



UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Tecnologia

João Emmanuel D' Alkmin Neves

**Mineração de Dados Aplicada a Simulação de Cenários
Complexos em Sistemas Multiagentes**

Limeira
2024

João Emmanuel D' Alkmin Neves

**Mineração de Dados Aplicada a Simulação de Cenários
Complexos em Sistemas Multiagentes**

Tese apresentada à Faculdade de Tecnologia da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Doutor em Tecnologia, na área de Sistemas de Informação e Comunicação.

Orientador: Prof. Dr. Paulo Sergio Martins Pedro (*in memoriam*)

Coorientadora: Profa. Dra. Marli de Freitas Gomes Hernandez

Este trabalho corresponde à versão final da Tese defendida por João Emmanuel D' Alkmin Neves e orientado pelo Prof. Dr. Paulo Sergio Martins Pedro (*in memoriam*) e coorientado pela Profa. Dra. Marli de Freitas Gomes Hernandez

Limeira
2024

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Tecnologia
Felipe de Souza Bueno - CRB 8/8577

N414m Neves, João Emmanuel D'Alkmin, 1977-
Mineração de dados aplicada a simulação de cenários complexos em sistemas multiagentes / João Emmanuel D'Alkmin Neves. – Limeira, SP : [s.n.], 2024.

Orientador: Paulo Sérgio Martins Pedro.
Coorientador: Marli de Freitas Gomes Hernandez.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Tecnologia.

1. Mineração de dados (Computação). 2. Agentes inteligentes (Software). 3. Simulação (Computadores). 4. Otimização. 5. Eficiência energética. I. Martins Pedro, Paulo Sérgio, 1967-2024. II. Hernandez, Marli de Freitas Gomes, 1959-. III. Universidade Estadual de Campinas. Faculdade de Tecnologia. IV. Título.

Informações Complementares

Título em outro idioma: Data mining applied to simulation of complex scenarios in multiagent systems

Palavras-chave em inglês:

Data mining

Intelligent agents (Computer software)

Computer simulation

Optimization

Energy efficiency

Área de concentração: Sistemas de Informação e Comunicação

Titulação: Doutor em Tecnologia

Banca examinadora:

Edson Luiz Ursini

Ieda Geriberto Hidalgo

Luiz Camolesi Júnior

Maria Cristina Aranda

Luiz Ariovaldo Fabri Junior

Data de defesa: 05-04-2024

Programa de Pós-Graduação: Tecnologia

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0002-9472-9753>

- Currículo Lattes do autor: <http://lattes.cnpq.br/4815333898210629>

FOLHA DE APROVAÇÃO

Abaixo se apresentam os membros da comissão julgadora da sessão pública de defesa de tese para o Título de Doutor em Tecnologia na área de concentração Sistemas de Informação e Comunicação, a que se submeteu o aluno João Emmanuel D' Alkmin Neves, em 05 de abril de 2024 na Faculdade de Tecnologia – FT/UNICAMP, em Limeira/SP.

Prof. Dr. Edson Luiz Ursini

Presidente da Comissão Julgadora

Profa. Dra. Ieda Geriberto Hidalgo

Universidade Estadual de Campinas - UNICAMP

Prof. Dr. Luiz Camolesi Júnior

Universidade Estadual de Campinas - UNICAMP

Profa. Dra. Maria Cristina Aranda

Faculdade de Tecnologia - Ministro Ralph Biasi - FATEC AMERICANA

Prof. Dr. Luiz Ariovaldo Fabri Junior

Pesquisador Independente - Prefeitura de Araras/SP

Ata da defesa, assinada pelos membros da Comissão Examinadora, encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação da Faculdade de Tecnologia.

DEDICATÓRIA

Dedico este trabalho, minha essência e minha vida, às mulheres que transcendem a espetacularidade:

À minha mãe, Cladis Ester (*in memoriam*), cuja presença agora repousa nas lembranças e na eternidade, minha fonte de inspiração eterna, a que sempre confiou no meu potencial, é a razão do que sou, de quem me tornei.

À minha esposa, Anamelia, minha luz que ilumina o caminho, constantemente me impulsionando, motivando a explorar o infinito juntos.

À minha irmã Daniela, minha fortaleza nas batalhas grandiosas da vida, que seja nas tempestades ou nos dias de bonança, ela permanece ao meu lado.

Mulheres espetaculares, cada uma em sua essência única, vocês são o alicerce de meu ser. É uma bênção que eu tenha a honra de compartilhar minha jornada com vocês. Meu amor por vocês transborda, uma sinfonia que ressoa no âmago da minha alma, eternamente dedicada a vocês.

AGRADECIMENTOS

Primeiramente, expresso minha gratidão a Deus, guia constante em todos os momentos, fonte de força em horas difíceis, Aquele que me ergue quando fraquejo.

À minha mãe, Cladis Ester (*in memoriam*), por seu amor, carinho, compreensão e luz. Abnegada, tantas vezes sacrificou seus próprios sonhos em prol dos meus. A você, minha mãe, o meu mais profundo agradecimento e eterno amor.

À minha esposa, Anamelia, por permanecer sempre ao meu lado, oferecendo sua força, companheirismo, respeito e incentivo. Seu afeto, tão essencial em minha vida, é motivo de gratidão por me fazer sentir amado em todos os momentos, independentemente do tamanho das adversidades.

Aos meus orientadores, Professor Dr. Paulo Sergio Martins Pedro (*in memoriam*) e Professora Dra. Marli de Freitas Gomes Hernandez, agradeço a oportunidade, orientação, atenção e apoio em todos os momentos. Suas experiências e sugestões moldaram meus conceitos, conduzindo-me à conclusão desta tese.

Aos ilustres membros da banca examinadora, Professor Dr. Edson Luiz Ursini, Professora Dra. Ieda Geriberto Hidalgo, Professor Dr. Luiz Camolesi Júnior, Professora Dra. Maria Cristina Aranda e Professor Dr. Luiz Ariovaldo Fabri Junior, manifesto minha sincera gratidão por dedicarem seu precioso tempo e expertise à avaliação desta tese.

Ao Professor Dr. Luiz Camolesi Júnior, pela oportunidade ao permitir a realização do Programa de Estágio Docente sob sua orientação.

Ao Professor Dr. Edson Luiz Ursini, pelo apoio durante a fase final do meu doutorado e pelos valiosos apontamentos que contribuíram para minha tese.

Ao Professor Me. Clerivaldo José Roccia, por sua amizade e conselhos.

À família e aos amigos, manifesto meu apreço pelos conselhos e confiança.

Aos professores e funcionários da Faculdade de Tecnologia - Unicamp, expresso minha gratidão pela dedicação, competência e ensinamentos.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Expresso também minha gratidão a todas as pessoas que, de diversas maneiras, colaboraram para a conclusão desta jornada chamada Doutorado.

*quebrou
grilhões
do
medo
e
voou*

(poesia Aldravia de Anamelia Freire)

RESUMO

A presente tese explorou a complexidade da tomada de decisões em ambientes desafiadores, onde a obtenção de dados é dificultada pela demora, altos custos e uso intensivo de recursos computacionais. Em um contexto de explosão na produção de dados durante a Era da Informação, o objetivo principal propôs uma abordagem inovadora integrando mineração de dados em simulações baseadas em agentes. Para alcançar esse propósito, a pesquisa desenvolveu uma técnica baseada na raspagem de dados não estruturados e atuais, utilizando a linguagem de programação Python, visando a obtenção eficiente de dados para simulações em sistemas multiagentes em cenários complexos usando o programa Netlogo. A metodologia empregada envolveu pesquisa aplicada, quali-quantitativa, exploratória e bibliográfica. Por meio de dois estudos de caso focados no contexto brasileiro, onde o primeiro investigou as razões fundamentais para a adoção de sistemas solares e o segundo abordou as principais causas da redução do consumo de energia elétrica, os resultados corroboraram a hipótese inicial, evidenciando que a técnica desenvolvida e integrada aos sistemas multiagentes contribuiu para aprimorar o processo de simulação. Em resumo, a tese propôs uma abordagem pioneira para a coleta, validação, análise e integração de dados em simulações de cenários complexos em sistemas multiagentes, com o objetivo de otimizar os processos e reduzir custos, tempo, recursos computacionais e erros.

Palavras-Chave: Mineração de dados, Agentes inteligentes, Simulação, Otimização e Eficiência energética.

ABSTRACT

The present thesis delved into the complexity of decision-making in challenging environments, where data acquisition is hindered by delays, high costs, and intensive computational resources. In a context of data explosion during the Information Age, the primary objective proposed an innovative approach integrating Data Mining into agent-based simulations. To achieve this aim, the research developed a technique based on scraping current and unstructured data, utilizing the Python programming language, aiming for efficient data acquisition for simulations in multi-agent systems in complex scenarios using the Netlogo program. The methodology employed encompassed applied research, qualitative and quantitative, exploratory, and bibliographic approaches. Through two case studies focused on the Brazilian context, where the first investigated the fundamental reasons for the adoption of solar systems and the second addressed the main causes of reduced electricity consumption, the results corroborated the initial hypothesis, showing that the technique developed and integrated into multi-agent systems contributed to improving the simulation process. In summary, the thesis proposed a pioneering approach for collecting, validating, analyzing, and integrating data in simulations of complex scenarios in multi-agent systems, with the aim of optimizing processes and reducing costs, time, computational resources, and errors.

Keywords: Data mining, Intelligent agents, Computer simulation, Optimization, and Energy efficiency.

LISTA DE ILUSTRAÇÕES

Figura 1 - Estrutura da Tese.....	42
Figura 2 - Áreas que originaram os agentes e Sistemas Multiagentes	45
Figura 3 - Esquema típico de um agente	46
Figura 4 - Categorias de agentes.....	54
Figura 5 - Esquema genérico de uma arquitetura deliberativa	62
Figura 6 - Esquema genérico de uma arquitetura reativa	63
Figura 7 - Esquema genérico de uma arquitetura híbrida	65
Figura 8 - Representação da arquitetura em camadas horizontais	66
Figura 9 - Representação da arquitetura em camadas verticais	67
Figura 10 - Esquema genérico de uma arquitetura BDI	69
Figura 11 - Esquema genérico de uma arquitetura de um agente social	71
Figura 12 - Estrutura de um Sistema Multiagente	75
Figura 13 - Divisão da Inteligência Artificial Distribuída.....	77
Figura 14 - Agente com capacidade de comunicação.....	80
Figura 15 - Arquiteturas de comunicação de agentes. (a) Direta. (b) Assistida	81
Figura 16 - Características da aprendizagem em SMA.....	89
Figura 17 - Quociente de inteligência dos dados	101
Figura 18 - Dimensão do banco de dados e capacidade analítica	102
Figura 19 - Tipos de capacidade analítica.....	103
Figura 20 - Estágios da Analítica.....	104
Figura 21 - Principais fases do <i>Big Data Analytics</i>	105
Figura 22 - Ciclo de descoberta de conhecimento em bases de dados	109
Figura 23 - Exemplo de uma base de dados de retenção de clientes.....	112
Figura 24 - Modelo de redução de dados.....	117
Figura 25 - Infraestrutura genérica de um armazém de dados.....	120
Figura 26 - Visão geral da raspagem de dados.....	121
Figura 27 - Modelo de um sistema de rastreamento da <i>web</i>	125
Figura 28 - Organograma de metodologia de pesquisa	127
Figura 29 - Interface do NetLogo.....	131
Figura 30 - Agente e ambiente	132
Figura 31 - Processo da coleta de dados.....	134
Figura 32 - Representação gráfica do método de LDA	138
Figura 33 - Representação simplificada do LDA	139
Figura 34 - Geração de energia solar fotovoltaica no Brasil.....	150
Figura 35 - Nuvem de palavras coletadas.....	152
Figura 36 - Número de tópicos. (a) Métrica: Cv. (b) Métrica: CUMass	153
Figura 37 - Layout pyLDavis (a) Distância entre tópicos (b) Palavras por tópico....	154
Figura 38 - Visão durante as simulações	160
Figura 39 - Instalações sistemas solares domésticos em 25 anos.....	164
Figura 40 - Impacto do raio de influência na instalação	165
Figura 41 - Comparação de instalações segundo o raio de influência	166

Figura 42 - Impacto do limiar de vizinhança	167
Figura 43 - Comparação de instalações segundo o limiar de vizinhança.....	168
Figura 44 - Nuvem de palavras coletadas	174
Figura 45 - Número de tópicos (a) Métrica: Cv (b) Métrica: CUMass	175
Figura 46 - Layout pyLDavis (a) Distância entre tópicos. (b) Palavras por tópico...	177
Figura 47 - Visão do ambiente de simulação	186
Figura 48 - Consumo médio de energia por pessoa anualmente.....	187
Figura 49 - Fatores de influência do consumo relacionados a habitação.....	189
Figura 50 - Custos da implementação em relação ao valor da habitação	190
Figura 51 - Fatores de influência do consumo relacionados aos habitantes	191
Figura 52 - Custos da aprendizagem em relação ao valor da habitação	192
Figura 53 - Fatores de influência do consumo e custos relacionados ao clima.....	194
Figura 54 - Custos dos fatores climáticos em relação ao valor da habitação	196
Figura 55 - Resultados considerando prazo x benefício	198
Figura 56 - Resultados considerando custo x benefício	200
Figura 57 - Resultados considerando redução x benefício	202
Figura 58 - Índice de Viabilidade Ajustado	205

LISTA DE TABELAS

Tabela 1 - Principais formas de obter dados de entrada em artigos de 2018	26
Tabela 2 - Principais formas de obter dados de entrada em artigos de 2019	28
Tabela 3 - Principais formas de obter dados de entrada em artigos de 2020	31
Tabela 4 - Principais formas de obter dados de entrada em artigos de 2021	33
Tabela 5 - Principais formas de obter dados de entrada em artigos de 2022	36
Tabela 6 - Principais formas de obter dados de entrada em artigos de 2023	38
Tabela 7 - Quantidade de cada tipo de obtenção de dados de entrada	39
Tabela 8 - Evolução dos principais paradigmas de programação	55
Tabela 9 - Diferenças entre as implementações da comunicação entre agentes	82
Tabela 10 - Capacidades de comunicação de diferentes agentes	86
Tabela 11 - O dado, a informação e o conhecimento	97
Tabela 12 - Os 5 Vs do Big Data	99
Tabela 13 - Síntese dos Estudos de Caso	145
Tabela 14 - Dados regionais do Brasil	149
Tabela 15 - Descrição dos dados de texto coletados	151
Tabela 16 - Decomposição dos dados de texto coletados	152
Tabela 17 - Ajuste e melhoria das medidas de coerência	154
Tabela 18 - Modelagem de tópicos	156
Tabela 19 - Parâmetros usados no modelo NetLogo	158
Tabela 20 - Principais razões para instalação	161
Tabela 21 - Maiores influências para instalação	162
Tabela 22 - Comparação entre maiores influências para instalação	163
Tabela 23 - Comparação dos meios de pagamento	163
Tabela 24 - Descrição dos dados de texto coletados	173
Tabela 25 - Decomposição dos dados de texto coletados	174
Tabela 26 - Ajuste e melhoria das medidas de coerência	176
Tabela 27 - Modelagem de tópicos	178
Tabela 28 - Tipos de Habitação	179
Tabela 29 - Informações sobre Iluminação e Aparelhos Elétricos	179
Tabela 30 - Informações sobre os Habitantes	179
Tabela 31 - Informações sobre o Clima	180
Tabela 32 - Indicadores do IVA	203

LISTA DE ABREVIATURAS E SIGLAS

API	Interface de Programação de Aplicativos.
BI	Inteligência de Negócios.
DOM	Modelo de Objeto de Documento.
LDA	Alocação Latente de Dirichlet.
IVA	Índice de Viabilidade Ajustado.
IA	Inteligência Artificial.
IAD	Inteligência Artificial Distribuída.
SMA	Sistema Multiagente.

SUMÁRIO

LISTA DE ILUSTRAÇÕES	10
LISTA DE TABELAS	12
LISTA DE ABREVIATURAS E SIGLAS	13
1. INTRODUÇÃO	16
1.1. MOTIVAÇÃO	18
1.2. OBJETIVOS.....	20
1.3. JUSTIFICATIVA.....	21
1.4. REVISÃO DA LITERATURA.....	23
1.5. ESTRUTURA DA TESE.....	41
2. AGENTES AUTÔNOMOS.....	43
2.1. CLASSES DOS AGENTES	47
2.2. AGENTES VERSUS OBJETOS	54
2.3. AMBIENTES	56
2.4. ARQUITETURA DOS AGENTES	60
2.5. APLICAÇÕES DOS AGENTES	72
3. SISTEMAS MULTIAGENTES	74
3.1. INTELIGÊNCIA ARTIFICIAL DISTRIBUÍDA VERSUS SMA.....	75
3.2. MOTIVAÇÃO DOS SMA	78
3.3. COMUNICAÇÃO EM SMA.....	79
3.3.1. ARQUITETURAS DO SISTEMA DE COMUNICAÇÃO	80
3.3.2. CARACTERÍSTICAS DA COMUNICAÇÃO	83
3.3.3. CARACTERIZAÇÃO DAS MENSAGENS	84
3.3.4. TIPO DE MENSAGENS	85
3.3.5. PROTOCOLOS E NÍVEIS DE COMUNICAÇÃO	87
3.3.6. ONTOLOGIAS	87
3.4. APRENDIZAGEM EM SMA.....	88
3.5. COORDENAÇÃO EM SMA.....	92
4. CIÊNCIA DE DADOS.....	96
4.1. BIG DATA	97
4.2. ANALÍTICA	100
4.3. MINERAÇÃO DE DADOS	106

4.4. RASPAGEM DE DADOS.....	121
5. METODOLOGIA.....	127
5.1. MATERIAIS.....	129
5.2. MÉTODOS.....	133
6. ESTUDOS DE CASO E DISCUSSÕES.....	145
6.1. ESTUDO DE CASO 1.....	146
6.1.1. COLETA DE DADOS.....	148
6.1.2. MODELAGEM DO CENÁRIO UTILIZANDO SMA.....	156
6.1.3. RESULTADOS E DISCUSSÕES.....	161
6.2. ESTUDO DE CASO 2.....	170
6.2.1. COLETA DE DADOS.....	171
6.2.2. MODELAGEM DO CENÁRIO UTILIZANDO SMA.....	180
6.2.3. RESULTADOS E DISCUSSÕES.....	197
7. CONSIDERAÇÕES FINAIS.....	209
7.1. RECOMENDAÇÕES PARA PESQUISAS FUTURAS.....	211
REFERÊNCIAS BIBLIOGRÁFICAS.....	214
APÊNDICES.....	225
APÊNDICE A - CÓDIGO BASE DA TÉCNICA DE RASPAGEM DE DADOS DESENVOLVIDO NA LINGUAGEM DE PROGRAMAÇÃO PYTHON.....	226

1. INTRODUÇÃO

Vivencia-se na contemporaneidade a chamada Era da Informação, período este, no qual, há uma crescente produção de informações. Estima-se que 92% dos dados disponíveis hoje foram criados nos últimos cinco anos (Foster *et al.*, 2020). Nesta era, tanto as pessoas quanto as instituições estão altamente conectadas e deixam rastros digitais, seja diretamente na Internet ou nos inúmeros dispositivos a ela conectados.

Esse enorme volume de dados, que abriga valiosas informações para as mais diversas instituições, dificilmente será bem aproveitado se não houver meios de utilizá-los, uma vez que em sua grande maioria esses dados são distribuídos, heterogêneos e não-estruturados. Por isso, atualmente, o interesse por criar formas para o uso de simulação utilizando os dados na área de tecnologia e informação tem aumentado e possibilitado o avanço de estudos e pesquisas abrangendo essa temática.

A simulação é uma poderosa ferramenta para análise dos processos produtivos, que possibilita o desenvolvimento e aperfeiçoamento de estudos em um ambiente controlado, considerando a sua aplicabilidade em sistemas dinâmicos, complexos e/ou que apresentam dificuldades para estudos analíticos. O valor não está nos dados em si, mas na atualidade e qualidade de indicadores para alimentar a simulação de cenários complexos e, assim, desenvolver simulações mais realistas e com resultados mais assertivos.

No contexto desta pesquisa, a simulação é uma ferramenta que permite a geração de cenários complexos, a partir dos quais pode-se orientar o processo de tomada de decisão, proceder análises e avaliações de sistemas e propor soluções para a melhoria de performance (Law; Kelton, 2000). Em outras palavras, a simulação pode ser entendida como um processo de experimentação com um modelo detalhado de um sistema real na qual adota-se o processo e projeção para a condução de experimentos para o entendimento e avaliação do comportamento (Chwif, 2014).

As principais vantagens da simulação incluem sua não interferência no funcionamento do sistema real; a avaliação e interpretação dos dados ocorre de forma lógica, sistemática e consistente; e, o modelo pode ser alterado e/ou aperfeiçoado facilmente, proporcionando uma aprendizagem adaptativa (Law; Kelton, 2000).

Com esse propósito, os programas serão considerados como “ferramentas inteligentes” e, dentro do universo das simulações serão utilizados os sistemas multiagentes. Os agentes autônomos se apresentam como um paradigma de computação apropriado para o desenvolvimento de aplicações para os ambientes de estudo neste trabalho. A utilização de múltiplos agentes tem se mostrado eficaz para simular e construir inúmeros cenários de grande complexidade.

De acordo com Garcia, Giret e Botti (2015), os sistemas multiagentes simulam o comportamento de um conjunto de agentes autônomos em um mesmo ambiente e esses sistemas são constituídos por múltiplos agentes que apresentam comportamentos inteligentes e interações com outros agentes do sistema. Esses comportamentos e interações apresentam ações em conjunto e/ou caráter competitivo, dependendo dos objetivos de cada um desses agentes.

A simulação de cenários complexos utilizando sistemas multiagentes é uma possibilidade muito interessante, uma vez que estes possibilitam a modelagem desses cenários para testar hipóteses sobre sistemas diversos; tais como simulações em ciências naturais e/ou sociais, que podem ser utilizadas e modificadas.

Os agentes dos sistemas multiagentes podem ser habilitados para atuar autonomamente visando atingir seus objetivos e, também, com capacidade de interagir com outros agentes a partir de protocolos de interação social. A Inteligência Artificial transforma os agentes em programas independentes e se apresenta como aliada, devida às inúmeras possibilidades de resolução envolvendo problemas complexos. Os agentes aceitam novas circunstâncias e apresentam a capacidade de aprenderem e adaptarem o seu comportamento baseado na experiência. Além disso, os agentes podem ainda, ter regras para mudar as suas próprias regras de comportamento (Musaeus; Musaeus, 2020).

Considerando o enorme volume de dados existentes, há uma crescente necessidade de extrair informações relevantes e identificar possibilidades novas de conhecimentos a partir destes dados. Para alcançar tais objetivos, apoiado nas pesquisas bibliográficas, a mineração de dados é utilizada neste estudo baseada na definição de Fayyad *et al.* (1996), como o processo não-trivial de identificar, em dados, padrões válidos, novos, potencialmente úteis e ultimamente compreensíveis.

Frente aos desafios intrínsecos à modelagem de cenários complexos em simulações multiagentes, esta tese visa explorar a seguinte problemática: como aprimorar o processo de simulação em contextos intrincados, caracterizados pela

escassez ou por obstáculos na obtenção de dados de entrada em sistemas multiagentes por meio de estratégias de mineração de dados? Essa indagação surge em resposta à crescente necessidade de inovações no campo das simulações multiagentes, destacando-se na modelagem de cenários de alta complexidade.

Considerando os desafios intrínsecos à modelagem de cenários complexos nas simulações multiagentes, esta tese formula a seguinte hipótese para orientar a pesquisa: O desenvolvimento de uma abordagem ímpar e pioneira de raspagem de dados, campo pertencente à mineração de dados, pode conduzir a uma técnica eficiente na simulação de ambientes de alta complexidade em sistemas multiagentes.

Essa hipótese surge como resposta à demanda crescente por soluções que transcendam os desafios inerentes à coleta e utilização de dados em cenários complexos. A integração da mineração de dados, especificamente a raspagem de dados, em sistemas multiagentes é vista como uma abordagem pioneira para superar as limitações associadas à coleta e refinamento de dados em ambientes dinâmicos e heterogêneos.

A premissa subjacente à hipótese é que o desenvolvimento de uma técnica baseada na raspagem de dados para a obtenção de dados de entrada para ser utilizada em simulação multiagente pode oferecer uma vantagem significativa na aquisição de informações relevantes, mesmo em contextos nos quais as dificuldades de adquirir de dados tradicionais se apresentam como obstáculos. A agilidade, adaptabilidade e capacidade de extrair dados de fontes não estruturadas da técnica de raspagem de dados são consideradas elementos-chave para enfrentar os desafios específicos da modelagem desses cenários.

1.1. MOTIVAÇÃO

A raspagem de dados, técnica comumente utilizada para extrair informações de fontes *online* e a simulação baseada em agentes, empregada para analisar fenômenos complexos, são poderosas em suas respectivas áreas. No entanto, a combinação dessas técnicas apresenta um potencial inovador e inédito, revelando-se como o cerne desta tese de doutorado.

No contexto das simulações que se valem de sistemas multiagentes para modelagem de cenários complexos, esta pesquisa estabelece uma abordagem única

de coleta de dados por meio do desenvolvimento de uma técnica baseada na raspagem de dados.

O ineditismo desta abordagem reside no desenvolvimento de uma técnica única na esfera da raspagem de dados destinada a adquirir informações de entrada para simulações de cenários complexos em sistemas multiagentes. Esse enfoque amplia consideravelmente o alcance de aplicação, convertendo-o em uma ferramenta distintiva. Essa transformação é alcançada mediante a adoção de diversas premissas e a integração de inovações técnicas no processo de coleta da raspagem de dados, visando a obtenção de informações mais precisas e aprimoradas.

A motivação da pesquisa surge da necessidade de interpretar e direcionar o vasto mar de informações característico da Era da Informação. Segundo Amaral (2018), é imperativo compreender e gerenciar a vasta quantidade de dados disponíveis, uma vez que não é suficiente apenas a interpretação das informações, mas também a habilidade de as direcionar de maneira eficaz e estratégica para atender aos objetivos específicos de análise e tomada de decisão.

Essa pesquisa pretende contribuir para um mundo mais consciente, utilizando a informação como ferramenta-chave para simulações de cenários complexos visando a melhoria da saúde do planeta e assim, conseqüente, o bem-estar humano. Segundo Castro e Ferrari (2016), ao modelar e analisar os efeitos de diferentes estratégias e políticas em contextos complexos, as simulações baseadas em agentes oferecem uma perspectiva holística sobre questões globais urgentes, desde a gestão de recursos naturais até a adaptação às mudanças climáticas.

Ao conceber uma abordagem pioneira, que faz uso da raspagem de dados para gerar os dados de entrada em simulações baseadas em agentes voltadas para cenários complexos, esta tese propõe uma abordagem que detém o potencial de aprimorar a eficiência, possibilitando melhorias nos desfechos das simulações em contextos intrincados.

