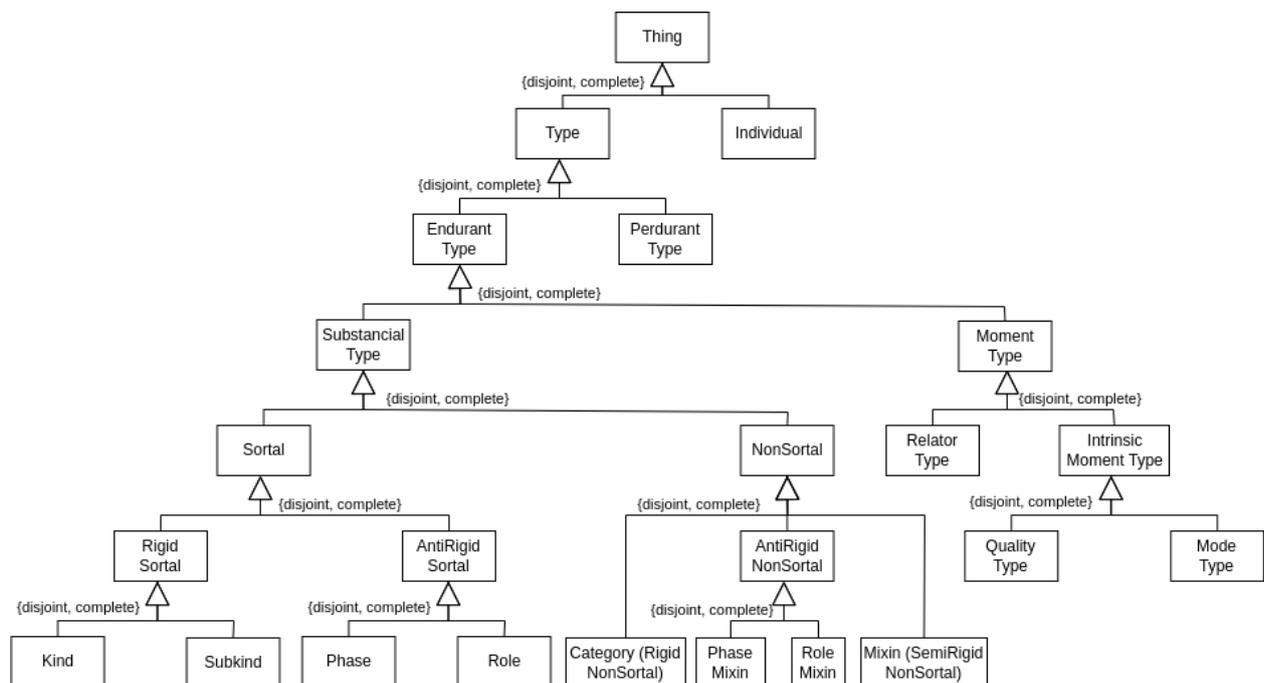
A UFO está organizada em três ontologias principais, que continuam sendo exploradas e desenvolvidas:

- UFO-A, composta por entidades que caracterizam tipos de indivíduos que mantêm suas partes temporais ao longo de sua existência, portanto, mantêm sua identidade. Eles são conhecidos como Endurantes (*Endurants*). São exemplos de Endurantes: fragata Niterói, submarino Tupi, caça Gripen 56, etc.
- UFO-B, constituída por entidades que caracterizam tipos de indivíduos que se desdobram no tempo enquanto acumulam partes temporais, porém, não há a manifestação de identidade, somente de disposições. Disposições são propriedades intrínsecas de um objeto que indicam sua capacidade ou propensão para realizar certas atividades ou comportamentos. Por exemplo, se considerarmos um objeto como um artefato explosivo, uma disposição associada a esse artefato pode ser a capacidade de explodir. Em termos ontológicos, a disposição de explodir seria uma característica inerente à natureza do artefato, indicando sua capacidade de desempenhar essa função específica. Os Perdurantes (*Perdurants*), como são conhecidos, são entidades que se estendem no tempo e só existem no passado. Dentre os diversos tipos de Perdurantes, destacam-se as batalhas históricas, que representam uma classe bem definida dentro dessa categoria. Exemplos de instâncias desse tipo incluem: a Independência da Bahia, em 2 de julho de 1823, quando as tropas brasileiras venceram os portugueses nos arredores de Salvador; a Batalha Naval do Riachuelo, em 11 de junho de 1865, marcada pela bravura de aguerridos marinheiros e fuzileiros navais durante a Guerra da Tríplice Aliança; e o Dia D, em 6 de junho de 1944, quando as tropas Aliadas desembarcaram na Normandia, na França, durante a Segunda Guerra Mundial.

- UFO-C, fundamentada em entidades sociais e intencionais, abrangendo conceitos de agentes e ações. Por exemplo, considera uma pessoa como um agente físico e uma empresa como um agente social. Ela estende os conceitos da UFO-B, apoiada pela UFO-A.

A UFO busca fornecer uma base comum para a representação e a integração de diferentes ontologias específicas de domínio. Ela foi projetada para capturar os conceitos fundamentais e as relações ontológicas que são comuns a várias áreas do conhecimento. A ideia por trás da UFO é criar uma estrutura conceitual unificada, que possa ser utilizada como uma linguagem comum e independente de domínio em ODCM (GUIZZARDI, 2005). A Figura 5 ilustra um fragmento da taxonomia da UFO-A.

Figura 5 – Fragmento da taxonomia da UFO-A.



Fonte: (GUIZZARDI et al., 2022).

A estrutura da UFO tem, em sua essência, categorias ontológicas, que se referem a tipos fundamentais de entidades usadas para modelar diferentes aspectos do mundo. Estas categorias fornecem uma base fundamental para a criação de modelos ontológicos ricos em semântica. Essas categorias ontológicas, ilustradas na Figura 5, são:

- Substanciais (*Substantials*) definem entidades independentes que estendem os *Endurants*. Por outro lado, Momentos (*Moments*) são entidades que representam propriedades ou características existencialmente dependentes de outras entidades, como *Sub-*

*tancials*. Exemplos de *Moments* incluem a cor de um uniforme, o tempo de serviço um militar em um momento específico e a posição de um alvo em um instante particular.

- Sortais (*Sortals*) são categorias ou tipos de entidades com critérios de identidade bem definidos. Eles representam objetos que podem ser identificados e individualizados de forma única no mundo real. Exemplos de *Sortals* incluem humanos, armas nucleares, aeronaves de asa rotativa, etc. *Sortals* normalmente possuem atributos essenciais e são frequentemente caracterizados por um conjunto de condições necessárias e suficientes para adesão.
- Não Sortais (*NonSortals*) são categorias ou tipos de entidades que não possuem critérios de identidade bem definidos. Eles representam conceitos que não podem ser individualizados de forma única. Estas entidades são tipicamente caracterizadas pela sua natureza relacional e contextual.
- A rigidez é metapropriedade ontológica fundamental dos tipos. Um tipo será considerado rígido (*Rigid*) quando possibilitar que um indivíduo seja instanciado em todos os cenários de sua existência. Por exemplo, um indivíduo que instancia os tipos Pessoa e Soldado nunca deixará de ser Pessoa enquanto existir, pois a característica de ser Pessoa é intrínseca e essencial à sua identidade. No entanto, o tipo Soldado pode ser considerado antirrígido (*AntiRigid*), pois um indivíduo pode deixar de ser Soldado ao mudar de posição ou função dentro da carreira militar, como ao se tornar Oficial. Assim, os tipos rígidos definem características essenciais, enquanto os antirrígidos representam papéis ou estados transitórios.

Além das categorias ontológicas, a UFO oferece metaclasses, que compõem um conjunto abrangente de estruturas, para modelar diferentes aspectos das entidades em um domínio específico de interesse. As metaclasses da UFO-A, ilustradas na Figura 5, são descritas a seguir:

- *Kind*: Uma metaclassa fundamental de entidades que compartilham características essenciais, que são necessárias e suficientes para a definição e identificação de todos os seus membros. Os *Kinds* representam categorias naturais de entidades. Por exemplo, o tipo Militar é um *kind*, pois todos os indivíduos que pertencem a essa classe compartilham características essenciais, como estarem formalmente ligados a uma força armada

e terem o dever de seguir um código de conduta específico. Essas características são invariantes, independentemente de variações contextuais;

- *Subkind*: Uma especialização de um *kind* que herda todas as suas características essenciais, mas refina essa classe geral com características adicionais que criam subcategorias mais específicas. Por exemplo, Oficial é um *subkind* de Militar, pois herda todas as características de Militar, mas também inclui propriedades adicionais, como a responsabilidade de comandar tropas e tomar decisões estratégicas;
- *Role*: Representa um conjunto de comportamentos, funções ou responsabilidades desempenhadas por uma entidade quando ela participa de um contexto ou situação específica. Diferentemente de *kinds* e *subkinds*, o papel que uma entidade desempenha é dependente do contexto. Por exemplo, um Militar pode desempenhar o papel de Instrutor durante um treinamento, mas em outro contexto, pode assumir o papel de Comandante de uma operação;
- *Phase*: Descreve os estágios temporais no ciclo de vida de uma entidade. As fases são mutuamente exclusivas e representam a mudança de estado de uma entidade ao longo do tempo. Por exemplo, um militar pode passar pelas fases de Formação, Ativo, Reserva, Tarefa por Tempo Certo e Reformado, sendo cada fase caracterizada por diferentes direitos, responsabilidades e condições;
- *Category*: Representa uma classe de entidades que compartilham características comuns, mas sem a exigência de uma essência ou natureza essencial compartilhada. Por exemplo, a categoria Forças Armadas agrupa entidades como Exército, Marinha e Força Aérea, com base em sua função de proteger a nação, embora não compartilhem uma essência comum;
- *Role Mixin*: Uma combinação de papéis que uma entidade pode desempenhar simultaneamente em diferentes contextos. Por exemplo, um militar pode ser simultaneamente um Operador de Comunicações e um Oficial Logístico, desempenhando ambos os papéis em situações específicas que exigem essas funções;
- *Phase Mixin*: Assim como o *role mixin*, o *phase mixin* permite que uma entidade esteja em múltiplas fases de seu ciclo de vida ao mesmo tempo. Por exemplo, um militar pode

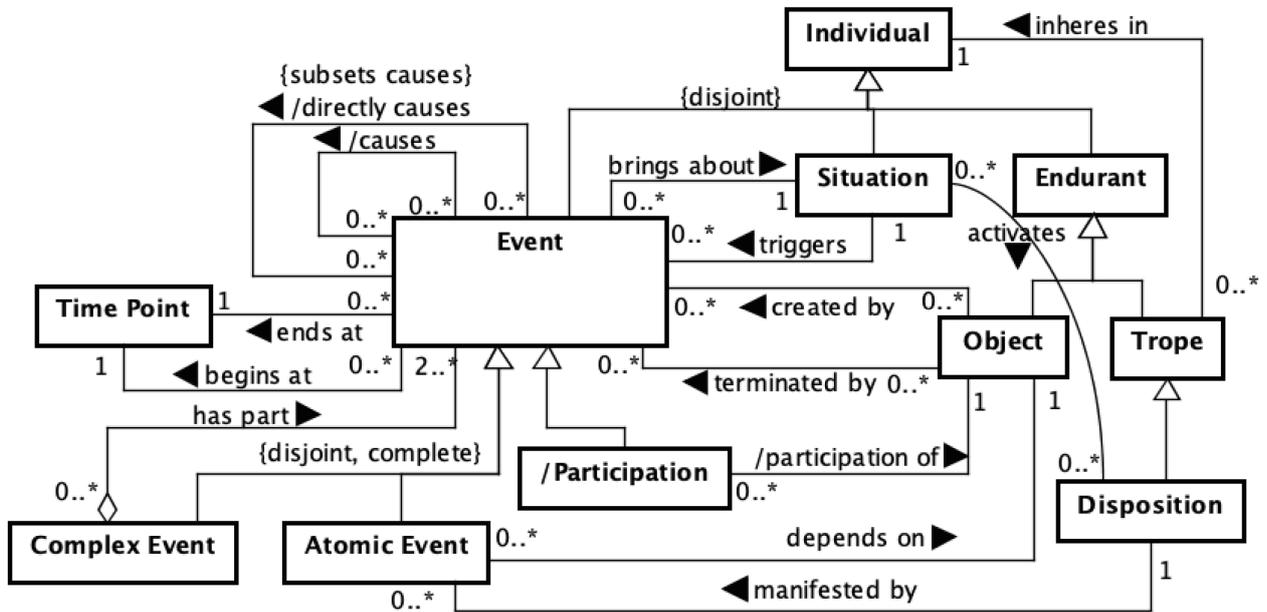
estar simultaneamente na fase de Reserva e em Tarefa por Tempo Certo, refletindo a sobreposição de estados no ciclo de vida;

- *Mixin*: Um *mixin* é uma metacategoria que permite agrupar propriedades ou comportamentos que podem ser compartilhados entre diferentes tipos de entidades, sem que essas entidades necessariamente pertençam à mesma categoria. Por exemplo, o *mixin* Identificação Militar pode incluir atributos como número de registro, posto e unidade de lotação, aplicáveis tanto a Oficiais quanto a Praças;
- *Relator*: Entidade ontológica que estabelece ou media uma relação entre duas ou mais entidades, criando uma dependência mútua entre elas. Por exemplo, uma Ordem de Operação Militar é um *relator* que conecta uma unidade militar a um objetivo estratégico, especificando os papéis e responsabilidades de cada participante;
- *Quality*: Refere-se a características ou propriedades que algo possui e que podem ser observadas, avaliadas ou medidas. Por exemplo, a eficiência operacional de uma unidade militar é uma qualidade que pode ser medida em termos de tempo de resposta ou cumprimento de metas em missões;
- *Mode*: Um *mode* é uma entidade que não pode existir de forma independente, pois depende da existência de outras entidades para ser. Por exemplo, a habilidade de pilotar um helicóptero militar é um *mode*, pois essa habilidade depende da existência de um militar treinado e não pode existir de forma independente dele.

A ontologia UFO-B, dentro da UFO, fornece uma estrutura robusta para modelar e entender eventos e processos dinâmicos em diversos contextos. O principal conceito presente na UFO-B é dos Eventos (*Events*). Eventos são considerados entidades que acontecem no tempo e podem ter participantes e sub-eventos. Eles são fundamentais para a descrição de dinâmicas e mudanças no sistema. As metaclasses da UFO-B são ilustradas na Figura 6.

A construção *Situation* da UFO é uma metacategoria ontológica que descreve configurações específicas de uma parcela da realidade, geralmente associadas ao estado de objetos e/ou qualidades em um dado instante. Na UFO, situações são consideradas elementos essenciais para a compreensão dos processos pelos quais eventos transformam a realidade (GUIZZARDI, 2005).

Figura 6 – Fragmento da taxonomia da UFO-B.



Fonte: (ALMEIDA; FALBO; GUIZZARDI, 2019).

Formalmente, situações podem ser definidas como partes concretas da realidade que se configuram como fatos quando ocorrem em um determinado momento. Essas situações são compostas por indivíduos, incluindo suas qualidades. Por exemplo, a situação de um veículo blindado danificado em combate consiste na combinação do veículo (o objeto), o local onde o combate ocorreu (o contexto) e o dano sofrido, que é uma qualidade associada ao veículo no momento específico (BARATELLA et al., 2022).

Em UFO-B, os eventos são transformações que levam de uma situação para outra. Uma situação que desencadeia (*trigger*) um evento começa ao mesmo tempo que esse evento. Já uma situação que é provocada (*brought about*) por um evento começa quando este termina. No contexto deste estudo, o conceito de *situation* é utilizado para compreender a causalidade e a sequência de transformações na realidade dentro do domínio da autonomia de meios operativos.

### 3.2.4 OntoUML

A OntoUML constitui uma das contribuições mais relevantes do trabalho de Guizzardi (2005), desenvolvida como uma metodologia específica para a construção de ontologias fundamentada na UFO. Essa linguagem, ontologicamente bem fundamentada, é amplamente utilizada no contexto da ODCM.

Essa linguagem faz uso de uma extensão dos fragmentos dos diagramas de classe da UML,

possibilitando construções específicas para a ODCM. Contudo, OntoUML suporta apenas a especificação de *Types*. Por isso, o modelo foi constituído somente por diagrama de classes - UML, sem a especificação de *Individuals*.

A UFO e a linguagem OntoUML têm sido amplamente aplicadas em diversos projetos e áreas de pesquisa. Seus usos abrangem desde engenharia de software e sistemas de informação até ciências da saúde e ciência da computação em geral. Essas abordagens têm auxiliado desenvolvedores na tarefa de representar o conhecimento de forma explícita e semanticamente rica.

O trabalho de Guizzardi (2005) tem desempenhado um papel significativo na promoção do uso da OntoUML na área de ciência da computação. A OntoUML tem sido amplamente empregada em pesquisas relacionadas a ODCM e recebeu reconhecimento internacional por sua relevância e impacto (VERDONCK et al., 2019), o que justifica sua adoção como linguagem para o desenvolvimento do modelo conceitual proposto neste estudo.

### 3.3 APRENDIZADO DE MÁQUINA

O AM tem suas raízes na IA, uma vez que surgiu como uma subárea voltada ao desenvolvimento de sistemas capazes de aprender a partir de dados, sem a necessidade de programação explícita para cada tarefa. A história do AM é permeada por uma combinação de teorias fundamentais, avanços tecnológicos e aplicações práticas que revolucionaram nossa interação com a tecnologia e o processamento de dados.

O conceito de AM emergiu nas décadas de 1940 e 1950, quando pesquisadores visionários começaram a explorar a possibilidade de máquinas aprenderem com dados e experiências. Contudo, as limitações computacionais da época impuseram barreiras significativas. O verdadeiro impulso veio posteriormente, com avanços no poder computacional e o manuseio de estruturas de dados mais extensas.

As Redes Neurais Convolucionais (*Convolutional Neural Networks* – CNN) e as Máquinas de Vetores de Suporte (*Support Vector Machines* – SVM) desempenharam um papel fundamental na consolidação do AM como uma disciplina independente nas décadas de 1980 e 1990. Os estudos pioneiros de LeCun et al. (1998) e Vapnik (1998) foram marcos importantes no desenvolvimento desses métodos, que viriam a se tornar amplamente utilizados em diversas aplicações de AM.

O início dos anos 2000 marcou uma virada significativa com a ascensão do fenômeno

conhecido como *Big Data*. Caracterizado por seus múltiplos “V” — volume, variedade, velocidade, veracidade e valor —, esse novo cenário foi impulsionado pela geração massiva e contínua de dados em diferentes formatos, a partir de diversas fontes. Esse aumento exponencial na disponibilidade de dados viabilizou o treinamento de modelos mais complexos e expressivos. Pesquisas como a de Hinton et al. (2012) desempenharam um papel fundamental nesse contexto ao demonstrar que redes neurais profundas, quando treinadas com grandes volumes de dados e técnicas como a pré-inicialização por camadas (*pre-training*), podiam superar significativamente os métodos existentes em tarefas complexas. Esse trabalho marcou um ponto de inflexão no campo, ao obter resultados de destaque em competições de reconhecimento de imagem, e impulsionou o uso das Redes Neurais Profundas em uma ampla gama de aplicações, incluindo processamento de linguagem natural, sistemas de recomendação e jogos.

Atualmente, vivemos na era da IA aplicada, onde o AM é essencial para uma variedade de soluções tecnológicas. Modelos de linguagem avançados, como os *Large Language Models* (LLMs), são exemplos emblemáticos de como a capacidade de processar informações e gerar texto alcançou patamares extraordinários, moldando aplicações desde assistentes virtuais sofisticados até a criação de conteúdo gerado por máquina (NAVEED et al., 2023).

De maneira mais específica, as pesquisas em AM têm se concentrado no desenvolvimento de algoritmos e métodos para capacitar sistemas computacionais a realizarem tarefas preditivas e descritivas de maneira autônoma. Esses sistemas aprendem a partir de dados, construindo modelos que identificam padrões e realizam inferências sem necessidade de programação explícita para cada tarefa. O processo de aprendizado ocorre através do treinamento de modelos, no qual os algoritmos ajustam parâmetros com base em um conjunto de dados composto por atributos (variáveis) preditivos ( $X_1, X_2, \dots, X_n$ ) e, em muitos casos, um atributo alvo ( $Y$ ) ou variável resposta (MONARD; BARANAUSKAS, 2003).

Os atributos preditivos representam características ou variáveis que descrevem as instâncias (observações) do conjunto de dados, formando o vetor de entrada do modelo. O atributo alvo, por sua vez, é utilizado para rotular as instâncias, podendo conter valores categóricos, como classes, ou numéricos, para os problemas de regressão. Em problemas supervisionados, o treinamento busca minimizar a diferença entre os valores preditos pelo modelo e os valores reais do atributo alvo, otimizando a capacidade preditiva do sistema (FACELI et al., 2021).

Por outro lado, quando os conjuntos de dados não possuem um atributo alvo (cenário tí-

pico de tarefas descritivas), o aprendizado não supervisionado é empregado. Esse paradigma foca na identificação de padrões, como agrupamentos ou associações, entre as instâncias do conjunto de dados. Assim, o AM oferece uma ampla gama de técnicas e métodos para abordar tanto problemas supervisionados quanto não supervisionados, com aplicações que vão desde a previsão de eventos até a descoberta de conhecimento em bases de dados complexas (FACELI et al., 2021).

Um sistema de informação baseado em AM é concebido com o propósito de adquirir conhecimento sobre um domínio específico, permitindo a identificação de padrões relevantes e a tomada de decisões automatizadas. Esse processo baseia-se na utilização de dados históricos que representam situações anteriormente observadas e solucionadas com êxito. O aprendizado ocorre de forma automática por meio da aplicação do raciocínio indutivo, um método de inferência lógica que consiste em derivar generalizações ou modelos a partir de um conjunto de exemplos concretos. Por exemplo, ao analisar um conjunto de dados em que variáveis de entrada, que representam atributos do objeto de estudo, estão associadas a variáveis de saída esperadas, correspondentes aos valores atribuídos a esses atributos, o processo de indução permite construir um modelo capaz de capturar os padrões subjacentes presentes nos exemplos fornecidos. Esse modelo pode, então, ser utilizado para realizar previsões ou classificações em novas instâncias, mesmo na ausência de observações prévias sobre esses casos (MONARD; BARANAUSKAS, 2003).

Em AM, a notação utilizada para representar as variáveis preditivas pode variar dependendo do nível de agregação dos dados. Seguindo a convenção utilizada em (JAMES et al., 2023), este trabalho distingue entre a matriz completa de atributos, vetores de variáveis individuais e valores específicos dentro da matriz. A matriz de variáveis preditivas ( $X$ ), com  $X \in \mathbb{R}^{n \times p}$ , contém todas as variáveis preditivas do conjunto de dados, onde  $n$  representa o número de observações (linhas da matriz) e  $p$  representa o número de variáveis preditivas (colunas da matriz). Cada linha da matriz  $X$  representa um vetor de atributos para uma observação específica, enquanto cada coluna representa uma variável preditiva para todas as observações, conforme representado na matriz da equação 3.1.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (3.1)$$

Essa matriz tem dimensão  $n \times p$ , onde cada elemento  $x_{ij}$  representa o valor específico da variável  $j$  para a observação  $i$ . Em geral, utiliza-se  $x_{ij}$  para representar o valor da  $j$ -ésima variável preditiva na  $i$ -ésima observação, onde: o índice  $i$  referencia uma amostra ou observação específica ( $i = 1, 2, \dots, n$ ), enquanto o índice  $j$  referencia uma variável preditiva ( $j = 1, 2, \dots, p$ ).

A variável resposta  $Y$  representa o valor a ser previsto com base nas variáveis preditivas  $X$ . Essa variável é organizada como  $Y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$ , onde cada  $y_i$  representa o valor observado da variável resposta para a  $i$ -ésima observação.

Adicionalmente, para garantir uniformidade e clareza na apresentação dos modelos preditivos ao longo deste trabalho, adotam-se as seguintes convenções de notação unificada:

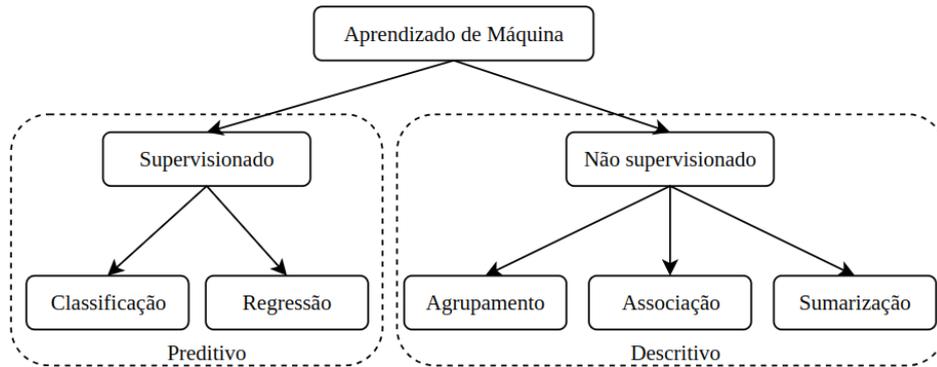
- $x = (x_1, x_2, \dots, x_p) \in \mathbb{R}^p$ : vetor de atributos de uma única observação extraído da matriz  $X$ ;
- $y \in \mathbb{R}$ : valor escalar correspondente à variável resposta da observação;
- $\hat{y} \in \mathbb{R}$ : valor previsto para  $y$  por um modelo de regressão;
- $f(x)$ : modelo preditivo (função de regressão) treinado com base nos dados;

### 3.3.1 Paradigmas de aprendizado

O AM pode ser categorizado de acordo com paradigmas e tarefas associadas. A Figura 7 ilustra a disposição clássica dos tipos de aprendizado.

O paradigma de aprendizado supervisionado é amplamente aplicado em tarefas preditivas, tanto de classificação quanto de regressão. Nesse tipo de aprendizado, o algoritmo é treinado com base em um conjunto de dados de entrada e saída previamente conhecidos, ou seja, rotulados, permitindo que aprenda a mapear os dados de entrada para os resultados esperados. Em tarefas de classificação, os rótulos correspondem a classes ou valores nominais,

Figura 7 – Hierarquia clássica de aprendizado.



Fonte: (FACELI et al., 2021).

enquanto em tarefas de regressão, os rótulos são valores numéricos ordenados. O objetivo é induzir, a partir dos dados rotulados, um modelo preditivo capaz de estimar o valor do atributo alvo de novas instâncias com base nos valores dos atributos preditivos.

Por outro lado, o paradigma de aprendizado não supervisionado é voltado a tarefas descritivas, como agrupamento, associação e sumarização. Ao contrário do aprendizado supervisionado, os algoritmos não supervisionados não utilizam dados rotulados. Em vez disso, exploram os dados de forma autônoma, buscando identificar padrões, agrupamentos naturais, relações e características relevantes. Um exemplo clássico é o agrupamento (*clustering*) de dados, em que o algoritmo descobre subconjuntos de dados semelhantes, sem conhecimento prévio sobre as categorias existentes.

Além desses, a literatura apresenta outros paradigmas, como o aprendizado semissupervisionado e o aprendizado por reforço. O primeiro é aplicado quando apenas parte dos dados está rotulada, combinando técnicas supervisionadas e não supervisionadas para aprimorar a capacidade preditiva do modelo. Já o aprendizado por reforço baseia-se na interação contínua com o ambiente, na qual o modelo recebe recompensas ou penalidades conforme suas decisões. Esse tipo de abordagem é especialmente comum em aplicações de robótica e jogos (FACELI et al., 2021).

O uso de métodos de aprendizado supervisionado para tarefas de regressão justifica-se pela natureza do problema investigado neste estudo, que envolve a predição de valores numéricos contínuos com base em variáveis históricas e observáveis. Em cenários nos quais se busca estimar quantitativamente um fenômeno, como o consumo de recursos, a demanda futura, a capacidade preditiva de sistemas ou medidas associadas a grandezas mensuráveis, os modelos de regressão constituem uma abordagem robusta para identificar padrões e capturar

relações estatísticas presentes nos dados.

Dada a natureza quantitativa da autonomia dos meios operativos, bem como a disponibilidade de um conjunto de dados rotulado e confiável, o uso de técnicas de regressão supervisionada configura-se como uma escolha metodológica adequada. Essa abordagem permite não apenas a geração de previsões com qualidade, mas também a realização de análises críticas acerca dos fatores que influenciam o comportamento da autonomia.

### 3.3.2 Modelos de Regressão

Os métodos de aprendizado supervisionado voltados a tarefas de regressão, abordados nesta seção, são diretamente relevantes para este estudo. Eles estão alinhados aos objetivos da pesquisa, que envolvem a realização de previsões com base em conjuntos de dados que contêm informações históricas e verificáveis sobre o fenômeno analisado. Esses dados servem como rótulos, oferecendo uma estrutura clara para a extração de padrões realizada pelo modelo preditivo.

#### 3.3.2.1 Regressão Linear

A regressão linear é um modelo preditivo utilizado no AM que visa prever uma variável dependente  $y$  com base em um conjunto de variáveis independentes  $x \in \mathbb{R}^p$ , também conhecidas como *features* ou atributos. A ideia central é modelar a relação entre essas variáveis por meio de uma equação linear. A forma geral da regressão linear simples é apresentada na equação 3.2:

$$\hat{y} = w^T x = w_0 + \sum_{j=1}^p w_j x_j \quad (3.2)$$

Onde  $w_0$  representa o intercepto (valor de  $y$  quando todos os atributos são zero), e  $w_j$  os pesos associados a cada variável preditiva. O objetivo do modelo é estimar o vetor de pesos  $w$  de forma a minimizar a diferença entre os valores preditos  $\hat{y}$  e os valores reais  $y$  (MAASS; STOREY, 2021).

#### 3.3.2.2 *K-Nearest Neighbors*

O algoritmo *K-Nearest Neighbors* (KNN) é uma técnica de aprendizado supervisionado que pode ser empregada tanto em classificação quanto em regressão. Baseia-se no princípio

de que exemplos similares tendem a estar próximos uns dos outros no espaço vetorial de características. Para realizar uma predição, o KNN identifica os  $K$  exemplos mais próximos de uma nova instância  $x$  utilizando uma métrica de distância, como a Euclidiana (equação 3.3) ou Manhattan (equação 3.4). Onde  $x_{ij}$  representa o valor da  $j$ -ésima variável preditiva na  $i$ -ésima observação, e  $p$  o número total de atributos preditivos. No contexto de regressão, o valor predito  $\hat{y}$  é obtido pela média ou mediana dos valores alvo dos  $K$  vizinhos mais próximos (MAASS; STOREY, 2021).

$$d(x_1, x_2) = \sqrt{\sum_{j=1}^p (x_{1j} - x_{2j})^2} \quad (3.3)$$

$$d(x_1, x_2) = \sum_{j=1}^p |x_{1j} - x_{2j}| \quad (3.4)$$

### 3.3.2.3 Árvore de Decisão

Métodos simbólicos aplicados a tarefas preditivas de AM proporcionam maior interpretabilidade dos resultados, sendo acessíveis tanto para desenvolvedores quanto para especialistas de domínio. As árvores de decisão são exemplos notáveis dessa abordagem, permitindo representar explicitamente o conhecimento extraído dos dados.

Este trabalho adota o modelo CART (*Classification and Regression Trees*) (BREIMAN et al., 1984), que suporta tanto tarefas de classificação quanto de regressão. A predição para uma variável alvo  $y$  baseia-se na divisão do espaço de entrada  $\mathbb{R}^p$  em regiões retangulares, às quais são atribuídos valores médios como predições.