A aplicação desta técnica visa ampliar a compreensão de fenômenos complexos em uma variedade de domínios, caracterizando-se por custos reduzidos e uma coleta de dados rápida, com informações atuais, precisas e confiáveis. De acordo com Bahga e Madiseti (2016), a capacidade de interpretar e direcionar as informações de forma eficiente tornou-se uma competência essencial para enfrentar os desafios contemporâneos da análise de dados e da tomada de decisões.

A contribuição deste estudo incorpora dados dinâmicos e diversificados para

as informações de entrada; a raspagem de dados enriquece a simulação com uma camada de realismo e atualidade. Essa abordagem inovadora promete avanços significativos na coleta, processamento e integração de dados em simulações multiagentes, abrindo novas perspectivas para pesquisas e aplicações.

A relevância do emprego de conjuntos de dados contemporâneos reside na capacidade de enriquecer a representação de sistemas simulados em ambientes multiagentes, proporcionando uma contextualização mais precisa e alinhada com a realidade em constante evolução. A utilização de dados atualizados possibilita a incorporação de dinâmicas e tendências recentes, o que contribui para uma simulação mais fidedigna e robusta, ao refletir com maior precisão os padrões e comportamentos observados no mundo real (Coelho, 2018).

Além das vantagens multifacetadas resultantes da implementação da técnica de raspagem de dados na simulação de cenários complexos, como a coleta rápida e eficaz de dados atualizados e a habilidade de extrair informações de fontes não estruturadas, esta pesquisa destaca-se pelo uso de linguagens de programação avançadas e plataformas de simulação. Ao desenvolver uma abordagem no contexto da raspagem de dados para servir à simulação de cenários complexos, estabelece-se um novo padrão para a pesquisa e aplicação prática nesse campo.

1.2. OBJETIVOS

O objetivo geral desta tese é desenvolver e validar uma abordagem inovadora de coleta de dados para simulações que utilizam sistemas multiagentes na modelagem de cenários complexos, por meio da criação de uma técnica no contexto da raspagem de dados. Por meio disso, visa-se aprimorar o processo de simulação, proporcionando realismo e atualidade às informações coletadas, e, assim, contribuir significativamente para o avanço no campo das simulações multiagentes.

Os objetivos específicos incluem:

- Conduzir uma revisão abrangente da literatura sobre simulações baseadas em sistemas multiagentes, investigando características, aplicações, métodos de obtenção de dados e limitações presentes na literatura científica.
- Estabelecer uma base teórica sólida e robusta, integrando conhecimentos de

linguagens de programação, bibliotecas avançadas e plataformas de simulação. Essa fundamentação visa proporcionar uma melhor compreensão dos elementos essenciais para o desenvolvimento da abordagem proposta.

- Identificar e analisar criticamente as áreas de aplicação mais promissoras para a abordagem inovadora de coleta de dados proposta, considerando suas contribuições potenciais para a tomada de decisões em diferentes contextos.
- Realizar estudos de caso em diferentes cenários complexos para validar a eficácia e aplicabilidade da abordagem desenvolvida, evidenciando sua capacidade de melhorar processos, reduzir custos e aumentar a eficiência das simulações multiagentes.
- Analisar os resultados obtidos, destacando as contribuições específicas da abordagem criada para a melhoria das simulações multiagentes. Estabelecer padrões de excelência para pesquisas e aplicações práticas nesse campo, visando promover avanços significativos na coleta, processamento e integração de dados em simulações complexas.

Estes objetivos específicos convergem para o objetivo geral de proporcionar avanços substanciais na coleta de dados para simulações multiagentes, contribuindo para o aprimoramento do realismo dessas simulações em cenários complexos.

1.3. JUSTIFICATIVA

A justificativa para a realização desta tese encontra-se na crescente demanda por avanços significativos no domínio das simulações multiagentes, particularmente na modelagem de cenários complexos. A introdução de uma abordagem inovadora de coleta de dados propõe uma abordagem única e altamente relevante para superar os desafios intrínsecos a esse campo em constante evolução.

A necessidade premente de aprimorar a capacidade de tomada de decisões em ambientes complexos motiva a busca por estratégias inovadoras. Essa imperatividade advém da crescente complexidade dos problemas contemporâneos, que exigem soluções que transcendam os métodos convencionais e incorporem abordagens adaptativas e proativas (Coelho, 2018).

A proposta de incorporar a técnica de raspagem de dados nesse contexto específico representa um passo significativo em direção à consecução desse objetivo. A habilidade de extrair informações dinâmicas e diversificadas de uma ampla gama de fontes, incluindo portais governamentais, páginas de notícias, organizações não governamentais, redes sociais e plataformas de comércio eletrônico, é essencial para estabelecer uma base robusta para simulações multiagentes. Ao integrar dados não estruturados em modelos de simulação multiagentes, é possível capturar uma variedade de aspectos do comportamento humano, social e econômico, enriquecendo assim a capacidade de prever e analisar o comportamento de sistemas dinâmicos em diversos contextos (Dixon, 2019).

A relevância desta pesquisa transcende os benefícios para a tomada de decisões assertivas. A criação de uma técnica baseada na área de raspagem de dados para simulações multiagentes não apenas otimiza o tempo dedicado a processos decisórios, mas também contribui para a identificação eficaz de custos e esforços, redução de riscos e erros, aprimoramento do conhecimento sobre cenários e possibilidades, além de proporcionar um aumento significativo no controle das ações a serem determinadas. Isso significa que é possível controlar de forma mais eficiente as interações entre os agentes e as decisões que são tomadas ao longo do processo de simulação. Essa maior capacidade de manipulação e direcionamento das ações resulta em uma governança mais eficaz, garantindo uma condução mais precisa e otimizada das simulações multiagentes (Jezic; Chen-Burger; Kusek, 2020).

A busca por melhores soluções em cenários complexos requer uma abordagem que vá além das limitações para a coleta de dados. A incorporação da raspagem de dados oferece flexibilidade na coleta de informações, possibilitando o acesso e a síntese de dados que, de outra forma, seriam de difícil aquisição ou inacessíveis. Essa capacidade de transcender limitações e integrar uma variedade mais ampla de dados enriquece substancialmente as simulações, fornecendo uma base de dados robusta e alinhada com a complexidade do ambiente simulado (Bahga; Madiseti, 2016).

Ademais, a introdução desta abordagem única não apenas representa um avanço tecnológico e científico, mas também eleva o nível de realismo e atualidade nas simulações multiagentes. A capacidade de coletar dados rapidamente é inestimável para simulações que demandam um fluxo constante de dados contemporâneos para manter sua relevância e precisão.

A contribuição central desta pesquisa reside na melhoria substancial do

processo de simulação de cenários complexos por meio da inovadora técnica de raspagem de dados. Além disso, o estudo avança ao empregar linguagens de programação, bibliotecas avançadas e plataformas de simulação, transformando a raspagem de dados em um instrumento poderoso para aprimorar a qualidade e o realismo das simulações, ampliando seu escopo de aplicação e elevando-a de uma simples ferramenta de agregação de informações para um mecanismo refinado que melhora significativamente o fornecimento de dados aos modelos de simulação.

1.4. REVISÃO DA LITERATURA

Importantes estudos foram realizados sobre a simulação de cenários complexos, inclusive utilizando agentes inteligentes. Foram desenvolvidas pesquisas de simulação baseada em sistemas multiagentes em diversas áreas do conhecimento, como ciências sociais, biologia, engenharia, logística, jogos e entretenimento, entre outras. Na ciência social, por exemplo, ela tem sido amplamente utilizada para modelar fenômenos complexos, como comportamento de mercado, opinião pública e tomada de decisão em grupos. Na biologia, ela tem sido utilizada para modelar sistemas biológicos e ecológicos complexos, como a dinâmica populacional de espécies e interações entre espécies. Por sua vez, na engenharia e robótica é usada para projetar e testar robôs e sistemas de controle automatizados que precisam tomar decisões em tempo real em ambientes complexos e dinâmicos.

Este campo de estudo tem sido amplamente investigado, revelando uma rica gama de descobertas, porém, ainda existem lacunas significativas a serem preenchidas. Uma dessas lacunas reside na inovação dos métodos de aquisição de dados para alimentar simulações, uma vez que até o momento atual, as abordagens convencionais continuam a ser predominantes. Este aspecto será minuciosamente examinado e discutido nesta seção, onde serão apresentadas e analisadas as técnicas tradicionais utilizadas para a obtenção dos dados de entrada nas simulações.

Nesse contexto, o objetivo principal desta seção é realizar uma análise das pesquisas com simulações baseada em sistemas multiagentes, bem com a técnica para a obtenção dos dados de entrada. Posteriormente, essas entradas serão agrupadas, analisadas e contrastadas com a proposta desta pesquisa em utilizar técnicas de mineração de dados.

A obtenção de dados é uma etapa crucial em qualquer pesquisa científica, pois a qualidade e a confiabilidade dos resultados dependem diretamente da qualidade dos dados coletados. Existem diversos métodos para a obtenção de dados, cada um com suas características, sendo que os principais tipos de obtenção de dados utilizados atualmente são dados históricos de arquivos, simulação de cenários, pesquisas, entrevistas e sensores. Em consonância com Fontes e Souza (2021):

- Os dados históricos de arquivos são informações registradas em documentos, arquivos, cartórios, bibliotecas, entre outros, e que se referem a um período anterior ao momento atual. Essa abordagem é comumente utilizada em estudos que envolvem análise temporal, e podem ser de grande valia para análises de longo prazo.
- A projeção de cenários é uma técnica que envolve a elaboração de modelos matemáticos que possibilita a geração de dados para representar o comportamento de sistemas sob variadas condições. Essa abordagem é eficaz para explorar os desfechos potenciais de eventos específicos.
- As pesquisas são estudos que envolvem a coleta e análise de dados quantitativos ou qualitativos, e são realizadas com o objetivo de responder a uma pergunta de pesquisa ou hipótese. Essa abordagem pode envolver métodos, como questionários, observação, análise documental, entre outros.
- As entrevistas são uma técnica de coleta de dados qualitativos que envolve a realização de perguntas estruturadas ou não estruturadas a um indivíduo ou grupo de pessoas. O objetivo é obter informações sobre as percepções e experiências dos entrevistados em relação ao tema de pesquisa.
- Os sensores são dispositivos que permitem a medição de variáveis físicas, químicas ou biológicas. Essa abordagem é comumente utilizada em estudos de monitoramento ambiental, permitindo a coleta de dados em tempo real.

Coelho (2018) desempenhou um papel significativo na transição paradigmática da análise de sistemas sociais e dinâmicos, onde ele introduziu uma abordagem baseada na simulação de agentes, rompendo com os tradicionais métodos matemáticos. Essa pesquisa concentrou-se na representação de agentes individuais e inteligentes, modelando comportamentos semelhantes aos seres humanos simplificados. O pesquisador utilizou dados históricos de arquivos para simular

eventos do passado, como o destino do povo Anasazi no século XIV. O autor destacou a adaptabilidade desses modelos baseados em agentes, aplicáveis em uma ampla gama de disciplinas, desde a arqueologia até a indústria cinematográfica.

Badham *et al.* (2018) contribuíram para a compreensão da complexidade dos sistemas de saúde e do gerenciamento de doenças não transmissíveis. Para investigar o comportamento de sistemas complexos, empregaram a modelagem baseada em agentes. Os autores utilizaram dados históricos de arquivos, pesquisas e entrevistas para a aquisição de dados visando a calibração e validação do modelo. O estudo concluiu que a modelagem baseada em agentes possui potencial para melhorar a pesquisa em saúde pública, encorajando a sua adoção mais ampla.

Gómez-Marín, Arango-Serna e Serna-Urán (2018) exploraram a complexidade da distribuição urbana de mercadorias, uma área complexa devido ao crescimento populacional e ao aumento da demanda nas cidades. Para abordar esses desafios, os pesquisadores propuseram um modelo conceitual baseado em sistemas multiagentes. A obtenção de dados envolveu dados históricos de arquivos e pesquisas de campo. Os resultados destacaram como os agentes influenciam a redistribuição em cenários dinâmicos, considerando variações nas condições de tráfego.

Yazan *et al.* (2018) exploraram a produção de biogás a partir de esterco como uma solução energética e ambiental. Eles optaram por uma abordagem para a obtenção de dados que envolvia projeção de cenários e pesquisas para simular as interações entre fornecedores de esterco (agricultores) e produtores de biogás na Holanda utilizando sistemas multiagentes. Os resultados evidenciaram a variação significativa no preço do esterco e seu impacto na economia dos agricultores e produtores de biogás, concluindo que a cooperação entre as partes resultaria em benefícios econômicos mútuos, ilustrando o potencial da modelagem baseada em agentes para informar políticas energéticas e ambientais.

Keane, Egan e Hawk (2018) investigaram o impacto da distribuição de naloxona como medida para combater as overdoses de opioides nos EUA. Os autores obtiveram os dados para a modelagem baseada em agentes por meio de entrevistas e projeção de cenários para avaliar a eficácia da distribuição. O estudo destacou que a distribuição de naloxona poderia reduzir significativamente as mortes por overdose.

Manley e Cheng (2018) realizaram uma pesquisa sobre análise do papel da cognição espacial na previsão do fluxo de tráfego urbano. Os autores enfatizaram a complexidade do sistema urbano, influenciado pelas interações de múltiplos

indivíduos, negligenciada pelos modelos tradicionais de tráfego urbano. A abordagem para a obtenção de dados utilizou a projeção de cenários, dados históricos de arquivos e pesquisas para desenvolver um modelo baseado em agentes que considerava a cognição espacial na seleção de rotas e a diversidade no conhecimento espacial. O estudo concluiu que a cognição espacial desempenha um papel crucial na previsão do fluxo de tráfego urbano, destacando a importância de investigações nessa área.

Woltmann *et al.* (2018) examinaram o fornecimento de energia verde para empresas de manufatura, com ênfase nas energias renováveis. Os pesquisadores desenvolveram uma simulação baseada em agentes para explorar combinações de fontes de energia renovável, armazenamento de energia e consumidores, concebendo um sistema multiagente. A pesquisa obteve os dados para simulação por meio de dados históricos de arquivos e projeção de cenários, principalmente com base em perfis históricos dos consumidores e dados meteorológicos locais. O estudo enfatizou o papel crucial das Pequenas e Médias Empresas na economia global e na revolução energética, visando a busca por objetivos de sustentabilidade.

A Tabela 1 lista as principais formas de obtenção de dados de entrada em artigos de 2018 visando a realização de simulação de cenários complexos utilizando modelagem baseada em agentes e sistemas multiagentes.

Tabela 1 - Principais formas de obter dados de entrada em artigos de 2018

Título	Autores	Obtenção de dados de entrada
Modelação Computacional Baseada em Agentes: Enfrentar a Complexidade	Coelho	Dados históricos de arquivos
Developing agent-based models of complex health behaviour	Badham <i>et al.</i>	Dados históricos de arquivos, pesquisas e entrevistas
Agent-based microsimulation conceptual model for urban freight distribution	Gómez-Marín; Arango-Serna; Sema-Urán	Dados históricos de arquivos e pesquisas
Cooperation in manure-based biogas production networks: An agent-based modeling approach	Yazan <i>et al.</i>	Projeção de cenários e pesquisas
Effects of naloxone distribution to likely bystanders: Results of an agent-based model	Keane; Egan; Hawk	Projeção de cenários e entrevistas
Exploring the role of spatial cognition in predicting urban traffic flow through agent-based modelling	Manley; Cheng	Dados históricos de arquivos, projeção de cenários e pesquisas
Agent Based Simulation Model of Virtual Power Plants for greener Manufacturing	Woltmann <i>et al.</i>	Dados históricos de arquivos e projeção de cenários

Fonte: Autor (2023)

Akhtar *et al.* (2019) concentraram seus esforços em um estudo sobre as inundações recorrentes na região de Punjab, Paquistão, reconhecendo as altas perdas humanas e econômicas associadas a esses desastres naturais. O objetivo dos pesquisadores foi desenvolver um sistema crítico de monitoramento e resposta baseado em um sistema multiagente, utilizando métodos de Engenharia de Software baseados em matemática rigorosa. Os autores utilizaram dados históricos de arquivos e pesquisas como fonte principal de informações para simulação. Como resultado, apresentaram um sistema formal para especificação, análise, modelagem e verificação, distribuído com base em multiagentes.

Deng e Yang (2019) abordaram o desafio da regulação de saída cooperativa para sistemas multiagentes lineares com falhas no atuador, considerando falhas de interrupção e perda de eficácia. Os pesquisadores conceberam e implementaram um observador distribuído com o propósito de estimar o estado dos fatores externos, além de propor um controlador adaptativo distribuído que mostrou ser tolerante a falhas, fundamentado nesse observador. Por meio da técnica de projeção de cenários para a obtenção de dados de entrada, os autores validaram a eficácia do método em lidar com o problema investigado.

Long, Su e Liu (2019) publicaram um estudo sobre a controlabilidade de sistemas multiagentes de segunda ordem com duas escalas de tempo. O sistema multiagente foi dividido em um subsistema lento e um rápido, eliminando o parâmetro de perturbação singular. Para a obtenção dos dados de entrada os autores utilizaram projeção de cenários e com isso estabeleceram critérios necessários e/ou suficientes para a controlabilidade de segunda ordem desses sistemas. Seus critérios, baseados nos autovalores das matrizes do sistema, simplificaram a análise. A eficácia desses critérios foi evidenciada por meio de um exemplo de simulação.

Wang e Wang (2019) estudaram o controle de contenção para sistemas multiagentes em tempo discreto com dinâmicas de primeira ordem. Os autores apresentaram condições necessárias e suficientes para o controle de contenção sem atraso temporal a partir de dados de entrada provindo da projeção de cenários. Os pesquisadores destacaram a influência do tamanho do passo e da topologia de interação na estabilidade dos sistemas multiagentes em tempo discreto.

Yang *et al.* (2019) se dedicaram a investigar a regulação cooperativa robusta de sistemas multiagentes lineares e incertos. Os autores propuseram controladores distribuídos baseados em controle adaptativo e esquema de transmissão acionado

por eventos. Usando dados de projeção de cenários, os pesquisadores evidenciaram que esses esquemas alcançaram a regulação cooperativa robusta de saída.

Farias *et al.* (2019) analisaram recursos hídricos por meio da modelagem baseada em agentes, com foco na bacia hidrográfica São Gonçalo e Lagoa Mirim. Utilizando dados históricos de arquivos e projeção de cenários, os autores representaram regiões e rios como agentes. As simulações permitiram análises sobre o consumo de água, produção por região e fluxo de água entre rios, validando a eficácia da modelagem baseada em agentes na gestão de recursos hídricos.

Nagoev *et al.* (2019) se propuseram a desenvolver um modelo de simulação para o reconhecimento de objetos estáticos com base em arquiteturas multiagentes. Utilizando dados históricos de arquivos e projeção de cenários, os autores criaram um modelo que reconhece objetos em transmissões de vídeo e aprende autonomamente com base nas informações multimodais. Esse modelo autônomo tem aplicações promissoras em sistemas de inteligência artificial e robótica, permitindo o reconhecimento de fluxos de dados não estruturados.

A Tabela 2 enumera as principais formas de aquisição de dados de entrada empregadas em artigos publicados em 2019, voltados à simulação de cenários complexos por meio da utilização de modelagem baseada em agentes e sistemas multiagentes.

Tabela 2 - Principais formas de obter dados de entrada em artigos de 2019

Título	Autores	Obtenção de dados de entrada
Hierarchical Coloured Petri-Net Based Multi-Agent System for FMPR	Akhtar <i>et al.</i>	Dados históricos de arquivos e pesquisas
Distributed adaptive fault-tolerant control approach to cooperative output regulation for linear multi-agent systems	Deng; Yang	Projeção de cenários
Second-order controllability of two-time-scale multi-agent systems	Long; Su; Liu	Projeção de cenários
Necessary and sufficient conditions for containment control of multi-agent systems with time delay	Wang; Wang	Projeção de cenários
Robust cooperative output regulation of multi-agent systems via adaptive event-triggered control	Yang <i>et al.</i>	Projeção de cenários
Modelagem Baseada em Agentes para Análise de Recursos Hídricos	Farias <i>et al.</i>	Dados históricos de arquivos e projeção de cenários
A Simulation Model for the Cognitive Function of Static Objects Recognition based on Machine-Learning Multi-agent Architectures	Nagoev <i>et al.</i>	Dados históricos de arquivos e projeção de cenários

Bortoletto, Moitinho e Waldman (2020) exploraram o impacto das mudanças no comportamento do usuário e no fluxo de dados nas operadoras de telecomunicações. Os pesquisadores utilizaram projeção de cenários e dados históricos de arquivos para analisar o comportamento de um duopólio e comparar os resultados da simulação com um modelo analítico previamente estudado. Os resultados da simulação, juntamente com os resultados analíticos, proporcionaram uma compreensão mais profunda da influência dos usuários em sistemas de telecomunicações.

Campos, Cunha e Bueno (2020) se concentraram na disseminação de informações em sistemas socioecológicos. Os autores integraram abordagens de dinâmica de sistemas, modelagem baseada em agentes e análise de redes sociais para analisar o perímetro de irrigação em Minas Gerais. Pesquisas e dados históricos de arquivos foram utilizados para destacar como agentes centrais podem estrategicamente influenciar o fluxo de informação. A pesquisa enfatizou a importância de políticas voltadas para melhorar a disseminação de informações em sistemas, especialmente enfocando líderes locais.

Castro *et al.* (2020) investigaram a integração de sistemas multiagentes e aplicações de Internet das Coisas em um cenário de estacionamento inteligente. Os pesquisadores utilizaram sensores, pesquisas e entrevistas para mostrar como essa integração abrangente pode ser crucial para resolver desafios em cenários urbanos modernos, abordando questões de mobilidade, heterogeneidade e baixo acoplamento tecnológico.

Santos Filho, Mattos e Medeiros (2020) apresentaram um estudo focado na alocação eficiente de tráfego em ambientes de nuvem. Os autores projetaram cenários na tentativa de otimizar o uso da nuvem e atender à demanda variável dos clientes. Os resultados indicaram uma redução na ociosidade da nuvem e uma alocação eficiente de recursos.

Hjorth *et al.* (2020) investigaram o crescente interesse na modelagem de agentes multinível. O objetivo do artigo foi apresentar uma extensão que facilita a criação de modelagens multiníveis na linguagem NetLogo. Os pesquisadores comprovaram a utilidade da extensão em três cenários utilizando dados históricos de arquivos para efetuar a simulação. Os resultados destacaram a facilidade de criar modelagens com essa extensão.

Krupnik, Mordatch e Tamar (2020) conduziram uma análise da aprendizagem por reforço orientada a modelo em problemas de controle contínuo de alta fidelidade

com agentes. Os dados históricos de arquivos e a projeção de cenários foram utilizados para embasar essa análise. Os resultados indicaram que essa técnica apresentou maior eficiência amostral em comparação com uma linha de base sem modelo. Além disso, observou-se que a abordagem poderia aprender comportamentos cooperativos e adversários a partir dos mesmos dados.

Minarsch *et al.* (2020) introduziram um sistema de competição entre agentes de negociação, que refletia uma economia de troca Walrasiana. Cada agente no estudo possuía ativos digitais e preferências sobre eles, negociando individualmente em busca de maximizar suas utilidades por meio de transações registradas em uma *blockchain* simulada. Os dados históricos de arquivos foram utilizados para embasar a análise. O foco principal foi compreender os comportamentos de mercado, a eficácia dos mecanismos de mercado e as estratégias dos próprios agentes de negociação, visando uma futura plataforma de negociação de ativos digitais.

Nogare e Chitnis (2020) conduziram um estudo sobre a auto-organização da peixe-zebra orientado pela linha lateral posterior. Para tal, utilizaram o ambiente de programação baseado em agente NetLogo, permitindo a visualização e exploração das interações do peixe-zebra e seu ambiente. A obtenção de dados baseou-se em dados históricos de arquivos e projeção de cenários. As simulações realizadas no estudo contribuíram para a compreensão das interações complexas e da formação de padrões emergentes do peixe-zebra a partir da linha lateral posterior.

Pereira e Carvalho (2020) dedicaram-se ao estudo da aplicação combinada das técnicas de campos potenciais artificiais e sistemas multiagentes, especificamente em veículos aéreos não tripulados e veículos terrestres não tripulados. A obtenção de dados para a pesquisa foi realizada por meio de experimentos com projeção de cenários. Os experimentos avaliaram a cooperação de veículos com diferentes capacidades de movimentação, o que representou uma contribuição significativa em relação a pesquisas anteriores.

Tavares, Billa e Adamatti (2020) se dedicaram ao desenvolvimento de um modelo em sistemas multiagentes para o estudo de emoções e personalidades. A obtenção de dados para o estudo baseou-se em dados históricos de arquivos, pesquisas e entrevistas. Com a implementação desse estudo, os autores buscaram uma visão mais abrangente, representativa e fiel do comportamento humano, com potencial aplicação em diversos domínios.

A Tabela 3 apresenta um registro das principais abordagens de aquisição de

dados de entrada utilizadas em estudos publicados em 2020, que se concentraram na simulação de cenários complexos por meio da aplicação de modelagem baseada em agentes e sistemas multiagentes.

Tabela 3 - Principais formas de obter dados de entrada em artigos de 2020

Título	Autores	Obtenção de dados de entrada
Utilização de Sistemas Multiagentes para a Simulação do Comportamento de Usuários em Redes de Telecomunicações	Bortoletto; Moitinho; Waldman	Dados históricos de arquivos e projeção de cenários
Disseminação de Informações em Sistemas Socioecológicos: Análise de um Modelo Híbrido de Dinâmica de Sistemas e Modelagem Baseada em Agentes	Campos; Cunha; Bueno	Dados históricos de arquivos e pesquisas
Integrando Sistemas Multi-Agentes Embarcados, Simulação Urbana e Aplicações de IoT	Castro <i>et al.</i>	Sensores, pesquisas e entrevistas
Agentes Inteligentes baseados em Aprendizado por Reforço para Alocação Dinâmica de Tráfego em Nuvens	Santos Filho; Mattos; Medeiros	Projeção de cenários
LevelSpace: A NetLogo Extension for Multi-Level Agent-Based Modeling	Hjorth <i>et al.</i>	Dados históricos de arquivos
Multi-Agent Reinforcement Learning with Multi-Step Generative Models	Krupnik; Mordatch; Tamar	Dados históricos de arquivos e projeção de cenários
Trading Agent Competition with Autonomous Economic Agents	Minarsch <i>et al.</i>	Dados históricos de arquivos
NetLogo agent-based models as tools for understanding the self-organization of cell fate, morphogenesis and collective migration of the zebrafish posterior Lateral Line primordium	Nogare; Chitnis	Dados históricos de arquivos e projeção de cenários
Campos Potenciais Artificiais e Consenso Aplicados ao Problema de Navegação de Sistemas Multiagentes Heterogêneo	Pereira; Carvalho	Projeção de cenários
Uma extensão ao modelo de emoções e de personalidades baseado em sistemas multiagentes utilizando distribuição normal, lógica Fuzzy e matemática intervalar	Tavares; Billa; Adamatti	Dados históricos de arquivos, pesquisas e entrevistas

Fonte: Autor (2023)

Zhang *et al.* (2021) investigaram o aprendizado por reforço orientado a modelos em jogos Markov com foco em sistemas multiagentes. Utilizando projeção de cenários, os pesquisadores examinaram um cenário básico de jogos Markov com dois jogadores. A pesquisa apresentou resultados substanciais para o campo de aprendizado de máquina, especialmente em cenários complexos em contextos de aprendizado multiagente, onde encontrar soluções eficientes é desafiador.

Almagor *et al.* (2021) concentraram-se nas intervenções para aumentar a atividade física das crianças. Por meio de projeção de cenários e dados históricos de arquivos, os autores simularam o impacto de intervenções como brincadeiras ao ar livre, educação física escolar e viagens ativas. As descobertas do estudo destacaram como é possível melhorar a saúde das crianças, mas também enfatizaram a importância de considerar fatores socioeconômicos ao efetuar o planejamento.

Lange *et al.* (2021) abordaram a viabilidade de modelos de negócios circulares em redes empresariais. Com base em projeção de cenários, entrevistas e pesquisas, os pesquisadores desenvolveram um modelo baseado em agentes e aplicaram-no a uma rede de empresas comerciais. As simulações mostraram como negócios circulares podem ser otimizados antes da implementação.

Srikrishnan e Keller (2021) focaram na melhoria da precisão de modelos baseados em agentes utilizando dados históricos de arquivos. A partir do estudo, os pesquisadores destacaram os desafios de calibrar esses modelos, onde a qualidade dos dados pode afetar significativamente os resultados e as conclusões.

Kou *et al.* (2021) analisaram a propagação da COVID-19 em cenários dinâmicos. Usando dados históricos de arquivos e projeção de cenários, os pesquisadores avaliaram o impacto das medidas de intervenção, como vacinação e isolamento. As simulações forneceram informações críticas para a tomada de decisões em saúde pública durante a pandemia.

Moya *et al.* (2021) investigaram o impacto do distanciamento ideológico entre eleitores e partidos políticos. Apoiado em dados históricos de arquivos, pesquisas e projeção de cenários, os pesquisadores desenvolveram um modelo baseado em agentes para explorar como as estratégias políticas afetam as percepções ideológicas. O artigo evidenciou relevância ao entender o comportamento político em contextos de ideologia entre eleitores e partidos políticos.

Xie *et al.* (2021) examinaram o potencial da irrigação durante a estação seca na África Subsaariana, usando dados históricos de arquivos. As simulações identificaram oportunidades de investimento em pequenas irrigações, o que pode ser fundamental para enfrentar desafios de segurança alimentar na região.

Burg *et al.* (2021) investigaram a tendência de agricultores suíços em adotar instalações de biogás. Baseado em dados históricos de arquivos, pesquisas e entrevistas, os autores destacaram a importância de investimentos como o principal fator para os agricultores adotarem essas instalações, destacando que as

descobertas têm implicações para a transição energética e a produção de biogás.

Benson *et al.* (2021) analisaram a migração de enguias em ambientes aquáticos. Fundamentado em dados históricos de arquivos e sensores, os pesquisadores desenvolveram um modelo baseado em agentes para entender como as enguias usam informações ambientais para orientar sua migração.

Ghaffarian *et al.* (2021) concentraram-se na gestão de riscos de desastres e na recuperação pós-desastre. Utilizando dados históricos de arquivos e projeção de cenários, os autores desenvolveram um modelo baseado em agentes para simular a recuperação em áreas urbanas após um desastre.

A Tabela 4 expõe uma compilação das abordagens preeminentes empregadas para a aquisição de dados de entrada em estudos publicados no ano de 2021.

Tabela 4 - Principais formas de obter dados de entrada em artigos de 2021

Título	Autores	Obtenção de dados de entrada
Model-Based Multi-Agent RL in Zero-Sum Markov Games with Near-Optimal Sample Complexity	Zhang <i>et al.</i>	Projeção de cenários
How can an agent-based model explore the impact of interventions on children's physical activity in an urban environment?	Almagor <i>et al.</i>	Dados históricos de arquivos e projeção de cenários
Agent-based modelling and simulation for circular business model experimentation	Lange <i>et al.</i>	Projeção de cenários, pesquisas e entrevistas
Small increases in agent-based model complexity can result in large increases in required calibration data	Srikrishnan; Keller	Dados históricos de arquivos
A multi-scale agent-based model of infectious disease transmission to assess the impact of vaccination and non-pharmaceutical interventions: The COVID-19 case	Kou <i>et al.</i>	Dados históricos de arquivos e projeção de cenários
Simulating the influence of terror management strategies on the voter ideological distance using agent-based modeling	Moya <i>et al.</i>	Dados históricos de arquivos, projeção de cenários e pesquisas
Mapping development potential of dry-season small-scale irrigation in Sub-Saharan African countries under joint biophysical and economic constraints - An agent-based modeling approach with an application to Ethiopia	Xie <i>et al.</i>	Dados históricos de arquivos
Farmer's willingness to adopt private and collective biogas facilities: An agent-based modeling approach	Burg <i>et al.</i>	Dados históricos de arquivos, pesquisas e entrevistas
Agent-based modelling of juvenile eel migration via selective tidal stream transport	Benson <i>et al.</i>	Dados históricos de arquivos e sensores
Agent-based modelling of post-disaster recovery with remote sensing data	Ghaffarian <i>et al.</i>	Dados históricos de arquivos e projeção de cenários

Fonte: Autor (2023)

Barnett-Neefs, Wiedmann e Ivanek (2022) investigaram a contaminação persistente por listeriose em armazéns usando modelos baseados em agentes. O estudo visou desenvolver ferramentas de apoio à decisão baseadas em simulação para prever locais propensos à contaminação e ações corretivas. A partir de dados históricos e simulações de cenários, os pesquisadores adaptaram um modelo baseado em agentes para dinâmicas de contaminação. Os resultados indicaram que modelos baseados em agentes são ferramentas valiosas na identificação de padrões de contaminação e na avaliação de ações corretivas.

Mehdzadeh, Nordfjaern e Klockner (2022) realizaram uma revisão sistemática da aplicação de modelos baseados em agentes em estudos de transição de mobilidade. Com base em dados históricos de arquivos, sensores e projeção de cenários, analisaram diversos estudos nesse campo. A pesquisa identificou lacunas e desafios na literatura atual sobre transição de mobilidade.

Parviero *et al.* (2022) investigaram a propagação de novos produtos por meio de um modelo baseado em agentes. Os autores utilizaram dados históricos de arquivos, pesquisas e entrevistas para entender como influências e fatores externos afetam as previsões e inferências. A pesquisa proporcionou avanços significativos na predição e validação de novos produtos no mercado.

Liu *et al.* (2022) exploraram a urbanização sustentável de paisagens culturais em Nansha, China. Por meio de entrevistas e pesquisas, coletaram dados para criar um modelo baseado em agentes que avaliou estratégias de urbanização. A pesquisa concluiu que o desenvolvimento rural sustentável deve considerar adaptações locais e a participação das partes interessadas.

Stevens *et al.* (2022) avaliaram a viabilidade financeira para serviços de transporte público. Os pesquisadores utilizaram dados históricos de arquivos e projeção de cenários para a modelagem da demanda de passageiros e as rotas dos veículos públicos. Os resultados indicaram que os serviços de transporte público se tornavam mais lucrativos quando se considerava o uso de veículos automatizados e o compartilhamento de viagens dinâmicas.

Kjær e Schaubert (2022) desenvolveram um modelo baseado em agentes para estudar a dinâmica de uma doença crônica em cervos-de-cauda-branca. Com base em sensores e projeção de cenários, os autores obtiveram os parâmetros comportamentais e demográficos, considerando rotas de transmissão direta e indireta. A partir disso, a simulação revelou o impacto da estrutura da paisagem e do modo de

transmissão e prevalência da doença.

Gracia-Lázaro, Dercole e Moreno (2022) investigaram os fatores determinantes dos processos de saída de uniões econômicas, sendo influenciados pelo contexto da saída do Reino Unido da União Europeia. A partir de um modelo baseado em agentes que incorporou projeções de cenários e dados históricos de arquivos, os pesquisadores identificaram que os principais impulsionadores das retiradas de nações dessas uniões econômicas foram impostos, taxas alfandegárias e desigualdade de riqueza.

Fan *et al.* (2022) abordaram o impacto da pandemia da Covid-19 na tomada de decisões relacionadas a viagens públicas. Os autores desenvolveram um modelo que incluiu projeções de cenários para avaliar a percepção de risco dos agentes e informações de risco governamental. Os resultados revelaram a diminuição na probabilidade de viagens à medida que a percepção de risco aumentava e destacaram a importância da credibilidade governamental na gestão da pandemia.

Chersoni, Dellavalle e Fontana (2022) desenvolveram um modelo abrangente para analisar os investimentos em isolamento térmico na União Europeia. A fim de alimentar esse modelo, os pesquisadores utilizaram dados históricos provenientes de arquivos da pesquisa de consumo de 2016, que ofereceu informações cruciais sobre o comportamento dos mercados de eletricidade destinados aos consumidores na União Europeia. A abordagem adotada não se limitou apenas a fatores econômicos, mas também abrangeu aspectos comportamentais e sociais. Ao considerar uma gama diversificada de influências, como situações financeiras percebidas e preocupações ambientais dos proprietários, os pesquisadores foram capazes de capturar a complexidade desse cenário. Os resultados destacaram a interconexão desses motivos, ressaltando que políticas eficazes não devem se concentrar exclusivamente em incentivos financeiros, mas também em campanhas pró-ambientais.

Butts *et al.* (2022) investigaram o comportamento de veados silvestres fundamentada em dados históricos de arquivos e sensores. A modelagem baseada em agentes permitiu não apenas a simulação do movimento de grupos de veados silvestres, mas também a análise da propagação de doenças infecciosas nas populações desses animais. Essa pesquisa tem implicações importantes para a gestão da saúde animal e a ecologia da fauna selvagem.

A Tabela 5 apresenta uma síntese das estratégias de coleta de dados utilizadas na obtenção de informações de entrada em pesquisas publicadas durante 2022.

Tabela 5 - Principais formas de obter dados de entrada em artigos de 2022

Título	Autores	Obtenção de dados de entrada
Examining Patterns of Persistent Listeria Contamination in Packinghouses Using Agent-Based Models	Barnett-Neefs; Wiedmann; Ivanek	Dados históricos de arquivos e projeção de cenários
A systematic review of the agent-based modelling/simulation paradigm in mobility transition	Mehdizadeh; Nordfjaern; Klockner	Dados históricos de arquivos, projeção de cenários e sensores
An agent-based model with social interactions for scalable probabilistic prediction of performance of a new product	Parviero <i>et al.</i>	Dados históricos de arquivos, pesquisas e entrevistas
Using agent-based modeling to assess multiple strategy options and trade-offs for the sustainable urbanization of cultural landscapes: A case in Nansha, China	Liu <i>et al.</i>	Entrevistas e pesquisas
An agent-based model for assessing the financial viability of autonomous mobility on-demand systems used as first and last-mile of public transport trips: A case-study in Rotterdam, the Netherlands	Stevens <i>et al.</i>	Dados históricos de arquivos e projeção de cenários
The effect of landscape, transmission mode and social behavior on disease transmission: Simulating the transmission of chronic wasting disease in white-tailed deer (<i>Odocoileus virginianus</i>) populations using a spatially explicit agent-based model	Kjær; Schaubert	Sensores e projeção de cenários
Dynamics of economic unions: An agent-based model to investigate the economic and social drivers of withdrawals	Gracia-Lázaro; Dercole; Moreno	Dados históricos de arquivos e projeção de cenários
An agent-based model for public travel decision-making considering risk perception on the pandemic	Fan <i>et al.</i>	Projeção de cenários
Modelling thermal insulation investment choice in the EU via the behaviourally informed agent-based model	Chersoni; Dellavalle; Fontana	Dados históricos de arquivos
Data-driven agent-based model building for animal movement through Exploratory Data Analysis	Butts <i>et al.</i>	Dados históricos de arquivos e sensores

Fonte: Autor (2023)

Derkenbaeva *et al.* (2023) realizaram um estudo com agentes inteligentes que examinou as escolhas energéticas dos proprietários de imóveis em Amsterdã, com base em dados históricos de arquivos e projeção de cenários. Os resultados destacaram a necessidade de análises a nível de bairro para uma melhor escolha energética, uma vez que contextos diferentes influenciam as decisões.

Mussawar *et al.* (2023) investigaram a sustentabilidade energética enfatizando a influência em um contexto urbano. Por meio de projeção de cenários, o estudo utilizou simulações multiagentes para avaliar o desempenho energético em edifícios

comerciais. Descobriu-se que a autossuficiência energética pode variar em até 45% dependendo dos materiais utilizados na construção dos prédios.

Kaziyeva *et al.* (2023) conceberam um modelo baseado em agentes destinado à simulação dos fluxos de tráfego de pedestres em ambientes urbanos. Os dados de entrada foram adquiridos mediante pesquisas e entrevistas. As simulações, focadas na cidade de Salzburg, revelaram um tráfego intenso nas áreas centrais e ao longo do rio, fornecendo *insights* cruciais para o aprimoramento das estratégias urbanas.

Yoon *et al.* (2023) exploraram o risco de inundações por meio de simulações baseadas em agentes. Por meio da projeção de cenários, os pesquisadores desenvolveram um modelo destinado a simular o risco de inundação em cenários urbanos. A pesquisa destacou a importância de considerar fatores-chave, como escolhas de localização e infraestrutura, na análise do risco de inundação.

Gong *et al.* (2023) abordaram as motivações por trás do movimento humano, fazendo uso de dados históricos de arquivos, sensores e projeção de cenários. O modelo baseado em agentes proporcionou uma compreensão mais precisa do movimento em ambientes urbanos, evidenciando a sensibilidade geográfica dos destinos e os efeitos das motivações dos passageiros como elementos cruciais.

Hunter *et al.* (2023) se dedicaram a examinar o papel do rastreamento de contatos na contenção da COVID-19 na Irlanda. Fundamentada em dados históricos de arquivos e projeções de cenários, a pesquisa baseada em multiagentes destacou a importância do rastreamento de contatos como uma ferramenta essencial na gestão de surtos de doenças infecciosas. A ausência do rastreamento resultaria em um acentuado aumento nos casos e nas taxas de mortalidade.

James e Bradshaw (2023) desenvolveram um modelo baseado em agentes com o propósito de fornecer percepções relevantes para a indústria do mel, considerando a colheita sustentável e a adaptação às mudanças climáticas. Utilizando dados históricos de arquivos, o estudo mostrou uma gama diversificada de abordagens para abordar os desafios em sistemas socioecológicos complexos.

Wise *et al.* (2023) exploraram uma modelagem multiagente sobre a propagação de COVID-19 em diferentes bairros do Zimbábue, combinando dados históricos de arquivos e projeção de cenários. A pesquisa destacou que a população infectada afeta a dinâmica geográfica da disseminação da doença, enfatizando a importância da representação precisa em simulações epidemiológicas.

Qiao *et al.* (2023) investigaram a dinâmica de transmissão do COVID-19 em

canteiros de obras baseada em simulações multiagentes. A partir da projeção de cenários, pesquisas e entrevistas, a pesquisa forneceu orientações para tomadores de decisão, auxiliando na mitigação da propagação de doenças infecciosas.

De Bosscher, Mohammadi Ziabari e Sharpanskykh (2023) exploraram as operações de terminais de aeroportos por meio de agentes, com dados históricos de arquivos. Os resultados destacaram a influência de estratégias inadequadas de pessoal e altas taxas de ocupação em voos, afetando a jornada dos passageiros e o fluxo nos postos de controle de segurança.

A Tabela 6 resume as estratégias para a coleta de dados empregadas para adquirir informações de entrada em estudos publicados no ano de 2023.

Tabela 6 - Principais formas de obter dados de entrada em artigos de 2023

Título	Autores	Obtenção de dados de entrada
Simulating households' energy transition in Amsterdam: An agent-based modeling approach	Derkenbaeva <i>et al.</i>	Dados históricos de arquivos e projeção de cenários
Performance and prospects of urban energy communities conditioned by the built form and function: A systematic investigation using agent-based modeling	Mussawar <i>et al.</i>	Projeção de cenários
Large-scale agent-based simulation model of pedestrian traffic flows	Kaziyeva <i>et al.</i>	Entrevistas e pesquisas
Structural model choices regularly overshadow parametric uncertainty in agent-based simulations of household flood risk outcomes	Yoon <i>et al.</i>	Projeção de cenários
Agent-based modelling with geographically weighted calibration for intra-urban activities simulation using taxi GPS trajectories	Gong <i>et al.</i>	Dados históricos de arquivos, projeção de cenários e sensores
Assessing the impact of contact tracing with an agent-based model for simulating the spread of COVID-19: The Irish experience	Hunter <i>et al.</i>	Dados históricos de arquivos e projeção de cenários
Agent-based model development of a complex socio-ecological system: Methods for overcoming data and domain limitations	James; Bradshaw	Dados históricos de arquivos
Scale matters: Variations in spatial and temporal patterns of epidemic outbreaks in agent-based models	Wise <i>et al.</i>	Dados históricos de arquivos e projeção de cenários
An interactive agent-based modelling framework for assessing COVID-19 transmission risk on construction site	Qiao <i>et al.</i>	Projeção de cenários, pesquisas e entrevistas
A comprehensive study of agent-based airport terminal operations using surrogate modeling and simulation	De Bosscher; Mohammadi Ziabari; Sharpanskykh	Dados históricos de arquivos

Fonte: Autor (2023)

Esta revisão sistemática da literatura, que abrange o período de 2018 a 2023, examina um total de 54 artigos que discutem a aplicação de sistemas multiagentes em simulações de cenários complexos. O levantamento destaca a variedade de abordagens empregadas para a aquisição de dados de entrada.

Notavelmente, um enfoque substancial recai sobre a combinação de técnicas para aprimorar a precisão dos modelos. A maioria das pesquisas, especificamente 15, adota uma abordagem híbrida que integra dados históricos de arquivos com a projeção de cenários. Uma segunda vertente significativa, representada por 10 artigos, se concentra exclusivamente na projeção de cenários.

Após minuciosa análise, constata-se que, até o presente momento, não foi identificada a implementação da técnica de raspagem de dados para atender ao propósito mencionado. A Tabela 7 apresenta um resumo dos artigos analisados.

Tabela 7 - Quantidade de cada tipo de obtenção de dados de entrada

Obtenção de dados de entrada	Quantidade por ano						TOTAL
	2018	2019	2020	2021	2022	2023	
Dados históricos de arquivos	1	0	2	2	1	2	8
Dados históricos de arquivos e pesquisas	1	1	1	0	0	0	3
Dados históricos de arquivos, pesquisas e entrevistas	1	0	1	1	1	0	4
Dados históricos de arquivos e projeção de cenários	1	2	3	3	3	3	15
Dados históricos de arquivos, projeção de cenários e pesquisas	1	0	0	1	0	0	2
Dados históricos de arquivos, projeção de cenários e sensores	0	0	0	0	1	1	2
Dados históricos de arquivos e sensores	0	0	0	1	1	0	2
Entrevistas e pesquisas	0	0	0	0	1	1	2
Entrevistas, pesquisas e projeção de cenários	1	0	0	0	0	0	1
Projeção de cenários	0	4	2	1	1	2	10
Projeção de cenários e pesquisas	1	0	0	0	0	0	1
Projeção de cenários, pesquisas e entrevistas	0	0	0	1	0	1	2
Sensores e projeção de cenários	0	0	0	0	1	0	1
Sensores, pesquisas e entrevistas	0	0	1	0	0	0	1
TOTAL	7	7	10	10	10	10	54

Fonte: Autor (2023)

Ressalta-se também que essas formas de aquisição de dados apresentam elementos dificultadores. De acordo com Fontes e Souza (2021), as principais desvantagens dessas abordagens são:

- Os dados históricos de arquivos estão sujeitos a vieses e erros de registro, sendo que eles podem estar incompletos ou imprecisos, devido a problemas como a falta de padronização no registro de dados e a perda de informações ao longo do tempo. Além disso, os dados históricos podem ser afetados por vieses de seleção, pois apenas as informações que foram registradas e preservadas ao longo do tempo estão disponíveis para análise.
- A projeção de cenários é altamente dependente da qualidade dos dados gerados pelos modelos, sendo crucial garantir que se gere resultados confiáveis, uma vez que se não refletir com precisão o comportamento do sistema real, ela pode produzir dados imprecisos e levar a conclusões erradas.
- As pesquisas podem ser influenciadas pela posição e perspectiva do pesquisador, o que pode afetar a qualidade e a confiabilidade dos dados coletados. Além disso, as respostas fornecidas pelos entrevistados podem não ser confiáveis, já que muitas vezes as pessoas tendem a fornecer respostas que acham que o pesquisador quer ouvir, ou que sejam socialmente aceitáveis.
- As entrevistas podem apresentar limitações afetadas pelo tamanho da amostra e pela representatividade da população estudada. E, também, o custo elevado e o tempo para a obtenção das respostas são elementos complicadores.
- Os sensores podem apresentar algumas desvantagens, como serem caros e difíceis de implementar, especialmente em ambientes hostis ou inacessíveis. Além disso, os sensores podem produzir grandes quantidades de dados que podem ser difíceis de gerenciar e analisar. Outra limitação é que a precisão dos sensores pode ser afetada por fatores externos, como interferência eletromagnética ou condições climáticas adversas.

A tese propõe uma nova abordagem para a coleta de dados em simulações com sistemas multiagentes em cenários complexos, utilizando uma técnica baseada em raspagem de dados. Essa abordagem enriquece a simulação ao incorporar dados dinâmicos de fontes diversas, proporcionando realismo e atualidade. Assim, a pesquisa estabelece um novo padrão ao aplicar a raspagem de dados em simulações

multiagentes. Essa abordagem não apenas inova na coleta e integração de dados, mas também estabelece um referencial para futuras pesquisas e aplicações práticas nesse campo.

Em síntese, as vantagens desta técnica incluem a coleta de dados rápida e eficaz, a capacidade de extrair dados de fontes não estruturadas e a melhoria na validade e confiabilidade das simulações multiagentes ao integrar dados externos. Esta pesquisa não apenas inova, mas também estabelece um novo padrão para pesquisas e aplicações práticas, propondo melhorias na forma que os dados são coletados, processados e integrados em simulações de alta complexidade.

1.5. ESTRUTURA DA TESE

A presente tese está organizada em oito capítulos. O Capítulo 1 apresenta a contextualização do tema, assim como o problema abordado, a hipótese que será trabalhada, a motivação, os objetivos, revisão da literatura e a estrutura da tese.

O Capítulo 2 versa sobre os agentes autônomos, bem como suas definições, classes, classificações, arquitetura, aplicações e elucida o conceito de ambiente.

O Capítulo 3 expõe as definições sobre sistemas multiagentes, assim como a comunicação, a aprendizagem e a coordenação desses sistemas.

O Capítulo 4 discorre sobre os conceitos de Ciência de Dados, *Big Data*, mineração de dados e a raspagem de dados.

O Capítulo 5 explana a metodologia científica empregada, assim como, os materiais, métodos e técnicas utilizadas nas experimentações.

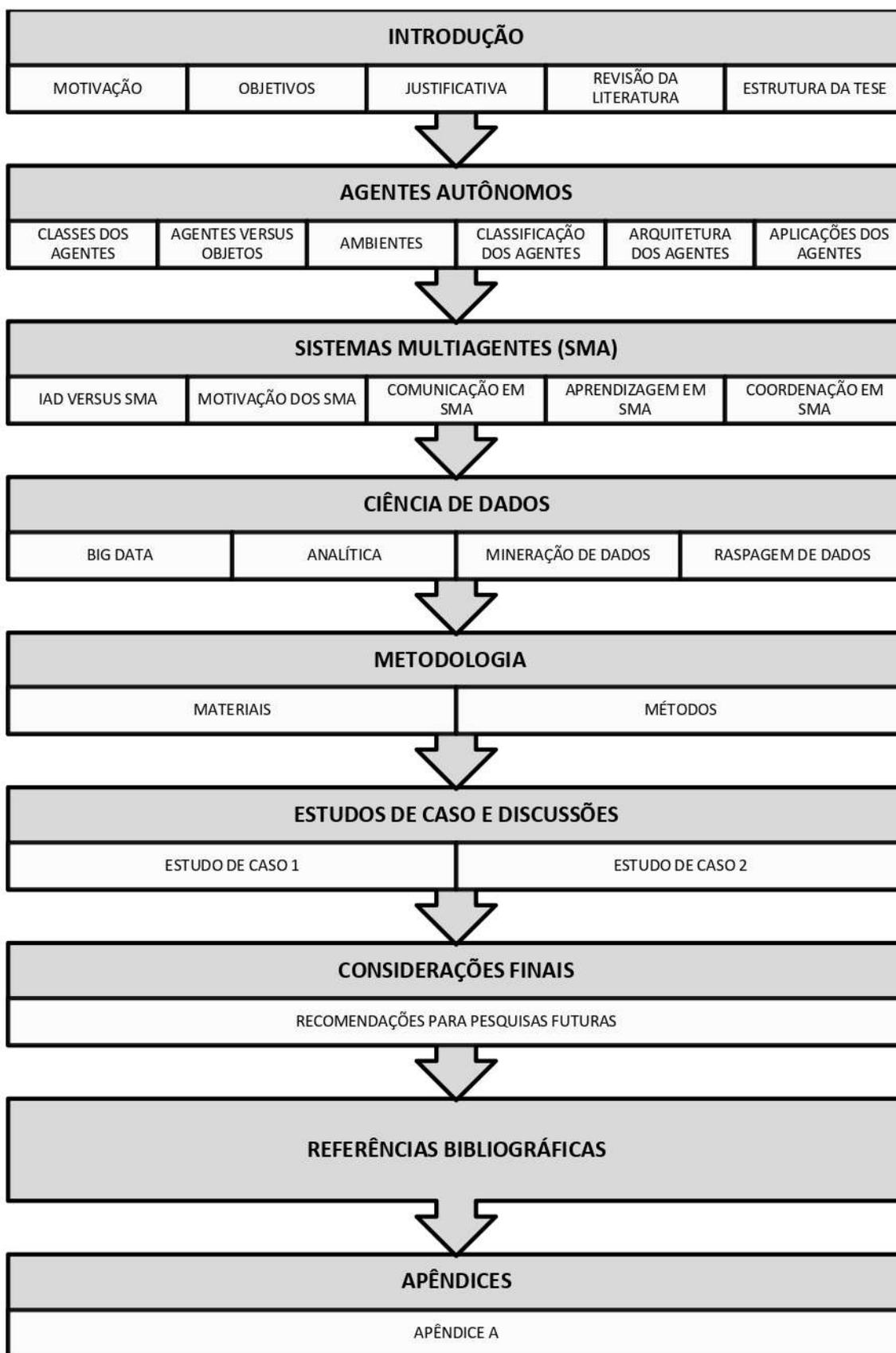
O Capítulo 6 concentra-se nos Estudos de Caso, com o Brasil como objeto de estudo. O primeiro estudo de caso investigou os determinantes da adoção de sistemas solares, enquanto o segundo analisou os principais fatores que contribuíram para a redução do consumo de energia elétrica. Além disso, os estudos de caso incluíram a contextualização, a análise dos resultados e as discussões sobre as experimentações.