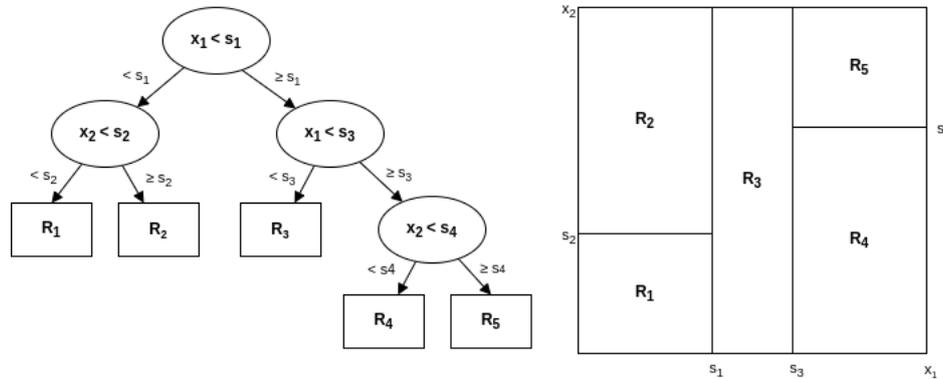
Cada instância de dados pode ser representada por um vetor  $x = (x_1, x_2, \dots, x_p) \in \mathbb{R}^p$ , e o algoritmo procura por divisões do tipo  $\{x_1 < s\}$  e  $\{x_1 \geq s\}$  que minimizem o erro quadrático médio (MSE) ou maximizem o coeficiente de determinação ( $R^2$ ), conforme descrito na Seção 3.3.3 (JAMES et al., 2023).

A estrutura da árvore consiste em um grafo acíclico direcionado, composto por nós de decisão, com testes condicionais, e nós folha que associam valores de predição. A Figura 8 ilustra uma árvore de decisão e as regiões correspondentes no espaço de entrada.

Segundo Morettin e Singer (2023), a construção de uma árvore de decisão para regressão segue os seguintes passos:

1. Escolher uma variável  $x_j$  e um ponto de corte  $s$  que minimize o erro de predição ao

Figura 8 – Uma árvore de decisão e a representação da divisão do espaço de entrada das variáveis.



Fonte: (JAMES et al., 2023).

dividir os dados em  $\{x_j < s\}$  e  $\{x_j \geq s\}$ ;

2. Avaliar todos os pares  $(j, s)$  e selecionar aquele que resulta na melhor divisão;
3. Repetir o processo recursivamente sobre as regiões criadas;
4. Parar o crescimento da árvore com base em critérios como profundidade máxima ou número mínimo de instâncias por nó.

Um aspecto significativo desse método é a sua eficácia em lidar com problemas de natureza não linear, uma vez que se trata de modelos não paramétricos. Modelos não paramétricos são flexíveis em relação à forma da função que relaciona as variáveis de entrada às saídas. Eles não impõem restrições rígidas à natureza da relação e podem se adaptar a padrões complexos e não lineares nos dados (MORETTIN; SINGER, 2023).

### 3.3.2.4 *Random Forest*

O método *Random Forest*, proposto por Breiman (2001), é uma técnica baseada em conjuntos que constrói múltiplas árvores de decisão utilizando amostragem com reposição (*bootstrap*) e média de predições. A predição final  $\hat{y}(x)$  é dada pela média das saídas individuais das  $B$  árvores  $\hat{y}_b(x)$ , conforme a equação 3.5. O modelo é robusto ao *overfitting* e fornece medidas de importância das variáveis (BIAU; SCORNET, 2016).

$$\hat{y}(x) = \frac{1}{B} \sum_{b=1}^B \hat{y}_b(x) \quad (3.5)$$

### 3.3.2.5 Gradient Boosting

O *Gradient Boosting* (GB) (FRIEDMAN, 2001) é uma técnica de AM baseada em *ensemble*, cuja construção do modelo ocorre de forma sequencial. O objetivo central é combinar diversas árvores de decisão fracas (*weak learners*) em um modelo mais robusto e eficaz na geração de estimativas.

A ideia principal do algoritmo é ajustar sucessivamente modelos aos resíduos (erros) do modelo anterior. Em cada iteração, uma nova árvore é treinada para corrigir os erros cometidos pelo modelo atual, resultando em uma melhoria incremental da predição. O modelo de GB pode ser representado pela equação 3.6.

$$F_M(x) = F_0(x) + \sum_{m=1}^M \beta_m T_m(x) \quad (3.6)$$

Onde  $F_0(x)$  é a estimativa inicial (por exemplo, a média dos valores da variável alvo),  $T_m(x)$  representa a  $m$ -ésima árvore de decisão treinada,  $\beta_m$  é o peso atribuído à contribuição da árvore  $T_m$ ,  $M$  é o número total de iterações (ou árvores),  $x \in \mathbb{R}^p$  representa uma instância do conjunto de dados.

Diferentemente do modelo *Random Forest*, no qual as árvores são construídas de forma independente e paralela, o GB constrói as árvores em sequência, cada uma treinada para minimizar a função de perda dos resíduos da anterior. Essa abordagem permite que o modelo final apresente melhor qualidade nas estimativas, embora também esteja mais suscetível ao *overfitting* se não forem aplicadas técnicas de regularização adequadas (PINHEIRO; PATETTA, 2021).

### 3.3.2.6 Aprendizado em Comitê (*Ensemble Learning*)

O *Ensemble Learning* combina múltiplos modelos de AM com o objetivo de melhorar a capacidade preditiva em relação a modelos individuais. Essa estratégia é útil para reduzir viés, variância e sobreajuste (FACELI et al., 2021). Na literatura, são amplamente utilizados diversos métodos de aprendizado por comitê aplicados a problemas de regressão, destacando-se principalmente o *Bagging*, o *Boosting*, o *Voting Regression* e o *Stacking* (BREIMAN, 1996; FRIEDMAN, 2001; WOLPERT, 1992).

O *Bagging* (Bootstrap Aggregating) consiste em treinar múltiplos modelos em paralelo a partir de subconjuntos aleatórios dos dados, buscando reduzir a variância do modelo final —

sendo o *Random Forest* seu exemplo mais representativo (BREIMAN, 2001). Já o *Boosting*, como o GB (FRIEDMAN, 2001), constrói modelos de forma sequencial, onde cada novo modelo corrige os erros do anterior, reduzindo o viés e reduzindo os erros de predição. O *Voting Regression* (CHEN; LUC, 2022; PEDREGOSA et al., 2011), por sua vez, combina modelos conceitualmente distintos, como regressão linear, árvores de decisão, KNN, entre outros, realizando uma média — simples ou ponderada — das predições individuais. Por fim, o *Stacking* (WOLPERT, 1992) adota uma estratégia hierárquica, na qual os modelos base produzem predições que são posteriormente utilizadas como entrada para um modelo de nível superior (meta-modelo), que aprende a combinar essas predições de forma otimizada. Essa técnica é especialmente útil quando os modelos base apresentam diferentes padrões de erro e níveis de especialização sobre os dados.

Segundo Mienye e Sun (2022), o sucesso dos métodos de aprendizado por comitê depende criticamente da diversidade entre os modelos base e da combinação eficaz de suas previsões. Nesse sentido, a abordagem *ATOp* é projetada para maximizar essa diversidade tanto pela adoção de algoritmos heterogêneos, quanto por meio da segmentação semântica dos atributos preditivos com base em uma ontologia operacional. Essa segmentação é orientada pelas diferentes *situações* mapeadas na ontologia, permitindo que cada subconjunto de atributos represente uma perspectiva distinta do fenômeno modelado — como aspectos relacionados à logística, suprimentos ou força de trabalho. A combinação dessas múltiplas perspectivas, por meio de modelos especializados em cada situação, contribui para um aprendizado mais completo e complementar, que é posteriormente integrado via meta-aprendizagem no modelo do tipo *Stacking*. A adoção dessa estratégia é destacada por Mienye e Sun (2022) como especialmente promissora em contextos complexos e de significativa heterogeneidade dos atributos, características presentes nos dados utilizados neste trabalho.

### 3.3.3 Avaliação de Modelos de Regressão e Testes Estatísticos

As métricas de avaliação de modelos de AM são utilizadas para quantificar a capacidade preditiva de um modelo em relação a um conjunto de dados. Essas métricas avaliam a capacidade do modelo de realizar predições com baixo erro e boa estabilidade, servindo como guias para ajuste e otimização durante o processo de treinamento (MAASS; STOREY, 2021). Nesse contexto, esta seção destacará as principais métricas empregadas na avaliação de tarefas de regressão.

Ao avaliar a capacidade preditiva de diferentes modelos de regressão, não basta apenas comparar seus valores médios de métricas como  $R^2$  e RMSE. É fundamental aplicar métodos estatísticos que permitam verificar se as diferenças observadas são estatisticamente significativas ou se podem ser atribuídas a variações naturais dos dados.

Nesse sentido, este trabalho adota a aplicação de testes estatísticos não paramétrico para a comparação dos modelos preditivos avaliados. O objetivo é garantir uma análise robusta e confiável das diferenças de capacidade preditiva, fortalecendo a validade dos resultados apresentados.

Dentre os testes estatísticos disponíveis, os testes de *Friedman* (DEMŠAR, 2006) e *Durbin-Conover* (POHLERT, 2014) foram selecionados por serem apropriados para a comparação de três ou mais métodos avaliados sob condições experimentais idênticas. Esse é o caso deste estudo, no qual todos os modelos foram testados utilizando os mesmos *folds* de validação cruzada.

### 3.3.3.1 Coeficiente de Determinação ( $R^2$ )

O coeficiente de determinação, frequentemente denotado como  $R^2$ , é uma métrica estatística que fornece uma medida da proporção da variabilidade no atributo alvo (dependente) que é explicada pelos atributos preditivos (independentes) em um modelo de regressão. O  $R^2$  avalia a qualidade do ajuste do modelo aos dados observados, conforme equação 3.7.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.7)$$

Onde:

- $n$ : número total de observações;
- $y_i$ : valores reais observados para a  $i$ -ésima observação;
- $\hat{y}_i$ : valores preditos pelo modelo para a  $i$ -ésima observação;
- $\bar{y}$ : média dos valores reais observados.

O  $R^2$  tem valores no intervalo de 0 a 1, quanto mais próximo esse valor estiver de 1, melhor o modelo se ajusta aos dados, denotando que o modelo consegue explicar certo nível de variabilidade na variável dependente. No entanto, o  $R^2$  não indica causalidade nem verifica

a relevância das variáveis independentes, sendo frequentemente utilizado em conjunto com outras métricas de avaliação (JAMES et al., 2013).

### 3.3.3.2 Erro Quadrático Médio (*Mean Squared Error* - MSE)

O Erro Quadrático Médio (*Mean Squared Error* - MSE) é uma métrica que contabiliza a média da soma dos quadrados das diferenças (erros) entre os valores preditos e os valores reais. Dessa forma, o MSE penaliza mais fortemente erros grandes, conforme descrito na equação 3.8 (JAMES et al., 2013).

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.8)$$

Onde:

- $n$ : representa o número total de observações;
- $y_i$ : são os valores reais observados para a  $i$ -ésima observação;
- $\hat{y}_i$ : são os valores preditos pelo modelo para a  $i$ -ésima observação.

### 3.3.3.3 Raiz do Erro Quadrático Médio (*Root Mean Squared Error* - RMSE)

Devido ao fato de o MSE elevar os valores ao quadrado, dificultando a interpretação do erro, derivou-se a métrica Raiz do Erro Quadrático Médio (*Root Mean Squared Error* - RMSE). O RMSE corresponde à raiz quadrada do MSE e fornece uma medida do erro médio na mesma unidade de medida da variável predita, facilitando sua interpretação, conforme descrito na equação 3.9 (CARVALHO; MENEZES; BONIDIA, 2024).

$$RMSE = \sqrt{MSE} \quad (3.9)$$

### 3.3.3.4 Testes de *Friedman* e *Durbin-Conover*

O teste de *Friedman* é um teste estatístico não paramétrico amplamente utilizado para comparar três ou mais métodos avaliados sob as mesmas condições experimentais. Ele é considerado uma alternativa ao teste de Análise de Variância (*Analysis of Variance* - ANOVA) para medidas repetidas, o qual é comumente aplicado em experimentos nos quais os mesmos sujeitos, amostras ou instâncias são submetidos a diferentes tratamentos ou algoritmos. O

teste de *Friedman* mostra-se apropriado especialmente quando os pressupostos de normalidade dos dados exigidos pelo ANOVA não podem ser garantidos, oferecendo assim uma abordagem robusta para a comparação de desempenho em cenários com distribuições não normais.

De acordo com (DEMŠAR, 2006), a aplicação do teste de *Friedman* é indicada em cenários onde:

- Existem três ou mais algoritmos ou abordagens a serem comparados;
- Cada algoritmo é avaliado sob os mesmos conjuntos de dados ou partições (*folds*), garantindo dependência dos resultados;
- Não se pode assumir que os dados seguem uma distribuição normal.

O teste baseia-se no ranqueamento dos métodos dentro de cada bloco experimental (por exemplo, cada *fold* de validação cruzada), atribuindo a posição 1 ao método de melhor capacidade preditiva. Posteriormente, avalia-se se as posições médias dos métodos diferem de forma estatisticamente significativa.

As hipóteses formuladas para o teste são:

- Hipótese nula ( $H_0$ ): não há diferenças estatísticas entre os resultados médios dos métodos avaliados;
- Hipótese alternativa ( $H_1$ ): existe diferença estatisticamente significativa entre os resultados de pelo menos dois métodos.

Neste trabalho, o teste de *Friedman* é empregado para comparar a capacidade preditiva dos modelos de regressão avaliados, considerando as métricas  $R^2$  e RMSE obtidas ao longo dos mesmos *folds* de validação cruzada. Assim, é possível verificar se as diferenças observadas entre os modelos são estatisticamente significativas, fortalecendo as conclusões sobre a superioridade ou equivalência das abordagens propostas. A estatística de teste de *Friedman* segue aproximadamente uma distribuição qui-quadrado, sendo o  $p$ -valor (*p-value*) utilizado para a tomada de decisão. Um valor de  $p \leq 0,05$  indica a rejeição da hipótese nula, sugerindo diferença significativa entre os métodos.

Uma vez realizado o teste de Friedman e identificada uma diferença estatisticamente significativa entre os modelos avaliados, torna-se necessária a aplicação de um teste post hoc,

com o objetivo de determinar quais comparações específicas entre os modelos apresentam diferenças significativas. Neste trabalho, optou-se pela utilização do teste de *Durbin-Conover* (POHLERT, 2014), que realiza comparações pareadas entre todos os métodos avaliados. O teste de *Durbin-Conover* é particularmente apropriado para esse contexto, pois apresenta maior poder estatístico em comparação a abordagens tradicionais, como o teste de *Nemenyi* (DEMŠAR, 2006), especialmente quando aplicado a conjuntos de dados de tamanho reduzido. Conforme discutido em (DEMŠAR, 2006), a realização de testes *post-hoc* apropriados é essencial após o teste de *Friedman* para assegurar a validade estatística da análise comparativa entre modelos de AM.

## 4 TRABALHOS RELACIONADOS

Neste capítulo, serão discutidos os principais trabalhos relacionados à combinação entre MC e AM, destacando, ainda, as contribuições para temas como análise de processos de ciência de dados, similaridade semântica de representações de dados e qualidade dos dados. A seguir, são analisados os estudos que fundamentam e se relacionam com o desenvolvimento da abordagem *ATOp*, proposta nesta dissertação.

### 4.1 ANÁLISE DE PROCESSOS DE CIÊNCIA DE DADOS

A interação entre MC e AM emergiu como um tópico de pesquisa em expansão (BORK; ALI; ROELEN, 2023). Vários estudos, inclusive, investigam abordagens a partir de uma perspectiva especulativa, examinando suas potencialidades e limitações.

Nesse contexto, destacam-se as pesquisas de Lukyanenko, Parsons e Storey (2018), Lukyanenko et al. (2019), que discutem como a MC pode ser integrada ao Processo Padrão Inter-Indústrias para Mineração de Dados (*Cross Industry Standard Process for Data Mining - CRISP-DM*) (WIRTH; HIPPEL, 2000), para aprimorar a execução e aplicabilidade dos algoritmos de AM. O CRISP-DM foi desenvolvido para Mineração de Dados, mas é largamente aplicado no AM.

Nesses trabalhos, aliás, os autores defendem que o uso de MC apoia a interpretação dos modelos e a incorporação de conhecimento de domínio. As principais contribuições da MC para o AM apontadas pelos autores são resumidas na Tabela 1.

A abordagem *ATOp*, desenvolvida no contexto deste trabalho, está alinhada à incorporação de conhecimento de domínio de maneira sistemática. Essa abordagem utiliza a ODCM para explicitar informações semânticas do domínio, que são posteriormente utilizadas no processo de desenvolvimento de modelos preditivos. Essa utilização permite uma maior contextualização e refinamento das análises realizadas.

Nesse mesmo contexto, o trabalho (SOTHILINGAM; YU; SENDEROVICH, 2019) propõe a utilização de MC para estruturar e aprimorar a integração de AM em ambientes empresariais. Os autores argumentam que os métodos tradicionais de análise de sistemas não são suficientes para lidar com as particularidades dos processos de AM, como a necessidade de iteração contínua, a colaboração entre equipes multidisciplinares e a garantia da

Tabela 1 – Contribuições da MC para o AM, com base nas fases do CRISP-DM.

Fase	Contribuições
Entendimento do Negócio	Representação, simplificação, abstração e identificação dos conceitos do domínio
Entendimento dos Dados	Avaliação da qualidade dos dados
Preparação dos Dados	Seleção, limpeza e transformação dos dados
Modelagem	Representação de modelos de domínio
Avaliação	Transparência, compreensibilidade e entendimento do modelo
Implantação	Documentação dos objetivos e metas do projeto

Fonte: Adaptado de (LUKYANENKO; PARSONS; STOREY, 2018).

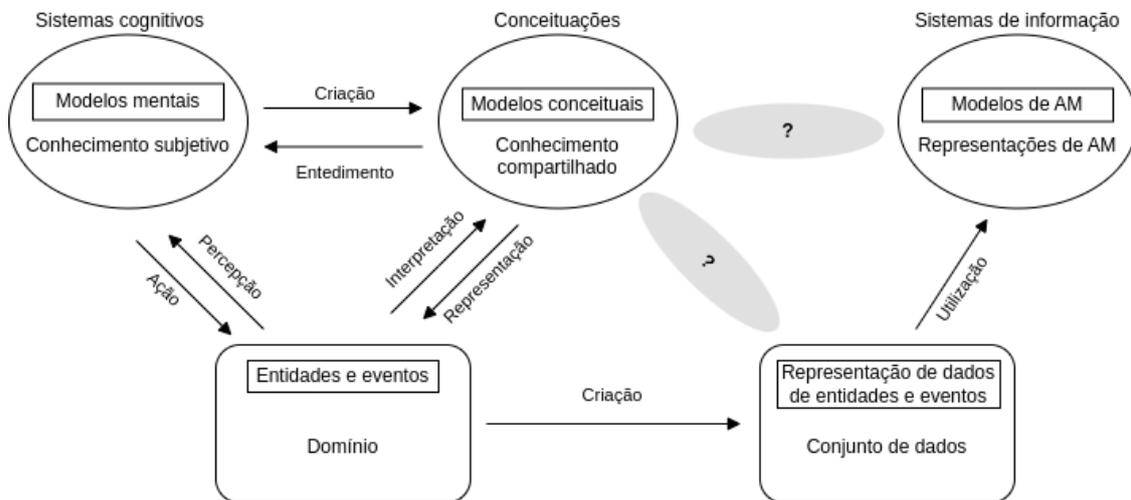
qualidade dos modelos ao longo do tempo. Para isso, exploram técnicas de modelagem, como *Business Process Modeling Notation* (BPMN) e *Data Flow Diagrams* (DFD), identificando lacunas nessas abordagens, especialmente no que diz respeito à captura de dependências temporais, *feedbacks* iterativos e interação entre múltiplos papéis no fluxo de trabalho de AM. O trabalho sugere a necessidade de um novo formalismo que integre melhor esses aspectos, permitindo um desenvolvimento e implantação mais sistemáticos de ML dentro das organizações.

A proposta dos autores se alinha com a abordagem *ATOp*, proposta nesta pesquisa, pois ambos os trabalhos buscam formalizar e estruturar melhor os processos de AM para aumentar sua eficiência e maturidade. No caso da abordagem *ATOp*, é empregada uma MC orientada por ontologias (especificamente a UFO), para segmentar semanticamente conjuntos de dados de treinamento, melhorando a qualidade dos modelos de AM. Enquanto (SOTHILINGAM; YU; SENDEROVICH, 2019) foca na organização do fluxo de desenvolvimento e implantação de AM. A abordagem *ATOp* complementa a estratégia empregada por (SOTHILINGAM; YU; SENDEROVICH, 2019), ao estruturar a representação dos dados utilizados no treinamento dos modelos, permitindo que diferentes situações ou contextos sejam explicitamente representados e diferenciados.

No estudo (MAASS; STOREY, 2021), os autores ampliam as possibilidades de interação entre MC e AM. Eles investigam a combinação mútua dessas duas áreas de pesquisa, identificando lacunas na concepção de modelos que integrem ambas de forma coesa. A Figura

9 ilustra o processo de evolução do modelo mental para os modelos conceitual e de AM proposto pelos autores.

Figura 9 – Relacionamentos entre modelos mentais, conceituais e de AM.



Fonte: (MAASS; STOREY, 2021).

De fato, a deficiência de métodos e técnicas decorre da ausência de definições que caracterizem as relações entre a MC e os modelos de AM. Essas definições são consideradas essenciais para a unificação dessas duas abordagens.

Além de explorar conceitualmente esse tema, Maass e Storey (2021) apresentam um *framework* para mediar a sequência de atividades envolvidas nos processos de AM, que incorpora o uso de ontologias, demonstrando uma preocupação com a melhoria da qualidade dos dados coletados. Nesse *framework*, os autores também exploram a aplicação de restrições sobre os dados proporcionada pelo MC, a partir da análise do domínio, e apresentam resultados positivos em relação à capacidade preditiva de algoritmos de AM apoiados por essa abordagem.

Com um viés fortemente acoplado ao trabalho (MAASS; STOREY, 2021), a abordagem *ATOp* faz uso da ODCM para organizar os dados relacionados ao consumo de recursos logísticos em plataformas navais militares. Porém, a abordagem *ATOp* propõe um aplicação real em um contexto restrito de uso militar, onde são fornecidos predições de maior qualidade sobre a autonomia dessas plataformas.

No trabalho (AMARAL; BAIÃO; GUIZZARDI, 2021), os autores exploram os múltiplos benefícios das ontologias de fundamentação e da ODCM no campo da Mineração de Dados. Eles destacam o uso da OntoUML para resolver conflitos semânticos e melhorar a qualidade dos dados para a mineração. Os autores realizam uma análise abrangente acerca da desco-

berta de conhecimento em bases de dados, exploram as vantagens da MC em projetos de AM e oferecem uma perspectiva sobre a aplicação dessa abordagem em todas as fases do ciclo de vida de um projeto em Ciência de Dados. Essas fases compreendem o Entendimento do Problema, Pré-processamento, Mineração de Dados e Pós-processamento.

No referido estudo, os autores defendem a utilização de linguagens de modelagem orientadas por ontologias fundamentadas em UFO. Com relação à Mineração de dados, eles discutem os benefícios do emprego das ontologias de fundamentação para algumas técnicas de mineração, como a Classificação. A Tabela 2 lista um resumo desses benefícios.

Tabela 2 – Benefícios do uso de ontologias de fundamentação nas etapas do ciclo de vida de projetos de Ciência de Dados.

<b>Etapa</b>	<b>Benefícios</b>
Entendimento do Problema	Transparência semântica Gerência para domínios complexos Os modelos de dados são mais uniformes
Pré-processamento	Interoperabilidade semântica Compromissos ontológicos mais claros
Classificação	Orientação sistemática no desenvolvimento de classificadores Aumento da precisão da classificação
Pós-processamento	Melhor compreensão dos padrões descobertos Orientação sistemática na validação dos padrões descobertos A compreensão entre os padrões descobertos e o domínio é baseada em metapropriedades ontológicas

Fonte: (AMARAL; BAIÃO; GUIZZARDI, 2021).

A estratégia de propor soluções para a resolução de conflitos semânticos nos dados é diretamente aplicada na abordagem *ATOp*, na qual a OntoUML foi utilizada para modelar o domínio militar, oferecendo uma base sólida para a integração dos dados e para a geração de um modelo preditivo mais eficiente. A aplicação destas metodologias no domínio militar reforça a relevância das ontologias na geração de soluções tecnológicas de melhor capacidade preditiva e interoperabilidade.

## 4.2 SIMILARIDADE SEMÂNTICA E REPRESENTAÇÕES DE DADOS

As questões relacionadas à semântica e à análise de dados têm levado pesquisadores da área de bioinformática a revisarem e discutirem métodos de similaridade semântica e estratégias de representação de dados em aplicações de Ciência de Dados, com destaque para o uso de ontologias. Nesse contexto, portanto, os esforços de pesquisa têm avançado no sentido de definir métricas de similaridade capazes de apoiar o AM.

No trabalho (KULMANOV et al., 2020), os autores propõem uma abordagem que integra diversos métodos para o cálculo da similaridade semântica entre as entidades pertencentes à *Gene Ontology* (GO)<sup>1</sup>. Segundo os autores, a similaridade semântica é amplamente aplicada à GO para medir a semelhança entre termos GO e entre genes anotados com termos GO.

Na bioinformática, a GO vale-se de um conjunto diversificado de medidas de similaridade semântica para suas aplicações. Essas medidas podem ser categorizadas em duas principais abordagens: as baseadas em informação, como as medidas de Resnik (1995), de Lin et al. (1998) e de Jiang e Conrath (1997), que quantificam a similaridade entre termos da GO com base em sua distribuição estatística em um *corpus*; e as baseadas em grafos, como as medidas de simGIC e de simUI (PESQUITA et al., 2008), que exploram a estrutura hierárquica da GO para calcular a similaridade entre termos. Essas métricas encontram amplo uso em diversas áreas biomédicas, incluindo a recuperação de informações, sendo empregadas na avaliação da qualidade das anotações GO e no aprimoramento da precisão de classificações baseadas nessa ontologia, contribuindo para uma análise adequada dos dados biológicos (PESQUITA et al., 2009).

Além de integrar as medidas de similaridade, Kulmanov et al. (2020) enfatizam a relevância dos *embeddings* de ontologias e sua aplicação em modelos de AM. Os *embeddings* de ontologias consistem em mapeamentos que preservam a estrutura de um grafo gerado a partir de uma ontologia em um espaço vetorial. Estes podem ser gerados por meio de diversos métodos, como DeepWalk (PEROZZI; AL-RFOU; SKIENA, 2014), Node2Vec (GROVER; LESKOVEC, 2016) e TransE (BORDES et al., 2013). A utilização dos *embeddings* de ontologias permite incorporar informações semânticas em modelos de AM, proporcionando um caminho promissor para o aprimoramento de técnicas de agrupamento e classificação.

No contexto da abordagem *ATOp*, proposta nesta dissertação, as ontologias são usadas para definir regras de agrupamento semântico de entidades presentes em uma ontologia

---

<sup>1</sup> <https://geneontology.org>

militar, provendo melhoria na qualidade dos dados, e conseqüentemente, melhorando a predição da autonomia de plataformas militares. De forma semelhante ao estudo conduzido por Kulmanov et al. (2020), as ontologias são utilizadas como um mecanismo para extração de informações compartilhadas dentro de um domínio, com o objetivo de aprimorar a análise de dados.

Ainda no contexto de similaridade semântica, o estudo de Maddalena e Baião (2022) propôs uma métrica de similaridade ontologicamente bem-fundamentada para melhorar o qualidade da técnica de Agrupamento (*Clustering*), amplamente utilizada em métodos não supervisionados de AM. Essa abordagem foi aplicada à *Disease Ontology* (DO)<sup>2</sup>.

A abordagem proposta por Maddalena e Baião (2022) consiste na aplicação de ontologias e estruturas semânticas para aprimorar técnicas de AM, sendo implementada em múltiplas etapas. Inicialmente, os dados são coletados e selecionados com base na semântica e no contexto das informações. Em seguida, as comorbidades presentes no conjunto de dados são manualmente mapeadas para a DO, associando cada comorbidade a uma doença representada na ontologia. Posteriormente, são calculadas as similaridades entre os conjuntos de doenças, utilizando métricas de similaridade de grupo que permitem a comparação entre os conjuntos de termos da DO.

Além disso, na pesquisa (MADDALENA; BAIÃO, 2022), os autores investigaram a aplicação dessa abordagem na etapa de pós-processamento de dados, onde são avaliados os resultados do agrupamento usando métricas de qualidade de *cluster*. Isso inclui a análise de como os agrupamentos resultantes dividiram os subconjuntos de doenças e como as similaridades semânticas impactaram a formação dos *clusters*.

De forma análoga, no presente trabalho, a abordagem *ATOp* emprega uma ontologia para agrupar e interpretar dados complexos, como as entidades representadas em uma ontologia militar. Essa abordagem permite a criação de subconjuntos de dados mais coesos, o que resulta em uma melhoria significativa na capacidade preditiva dos modelos.

### 4.3 QUALIDADE DOS DADOS

Um tema relevante que tem despertado crescente interesse entre os pesquisadores é a aplicação da interação entre MC e AM como estratégia para aprimorar a qualidade dos dados. Tal abordagem, com efeito, é considerada um ponto de partida fundamental no

---

<sup>2</sup> <https://disease-ontology.org>

desenvolvimento de modelos preditivos eficazes, uma vez que a qualidade dos dados utilizados no treinamento influencia diretamente a capacidade do modelo de AM em generalizar padrões (FACELI et al., 2021).

Nesse contexto, portanto, o trabalho de Silva, Mutz e Ruy (2022) investiga como as ontologias e o AM podem ser combinados para criar recursos tecnológicos capazes de promover uma melhoria na qualidade dos dados aplicados à predição da evasão escolar em diferentes instituições de ensino. A abordagem apresentada nesse estudo realiza a integração de dados com base em ontologias, permitindo a padronização das definições de variáveis por meio de um vocabulário comum e controlado. Além disso, torna explícitas e claras as relações semânticas entre variáveis provenientes de diferentes fontes, favorecendo a interpretação e o uso por parte dos usuários. Complementarmente, a abordagem incorpora técnicas AM para identificar padrões nos dados e desenvolver modelos preditivos voltados à análise da evasão escolar.

Embora a estratégia apresentada por Silva, Mutz e Ruy (2022) esteja em consonância com o escopo da abordagem *ATOp*, os autores se concentram especificamente na etapa de pré-processamento, propondo um mapeamento dos dados para uma ontologia. Já na abordagem *ATOp*, é estabelecida uma base para a aplicação similar de ontologias no domínio militar, onde o uso de ODCM possibilita a extração de conhecimento do domínio para aprimoramento do modelo preditivo aplicado a uma tarefa de regressão.

De acordo com Coste, Li e Khedri (2023), foi desenvolvido um modelo híbrido que combina ontologias semânticas com AM para a previsão da ruptura de gelo em rios. A abordagem consiste na construção da *Ice Season Ontology*, uma ontologia que estrutura e organiza variáveis relacionadas a eventos sazonais do gelo, como datas de congelamento, fluxos e níveis de água. A *Ice Season Ontology* foi elaborada seguindo os princípios metodológicos da abordagem *Methontology* (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997), sendo sua implementação realizada por meio da linguagem OWL<sup>3</sup>. Utilizando técnicas de análise de grafos, os autores identificam as variáveis mais centrais e eliminam aquelas com menor influência, refinando assim a entrada dos modelos de AM. Após a aplicação desse processo, cinco algoritmos foram testados, sendo o *Random Forest* o mais preciso. O estudo demonstra que a integração de ontologias na seleção de variáveis pode aumentar a eficiência de modelos preditivos e facilitar a interpretação dos resultados para tomadores de decisão.

---

<sup>3</sup> <https://www.w3.org/OWL/>

A pesquisa (COSTE; LI; KHEDRI, 2023) tem uma relação direta com a abordagem *ATOp*, pois ambos exploram o uso de ontologias para melhorar a capacidade preditiva de modelos de AM. Enquanto (COSTE; LI; KHEDRI, 2023) utiliza uma ontologia para selecionar e estruturar variáveis preditivas, a *ATOp* emprega uma estratégia similar ao utilizar agrupamentos de entidades derivados de ontologias para promover a separação semântica dos conjuntos de treinamento. Além disso, o estudo de (COSTE; LI; KHEDRI, 2023) emprega técnicas de *ensemble* para tarefas de regressão, alinhando-se a um dos principais enfoques deste trabalho. Esse resultado reforça o potencial dos modelos híbridos na aprimoramento do poder de predição do AM, o que também é demonstrado na abordagem *ATOp*.

#### 4.4 CONSIDERAÇÕES FINAIS

Após a análise dos trabalhos apresentados neste capítulo, portanto, constatou-se que a combinação entre MC e AM ainda enfrenta lacunas não totalmente solucionadas, evidenciando a escassez de estudos nessa área. Ademais, a utilização de ODCM como suporte aos processos de AM continua sendo uma questão pouco explorada. A Tabela 3 apresenta uma comparação dos principais temas abordados nos estudos correlatos, evidenciando que esta dissertação trata de questões que também foram objeto de investigação por outros pesquisadores.

Os estudos analisados neste capítulo evidenciam a crescente relevância da combinação entre ontologias e AM em diversos domínios de aplicação. A abordagem *ATOp*, desenvolvida nesta dissertação, configura uma aplicação concreta dessa interação no contexto militar, apresentando resultados que demonstram um aumento significativo na qualidade das predições da autonomia de plataformas navais militares. Dessa forma, enfim, este trabalho se posiciona na fronteira de uma linha de pesquisa emergente, que explora o potencial das ontologias como instrumentos para aprimorar a eficácia das técnicas de AM.

Embora uma comparação direta entre as pesquisas analisadas não seja plenamente viável, este estudo se distingue dos demais ao apresentar características singulares tanto no recorte do universo discursivo quanto na abordagem metodológica adotada. Em contraste com os trabalhos revisados, propõe-se aqui a combinação de informações capturadas e estruturadas por meio das metacategorias da UFO com os dados utilizados no modelo de AM. Tais informações, que dificilmente seriam inferidas de forma isolada por um modelo de AM, conferem à solução proposta boa adaptabilidade e abrangência conceitual.

Tabela 3 – Comparação entre os trabalhos relacionados e a abordagem *ATOp*.

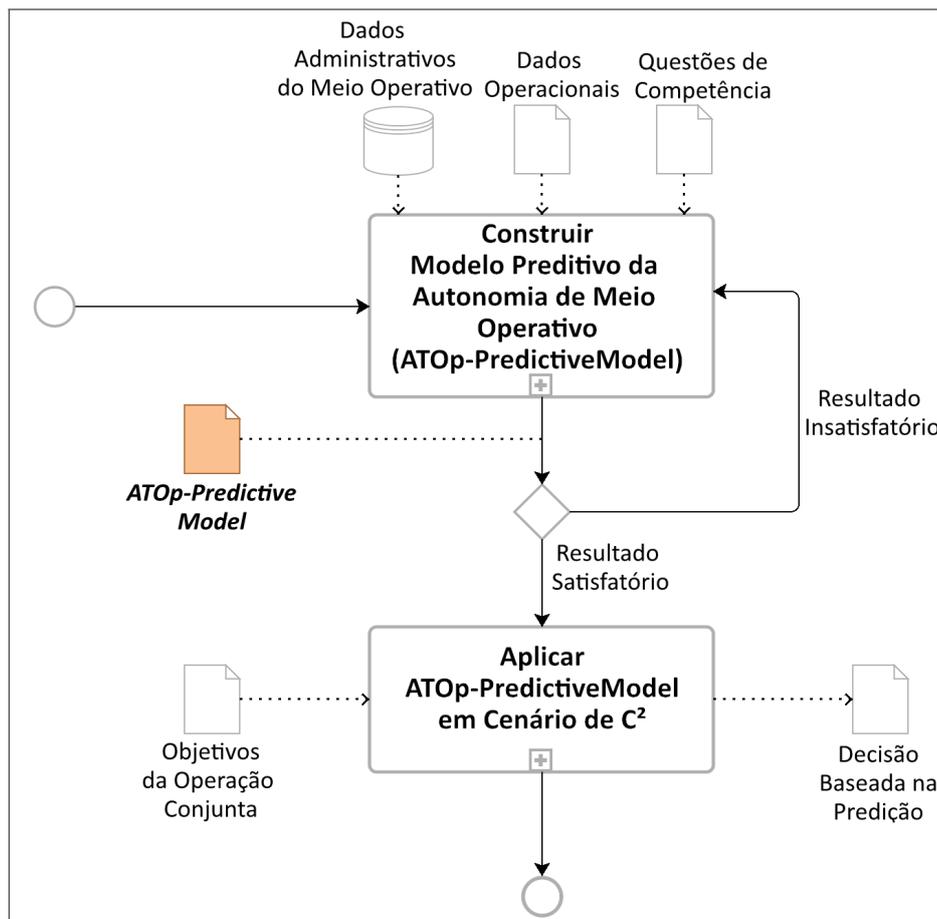
Trabalho	Domínio de Aplicação	Ontologia de Fundamentação	AM	Estratégia MC+AM
Lukyanenko, Parsons e Storey (2018), Lukyanenko et al. (2019)	Multidomínio	Não aplicável	Multimétodo	<i>Framework</i> baseado em BPMN, ER e UML
Sothilingam, Yu e Senderovich (2019)	Multidomínio	Não aplicável	Multimétodo	<i>Framework</i> baseado em BPMN e DFD
Kulmanov et al. (2020)	Saúde	BFO	Modelos de <i>Clustering</i>	<i>Embeddings</i> de ontologias
Maass e Storey (2021)	Multidomínio	UFO	Multimétodo	<i>Framework</i> baseado em BPMN, ER, UML e UFO
Amaral, Baião e Guizzardi (2021)	Multidomínio	UFO	Multimétodo	Padrões de <i>design</i> de ontologias
Maddalena e Baião (2022)	Saúde	BFO	Modelos de <i>Clustering</i>	Métrica de similaridade semântica de grupos de termos da ontologia
Silva, Mutz e Ruy (2022)	Educação	UFO	Modelos de Classificação	Mapeamento de dados para ontologias
Coste, Li e Khedri (2023)	Geociência	Não aplicável	Modelos de Regressão	Medidas de grafos para refinar a seleção de atributos
Abordagem <i>ATOp</i>	Militar	UFO	Modelos de Regressão	Agrupamento de entidades da ontologia baseado em metacategorias da UFO

## 5 ABORDAGEM ATOP

Este capítulo apresenta a abordagem *ATOp*, cuja finalidade é estruturar e integrar atividades que promovam interações sinérgicas entre ontologias bem fundamentadas e AM. O objetivo central da abordagem é aprimorar o processo de treinamento de modelos preditivos, assegurando maior qualidade das predições por meio da exploração de um modelo conceitual semanticamente enriquecido.

A abordagem *ATOp* foi estruturada em dois macroprocessos principais, de modo a representar a transição entre o desenvolvimento do modelo preditivo e sua subsequente aplicação. Essa organização visa proporcionar maior clareza na execução das etapas, assegurando uma visão sistemática do fluxo de atividades. A estrutura proposta está ilustrada na Figura 10.

Figura 10 – Macroprocessos da abordagem *ATOp*.



A divisão da abordagem *ATOp* nesses dois macroprocessos reflete a necessidade de uma estrutura clara e bem definida, garantindo que a modelagem semântica, a integração de dados e a construção do modelo preditivo estejam devidamente conectadas à sua aplicação prática.

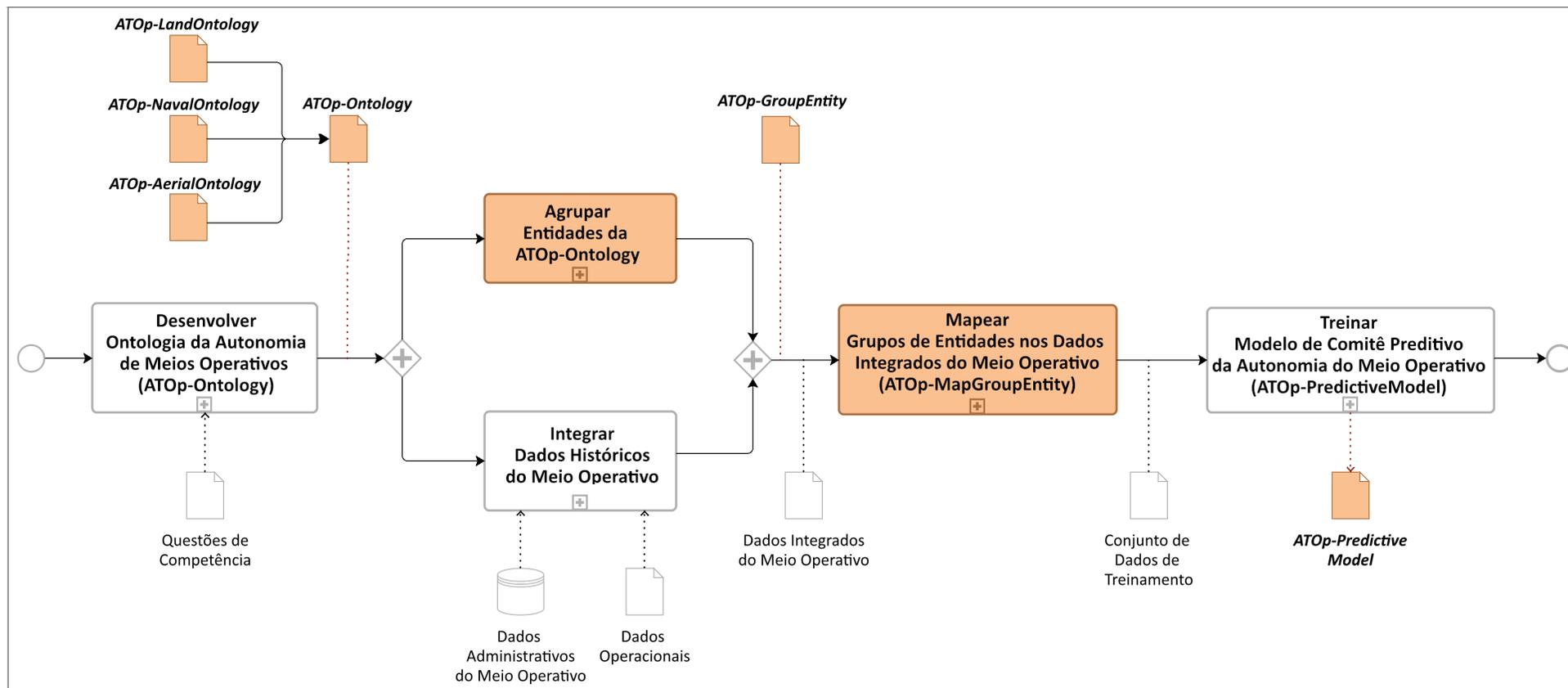
Importante destacar que, caso o modelo desenvolvido não atinja um erro dentro de limites aceitáveis, um resultado satisfatório, o processo retorna ao início do ciclo de desenvolvimento, permitindo ajustes e refinamentos necessários. Essa separação em macroprocessos facilita a implementação da abordagem e torna seu fluxo mais compreensível, favorecendo sua adoção em contextos operacionais complexos.

## 5.1 MACROPROCESSO 1: CONSTRUIR MODELO PREDITIVO DA AUTONOMIA DE MEIO OPERATIVO (*ATOP-PREDICTIVEMODEL*)

Esse macroprocesso abrange as atividades necessárias para estruturar semanticamente os dados e construir o modelo preditivo capaz de estimar a autonomia dos meios operativos. A Figura 11 ilustra os processos atinentes a este macroprocesso.

1. **Desenvolver Ontologia da Autonomia de Meios Operativos (*ATOp-Ontology*)**: definir um modelo conceitual baseado na UFO e modelado em *OntoUML*, com o objetivo de capturar as relações semânticas do domínio (Seção 5.1.1).
2. **Integrar Dados Históricos do Meio Operativo**: coletar e pré-processar os dados estruturados e não estruturados, oriundos, respectivamente, de fontes administrativas e operacionais, assegurando sua qualidade e coerência (Seção 5.1.2).
3. **Agrupar Entidades da *ATOp-Ontology***: organizar os grupos de entidades com base em critérios fundamentados nos conceitos ontológicos definidos pelas metacategorias da *OntoUML*, permitindo uma organização semântica dos dados (Seção 5.1.3).
4. **Mapear Grupos de Entidades nos Dados Integrados do Meio Operativo (*ATOp-MapGroupEntity*)**: estabelecer a associação entre os atributos presentes nos dados integrados e os grupos de entidades da ontologia, viabilizando a organização dos dados em subconjuntos disjuntos, caracterizados pelas especificidades de cada grupo (Seção 5.1.4).
5. **Treinar Modelo de Comitê Preditivo da Autonomia do Meio Operativo (*ATOp-PredictiveModel*)**: aplicar técnicas de AM, utilizando *ensemble learning*, para a construção do modelo preditivo baseado na separação semântica dos dados, incluindo treinamento, teste, validação e avaliação de métricas para o modelo preditivo (Seção 5.1.5).

Figura 11 – Fases que estruturam o Macroprocesso 1 da abordagem *ATOp*.



### 5.1.1 Desenvolver Ontologia da Autonomia de Meios Operativos (*ATOp-Ontology*)

A primeira etapa da abordagem consiste na definição da *ATOp-Ontology*, responsável por estruturar semanticamente os conceitos relacionados à autonomia de meios operativos em cenário de operações militares. Inicialmente, realiza-se uma verificação quanto à existência de ontologias já desenvolvidas que tratem do mesmo domínio. Caso uma ontologia existente seja identificada, ela poderá ser reutilizada, total ou parcialmente, desde que atenda a certos requisitos conceituais e estruturais compatíveis com os objetivos da *ATOp*.

Para ser considerada adequada ao emprego na abordagem *ATOp*, a ontologia deve estar fundamentada UFO, a fim de assegurar correção ontológica e rigor semântico. Deve também representar explicitamente eventos complexos, definidos como construções compostas por múltiplas ocorrências encadeadas, com base na metacategoria *Event*.

Além disso, exige-se a modelagem de situações que permitam a explicitação de relações causais, fundamentadas na metacategoria *Situation*, pois tais relações são essenciais para a análise de contexto e a inferência de padrões logísticos operacionais. A ontologia deve ainda ser capaz de representar entidades do domínio, como objetos, agentes e papéis, de forma contextualizada e alinhada aos compromissos ontológicos do domínio operacional. Finalmente, é fundamental que apresente coerência com os dados empíricos disponíveis, especialmente aqueles provenientes dos sistemas SINGRA e RFC, assegurando sua aplicabilidade prática e integração ao *pipeline* preditivo proposto.

A Figura 12 ilustra a estrutura ontológica exigida para a *ATOp-Ontology*, baseada em um padrão ontológico que articula eventos complexos e situações causais, conforme os princípios definidos pela UFO. A existência desse tipo de construção na ontologia é considerada um requisito essencial, pois possibilita uma adequada representação dos encadeamentos lógicos e temporais entre ocorrências, suas condições de ativação e seus efeitos, conferindo maior expressividade semântica à modelagem e melhor alinhamento com o propósito preditivo da abordagem.

A fundamentação da *ATOp-Ontology* nas metacategorias *Event* e *Situation*, conforme definidas pela UFO, é essencial para capturar a dinâmica do domínio operacional. O metacategoria *Situation* representa uma configuração específica da realidade que pode ser compreendida como um todo coerente, relacionando entidades que coexistem ou interagem em determinado contexto. Trata-se de um recorte semântico do mundo que ocorre em um



toUML, desempenham um papel crucial na representação e organização do conhecimento do domínio em questão, pois permitem a modelagem semântica adequada para identificação de situações pertinentes aos objetivos deste estudo. Essa abordagem facilita não apenas a compreensão profunda do domínio modelado, mas também a validação conceitual, promovendo maior alinhamento entre os elementos ontológicos e a realidade observada (BERNASCONI et al., 2022).

A metodologia SABiO, os princípios e as metacategorias da UFO formam um conjunto integrado de diretrizes para o desenvolvimento de ontologias. Essa coesão proporciona uma base sólida para a criação de modelos ontológicos que são tecnicamente robustos e semanticamente ricos. Assim, é possível atender, de maneira eficaz, às necessidades específicas do domínio de aplicação, garantindo coerência semântica.

### 5.1.2 Integrar Dados Históricos do Meio Operativo

Essa etapa pode ser realizada em paralelo à fase de Agrupar Entidades da *ATOp-Ontology*. Ela consiste na coleta e integração de dados administrativos de consumo, que são geralmente estruturados, com informações operacionais não estruturadas. Esses dados são obtidos a partir de sistemas de informação e relatórios técnicos elaborados por profissionais que atuaram nas plataformas analisadas.

O processo inclui atividades centrais da engenharia de dados, como extração, limpeza, padronização, normalização e transformação, visando preparar os dados de forma adequada para o treinamento de modelos de AM. Tais práticas asseguram o alinhamento e a compatibilidade das informações com a ontologia desenvolvida, possibilitando uma integração semântica eficaz entre fontes heterogêneas.

A limpeza dos dados envolve a identificação e a correção de inconsistências, como a remoção de valores ausentes ou anômalos. A padronização e a normalização garantem que as variáveis estejam em escalas comparáveis, reduzindo o impacto de discrepâncias numéricas. Já as transformações, como codificação de variáveis categóricas ou geração de novos atributos, asseguram que os dados estejam alinhados com os requisitos específicos dos algoritmos de AM, contribuindo para a qualidade das predições do modelo final.

### 5.1.3 Agrupar Entidades da *ATOp-Ontology*

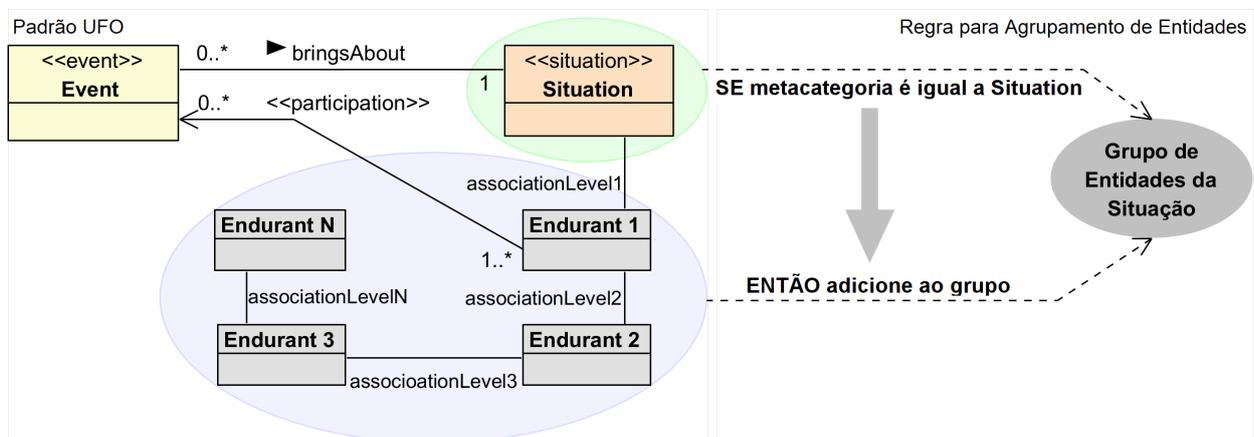
Para a realização desta atividade, é imprescindível a definição e aplicação de regras específicas que orientem o processo de agrupamento das entidades descritas na *ATOp-Ontology*. Essas regras têm como objetivo estabelecer critérios claros e coerentes para identificar correspondências ontológicas entre as entidades, considerando atributos ou relações pertinentes.

Como parte desta atividade, serão estabelecidos os critérios de agrupamento que orientarão a organização das entidades em grupos com base em características compartilhadas. A aplicação desses critérios visa assegurar uma estruturação lógica e semanticamente coerente, de modo a facilitar a análise, interpretação e utilização dos grupos formados no contexto da ontologia.

Com base na teoria de *Situação* definida na UFO (Seção 3.2.3), este estudo explora a possibilidade de utilizar uma situação como representação de um padrão a ser identificado por um modelo de AM no domínio específico analisado. A noção de *Situação* na UFO oferece uma estrutura semântica robusta que permite capturar configurações contextuais e relacionais complexas entre entidades, potencialmente auxiliando no aprimoramento da capacidade de reconhecimento de padrões dos modelos.

Para viabilizar essa abordagem, foi desenvolvida uma regra denominada *ATOp-GroupEntityRule*. Essa regra, fundamentada nos princípios ontológicos da UFO, permite a aplicação direta em ontologias bem fundamentadas, assegurando a coerência semântica e a integração com os preceitos da ontologia. A regra fortalece a fundamentação conceitual dos modelos e pode ser utilizada na abordagem *ATOp*, conforme ilustrado na Figura 13.

Figura 13 – Representação da regra *ATOp-GroupEntityRule*.



A *ATOp-GroupEntityRule* consiste em localizar dentro dos modelos conceituais as en-

tidades representadas com a metacategoria *Situation* da UFO. Após essa identificação, é criado um grupo contendo as entidades que possuem algum tipo de associação com a situação explicitada. Ao final do processo, serão criados grupos com situações que representam configurações do universo de discurso, caracterizadas por suas características particulares. O pseudocódigo da *ATOp-GroupEntityRule* é fornecido pelo *ATOp-GroupEntityAlgorithm*, ilustrado no Algoritmo 1.

O *ATOp-GroupEntityAlgorithm* é um procedimento desenvolvido para percorrer um modelo conceitual e agrupar entidades com base na metacategoria ontológica *Situation* da UFO. Esse agrupamento tem como objetivo identificar configurações semânticas relevantes dentro do universo de discurso, promovendo uma organização lógica e estruturada que reflete as relações contextuais presentes no modelo.

---

**Algoritmo 1:** Percorrer a *ATOp-Ontology* e criar Grupos de Entidades

---

**Entrada:** Modelo  $M$ , com um conjunto de construtos  $C = \{c_1, \dots, c_m\}$   
**Saída:** Conjunto de grupos  $G = \{G_1, \dots, G_n\}$   
 $G \leftarrow \emptyset$   
**Para Cada**  $c_j \in C$  **Faça**  
    **Se**  $Situation(c_j)$  **Então**  
         $G_i \leftarrow \text{criarGrupo}(c_j)$   
        **Para Cada**  $e_k \in C$  **Faça**  
            **Se**  $Endurant(e_k) \wedge isAssociated(e_k, c_j)$  **Então**  
                 $G_i \leftarrow G_i \cup \{e_k\}$   
         $G \leftarrow G \cup \{G_i\}$   
**return**  $G$

---

O objetivo principal do *ATOp-GroupEntityAlgorithm* é criar grupos de entidades a partir das situações definidas na ontologia bem fundamentada, garantindo uma organização estruturada e semanticamente alinhada. Cada grupo gerado reflete as características contextuais específicas de uma *Situation*, incluindo as entidades do tipo *Endurant* que estejam associadas a ela, seja de forma direta ou indireta, conforme detalhado a seguir.

- **Parâmetros de Entrada:**

1. **Modelo  $M$ :** Representa a ontologia bem fundamentada a ser analisada, composto por um conjunto de construtos  $C = \{c_1, \dots, c_m\}$ .

- **Resultado:** Conjunto de grupos de entidades  $G$ , composto por subconjuntos  $G_i$ , em que cada grupo  $G_i$  está vinculado a um construto classificado ontologicamente como uma *Situation* e agrega todas as entidades do tipo *Endurant* que estejam associadas

a essa *Situation*, a qual representa uma condição operacional específica da plataforma militar no contexto de uma operação.

• **Etapas do Algoritmo:**

1. **Inicialização do Conjunto de Grupos ( $G$ ):** O algoritmo inicia com um conjunto vazio  $G$ , que será gradualmente preenchido à medida que as *Situations* forem identificadas e processadas.
2. **Identificação de Situações no Modelo:** Para cada construto  $c_j \in C$ , o algoritmo verifica se ele satisfaz o predicado  $Situation(c_j)$ , ou seja, se é classificado ontologicamente como uma *Situation*.
3. **Criação de um Novo Grupo para Cada *Situation*:** Quando uma *Situation* é identificada, um novo grupo  $G_i$  é criado com base em  $c_j$ . Esse grupo representará as entidades relacionadas à situação em questão.
4. **Identificação e Adição de *Endurants* Associados:** Para cada construto  $e_k \in C$  que satisfaça os predicados  $Endurant(e_k)$  e  $isAssociated(e_k, c_j)$ , ou seja, que seja um *Endurant* e esteja direta ou indiretamente associado à *Situation*  $c_j$ , o algoritmo adiciona  $e_k$  ao grupo  $G_i$ . Essa associação pode ocorrer por meio de propriedades, relacionamentos ou cadeias de dependências semânticas no modelo.
5. **Atualização do Conjunto de Grupos ( $G$ ):** Após a construção de cada grupo  $G_i$ , este é adicionado ao conjunto global  $G$ , que agrega todos os agrupamentos realizados.
6. **Retorno do Resultado:** Ao final da execução, o algoritmo retorna o conjunto  $G$ , composto pelos grupos de entidades estruturados com base nas *Situações* da ontologia bem fundamentada.

Neste contexto, considera-se que uma entidade do tipo *Endurant* está associada a uma *Situation* não apenas quando há uma relação direta entre elas (como uma propriedade ou associação imediata), mas também quando essa associação ocorre de forma indireta, por meio de uma cadeia de relacionamentos intermediários. Essas associações podem incluir relações como formal, material, participação, mediação, agregação, composição ou qualquer outra ligação semântica que conecte a entidade à *Situation*, mesmo que em níveis de profundidade arbitrários ( $1, 2, 3, \dots, N \in \mathbb{N}$ ). Essa abordagem assegura uma coleta abrangente

e semanticamente rica das entidades relevantes em torno de cada situação identificada no modelo.

O *ATOp-GroupEntityAlgorithm* apresenta como benefício principal sua estreita relação com a organização semântica e estrutural dos modelos conceituais, permitindo o agrupamento eficiente de entidades e facilitando análises no domínio de estudo. Ademais, sua abordagem está fundamentada nos princípios da UFO, garantindo coerência ontológica e suporte ao raciocínio baseado em situações. Os grupos gerados pelo algoritmo são particularmente relevantes para o AM, uma vez que oferecem dados estruturados e semanticamente alinhados ao contexto do domínio.

#### 5.1.4 Mapear Grupos de Entidades nos Dados Integrados do Meio Operativo (*ATOp-MapGroupEntity*)

Esta atividade tem como objetivo associar os grupos de entidades identificados na *ATOp-Ontology* aos dados integrados do meio operativo, assegurando que a segmentação semântica do modelo conceitual seja refletida no conjunto de treinamento do modelo preditivo. Esse mapeamento é essencial para garantir que as características semânticas extraídas da ontologia sejam efetivamente utilizadas para aprimorar a predição da autonomia de meios operativos.

Nesse contexto, a correspondência entre as entidades definidas na ontologia e os atributos presentes nos dados operacionais constitui a etapa inicial da atividade *ATOp-MapGroupEntity*, sendo fundamental para assegurar a fidelidade semântica entre o modelo conceitual e o processo analítico subsequente. Essa associação viabiliza a incorporação efetiva do conhecimento formalizado na ontologia, representando situações, recursos e comportamentos do meio operativo à estrutura dos dados utilizados pelos modelos de AM. Ao estabelecer uma ligação direta entre os elementos ontológicos e as associações definidas empiricamente, a abordagem assegura que os padrões extraídos pelos algoritmos estejam semanticamente alinhados aos significados e aos contextos próprios do domínio em estudo.

A estratégia utilizada para essa tarefa está fundamentada na *ATOp-GroupEntityRule*, que já estabeleceu grupos semânticos dentro da *ATOp-Ontology*. Com base nesses grupos, esta etapa busca encontrar correspondências diretas entre os atributos dos dados integrados e os grupos de entidades da ontologia, criando subconjuntos de dados estruturados.

Para viabilizar o mapeamento, é necessário percorrer o conjunto de dados integrados e atribuir cada variável preditiva a um dos grupos previamente definidos pela *ATOp-GroupEn-*

*ntityRule* (Seção 5.1.3). Esse processo pode ser formalmente descrito pelo *ATOp-MapGroup-EntityAlgorithm*, que realiza a correspondência entre os atributos dos dados e os grupos ontológicos, conforme descrito pelo Algoritmo 2.

---

**Algoritmo 2:** Segmentação Semântica de Dados com *ATOp-Ontology*

---

**Entrada:** Conjunto de grupos de entidades  $G = \{G_1, \dots, G_n\}$ ;  
 Conjunto de dados  $D$ , composto por atributos observáveis.  
**Saída:** Coleção de subconjuntos de dados  $\mathcal{S}$ , semanticamente segmentados.  
 $\mathcal{O} \leftarrow \emptyset$ ; // Mapa de ocorrência de atributos por grupo  
 $\mathcal{S}_G \leftarrow \emptyset$ ; // Mapeamento de subconjuntos por grupo  
**Para Cada**  $G_i \in G$  **Faça**  
 |  $S_{G_i} \leftarrow \emptyset$   
 | **Para Cada**  $e \in G_i$  **Faça**  
 | | Identificar o conjunto de atributos  $A_e \subseteq D$  associados à entidade  $e$ ;  
 | | **Para Cada**  $a \in A_e$  **Faça**  
 | | |  $S_{G_i} \leftarrow S_{G_i} \cup \{a\}$ ;  
 | | | **Se**  $a \in \text{domínio de } \mathcal{O}$  **Então**  
 | | | |  $\mathcal{O}[a] \leftarrow \mathcal{O}[a] \cup \{G_i\}$ ;  
 | | | **Senão**  
 | | | |  $\mathcal{O}[a] \leftarrow \{G_i\}$ ;  
 | | **Se**  $S_{G_i} \neq \emptyset$  **Então**  
 | | |  $\mathcal{S}_G[G_i] \leftarrow S_{G_i}$ ;  
 $\mathcal{R} \leftarrow \emptyset$ ; // Grupos que compartilham atributos  
**Para Cada**  $a \in \text{domínio de } \mathcal{O}$  **Faça**  
 |  $G_a \leftarrow \mathcal{O}[a]$   
 | **Se**  $|G_a| \geq 2$  **Então**  
 | | **Se**  $G_a \notin \text{domínio de } \mathcal{R}$  **Então**  
 | | |  $\mathcal{R}[G_a] \leftarrow \emptyset$ ;  
 | | |  $\mathcal{R}[G_a] \leftarrow \mathcal{R}[G_a] \cup \{a\}$ ;  
 $\mathcal{S} \leftarrow \emptyset$ ; // Coleção final de subconjuntos segmentados  
**Para Cada**  $G_a \in \text{domínio de } \mathcal{R}$  **Faça**  
 | Adicionar  $\mathcal{R}[G_a]$  a  $\mathcal{S}$ ;  
 | **Para Cada**  $G_j \in G_a$  **Faça**  
 | |  $\mathcal{S}_G[G_j] \leftarrow \mathcal{S}_G[G_j] \setminus \mathcal{R}[G_a]$ ;  
**Para Cada**  $G_i \in \text{domínio de } \mathcal{S}_G$  **Faça**  
 | **Se**  $\mathcal{S}_G[G_i] \neq \emptyset$  **Então**  
 | | Adicionar  $\mathcal{S}_G[G_i]$  a  $\mathcal{S}$ ;  
**return**  $\mathcal{S}$

---

O principal objetivo do *ATOp-MapGroupEntityAlgorithm* é identificar e agrupar atributos que se relacionam a entidades específicas da *ATOp-Ontology*, de modo a formar subconjuntos de dados semanticamente coesos. Essa estruturação favorece uma organização sistemática dos dados, facilitando sua análise e aplicação nos modelos preditivos. Além de assegurar a segmentação de informações pertinentes, a abordagem contribui para a elimi-

nação de redundâncias e para a promoção da qualidade estrutural dos dados, elementos essenciais à integridade da informação no contexto da MC orientada por ontologias.