O Capítulo 7 traz as considerações finais do projeto e as perspectivas de pesquisas futuras.

O último Capítulo finaliza com as referências bibliográficas utilizadas na tese.

A estrutura da tese é apresentada na Figura 1, com os títulos dos capítulos e subcapítulos organizados de forma hierárquica.

Figura 1 - Estrutura da Tese



Fonte: Autor (2024)

2. AGENTES AUTÔNOMOS

Os agentes são aplicados em diversas áreas, desde a interação humano-computador até processos de controle industrial. Devido às diversas aplicações em diferentes áreas, há várias definições sobre o conceito de agentes e, não existe um consenso entre os autores sobre essa definição.

Segundo Russel e Norvig (1995), um agente é definido como uma sequência ordenada de elementos: $\langle P, A, I, in, out \rangle$. Nessa sequência, P representa o conjunto de objetos de entrada, ou seja, as percepções do agente, enquanto A é o conjunto de objetos de saída, ou seja, as ações que o agente pode realizar. I refere-se ao conjunto de estados internos do agente. Além disso, há duas funções importantes: "in", que mapeia as percepções do ambiente para o agente, e "out", que mapeia as ações realizadas pelo agente. Por sua vez, o ambiente do agente é descrito por outra sequência ordenada: $\langle A, P, W, see, do \rangle$. Aqui, A representa o conjunto de operações de saída disponíveis para o agente, P é o conjunto de operações de entrada que o agente pode receber, e W é o conjunto de estados possíveis no ambiente. A função see é responsável por mapear as operações que o agente pode realizar no ambiente, enquanto a função do mapeia as operações que o agente já executou no ambiente. Essas definições de agente e ambiente estabelecidas por Russel e Norvig (1995) são fundamentais para a compreensão da interação entre um agente e seu ambiente em sistemas inteligentes.

De acordo com Zhang (2017), os agentes são entidades de *software* que possuem persistência e um objetivo específico. A característica da persistência distingue os agentes das sub-rotinas, uma vez que os primeiros possuem sua própria maneira de executar as funções designadas. Além disso, seu objetivo os diferencia das aplicações multifuncionais, já que são desenvolvidos com códigos mais concisos.

Para Zappe (2012), agentes inteligentes desempenham sempre três funções: perceber o ambiente; agir para afetar as condições do ambiente; interpretar as percepções e assim, resolver problemas e realizar ações definidas previamente.

De acordo com Bai, Ren e Fujita (2017), um agente é algo que tem a percepção de ambientes via sensores e age nesse ambiente por meio de atuadores. De acordo com Bazzan e Klugl (2011), um agente pode ser um sistema computacional baseado em *software* ou uma peça de *hardware* que possui as seguintes propriedades:

autonomia, proatividade, reatividade e habilidade social de interação.

Para Ahlbrecht, Dix e Fiekas (2020), agentes são sistemas computacionais que se encontram em ambientes dinâmicos e complexos, que agem autonomamente nesses ambientes para realizar os objetivos que foram projetados.

Os pesquisadores Kudenko, Kazakov e Alonso (2005) definem agentes autônomos como sistemas situados em um ambiente, capazes de perceber e agir ao longo do tempo, com o objetivo de realizar ações projetadas.

De acordo com o artigo publicado pela IBM (1997), agentes inteligentes são entidades de *software* que operam em um ambiente com um certo grau de autonomia, atingindo os objetivos estabelecidos em sua programação.

Jezic, Chen-Burger e Kusek (2020) afirmam que um agente é um sistema com a capacidade de agir ou que possui o poder e a autoridade para agir em um ambiente específico.

Segundo Wooldridge (2009), um agente é um sistema computacional inserido em um determinado ambiente, com o qual interage por meio de ações pré-programadas.

Zappe (2012) destaca que o estudo e desenvolvimento de agentes autônomos surgiram a partir da contribuição de diversas áreas científicas, como Inteligência Artificial, Sistemas Distribuídos e Redes de Computadores, Sociologia, Engenharia de Software, Teoria dos Jogos e Economia.

- Inteligência Artificial: Microaspectos, tais como a resolução de problemas, raciocínio lógico, utilização de conhecimento prévio, planejamento, aprendizagem contínua etc.
- Sistemas Distribuídos e Redes de Computadores: Arquiteturas para agentes, sistemas multiagentes, comunicação, coordenação e cooperação.
- Sociologia: Macroaspectos, tais como a desenvolvimento de sociedades virtuais e a interação entre os agentes autônomos.
- Engenharia de Software: agentes autônomos como abstração, programação orientada a agentes.
- Teoria dos Jogos e Economia: Negociação, resolução de conflitos, e mecanismos de mercado.

A influência dessas áreas pode ser observada na Figura 2.

Figura 2 - Áreas que originaram os agentes e Sistemas Multiagentes



Fonte: Zappe (2012)

Inicialmente a Inteligência Artificial apresentou grande influência nos estudos e pesquisas dos agentes autônomos e sistemas multiagentes, porém, atualmente, esse campo evoluiu a tal ponto de ser considerado uma nova área de pesquisa e não apenas uma subárea da Inteligência Artificial.

Segundo Russel e Norvig (1995), observa-se ao longo dos últimos anos um crescimento exponencial nas pesquisas sobre agentes autônomos ou agentes inteligentes. Todavia, como toda nova área, ainda há desafios para consolidar esse ramo de pesquisa, sendo essencialmente:

- A ausência de um paradigma de programação bem definido e elaborado para o desenvolvimento de sistemas distribuídos.
- A tentativa de paradigmas voltados a agentes tentarem resolver situações problema da mesma forma que o paradigma de orientação a objetos.
- A utilização do termo agente de forma vaga e, muitas vezes errônea, para descrever qualquer sistema da atualidade.

Em consonância com as pesquisas de Wooldridge (2009), Zappe (2012), Russel e Norvig (1995) e Zhang (2017), esse trabalho adotará a seguinte definição para agentes:

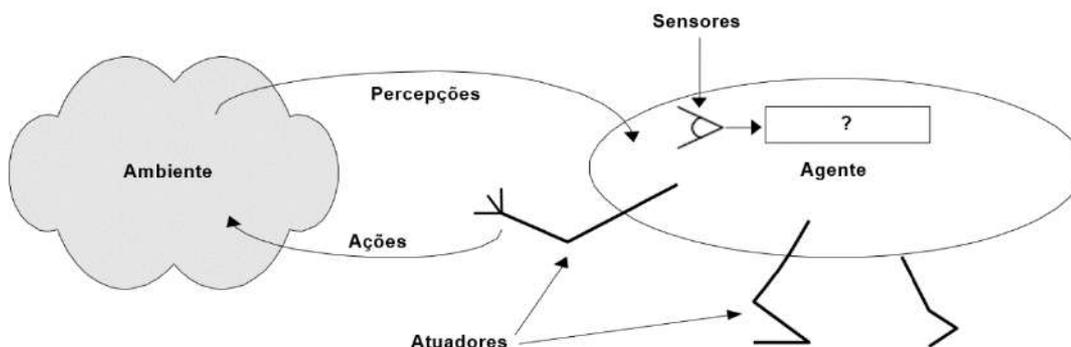
Um agente é um sistema computacional com capacidade de decisão, autonomia e comunicação de alto nível, que está estabelecido em um determinado ambiente. Destacando que:

1. Agem de forma autônoma por meio de atuadores.
2. Realizam a comunicação com outros agentes e/ou pessoas por meio de interações com ações pré-programadas.
3. Reconhecem o ambiente em que estão estabelecidos por meio de sensores.

Esses ambientes podem ser: um estabelecimento real (universidade, hospital etc.), um ambiente simulado ou mesmo um equipamento (computador de mesa ou computador portátil). Os agentes mais comuns são os de *software*, ou seja, existem apenas inseridos em um ambiente virtual, porém há agentes físicos, conhecidos como agentes robóticos (Sebem; Leal, 2019).

Independentemente dos tipos de agentes ou ambiente, é essencial que os mesmos reconheçam o ambiente em que estão inseridos e, com isso, tenham condições de agir de forma autônoma (Nedić; Pang; Scutari, 2018). Assim, os agentes devem possuir sensores e atuadores apropriados à execução das atividades para as quais foram projetados. A Figura 3 representa o esquema típico de um agente.

Figura 3 - Esquema típico de um agente



Fonte: Nedić; Pang; Scutari (2018)

Os humanos percebem os ambientes por meio dos cinco sentidos (a visão, o

olfato, o paladar, a audição e o tato) e, agem nesses ambientes, usando os seus atuadores (braços, pernas etc.). Os agentes robóticos reconhecem os ambientes por meio de sensores, tais como: sensores de proximidade, sensores de aceleração, sensores de calor, entre outros e, agem nos ambientes a partir de atuadores, que podem ser motores, rodas, braços robóticos, entre outros (Pereira; Carvalho, 2020). Na esfera dos agentes de *software*, a definição dos sensores e atuadores do agente apresenta uma complexidade acrescida. Por exemplo, um agente desenvolvido para jogar xadrez possui como sensores as funções que analisam quais são as peças existentes no tabuleiro, bem como suas respectivas posições e, por sua vez, os atuadores são as funções que realizam a movimentação das peças do agente pelo tabuleiro (Li; Duan, 2014).

2.1. CLASSES DOS AGENTES

Um agente é desenvolvido a partir de um conjunto de características. A escolha de quais atributos devem estar presentes em cada agente depende da funcionalidade e dos objetivos que o mesmo deve desempenhar (Bortoletto; Moitinho; Waldman, 2020). Em consonância com a pesquisa de Bai, Ren e Fujita (2017), o conjunto de atributos pode ser utilizado para agrupar os agentes em classes, sendo as principais descritas a seguir.

A) AUTONOMIA

A autonomia é um princípio fundamental no qual os agentes operam com base em suas próprias regras de tomada de decisão, buscando alcançar seus objetivos, enquanto são supervisionados por seres humanos (Li; Duan, 2014). Esse atributo é consensual aos investigadores da área dos agentes inteligentes. Por sua vez, para Jezic, Chen-Burger e Kusek (2020) não há uma autonomia plena dos agentes. Isso ocorre porque o desenvolvimento e o início do funcionamento dos agentes ocorrem por meio de um humano, além de que, usualmente, a interação com humanos seja um requisito desejável. Dessa forma, os agentes possuem a autonomia para executar seus objetivos após a sua criação, mas também devem ser capazes de receber

instruções de humanos, quando isso é requisitado (Santos Filho; Mattos; Medeiros, 2020). Além disso, alguns pesquisadores defendem que a autonomia aumenta com a proatividade.

Zape (2012) classifica a autonomia em cinco tipos distintos:

- **Autonomia Absoluta:** Os agentes têm controle completo sobre as suas percepções e ações.
- **Autonomia Social:** Os agentes reconhecem os outros agentes presentes no ambiente e é sociável, procurando cooperar com as ações em conjunto, desde que as mesmas não entrem em conflito com seus objetivos.
- **Autonomia de Interface:** Em muitos sistemas, é inviável a implementação da autonomia absoluta. Nesses casos, o ideal é construção de agentes que possuem a autonomia possível de ações, respeitando as regras e interface com o exterior.
- **Autonomia de Execução:** Autonomia que permite ao agente executar suas ações no ambiente.
- **Autonomia de Projeto:** Grau implementado pelos projetistas dos agentes em sua construção. Quanto maior a autonomia de projeto, maior será a heterogeneidade de agentes.

De acordo com Bazzan e Klugl (2011), os agentes com uma maior autonomia tendem a manter as suas ações independentemente dos seus utilizadores. Para tanto, os agentes apresentam três requisitos: capacidade de executar ações periódicas, execução espontânea e iniciativa própria. Esses requisitos permitem ao agente realizar ações independentes de evento exterior, e que tendem a beneficiar o seu utilizador (Farias *et al.*, 2019).

B) MOBILIDADE

A mobilidade é a propriedade de um agente se movimentar de um local para outro. Esse atributo é muito utilizado em agentes que atuam em pesquisas de informações na Internet (Campos; Cunha; Bueno, 2020). Todavia, esse atributo pode causar problemas de congestionamento na rede, uma vez que os agentes transitam

entre diferentes máquinas através na mesma rede (Castro *et al.*, 2020). Dependendo do balanceamento entre a dimensão do código do agente e a dimensão dos dados que este troca com o exterior, a mobilidade pode até ser vantajosa em termos de sobrecarga da rede. Além disso, Nedić, Pang e Scutari (2018), destacam a importância desse atributo estar acompanhado de processos de segurança, autorização e validação para garantir que os recursos da máquina estarão sempre protegidos de ameaças externas, como vírus e invasões.

Em agentes físicos, os quais possuem um corpo, a mobilidade refere-se a propriedade do agente se deslocar fisicamente em seu ambiente, movimentando-se de um local para outro. Essa mobilidade física apresenta um vasto conjunto de desafios, tais como: percepção ativa, integração de sensores, determinação de trajetórias, técnicas de localização, navegação e mapeamento do ambiente (Slavkovik, 2019).

C) REATIVIDADE

A reatividade é a propriedade de um agente reagir às mudanças em seu ambiente. Para isso, o agente deve ter a percepção de seu ambiente e atuar sobre o mesmo. No entanto, para Wooldridge (2009) e Russel e Norvig (1995), a reatividade possui duas visões distintas. A primeira refere-se à capacidade dos agentes reagirem às mudanças em seu ambiente o mais rápido possível. Já a segunda, diz respeito à capacidade dos agentes utilizarem de um planejamento reativo e regras de condição e ação para definirem seus comportamentos.

Embora a capacidade de reatividade seja fundamental, um agente inteligente não deve se limitar apenas a essa característica. É crucial reconhecer que, se um agente se restringir unicamente a reações ao ambiente, poderá enfrentar dificuldades em alcançar seus objetivos predefinidos, uma vez que suas ações ficarão inteiramente subjugadas às demandas ambientais (Tavares; Billa; Adamatti, 2020). Portanto, para que o agente possa desenvolver um comportamento verdadeiramente orientado a objetivos, torna-se essencial a inclusão e o equilíbrio de outras propriedades em sua arquitetura.

Além da reatividade, é imprescindível considerar outras características dos agentes inteligentes, tais como a proatividade, a autonomia e a capacidade de

comunicação. A proatividade permite que o agente tome a iniciativa na busca por seus objetivos, antecipando-se às mudanças no ambiente e tomando ações preventivas ou planejadas para alcançá-los. A autonomia confere ao agente a capacidade de tomar decisões independentes, sem a necessidade de intervenção externa, o que contribui para uma maior flexibilidade e adaptabilidade em diferentes cenários. Por fim, a habilidade de comunicação possibilita ao agente interagir com outros agentes e com o ambiente, trocando informações relevantes e coordenando esforços para atingir metas comuns (Wooldridge, 2009).

Portanto, para garantir que um agente inteligente seja eficaz na consecução de seus objetivos, é fundamental considerar não apenas a reatividade, mas também outras propriedades essenciais, como proatividade, autonomia e comunicação. A integração harmoniosa dessas características pode resultar em um comportamento mais dinâmico, adaptável e eficiente por parte do agente em seu ambiente (Russel; Norvig, 1995).

D) PROATIVIDADE

A proatividade é a propriedade de um agente apresentar iniciativa por meio de um comportamento independente. As ações executadas pelo agente são selecionadas de acordo com seus objetivos e não apenas às mudanças que ocorrem no ambiente (Kudenko; Kazakov; Alonso, 2005). Essa propriedade é, de certa forma, mais facilmente implementada quando os agentes estão inseridos em ambientes estáticos, ou seja, em ambientes que não apresentam mudanças durante a execução dos objetivos do agente. Essa maior facilidade ocorre em virtude de que o agente apenas realizará suas funções programadas, sem necessidade de analisar constantemente o ambiente, pois o mesmo não apresenta quaisquer mudanças ou incertezas (Campos; Cunha; Bueno, 2020).

Todavia, em ambientes dinâmicos, os agentes devem reagir às mudanças do ambiente e verificar se os objetivos iniciais ainda são válidos. Isso quer dizer que os agentes necessitam ser reativos para responder às mudanças ocorridas no ambiente. Consequentemente, o desafio é equilibrar as propriedades reativas e proativas (Jezic; Chen-Burger; Kusek, 2020).

E) COMUNICAÇÃO

Quando há vários agentes presentes em um mesmo ambiente, há a necessidade de implantar a capacidade de comunicação entre os mesmos. Os agentes devem apresentar uma linguagem que possibilite ser compreendida pelos outros agentes presentes no ambiente, ou seja, precisam partilhar a semântica do que será veiculado. Todavia, a capacidade de comunicação não se restringe apenas à capacidade de troca de informações entre os agentes. Em sua pesquisa, Bazzan e Klugl (2011) evidenciam que os agentes podem se comunicar com outras entidades, incluindo humanos.

Para que os agentes se comuniquem em ambientes físicos ou externos, faz-se necessário que os mesmos possuam informações e protocolos de comunicação para realizar o transporte das mensagens e atuadores apropriados para envio dessas mensagens, como placas de rede e alto-falantes. A disponibilidade de atuadores, protocolos e linguagem para realizar a comunicação é essencial para o desenvolvimento de outras capacidades mais elaboradas, como a habilidade social e a cooperação (Li; Duan, 2014).

F) HABILIDADE SOCIAL

A habilidade social está relacionada à capacidade do agente de efetuar processos de interação e trocar mensagens de alto nível com outros agentes. Esses processos abrangem a coordenação, a cooperação, a competição e a negociação; e para executá-los é preciso compreender os objetivos dos outros agentes (Ahlbrecht; Dix; Fiekas, 2020). Essa compreensão é vital, uma vez que os outros agentes e/ou humanos podem não ter os mesmos objetivos, ou, até mesmo, possuir finalidades totalmente opostas. Assim, muitas vezes é necessário que o agente examine, negocie e coopere com outros agentes para conseguir atingir seus objetivos.

De acordo com Bai, Ren e Fujita (2017), um agente deve ser capaz de se comunicar utilizando uma linguagem de alto-nível que seja comum com os outros agentes para que possam trocar informações. Todavia, o desafio é definir qual a linguagem ideal, pois há diversos tipos e nenhuma é aceita consensualmente entre os pesquisadores da área como sendo a melhor para a comunicação entre os agentes.

Kudenko, Kazakov e Alonso (2005), ressaltam que além da importância de uma linguagem comum entre os agentes, a habilidade social deve garantir que os agentes compartilhem de uma mesma semântica nas mensagens e utilizar o mesmo vocabulário. Esse conceito é baseado na ontologia que, na Ciência da Computação, é um modelo de dados que representa uma coleção de conceitos de um domínio e os seus respectivos relacionamentos. Dessa forma, todos os agentes interpretariam os conceitos da mesma forma durante a sua comunicação (Sebem; Leal, 2019).

O equilíbrio entre a habilidade social, a proatividade e a reatividade também é importante, principalmente quando vários agentes devem executar um trabalho cooperativo com objetivos em comum (Tavares; Billa; Adamatti, 2020). Nesse caso, cada agente precisa balancear a reação às mudanças no ambiente, a proatividade de decidir as ações individuais para desempenhar suas funções e, também, apresentar a habilidade social para desempenhar as ações coletivas (Bai; Ren; Fujita, 2017).

G) COOPERAÇÃO

A cooperação é a capacidade que os agentes possuem de trabalhar em conjunto a fim de concluírem tarefas de interesse comum. Segundo Zhang (2017), esse atributo promove o desenvolvimento de ambientes multiagentes, onde diferentes agentes combinam suas ações em prol de objetivos em comuns, os quais não seriam atingidos individualmente.

Para que haja a cooperação, o agente deve possuir habilidade social, que permite a interação com outros agentes e, até mesmo com os humanos, por meio de uma linguagem de comunicação em comum (Jezic; Chen-Burger; Kusek, 2020). Adicionalmente, os agentes devem utilizar estratégias apropriadas, que variam de acordo com a ação conjunta que será executada em conjunto com os outros agentes.

Entre as estratégias fundamentais para promover a cooperação entre agentes, Nedić, Pang e Scutari (2018) ressaltam uma série de abordagens que variam desde estratégias fixas até a implementação de organizações hierárquicas, a capacidade de aprendizagem de trabalho em equipe e o estabelecimento de normas e leis sociais. Essas estratégias desempenham um papel crucial na facilitação da interação e colaboração eficaz entre agentes em diversos contextos, desde sistemas multiagentes computacionais até ambientes sociais e econômicos complexos. A compreensão e

aplicação adequada dessas estratégias são essenciais para promover a coordenação e alcançar objetivos comuns em ambientes onde múltiplos agentes interagem dinamicamente.

H) APRENDIZAGEM

A aprendizagem é a capacidade que os agentes possuem de aprender e, se aprimorarem, continuamente ao longo do tempo. É uma característica vital para o bom desempenho dos agentes e, é frequentemente confundida com a inteligência (Bazzan; Klugl, 2011).

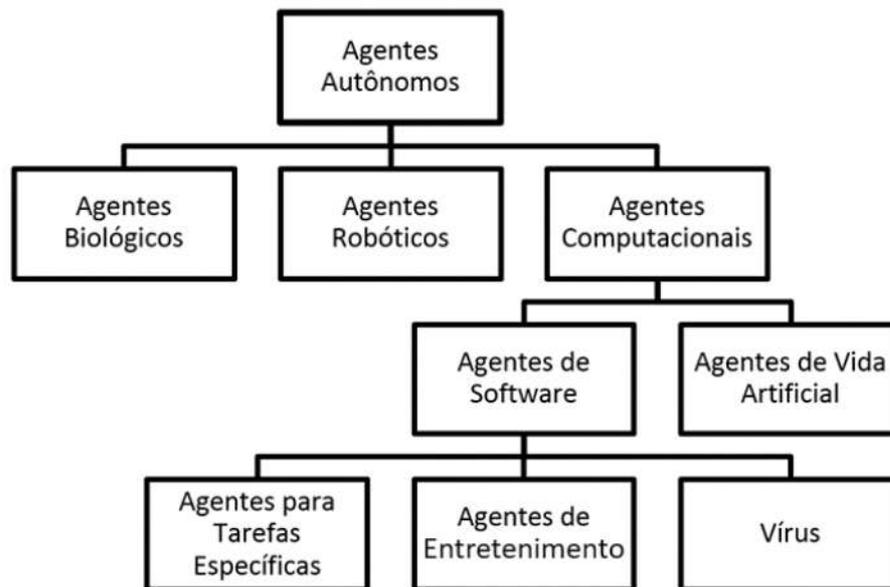
Conceitualmente, um agente atingirá o seu maior grau de eficiência quando possuir a capacidade de avaliar as inúmeras variáveis existentes (situações problemas, ambiente, outros agentes, humanos etc.) e executar as melhores ações para atingir seus objetivos. Todavia essas variáveis se alteram ao longo do tempo, tornando-se diferentes daquilo que eram quando o agente foi projetado (Santos Filho; Mattos; Medeiros, 2020). Adicionalmente, há uma grande dificuldade de se projetar um agente prevendo todas as mudanças que ocorrerão no futuro.

Dessa forma, mesmo quando um agente não apresente previamente a ação que resulte na maior probabilidade de sucesso para atingir seu objetivo, espera-se que o mesmo procure, encontre e experimente diferentes alternativas. Afinal, o intuito é aprender continuamente por meio das experiências de sucessos ou fracassos. Contudo, para Jezic, Chen-Burger e Kusek (2020), o desenvolvimento do processo de aprendizagem em agentes é uma tarefa complexa, pois é difícil identificar os motivos que ocasionam os sucessos e os fracassos nas ações dos agentes. Além disso, mesmo quando se identifica corretamente a causas dos acertos ou falhas, é complicado utilizar essa informação para a elaboração de planos alternativos com maior probabilidade de sucesso.

De acordo com Li e Duan (2014), a aprendizagem também pode ser um processo interativo em que os desenvolvedores podem treinar os agentes para executar atividades individuais e/ou cooperativas. Nesses casos, os desenvolvedores podem fornecer as instruções corretas aos agentes por meio do conjunto de algoritmos previamente instalados nos agentes ou informar a ação correta somente quando os agentes se depararem com situações que não consigam solucionar.

A partir da definição dos atributos fundamentais e de acordo com o ambiente em que se encontram, Russel e Norvig (1995), apresentam a taxonomia para agente, conforme exibido na Figura 4.

Figura 4 - Categorias de agentes



Fonte: Russel; Norvig (1995)

2.2. AGENTES VERSUS OBJETOS

Em consonância com Wooldridge (2009), um dos desafios da área de pesquisas em Agentes Inteligentes consiste em estabelecer a diferença entre Agentes e Objetos. O conceito de Objeto, de acordo com Russel e Norvig (1995), consiste em uma entidade computacional que encapsula um estado (informação) e apresenta a possibilidade de executar métodos e se comunicar por meio de mensagens.

As principais diferenças entre os conceitos de Agentes e Objetos, segundo Wooldridge (2009), são:

- Agentes apresentam capacidade de decisão autônoma.
- Agentes podem ter comportamento reativo, proativo e habilidade social.
- Agentes possuem fluxo de controle e objetivos próprios e características ímpares que os diferenciam de outros agentes presentes no mesmo ambiente

e que compõe um determinado sistema multiagente.

Os objetos não apresentam autonomia sobre o seu comportamento, ou seja, sempre que houver um método disponível, qualquer outro objeto pode solicitar e o objeto inicial obrigatoriamente o executará. Por sua vez, os agentes apresentam controle sobre o seu comportamento e, assim, caso outro agente deseje a execução de uma determinada ação, deverá solicitar ao agente em questão, que analisará a situação para executar ou não a ação solicitada. Com relação à proatividade, reatividade e habilidade social, o modelo orientado a objetos não apresenta qualquer referência a esses comportamentos. Quanto a questão do fluxo de controle e objetivos próprios e características ímpares, os objetos baseiam-se no fluxo de controle único, com objetos com objetivos e características padronizadas e comuns (Bortoletto; Moitinho; Waldman, 2020).

De acordo com Bazzan e Klugl (2011), embora existam distinções conceituais entre agentes e objetos, o modelo orientado a objetos mostra-se apropriado para a implementação de agentes inteligentes. Portanto, linguagens de programação como Java, C++ e suas extensões correspondentes são amplamente empregadas nesse contexto. Alinhando-se com as observações de Zhang (2017), ao estender essa perspectiva, é possível considerar os agentes inteligentes como um paradigma adicional, denominado Programação Orientada a Agentes.

A Engenharia de Software, ao longo da história, observa que novos paradigmas surgem no sentido de aperfeiçoar o encapsulamento das unidades de programação e o paradigma da Programação Orientada a Agentes confirma essa tendência (Ahlbrecht; Dix; Fiekas, 2020). A Tabela 8 mostra a evolução de diversos paradigmas de programação e suas principais características.

Tabela 8 - Evolução dos principais paradigmas de programação

	Programação Estruturada	Programação Orientada a Objetos	Programação Orientada a Agentes
Comportamento das Unidades	Local	Local	Local
Estado das Unidades	Externa	Local	Local
Invocação das Unidades	Externa (chamada)	Externa (mensagem)	Local (regras e objetivos)

Fonte: Ahlbrecht; Dix; Fiekas (2020)

A tendência de encapsulamento é cada vez maior, conforme observada na Tabela 8, que mostra a evolução dos principais paradigmas de programação. As unidades de programação estão se tornando cada vez mais autônomas e a Programação Orientada a Agentes apresenta maior autonomia às unidades de programação, denominadas agentes, proporcionando maior controle sobre o seu comportamento (Ahlbrecht; Dix; Fiekas, 2020).

2.3. AMBIENTES

Segundo Zappe (2012), ambiente é o meio onde os agentes inteligentes operam e, muitas vezes, o transformam. As características do ambiente desempenham um papel fundamental ao definir a arquitetura e o modo de atuação dos agentes. Ao desenvolver um projeto voltado para Programação Orientada a Agentes, é imprescindível realizar uma análise criteriosa das ações disponíveis e da complexidade dos objetivos a serem executados pelo agente (Coelho, 2018). Nesse cenário, a distinção entre ambientes não implica que um ambiente real seja mais complexo do que um ambiente simulado. Às vezes, agentes de *softwares* desenvolvidos para ambientes simulados apresentam um conjunto de ações mais complexo do que agentes robóticos construídos para atuar em ambientes reais.

De acordo com os estudos de Russel e Norvig (1995), é possível classificar as principais propriedades dos ambientes da seguinte maneira:

- Completamente observável versus Parcialmente observável: um ambiente é classificado como completamente observável como sendo aquele onde um agente obtém, através dos seus sensores, informações atualizadas, precisas e completas sobre o ambiente.
- Agente único versus Multiagente: em um ambiente multiagente há a presença de diversos agentes, diferentemente de um ambiente com agente único.
- Cooperativo versus Competitivo: um ambiente multiagente pode apresentar agentes que possuem objetivos em comum ou não e, com isso, a forma de atuação entre esses agentes será influenciada.
- Determinístico versus Estocástico: em um ambiente determinístico, cada ação apresenta um efeito único, não existindo qualquer incerteza em relação ao resultado de sua execução.

- Episódico versus Sequencial: em um ambiente episódico, o agente apenas realiza as suas ações sem depender de ações anteriores, diferentemente dos ambientes sequenciais.
- Estático versus Dinâmico: no estático, o ambiente permanece inalterado enquanto o agente executa suas ações. Em contrapartida, em um ambiente dinâmico outros agentes agem e tendem a alterar o ambiente simultaneamente.
- Discreto versus Contínuo: um ambiente é considerado discreto se existe um número limitado de ações e percepções possíveis para o agente. Nesse contexto, um ambiente pode ser contínuo quanto às percepções do agente e discreto com relação às ações do mesmo.

Quanto mais um ambiente for completamente observável, menos complexo será projetar e construir agentes para atuarem no mesmo. A qualidade das decisões tomadas por um agente está relacionada à qualidade e a quantidade de informações que o mesmo detém. Ambientes completamente observáveis possuem todos os aspectos relevantes ao ambiente acessíveis, proporcionando assim, tomadas de decisões mais assertivas pelo agente.

Li e Duan (2014), ressaltam também que a acessibilidade não pode ser considerada uma escala discreta como apenas completamente observável ou parcialmente observável. Em ambientes parcialmente observáveis deve-se considerar o grau de acessibilidade do ambiente em relação ao agente. Por exemplo, um agente robótico atuando em um ambiente físico, tende a tomar decisões mais assertivas se possuir câmaras e outros equipamentos sensoriais (sensores de aproximação, infravermelho etc.) a fim de garantir uma boa acessibilidade ao ambiente físico (Russel; Norvig, 1995). Nesse contexto, deve-se também considerar a precisão dos sensores do agente. Esses sensores são responsáveis pela coleta das informações que permitem a atuação do agente, porém se não possuírem qualidade para tal, os erros na percepção afetarão a execução das tarefas do agente.

Em ambientes com um único agente, esse agente é capaz de resolver todas as situações problemas e, assim, atingir os objetivos sozinho. Por outro lado, há ambientes que apresentam outros agentes. Esses ambientes são denominados multiagentes e apresentam, em linhas gerais, propriedade de comunicação entre as agentes e agentes com funções e objetivos próprios (Nedić; Pang; Scutari, 2018).

Os ambientes que apresentam múltiplos agentes, que são denominados

multiagentes, podem ser classificados como ambientes competitivos ou cooperativos. Em um ambiente competitivo, um agente visa maximizar suas funções e desempenho enquanto tenta minimizar as ações e o desempenho dos agentes concorrentes (Kudenko; Kazakov; Alonso, 2005). Por sua vez, em ambientes cooperativos, um agente visa maximizar seu desempenho juntamente com o desempenho dos agentes, pois possuem objetivos em comum. Adicionalmente essa cooperação é benéfica para todos os agentes desse ambiente, uma vez que todos os agentes obtêm ganhos adicionais ao trabalharem juntos (Bazzan; Klugl, 2011).

A ausência de determinismo de um ambiente, torna-o um ambiente estocástico, ou seja, há diversas incertezas, aleatoriedades e imprevisibilidade sobre o estado futuro do ambiente. Esse tipo de ambiente apresenta grande complexidade para os agentes, impondo sérias restrições à sua capacidade de efetuar previsões relacionadas ao estado futuro. Essa característica pode resultar em ações falhas ou imprevisíveis, e, assim, a construção desses agentes é uma tarefa complexa, uma vez que é necessário incluir propriedades de recuperação e tolerância a falhas no projeto desses agentes (Bai; Ren; Fujita, 2017).

Nos ambientes episódicos, a experiência do agente pode ser dividida em episódios independentes. O conceito de episódio implica que cada percepção resulta em uma única ação, sem dependência em relação aos episódios anteriores. Além disso, a ação em cada episódio depende exclusivamente desse episódio em si. Em outras palavras, a escolha da ação em cada episódio é influenciada apenas por esse episódio em particular. Por outro lado, nos ambientes sequenciais, há várias percepções que podem compor uma única ação, e os episódios subsequentes dependem das ações realizadas no episódio anterior (Zhang, 2017). Portanto, as ações se desenvolvem em uma sequência dependente de episódios.

Em ambientes dinâmicos, o ambiente muda enquanto o agente delibera e o mesmo deve avaliar essas mudanças constantemente. De acordo com Bazzan e Klugl (2011), deve-se considerar ainda, que o efeito de uma ação de um agente em um instante poderá ser outra em um instante diferente. Adicionalmente, uma ação realizada em um instante, que possui um tempo de análise, decisão e execução por parte do agente, pode ter um resultado distinto do esperado no instante inicial.

O dinamismo dos ambientes também implica que os agentes devem obter constantemente as informações do ambiente e, para isso, os mesmos devem possuir sensores para efetuar essa coleta (Zappe, 2012). Outra característica dos ambientes

dinâmicos, é que a presença de outros agentes atuantes no ambiente pode interferir nas ações efetuadas pelo agente. Nesse caso, o agente não está isolado e deve comunicar, coordenar, competir, cooperar ou negociar com esses outros agentes para atingir seus objetivos.

Um ambiente estático é menos complexo que os ambientes dinâmicos, pois o agente ao recolher informações sobre o ambiente, assume que o mesmo não será modificado. Para Jezic, Chen-Burger e Kusek, (2020), nos ambientes estáticos os agentes também não necessitam coordenar suas ações ou interagir com outros agentes. Além disso, há os chamados ambientes semidinâmicos, que se apresentam quando o ambiente não se altera enquanto o agente está atuando e o desempenho do agente pode mudar enquanto ele está atuando. Segundo Wooldridge (2009), os ambientes semidinâmicos, podem ser considerados como ambientes estáticos, com o desempenho piorando com passar do tempo.

Um ambiente discreto é caracterizado por ter um conjunto definido de estados, enquanto um ambiente contínuo apresenta um número infinito de estados possíveis. Os ambientes discretos são menos complexos do que os ambientes contínuos, pois os sistemas computacionais são, por natureza, discretos. Portanto, o desenvolvimento de ambientes contínuos requer a realização de aproximações, adaptações e discretizações dos elementos com um número infinito de estados (Tavares; Billa; Adamatti, 2020). Em um ambiente contínuo, os agentes sempre terão informações aproximadas ao tomar decisões. Por exemplo, o jogo de xadrez é um ambiente discreto, pois possui um número finito de jogadas possíveis. Por outro lado, os ambientes físicos do mundo real são considerados contínuos (Ahlbrecht; Dix; Fiekas, 2020). A continuidade de um ambiente pode ser compreendida em diferentes níveis:

- Estado do Mundo: o estado do mundo pode ser contínuo ou discreto, ou seja, pode existir um número finito ou infinito de estados para esse ambiente.
- Percepções do Agente: as percepções do agente podem ser diferentes do ambiente em si, ou seja, mesmo que um agente esteja inserido em um ambiente contínuo, as suas percepções podem ser discretas.
- Ações do Agente: assim como ocorre nas percepções do agente, um agente que se encontra inserido em um determinado ambiente contínuo pode ter sua capacidade de ação discretas, ou seja, esse agente dispõe de um conjunto limitado de ações.

Em consonância com Zhang (2017), a partir da análise das propriedades dos ambientes, observa-se que o desenvolvimento de agentes se torna mais complexo quando o ambiente apresenta as seguintes características: parcialmente observável, multiagente, competitivo, estocástico, sequencial, dinâmico e contínuo. Ambientes com essas propriedades são denominados ambientes abertos e, ainda que esses ambientes sejam os mais complexos, se verifica que os ambientes físicos do mundo real são, quase em sua totalidade, abertos. Dessa forma, em ambientes físicos do mundo real, o projeto dos agentes deve contemplar formas que possibilite a operação e o desenvolvimento dos mesmos em ambientes abertos (Slavkovik, 2019).

2.4. ARQUITETURA DOS AGENTES

Uma arquitetura de *software* pode ser descrita como uma configuração dos componentes que compõem um sistema e as conexões que coordenam as atividades entre esses componentes (Li; Duan, 2014). Quando se trata de arquitetura de agentes, pode-se tratar da arquitetura interna de cada agente e da arquitetura do sistema multiagente como um todo, ou seja, a maneira com que os agentes estão organizados em um sistema e como são estruturadas suas interações.

Da mesma forma que existem diversas arquiteturas de *software*, o mesmo ocorre com relação às arquiteturas de agentes, que possuem determinadas características que permitem avaliar a sua qualidade e eficácia. Por exemplo, os agentes robóticos possuem uma arquitetura de *software* e uma arquitetura de *hardware*, que é composta pelos componentes físicos e suas interligações.

De acordo com Bazzan e Klugl (2011), a arquitetura de um agente pode ser desenvolvida por meio de um procedimento que define os agentes por técnicas e algoritmos. Para Kudenko, Kazakov e Alonso (2005), a qualidade de uma arquitetura de agentes é subjetiva, pois, a mesma está relacionada com as especificações das aplicações de agentes que se pretende desenvolver. Sendo que uma arquitetura de agentes bem desenvolvida para um determinado domínio poderá ser inadequada para outros domínios. Wooldridge (2009) propõe conceitos fundamentais para a criação de arquiteturas de agentes de alta qualidade:

- Simplicidade: projetar a arquitetura e seus componentes para que sejam fáceis de compreender, implementar e manter.

- Funcionalidade: optar por ferramentas de desenvolvimento que foquem em aspectos específicos do problema a ser abordado.
- Expansibilidade: a arquitetura deve possibilitar que futuras ampliações ocorram facilmente.
- Portabilidade: capacidade de operar em diferentes arquiteturas.

A seleção da arquitetura de cada projeto depende da complexidade das atividades que o agente irá executar e do ambiente em que ele irá operar. Além disso, deve-se analisar como será a coexistência (competitiva, cooperativa etc.) entre os agentes. Baseado nos estudos de Wooldridge (2009), as arquiteturas para a construção dos agentes podem ser divididas em: arquiteturas deliberativas, arquiteturas reativas, arquiteturas híbridas, arquitetura em camadas, arquitetura BDI e arquitetura de agente social.

A) *ARQUITETURAS DELIBERATIVAS*

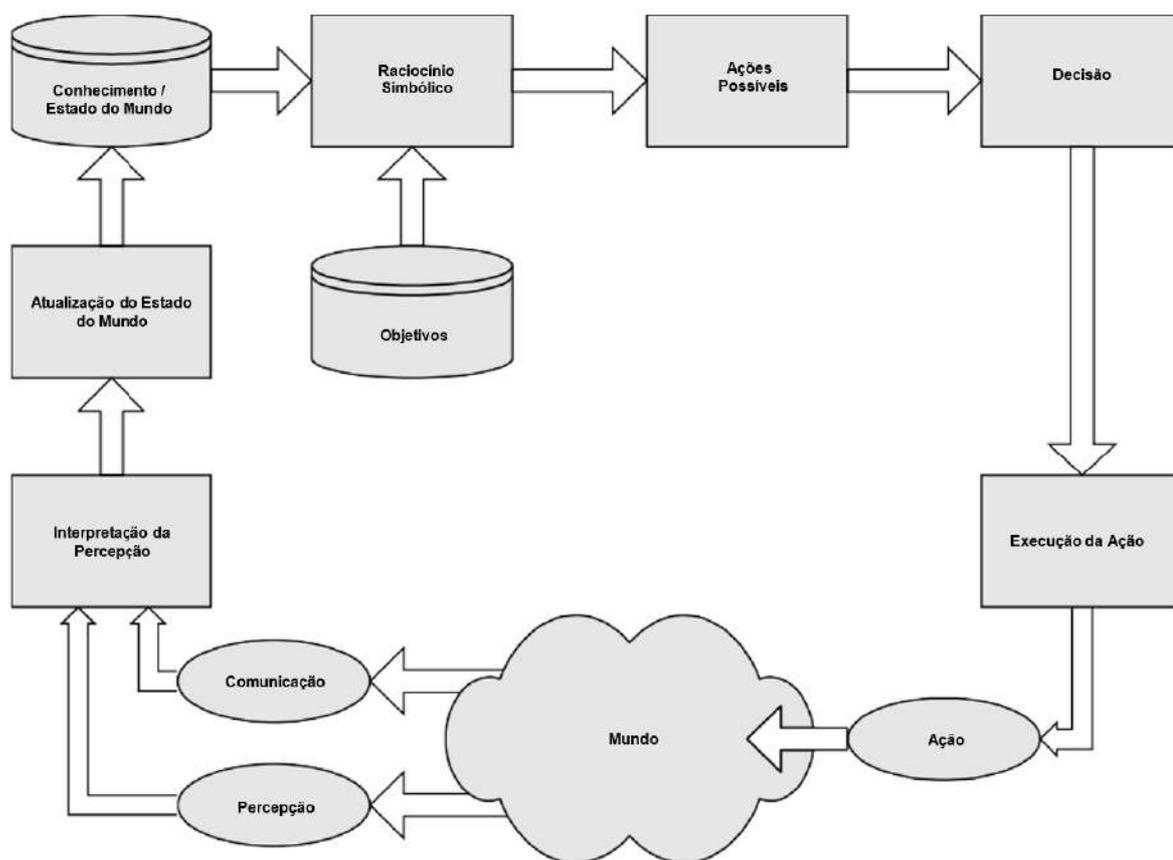
As arquiteturas deliberativas são reconhecidas como uma abordagem clássica no domínio da Inteligência Artificial, caracterizadas pela concepção de agentes que operam com níveis de autonomia relativamente baixos em relação aos seus ambientes. Dentro dessas estruturas, o funcionamento do sistema é orientado pelo emprego de conhecimento prévio, o qual direciona as decisões dos agentes, incluindo a seleção da próxima ação a ser executada, por meio de um processo de raciocínio simbólico (Zhang, 2017).

A característica central das arquiteturas deliberativas reside em sua dependência em relação a modelos simbólicos que representam o conhecimento e as regras de tomada de decisão dos agentes. Conforme destacado por Nedić, Pang e Scutari (2018), esses modelos são construídos com base em lógica formal ou em estruturas de representação simbólica, capacitando os agentes a realizar inferências sobre o ambiente e a selecionar ações apropriadas de acordo com objetivos definidos. A ênfase na utilização do raciocínio simbólico evidencia a natureza deliberada e planejada das interações dos agentes com o ambiente, favorecendo uma abordagem mais controlada e preditiva em relação ao comportamento dos sistemas inteligentes.

Dentro dessa arquitetura, o agente é dotado de uma representação interna do

mundo e de um estado explícito que pode ser modificado. Após a interpretação das percepções provenientes do ambiente, o agente utiliza essas informações para manter atualizada a sua representação interna do estado do mundo. Esse estado do mundo, combinado com os objetivos do agente, é empregado para gerar as ações e selecionar as melhores ações a serem executadas pelo agente (Wooldridge, 2009). Um esquema de uma arquitetura deliberativa é ilustrado na Figura 5.

Figura 5 - Esquema genérico de uma arquitetura deliberativa



Fonte: Wooldridge (2009)

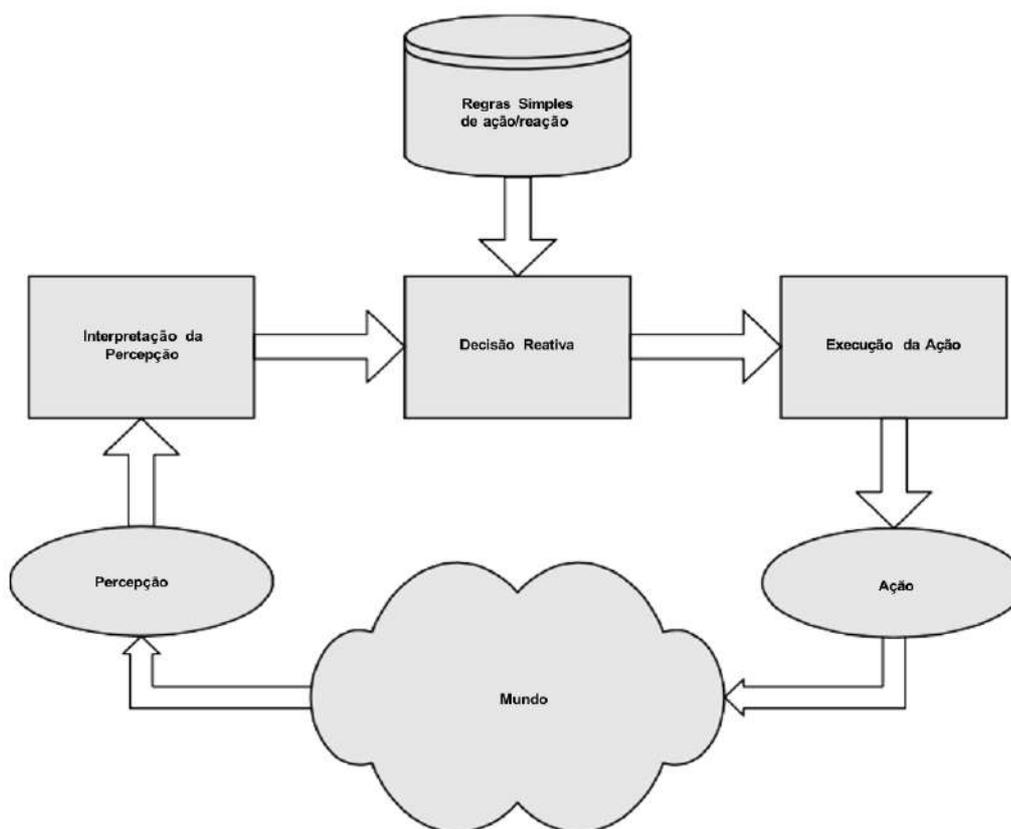
Para Nedić, Pang e Scutari (2018), o desenvolvimento de agentes baseados na arquitetura deliberativa apresenta dois desafios:

1. Como traduzir em detalhes o mundo real para um código e desenvolver a percepção para manter essa estrutura de mundo atualizada.
2. Como programar o raciocínio simbólico utilizando essa informação do mundo para o agente decidir as ações a serem executadas em cada momento.

B) ARQUITETURAS REATIVAS

As arquiteturas reativas se caracterizam por não utilizar um modelo de raciocínio simbólico. Essa arquitetura propõe que um agente pode desenvolver conhecimento a partir de interações com o seu ambiente sem a necessidade de um modelo pré-estabelecido (Li; Duan, 2014). O agente reativo decide baseado em um conjunto de informações limitado e regras simples de ação/reação. A informação vinda dos sensores é utilizada no processo de decisão, ou seja, não é criada qualquer representação simbólica do ambiente. A Figura 6 ilustra uma arquitetura reativa.

Figura 6 - Esquema genérico de uma arquitetura reativa



Fonte: Li; Duan (2014)

Uma das características dessa arquitetura é que a tomada de decisão do agente é realizada por meio de um conjunto de comportamentos direcionados para a execução de atividades, ou seja, cada comportamento é uma função individual de ação, que vai gradativamente tendo a percepção do ambiente e traduz isso em uma ação a ser executada (Ahlbrecht; Dix; Fiekas, 2020).

Além disso, outra característica é que vários comportamentos podem ser acionados simultaneamente. Dessa forma, deve existir um mecanismo para selecionar qual será a ação ótima a ser executada naquele momento. Assim, esse mecanismo se apresenta como uma estrutura hierárquica, que organiza os comportamentos por níveis, onde os níveis mais baixos são prioritários, uma vez que as ações de níveis mais altos representam comportamentos mais abstratos (Bai; Ren; Fujita, 2017).

Segundo Wooldridge (2009), a arquitetura reativa apresenta inúmeras vantagens, tais como a simplicidade, a otimização de recursos computacionais e a robustez contra falhas. Todavia, há também desvantagens a serem consideradas como a decisão dos agentes ser unicamente baseada em informações da percepção atual, a existência de uma hierarquia de comportamentos pré-fixada e os agentes serem incapazes de realizar ações com planos de longo-prazo (Zappe, 2012).

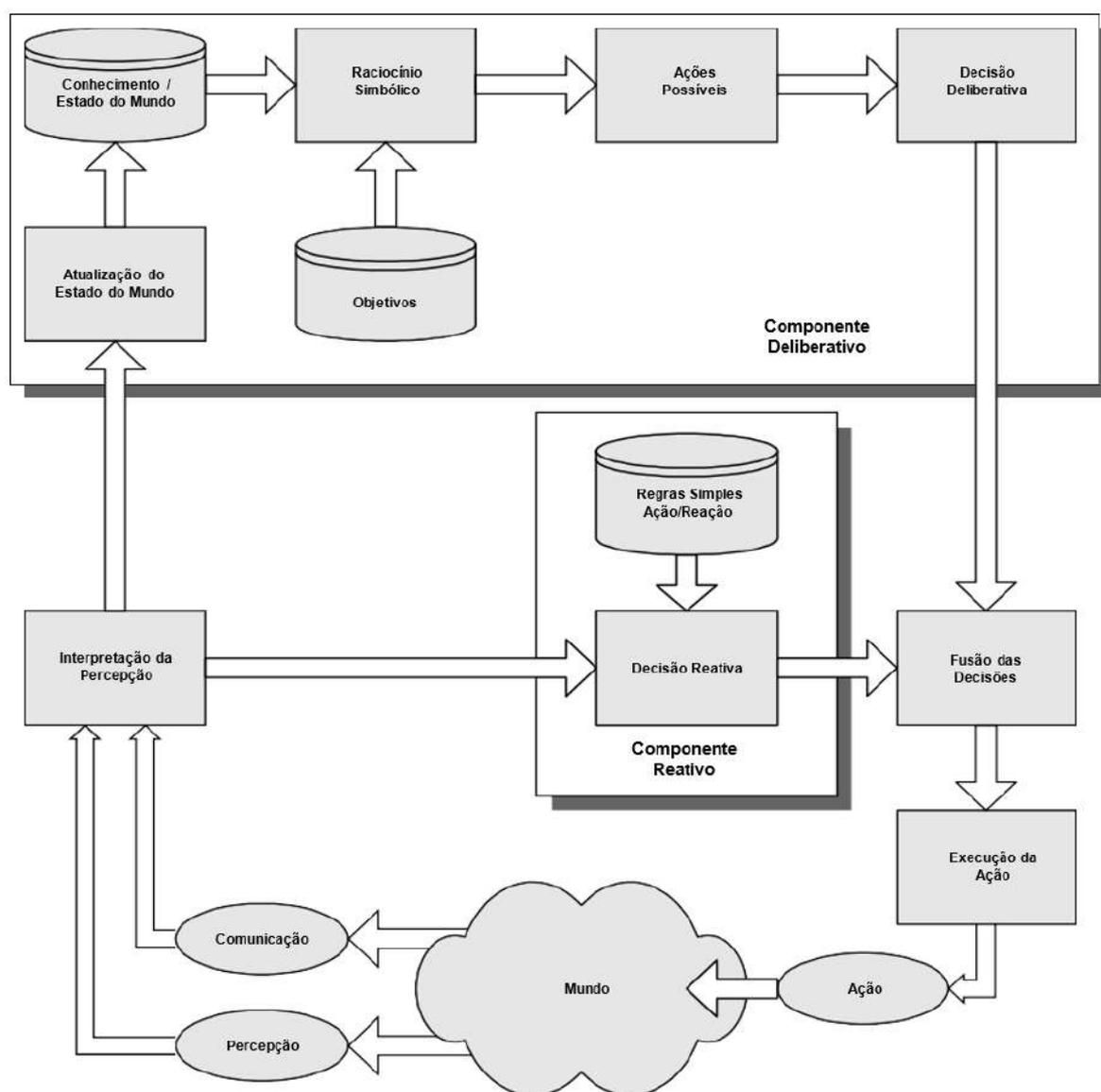
C) ARQUITETURAS HÍBRIDAS

As arquiteturas híbridas combinam características das arquiteturas deliberativas e reativas. Os agentes reativos apresentam limitações para implementar comportamento orientado por objetivos. Por sua vez, os agentes deliberativos, que são baseados em mecanismos de raciocínio complexo simbólico, apresentam limitações com relação a reações imediatas a estímulos do ambiente (Bazzan; Klugl, 2011).