O Algoritmo 2 realiza a segmentação semântica de um conjunto de dados integrados  $D$ , com base nos grupos de entidades  $G$  previamente definidos pelo *ATOp-GroupEntityAlgorithm* (vide Seção 5.1.3). A abordagem consiste em identificar, para cada grupo  $G_i$ , os atributos dos dados  $D$  que estejam associados às entidades pertencentes ao grupo, organizando-os em subconjuntos  $S_{G_i}$ .

Um mapeamento auxiliar  $\mathcal{O}$  é construído com o objetivo de registrar a ocorrência de cada atributo e os grupos aos quais ele está vinculado. A partir desse mapeamento, são identificados os atributos compartilhados por dois ou mais grupos. Tais atributos são agrupados em subconjuntos especiais  $\mathcal{R}[G_A]$ , que representam relações semânticas comuns a múltiplos grupos. Para evitar sobreposição semântica, esses atributos compartilhados são removidos dos subconjuntos individuais  $S_{G_i}$  aos quais pertenciam originalmente.

Ao final da execução, a coleção  $\mathcal{S}$  é composta tanto pelos subconjuntos individuais resultantes quanto pelos subconjuntos compartilhados, formando uma estrutura segmentada e semanticamente disjunta, adequada para suportar as etapas posteriores da abordagem *ATOp*. A seguir, detalham-se os elementos e fases do algoritmo:

- **Parâmetros de Entrada:**

1. **Grupos de Entidades  $G$ :** Conjunto  $G = \{G_1, \dots, G_n\}$ , onde cada grupo  $G_i$  representa um conjunto de entidades relacionadas a uma *Situação* ontológica.
2. **Conjunto de Dados  $D$ :** Conjunto de atributos extraídos de fontes empíricas, os quais serão mapeados semanticamente às entidades dos grupos  $G$ .

- **Resultado:** Coleção  $\mathcal{S}$ , composta por subconjuntos semanticamente segmentados dos dados  $D$ , incluindo subconjuntos exclusivos de cada grupo e subconjuntos de atributos compartilhados entre grupos.

- **Fases do Algoritmo:**

1. **Mapeamento de Atributos por Grupo de Entidades:**

- a) **Inicialização:** São inicializadas as estruturas auxiliares  $\mathcal{S}$ ,  $\mathcal{O}$  e  $S_G$ . A primeira armazenará o resultado final; a segunda, o mapeamento de ocorrências de atributos; e a terceira, os subconjuntos temporários por grupo.

b) **Iteração sobre os Grupos:** Para cada grupo  $G_i \in G$ , identificam-se os atributos  $A_e \subseteq D$  associados a cada entidade  $e \in G_i$ . Esses atributos são adicionados ao subconjunto  $S_{G_i}$ , e cada ocorrência é registrada em  $\mathcal{O}$ , associando o atributo aos grupos correspondentes.

## 2. Identificação e Tratamento de Atributos Compartilhados:

a) **Construção de Subconjuntos Compartilhados:** Com base em  $\mathcal{O}$ , o algoritmo identifica os atributos que ocorrem em dois ou mais grupos distintos. Esses atributos são organizados em subconjuntos compartilhados  $\mathcal{R}[G_A]$ , indexados pelo conjunto de grupos que os compartilham.

b) **Remoção de Redundância Semântica:** Os atributos compartilhados são removidos dos subconjuntos individuais  $S_{G_i}$ , garantindo que cada atributo esteja presente apenas uma vez no subconjunto apropriado da coleção  $\mathcal{S}$ , evitando duplicidade semântica.

3. **Construção da Coleção Final:** A coleção  $\mathcal{S}$  é formada pela união dos subconjuntos  $S_{G_i}$  não vazios e dos subconjuntos compartilhados  $\mathcal{R}[G_A]$ , assegurando uma segmentação semântica coerente e disjunta dos dados.

4. **Retorno do Resultado:** O algoritmo retorna  $\mathcal{S}$ , composta por:

- **Subconjuntos Individuais  $S_{G_i}$ :** Atributos específicos de cada grupo  $G_i$ , não compartilhados.
- **Subconjuntos Compartilhados  $\mathcal{R}[G_A]$ :** Atributos comuns a dois ou mais grupos  $G_i \in G$ .

Essa separação lógica e semântica dos atributos garante uma estrutura de dados bem definida, alinhada com os agrupamentos ontológicos, e otimizada para uso nas fases seguintes do macroprocesso da abordagem *ATOp*.

O *ATOp-MapGroupEntityAlgorithm* aprimora a organização dos dados ao aplicar critérios semânticos derivados da ontologia, assegurando que os subconjuntos extraídos reflitam fielmente as relações estabelecidas entre entidades e atributos. A estrutura auxiliar  $\mathcal{O}$  possibilita a rastreabilidade eficiente das associações, enquanto a divisão entre subconjuntos exclusivos e compartilhados evita redundâncias e promove uma segmentação de atributos com características comuns de pertencerem a diversas situações operacionais. Como resul-

tado, obtém-se uma estrutura de dados semanticamente organizada, adequada para análises avançadas e suporte à decisão no domínio da aplicação.

Após a execução do *ATOp-MapGroupEntityAlgorithm*, os dados de treinamento passam por um processo de organização estruturada, no qual são segmentados em subconjuntos. Essa segmentação é baseada nas relações semânticas estabelecidas entre os atributos dos dados e os grupos de entidades identificados na *ATOp-Ontology*, indicando que cada subconjunto possui algo em comum, refletido por uma situação operacional específica do meio operativo ou uma sobreposição de situações.

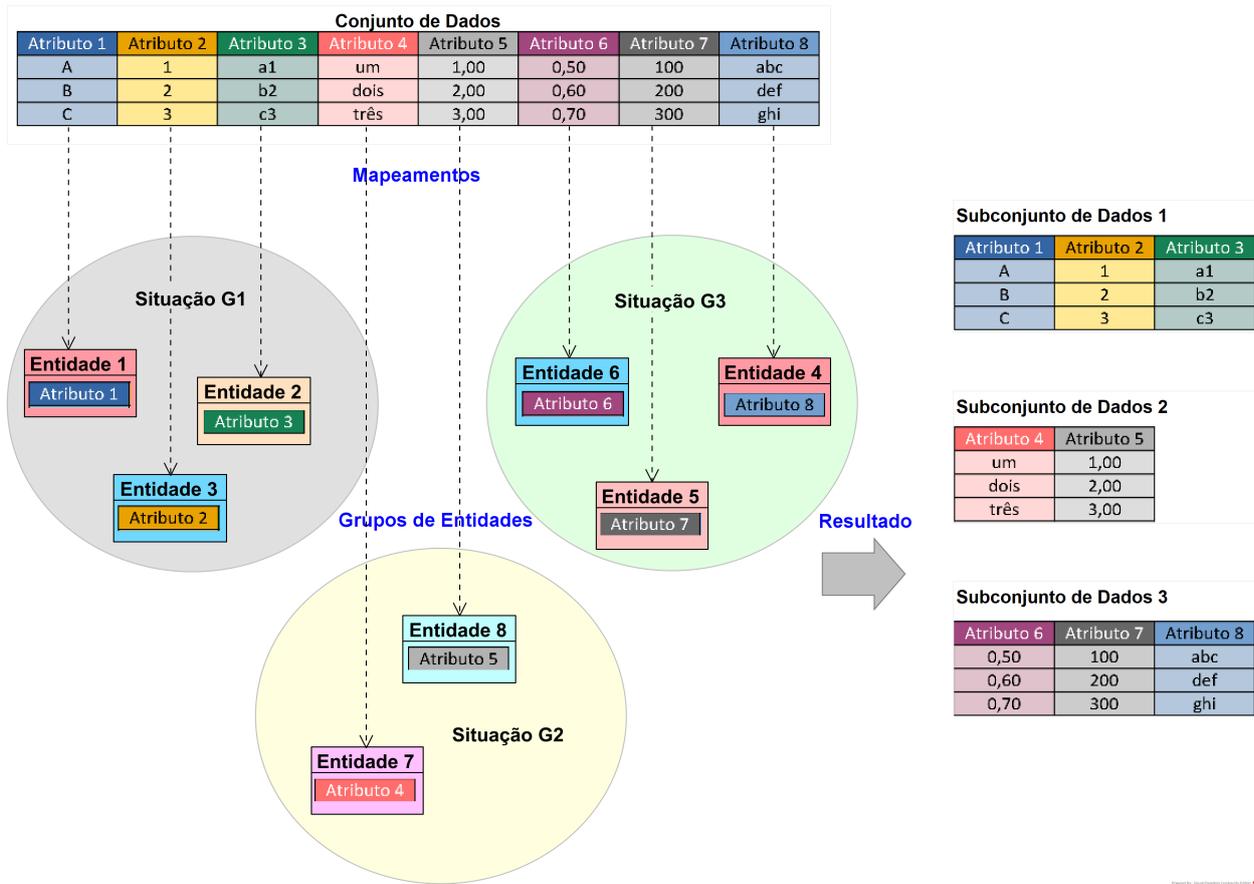
Dessa forma, é realizada uma análise detalhada que associa os atributos dos dados integrados, tal como especificado na atividade da Seção 5.1.2. O resultado desse processo de mapeamento é a criação de subconjuntos de dados semanticamente estruturados, onde cada subconjunto é caracterizado pelas especificidades dos grupos de entidades associados, conforme exemplificado pela Figura 14. O processo nela representado opera conforme descrito no Algoritmo 2 (*ATOp-MapGroupEntityAlgorithm*). Os atributos presentes nos conjuntos de dados são associados a uma mesma situação operacional. Assim, são gerados subconjuntos de dados, onde cada subconjunto é composto por atributos que compartilham características semânticas alinhadas às identificadas no grupo de entidades ao qual foram mapeados.

Quando há atributos correspondentes com sobreposição entre grupos de entidades, são aplicadas as operações da Teoria dos Conjuntos (JECH, 2003) com o objetivo de identificar e estruturar as relações existentes entre os conjuntos de dados. Especificamente, o *ATOp-MapGroupEntityAlgorithm* utiliza as operações de interseção, união e diferença para analisar essas relações.

A aplicação dessas operações permite derivar novos subconjuntos de dados que revelam padrões relevantes, seja evidenciando características compartilhadas (interseção), abrangendo a totalidade das informações combinadas (união), ou destacando aspectos exclusivos de um grupo em relação a outro (diferença). Assim, o uso conjunto dessas operações possibilita uma análise mais completa e refinada das relações entre os dados envolvidos.

Para ilustrar esse processo, o exemplo a seguir ilustra a aplicação do método sobre um agrupamento de entidades organizadas em três grupos. O resultado obtido, que pode ser observado na Figura 15, destaca a estrutura resultante das relações entre os conjuntos analisados.

Considera-se um grupo de entidade  $G = \{G_1, G_2, G_3\}$ . Cada grupo contém as seguintes

Figura 14 – Representação do mapeamento realizado na *ATOp-MapGroupEntity*.

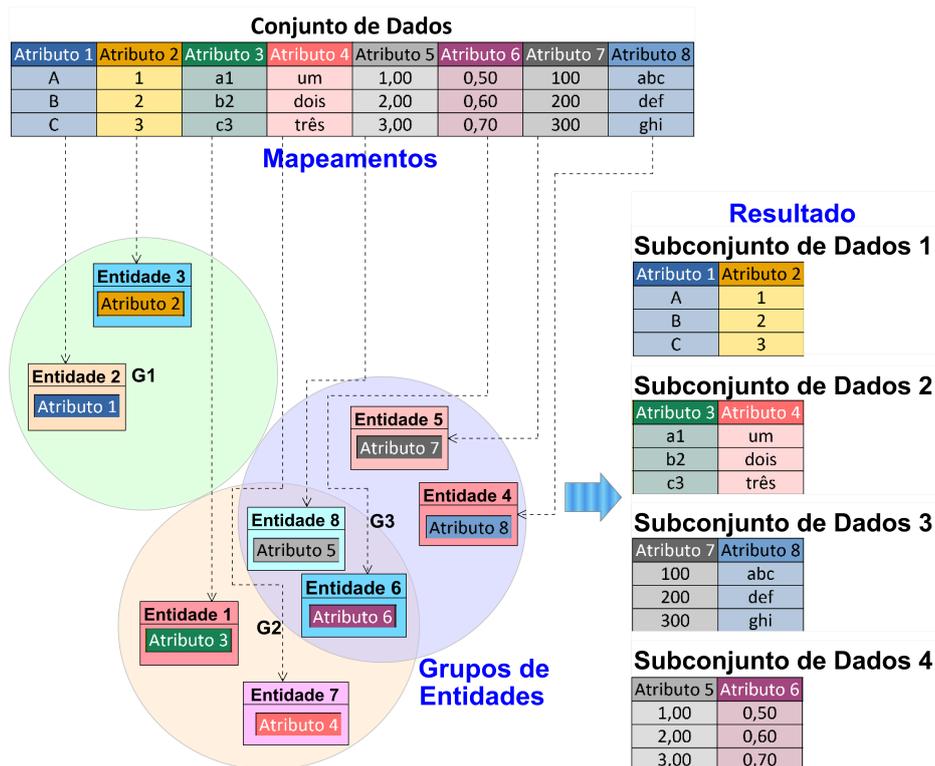
entidades:  $G_1 = \{e_2, e_3\}$ ;  $G_2 = \{e_1, e_6, e_7, e_8\}$ ; e  $G_3 = \{e_4, e_5, e_6, e_8\}$ . As sobreposições (interseções) de entidades ocorrem da seguinte forma:  $e_6$  e  $e_8$  pertencem a  $G_2$  e  $G_3$ . Por fim, cada atributo está associado a uma única entidade:  $a_{e_1} = \{a_3\}$ ,  $a_{e_2} = \{a_1\}$ ,  $a_{e_3} = \{a_2\}$ ,  $a_{e_4} = \{a_8\}$ ,  $a_{e_5} = \{a_7\}$ ,  $a_{e_6} = \{a_6\}$ ,  $a_{e_7} = \{a_4\}$ ,  $a_{e_8} = \{a_5\}$ .

Durante a execução do *ATOp-MapGroupEntityAlgorithm* cada grupo é mapeado para os atributos de suas entidades, gerando os subconjuntos individuais:  $S_{G_1} = \{a_1, a_2\}$ ;  $S_{G_2} = \{a_3, a_4, a_5, a_6\}$ ; e  $S_{G_3} = \{a_5, a_6, a_7, a_8\}$ . Em seguida, são identificados os atributos pertencentes a múltiplos grupos (interseção):  $a_5$  e  $a_6$  aparecem em  $S_{G_2}$  e  $S_{G_3}$ . Após esse passo, o Algoritmo 2 extrai os atributos compartilhados e os agrupa em subconjuntos de interseção, que neste caso, há uma interseção dupla ( $S_{G_2} \cap S_{G_3}$ ), resultando em:  $\mathcal{R}[S_{G_2 \cap G_3}] = \{a_5, a_6\}$ .

Após a criação dos subconjuntos de relações, os atributos compartilhados entre múltiplos grupos são removidos dos subconjuntos individuais, permanecendo exclusivamente nos subconjuntos compartilhados correspondentes. Essa reorganização não apenas preserva as relações semânticas entre as situações representadas na *ATOp-Ontology*, como tam-

bém estabelece uma estrutura explícita para os atributos comuns, evidenciando padrões recorrentes de atributos entre diferentes contextos situacionais. Dessa forma, o *ATOp-MapGroupEntityAlgorithm* retorna o Subconjunto de Dados  $S = \{S_{G_1} = \{a_1, a_2\}, S_{G_2} = \{a_3, a_4\}, S_{G_3} = \{a_7, a_8\}, \mathcal{R}[S_{G_2 \cap G_3}] = \{a_5, a_6\}\}$ , onde os atributos presentes em múltiplos grupos aparecem somente nos subconjuntos de relações.

Figura 15 – Representação do mapeamento com sobreposição de atributos correspondentes.



A forma de utilização dos subconjuntos de dados gerados pelo *ATOp-MapGroupEntityAlgorithm* no processo de treinamento do modelo pode ser determinada em conformidade com as especificidades do projeto de ciência de dados em questão. Tais decisões podem envolver a aplicação de operações adicionais oriundas da teoria dos conjuntos (JECH, 2003), com o intuito de alinhar a segmentação dos dados a determinado viés inerente ao modelo preditivo em desenvolvimento. No escopo da abordagem *ATOp*, as relações identificadas — como as interseções entre subconjuntos de dados, previstas pelo *ATOp-MapGroupEntityAlgorithm* — são utilizadas como etapa preparatória para o treinamento do modelo *ATOp-PredictiveModel*, conforme descrito na Seção 5.1.5.

O mapeamento dos grupos de entidades nos dados integrados do meio operativo permite uma estruturação semântica do conjunto de treinamento. Essa atividade não apenas organiza as informações relevantes, mas também categoriza os atributos de forma a refletir

as diferentes situações operacionais representadas no meio em estudo. O próximo passo na abordagem *ATOp* consiste no treinamento do modelo preditivo utilizando esses subconjuntos de dados segmentados semanticamente.

### 5.1.5 Treinar Modelo de Comitê Preditivo da Autonomia do Meio Operativo (*ATOp-PredictiveModel*)

Esta etapa consiste na seleção e no refinamento de um comitê (*ensemble*) de modelos de AM, que serão utilizados para realizar a análise dos dados previamente preparados nas etapas anteriores do processo. O principal objetivo desta atividade é identificar, ajustar e combinar modelos complementares, de modo a construir um sistema preditivo robusto e eficiente. O produto final consistirá em um modelo preditivo apto a estimar, com elevada qualidade, a autonomia do meio operativo, considerando as características extraídas e transformadas durante as fases de preparação e processamento dos dados.

O *pipeline* de treinamento do *ATOp-PredictiveModel* usa como entrada os subconjuntos de dados oriundos da atividade Mapear Grupos de Entidades nos Dados Integrados do Meio Operativo (Seção 5.1.4). Esses subconjuntos de treinamento são usados como entrada para algoritmos de AM, gerando um modelo preditivo que utiliza a combinação dessas múltiplas perspectivas via meta-aprendizagem no modelo do tipo *Stacking* (WOLPERT, 1992; MIENYE; SUN, 2022).

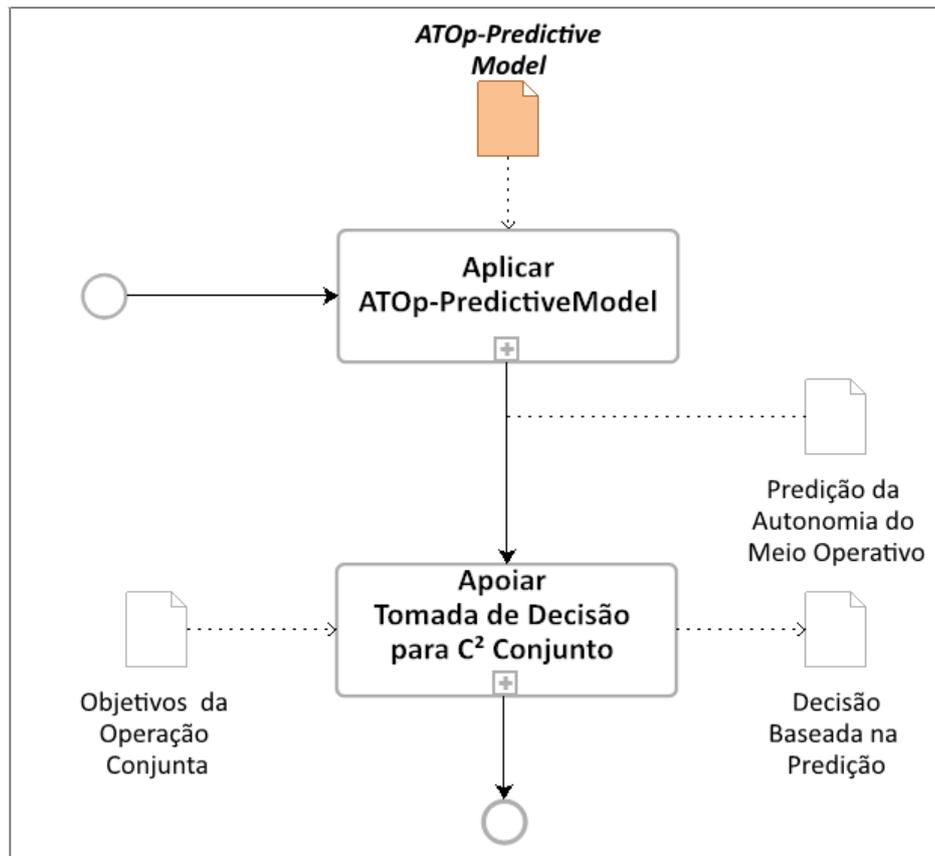
Como resultado desse macroprocesso, obtém-se um modelo preditivo devidamente treinado e validado, capaz de estimar a autonomia dos meios operativos com maior poder de predição. Esse modelo, fundamentado na segmentação semântica dos dados, está preparado para ser aplicado em cenários operacionais, fornecendo predições que apoiam a tomada de decisão. Contudo, caso os resultados de validação indiquem um erro insatisfatório, torna-se necessário revisar todo o processo, desde a modelagem semântica até a construção do modelo, a fim de garantir a qualidade e a confiabilidade das predições geradas.

## 5.2 MACROPROCESSO 2: APLICAR *ATOP-PREDICTIVEMODEL* EM CENÁRIO DE C<sup>2</sup>

Após a construção do modelo preditivo, o segundo macroprocesso foca na aplicação desse modelo para estimar a autonomia dos meios operativos e fornecer suporte à tomada de decisão. O resultado desse macroprocesso consiste na geração de informações estratégicas

fundamentadas, destinadas a subsidiar o processo decisório no contexto operacional. A utilização dessas informações possibilita uma gestão mais eficiente da autonomia dos meios operativos, permitindo a otimização de recursos e a melhoria na capacidade de planejamento e execução de operações. A Figura 16 ilustra os processos atinentes a este macroprocesso, cujas etapas são detalhadas nas subseções a seguir.

Figura 16 – Etapas que compõem o Macroprocesso 2 da abordagem *ATOp*.



### 5.2.1 Aplicar *ATOp-PredictiveModel*

Este processo tem como objetivo gerar estimativas de qualidade sobre a autonomia de meios operativos, a partir da análise de dados futuros coletados em cenários reais. O principal objetivo dessa aplicação é prever o período, em dias, em que um meio operativo pode permanecer em operação sem a necessidade de reabastecimento de suprimentos essenciais.

Essa predição desempenha um papel estratégico na gestão dos recursos disponíveis em uma Força Armada, permitindo a otimização do planejamento logístico e operacional. Ao oferecer informações fundamentadas, a atividade contribui para a tomada de decisões relacio-

nadas à alocação de meios operativos, o planejamento de missões prolongadas e a manutenção da prontidão das unidades operativas.

### 5.2.2 Apoiar a Tomada de Decisão para o C<sup>2</sup> Conjunto

Esta atividade tem como objetivo principal viabilizar a utilização das previsões geradas pelo *ATOp-PredictiveModel* como insumo estratégico para a tomada de decisão em operações conjuntas, dentro do contexto de C<sup>2</sup>. Para isso, é essencial que as previsões sejam interpretadas de forma clara e integradas ao fluxo de informações do sistema de C<sup>2</sup>. O resultado das previsões permite uma avaliação antecipada da autonomia dos meios operativos, auxiliando no planejamento logístico e na mitigação de riscos operacionais. Além disso, a utilização dessas previsões fortalece a capacidade de adaptação das Forças Armadas a cenários dinâmicos, otimizando a distribuição e o emprego dos recursos disponíveis.

O processo de tomada de decisão no contexto de C<sup>2</sup> exige informações exatas e previsões confiáveis para garantir o emprego eficiente dos meios operativos e a otimização dos recursos disponíveis. A abordagem *ATOp* permite a incorporação da etapa semântica do domínio aos modelos preditivos, possibilitando uma análise mais contextualizada da autonomia dos meios operativos.

A incorporação do *ATOp-PredictiveModel* ao sistema de C<sup>2</sup> proporciona um suporte estruturado para a formulação de estratégias operacionais voltadas à otimização da alocação de recursos logísticos em operações militares, levando em consideração as restrições e capacidades específicas de cada meio empregado. Além disso, sua aplicação poderá fortalecer a interoperabilidade entre as diferentes Forças Armadas, por meio da etapa semântica unificada, promovendo uma coordenação mais eficiente e garantindo respostas rápidas e exatas diante de variações na autonomia dos meios operativos no contexto operacional.

## 6 APLICAÇÃO DA ABORDAGEM *ATOP* NO CENÁRIO DE PLATAFORMAS NAVAIS MILITARES

Neste Capítulo, apresenta-se o detalhamento do processo de emprego da abordagem *ATOp*, conforme descrita no Capítulo 5, descrevendo o passo a passo da sua execução. Esse processo inclui a aplicação em um cenário simulado, que ilustra a necessidade de suporte à tomada de decisão no contexto  $C^2$ .

A seguir, as seções deste Capítulo descrevem a aplicação da abordagem *ATOp* no desenvolvimento do modelo preditivo no cenário que tem como foco principal um meio operativo específico, do tipo **Plataforma Naval Militar** (*MilitaryNavalPlatform*), utilizado como base para o presente estudo. Por conseguinte, são expostas todas as atividades realizadas e os artefatos gerados a partir da aplicação dessa abordagem. Tais artefatos estão disponíveis no repositório associado a esta pesquisa<sup>1</sup>.

### 6.1 CONCEPÇÃO

A concepção da *ATOp*, que teve como base as atividades do Macroprocesso 1 (Seção 5.1) da abordagem *ATOp*, foi realizada com base em dados históricos de um meio operativo da Marinha do Brasil (MB), abrangendo o período de 2013 a 2017, durante sua participação em operações militares em cenários caracterizados por  $C^2$ . Esses eventos exigiram o emprego do meio operativo, resultando no consumo de suprimentos, como combustíveis e gêneros alimentícios durante o período de coleta dos dados. Os dados coletados incluem informações estruturadas, extraídas do Sistema de Informações Gerenciais do Abastecimento (SINGRA), e não estruturadas, provenientes dos Relatórios de Fim de Comissão (RFC), elaborados pelo meio operativo ao término de cada operação.

#### 6.1.1 Desenvolvimento da *ATOp-NavalOntology*

A *ATOp-NavalOntology* desenvolvida neste trabalho, conforme detalhado na Seção 5.1.1, fundamenta-se nas categorias ontológicas definidas pela UFO. Para a construção da ontologia, foi utilizada a linguagem OntoUML, implementada por meio do *plug-in* v.0.5.3, integrado à ferramenta *Visual Paradigm Community Edition*<sup>2</sup>.

<sup>1</sup> <https://github.com/Valquire/ATOp.git>

<sup>2</sup> <https://www.visual-paradigm.com/editions/community/>

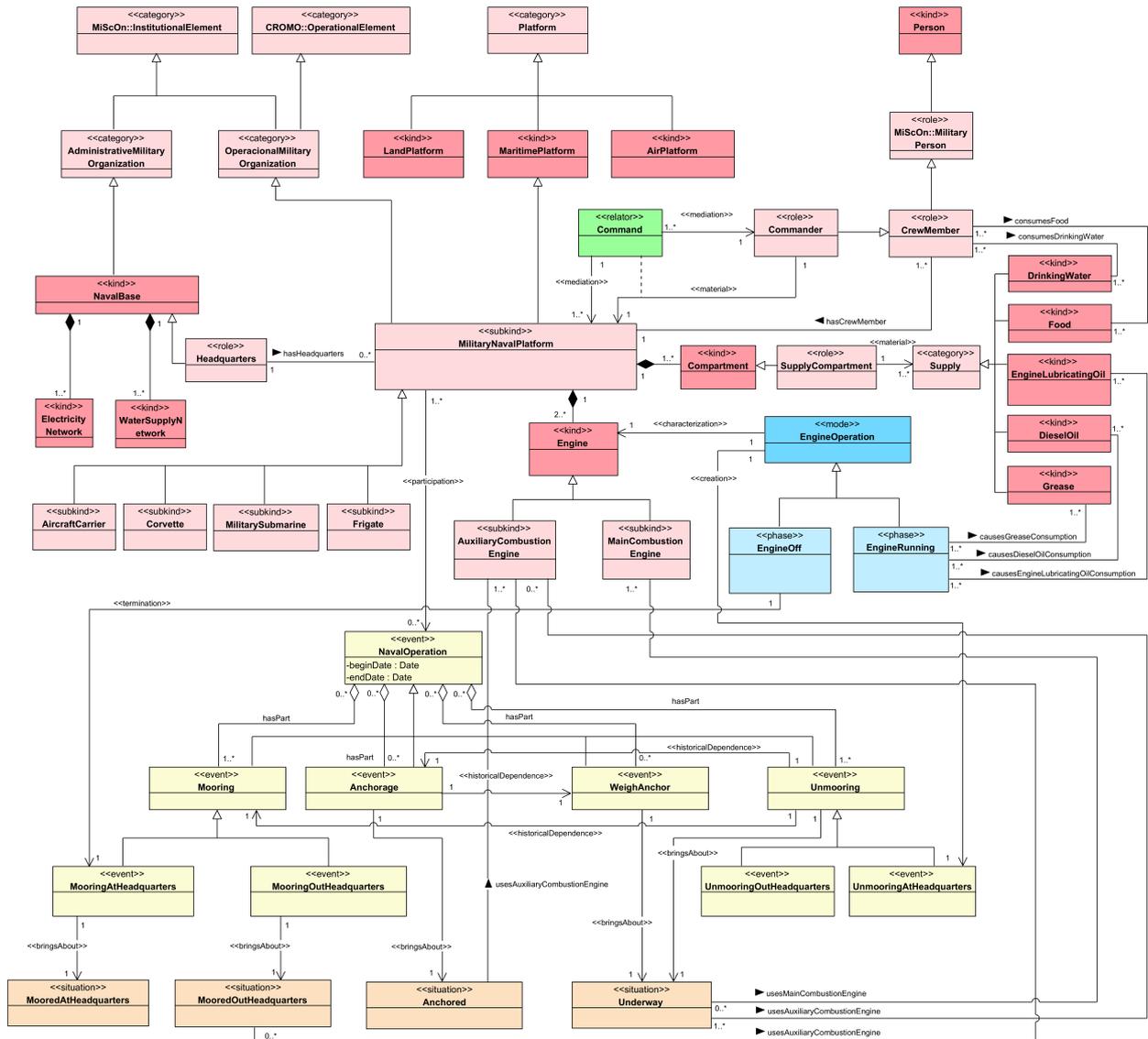
Há uma ampla variedade de conceitos que abrangem não apenas o domínio de interesse, mas também técnicas amplamente reconhecidas na área de MC, as quais desempenham um papel crucial no sucesso do modelo conceitual. No entanto, as informações oriundas do domínio em questão constituem as principais entradas para o processo de construção do modelo. Em razão da significativa relevância dessas informações, bem como da adoção de boas práticas de modelagem, foram formuladas questões de competência, em conformidade com a metodologia SABiO. Essa metodologia enfatiza a elaboração de questões de competência (QC) como etapa essencial no desenvolvimento de uma ontologia de referência.