Os agentes híbridos apresentam uma arquitetura composta por níveis, onde se torna possível estruturar as funcionalidades e controles dos agentes. A função é categorizar as funcionalidades do agente em camadas hierárquicas onde a camada reativa tem uma certa prioridade sobre a camada deliberativa, a fim de permitir uma resposta mais rápida aos eventos importantes detectados no ambiente (Farias *et al.*, 2019). Dessa forma, após a interpretação da percepção e da comunicação vindas do mundo, o agente as interpreta por meio do componente reativo e do componente deliberativo. O funcionamento do componente reativo baseia-se em um conjunto de regras ação/reação que associa determinados estímulos a determinadas decisões. O componente deliberativo implica na construção de um estado simbólico do mundo e na utilização de um raciocínio complexo simbólico a fim de decidir as ações a serem

executadas a cada momento para atingir os objetivos. Um mecanismo analisa as decisões deliberativas e reativas a fim de selecionar a ação final a ser executada. A Figura 7 ilustra o esquema de uma arquitetura híbrida (Wooldridge, 2009).

Figura 7 - Esquema genérico de uma arquitetura híbrida



Fonte: Wooldridge (2009)

Diversos pesquisadores utilizam arquiteturas híbridas para agentes executarem diferentes atividades. Kudenko, Kazakov e Alonso (2005), por exemplo, propõem a utilização dessa arquitetura para controlar robôs em ambientes imprevisíveis. Nesse caso, o componente deliberativo tem a função de guiar o comportamento do componente reativo, por meio de sugestões de planos a serem

realizados.

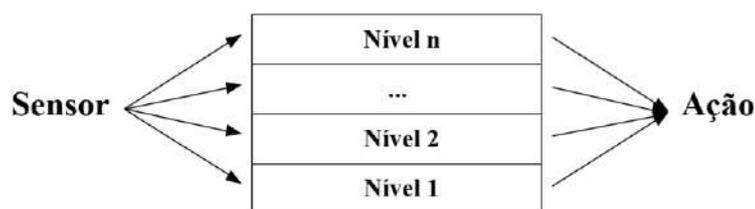
Uma das inovações desse trabalho, é o fato de existir uma interação entre os dois componentes, podendo o componente deliberativo alterar as regras de ação/reação que o componente reativo utiliza. Assim, o componente reativo, em certas emergências, se sobrepõe ao componente deliberativo, assumindo o controle dos robôs (Zhang, 2017).

D) ARQUITETURA EM CAMADAS

A arquitetura em camadas apresenta inúmeros subsistemas, que se organizam em hierarquias e interagem por níveis (Wooldridge, 2009). Há dois tipos de camadas de acordo com o fluxo de controle:

- Camadas Horizontais: as camadas de *software* estão diretamente ligadas aos sensores de entrada e às ações de saída, ou seja, cada camada atua como um agente, uma vez que apresenta sugestões quanto à ação a ser efetuada pelo agente. A Figura 8 ilustra uma representação da arquitetura em camadas horizontais.

Figura 8 - Representação da arquitetura em camadas horizontais

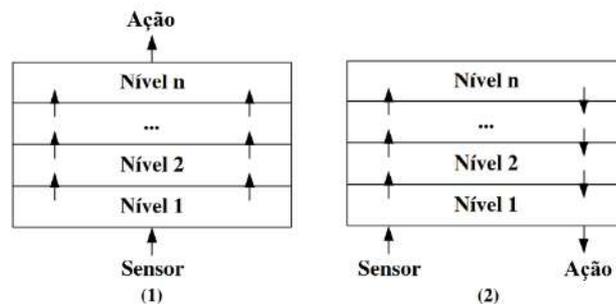


Fonte: Wooldridge (2009)

- Camadas Verticais: os sensores de entrada e as ações de saída têm, no mínimo, uma camada entre eles. A arquitetura em camadas verticais pode incluir uma única passagem de controle ou duas passagens de controle. O fluxo de controle na arquitetura com uma única passagem passa sequencialmente pelas diferentes camadas, até ao último nível, onde é gerada a ação a ser executada. Por outro lado, na arquitetura com duas passagens de controle, a informação sobe as camadas até atingir o nível superior (primeira passagem) e, em sentido inverso o fluxo de controle desce e atravessa as camadas, dando

origem à ação a ser executada (segunda passagem). A Figura 9 ilustra uma representação da arquitetura em camadas verticais.

Figura 9 - Representação da arquitetura em camadas verticais



Fonte: Wooldridge (2009)

A grande vantagem da arquitetura horizontal em relação à arquitetura vertical é a simplicidade, uma vez que se for pedido a um agente n comportamentos diferentes, será necessário implementar n camadas diferentes. No entanto, uma vez que as camadas competem umas com as outras, será necessário a inserção de um mecanismo mediador para resolver o problema, porém isso pode incorrer em uma lentidão no processo de decisão do agente (Wooldridge, 2009).

Essa situação não acontece em uma arquitetura vertical, pois o fluxo de controle passa sequencialmente camada por camada, até que ao chegar à última irá gerar a ação a ser executada. Por sua vez, essa arquitetura apresenta baixa flexibilidade, uma vez que para o agente tomar uma decisão é preciso que o fluxo de controle atravesse cada camada, ou seja, caso exista uma falha em alguma camada, o agente terá problemas para gerar e executar a ação (Russel; Norvig, 1995).

E) ARQUITETURA BDI

De acordo com Russel e Norvig (1995), a arquitetura BDI, do inglês *Belief-Desire-Intention*, é essencialmente deliberativa, onde o estado de processamento dos agentes é descrito por meio de um conjunto denominado de estados mentais.

Em linhas gerais, os estados mentais apresentam noções de crença (*belief*), desejo (*desire*) e intenção (*intention*). A crença se refere ao que o agente acredita a cada momento, bem como a descrição do estado do mundo desse agente. Em outras

palavras, as crenças representam informação. O desejo se refere ao que o agente deseja obter. Por sua vez, a forma de alcançar os desejos pode não ser conhecida a todo instante, assim, os desejos podem ser inconsistentes em determinados momentos (Sebem; Leal, 2019).

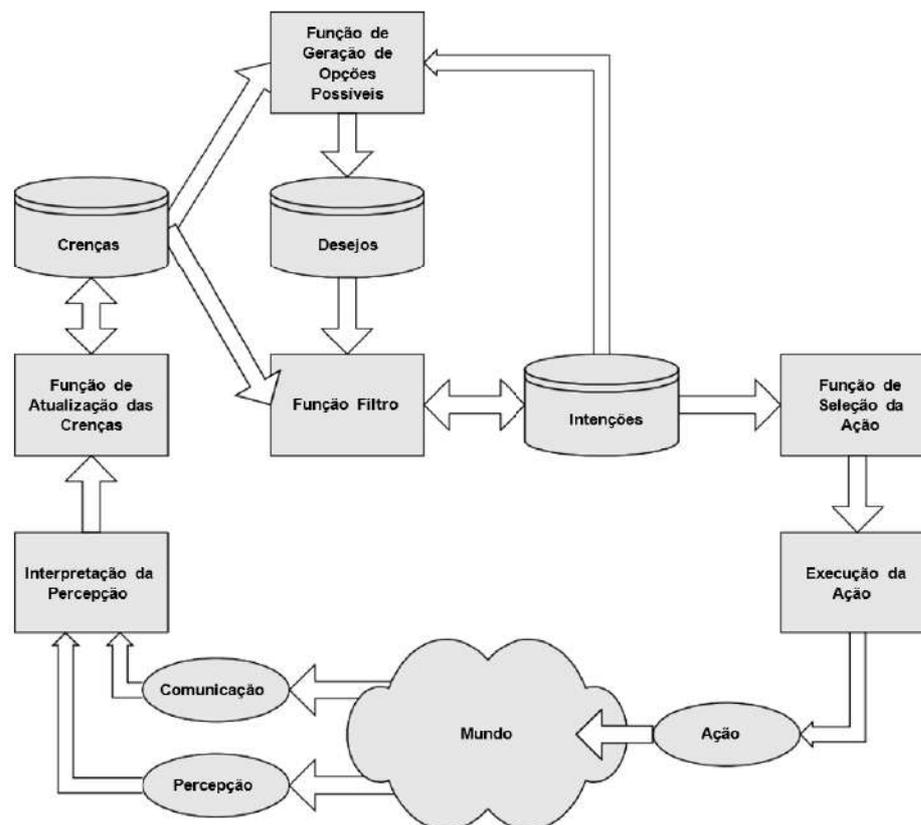
Os objetivos dos agentes resultam de processos decisórios, onde há a escolha de subconjuntos de desejos que são viáveis. A intenção se refere ao conjunto de que o agente selecionou para a realização dos objetivos.

As intenções dos agentes inteligentes devem ser consistentes e resultam de um processo de deliberação. De acordo com Bai, Ren e Fujita (2017), o comportamento dos agentes na arquitetura BDI é fundamentado na interação dinâmica entre diferentes estados mentais. Esses estados mentais desempenham papéis distintos na determinação das ações dos agentes. A relação entre esses estados mentais é crucial para compreender como os agentes tomam decisões e se comportam em um ambiente dinâmico. Essa classificação permite uma análise mais aprofundada da complexidade do comportamento dos agentes em sistemas baseados na arquitetura BDI.

- **Realismo Forte:** os agentes são modelados com uma representação mais abrangente e complexa do ambiente, envolvendo crenças, desejos e intenções mais elaboradas. Os agentes neste contexto possuem conhecimento não apenas sobre si mesmos, seus objetivos individuais, mas também sobre as capacidades dos outros agentes no ambiente. Esta abordagem resulta em uma representação mais precisa e detalhada das interações entre os agentes e do ambiente em si.
- **Realismo:** os agentes são modelados com um equilíbrio entre crenças, desejos e intenções sobre o ambiente e seus próprios objetivos. Aqui, os agentes tendem a agir de forma mais reativa, respondendo diretamente a estímulos sem considerar uma compreensão mais profunda do ambiente ou objetivos a longo prazo. Esta abordagem, embora menos complexa que o Realismo Forte, ainda permite uma interação significativa do agente com seu ambiente.
- **Realismo Fraco:** os agentes são representados com um conjunto limitado de crenças, desejos e intenções sobre o ambiente. Suas capacidades são restritas à manipulação dessas crenças, desejos e intenções, e sua ação é principalmente reativa, respondendo apenas a eventos perceptíveis e ajustando seu comportamento de acordo com mudanças nessas crenças,

desejos e intenções. Esta abordagem é adequada para sistemas onde os agentes possuem uma compreensão simplificada do ambiente e operam dentro de limitações perceptivas e de processamento de informações.

Figura 10 - Esquema genérico de uma arquitetura BDI



Fonte: Bai; Ren; Fujita (2017)

Analisando a Figura 10, que mostra a arquitetura genérica de um agente BDI, observa-se diversos componentes que a compõe. De acordo com Bai, Ren e Fujita (2017), essa arquitetura é composta por:

- Conjunto de Crenças: representa as informações que o agente possui sobre o ambiente a cada momento.
- Função de Revisão de Crenças: responsável por atualizar o conjunto de crenças a partir de novas percepções recebidas e levando em conta, também, as crenças atuais do agente.
- Função de Geração de Opções: responsável por determinar os desejos disponíveis baseados nas crenças e intenções atuais para o agente. É um processo que permite ao agente selecionar quais as melhores ações que

satisfaçam suas intenções atuais. Essa função deve gerar desejos consistentes com as crenças e intenções atuais visando que o agente reconheça as alterações mais propícias do ambiente para atingir as intenções que eram inatingíveis.

- Desejos: representam as opções possíveis e disponíveis para o agente.
- Função de Filtro: representa o processo de deliberação, que determina as intenções do agente baseados em suas crenças e desejos atuais aliadas às suas intenções prévias. É um processo em que o agente decide o que fazer de acordo com as suas intenções. Com isso, pode haver a desistência de intenções com um custo elevado e, também, a adoção de novas intenções que resultam em novas opções.
- Intenções: representa o conjunto de intenções.
- Função de Seleção de Ação: determina qual a ação a ser realizada em cada instante baseada nas intenções atuais do agente.

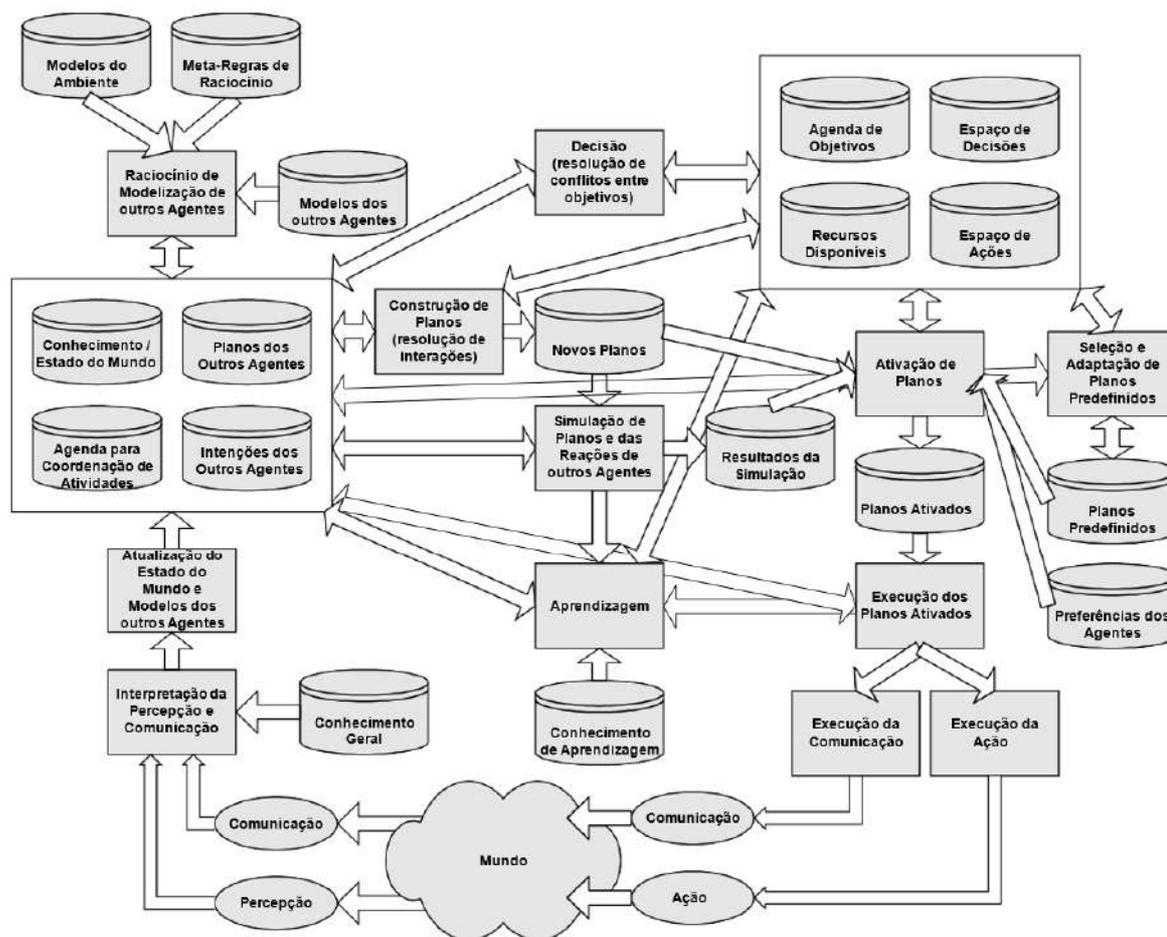
F) ARQUITETURA DE AGENTE SOCIAL

Uma arquitetura abrangente e sofisticada é necessária para o agente, pois ele deve ter internamente os modelos de outros agentes e ser capaz de mantê-los atualizados com base em comunicações e percepções. De acordo com Ahlbrecht, Dix e Fiekas (2020), o agente precisa raciocinar utilizando o conhecimento contido nesses módulos, como intenções, expectativas, previsões e reações antecipadas, a fim de tomar decisões e criar planos considerando os outros agentes. À medida que as atividades executadas se tornam mais complexas, o nível de complexidade das estruturas de conhecimento do agente aumenta.

Na Figura 11, é apresentada uma arquitetura genérica de um agente social, na qual fica evidente que o agente interpreta a percepção e a comunicação do ambiente com base em seu conhecimento. A partir dessas informações, o agente mantém atualizada uma estrutura que contém: a) o estado do mundo nesse momento, b) a agenda para a coordenação de atividades com outros agentes, e, c) as intenções e planos dos outros agentes. As informações sobre outros agentes são analisadas por um componente de raciocínio e modelização. Esse componente utiliza modelos pré-definidos de agentes e do ambiente e de regras de raciocínio. O módulo de

modelização dos agentes pode ser ativado para inferir as intenções e planos dos outros agentes (Castro *et al.*, 2020). O módulo de decisão é responsável por selecionar quais os objetivos devem ser atingidos a partir das crenças atuais do agente e das intenções e os objetivos selecionados são colocados em uma agenda.

Figura 11 - Esquema genérico de uma arquitetura de um agente social



Fonte: Ahlbrecht; Dix; Fiekas (2020)

O agente é equipado com uma diversidade de planos predefinidos e dispõe de um componente encarregado de selecionar esses planos de acordo com o contexto atual. No entanto, em determinadas circunstâncias, a aplicação dos planos pré-estabelecidos pode tornar-se inviável, necessitando assim da concepção de novos planos que considerem as crenças e os recursos disponíveis para o agente (Ahlbrecht; Dix; Fiekas, 2020).

Adicionalmente, o agente evidencia a habilidade de simular os efeitos dos novos planos, levando em conta o estado atual do ambiente, suas próprias crenças,

intenções e os planos dos demais agentes envolvidos no contexto. Os resultados dessa simulação podem ser registrados e empregados no processo de tomada de decisão (Coelho, 2018).

Essa capacidade de adaptação e simulação reforça a autonomia do agente e sua capacidade de responder eficazmente a mudanças no ambiente, possibilitando a tomada de decisões mais informadas e adaptativas em cenários dinâmicos (Ahlbrecht; Dix; Fiekas, 2020).

Uma perspectiva adicional para aprimorar a eficiência dos agentes inteligentes é a integração da capacidade de aprendizado, permitindo que os agentes utilizem os resultados de suas ações para informar e melhorar suas decisões futuras. Essa abordagem possibilita a incorporação de novos planos e objetivos nos módulos de ações e decisões, respectivamente (Nedić; Pang; Scutari, 2018).

A determinação do plano a ser ativado é influenciada por diversos fatores, incluindo os resultados da simulação, a disponibilidade de planos e os objetivos individuais do agente. O plano selecionado é então implementado por meio de ações no ambiente e de interações com outros agentes. Posteriormente, as crenças do agente são atualizadas com base nos resultados das ações executadas (Castro *et al.*, 2020).

Essa capacidade de aprendizado dinâmico não apenas aumenta a adaptabilidade dos agentes às mudanças no ambiente, mas também contribui para uma tomada de decisão mais eficiente e eficaz ao longo do tempo. Além disso, possibilita uma melhor adaptação a cenários complexos e dinâmicos, fornecendo uma base sólida para a otimização contínua do desempenho dos agentes em diversas aplicações (Coelho, 2018).

2.5. APLICAÇÕES DOS AGENTES

Em consonância com os estudos de Jezic, Chen-Burger e Kusek, (2020), dentre as possibilidades de aplicações dos agentes autônomos destacam-se:

- Aplicações Industriais: aplicações na área das telecomunicações, controle de fabricação, controle de processos, controle de tráfego aéreo, distribuição de energia elétrica e sistemas de transportes.
- Agentes de Pesquisa de Informação: aplicações com agentes de pesquisa,

recuperação e filtragem de informação na Internet, aplicações de agentes em assistentes pessoais, gestão de correio eletrônico e filtragem de notícias.

- Comércio Eletrônico: aplicações de mercados eletrônicos para B2C (abreviação da expressão em inglês *business-to-customer* que em tradução livre significa “de empresa para consumidor”) e para B2B (abreviação da expressão em inglês *business-to-business*, que em tradução livre significa “de empresa para empresa”).
- Aplicações de Entretenimento: aplicações de agentes em jogos, desenvolvimento de personagens virtuais e histórias interativas.
- Aplicações Médicas: sistemas de diagnósticos médicos, controle de robôs e equipamentos hospitalares e sistemas de tratamento de pacientes distribuídos.
- Controle de Robôs: controle de robôs para uma navegação mais segura e eficiente.
- Competições de Investigação Científica: competições científicas visando a investigação de uma determinada área a partir de um problema ou plataforma comum.
- Agentes para Simulação: simuladores de condução, voo, combate aéreo, investimentos e inúmeras áreas da robótica.

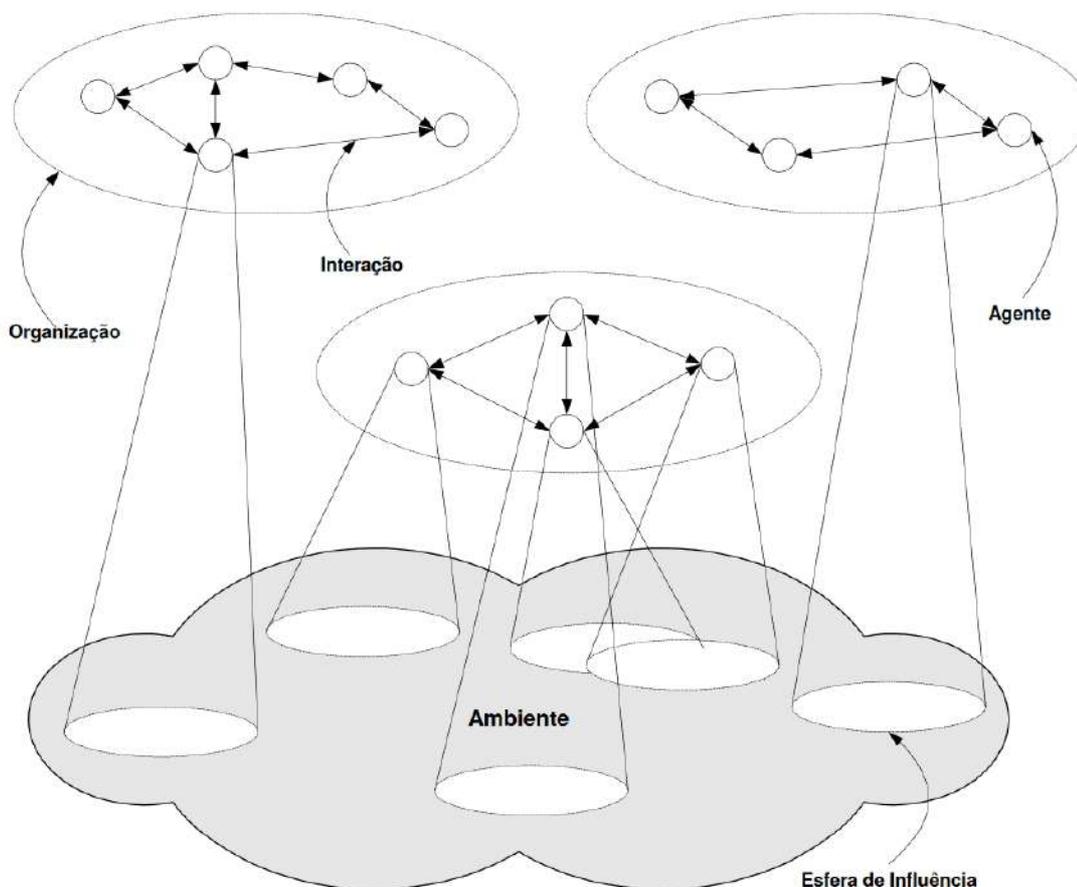
3. SISTEMAS MULTIAGENTES

O Sistema Multiagente (SMA) é um paradigma computacional no qual dois ou mais agentes interagem de forma autônoma e coordenada para realizar atividades específicas e/ou atingir objetivos predefinidos. A implementação de SMA está focada no desenvolvimento de princípios e modelos que permitam a criação de sociedades com agentes autônomos capazes de interagir para atingir os seus propósitos (Boissier *et al.*, 2020). Um dos pontos principais para o desenvolvimento de sociedades com agentes, autônomos é conseguir gerir as interações e as dependências dos diferentes agentes em um mesmo ambiente. Assim, a coordenação é um papel fundamental nos SMA, uma vez que esses são sistemas distribuídos (Yang *et al.*, 2019).

De acordo com Schwartz (2014), há dois procedimentos principais de coordenação: procedimentos aplicáveis em domínios contendo agentes competitivos e procedimentos aplicáveis a domínios contendo agentes cooperativos. No primeiro caso, onde os agentes são preocupados apenas com os seus benefícios, a coordenação por negociação é o método mais aplicado em projetos. Já o segundo procedimento, onde os agentes apresentam uma preocupação pelo bem do ambiente multiagente, há a preocupação com o desenvolvimento de agentes que construam equipes para atingir os objetivos em comum. Nesse contexto, há a relevância de métodos que definem a organização estrutural da sociedade de agentes, a troca de papéis, a alocação de atividades para os diferentes agentes e o planejamento do ambiente multiagente.

Segundo Beer, Fasli e Richards (2011), Os SMA apresentam diversos agentes homogêneos ou heterogêneos, que interagem e/ou atuam em conjunto. Cada agente é um elemento com capacidade autônoma de resolução de problemas que opera assincronamente e respeita os outros agentes. Para que um agente possa atuar como parte do sistema, é necessária uma infraestrutura que permita a interação e/ou comunicação entre os agentes que compõe o SMA. A Figura 12 mostra a estrutura de um SMA.

Figura 12 - Estrutura de um Sistema Multiagente



Fonte: Beer; Fasli; Richards (2011)

O SMA contém múltiplos agentes, cada um com diferentes capacidades de percepção e ação no ambiente. Em geral, cada agente terá uma determinada influência sobre o ambiente, ou seja, será capaz de influenciar diferentes partes do ambiente (Chwif, 2014).

A área dos SMA derivou da Inteligência Artificial Distribuída e a investigação dos SMA é focada no desenvolvimento de modelos computacionais para descrever, implementar e analisar as interações entre agentes em sociedades (Deng; Yang, 2019).

3.1. INTELIGÊNCIA ARTIFICIAL DISTRIBUÍDA VERSUS SMA

A Inteligência Artificial (IA) é uma área da Ciência da Computação, cuja ênfase está no estudo de sistemas inteligentes que exibam características associadas à

inteligência do comportamento humano, tais como aprendizagem, compreensão de linguagem, resolução de problemas, entre outros (Garcia; Giret; Botti, 2015).