De acordo com a SABiO, as questões de competência ajudam a identificar os principais aspectos do domínio, auxiliando na definição do escopo da ontologia e na validação de sua eficácia na captura e representação do conhecimento. Portanto, as seguintes questões de competência foram estabelecidas para a *ATOp-NavalOntology*: QC1 - O que é uma Plataforma Naval Militar (*MilitaryNavalPlatform*)? QC2 - Quais tipos de suprimentos podem ser consumidos por uma *MilitaryNavalPlatform*? QC3 - Quem consome os suprimentos que impactam a autonomia de uma *MilitaryNavalPlatform*? QC4 - A situação de uma *MilitaryNavalPlatform* pode afetar o consumo de suprimentos que impactam sua autonomia? A Figura 17 ilustra a ontologia de referência desenvolvida para o domínio da autonomia de Plataformas Navais Militares.

Alinhada às atividades previstas no processo de apoio ao desenvolvimento da metodologia SABiO, a construção da *ATOp-Ontology* reutilizou conceitos previamente estabelecidos em ontologias existentes. Os conceitos de *Elemento Institucional* (*InstitutionalElement*) e *Pessoa Militar* (*MilitaryPerson*) foram reaproveitados da Ontologia de Cenário Militar (*MiScOn*) (DEMORI. et al., 2023). De forma análoga, o conceito de *Elemento Operacional* (*OperationalElement*) foi reutilizado a partir da Ontologia de Rádio Cognitivo para Operações Militares (*CROMO*)<sup>3</sup>. Complementarmente, o conceito de *Plataforma* (*Platform*), modelado com base em princípios da Web Semântica (KARS; OĞUZTÜZÜN, 2011), também foi incorporado à ontologia. A este conceito, agregou-se a especialização *Plataforma Marítima* (*MaritimePlatform*), representada com a metacategoria *Kind* da UFO, conferindo-lhe uma tipificação ontológica adequada ao seu papel no domínio naval.

A *ATOp-NavalOntology* introduz o conceito de Plataforma Naval Militar (*MilitaryNavalPlatform*), que combina o princípio de identidade de uma Plataforma Marítima (*Maritime-*

<sup>3</sup> [https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/trabalhoConclusao/viewTrabalhoConclusao.xhtml?popup=true&id\\_trabalho=12861975#](https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/trabalhoConclusao/viewTrabalhoConclusao.xhtml?popup=true&id_trabalho=12861975#)

Figura 17 – *ATOp-NavalOntology* baseada em UFO.

*Platform*) com características-chave de uma Organização Militar Operativa (*OperationalMilitaryOrganization*). Essa Organização Militar Operativa (*OperationalMilitaryOrganization*) é um conceito que carrega a essência de uma entidade empregada em uma operação militar. Para ilustrar a dupla natureza do conceito de Plataforma Militar Naval (*MilitaryNavalPlatform*), foi representado o conceito de Comando (*Command*), metacategorizado como um *Relator* da OntoUML. O papel do Comandante (*Commander*), um Membro da Tripulação (*CrewMember*), é representado como um dos papéis pertencentes à tripulação de uma Plataforma Naval Militar (*MilitaryNavalPlatform*), herdando o papel de Pessoa Militar (*MilitaryPerson*). Esse Comandante (*Commander*) é formalmente nomeado para desempenhar este papel dentro da Plataforma Naval Militar (*MilitaryNavalPlatform*), que é hierarquica-

mente governada por doutrinas militares. O aspecto operacional está relacionado à presença de um Comandante (*Commander*) liderando as ações da tripulação durante uma operação militar. Adicionalmente, quando não está envolvido em uma operação militar, o pessoal militar desempenha funções administrativas para manter a prontidão da Plataforma Naval Militar para uma operação.

Uma Base Naval (*NavalBase*) é uma Organização Militar Administrativa (*AdministrativeMilitaryOrganization*) que herda características do Elemento Institucional (*InstitutionalElement*). Uma Base Naval (*NavalBase*) possui instalações como a Rede de Eletricidade (*ElectricityNetwork*) e a Rede de Abastecimento de Água (*WaterSupplyNetwork*), que suportam uma Plataforma Naval Militar (*MilitaryNavalPlatform*) caso esta sirva como a sede da Plataforma. Os principais componentes relacionados ao armazenamento e consumo de Suprimentos (*Supplies*) em uma Plataforma Naval Militar (*MilitaryNavalPlatform*) são representados por Compartimento (*Compartment*) e Motor (*Engine*). O Compartimento (*Compartment*), por meio do papel *SupplyCompartment*, armazena suprimentos da categoria *Supply*, como Água Potável (*DrinkingWater*), Alimento (*Food*), Óleo Lubrificante (*LubricatingOil*), Óleo Diesel (*DieselOil*) e Graxa (*Grease*), enquanto o Motor (*Engine*) possui Motor de Combustão Auxiliar (*AuxiliaryCombustionEngine*) e Motor de Combustão Principal (*MainCombustionEngine*). O Funcionamento do Motor (*EngineOperation*) tem duas fases — Fase Ligado (*EngineRunning Phase*) e Fase Desligado (*EngineOff Phase*) — que afetam o consumo de suprimentos.

Além de representar a identidade e dependência relacional da Plataforma Naval Militar (*MilitaryNavalPlatform*), assim como as entidades que a compõem, e demonstrar as relações de consumo de suprimentos que impactam sua autonomia, a *ATOp-NavalOntology* também inclui a participação em eventos. Uma Plataforma Naval Militar (*MilitaryNavalPlatform*) pode participar de um Evento (*Event*), representado por uma Operação Naval (*NavalOperation*), que é uma operação militar focada no ambiente marítimo. Para este Evento (*Event*), dois atributos foram escolhidos para representar relações temporais: *beginDate*, que é a data de início, e *endDate*, que é a data de término, para que todos os outros eventos herdem essas características deste Evento Complexo (*Complex Event*), conforme definido pela UFO (ALMEIDA; FALBO; GUIZZARDI, 2019).

Como um Evento Complexo (*Complex Event*), representado por uma Operação Naval (*NavalOperation*), ele compreende outros eventos com os quais mantém uma mereologia

extensional (ALMEIDA; FALBO; GUIZZARDI, 2019). Esses eventos também são representados na *ATOp-NavalOntology*. Os eventos de Atracação (*Mooring*), Desatracação (*Unmooring*), Fundeio (*Anchorage*) e Suspende (*WeighAnchor*) têm uma dependência histórica. Para que ocorra a Atracação (*Mooring*), a Desatracação (*Unmooring*) deve ter ocorrido anteriormente. Da mesma forma, para que ocorra o Fundeio (*Anchorage*), a Desatracação (*Unmooring*) deve ter ocorrido. Para que ocorra o Suspende (*WeighAnchor*), o Fundeio (*Anchorage*) deve ter ocorrido.

Diante dessa estrutura de dependência entre eventos, o Evento (*Event*) de Atracação (*Mooring*) é especializado em dois eventos: a Atracação na Base Sede (*MooringAtHeadquarters*) e a Atracação Fora de Sede (*MooringOutHeadquarters*). Esses eventos criam as Situações (*Situations*) de estar Atracado na Base Sede (*MooredAtHeadquarters*) e Atracado Fora de Sede (*MooredOutHeadquarters*). O Evento (*Event*) de Atracação na Base Sede (*MooringAtHeadquarters*) encerra o Funcionamento do Motor (*EngineOperation*), portanto, não há o consumo de Suprimentos (*Supplies*) pelo Motor (*Engine*). A Situação (*Situation*) de estar Atracado Fora de Sede (*MooredOutHeadquarters*) aciona o Motor de Combustão Auxiliar (*AuxiliaryCombustionEngine*). Essa ativação é mantida enquanto a Fase Ligado (*EngineRunning Phase*) permanece ativa, provocando o consumo de suprimentos relacionados ao Motor (*Engine*).

Igualmente, o Evento (*Event*) de Fundeio (*Anchorage*) cria a Situação (*Situation*) de estar Fundeado (*Anchored*). Essa Situação (*Situation*) envolve configurações onde o Motor de Combustão Auxiliar (*AuxiliaryCombustionEngine*) permanece ativado, causando consumo de suprimentos. Para o Evento de Desatracação (*Unmooring Event*), há uma especialização em Desatracação da Sede (*UnmooringAtHeadquarters*) e Desatracação Fora de Sede (*UnmooringOutHeadquarters*). Os eventos Suspende (*WeighAnchor*), Desatracação Fora de Sede (*UnmooringOutHeadquarters*) e Desatracação da Sede (*UnmooringAtHeadquarters*) provocam uma Situação de estar Em Movimento (*Underway Situation*) para a Plataforma Naval (*MilitaryNavalPlatform*), mas isso não significa que a Situação Em Movimento necessita que esses eventos ocorram ao mesmo tempo, pois existem relação temporais entres eles. Essa Situação de estar Em Movimento (*Underway Situation*) envolve o uso tanto do Motor de Combustão Auxiliar (*AuxiliaryCombustionEngine*) quanto do Motor de Combustão Principal (*MainCombustionEngine*) durante sua Fase Ligado (*EngineRunning Phase*), também causando o consumo de suprimentos pelo Motor (*Engine*).

A representação das entidades e suas relações, especialmente da Plataforma Naval Militar (*MilitaryNavalPlatform*), ilustram como o consumo de Suprimentos (*Supplies*) se correlaciona com a Situação (*Situation*) da Plataforma. Eventos como a Operação Naval (*NavalOperation*) afetam as taxas de consumo de Suprimentos (*Supplies*). O objetivo da *ATOp-NavalOntology* é definir a Plataforma Naval Militar (*MilitaryNavalPlatform*), explicar sua herança de identidade e identificar as entidades relacionadas que influenciam o consumo de material e a sua autonomia.

A escolha desta metodologia para representar este conceito específico de domínio se deve às capacidades expressivas e semânticas que a linguagem OntoUML fornece ao modelo conceitual, baseada em todas as teorias sobre relações ontológicas presentes na UFO (conforme descritas na Seção 3.2.3), que podem ser empregadas em vários campos do conhecimento. O conceito de Situação (*Situation*), conforme definido na ontologia UFO e adotado neste estudo, permite a representação e exploração dos efeitos do consumo de suprimentos sobre a Plataforma Naval Militar (*MilitaryNavalPlatform*) ao longo de seu ciclo de existência.

### 6.1.2 Integração Dados Históricos da *MilitaryNavalPlatform*

A coleta de dados foi conduzida em duas etapas distintas, com o objetivo de garantir uma abrangente representação das informações necessárias para o estudo. Na fase inicial, relacionados ao abastecimento da Plataforma Naval Militar foram extraídos do Sistema de Informações Gerenciais do Abastecimento (SINGRA), o qual disponibiliza um conjunto de registros organizados de forma sistemática, possibilitando o processamento eficiente e a realização de análises quantitativas subsequentes. Os detalhes desta atividade foram descritos na Seção 5.1.2.

Na segunda fase, foram coletados dados não estruturados, oriundos dos Relatórios de Fim de Comissão (RFC), os quais apresentam informações textuais e qualitativas que exigem tratamento específico para serem transformadas em dados utilizáveis. Após a coleta dos dados em ambas as fases, iniciou-se o processo de integração dessas informações, a fim de criar uma base de dados coesa e homogênea para a análise subsequente, conforme detalhado no código-fonte que está disponível no repositório associado a esta pesquisa<sup>4</sup>. Esse procedimento foi essencial para garantir a qualidade estrutural dos dados e possibilitar a aplicação de técnicas de AM apropriadas.

---

<sup>4</sup> <https://github.com/Valquire/ATOp.git>

Tabela 4 – Esquema dos dados de abastecimento da plataforma naval militar.

<b>Campo</b>	<b>Descrição</b>	<b>Tipo</b>
ANO	Ano da requisição do material	Discreto
PERIODO	Data da requisição do material	Discreto
MES	Mês da requisição do material	Discreto
SIGLA_OM	Sigla da Organização Militar	Nominal
RM_TYPE	Tipo da requisição do material	Nominal
SITUACAO_RM	Situação da requisição do material	Nominal
DESC_SITUACAO_RM	Observações da requisição do material	Nominal
GRUPO_JURISDICA0	Identifica a natureza do material	Nominal
DESCRICA0_CLG	Classifica os materiais combustíveis	Nominal
PI	Identificador numérico do material	Discreto
NOME_PORTUGUES	Nome do material requisitado	Nominal
UF	Unidade de Fornecimento do material	Nominal
QTDE_ITENS_RM	Quantidade de itens da requisição	Contínuo
PRECO_UNITARIO_VENDA	Preço de venda unitário	Contínuo

O conjunto de dados extraído do SINGRA consiste em 1.823 registros únicos, armazenados no formato *Comma-Separated Values* (CSV). Os dados coletados referem-se ao período compreendido entre janeiro de 2013 e dezembro de 2017. A Tabela 4 ilustra o esquema dos dados coletados.

Uma análise preliminar sobre os dados permitiu a identificação de um ajuste necessário na padronização da unidade de fornecimento (UF) dos materiais. Além das unidades de medida quilograma (*kg*) e litro (*l*), os valores dos registros estavam associados a outros tipos de unidades. A Tabela 5 apresenta a relações de conversões utilizadas.

Após todas essas conversões, as unidades de fornecimento dos itens foram padronizadas, resultando em representações exclusivamente em quilograma ou litro. Para a preparação e organização desses dados, foi utilizado o campo “GRUPO\_JURISDICA0”, que permitiu a segregação dos registros em dois grupos principais: “SUBSISTÊNCIA”, contendo 1.478 registros, e “COMBUSTÍVEIS”, com 348 registros.

A preparação dos dados coletados constituiu uma etapa essencial de engenharia de dados

Tabela 5 – Conversões de unidades de fornecimento de materiais.

Unidade Original	Equivalente Padronizado
Balde	20,0 <i>l</i>
Galão	3,79 <i>l</i>
Garrafa	0,90 <i>l</i>
Lata	5,00 <i>kg</i>
Pacote	0,50 <i>kg</i>
Quarto de Galão Americano	0,95 <i>l</i>
Tambor	200,0 <i>l</i>

e foi conduzida com o suporte da biblioteca *Pandas*<sup>5</sup>, da linguagem *Python*. Essa fase incluiu procedimentos como limpeza, padronização, eliminação de atributos redundantes e agregação temporal, assegurando que os dados estivessem prontos para integração semântica e aplicação em modelos preditivos.

Inicialmente, foram criados os *DataFrames* “CLG” e “ALIMENTOS”, os quais organizaram, respectivamente, os insumos do grupo “COMBUSTÍVEIS” — agregando categorias como combustível, lubrificante e graxa — e os itens da classe “SUBSISTÊNCIA”, estruturados com base na pirâmide alimentar brasileira (PHILIPPI et al., 1999). Posteriormente, esses dois *DataFrames* foram integrados a dados operacionais oriundos dos Relatórios de Fim de Comissão (RFCs), formando um *DataFrame* consolidado com atributos como “MILITARES”, “DIAS\_PORTO\_SEDE”, “DIAS\_PORTO\_FORA\_SEDE”, “MILHAS” e “DIAS\_MAR”.

Essa estrutura integrada foi então segmentada em subconjuntos semanticamente coerentes com apoio da ontologia *ATOp-NavalOntology*, gerando *DataFrames* associados aos mapeamentos de atributos realizado sobre os grupos de entidades identificados por meio das situações *MooredAtHeadquarters*, *Underway*, *Anchored* e *MooredOutHeadquarters*. Assim, o uso de *DataFrames* não apenas viabilizou a organização eficiente dos dados tabulares, como também serviu de elo entre a MC baseada em ontologia e as etapas de experimentação preditiva com os modelos de AM.

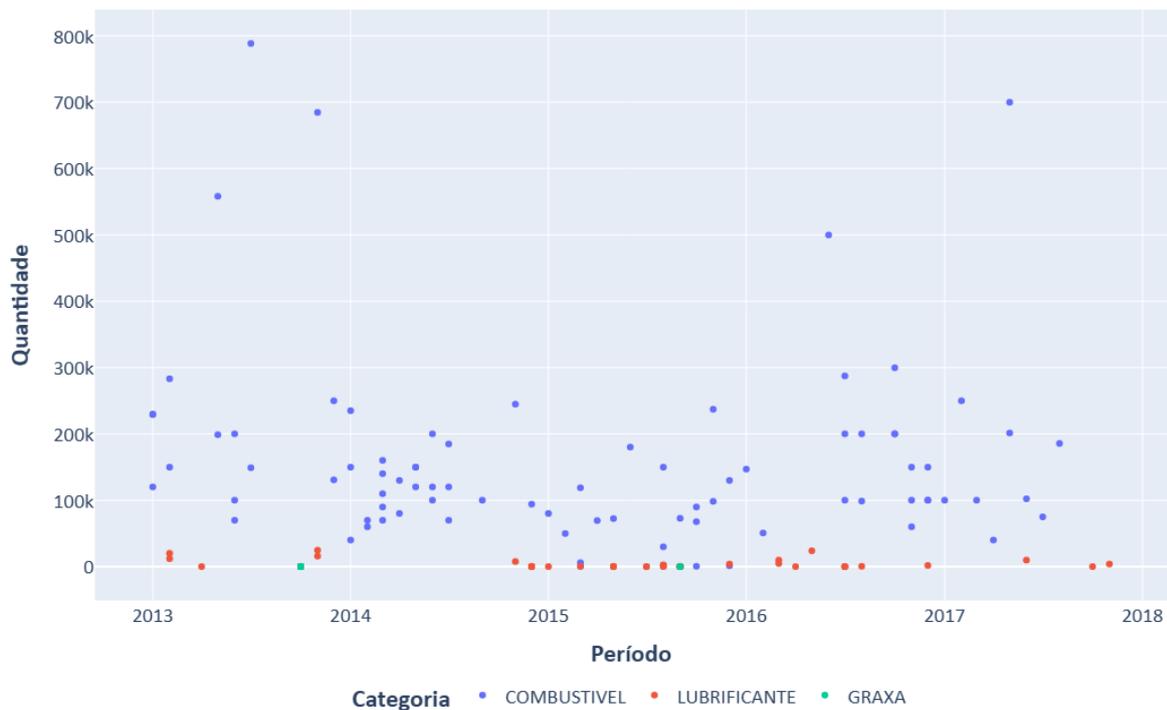
O grupo denominado “COMBUSTÍVEIS” originou o *DataFrame* denominado “CLG”, que consolida os dados relacionados a este setor, organizados em três categorias principais: Combustível, Lubrificante e Graxa. De maneira análoga, a partir dos dados do grupo

<sup>5</sup> <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>

“SUBSISTÊNCIA”, foi elaborado o *DataFrame* denominado “ALIMENTOS”, o qual reúne as informações pertinentes a este segmento. Dado que não havia categorias pré-definidas para este grupo, as categorias foram estabelecidas com base na pirâmide alimentar brasileira, conforme descrito por Philippi et al. (1999).

A distribuição dos dados extraídos dos *DataFrames* “CLG” e “ALIMENTOS”, referentes ao período de 2013 a 2017, é apresentada nas Figuras 18 e 19. Os dados estão organizados por mês, e cada ponto nas representações gráficas corresponde a uma requisição de material registrada pela plataforma naval militar. Essa estrutura permite uma visualização clara da evolução do consumo ao longo do tempo.

Figura 18 – Visualização do consumo Combustível, Lubrificante e Graxa.



A junção desses dois *DataFrames* teve como base o ano e o mês da requisição do material de consumo. Cada registro possui o valor acumulado da variável no trimestre. Foi imputada a média para valores desconhecidos.

Os dados extraídos dos RFCs incluem 155 registros únicos, relacionados às operações militares nas quais uma Plataforma Naval Militar, pertencente à MB, esteve envolvida durante o período analisado. A Tabela 6 apresenta o esquema desses dados.

A fim de facilitar a integração entre os dois conjuntos de dados, o conjunto de dados

Figura 19 – Visualização do consumo de Alimentos.

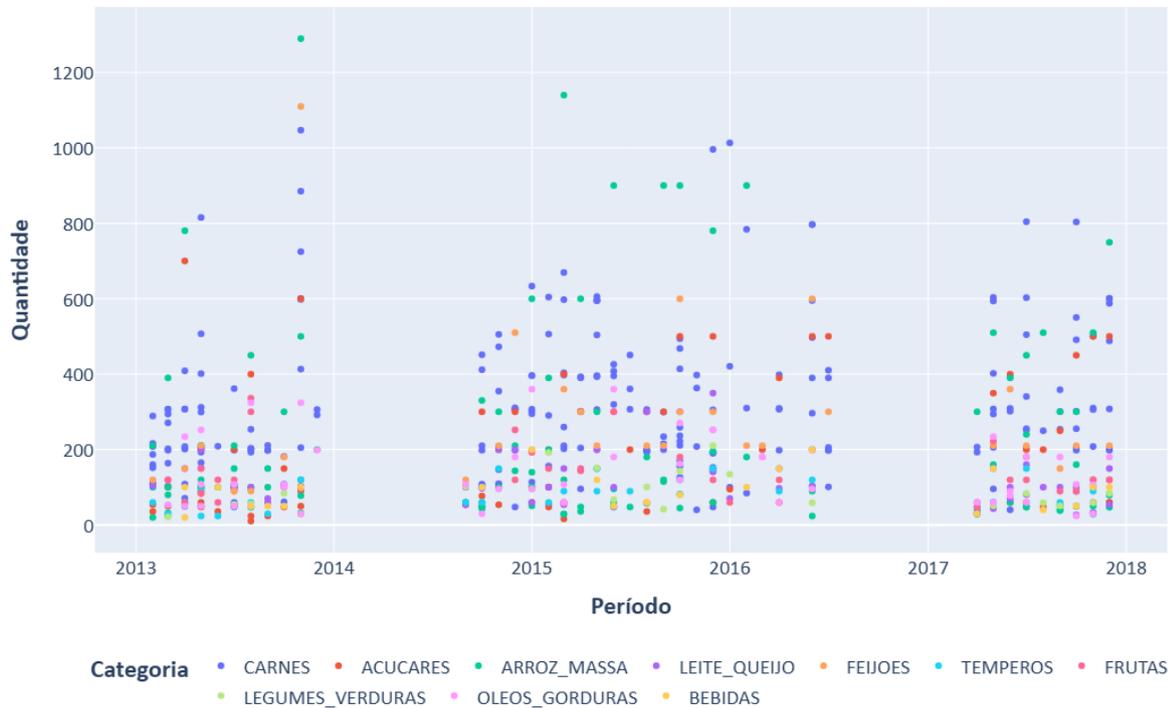


Tabela 6 – Esquema dos dados operacionais da plataforma naval militar.

<b>Campo</b>	<b>Descrição</b>	<b>Tipo</b>
DT_INICIO	Data de início da operação militar	Discreto
DT_FIM	Data de fim da operação militar	Discreto
OPERACAO	Nome da operação militar	Nominal
MILITARES	Quantidade de militares	Discreto
DIAS_PORTO_FORA_SEDE	Quantidade de dias em portos fora da sede	Discreto
MILHAS	Quantidade de milhas navegadas	Contínuo
DIAS_MAR	Quantidade de dias de operação no mar	Discreto

RFC passou por uma transformação com base no ano e no mês de cada observação. A variável “DIAS\_PORTO\_SEDE” é derivada da combinação das seguintes variáveis: “ANO”, “TRIMESTRE”, “DIAS\_PORTO\_FORA\_SEDE” e “DIAS\_MAR”. Os registros da variável MILITARES representam a média para cada trimestre, enquanto as demais variáveis apresentam registros de valores acumulados.

A integração entre os conjuntos de dados *SINGRA* e *RFC*, após as devidas transfor-

mações, foi realizada por meio dos campos *ANO* e *TRIMESTRE*, utilizados como chaves de junção, pois estão presentes em ambos os conjuntos de dados. O *DataFrame* resultante desse processo — contendo dados agregados, conforme descrito na Seção 6.2 — é apresentado na Figura 20. Esse *DataFrame* é composto por 518 registros únicos, distribuídos entre 20 atributos, incluindo o rótulo *DIAS\_MAR*. As variáveis *ANO*, *TRIMESTRE*, *MILITARES*, *DIAS\_PORTO\_SEDE*, *DIAS\_PORTO\_FORA\_SEDE* e *DIAS\_MAR* são do tipo discreto, enquanto as demais são variáveis contínuas.

### 6.1.3 Agrupamento das Entidades da *ATOp-NavalOntology*

Com o desenvolvimento da *ATOp-NavalOntology*, tornou-se viável a aplicação da *ATOp-GroupEntityRule* com o suporte do Algoritmo 1, previamente apresentado na Seção 5.1.3. Após a execução do referido algoritmo, foram identificados quatro grupos, incluindo as redundâncias de entidades presentes em cada um deles. A seguir, apresenta-se uma descrição detalhada desses grupos:

- $G_1 = \{MooredAtHeadquarters, MilitaryNavalPlatform, CrewMember, Food, DrinkingWater, EngineOff\}$ , demonstra a situação *MooredAtHeadquarters* (Atracado na Base Sede);
- $G_2 = \{Underway, MilitaryNavalPlatform, CrewMember, Food, DrinkingWater, MainCombustionEngine, EngineLubricatingOil, DieselOil, Grease, EngineRunning, AuxiliaryCombustionEngine\}$ , representa a situação *Underway* (Em Movimento);
- $G_3 = \{Anchored, MilitaryNavalPlatform, CrewMember, Food, DrinkingWater, EngineLubricatingOil, DieselOil, Grease, EngineRunning, AuxiliaryCombustionEngine\}$ , indica a situação *Anchored* (Fundeadado).
- $G_4 = \{MooredOutHeadquarters, MilitaryNavalPlatform, CrewMember, Food, DrinkingWater, EngineLubricatingOil, DieselOil, Grease, EngineRunning, AuxiliaryCombustionEngine\}$ , identifica a situação *MooredOutHeadquarters* (Atracado Fora de Sede).

Além da identificação dos grupos de entidades, foram também observadas relações, principalmente de interseção, entre esses grupos, evidenciando a presença de entidades compartilhadas entre diferentes agrupamentos. Essas interseções são particularmente relevantes, pois podem indicar padrões de redundância, correlações ou relações semânticas entre as entidades

Figura 20 – Amostra de vinte e cinco registros extraídos do Conjunto Derivado.

ANO	TRIMESTRE	COMBUSTIVEL (l)	GRAXA (kg)	LUBRIFICANTE (l)	ACUCARES (kg)	ARROZ_MASSA (kg)	BEBIDAS (l)	CARNES (kg)	FEIJOS (kg)	FRUTAS (kg)	LEGUMES_VERDURAS (kg)	LEITE_QUEIJO (kg)	OLEOS_GORDURAS (l)	TEMPEROS (kg)	MILITARES	DIAS_PORTO_SEDE	DIAS_PORTO_FORA_SEDE	MILHAS	DIAS_MAR
2013	1	1.274.618,59	226,80	47.008,75	579,08	1.529,20	228,82	5.353,76	692,38	730,52	245,50	477,86	468,32	174,78	227	32	18	9.220,96	40
2013	2	1.389.391,59	226,80	15.148,75	1.103,04	1.882,00	296,27	5.022,77	796,19	558,00	335,60	688,13	901,16	198,00	205	33	17	6.914,57	42
2013	4	1.328.018,59	302,20	49.864,37	1.276,60	2.620,00	226,27	5.423,98	1.576,19	688,52	223,87	389,53	760,00	289,56	222	60	7	6.240,00	26
2014	1	1.125.000,00	226,80	22.513,12	1.017,23	1.719,59	228,82	5.593,81	858,57	735,77	335,60	560,79	621,49	254,33	261	0	35	13.629,25	56
2014	1	1.125.000,00	226,80	22.513,12	1.017,23	1.719,59	228,82	5.593,81	858,57	735,77	335,60	560,79	621,49	254,33	261	0	35	13.629,25	56
2014	3	475.000,00	151,20	15.008,75	392,54	633,20	152,55	1.924,60	406,19	451,26	213,87	340,93	315,16	144,78	261	10	32	13.145,04	51
2014	4	625.491,70	225,30	19.018,82	1.401,11	2.635,92	232,68	5.795,69	1.485,23	546,34	464,10	980,11	1.264,58	353,31	226	65	9	4.222,66	19
2014	3	625.346,49	227,24	19.025,69	1.400,47	2.635,41	232,20	5.787,86	1.485,02	546,14	465,44	978,34	1.262,55	353,88	226	64	7	4.221,16	20
2014	3	625.227,84	225,99	18.989,68	1.397,15	2.634,93	232,46	5.783,42	1.483,80	545,99	464,74	978,63	1.262,21	353,59	225	63	8	4.222,57	20
2014	3	625.447,04	229,02	19.007,49	1.397,16	2.635,96	232,84	5.798,51	1.489,26	548,54	466,62	974,97	1.263,12	356,04	227	62	10	4.208,50	20
2014	4	625.308,86	227,44	19.007,48	1.398,38	2.632,31	232,39	5.787,37	1.484,89	544,09	461,71	974,00	1.264,18	352,57	226	64	9	4.217,52	20
2014	3	625.291,03	226,43	18.973,41	1.399,51	2.636,30	232,42	5.795,34	1.488,61	543,00	465,84	977,85	1.262,91	355,69	225	63	8	4.224,39	20
2014	4	625.449,38	226,30	18.993,10	1.398,48	2.635,25	231,66	5.793,61	1.486,67	543,74	464,35	980,25	1.268,94	354,43	225	62	8	4.209,81	20
2014	3	625.430,56	226,61	18.979,35	1.396,67	2.640,28	232,58	5.791,33	1.487,16	546,27	464,62	977,73	1.259,28	354,21	225	65	9	4.213,94	20
2014	3	625.300,78	226,21	19.005,33	1.397,63	2.637,16	232,29	5.786,56	1.488,05	546,23	465,09	980,54	1.258,62	352,52	226	63	9	4.218,31	20
2014	4	625.570,98	227,19	19.006,01	1.399,64	2.634,28	232,59	5.786,39	1.485,35	546,71	464,29	979,70	1.263,37	354,29	226	63	9	4.215,84	20
2014	4	625.268,89	227,32	19.008,93	1.397,97	2.634,76	233,30	5.790,03	1.490,50	545,38	463,53	977,57	1.263,00	353,77	226	63	8	4.207,82	20
2015	4	625.388,00	226,80	19.008,75	1.398,48	2.634,80	232,55	5.793,64	1.486,19	545,26	464,67	977,33	1.263,16	354,78	226	64	9	4.215,39	20
2015	2	322.166,00	226,80	15.545,53	978,15	2.240,40	246,27	7.275,60	796,19	1.139,26	330,27	686,73	954,32	409,56	240	61	4	5.178,05	27
2015	3	625.291,23	227,07	19.019,07	1.394,99	2.639,27	232,59	5.794,44	1.483,82	549,23	464,60	977,29	1.264,96	355,79	225	63	8	4.219,63	21
2015	3	625.342,46	226,72	19.003,99	1.400,28	2.632,40	232,73	5.793,71	1.491,04	544,95	462,65	976,76	1.266,39	355,26	226	64	9	4.221,55	20
2015	3	625.420,06	225,15	19.032,51	1.396,29	2.633,56	232,94	5.801,16	1.485,87	540,51	463,42	975,88	1.260,85	354,33	226	63	9	4.210,59	20
2016	4	1.360.000,00	226,80	17.008,75	1.017,23	1.719,59	228,82	5.593,81	858,57	735,77	335,60	560,79	621,49	254,33	248	0	54	8.606,30	38
2017	1	450.000,00	226,80	22.513,12	1.017,23	1.719,59	228,82	5.593,81	858,57	735,77	335,60	560,79	621,49	254,33	247	0	46	11.882,06	44
2017	2	1.043.560,00	226,80	25.008,75	886,62	1.536,80	256,27	4.006,80	856,19	390,00	274,14	455,80	522,00	229,56	231	59	5	5.526,01	27

classificadas. A seguir, são apresentadas as interseções identificadas, detalhando as entidades envolvidas e suas respectivas associações dentro dos grupos:

- $G_1 \cap G_2 \cap G_3 \cap G_4 = \{MilitaryNavalPlatform, CrewMember, Food, DrinkingWater\}$ ;
- $(G_2 \cap G_3 \cap G_4) \setminus G_1 = \{Grease, EngineLubricatingOil, DieselOil, EngineRunning, AuxiliaryCombustionEngine\}$ .

As interseções entre esses grupos são reaproveitadas na tarefa *ATOp-MapGroupEntity* (conforme a Seção 5.1.4). O objetivo dessa tarefa é mapear grupos de entidades nos dados integrados, viabilizando a identificação e classificação das características relevantes para o modelo preditivo, além de facilitar a segmentação dos subconjuntos de dados. O resultado desse processo é apresentado na Figura 21. Os grupos de entidades foram empregados na etapa de mapeamento dos dados de treinamento, originando os subconjuntos utilizados na fase de treinamento do *ATOp-PredictiveModel*.

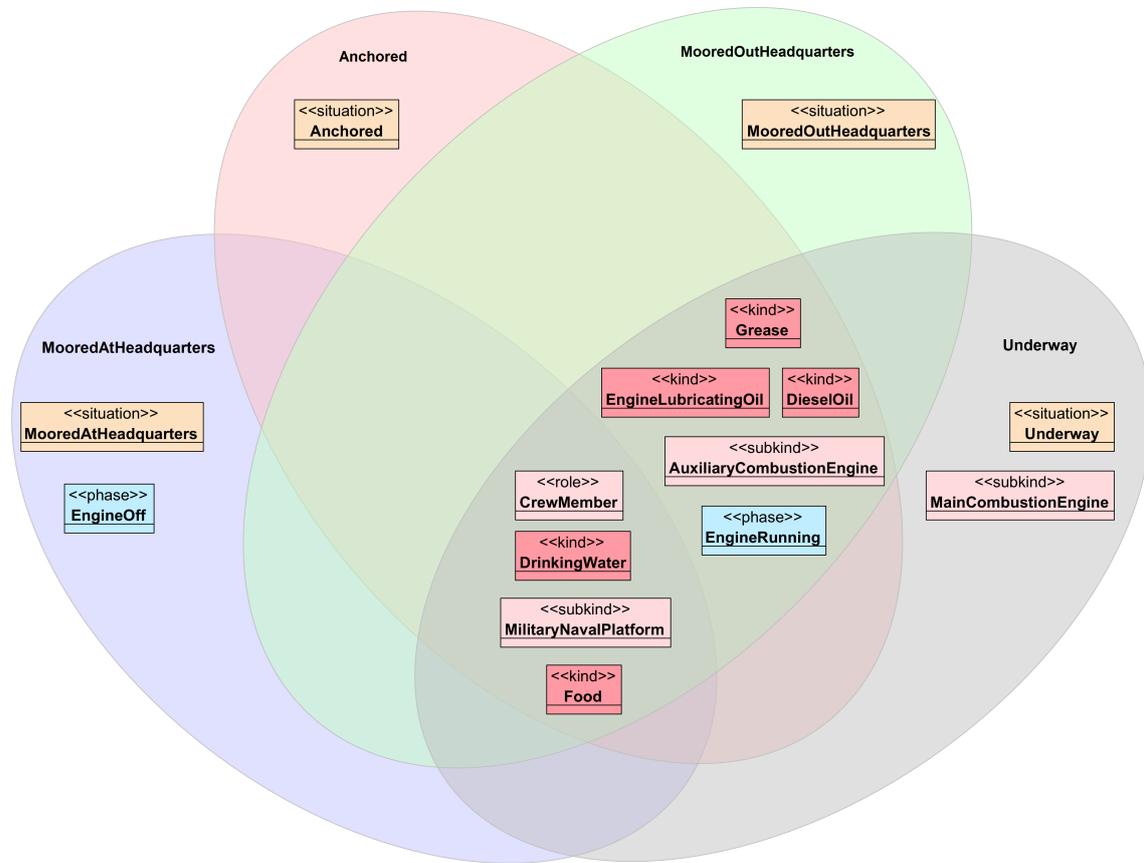
#### 6.1.4 Mapeamento dos Grupos de Entidades nos Dados Integrados

O processo inicial de aplicação do *ATOp-MapGroupEntityAlgorithm* sobre a *ATOp-NavalOntology* permitiu estruturar os atributos conforme os grupos de entidades identificados na ontologia, refletindo as diferentes situações operacionais do meio em estudo. A correspondência entre as entidades ontológicas e os atributos dos dados integrados foi definida a partir de uma análise empírica conduzida sobre os conjuntos de dados coletados. Essa análise envolveu a inspeção detalhada das variáveis disponíveis, considerando seus significados, contextos de uso e relações com os elementos do domínio da autonomia de plataformas navais militares.

As associações foram determinadas com base em afinidades conceituais entre os atributos e as entidades da ontologia, orientadas pelo conhecimento do domínio e pela estrutura semântica previamente modelada. Embora não automatizado, esse processo foi realizado de forma empírica, a partir de levantamento com especialistas em plataformas navais militares, assegurando que os atributos fossem agrupados de maneira semanticamente coerente com os conceitos operacionais representados na ontologia. A Tabela 7 sintetiza os resultados obtidos na etapa de correspondência entre entidades ontológicas e atributos dos dados.

A aplicação do *ATOp-MapGroupEntityAlgorithm* sobre os grupos de entidades identificados na *ATOp-NavalOntology* resultou na segmentação dos dados em subconjuntos seman-

Figura 21 – Diagrama de Venn da aplicação da *ATOp-GroupEntityRule* sobre a *ATOp-NavalOntology*.



ticamente estruturados, garantindo que os atributos preditivos estivessem organizados de acordo com as relações ontológicas estabelecidas. O algoritmo garantiu que cada atributo fosse corretamente atribuído aos grupos individuais e interseções onde ocorre, removendo a redundância necessária para manter a coerência semântica da segmentação. Após a execução do algoritmo, os subconjuntos individuais ficaram vazios devido à remoção dos atributos que eram compartilhados por mais de um subconjunto.

O algoritmo também foi capaz de gerar os subconjuntos de relações de atributos. Esses subconjuntos representam os atributos que são compartilhados entre múltiplos grupos. Como resultado, foram obtidos os seguintes subconjuntos de relações:

- $S_{G_1 \cap G_2 \cap G_3 \cap G_4} = \{ANO, TRIMESTRE, DIAS\_PORTO\_SEDE, MILHAS, DIAS\_MAR, DIAS\_PORTO\_FORA\_SEDE, MILITARES, ACUCARES, ARROZ\_MASSA, BEBIDAS, CARNES, FEIJOES, FRUTAS, LEGUMES\_VERDURAS, LEITE\_QUEIJO,$

Tabela 7 – Correspondência entre Entidades e Atributos.

<b>Entidade</b>	<b>Atributos</b>
<i>Anchored</i>	<i>NULL</i>
<i>AuxiliaryCombustionEngine</i>	<i>NULL</i>
<i>CrewMember</i>	<i>MILITARES</i>
<i>DieselOil</i>	<i>COMBUSTIVEL</i>
<i>DrinkingWater</i>	<i>NULL</i>
<i>EngineLubricatingOil</i>	<i>LUBRIFICANTE</i>
<i>EngineOff</i>	<i>NULL</i>
<i>EngineRunning</i>	<i>NULL</i>
<i>Food</i>	<i>ACUCARES, ARROZ_MASSA, BEBIDAS, CARNES, FEIJOES, FRUTAS, LEGUMES_VERDURAS, LEITE_QUEIJO, OLEOS_GORDURAS, TEMPEROS</i>
<i>Grease</i>	<i>GRAXA</i>
<i>Underway</i>	<i>NULL</i>
<i>MainCombustionEngine</i>	<i>NULL</i>
<i>MilitaryNavalPlatform</i>	<i>ANO, TRIMESTRE, DIAS_PORTO_SEDE, MILHAS, DIAS_MAR, DIAS_PORTO_FORA_SEDE</i>
<i>MooredAtHeadquarters</i>	<i>NULL</i>
<i>MooredOutHeadquarters</i>	<i>NULL</i>

*OLEOS\_GORDURAS, TEMPEROS*};

- $S_{(G2 \cap G3 \cap G4) \setminus G1} = \{COMBUSTIVEL, LUBRIFICANTE, GRAXA\}$ .

A aplicação do *ATOp-MapGroupEntityAlgorithm* permitiu uma segmentação dos dados, garantindo que todos os atributos fossem corretamente alocados aos grupos e interseções correspondentes. O algoritmo manteve a coerência semântica da estrutura ontológica, assegurando que os atributos compartilhados fossem corretamente preservados nos subconjuntos de relações. Dessa forma, os dados organizados refletem a segmentação esperada, possibilitando análises mais estruturadas e alinhadas com o modelo conceitual desenvolvido. Como resultado tem-se o subconjunto de dados  $S = \{S_1 = \{S_{G1 \cap G2 \cap G3 \cap G4}\}, S_2 = \{S_{(G2 \cap G3 \cap G4) \setminus G1}\}$ .

Com essa organização refinada, os subconjuntos resultantes foram utilizados como entrada na fase de treinamento dos modelos de AM (*ATOp-PredictiveModel*). A segmentação baseada na estrutura conceitual da *ATOp-NavalOntology* aprimora a capacidade dos modelos de capturar padrões específicos de cada situação operacional, proporcionando maior qualidade nas previsões geradas. A implementação detalhada do *ATOp-MapGroupEntityAlgorithm*

e o processo de segmentação dos subconjuntos podem ser acessados no repositório da pesquisa<sup>6</sup>.

## 6.2 RESULTADOS DO TREINAMENTO

Esta seção apresenta os experimentos realizados na etapa de treinamento do *ATOp-PredictiveModel*, componente essencial da abordagem *ATOp*. O objetivo principal desta etapa é avaliar os ganhos preditivos e computacionais decorrentes da aplicação de comitês de regressão treinados com base em subconjuntos semânticos de atributos, definidos a partir do modelo conceitual ontológico. Essa estruturação semântica busca explorar o conhecimento de domínio para orientar a indução de modelos mais robustos e especializados, otimizando sua capacidade de generalização e eficiência.

Para contextualizar e validar os resultados obtidos com o *ATOp-PredictiveModel*, foram realizados três blocos experimentais. O primeiro foi o *Treinamento de Modelos Individuais*, seguido pelo *Treinamento de Comitês de Modelos sem Direcionamento Semântico* e por último o *Treinamento do ATOp-PredictiveModel*. Contudo, apenas o último integra diretamente a abordagem *ATOp*, enquanto os demais — embora metodologicamente relevantes — foram incluídos nesta seção para fins comparativos e para demonstrar a evolução da estratégia preditiva adotada.

Em todos os blocos, os modelos foram avaliados com base em dois conjuntos: o conjunto real, composto por dados observacionais, e o conjunto derivado, formado pela união do conjunto real com instâncias sintéticas geradas por técnicas de sobreamostragem (*oversampling*) utilizando a biblioteca *Imbalanced Learning Regression* (BRANCO; TORGO; RIBEIRO, 2019; WU; KUNZ; BRANCO, 2022). O objetivo desse enriquecimento é mitigar a escassez amostral em regiões do domínio, promovendo maior robustez nos modelos e melhor capacidade de generalização.

A implementação dos experimentos foi realizada com a linguagem *Python* (versão 3.12.8), com o uso das bibliotecas *Pandas* e *Scikit-Learn* (PEDREGOSA et al., 2011), no ambiente interativo *Jupyter Notebook* (versão 7.2.2) da distribuição *Anaconda*<sup>7</sup> (versão 24.11.1). A análise estatística foi complementada com a ferramenta *jamovi* (versão 2.6)<sup>8</sup>, conforme des-

<sup>6</sup> <https://github.com/Valquire/ATOp.git>

<sup>7</sup> <https://anaconda.org/>

<sup>8</sup> The jamovi project (2024). jamovi. (Version 2.6) [Computer Software]. Retrieved from <https://www.jamovi.org>.

crito na Seção 5.1.5.

Os experimentos foram executados em um computador pessoal com sistema operacional *Windows 11*, processador *Intel Core i7-1360P* (2,20 GHz), 32 GB de RAM e placa gráfica integrada *Intel Iris Xe Graphics*. Todas as etapas, desde a implementação dos modelos e otimização de hiperparâmetros até a validação cruzada e cálculo das métricas ( $R^2$ , RMSE) e do tempo de treinamento, foram realizadas nesse ambiente. O código-fonte encontra-se disponível no repositório oficial da pesquisa<sup>9</sup>.

O aspecto mais distintivo entre os blocos experimentais está no modo de organização dos dados de entrada. Enquanto o *Treinamento de Modelos Individuais* e o *Treinamento de Comitês sem Direcionamento Semântico* utilizaram o conjunto completo de atributos sem distinção conceitual, o *Treinamento do ATOp-PredictiveModel* adotou uma abordagem fundamentada na segmentação semântica, derivada do modelo ontológico. Parte-se da hipótese de que essa organização orientada ao conhecimento contribui para maior especialização dos modelos, promovendo melhorias significativas na capacidade preditiva e na redução do tempo de treinamento.

Ao final, realiza-se uma análise comparativa entre os modelos, utilizando métricas estatísticas e testes de significância para validar as diferenças de desempenho. Essa abordagem visa demonstrar as vantagens metodológicas da proposta *ATOp* em relação a métodos tradicionais, destacando tanto os ganhos preditivos quanto a eficiência computacional obtida com a orientação semântica.

### 6.2.1 Treinamento de Modelos Individuais

O *Treinamento de Modelos Individuais* consistiu na aplicação isolada de algoritmos de AM sobre o conjunto completo de atributos, sem qualquer direcionamento semântico oriundo da *ATOp-NavalOntology*. Seu objetivo foi estabelecer uma linha de base de poder preditivo para comparação com abordagens mais elaboradas.

Neste bloco de treinamento, foram implementados algoritmos individuais de AM, incluindo Regressão Linear (RL), Árvore de Decisão (AD), *Random Forest* (RF), *Gradient Boosting* (GB) e *K-Nearest Neighbors* (KNN). As implementações foram realizadas com o suporte da biblioteca *Scikit-Learn*, sendo que as escolhas relacionadas à otimização de cada algoritmo foram objeto de análise específica neste trabalho.

<sup>9</sup> <https://github.com/Valquire/ATOp.git>

Para assegurar um bom nível de assertividade nos resultados produzidos pelos algoritmos, empregou-se a técnica de otimização *bayesiana* (FRAZIER, 2018) na seleção de hiperparâmetros via *Optuna* (AKIBA et al., 2019). Essa abordagem utiliza métodos probabilísticos para determinar a combinação ótima de parâmetros, maximizando ou minimizando uma função objetivo. Tal função, geralmente associada a processos computacionalmente intensivos como o treinamento de modelos de AM, demanda estratégias eficientes para reduzir o custo computacional e melhorar a capacidade preditiva. A Tabela 8 apresenta os hiperparâmetros submetidos ao processo de otimização, bem como os respectivos valores padrão iniciais utilizados adotados pela biblioteca *Scikit-Learn*.

Tabela 8 – Hiperparâmetros empregados em cada modelo.

Modelo	Hiperparâmetros
Árvore de Decisão	<code>max_depth = None, min_samples_split = 2, min_samples_leaf = 1</code>
Random Forest	<code>max_depth = None, min_samples_split = 2, min_samples_leaf = 1, n_estimators = 100</code>
Gradient Boosting	<code>max_depth = 3, n_estimators = 100, learning_rate = 0.1</code>
K-Nearest Neighbors	<code>n_neighbors = 5, weights = uniform</code>

O modelo de Regressão Linear implementado pela biblioteca *Scikit-Learn* não requer ajuste de hiperparâmetros, uma vez que se trata de um método determinístico. Isso significa que, sob as mesmas condições iniciais e entradas, o modelo sempre produzirá o mesmo resultado, sem a influência de fatores aleatórios.

Para os modelos baseados em árvores, como AD, RF e GB, o hiperparâmetro *max\_depth* controla a profundidade máxima da árvore, influenciando a capacidade de aprendizado e o risco de *overfitting*. Da mesma forma, *min\_samples\_split* define o número mínimo de amostras necessárias para dividir um nó interno, enquanto *min\_samples\_leaf* estabelece a quantidade mínima de amostras que um nó folha deve conter, evitando a criação de folhas muito pequenas e melhorando a generalização.

No caso específico do *Random Forest*, um *ensemble* de árvores de decisão, o hiperparâmetro *n\_estimators* representa o número de árvores na floresta, onde um valor maior pode aumentar a robustez do modelo, mas também eleva o custo computacional. Já no GB, além de *n\_estimators*, que define o número de “*weak learners*” utilizados no *boosting*, o *le-*

*arning\_rate* controla a taxa de aprendizado do modelo, sendo que valores menores tornam o treinamento mais lento, mas favorecem a generalização.

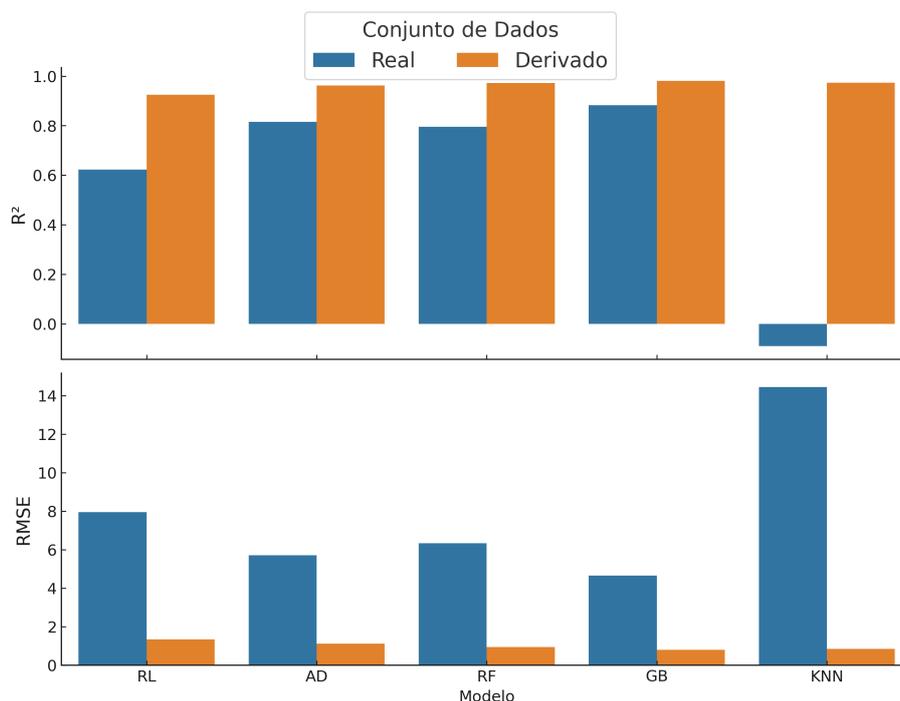
Para o *K-Nearest Neighbors*, o hiperparâmetro *n\_neighbors* define o número de vizinhos considerados na predição, onde valores pequenos tornam o modelo mais sensível a ruídos e valores maiores suavizam as predições. O hiperparâmetro *weights* determina a influência dos vizinhos na decisão, podendo ser “*uniform*”, onde todos têm o mesmo peso, ou “*distance*”, que atribui maior importância aos vizinhos mais próximos.

A técnica de otimização bayesiana foi empregada em todos os experimentos realizados neste trabalho como estratégia central para a seleção automática de hiperparâmetros. Essa escolha visou garantir uniformidade metodológica e comparabilidade entre os diferentes cenários avaliados, ao mesmo tempo em que se buscou maximizar a capacidade preditiva dos modelos. Ao utilizar esse método em todas as etapas experimentais, foi possível explorar de forma eficiente o espaço de busca de configurações, reduzindo a necessidade de intervenção manual e assegurando que cada modelo fosse avaliado em suas melhores condições operacionais.

**Resultados.** Como forma de apresentar os resultados obtidos, os índices alcançados por cada modelo durante este bloco experimental foram consolidados na Tabela 9, utilizando  $R^2$  e RMSE como métricas. Além disso, a Figura 22 ilustra a comparação dos resultados do *Treinamento de Modelos Individuais*, utilizando tanto o conjunto de dados reais quanto o conjunto de dados derivados.

Tabela 9 – Resultado do *Treinamento de Modelos Individuais*.

Modelo	Conjunto Real			Conjunto Derivado		
	$R^2$	RMSE	T (s)	$R^2$	RMSE	T (s)
RL	0,6225	7,9521	6,71	0,9253	1,3357	7,35
AD	0,8149	5,7081	<b>5,55</b>	0,9626	1,1094	4,66
RF	0,7958	6,3264	19,82	0,9725	0,9348	67,13
<b>GB</b>	<b>0,8825</b>	<b>4,6587</b>	15,24	<b>0,9807</b>	<b>0,7981</b>	120,63
KNN	-0,0899	14,4341	4,71	0,9728	0,8516	<b>4,54</b>

Figura 22 – Visualização dos resultados do *Treinamento de Modelos Individuais*.

RL: Regressão Linear, AD: Árvore de Decisão, RF: Random Forest, GB: Gradient Boosting, KNN: K-Nearest Neighbors

Os gráficos de barras agrupadas permitem observar diferenças significativas no poder preditivo dos modelos de regressão individuais. Dentre os modelos avaliados, o GB apresentou o melhor capacidade preditiva global, com  $R^2$  de 0,8825 e RMSE de 4,66 no conjunto real, e  $R^2$  de 0,9807 e RMSE de 0,7981 no conjunto derivado. Esses valores indicam que o modelo foi capaz de explicar mais de 88% da variância nos dados reais, além de manter um erro médio baixo, demonstrando boa capacidade de generalização.

Essa superioridade pode ser atribuída às características inerentes ao GB, que combina sequencialmente múltiplos modelos fracos, geralmente árvores de decisão, com o objetivo de corrigir os erros cometidos pelas iterações anteriores. Essa abordagem iterativa permite ao modelo capturar padrões complexos nos dados, mesmo em situações de não linearidade ou presença de interações entre variáveis, o que pode ser limitante para métodos mais simples como a Regressão Linear. Além disso, o GB realiza uma regularização eficaz durante o processo de treinamento, o que contribui para reduzir o sobreajuste (*overfitting*) e melhorar o desempenho preditivo em dados não vistos. A robustez e flexibilidade desse método justificam, portanto, os melhores indicadores de desempenho observados em comparação aos demais algoritmos avaliados.

### 6.2.2 Treinamento de Comitês de Modelos sem Direcionamento Semântico

O *Treinamento de Comitês de Modelos sem Direcionamento Semântico* explorou a combinação dos modelos com melhor desempenho individual, por meio de estratégias de comitê (*ensemble*), ainda utilizando o conjunto completo de atributos de forma indiferenciada, sem segmentação conceitual. Este bloco experimental teve como objetivo avaliar o impacto da combinação de algoritmos de regressão por meio da formação de comitês de modelos baseados na técnica *Stacking Regressor (SR)* (MIENYE; SUN, 2022). Para isso, foram conduzidas duas estratégias experimentais distintas, com diferentes níveis de aleatoriedade quanto à formação dos comitês e à segmentação dos dados de entrada.

Foram consideradas quatro dimensões de aleatoriedade que podem influenciar a capacidade preditiva dos comitês de modelos:

1. **Divisão do número de subconjuntos** de atributos preditivos (por exemplo, divisão em 2 ou mais subconjuntos);
2. **Escolha dos algoritmos** associados a cada subconjunto;
3. **Distribuição da quantidade de atributos** entre os subconjuntos (balanceada ou desbalanceada);
4. **Seleção dos atributos** que compõem cada subconjunto (quais colunas vão para cada parte).

A primeira abordagem adotou uma estratégia parcialmente controlada: foram consideradas todas as permutações ordenadas de pares de algoritmos tomados dois a dois entre os cinco avaliados no primeiro bloco experimental (RL, AD, RF, GB e KNN), totalizando 20 combinações distintas. Nessa abordagem, apenas a primeira dimensão de aleatoriedade foi controlada — a divisão foi sempre em dois subconjuntos de igual tamanho, correspondendo a dois modelos no comitê. As demais decisões (escolha dos algoritmos por subconjunto, quantidade e conteúdo dos atributos atribuídos a cada um) foram realizadas de forma totalmente aleatória. Para mitigar o viés gerado por essa aleatoriedade, cada configuração foi executada cinco vezes com diferentes partições, e os resultados médios foram utilizados na análise final. Essa estratégia buscou simular um cenário onde os comitês são compostos por pares bem definidos, mas com variabilidade interna na alocação dos dados.

A segunda abordagem, de caráter mais exploratório, foi executada sem qualquer direcionamento nas quatro dimensões mencionadas. Tanto o número de subconjuntos quanto a seleção dos algoritmos, a quantidade de atributos por subconjunto e a alocação específica dos atributos foram definidos aleatoriamente em cada execução. Essa abordagem teve como objetivo avaliar a robustez dos comitês frente à aleatoriedade total em sua formação, representando situações em que não há controle conceitual sobre a estrutura do modelo nem sobre a segmentação dos dados.

**Resultados.** Os resultados obtidos nas duas estratégias conduzidas no *Treinamento de Comitês de Modelos sem Direcionamento Semântico*, conforme descrito anteriormente, foram organizados e consolidados nas Tabelas 10 e 11, as quais apresentam os valores das métricas adotadas para cada configuração de comitê de modelos. A Figura 23 oferece uma representação visual dos mesmos resultados apresentados na Tabela 10, permitindo uma análise comparativa mais intuitiva dessa abordagem.

A Tabela 10 apresenta os resultados da primeira estratégia, baseada em permutações de pares de algoritmos com divisão dirigida em dois subconjuntos balanceados de atributos. Dentre os pares analisados, o comitê “*GB + AD*” apresentou o melhor poder preditivo no conjunto real, com  $R^2 = 0,8931$  e  $RMSE = 4,5840$ , além de se destacar também no conjunto derivado, atingindo  $R^2 = 0,9654$  e  $RMSE = 1,1322$ .

Tabela 10 – Resultado do *Treinamento de Comitês de Modelos sem Direcionamento Semântico* com divisão em dois subconjuntos.

Modelo	Conjunto Real			Conjunto Derivado		
	$R^2$	RMSE	T (s)	$R^2$	RMSE	T (s)
<b>GB + AD</b>	<b>0,8931</b>	<b>4,5840</b>	341,30	0,9654	1,1322	2712,66
RL + GB	0,8916	4,6299	353,88	0,9533	1,2278	3901,22
GB + RL	0,8853	4,7232	320,28	0,9544	1,2115	2619,61
AD + GB	0,8764	4,7973	292,24	0,9637	1,1474	3019,61
RF + GB	0,8735	4,9462	782,92	0,9736	1,0071	4734,77
<b>GB + KNN</b>	0,8482	5,2730	381,54	<b>0,9752</b>	<b>0,9655</b>	2577,09

Modelo	Conjunto Real			Conjunto Derivado		
	R <sup>2</sup>	RMSE	T (s)	R <sup>2</sup>	RMSE	T (s)
KNN + GB	0,8465	5,3194	474,48	0,9732	1,0006	2641,40
GB + RF	0,7972	5,8784	783,06	0,9742	1,0090	4360,46
RL + AD	0,7971	6,1172	25,15	0,9088	1,7532	138,00
AD + RF	0,7674	6,1568	524,12	0,9392	1,4416	1399,00
RF + AD	0,7667	6,4716	429,62	0,9444	1,3780	1528,81
AD + RL	0,7645	6,2161	<b>24,77</b>	0,8981	1,8613	<b>47,00</b>
KNN + AD	0,7625	6,6520	77,85	0,9635	1,2130	110,26
AD + KNN	0,7315	6,7878	94,70	0,9689	1,1145	109,55
RL + RF	0,7107	7,0320	420,74	0,9277	1,5358	6004,24
RF + RL	0,7067	7,0714	559,02	0,9317	1,4973	1583,92
KNN + RF	0,6600	7,8133	650,33	0,9606	1,2416	1607,15
RF + KNN	0,6548	7,8881	588,95	0,9611	1,2365	1687,20
RL + KNN	-0,7215	18,1176	66,46	0,7962	2,7894	89,29
KNN + RL	-0,7215	18,1176	78,75	0,7962	2,7894	87,22

Já a Tabela 11 apresenta os resultados da segunda estratégia, que adotou formações completamente aleatórias tanto na escolha dos algoritmos quanto na segmentação dos atributos. Apesar da alta variabilidade, alguns comitês apresentaram qualidade preditiva competitiva, como “*GB + KNN*”, que obteve  $R^2 = 0,8644$  no conjunto real e  $R^2 = 0,9433$  no conjunto derivado, corroborando a viabilidade de combinações aleatórias bem-sucedidas.