O paradigma de transformação de sistemas complexos em entidades, com o objetivo de alcançar maior eficiência, não é uma novidade na área da Ciência da Computação. Ao longo dos anos, os pesquisadores em IA têm buscado desenvolver agentes dotados de comportamento inteligente. No entanto, nos últimos anos, observou-se um aumento significativo no interesse por paralelismo e sistemas distribuídos. Esse fenômeno foi impulsionado pelo avanço no desenvolvimento de *hardware* mais poderoso, pela expansão das redes de computadores e pela facilidade de formação de grupos de estudo dedicados à resolução de problemas complexos. Como resultado, a Inteligência Artificial Distribuída (IAD) emergiu como uma das áreas de maior crescimento na pesquisa em ciência da computação (Bordini *et al.*, 2006).

Conforme mencionado por Dixon (2019), a pesquisa em IAD pode ser categorizada em duas abordagens distintas. A primeira abordagem, que teve origem na década de 70 e continua sendo objeto de estudo atualmente, enfoca os agentes deliberativos. Esses agentes possuem uma representação explícita do ambiente, geralmente na forma de um modelo simbólico, e suas decisões são fundamentadas em raciocínio simbólico (Ning; Han; Zuo, 2019). A segunda abordagem, emergente nos anos 90 e predominante na pesquisa contemporânea, destaca a diversificação das classes de agentes. Nessa perspectiva, os agentes exibem uma progressão em termos de inteligência e autonomia. Em outras palavras, essa linha de pesquisa enfatiza a transição da deliberação para a execução e do raciocínio para ação prática (Krupnik; Mordatch; Tamar, 2020).

A IAD vai além das limitações das arquiteturas de computadores paralelos, linguagens de programação paralela, sistemas operacionais distribuídos ou coordenação de processos simultâneos para resolver problemas. Seu objetivo é abordar desafios por meio de múltiplos agentes que executam subtarefas e se comunicam usando uma linguagem de alto nível (Dixon, 2019).

A investigação da Computação Distribuída era estritamente baseada entre a ligação física dos processadores com a transmissão de dados. Essas questões dominaram as pesquisas de tal forma que o termo processamento distribuído era utilizado apenas para designar esse conjunto de problemas (Zhang *et al.*, 2021). Todavia, a partir das últimas décadas as questões relacionadas ao processamento distribuído começaram a ser tratadas em um nível maior de abstração. Assim, para

Boissier *et al.* (2020), o foco passou a ser as estratégias de decomposição e coordenação do processamento em um sistema distribuído. Simultaneamente, a investigação em IA inspirada em conceitos clássicos de organizações e sociedades, começou a desenvolver soluções baseadas na construção de sistemas inteligentes.

Para Dixon (2019), a resolução de problemas abordados pela IAD divide-se em dois grupos: Resolução Distribuída de Problemas e SMA. A Figura 13 mostra essa divisão da IAD.

Figura 13 - Divisão da Inteligência Artificial Distribuída



Fonte: Dixon (2019)

- **Resolução Distribuída de Problemas:** consiste em abordar um problema específico por meio de sua divisão em módulos distintos. Dessa forma, o planejamento das ações surge a partir da decomposição do problema em várias partes menores, que são distribuídas entre os agentes envolvidos. Essa abordagem permite que os agentes cooperem na resolução do problema, compartilhando conhecimentos e resultados ao longo do processo.
- **SMA:** desempenha um papel crucial na coordenação de comportamentos inteligentes em grupos de agentes autônomos. Por meio dessa abordagem, busca-se estabelecer formas de coordenação que envolvem a partilha de conhecimento, definição de objetivos, desenvolvimento de habilidades e elaboração de planos, com o objetivo final de realizar ações ou solucionar problemas. Essa estrutura permite que os agentes atuem tanto em prol de um objetivo global comum, como também persigam objetivos individuais.

De acordo com Schwartz (2014), pode afirmar-se que a IAD está preocupada essencialmente com quatro assuntos:

- Granularidade dos agentes.
- Heterogeneidade dos agentes.
- Métodos para a distribuição do controle entre os agentes.
- Possibilidades de comunicação entre os agentes.

Essencialmente os SMAs, foco de estudo deste trabalho, estuda os agentes heterogêneos que se agrupam em comunidades sem um controle centralizado e que se comunicam por meio de uma linguagem de alto-nível (Wang; Wang, 2019).

3.2. MOTIVAÇÃO DOS SMA

A principal motivação dos SMA está relacionada com o fato de que maioria dos problemas da área são inerentemente distribuídos de uma ou várias formas (Uhrmacher; Weyns, 2009). Além disso, outras motivações são:

- A dimensão do problema ser elevada para ser resolvida por um único agente.
- Permitir a conexão entre múltiplos sistemas legados.
- Fornecer soluções para problemas onde as informações se encontram distribuídas.
- Permitir uma interface mais natural entre homem-máquina onde ambos atuem como agentes no sistema.
- Oferecer simplicidade e clareza aos projetos.

Em SMA, os agentes podem trabalhar em conjunto para atingir um objetivo geral ou atuar em objetivos individuais, para isso pode ser necessário se relacionar com outros diferentes agentes do sistema visando atingir seus próprios objetivos.

Segundo Goldberg (2021), os SMA oferecem uma maior rentabilidade de recursos para solução de problemas onde a atividade é distribuída e entre as diversas razões para isso, evidenciam-se:

- O domínio do problema, onde exige a distribuição espacial dos agentes.
- O paralelismo, que atribuindo a diversos agentes para a execução ser mais rápida.
- A robustez, pois utiliza diversos e diferentes agentes evitando falhas;
- A escalabilidade, que permite o aumento dos agentes.
- A simplificação da programação, pois divide o problema em diversos subproblemas.
- O estudo do comportamento social, pois permitem o trabalho em conjunto entre os agentes.
- A manutenção da privacidade da informação e o desenvolvimento dos conhecimentos individuais de cada agente.

A utilização de SMA para a solução de problemas de IA apresenta inúmeros benefícios, entre os quais Bordini *et al.* (2006), destacam:

- Resolução mais rápida de problemas devido ao processamento paralelo.
- Redução do tempo de comunicação devido ao processamento estar localizado junto à fonte de informação.
- Aumento da escalabilidade devido à interconexão de múltiplos sistemas com arquiteturas distintas.
- Aumento da capacidade de resposta devido aos sensores, sistemas de processamento e atuadores estarem localizados dentro dos agentes.
- Facilidade de desenvolvimento de sistemas devido a decomposição dos problemas em subproblemas com atuação dos agentes autônomos.

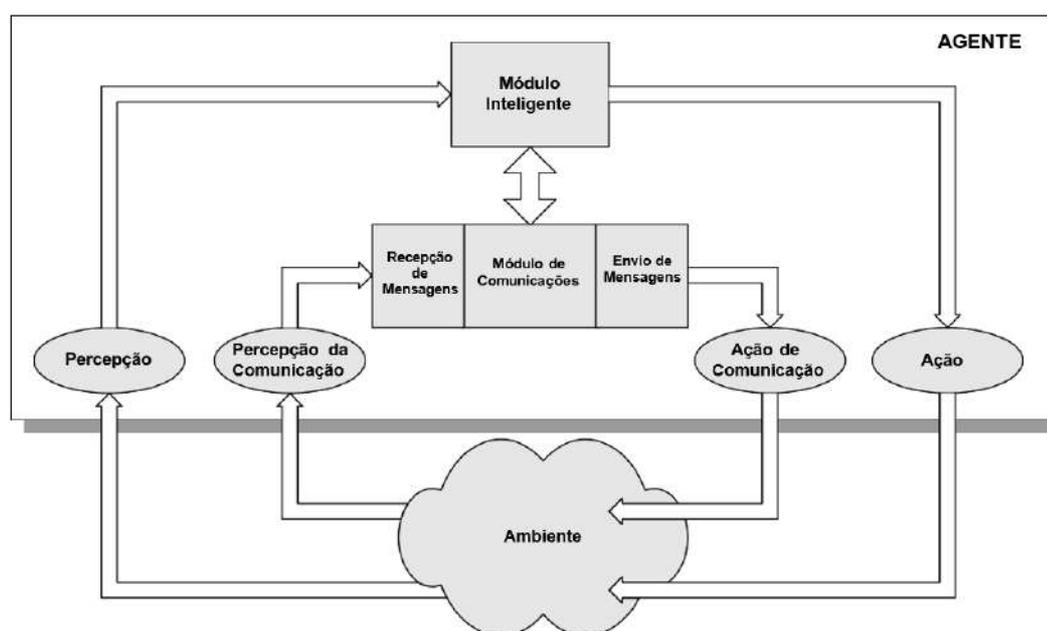
3.3. COMUNICAÇÃO EM SMA

A comunicação em SMA é tratada em um nível mais elevado do que em outras áreas da computação, pois esses sistemas utilizam comunicação de alto nível a partir de linguagens de comunicação próximas das linguagens utilizadas por humanos. Tendo em vista isso, os agentes autônomos apresentam capacidade de percepção,

processamento e atuação em um determinado ambiente. Ou seja, esses agentes apresentam comunicação e habilidade social, ou seja, capacidade para interagir com outros agentes e/ou humanos presentes em seu ambiente (Goldberg, 2021).

Visando incluir a capacidade de comunicação em um agente, deve-se incluir um módulo de comunicações em sua arquitetura que se divide entre componentes de percepção (recepção de mensagens) e de ação (envio de mensagens). Esse módulo está ligado ao módulo central do agente, ou seja, o módulo inteligente e, assim permitindo acesso às mensagens recebidas e definindo quais as mensagens serão enviadas (Beer; Fasli; Richards, 2011). A Figura 14 ilustra o esquema de agente com a capacidade de comunicação.

Figura 14 - Agente com capacidade de comunicação

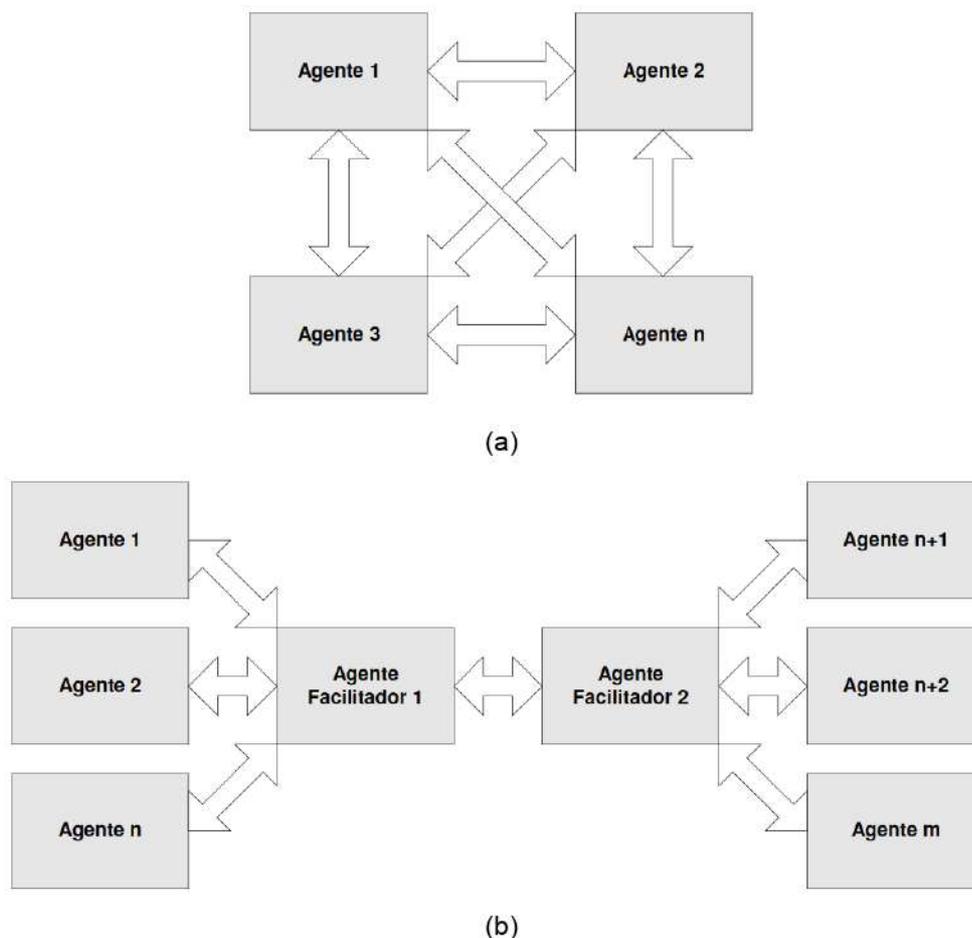


Fonte: Beer; Fasli; Richards (2011)

3.3.1. ARQUITETURAS DO SISTEMA DE COMUNICAÇÃO

De acordo com Waslander (2008), o sistema de comunicação entre os agentes pode apresentar duas arquiteturas básicas: comunicação direta e comunicação assistida. A Figura 15 representa as arquiteturas de comunicação direta e assistida dos agentes.

Figura 15 - Arquiteturas de comunicação de agentes. (a) Direta. (b) Assistida



Fonte: Waslander (2008)

- Comunicação Direta: os agentes realizam a própria comunicação sem qualquer intermediário. Para isso, enviam aos outros agentes as suas capacidades e/ou necessidades para cada agente decidir individualmente as suas decisões. Um dos principais problemas dessa arquitetura é a ausência de um elemento que coordene a comunicação, o que pode causar um bloqueio do sistema se, por exemplo, todos os agentes enviarem mensagens simultaneamente.
- Comunicação Assistida: os agentes utilizam os “agentes facilitadores”, para intermediar a comunicação com os outros agentes. Essa arquitetura resolve parcialmente os problemas causados pela ausência de coordenação da comunicação e diminui a complexidade na realização de comunicação entre os agentes. Isso ocorre porque os agentes não necessitam armazenar informações detalhadas sobre os outros agentes e nem conhecer o seu

endereço específico para efetuar a comunicação. Todos esses aspectos são responsabilidade do “agente facilitador” que efetua a comunicação. Todavia, a existência do “agente facilitador” causa a centralização no sistema e, isso gerar um estrangulamento no sistema de comunicações. Além disso, se esses agentes pararem de funcionar, o sistema de comunicações também deixará de funcionar.

A comunicação entre agentes está relacionada com o nível de implementação das comunicações. Essa comunicação pode ser implementada via memória compartilhada por todos os agentes da comunidade ou passagem de mensagens entre agentes (Uhrmacher; Weyns, 2009). A Tabela 9 exibe as principais diferenças entre essas implementações.

Tabela 9 - Diferenças entre as implementações da comunicação entre agentes

	Memória compartilhada por todos os agentes da comunidade	Passagem de mensagens entre agentes
Sincronização dos agentes	Díficeis de distribuir.	Fáceis de distribuir.
Influência da arquitetura do sistema de comunicação	Baixa. Se adapta da mesma forma em arquiteturas de comunicação direta quanto comunicação assistida.	Alta. Se for utilizado uma arquitetura de comunicação direta, é necessário assegurar que os agentes recebam as mensagens constantemente e prever a possibilidade de uma mensagem enviada não ser recebida pelo agente destinatário. A utilização de uma arquitetura de comunicação assistida pode resolver isso, porém essa implementação tem uma menor eficiência do sistema de comunicação.
Privacidade da informação	Baixa.	Alta.
Rapidez de transmissão	Baixa.	Alta.

Fonte: Uhrmacher; Weyns (2009)

De uma maneira geral, a passagem de mensagens entre agentes é o modo mais utilizado devido à maior privacidade e rapidez com que a mensagem é transmitida.

3.3.2. CARACTERÍSTICAS DA COMUNICAÇÃO

A comunicação tem como metas o compartilhamento do conhecimento, das informações, das crenças ou dos planos com outros agentes e a coordenação de atividades entre os agentes. Segundo Slavkovik (2019), para atingir essas duas metas, é preciso definir uma linguagem em comum de compartilhamento, que é caracterizada por:

- **Sintaxe:** estrutura da linguagem (símbolos) que contém as regras relativas à combinação das palavras e as relações existentes entre essas palavras.
- **Semântica:** significado dos símbolos e de suas combinações.
- **Vocabulário:** conjunto de símbolos utilizados.
- **Pragmática:** conjunto de regras e fórmulas de interpretação dos símbolos usados na comunicação.
- **Modelo do domínio do discurso:** significado que um conjunto de símbolos adota quando interpretado em um determinado contexto de conversação.

Para uma comunicação eficaz, é necessário considerar quatro pontos essenciais, conforme destacado por Chwif (2014):

- **Conteúdo da Comunicação:** o emissor deve selecionar cuidadosamente as informações a serem comunicadas, levando em consideração as características do canal de comunicação e a finalidade da mensagem. É essencial priorizar as informações mais relevantes para garantir uma transmissão eficaz.
- **Timing da Comunicação:** em contextos com largura de banda limitada, é crucial determinar o momento ideal para transmitir a mensagem. Em situações em que a informação não é urgentemente necessária, pode ser mais prudente aguardar e liberar a largura de banda para outras comunicações mais críticas.
- **Destinatários da Comunicação:** as mensagens podem ser direcionadas a um único destinatário, a vários destinatários ou a todos os agentes do sistema. A escolha do público-alvo é fundamental para garantir que a mensagem alcance os destinatários relevantes e contribua para os objetivos do sistema.

- Meios de Comunicação: em ambientes com múltiplos canais de comunicação disponíveis, é imprescindível que o emissor selecione o meio mais apropriado para cada tipo de mensagem. A escolha do meio de comunicação adequado aumenta a eficácia da transmissão e minimiza possíveis ruídos ou distorções na comunicação.

Considerando os pontos mencionados, é possível inferir que os agentes podem substancialmente aprimorar a qualidade e eficiência de suas comunicações em sistemas multiagentes, facilitando a troca de informações e contribuindo para o alcance dos objetivos do sistema. É essencial reconhecer que um agente comunica uma mensagem em um momento específico, contextualizando-a de acordo com o ambiente em que é transmitida e/ou recebida (Deng; Yang, 2019).

3.3.3. CARACTERIZAÇÃO DAS MENSAGENS

De acordo com Uhrmacher e Weyns (2009), é possível caracterizar o significado das mensagens trocadas entre agentes inteligentes, segundo as seguintes características:

- Descritiva versus Prescritiva: as descrições de fenômenos são extremamente complexas para os agentes compreenderem, sendo assim, usualmente os agentes se comunicam através da prescrição de comportamentos, ou seja, por meio de atividades e comportamentos.
- Significado Personalizado versus Convencional: a interpretação que um agente tem de uma mensagem ser diferente do que os outros agentes lhe interpretam, ou seja, há um significado personalizado nesse sentido. Todavia, em SMA é aconselhável a utilização de significados convencionais, ou seja, com uma única interpretação para todos, pois a qualquer momento são adicionados novos agentes.
- Subjetiva versus Objetiva: o significado é objetivo quando é interpretado externamente ao agente e apresenta um efeito explícito no ambiente.
- Perspectiva do Remetente versus Destinatário versus Sociedade: Independentemente do significado convencional ou objetivo da mensagem, a

mesma pode ser expressa de acordo com o ponto de vista do remetente, destinatário ou dos outros agentes atuantes no ambiente.

- **Semântica versus Pragmática:** a pragmática está focada em como os agentes utilizam a comunicação, isto é, como agem em função dessa comunicação. Para isso, considera o estado dos agentes, o ambiente, a sintaxe e a semântica da comunicação.
- **Contextualidade:** as interpretações das mensagens devem considerar o estado do agente, o estado e a história do ambiente, ou seja, as interpretações devem analisar o contexto da comunicação.
- **Cobertura:** as linguagens devem apresentar abrangência suficiente para que os agentes consigam expressar corretamente as suas mensagens durante a comunicação com os outros agentes.
- **Identidade:** o significado de uma mensagem depende da identidade de seus interlocutores e, também, do papel que cada um desempenha no sistema.
- **Cardinalidade:** uma mensagem privada enviada para um agente em particular deve ser entendida de forma diferente de uma mensagem pública, a qual foi enviada a todos os agentes do sistema.

3.3.4. TIPO DE MENSAGENS

Conforme observado por Yang *et al.* (2019), é crucial que todos os agentes possuam a capacidade de comunicação, independentemente das características funcionais que os definem dentro do sistema. A comunicação entre agentes é essencial para a troca de informações, coordenação de ações e alcance de objetivos compartilhados. No entanto, a extensão e a natureza dessa capacidade de comunicação podem variar de acordo com as características individuais de cada agente e sua função específica no contexto do sistema multiagente.

Em consonância com as investigações de Slavkovik (2019), é possível identificar dois tipos fundamentais de mensagens trocadas entre os agentes: perguntas e asserções. Enquanto as perguntas são utilizadas para solicitar informações ou esclarecimentos sobre determinado assunto, as asserções consistem

em afirmações ou declarações feitas pelos agentes com o objetivo de transmitir conhecimento ou expressar intenções.

Com base nas diferentes capacidades de comunicação, é viável classificar os agentes em quatro categorias distintas, cada uma caracterizada por seu nível e modalidade de interação comunicativa. A Tabela 10 apresenta uma síntese das características e capacidades de comunicação desses diferentes tipos de agentes, permitindo uma compreensão mais abrangente e detalhada do papel da comunicação na dinâmica dos sistemas multiagentes.

Tabela 10 - Capacidades de comunicação de diferentes agentes

	Agente Básico	Agente Passivo	Agente Ativo	Agente Interlocutor
Receber Asserções	Sim	Sim	Sim	Sim
Receber Perguntas	Não	Sim	Não	Sim
Enviar Asserções	Não	Sim	Sim	Sim
Enviar Perguntas	Não	Não	Sim	Sim

Fonte: Slavkovik (2019)

- **Agente Básico:** aceita informações do exterior apenas sob a forma de asserções, sendo incapaz de executar outras formas de comunicação.
- **Agente Passivo:** pode realizar diálogos, onde aceita informações do exterior sob a forma de asserções e perguntas do exterior, porém responde apenas na forma de asserções.
- **Agente Ativo:** pode realizar diálogos, onde aceita informações do exterior sob a forma de asserções e realiza perguntas e asserções.
- **Agente Interlocutor:** pode realizar diálogos em que assume papel interlocutor entre os outros agentes. Para isso, esse agente aceita informações do exterior sob a forma de perguntas e asserções e realiza perguntas e asserções.

3.3.5. PROTOCOLOS E NÍVEIS DE COMUNICAÇÃO

Os protocolos de comunicação são classificados em diferentes níveis: inferiores, intermediários e superiores. Os níveis inferiores estão relacionados à definição dos métodos de interconexão entre os agentes. Os níveis intermediários estabelecem o formato da informação a ser transmitida. Já os níveis superiores lidam com as especificações do significado da informação (Boissier *et al.*, 2020).

A estrutura de dados de um protocolo pode ser descrita de maneira sistemática, incluindo elementos essenciais como o emissor e o receptor, além da linguagem para a comunicação. Adicionalmente, são consideradas as funções de codificação e decodificação pertinentes à linguagem utilizada, bem como as ações a serem executadas pelo receptor em resposta às mensagens recebidas (Akhtar *et al.*, 2019).

3.3.6. ONTOLOGIAS

Em SMA para que haja as interações entre os agentes é necessário que exista uma linguagem de comunicação, uma plataforma de comunicação e que os agentes possuam um vocabulário em comum. Todavia, é possível que agentes diferentes possuam terminologias idênticas para significados diferentes e diferentes terminologias para o mesmo significado (Schwartz, 2014).

Essa situação pode ser solucionada por meio da adoção de uma ontologia compartilhada, conceituada por Garcia, Giret e Botti (2015) como a representação do conhecimento de um domínio específico disponibilizada a todos os componentes de um sistema de informação. O emprego de ontologias assegura tanto a compatibilidade, onde um conceito é atribuído à mesma expressão por todos os agentes, quanto a consistência, garantindo que uma expressão possua o mesmo significado entre todos os agentes, dentro de um sistema (Schwartz, 2014).

Em consonância com os trabalhos de Garcia, Giret e Botti (2015), a ontologia aplicada aos SMA consiste em uma representação de conceitos, características e relacionamentos em um determinado domínio que permite o entendimento do conhecimento comum entre pessoas e agentes do sistema (Long; Su; Liu, 2019). Dessa forma, torna-se possível a) uma máquina utilizar o conhecimento

disponibilizado por pessoas ou por outras máquinas; b) máquinas compartilhem conhecimento e; c) reutilização de conhecimento. Para Waslander (2008), há diversos tipos de ontologias:

- Ontologias Genéricas: modelo que apresenta termos genéricos usados como base para definir outros termos compostos.
- Ontologias de Domínio: modelo baseado nas ontologias genéricas, porém especializado em determinadas áreas de conhecimento.
- Ontologias de Representação: modelo que combina os termos que são utilizados nas ontologias genéricas e de domínio.
- Ontologias de Aplicação: apresenta as definições necessárias ao conhecimento de uma área específica da aplicação.

As ontologias são essenciais para o desenvolvimento e utilização de SMA, pois permitem a comunicação entre pessoas ou agentes autônomos, por meio de um vocabulário sobre uma determinada área de conhecimento importante para o sistema (Akhtar *et al.*, 2019).

3.4. APRENDIZAGEM EM SMA

A aprendizagem em SMA é um campo de investigação que visa analisar tópicos como arquitetura, comunicação, negociação, coordenação ou cooperação multiagente. De acordo com Goldberg (2021), as principais motivações para o estudo desse campo são:

- Desenvolver melhores SMA heterogêneos com conhecimentos distintos, tais como capacidade para aprender e ajustar o seu comportamento individual e de grupo automaticamente.
- Perceber melhor os processos naturais de aprendizagem em grupo, tais como agentes cooperando entre si e com pessoas.