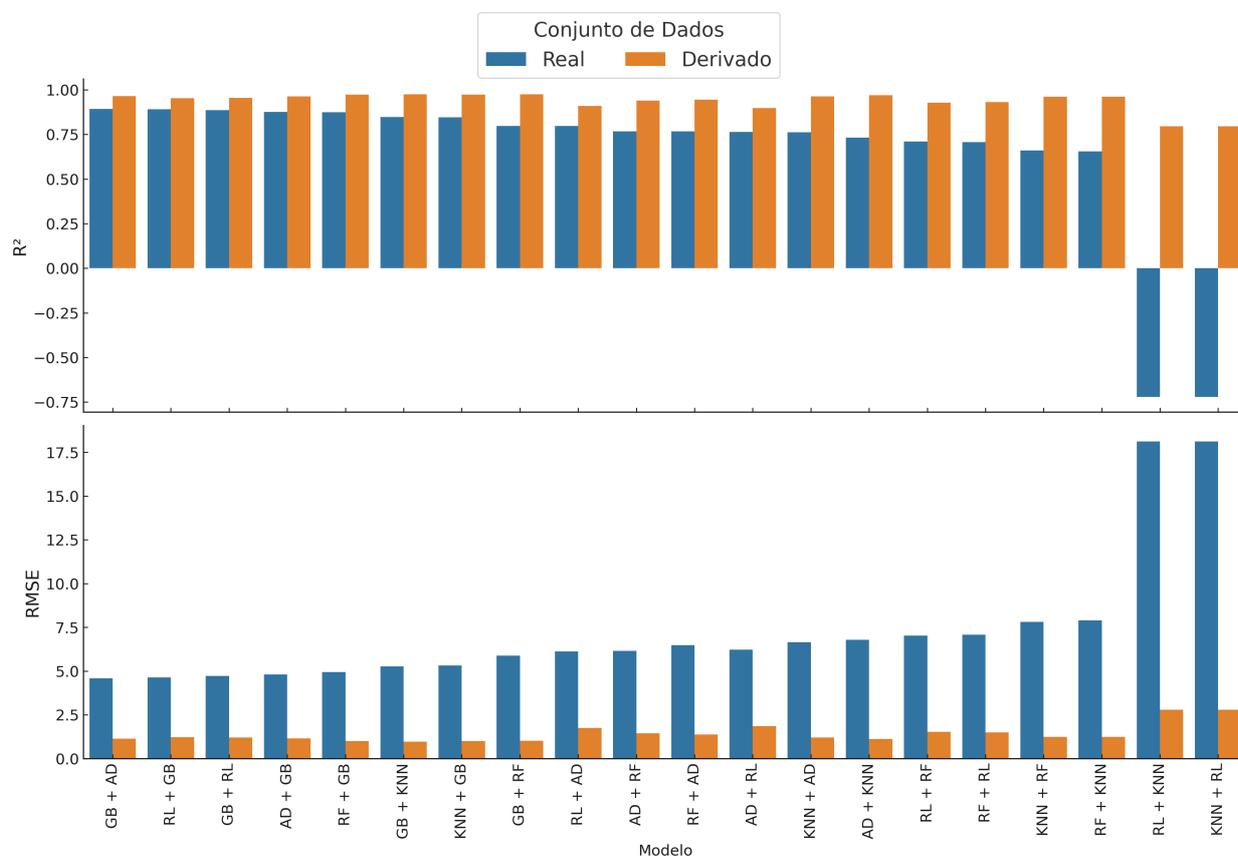
Tabela 11 – Resultado do *Treinamento de Comitês de Modelos sem Direcionamento Semântico* com formações completamente aleatórias.

Modelo	Conjunto Real			Conjunto Derivado		
	R <sup>2</sup>	RMSE	T (s)	R <sup>2</sup>	RMSE	T (s)
<b>GB + KNN</b>	<b>0,8644</b>	<b>5,0425</b>	239,34	<b>0,9433</b>	<b>1,5271</b>	2120,00
AD + KNN + RL + GB	0,8547	5,1653	226,31	0,8421	1,8033	2535,60
RL + GB + KNN	0,8347	5,5195	195,07	0,8101	2,0850	1860,58
KNN + GB + RL + AD + RF	0,7461	6,8043	463,48	0,7289	2,4849	3725,49
RF + RL + GB	0,7177	6,5290	435,01	0,6797	2,6056	2337,35
AD + RL + KNN + GB + RF	0,7109	6,9394	451,04	0,6786	2,5542	3654,13
RL + RF + KNN + GB	0,6879	6,9300	439,01	0,6626	2,6300	2817,12
RF + AD + KNN + RL	0,6707	7,5802	303,53	0,6529	2,6526	1868,68
RF + KNN	0,6699	7,8619	307,31	0,9074	1,9117	908,90
KNN + RL + GB + AD + RF	0,6664	7,4677	549,51	0,6297	2,8390	4130,04
GB + RF + KNN	0,6625	7,4930	460,92	0,9276	1,7242	2332,73
RL + RF	0,6586	7,5034	306,12	0,6196	2,8465	1691,73
KNN + RF	0,6453	7,9416	276,53	0,9041	1,9089	847,37
KNN + AD + RL + RF	0,6184	7,9627	315,31	0,6146	2,8242	1811,78
RL + KNN + RF + AD + GB	0,5490	9,0168	466,35	0,5339	3,2641	2731,94
RF + AD + RL + GB + KNN	0,5247	8,6652	490,90	0,5165	3,1322	2820,20
KNN + RL + AD + RF + GB	0,5158	8,6579	517,94	0,5135	3,1171	3004,88
AD + GB + RF + RL + KNN	0,4150	9,4159	478,61	0,4156	3,5501	2888,99

Modelo	Conjunto Real			Conjunto Derivado		
	R <sup>2</sup>	RMSE	T (s)	R <sup>2</sup>	RMSE	T (s)
RL + KNN	-0,7215	18,1176	<b>33,98</b>	0,2499	4,4254	<b>183,50</b>
KNN + RL	-0,7215	18,1176	36,28	0,2499	4,4254	185,79

Os resultados evidenciam que a aleatoriedade na formação dos comitês pode originar combinações com boa capacidade preditiva, como exemplificado pelo comitê formado por “*GB + KNN*”, que obteve o melhor resultado ( $R^2 = 0,8644$ ) para o conjunto de dados reais. Contudo, a variabilidade entre os comitês também resulta em combinações ineficientes, o que reforça a necessidade de adoção de estratégias de seleção capazes de identificar e priorizar as melhores configurações.

Figura 23 – Visualização dos resultados do *Treinamento de Comitês de Modelos sem Direcionamento Semântico* com divisão em dois subconjuntos.



RL: Regressão Linear, AD: Árvore de Decisão, RF: Random Forest, GB: Gradient Boosting, KNN: K-Nearest Neighbors

A Figura 23 apresenta a comparação gráfica das métricas  $R^2$  e RMSE obtidas nos dois conjuntos de dados, permitindo visualizar a capacidade preditiva relativo entre as diferentes configurações testadas na primeira estratégia no segundo bloco experimental. O objetivo do *Treinamento de Comitês de Modelos sem Direcionamento Semântico* consistiu em verificar se a combinação de diferentes algoritmos poderia proporcionar ganhos em termos de capacidade preditiva, em comparação com as abordagens individuais analisadas previamente. De modo geral, os resultados confirmam a superioridade de permutações que envolvem o GB.

### 6.2.3 Treinamento do *ATOp-PredictiveModel*

O *Treinamento do ATOp-PredictiveModel*, núcleo experimental deste trabalho, aplicou a categorização semântica dos atributos preditivos, baseada no mapeamento de entidades do modelo conceitual. Essa segmentação orientada permitiu treinar comitês de modelos especializados em subconjuntos específicos de dados, possibilitando uma abordagem modular e mais alinhada ao domínio, com potencial de ganho em poder preditivo e redução do tempo de treinamento. Neste experimento, que corresponde à aplicação do *ATOp-PredictiveModel* (*ATOp-PM*) desenvolvido nesta pesquisa, foram utilizados dois modelos individuais, previamente implementados no primeiro experimento, como base para a composição do comitê de modelos.

O diferencial para o comitê de modelos empregado no segundo bloco experimental reside na utilização dos subconjuntos de dados obtidos após a separação do conjunto de dados de treinamento. Nesse contexto, cada modelo individual foi treinado com os atributos correspondentes a uma destes subconjuntos. Especificamente, o primeiro modelo foi treinado utilizando exclusivamente os atributos relacionados ao subconjunto  $S_{(G2 \cap G3 \cap G4) \setminus G1} = \{COMBUSTIVEL, LUBRIFICANTE, GRAXA\}$ , enquanto o segundo modelo foi treinado com os atributos associados ao subconjunto  $S_{G1 \cap G2 \cap G3 \cap G4} = \{ANO, TRIMESTRE, DIAS\_PORTO\_SEDE, MILHAS, DIAS\_MAR, DIAS\_PORTO\_FORA\_SEDE, MILITARES, ACUCARES, ARROZ\_MASSA, BEBIDAS, CARNES, FEIJOES, FRUTAS, LEGUMES\_VERDURAS, LEITE\_QUEIJO, OLEOS\_GORDURAS, TEMPEROS\}$  (conforme definidos na Seção 6.1.4). Os atributos “ANO, TRIMESTRE, MILHAS” não foram considerados durante o treinamento, pois não contribuem para a construção do modelo. “DIAS\_MAR” é o atributo alvo, servindo como rótulo para o treinamento do modelo.

Assim, foi possível alcançar uma separação semântica implementada por meio do *ATOp-*

*PM*. Tal separação, por sua natureza, não poderia ser realizada de forma autônoma pelos algoritmos empregados. A informação derivada do modelo conceitual, devidamente representada na ontologia, foi incorporada ao processo algorítmico, permitindo a efetivação dessa separação. Os resultados obtidos, apresentados na sequência, evidenciam uma capacidade preditiva superior em relação ao experimentos que não contemplam essa separação semântica.

**Resultados.** No que diz respeito a este bloco experimental, os resultados obtidos encontram-se sistematicamente organizados na Tabela 12, que resume os valores das métricas alcançadas pelos modelos compostos segundo os critérios de avaliação definidos. Por outro lado, a Figura 24 apresenta uma visualização gráfica desses resultados, facilitando a comparação entre as diferentes configurações testadas.

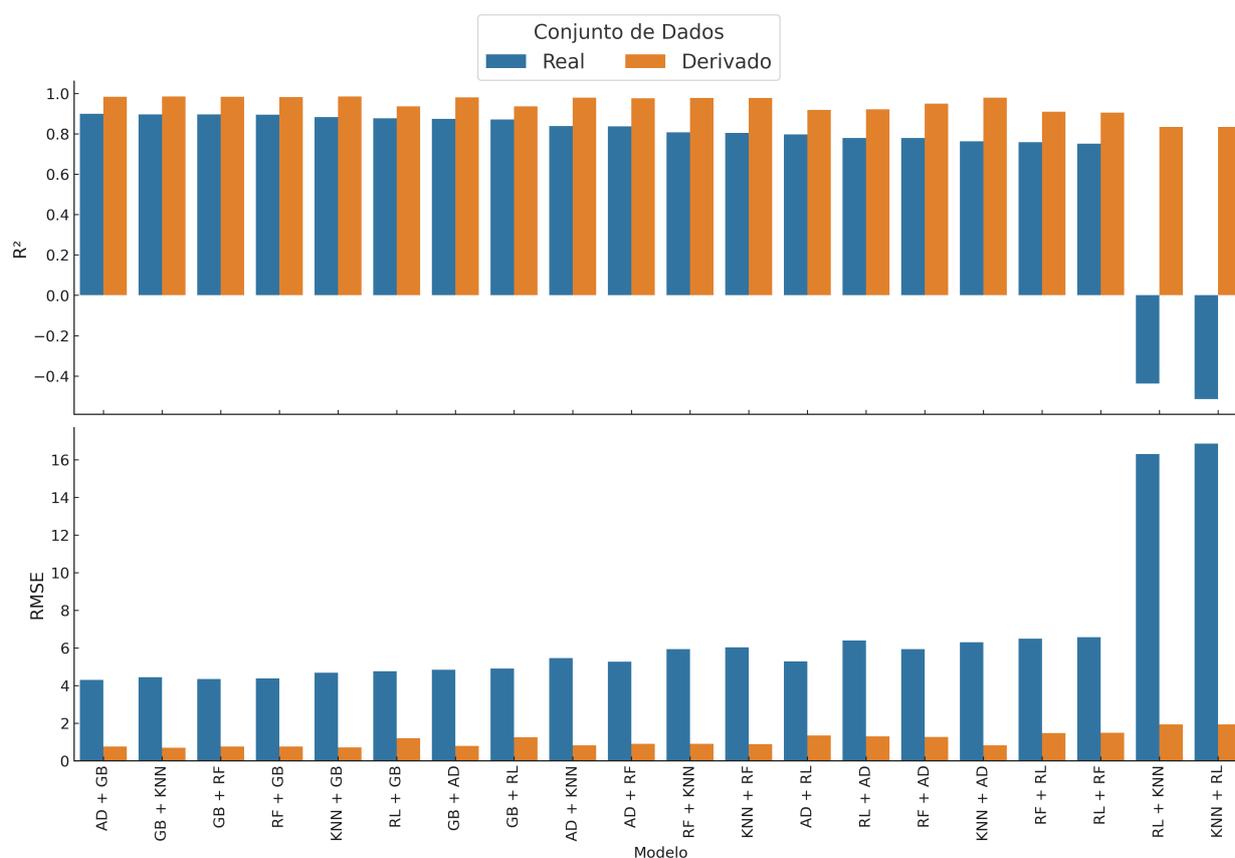
Tabela 12 – Resultado do *Treinamento do ATOp-PredictiveModel*.

Modelo	Conjunto Real			Conjunto Derivado		
	R <sup>2</sup>	RMSE	T (s)	R <sup>2</sup>	RMSE	T (s)
<b>AD + GB</b>	<b>0,8983</b>	<b>4,2958</b>	78,29	0,9831	0,7501	570,21
<b>GB + KNN</b>	0,8967	4,4315	100,87	<b>0,9853</b>	<b>0,6992</b>	594,37
GB + RF	0,8961	4,3508	188,60	0,9829	0,7548	1306,16
RF + GB	0,8950	4,3725	198,04	0,9824	0,7515	1231,41
KNN + GB	0,8833	4,6766	100,06	0,9849	0,7039	773,00
RL + GB	0,8772	4,7505	68,89	0,9367	1,2058	804,64
GB + AD	0,8745	4,8425	70,09	0,9805	0,7941	618,94
GB + RL	0,8709	4,9032	76,25	0,9361	1,2495	602,63
AD + KNN	0,8390	5,4470	20,31	0,9782	0,8160	20,61
AD + RF	0,8373	5,2574	100,71	0,9762	0,8984	433,52
RF + KNN	0,8076	5,9242	157,25	0,9774	0,8978	380,01

Modelo	Conjunto Real			Conjunto Derivado		
	R <sup>2</sup>	RMSE	T (s)	R <sup>2</sup>	RMSE	T (s)
KNN + RF	0,8039	6,0333	138,68	0,9767	0,8908	488,24
AD + RL	0,7974	5,2858	6,57	0,9175	1,3390	<b>9,79</b>
RL + AD	0,7791	6,3892	<b>5,59</b>	0,9215	1,3032	10,00
RF + AD	0,7790	5,9360	92,85	0,9490	1,2655	360,38
KNN + AD	0,7628	6,2914	17,87	0,9796	0,8224	23,16
RF + RL	0,7583	6,4797	137,62	0,9092	1,4760	245,84
RL + RF	0,7515	6,5620	98,96	0,9053	1,4810	327,34
RL + KNN	-0,4377	16,2952	15,76	0,8332	1,9260	23,99
KNN + RL	-0,5147	16,8532	17,11	0,8332	1,9260	22,12

Por fim, a Figura 24 apresenta os resultados deste bloco experimental, no qual foi avaliado a capacidade preditiva do modelo proposto, o *ATOp-PredictiveModel*, composto por comitês de modelos otimizados que integram a segmentação semântica dos dados ao processo de regressão. Assim como nos blocos experimentais anteriores, os modelos foram comparados em termos de R<sup>2</sup> e RMSE, considerando os conjuntos de dados reais e derivados. Dentre todos os pares testados, o comitê “*AD + GB*” apresentou o melhor poder preditivo no conjunto real, com R<sup>2</sup> de 0,8983 e RMSE de 4,2958, superando levemente os pares “*GB + KNN*” (R<sup>2</sup> = 0,8967) e “*GB + RF*” (R<sup>2</sup> = 0,8961). Estes três pares obtiveram valores muito próximos, com variações mínimas nas métricas de erro, o que indica que os modelos baseados em GB são os mais eficazes, mesmo com a abordagem otimizada.

No conjunto derivado, a melhor capacidade preditiva foi obtida pelo par “*GB + KNN*”, com R<sup>2</sup> de 0,9853 e RMSE de 0,6992, confirmando novamente o padrão observado nos experimentos anteriores: modelos que combinam o GB com outros algoritmos, especialmente KNN, tendem a se beneficiar da estrutura mais regular dos dados derivados. Modelos com pares assimétricos, como “*RL + KNN*” e “*KNN + RL*”, apresentaram R<sup>2</sup> negativos no conjunto

Figura 24 – Visualização dos resultados do *Treinamento do ATOp-PredictiveModel*.

RL: Regressão Linear, AD: Árvore de Decisão, RF: Random Forest, GB: Gradient Boosting, KNN: K-Nearest Neighbors

real, indicando poder preditivo inferior ao de um modelo constante, mas novamente tiveram uma capacidade preditiva aceitável no conjunto derivado ( $R^2 = 0,8332$ ), revelando a capacidade do *ATOp-PredictiveModel* de mitigar deficiências individuais através da combinação estratégica dos preditores.

Modelos como “*RL + KNN*” e “*KNN + RL*” combinaram algoritmos com capacidades preditivas distintas, sendo a RL baseada na suposição de uma relação linear entre as variáveis e o KNN dependente da distribuição espacial dos dados. Essa combinação configura um comitê assimétrico, composto por algoritmos com estratégias e limitações significativamente diferentes, o que pode resultar em conflitos internos ou ineficiências na agregação das previsões, especialmente na ausência de uma preparação adequada dos dados. No conjunto real, esses comitês apresentaram valores negativos de  $R^2$ , indicando desempenho inferior ao de um modelo que simplesmente prediz a média do atributo alvo, *DIAS\_MAR*. Esse resultado evidencia que a escolha inadequada de algoritmos compromete o desempenho do comitê, sobretudo em cenários de maior complexidade e sem segmentações orientadas por semântica.

Reforça-se, portanto, a importância de selecionar algoritmos metodologicamente compatíveis e complementares, especialmente em comitês assimétricos, nos quais abordagens conflitantes podem comprometer a eficácia do modelo composto.

Os resultados obtidos no *Treinamento do ATOp-PredictiveModel* reforçam a efetividade do modelo, que, ao incorporar a lógica semântica da divisão de atributos e explorar permutações otimizadas entre algoritmos, alcança capacidade preditiva superior aos modelos individuais, treinados no primeiro bloco experimental, e também aos comitês de modelos do segundo bloco experimental. Além disso, a abordagem mostra-se robusta à variação entre conjuntos de dados, com poder de predição adequado tanto nos dados reais quanto nos dados derivados gerados por técnicas de sobreamostragem.

#### 6.2.4 Comparação dos Modelos

Durante a construção de modelos de AM, é imprescindível comparar, de forma objetiva, a capacidade de gerar estimativas de qualidade, a fim de identificar a abordagem mais eficaz. O  $R^2$  e o RMSE são métricas complementares que permitem entender a capacidade preditiva do modelo. O  $R^2$  mostra o quanto da variação nos dados o modelo consegue explicar, enquanto o RMSE indica o erro médio das previsões. O RMSE é melhor para avaliar a qualidade das estimativas do modelo, enquanto o  $R^2$  mede o ajuste geral (CARVALHO; MENEZES; BONIDIA, 2024).

A avaliação dos modelos empregou a técnica de validação cruzada de ordem  $k$  (*k-fold cross-validation*), com  $k = 10$ . Na técnica *k-fold cross-validation*, o conjunto de dados reais é particionado em  $k$  subconjuntos de forma estratificada. A cada iteração, um desses subconjuntos é utilizado como conjunto de validação, enquanto os  $k - 1$  subconjuntos restantes são empregados no treinamento do modelo. Esse processo é repetido  $k$  vezes, garantindo que cada subconjunto seja utilizado como conjunto de validação exatamente uma vez (MORETTIN; SINGER, 2023).

**Teste Estatístico.** Para a realização da análise estatística, foram selecionados os melhores modelos identificados em cada um dos três blocos experimentais conduzidos ao longo desta Seção. Essa seleção teve como critério principal o desempenho obtido nas métricas  $R^2$  e RMSE, de modo a garantir uma comparação representativa entre os modelos individuais (primeiro bloco experimental - Seção 6.2.1), os comitês de modelos (segundo bloco experimental - Seção 6.2.2) e o modelo proposto *ATOp-PredictiveModel* (terceiro bloco experimental -

Seção 6.2.3).

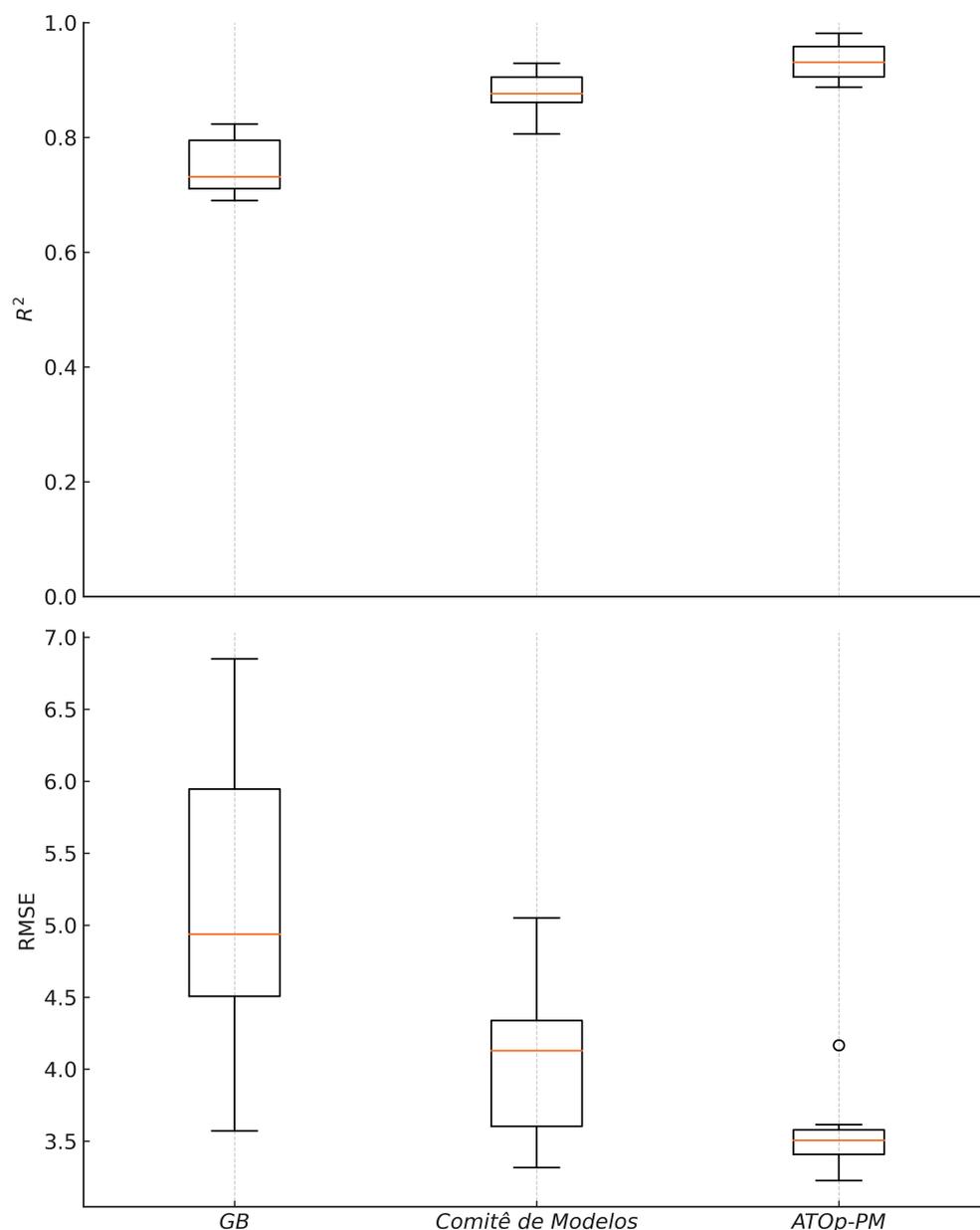
Especificamente, o modelo GB foi escolhido como representante do primeiro bloco experimental, enquanto o comitê de modelos baseado na combinação dos algoritmos “*GB + AD*” foi selecionado para o segundo bloco. Já o terceiro bloco experimental foi representado pelo modelo *ATOp-PredictiveModel* configurado com a combinação de algoritmos “*AD + GB*”, porém estruturado com base na segmentação semântica dos atributos. A utilização dos modelos de melhor desempenho em cada bloco visa assegurar que as conclusões estatísticas reflitam, de forma justa e rigorosa, o potencial máximo de cada abordagem avaliada.

Para garantir a validade estatística das diferenças observadas entre os melhores modelos de cada experimento, foi aplicado o teste de *Friedman* (conforme Seção 3.3.3) sobre os resultados obtidos em 10 *folds* de validação cruzada para o conjunto de dados reais, considerando as métricas  $R^2$  e RMSE. Para ambas as métricas, os resultados indicaram a rejeição da hipótese nula de capacidades preditivas equivalentes ( $R^2$ : estatística de 10,400,  $p$ -valor = 0,0055; RMSE: estatística de 10,400,  $p$ -valor = 0,0055). Esses valores evidenciam diferenças estatisticamente significativas entre os modelos avaliados, justificando a realização de testes *post-hoc* para identificar quais pares de modelos diferem significativamente.

A Figura 25 apresenta os *boxplots* correspondentes às métricas  $R^2$  e RMSE para os modelos avaliados no conjunto de dados reais. Em termos de capacidade preditiva, o modelo *ATOp-PM* (demonstrou os melhores resultados, com a maior mediana de  $R^2$  (0,921) e o menor valor de RMSE (3,45) entre as abordagens analisadas. O comitê de modelos “*GB + AD*” apresentou uma capacidade preditiva intermediária, com mediana de  $R^2$  igual a 0,893 e de RMSE igual a 4,05, situando-se próximo ao *ATOp-PM* “*AD + GB*”. Já o modelo *GB* obteve a menor mediana de  $R^2$  (0,765) e o maior valor de RMSE (4,95), indicando menor poder preditivo em comparação aos demais.

Com o intuito de identificar quais pares de modelos apresentaram diferenças estatisticamente significativas, foi aplicado o teste *post-hoc* de *Durbin-Conover* (POHLERT, 2014). Os resultados revelaram que o modelo individual *GB* difere significativamente tanto do comitê de modelos baseado na combinação “*GB + AD*” ( $p$ -valor = 0,007), quanto do modelo *ATOp-PredictiveModel*, configurado com a combinação “*AD + GB*” ( $p$ -valor < 0,001). Por outro lado, não foi observada diferença estatisticamente significativa entre o comitê de modelos “*GB + AD*” e o modelo *ATOp-PredictiveModel* “*AD + GB*” ( $p$ -valor = 0,236), para ambas as métricas avaliadas ( $R^2$  e RMSE), conforme apresentado na Tabela 13.

Figura 25 – Distribuição das métricas  $R^2$  e RMSE para os modelos avaliados no conjunto real.



Observou-se que os resultados das comparações par a par pelo teste de *Durbin-Conover* foram idênticos tanto para as métricas  $R^2$  quanto para RMSE. Esse comportamento uniforme reforça a robustez das diferenças de capacidade preditiva detectadas entre os modelos, indicando que a superioridade de determinadas abordagens não depende da métrica de avaliação utilizada, mas reflete uma melhoria geral no processo de predição da autonomia. Embora o teste estatístico de *Durbin-Conover* tenha indicado a existência de uma diferença estatisticamente significativa entre o modelo *GB* e as demais abordagens avaliadas — mas não entre o Comitê de Modelos e o *ATOp-PredictiveModel* —, a análise integrada dos *boxplots* com os

resultados dos testes estatísticos reforça a conclusão de que a abordagem *ATOp* proporciona ganhos relevantes no que se refere à capacidade preditiva.

Tabela 13 – Resultado do teste *post-hoc Durbin-Conover* para  $R^2$  e RMSE

Comparação	Estatística	p-valor
GB x Comitê de Modelos ( <i>GB + AD</i> )	3,06	0,007
GB x <i>ATOp-PM</i> ( <i>AD + GB</i> )	4,29	< 0,001
Comitê de Modelos ( <i>GB + AD</i> ) x <i>ATOp-PM</i> ( <i>AD + GB</i> )	1,22	0,236

Ressalta-se que, no presente estudo, a configuração dos modelos que compõem o comitê de modelos *GB + AD* foi deliberadamente ajustada para se aproximar da lógica de segmentação semântica adotada na abordagem *ATOp-PM*, especialmente no que se refere à quantidade de subconjuntos de separação dos atributos preditivos. Caso a formação desse comitê de modelos tivesse se baseado apenas em divisões aleatórias, sem esse direcionamento conceitual, é razoável supor que o comportamento estatístico observado poderia ter sido diferente, possivelmente ampliando a distância de poder preditivo entre o comitê de modelos “*GB + AD*” e o *ATOp-PM* “*AD + GB*”. Assim, o alinhamento intencional nas configurações reforça a validade da comparação realizada, além de evidenciar a robustez dos ganhos metodológicos promovidos pela estruturação semântica da abordagem *ATOp*.

**Conclusões.** De maneira geral, comparando os três blocos experimentais conduzidos, verifica-se que os modelos individuais (Seção 6.2.1), embora apresentem capacidade preditiva satisfatória em alguns casos — como no resultado isolado do GB —, não superaram os resultados alcançados pelas abordagens baseadas em comitê de modelos (Seção 6.2.2) e, especialmente, pela abordagem proposta *ATOp-PredictiveModel* (Seção 6.2.3). O terceiro bloco experimental (*ATOp-PredictiveModel*) se destacou por atingir, em múltiplas combinações de pares de modelos, os maiores valores de coeficiente de determinação ( $R^2$ ) e os menores valores de erro quadrático médio (RMSE), especialmente no conjunto de dados derivados — que, por apresentar maior volume e diversidade, forneceu mais exemplos para o aprendizado e uma distribuição mais representativa dos padrões do domínio. Esses resultados evidenciam o impacto positivo da estruturação semântica dos dados, promovida pela ODCM, sobre a capacidade de generalização dos modelos de regressão.

Além da capacidade preditiva, o tempo de treinamento dos modelos constitui uma medida importante para avaliar a viabilidade prática das abordagens propostas. Conforme

apresentado na Tabela 12, o *ATOp-PredictiveModel* demonstrou não apenas melhores valores de  $R^2$  e RMSE, mas também tempos de treinamento significativamente inferiores em comparação ao comitê de modelos do treinado no segundo bloco experimental, baseado em divisões aleatórias de atributos.

Como exemplo, o par “*RF + GB*”, que obteve excelentes resultados em ambos os experimentos, exigiu aproximadamente 782 segundos no segundo bloco experimental (Seção 6.2.2), contra apenas 198 segundos no terceiro bloco experimental (*ATOp-PredictiveModel* - Seção 6.2.2). Esse padrão de economia de tempo se repete em diversos outros pares de modelos, mesmo quando mantido o mesmo algoritmo de base.

A economia computacional proporcionada pela abordagem *ATOp* pode ser atribuída à segmentação semântica dos dados promovida pela ODCM. Ao invés de realizar divisões aleatórias dos atributos preditivos, o *ATOp-PredictiveModel* utiliza subconjuntos guiados conceitualmente, compostos por atributos com coerência semântica. Essa organização semântica favorece a aprendizagem por parte dos modelos, ao reduzir redundâncias e ruídos informacionais.

Dessa forma, além de melhorar a qualidade preditiva, a abordagem *ATOp* também se mostra mais eficiente em termos computacionais. Essa característica é especialmente relevante em cenários reais que exigem predições rápidas e robustas, como aplicações no domínio militar, onde decisões oportunas podem depender da agilidade e confiabilidade dos sistemas preditivos.

Em síntese, os resultados obtidos nesta pesquisa demonstram que a incorporação de aspectos semânticos por meio da MC orientada por ontologias contribui não apenas para a melhoria do poder preditivo dos modelos, mas também para a redução do tempo de treinamento, evidenciando a efetividade da abordagem *ATOp*. Tais achados reforçam a viabilidade de estratégias baseadas em ODCM como mecanismos complementares ao AM, especialmente em domínios complexos e sensíveis à qualidade dos dados, como o setor militar naval. Esses resultados consolidam as bases para as discussões que se seguem, onde serão exploradas as implicações práticas e teóricas da abordagem proposta.

### 6.3 CASO DE APLICAÇÃO

O caso de aplicação, que teve como base as atividades do Macroprocesso 2 (Seção 5.2) da abordagem *ATOp*, visa o emprego do *ATOp-PredictiveModel* para o apoio no planeja-

mento e execução da operação conjunta hipotética denominada “*TYPHOON*”. A operação “*TYPHOON*” tem como objetivo assegurar a mobilidade e permanência das unidades navais em um cenário de defesa marítima, garantindo a capacidade de sustentação prolongada das forças no teatro de operações. Para isso, é necessário empregar plataformas navais dos tipos fragata, apoio logístico e aeronaves de patrulha marítima, que operam sob restrições específicas de suprimentos e autonomia de navegação. O sucesso da operação depende da eficiente alocação desses meios para manter a sustentação da operação sem comprometer a disponibilidade de recursos essenciais, como combustível e alimentos.

O Comando Conjunto (CCj) de operações precisa decidir quais meios devem ser empregados para cada fase da missão, levando em consideração a autonomia de cada meio operativo e a necessidade de reabastecimento ao longo da operação. Para apoiar essa decisão, foi solicitado o cálculo do indicador de autonomia de meios operativos, que avalia o tempo máximo que cada unidade pode operar antes de exigir reabastecimento, considerando fatores como consumo médio de suprimentos e perfil da operação.

A tomada de decisão é auxiliada pelo Sistema Militar de Comando e Controle (SISMC<sup>2</sup>), que permite o intercâmbio seguro de informações entre os centros de comando e os meios operacionais envolvidos (conforme descrito na Seção 3.1.3). O SISMC<sup>2</sup> fornece dados estratégicos e operacionais, garantindo que o CCj tenha uma visão consolidada da autonomia de cada unidade e possa definir o melhor curso de ação.

### 6.3.1 Aplicação do *ATOp-PredictiveModel*

A abordagem *ATOp* foi aplicada na Operação Conjunta *TYPHOON* para a *MilitaryNavalPlatform* com o objetivo de estimar a autonomia dos meios operativos empregados nas operações navais, utilizando uma modelagem preditiva orientada por ontologia. O modelo *ATOp-PredictiveModel* fundamenta-se em dados históricos de consumo de suprimentos e em informações sobre a atuação das plataformas navais em operações militares, com o intuito de prever sua capacidade de sustentação ao longo do tempo, considerando aspectos semânticos relevantes extraídos por meio da estrutura ontológica. Para isso, os seguintes dados de entrada foram considerados:

- **Perfis operacionais de plataformas navais militares:** tempo médio de operação, consumo de combustível, consumo de alimentos e quantidade de militares embarcados;

- **Dados históricos de operações anteriores:** relatórios de operações e registros de abastecimento.

Os dados de saída produzidos pelo *ATOp-PredictiveModel* fornecem estimativas da autonomia operacional de cada meio empregado na operação. Essas estimativas possibilitam uma análise detalhada do tempo de operação disponível antes da necessidade de reabastecimento, considerando diferentes variáveis logísticas. A Figura 26 apresenta a representação gráfica de um protótipo de sistema, que exhibe os resultados obtidos a partir do *ATOp-PredictiveModel* neste caso de aplicação.

O protótipo do *ATOp-PredictiveModel* está implementado na camada de interface utilizando a biblioteca do *Framework Streamlit*<sup>10</sup>, a qual viabiliza o desenvolvimento da camada de apresentação com suporte à execução em qualquer aplicativo de navegação (*browser*). Essa abordagem proporciona uma interface interativa e acessível, permitindo a visualização e manipulação dos dados de forma intuitiva e eficiente.

A operacionalização vislumbrada da abordagem *ATOp*, realizada em conjunto com o SISMC<sup>2</sup>, integra informações provenientes do Sistema de Planejamento Operacional Militar (SIPLOM). O SIPLOM fornece um panorama atualizado da disponibilidade de meios e permite a análise de inteligência logística e das estimativas operacionais.

A integração com o SISMC<sup>2</sup> garante que as estimativas de autonomia geradas pela *ATOp* sejam automaticamente compartilhadas com os Centros de Comando e Controle (CC<sup>2</sup>) responsáveis pela coordenação das operações navais e aéreas. Dessa forma, os planejadores podem simular diferentes cenários de alocação de recursos e identificar o impacto de cada decisão antes de sua execução.

### 6.3.2 Apoio à Tomada de Decisão na Operação Conjunta *TYPHOON*

Com base nos resultados do modelo *ATOp* e na análise fornecida pelo SISMC<sup>2</sup> para cada uma das plataformas navais militares participantes, o Comandante da Operação Conjunta poderia tomar as seguintes decisões estratégicas:

1. **Redistribuição das rotas de patrulha:** para minimizar o consumo excessivo de combustível e evitar períodos críticos de baixa autonomia.

---

<sup>10</sup> <https://streamlit.io/>

Figura 26 – Tela do protótipo de modelo preditivo da autonomia.



2. **Revisão do plano de reabastecimento em alto-mar:** para ajustar os intervalos de apoio logístico para manter a prontidão operacional das plataforma navais militares.
3. **Priorização de meios de maior autonomia:** para realizar operações mais longas e alocação de meios de menor autonomia em áreas de menor exigência operacional.
4. **Ajuste na composição dos meios do tipo fragata:** garantindo que cada grupo tático tenha suporte adequado em termos de suprimentos e capacidade de resposta.

A abordagem *ATOp*, operacionalizada através do *SISMC<sup>2</sup>*, permite um planejamento mais preciso e eficiente, reduzindo custos operacionais e risco de falha logística durante a operação. Além disso, a modelagem baseada em ontologias possibilita a padronização das decisões estratégicas, garantindo maior interoperabilidade e rastreabilidade das ações realizadas durante a operação.

Além desse caso de aplicação, a abordagem proposta apresenta diversas possibilidades de ampliação e aplicação futura. Por exemplo, o *ATOp-PredictiveModel* pode ser estendido para prever demandas de manutenção preventiva de plataformas e equipamentos militares, otimizando cronogramas de reparo e reduzindo o risco de falhas inesperadas. Outra aplicação potencial envolve a adaptação do modelo para estimar a eficiência de diferentes estratégias logísticas em operações prolongadas, auxiliando na escolha das rotas mais seguras e econômicas para o transporte de suprimentos.

#### 6.4 AMEAÇAS À VALIDADE

Embora a aplicação da abordagem *ATOp* tenha apresentado resultados positivos na predição da autonomia de plataformas navais militares, é necessário reconhecer algumas ameaças à validade que podem ter influenciado os resultados obtidos. Uma limitação relevante refere-se à correspondência entre os atributos presentes nos dados integrados e as entidades modeladas na *ATOp-NavalOntology*. Esse mapeamento foi realizado com base em análise empírica e conhecimento do domínio, porém não passou por uma validação formal ou automatizada. A ausência de mecanismos sistemáticos para validar tais associações indica a necessidade de investigações adicionais, a fim de verificar em que medida essa etapa pode influenciar a coerência semântica da segmentação dos dados e o desempenho preditivo do modelo.

Além disso, os experimentos realizados neste estudo utilizaram exclusivamente dados históricos de plataformas navais da Marinha do Brasil, o que limita a generalização dos resultados para outros meios operativos, como unidades aéreas ou terrestres. A aplicação da abordagem *ATOp* em novos contextos exigiria a reformulação da ontologia e a adaptação da estratégia preditiva para refletir as especificidades de cada domínio. Outro fator que merece atenção é a dependência da abordagem em relação ao conhecimento especializado. A elaboração da ontologia e a definição dos agrupamentos semânticos dependem fortemente da participação de especialistas do domínio, o que pode dificultar a replicação da metodologia em ambientes com menor disponibilidade de expertise.

Ainda no que diz respeito à ontologia bem fundamentada, cabe destacar que a segmentação dos dados baseada em situações ontológicas, embora tecnicamente fundamentada, representa abstrações que podem não capturar toda a complexidade do ambiente operacional real. O nível de granularidade adotado na ontologia busca equilibrar o rigor semântico com a

viabilidade prática de aplicação, mas isso pode ocultar nuances relevantes do comportamento logístico em cenários mais dinâmicos.

Tais limitações abrem espaço para investigações futuras que explorem, por exemplo, mecanismos automatizados de validação do mapeamento entre atributos e entidades ontológicas, além da aplicação da abordagem *ATOp* em outros domínios militares. Também é possível avançar na definição de métricas específicas para aferição da coerência semântica entre os dados e a ontologia, bem como na integração da abordagem com sistemas de Comando e Controle operando com dados em tempo real. A superação dessas ameaças à validade pode contribuir para a consolidação da abordagem *ATOp* como um modelo robusto e replicável em contextos operacionais diversos.

## 7 CONCLUSÃO

Esta dissertação apresentou a abordagem *ATOp*, que promove a convergência entre uma ontologia bem fundamentada com técnicas de AM, com o objetivo de apoiar a tomada de decisão em cenários militares sustentados por sistemas de C<sup>2</sup>. Fundamentada na UFO e modelada com OntoUML, a *ATOp* busca mitigar os efeitos da heterogeneidade semântica dos dados operacionais, promovendo uma segmentação coerente por meio da metacategoria *Situation*.

A hipótese de que a modelagem semântica eleva a capacidade preditiva foi testada por meio de experimentos comparativos e validados estatisticamente, nos quais o modelo *ATOp-PredictiveModel* apresentou capacidade preditiva superior às abordagens isoladas tanto em termos de qualidade das predições, sendo avaliados R<sup>2</sup> e RMSE, quanto de custo computacional, com a medição do tempo de treinamento. Os resultados indicam que a combinação entre conhecimento de domínio, explicitado por meio da *ATOp-NavalOntology*, e especialização algorítmica aprimora o poder de predição em domínios complexos como o militar.

A aplicação prática da abordagem foi realizada com base em dados históricos da Marinha do Brasil, focando na Plataforma Naval Militar. Os dados foram extraídos de fontes estruturadas, como o sistema SINGRA, e de fontes não estruturadas, como os Relatórios de Fim de Comissão, sendo posteriormente integrados e preparados para análise. A ontologia *ATOp-NavalOntology*, desenvolvida para representar os conceitos centrais do domínio da autonomia, forneceu a base para a segmentação semântica dos dados, orientando a construção de subconjuntos especializados de atributos e viabilizando a aplicação do modelo preditivo no contexto da operação simulada *TYPHOON*.

Os objetivos específicos definidos no início deste trabalho foram integralmente alcançados. O primeiro, referente ao desenvolvimento de uma ontologia bem fundamentada que representasse os principais conceitos relacionados à autonomia de meios operativos no contexto militar, foi atendido por meio da criação da *ATOp-NavalOntology*, baseada na UFO e construída com apoio da metodologia SABiO. O segundo, relacionado à definição de uma trajetória de ciência de dados que incorporasse a ontologia ao processo de treinamento de modelos preditivos, também foi plenamente cumprido, culminando na formulação e execução do *ATOp-PredictiveModel*.

## 7.1 CONTRIBUIÇÕES

A dissertação contribui de forma relevante tanto para o avanço teórico quanto para a aplicação prática da integração entre ontologias bem fundamentadas e AM. A principal contribuição consiste na proposição da abordagem *ATOp*, que estrutura semanticamente dados operacionais e os integra a um pipeline preditivo baseado em comitês de modelos, otimizados por meta-aprendizagem. O desenvolvimento da *ATOp-NavalOntology*, alinhada à UFO e construída segundo boas práticas de engenharia de ontologias, representa um artefato reutilizável e adaptável a outros domínios militares. O modelo preditivo gerado, o *ATOp-PredictiveModel*, demonstrou superioridade estatística em relação a modelos isolados, confirmando a eficácia da segmentação semântica baseada em metacategorias ontológicas.

A pesquisa também apresenta contribuições práticas, ao propor uma solução aplicável ao SISMC<sup>2</sup>, em especial ao SIPLOM, com potencial para fortalecer a capacidade analítica e a flexível das decisões logísticas no nível estratégico-operacional. A estrutura modular e replicável da abordagem permite sua adaptação para contextos civis e industriais que enfrentem desafios semelhantes em termos de heterogeneidade de dados e necessidade de previsões adequadas.

## 7.2 CONSIDERAÇÕES E LIMITAÇÕES

Os resultados alcançados indicam que a abordagem *ATOp* representa uma estratégia promissora para aprimorar a qualidade de previsões em domínios logísticos complexos, como o das operações militares. A integração entre ontologias bem fundamentadas e AM demonstrou-se eficaz na organização semântica dos dados, na especialização de modelos e aprimoramento das estimativas.

Entretanto, a presente pesquisa apresenta algumas limitações. O estudo foi conduzido com base em dados de uma única plataforma naval, referentes a um período específico, 2013 a 2017, o que restringe a generalização dos resultados. Além disso, o acesso aos dados foi condicionado a autorizações institucionais, o que limitou o compartilhamento público e a replicabilidade ampla dos experimentos. A associação entre entidades ontológicas e atributos dos dados foi realizada de forma empírica, com base em conhecimento especializado, o que pode introduzir viés ou limitar a escalabilidade do processo. A abordagem ainda não foi testada em contextos interforças, Exército, Marinha e Força Aérea, nem validada em sistemas

de C<sup>2</sup> operacionais em tempo real, o que constitui uma lacuna a ser explorada em futuras pesquisas.

### 7.3 TRABALHOS FUTUROS

Com base nos resultados obtidos e nas limitações identificadas, diversos caminhos se abrem para a continuidade desta linha de pesquisa. A ampliação da aplicação da *ATOp* para os domínios terrestre e aéreo é uma das direções mais promissoras, o que implica a especialização da ontologia para esses contextos por meio da criação das variantes *ATOp-LandOntology* e *ATOp-AerialOntology*. Outro avanço relevante será a integração da *ATOp* em sistemas de C<sup>2</sup> operacionais, permitindo sua execução em tempo real como um serviço preditivo adaptável.

Além disso, recomenda-se a automação do processo de mapeamento entre dados e entidades ontológicas, com uso de técnicas de Processamento de Linguagem Natural (PLN) para extração de conhecimento a partir de documentos técnicos, como os relatórios operacionais. Também se destaca a necessidade de incorporar mecanismos de adaptação dinâmica dos modelos à medida que novos dados forem disponibilizados, ampliando a resiliência e a capacidade de generalização da abordagem. Por fim, o desenvolvimento de protótipos baseados em arquiteturas orientadas a serviços, utilizando tecnologias da Web Semântica, poderá viabilizar a interoperabilidade da *ATOp* com sistemas diversos e consolidá-la como uma solução de apoio inteligente à decisão no contexto da Defesa.

## REFERÊNCIAS

- AKIBA, T. et al. Optuna: A next-generation hyperparameter optimization framework. In: **Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. [S.l.: s.n.], 2019.
- ALMEIDA, J. P. A.; FALBO, R. A.; GUIZZARDI, G. Events as entities in ontology-driven conceptual modeling. In: **Conceptual Modeling: 38th International Conference, ER 2019, Salvador, Brazil, November 4–7, 2019, Proceedings 38**. Cham: Springer International Publishing, 2019. p. 469–483. ISBN 978-3-030-33223-5.
- AMARAL, G.; BAIÃO, F.; GUIZZARDI, G. Foundational ontologies, ontology-driven conceptual modeling, and their multiple benefits to data mining. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 11, n. 4, p. e1408, 2021.
- ARAYA, D.; KING, M. **The impact of artificial intelligence on military defence and security**. Waterloo, ON, Canada, 2022. CIGI Papers. Disponível em: <https://hdl.handle.net/10419/299735>.
- ARP, R.; SMITH, B.; SPEAR, A. D. **Building ontologies with basic formal ontology**. [S.l.]: Mit Press, 2015.
- BARATELLA, R. et al. Understanding and modeling prevention. In: SPRINGER. **International Conference on Research Challenges in Information Science**. [S.l.], 2022. p. 389–405.
- BERNASCONI, A. et al. Semantic interoperability: ontological unpacking of a viral conceptual model. **BMC bioinformatics**, Springer, v. 23, n. Suppl 11, p. 491, 2022.
- BIAU, G.; SCORNET, E. A random forest guided tour. **Test**, Springer, v. 25, p. 197–227, 2016.
- BOOCH, G. et al. The unified modeling language. **Unix Review**, v. 14, n. 13, p. 5, 1996.
- BORDES, A. et al. Translating embeddings for modeling multi-relational data. **Advances in neural information processing systems**, v. 26, 2013.
- BORK, D.; ALI, S. J.; ROELEN, B. Conceptual modeling and artificial intelligence: A systematic mapping study. **arXiv preprint arXiv:2303.06758**, 2023.
- BRANCO, P.; TORGO, L.; RIBEIRO, R. P. Pre-processing approaches for imbalanced distributions in regression. **Neurocomputing**, Elsevier, v. 343, p. 76–99, 2019.
- BRASIL, E.-M. C. d. F. A. **MD31-M-03 - Doutrina para o Sistema Militar de Comando e Controle**. 3<sup>a</sup>. ed. Brasília, DF: Ministério da Defesa, 2015. 48 p. Disponível em: [https://www.gov.br/defesa/pt-br/arquivos/doutrina\\_militar/lista\\_de\\_publicacoes/md31a\\_ma\\_03a\\_douta\\_sismca\\_3a\\_eda\\_2015.pdf](https://www.gov.br/defesa/pt-br/arquivos/doutrina_militar/lista_de_publicacoes/md31a_ma_03a_douta_sismca_3a_eda_2015.pdf).

- BRASIL, E.-M. C. d. F. A. **MD31-S-02 - Conceito de Operações do Sistema Militar de Comando e Controle**. 1<sup>a</sup>. ed. Brasília, DF: Ministério da Defesa, 2015. 66 p. Disponível em: [https://www.gov.br/defesa/pt-br/arquivos/ajuste-01/doutrina\\_militar/lista\\_de\\_publicacoes/md31a\\_sa\\_02a\\_conopsa\\_sismca\\_1a\\_eda\\_2015.pdf](https://www.gov.br/defesa/pt-br/arquivos/ajuste-01/doutrina_militar/lista_de_publicacoes/md31a_sa_02a_conopsa_sismca_1a_eda_2015.pdf).
- BRASIL, E.-M. C. d. F. A. **MD35-G-01 - Glossário das Forças Armadas**. 5<sup>a</sup>. ed. Brasília, DF: Ministério da Defesa, 2015. 294 p. Disponível em: <https://www.gov.br/defesa/pt-br/arquivos/ajuste-01/legislacao/emcfa/publicacoes/doutrina/md35-G-01-glossario-das-forcas-armadas-5-ed-2015-com-alteracoes.pdf>.
- BRASIL, E.-M. C. d. F. A. **MD42-M-02 - Doutrina de Logística Militar**. 3<sup>a</sup>. ed. Brasília, DF: Ministério da Defesa, 2016. 42 p. Disponível em: [https://www.gov.br/defesa/pt-br/arquivos/doutrina\\_militar/lista\\_de\\_publicacoes/md42-m-02-dout-log-mil-3a-ed-2016.pdf](https://www.gov.br/defesa/pt-br/arquivos/doutrina_militar/lista_de_publicacoes/md42-m-02-dout-log-mil-3a-ed-2016.pdf).
- BRASIL, E.-M. C. d. F. A. **MD30-M-01 - Doutrina de Operações Conjuntas**. 2<sup>a</sup>. ed. Brasília, DF: Ministério da Defesa, 2020. v. 1. 240 p. Disponível em: <https://www.gov.br/defesa/pt-br/arquivos/legislacao/emcfa/publicacoes/doutrina/md30-m-01-vol-1-2a-edicao-2020-dou-178-de-15-set.pdf>.
- BREIMAN, L. Bagging predictors. **Machine learning**, Springer, v. 24, p. 123–140, 1996.
- BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001.
- BREIMAN, L. et al. **Classification and Regression Trees**. 1st. ed. [S.l.]: Chapman & Hall/CRC, 1984. 368 p. ISBN 9780412048418.
- CARVALHO, A. de; MENEZES, A.; BONIDIA, R. **Ciência de Dados: Fundamentos e Aplicações**. 1<sup>a</sup>. ed. Rio de Janeiro, RJ: LTC, 2024. 376 p. ISBN 9788521638759.
- CHEN, P. P.-S. The entity-relationship model—toward a unified view of data. **ACM Trans. Database Syst.**, Association for Computing Machinery, New York, NY, USA, v. 1, n. 1, p. 9–36, mar 1976. ISSN 0362-5915. Disponível em: <https://doi.org/10.1145/320434.320440>.
- CHEN, S.; LUC, N. M. Rmse voting regressor: A weighting function based improvement to ensemble regression. **arXiv preprint arXiv:2207.04837**, 2022.
- COSTE, M. D.; LI, Z.; KHEDRI, R. A hybrid ontology-based semantic and machine learning model for the prediction of spring breakup. **Computer-Aided Civil and Infrastructure Engineering**, Wiley Online Library, 2023.
- DEMORI, A. et al. A semantic web approach for military operation scenarios development for simulation. In: INSTICC. **Proceedings of the 12th International Conference on Data Science, Technology and Applications - DATA**. [S.l.]: SciTePress, 2023. p. 390–397. ISBN 978-989-758-664-4. ISSN 2184-285X.
- DEMŠAR, J. Statistical comparisons of classifiers over multiple data sets. **Journal of Machine learning research**, v. 7, n. Jan, p. 1–30, 2006.
- FACELI, K. et al. **Inteligência Artificial: uma abordagem de Aprendizado de Máquina**. 2<sup>a</sup>. ed. Rio de Janeiro, RJ: LTC, 2021. 400 p. ISBN 9788521637493.
- FALBO, R. de A. Sabio: Systematic approach for building ontologies. **Onto. Com/odise@Fois**, v. 1301, 2014.

- FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. Methontology: from ontological art towards ontological engineering. **American Association for Artificial Intelligence**, 1997.
- FRAZIER, P. I. A tutorial on bayesian optimization. **arXiv preprint arXiv:1807.02811**, 2018.
- FRIEDMAN, J. H. Greedy function approximation: a gradient boosting machine. **Annals of statistics**, JSTOR, p. 1189–1232, 2001.
- GRAND-CLÉMENT, S. **Artificial intelligence beyond weapons: Application and impact of AI in the military domain**. [S.l.]: UNIDIR Report. Geneva, 2023.
- GROVER, A.; LESKOVEC, J. node2vec: Scalable feature learning for networks. In: **Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.: s.n.], 2016. p. 855–864.
- GUARINO, N. **Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy**. [S.l.]: IOS press, 1998. v. 46.
- GUARINO, N.; GUIZZARDI, G.; MYLOPOULOS, J. On the philosophical foundations of conceptual models. **Information Modelling and Knowledge Bases**, v. 31, n. 321, p. 1, 2020.
- GUIZZARDI, G. **Ontological foundations for structural conceptual models**. Tese (Doutorado) — University of Twente, oct 2005.
- GUIZZARDI, G. Ontological foundations for conceptual modeling with applications. In: SPRINGER. **Advanced Information Systems Engineering: 24th International Conference, CAiSE 2012, Gdansk, Poland, June 25-29, 2012. Proceedings 24**. [S.l.], 2012. p. 695–696.
- GUIZZARDI, G. et al. Ufo: Unified foundational ontology. **Applied ontology**, IOS Press, v. 17, n. 1, p. 167–210, 2022.
- HELLER, C. H. Near-term applications of artificial intelligence. **Naval War College Review**, JSTOR, v. 72, n. 4, p. 73–100, 2019.
- HINTON, G. et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. **IEEE Signal processing magazine**, IEEE, v. 29, n. 6, p. 82–97, 2012.
- JAMES, G. et al. **An introduction to statistical learning**. [S.l.]: Springer, 2013. v. 112.
- JAMES, G. et al. Tree-based methods. In: \_\_\_\_\_. **An Introduction to Statistical Learning: with Applications in Python**. Cham: Springer International Publishing, 2023. p. 331–366. ISBN 978-3-031-38747-0. Disponível em: [https://doi.org/10.1007/978-3-031-38747-0\\_8](https://doi.org/10.1007/978-3-031-38747-0_8).
- JECH, T. **Set theory: The third millennium edition, revised and expanded**. [S.l.]: Springer, 2003.

JIANG, J. J.; CONRATH, D. W. Semantic similarity based on corpus statistics and lexical taxonomy. **arXiv preprint cmp-lg/9709008**, 1997.

KARS, S.; OĞUZTÜZÜN, H. An ontology for a naval wargame conceptual model. In: SPRINGER. **Metadata and Semantic Research: 5th International Conference, MTSR 2011, Izmir, Turkey, October 12-14, 2011. Proceedings 5**. [S.l.], 2011. p. 1–11.

KENT, W. **Data and Reality: A Timeless Perspective on Perceiving and Managing Information in Our Imprecise World**. 3<sup>a</sup>. ed. 966 Woodmere Drive Westfield, NJ 07090 U.S.A.: Technics Publications, LLC, 2012. 143 p. ISBN 978-1-9355042-1-4.

KONDRATENKO, Y. et al. Towards implementing the strategy of artificial intelligence development: Ukraine peculiarities. In: **Proceedings of the Information Control Systems & Technologies (ICST 2023)**. [S.l.: s.n.], 2023.

KRESS, M. et al. Operational logistics. **The Art and Science of Sustaining Military Operations**, Springer, 2002.

KULMANOV, M. et al. Semantic similarity and machine learning with ontologies. **Briefings in Bioinformatics**, v. 22, n. 4, p. bbaa199, 10 2020. ISSN 1477-4054. Disponível em: <https://doi.org/10.1093/bib/bbaa199>.

LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, Ieee, v. 86, n. 11, p. 2278–2324, 1998.

LIN, D. et al. An information-theoretic definition of similarity. In: **Icml**. [S.l.: s.n.], 1998. v. 98, n. 1998, p. 296–304.

LUKYANENKO, R. et al. Using conceptual modeling to support machine learning. In: SPRINGER. **Information Systems Engineering in Responsible Information Systems: CAiSE Forum 2019, Rome, Italy, June 3–7, 2019, Proceedings 31**. [S.l.], 2019. p. 170–181.

LUKYANENKO, R.; PARSONS, J.; STOREY, V. C. Modeling matters: Can conceptual modeling support machine learning? **AI SIGSAND**, p. 1–12, 2018.

MAASS, W.; STOREY, V. C. Pairing conceptual modeling with machine learning. **Data & Knowledge Engineering**, Elsevier, v. 134, p. 101909, 2021.

MADDALENA, L.; BAIÃO, F. An application of the disease ontology (do) for clustering covid-19 hospitalizations in rio de janeiro. In: **15th Seminar on Ontology Research in Brazil, ONTOBRAS 2022 and 6th Doctoral and Masters Consortium on Ontologies, WTDO 2022**. [S.l.: s.n.], 2022. p. 9–12.

MASOLO, C. et al. The wonderweb library of foundational ontologies. **WonderWeb deliverable D**, v. 17, p. 2002, 2002.

MIENYE, I. D.; SUN, Y. A survey of ensemble learning: Concepts, algorithms, applications, and prospects. **Ieee Access**, IEEE, v. 10, p. 99129–99149, 2022.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.

- MORETTIN, P. A.; SINGER, J. d. M. **Estatística e ciência de dados**. 1<sup>a</sup>. ed. Rio de Janeiro, RJ: LTC, 2023. 454 p. ISBN 9788521638162.
- MYLOPOULOS, J. Conceptual modelling and telos. **Conceptual modelling, databases, and CASE: An integrated view of information system development**, John Wiley & Sons, p. 49–68, 1992.
- NAVEED, H. et al. A comprehensive overview of large language models. **arXiv preprint arXiv:2307.06435**, 2023.
- OURIQUES, L. M. C. et al. Siplom 3: A terceira geração do sistema de comando e controle da defesa. In: **Simpósio de Aplicações Operacionais em Áreas de Defesa 2012 (SIGE2012)**. [s.n.], 2012. Disponível em: [https://www.sige.ita.br/edicoes-antiores/2012/st/P\\_15.pdf](https://www.sige.ita.br/edicoes-antiores/2012/st/P_15.pdf).
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. **the Journal of machine Learning research**, JMLR. org, v. 12, p. 2825–2830, 2011.
- PEROZZI, B.; AL-RFOU, R.; SKIENA, S. Deepwalk: Online learning of social representations. In: **Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining**. [S.l.: s.n.], 2014. p. 701–710.
- PESQUITA, C. et al. Metrics for go based protein semantic similarity: a systematic evaluation. In: SPRINGER. **BMC bioinformatics**. [S.l.], 2008. v. 9, p. 1–16.
- PESQUITA, C. et al. Semantic similarity in biomedical ontologies. **PLoS computational biology**, Public Library of Science San Francisco, USA, v. 5, n. 7, p. e1000443, 2009.
- PHILIPPI, S. T. et al. Pirâmide alimentar adaptada: guia para escolha dos alimentos. **Revista de nutrição**, SciELO Brasil, v. 12, p. 65–80, 1999.
- PINHEIRO, C.; PATETTA, M. **Introduction to Statistical and Machine Learning Methods for Data Science**. [S.l.]: SAS Institute, 2021.
- POHLERT, T. **The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR)**. [S.l.], 2014. R package. Disponível em: <https://CRAN.R-project.org/package=PMCMR>.
- POPPER, K. **The logic of scientific discovery**. 2nd. ed. London: Routledge, 2002. 544 p. ISBN 9780203994627.
- PROVOST, F.; FAWCETT, T. Data science and its relationship to big data and data-driven decision making. **Big data**, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 1, n. 1, p. 51–59, 2013.
- RESNIK, P. Using information content to evaluate semantic similarity in a taxonomy. **arXiv preprint cmp-lg/9511007**, 1995.
- SILVA, E. M. d.; MUTZ, F. W.; RUY, F. B. Uso de ontologias no suporte a aplicação de machine learning: um caso no domínio de evasão escolar. In: **15th Seminar on Ontology Research in Brazil, ONTOBRAS 2022 and 6th Doctoral and Masters Consortium on Ontologies, WTDO 2022**. [S.l.]: Ontobras, 2022. p. 32–45.

- SIMON, S. J. The art of military logistics. **Communications of the ACM**, ACM New York, NY, USA, v. 44, n. 6, p. 62–66, 2001.
- SOTHILINGAM, R.; YU, E.; SENDEROVICH, A. Towards higher maturity for machine learning: A conceptual modelling approach. **The iJournal: Student Journal of the University of Toronto's Faculty of Information**, v. 5, n. 1, p. 80–97, 2019.
- SPEAR, A. D.; CEUSTERS, W.; SMITH, B. Functions in basic formal ontology. **Applied Ontology**, IOS Press, v. 11, n. 2, p. 103–128, 2016.
- VAPNIK, V. N. Statistical learning theory. Wiley, New York, 1998.
- VERDONCK, M. et al. Comparing traditional conceptual modeling with ontology-driven conceptual modeling: An empirical study. **Information Systems**, v. 81, p. 92–103, 2019. ISSN 0306-4379. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0306437918303727>.
- WIRTH, R.; HIPPEL, J. Crisp-dm: Towards a standard process model for data mining. In: MANCHESTER. **Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining**. [S.l.], 2000. v. 1, p. 29–39.
- WOLPERT, D. H. Stacked generalization. **Neural Networks**, v. 5, n. 2, p. 241–259, 1992. ISSN 0893-6080. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0893608005800231>.
- WU, W.; KUNZ, N.; BRANCO, P. Imbalancedlearningregression-a python package to tackle the imbalanced regression problem. In: SPRINGER. **Joint European Conference on Machine Learning and Knowledge Discovery in Databases**. [S.l.], 2022. p. 645–648.
- YIN, R. K. **Case study research: Design and methods**. [S.l.]: sage, 2009. v. 5.