Os SMA possuem uma complexidade considerável, sendo assim é de vital importância o desenvolvimento de agentes capazes de aprenderem e se adaptarem

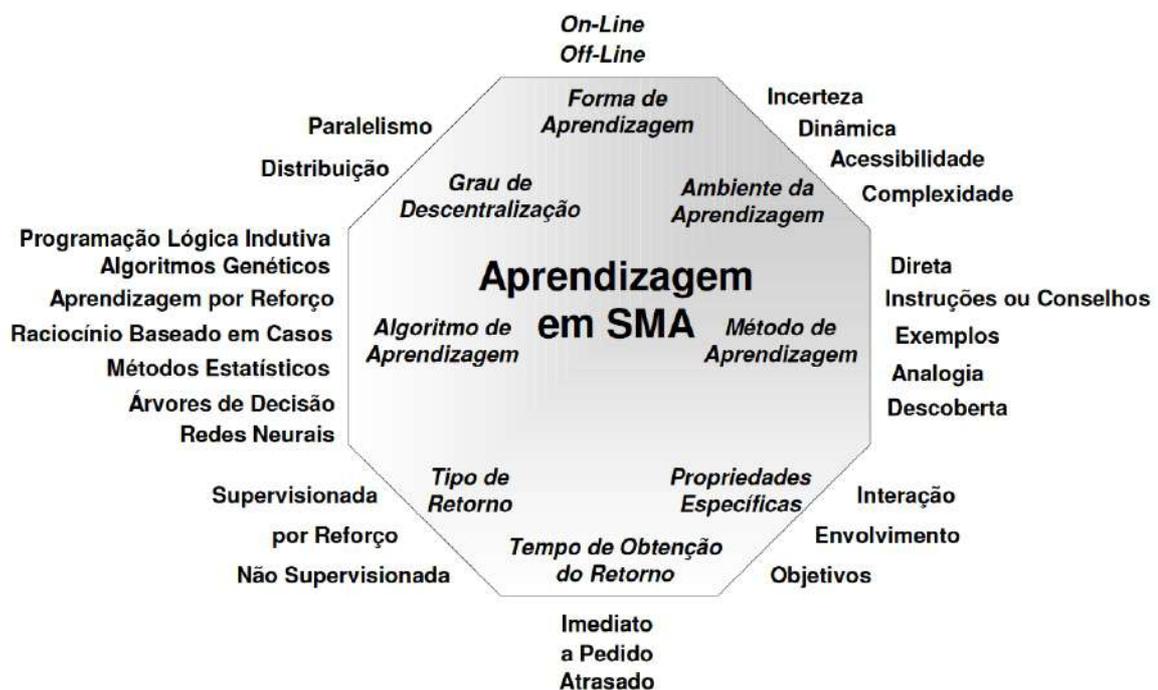
para melhor funcionarem no contexto do sistema (Minarsch *et al.*, 2020).

Para Beer, Fasli e Richards (2011), a aprendizagem em SMA engloba a aprendizagem de agentes isolados e em conjunto e, que essa aprendizagem, seja influenciada pela troca de informações, objetivos compartilhados, pontos de vista comuns sobre o ambiente, normas, convenções sociais etc. Nesse contexto, há dois tipos de aprendizagem:

- Aprendizagem Interativa: onde os agentes visam atingir os seus objetivos de aprendizagem coletivamente.
- Aprendizagem Individual: onde cada agente visa atingir os seus próprios objetivos de aprendizagem, todavia há a influência de outros agentes em sua aprendizagem, conhecimento, crenças e intenções.

Essas categorias de aprendizagem abrangem um largo espectro de aprendizagem que ocorrem no contexto de SMA (Slavkovik, 2019). A Figura 16 apresenta um diagrama que resume os principais tópicos de investigação em aprendizagem.

Figura 16 - Características da aprendizagem em SMA



Fonte: Slavkovik (2019)

Segundo Boissier *et al.* (2020), as abordagens de aprendizagem em SMA

podem ser classificadas com maior precisão a partir das características auxiliares:

- Grau de Descentralização: grau de distribuição e paralelismo no processo de aprendizagem.
- Forma de Aprendizagem: com base na disponibilidade de dados no processo de aprendizagem, é possível classificar as formas de aprendizagem em dois principais modos: *online* e *offline*. No modo *online*, a aprendizagem ocorre concomitantemente à operação dos agentes, enquanto no modo *offline*, a aprendizagem é realizada após a análise dos dados acumulados durante a operação dos agentes.
- Ambiente de Aprendizagem: é necessário considerar a dinâmica, a incerteza, a acessibilidade e a complexidade do ambiente.
- Propriedades Específicas da Interação: incluem diversas propriedades relativas à interação, tais como: o nível, a persistência, a frequência, a estruturação, a variabilidade da interação, entre outras.
- Propriedades Específicas do Envolvimento: o envolvimento pode ser caracterizado pela relevância e pelo papel desempenhado do agente.
- Propriedades Específicas dos Objetivos: considera o tipo de melhoria esperada pela utilização do processo de aprendizagem e a compatibilidade dos objetivos de aprendizagem dos agentes.
- Método de Aprendizagem: podem ser classificados em:
 - Aprendizagem Direta: implantação direta do conhecimento e/ou capacidade sem nenhum tipo de inferência.
 - Aprendizagem por Instruções: combina o conhecimento e/ou capacidade adquirida com o conhecimento e/ou capacidade prévias do agente.
 - Aprendizagem por Exemplos: melhoria do conhecimento e/ou a partir de exemplos positivos e/ou negativos.
 - Aprendizagem por Analogia: transforma o conhecimento utilizado na resolução de um problema para resolver um novo problema análogo.
 - Aprendizagem por Descoberta: adquire novo conhecimento e/ou capacidade a partir de observações, experiências, testes e análise de

resultados.

- Tipo de Retorno da Aprendizagem: indica a melhoria ocorrida até um determinado momento. Pode ser classificada em:
 - Aprendizagem Supervisionada: o retorno especifica uma atividade desejada pelo agente e o objetivo da aprendizagem é obter um resultado mais próximo possível do retorno.
 - Aprendizagem por Reforço: o retorno especifica apenas a utilidade da atividade atual do agente e o objetivo é maximizar essa utilidade.
 - Aprendizagem Não Supervisionada: não há retorno, então, o objetivo é encontrar atividades por meio de tentativas de erro e acerto.
- Tempo de Obtenção do Retorno: quando há retorno, a forma de se obter o mesmo pode ser:
 - Imediato: as ações recebem um retorno imediatamente, o que possibilita ao agente avaliar a qualidade de suas ações.
 - A Pedido: o agente requisita o retorno de uma determinada ação.
 - Atrasada: o retorno é enviado ao agente em um determinado momento, o qual não é controlado pelo agente.

Por meio dessas características é possível caracterizar o processo de aprendizagem em diversos níveis nos SMA (Boissier *et al.*, 2020). É possível observar que devido ao grande número de combinações e possibilidades, a determinação do procedimento de aprendizagem em cada projeto é um fator que deve ser muito bem planejado (Nagoev *et al.*, 2019).

Há diversos tipos de algoritmos computacionais desenvolvidos voltados para a aprendizagem de agentes autônomos. Segundo Waslander, (2008), destacam-se:

- Métodos Estatísticos.
- Raciocínio Baseado em Casos.
- Árvores de Decisão.
- Redes Neurais.
- Aprendizagem por Reforço.

- Programação Lógica Indutiva.
- Algoritmos Genéticos.

3.5. COORDENAÇÃO EM SMA

A coordenação em sistemas multiagentes refere-se ao processo pelo qual os agentes individuais interagem e colaboram entre si para alcançar objetivos comuns ou resolver problemas complexos (Slavkovik, 2019). Essa interação coordenada é essencial para garantir que os agentes trabalhem de forma eficiente e eficaz em conjunto, em vez de operarem de forma isolada e potencialmente conflitante. Segundo Bordini *et al.* (2006), existem duas dificuldades a serem consideradas para a adequada implementação da coordenação em SMA:

- **Agentes Autônomos com Objetivos Distintos:** são agentes individuais que operam autonomamente, muitas vezes com objetivos divergentes, devido à implantação de equipes heterogêneas com diferentes desenvolvedores. Esses agentes podem ter prioridades e metas únicas, resultando em uma variedade de comportamentos e estratégias de ação dentro do sistema.
- **Agentes Autônomos Tomando Decisões em Tempo Real:** são agentes autônomos que estão constantemente tomando decisões em um ambiente dinâmico e em constante mudança, muitas vezes sem o mesmo nível de dinamismo para coordenar suas atividades com outros agentes do ambiente. Isso pode levar a desafios na sincronização e colaboração entre os agentes, especialmente quando as decisões precisam ser tomadas rapidamente em resposta a eventos imprevistos.

Essas dificuldades proporcionaram a criação de diversos procedimentos para coordenação em SMA. Tais procedimentos, essencialmente, afirmam que os agentes devem ser capazes de compartilhar informação e atividades dinamicamente no contexto de SMA (Beer; Fasli; Richards, 2011). A definição de boas estratégias para coordenação proporciona que os agentes executem atividades de forma eficiente por meio de decisões conjuntas sobre quais agentes devem executar uma determinada atividade, quando essa atividade deve ser executada e a quem devem ser

comunicado os resultados (Zhang *et al.*, 2021).

De acordo com Bordini *et al.* (2006), a coordenação entre agentes é necessária devido às relações existentes entre eles, já que frequentemente dependem uns dos outros para alcançar seus objetivos. Os autores identificam as seguintes relações entre agentes autônomos:

- Independência: quando não há qualquer tipo de dependência entre os agentes.
- Unilateral: quando um agente depende do outro, porém a relação inversa não existe.
- Mútua: quando ambos os agentes dependem um do outro para atingirem um objetivo em comum.
- Dependência Recíproca: quando um agente depende do outro para atingir um objetivo e o outro agente depende do primeiro para atingir outro objetivo.

Dixon (2019) ressalta a significância da coordenação entre múltiplos agentes, identificando três motivos fundamentais que justificam essa necessidade:

- Dependência das Ações: este fenômeno ocorre quando as ações destinadas a alcançar os objetivos individuais dos agentes estão interligadas. Em outras palavras, as ações de um agente afetam as ações dos outros agentes e, portanto, a coordenação é crucial para garantir que essas ações estejam alinhadas e contribuam para o objetivo geral.
- Respeito às Restrições Globais: estas restrições surgem quando os agentes individuais são incapazes de cumprir as restrições globais por conta própria. Tais restrições são concebidas para serem cumpridas por um conjunto coordenado de agentes trabalhando em conjunto. Portanto, a coordenação é necessária para garantir que as ações dos agentes sejam compatíveis com as restrições impostas pelo sistema como um todo.
- Recursos: individualmente, nenhum agente possui recursos ou informações suficientes para executar uma atividade ou resolver um problema de forma completa. Portanto, a coordenação entre os agentes é essencial para combinar e compartilhar recursos e informações, possibilitando assim a realização de tarefas que seriam inacessíveis para um único agente.

Esses motivos destacam a importância da coordenação como um elemento fundamental para garantir o funcionamento eficiente e eficaz em SMA, permitindo a realização de tarefas complexas e a resolução de problemas que transcendem as capacidades individuais de cada agente.

Adicionalmente, conforme observado por Beer, Fasli e Richards (2011), mesmo quando os agentes evidenciam habilidades para operar de forma autônoma em SMA, a coordenação entre eles é fundamental para aprimorar sua performance. De acordo com os mesmos autores, essa coordenação oferece vantagens significativas por duas razões principais:

- **Eficiência:** por meio da divisão de tarefas ou troca de informações entre os agentes, os objetivos tendem a ser alcançados de forma mais rápida e eficaz. Esta colaboração permite uma distribuição mais eficiente dos recursos e habilidades dos agentes, maximizando a utilização dos recursos disponíveis.
- **Contenção do Caos:** a coordenação é essencial para prevenir o caos e a anarquia nos SMA. Dado que nenhum agente possui uma visão completa do sistema ou autoridade total, a falta de coordenação pode resultar em interações complexas e desordenadas entre os agentes, levando à desorganização e à ineficiência global do sistema.

Assim, a coordenação efetiva entre os agentes em sistemas multiagentes desempenha um papel crucial na promoção da eficiência operacional e na manutenção da ordem e estabilidade do sistema como um todo. A literatura acadêmica enfatiza a importância dessa coordenação, destacando diversas estratégias e mecanismos que facilitam esse processo.

Segundo Dixon (2019), a coordenação geralmente ocorre em situações de cooperação, onde os agentes trabalham em conjunto para alcançar objetivos compartilhados. Esse processo é facilitado pela comunicação entre os agentes, que permite a troca de informações relevantes e a sincronização de ações. No entanto, é importante ressaltar que a coordenação também pode ser alcançada por meio de abordagens não cooperativas, como negociação ou competição. Esses mecanismos permitem que os agentes ajam de forma independente, buscando maximizar seus próprios interesses, enquanto ainda contribuem para o objetivo geral do sistema.

Além disso, estudos como o de Long, Su e Liu (2019) evidenciam que a coordenação pode ser estabelecida mesmo na ausência de comunicação direta entre os agentes. Isso pode ser alcançado por meio do estabelecimento de uma estrutura organizacional clara, que define papéis, responsabilidades, cadeias de comando e hierarquias entre os agentes. Nesse contexto, cada agente opera dentro de seu papel designado, seguindo diretrizes predefinidas e contribuindo para os objetivos do sistema de acordo com suas competências e capacidades individuais.

Segundo Beer, Fasli e Richards (2011), o desenvolvimento de SMAs é fundamentado em duas abordagens principais: cooperativa e competitiva. Nos SMAs cooperativos, os agentes buscam otimizar o bem-estar do sistema como um todo, priorizando o desempenho global em detrimento do desempenho individual. Nessa perspectiva, a interação entre os agentes é orientada pela cooperação e pelo compartilhamento de recursos, visando alcançar objetivos comuns e maximizar o benefício coletivo.

Por outro lado, nos SMAs competitivos, cada agente é motivado por interesses individuais e busca melhorar sua própria performance, independentemente do impacto no desempenho geral do sistema. Nesse contexto, a interação entre os agentes é caracterizada pela competição por recursos e pela busca por vantagens individuais, podendo resultar em conflitos ou estratégias de maximização de benefícios pessoais (Wang e Wang, 2019).

Essas duas abordagens refletem diferentes dinâmicas de interação e cooperação entre os agentes em um sistema multiagente, influenciando diretamente as estratégias de tomada de decisão e o comportamento coletivo observado. É importante considerar as características específicas do ambiente e dos objetivos do sistema ao escolher entre abordagens cooperativas e competitivas no desenvolvimento de SMA (Krupnik, Mordatch e Tamar, 2020).

4. CIÊNCIA DE DADOS

Dado, do latim *datum*, significa presente no sentido de que algo foi oferecido, é um registro que permite obter conhecimento ou deduzir as consequências de um fato e, é suporte para inúmeras operações (Amaral, 2018). Dados, por sua vez, são um conjunto de valores e/ou ocorrências em um estado bruto que possibilita a obtenção de informações visando benefícios. Para Foster *et al.* (2020), dados brutos são valores recolhidos e armazenados da mesma forma que foram adquiridos, ou seja, sem qualquer tratamento. Esses valores podem ser conjuntos de caracteres, números, imagens, entre outros.

Os dados são registros aleatórios sem quaisquer análises. Eles constituem a matéria prima da informação, ou seja, é a informação não tratada que ainda não apresenta sentido ou relevância isoladamente. Com isso, podem representar os mais diversos significados, ou até mesmo nenhum, se analisados sozinhos e, sendo assim, não transmitem uma mensagem (Castro; Ferrari, 2016).

Em Tecnologia da Informação, dado é o atributo que descreve as características das entidades sobre as quais operam os algoritmos. Essas expressões, são apresentadas de modo que possam ser tratadas por máquina(s). Por base de dados, define-se todo o conjunto de dados que fazem parte do mesmo contexto e são armazenados para utilizações futuras.

A informação consiste na organização e estruturação dos dados, que por si só não possuem sentido ou contexto. A organização dos dados é fundamental para sua transformação em informação, a qual, por sua vez, constitui a base para a construção do conhecimento, sendo este processo mediado pela análise crítica do indivíduo (Barbieri, 2019).

O conhecimento, por sua vez, é resultado de inúmeras informações organizadas de forma lógica, tornando possível a compreensão de um evento ainda não conhecido, bem como as suas causas, histórico, resultados (Marr, 2020). A Tabela 11 exibe as principais características referentes ao dado, à informação e ao conhecimento.

Tabela 11 - O dado, a informação e o conhecimento

Dado	Informação	Conhecimento
Simple observações sobre o mundo.	Dados com relevância e propósito.	Informação valiosa da mente humana, que inclui análise, síntese e contexto.
Facilmente estruturado Facilmente obtido por máquinas Facilmente transferível Frequentemente quantificado	Requer unidade de análise Exige consenso em relação ao significado Exige mediação humana	De difícil estruturação De difícil captura por máquinas De difícil transferência

Fonte: Marr (2020)

Atualmente, décadas após os computadores entrarem no meio social, os dados passaram a se acumular. O mundo está repleto de informação e, também, a informação está se acumulando com maior rapidez.

Desde o século XIX, a sociedade utiliza amostragens para o tratamento de grandes quantidades de dados, todavia essa necessidade advém do período de escassez de informações. Antes das tecnologias digitais de alto desempenho, a amostragem era algo comum, hoje, porém, a quantidade de dados, a velocidade de processamento e a capacidade de armazenamento viabilizam a análise de grandes quantidades de dados. Ressalta-se que, apesar das mudanças na tecnologia viabilizarem essas transformações, a mudança de mentalidade sobre como os dados poderiam ser utilizados foi um fator crítico no impulsionamento da utilização do *Big Data* (Webber; Zheng, 2020).

4.1. BIG DATA

O conceito de *Big Data* é complexo e está intrinsecamente ligado à revolução tecnológica. Não há uma definição singular sobre essa terminologia, existindo várias definições traçadas por estudos e publicações. Para Bahga e Madiseti (2016), o conceito de *Big Data* vem evoluindo rapidamente e, sendo assim, sua definição acaba gerando questionamentos a respeito da abrangência da área. No *Big Data*, os dados tendem a revelar aspectos obdutos e, portanto, não buscados e, isso pode gerar conhecimento consecutivamente.

O presente estudo, está em consonância com Schönberger e Cukier (2014),

que descreve o *Big Data* como sendo uma nova geração de arquiteturas e tecnologias projetadas para extrair valor de grandes volumes de dados armazenados, por meio de alto processamento e velocidade e, assim, vislumbrando a descoberta de novas possibilidades. Possibilidades essas que permitem à sociedade adquirir informação de diferentes maneiras e produzir novos bens e serviços de valor significativo. Todavia, a análise desses dados e informações não é algo simples de ser feito, pois devido ao grande volume é necessário o desenvolvimento de novas ferramentas e técnicas para efetuar sua manipulação. As soluções via *Big Data* visam oferecer abordagens para o tratamento do constante crescimento e complexidade dos dados. Para isso, são considerados o volume, a velocidade, a variedade, a veracidade e o valor.

O conceito de volume é evidenciado pelo cotidiano contemporâneo, uma vez que diariamente há uma grande quantidade de envio de e-mails, transações bancárias, interações em redes sociais, registro de chamadas e tráfego de dados em linhas telefônicas. Além disso, a quantidade de dados vem crescendo exponencialmente e as fontes estão aumentando em número e conteúdo e, também, refletem a tendência de analisar grandes quantidades de dados ao invés de amostras, a fim de se obter maior valor (Kleppmann, 2017). Tendo em vista que a maioria dos dados armazenados foram criados nos últimos anos, é importante frisar que o conceito de volume está relacionado ao tempo, ou seja, o que é grande hoje pode não ser amanhã (Schönberger; Cukier, 2014).

A velocidade dos dados é fundamental para se conhecer as condições atuais e, a partir disso, realizar os processos como, por exemplo, a tomada de decisão. Para Erl, Khattak e Buhler (2015), a importância da velocidade é tanta que em breve haverá ferramentas com a capacidade de analisar os dados em tempo real. Atualmente, os dados são analisados após serem armazenados, contudo o tempo utilizado para realizar esse procedimento já impossibilita que haja uma análise em tempo real.

Os dados coletados também apresentam variedade de formatos, tais como dados em e-mails, redes sociais, fotografias, áudios, smartphones, cartões de crédito, entre outros (Amaral, 2018). Sendo assim, segundo Foster *et al.* (2020), são necessárias ferramentas e técnicas para obter essa variedade de dados, sejam de fontes ou de critérios, para normatizá-los e assim possibilitar as análises. Tendo em vista isso, o *Big Data* escalona a variedade de dados da seguinte forma:

- Dados estruturados: são armazenados em bancos de dados, sequenciados em formato de tabelas.

- Dados semiestruturados: apresentam padrões heterogêneos e são mais difíceis de serem identificados, pois possuem diversos padrões.
- Dados não estruturados: são constituídos por dados de fontes diversificadas tais como áudios, imagens e documentos *online*.

Estima-se que cerca de 92% dos dados gerados atualmente são não estruturados (Kelleher; Tierney, 2018).

O conceito de velocidade, previamente descrito, está bem alinhado ao conceito de veracidade, uma vez que há a constante necessidade de análise em tempo real, ou seja, dados condizentes daquele momento. A verificação dos dados obtidos para adequação e relevância ao propósito é um fator fundamental para se obter dados que adicionem valor ao processo de análise (Gorelik, 2019).

É preciso, a todo momento, focar qual será o valor que a coleta e análise dos dados a partir do *Big Data* trará para a finalidade almejada. Esse valor está relacionado aos benefícios que as soluções de *Big Data* irão trazer e, cada instituição precisa checar se os benefícios trazidos pela análise compensam o investimento nas soluções específicas para isso (Sharda *et al.*, 2019). A Tabela 12 apresenta um resumo dos 5 Vs do *Big Data*.

Tabela 12 - Os 5 Vs do *Big Data*

Atributo	Definição
Volume	Relacionado com a quantidade de dados a serem processados.
Variedade	Estrutura heterogênea das bases de dados.
Velocidade	Taxa em que são gerados os dados e a rapidez necessária para os mesmos serem processados e analisados.
Veracidade	Nível de confiança que pode ser atribuído aos dados recebidos da fonte.
Valor	Potencial financeiro que pode ser obtido por meio do uso das técnicas de <i>Big Data</i> .

Fonte: Sharda *et al.* (2019)

Ressalta-se que o conceito de *Big Data* viabiliza a coleta e o processamento de volumes significativos de dados, provenientes de diversas esferas da vida cotidiana. Posteriormente, esses dados são submetidos à análise com o intuito de extrair informações de relevância para a sociedade (Amaral, 2018).

Essa transformação resulta na geração contínua de volumes crescentes de dados. Consequentemente, percebe-se que todas as atividades humanas realizadas deixam um rastro digital que pode ser analisado. Como destacado por Kelleher e Tierney (2018), a comunicação digital deixa registros, de modo que as interações realizadas em meios digitais são suscetíveis à coleta, armazenamento e análise.

Segundo Schönberger e Cukier (2014), o uso do *Big Data* em pesquisas científicas implica em três grandes mudanças paradigmáticas:

- As amostragens perdem sua importância, uma vez que isso ocorria anteriormente à Era da Informação, onde a informação era escassa.
- A busca por dados de alta precisão deixa de ser uma prioridade crucial, uma vez que isso era motivado pela limitada disponibilidade de dados, tornando imperativo assegurar a qualidade das amostras disponíveis.
- A renúncia da busca pela causalidade, uma vez que o *Big Data* realiza suas previsões por meio de análise de uma extensa base de dados, estabelecendo correlações reais e não casuais.

Para Bahga e Madiseti (2016), o avanço da Tecnologia da Informação viabilizou o *Big Data*, porém, essa técnica é uma continuidade da busca humana por medir, registrar e analisar o mundo. Ressalta-se ainda que os principais conceitos para a compreensão do *Big Data* abrangem os princípios da análise quantitativa, conhecida por Analítica, que será tema da próxima seção deste estudo.

4.2. ANALÍTICA

A fim de uma melhor compreensão sobre o *Big Data*, é importante entender os fundamentos da análise quantitativa de dados, conhecida por analítica. De acordo com Barbieri (2019), significa a utilização ampla de dados, de análise estatística e quantitativa, modelos explanatórios e preditivos e gestão fatural para orientar decisões e agregar valor. É necessário, ainda, que esses dados sejam explorados visando melhorar a tomada de decisão.

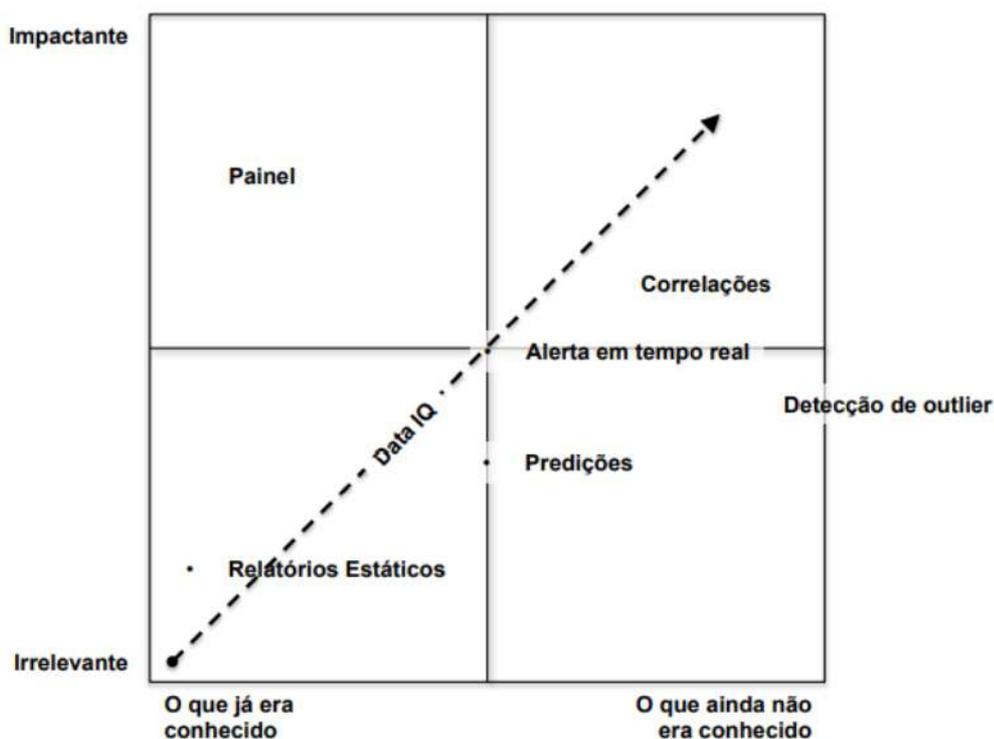
Segundo Marr (2020), vive-se a quarta revolução industrial, onde os dados são recursos indispensáveis dessa revolução. Com isso, possuir apenas recursos naturais não garante a competitividade de um país a longo prazo. O importante é capacidade

de desenvolvimento de produtos tecnológicos e, por isso, os dados passam a ser ativos fundamentais para essa finalidade.

A partir desse contexto, para Sharda *et al.* (2019), a Analítica tem como objetivo criar uma inteligência de dados que extrai percepções significativas a partir dos dados. Existem diferentes tipos de inteligência de dados necessários para resolver problemas, que podem ser medidos pelo Quociente de Inteligência de Dados. A Figura 17 mostra um diagrama sobre o Quociente de inteligência de dados, onde:

- O painel mostra a resposta em tempo real a uma pergunta e a atualiza constantemente.
- Os relatórios estáticos exibem respostas a perguntas anteriores, mas podem conter informações desatualizadas.
- Os outliers podem ser identificados para encontrar oportunidades e alertas extremos.
- As correlações podem ser identificadas e destacadas para identificar elementos essenciais que possam influenciar as tendências e padrões.
- As previsões são úteis, mas podem ser limitadas pela imprecisão dos dados.

Figura 17 - Quociente de inteligência dos dados

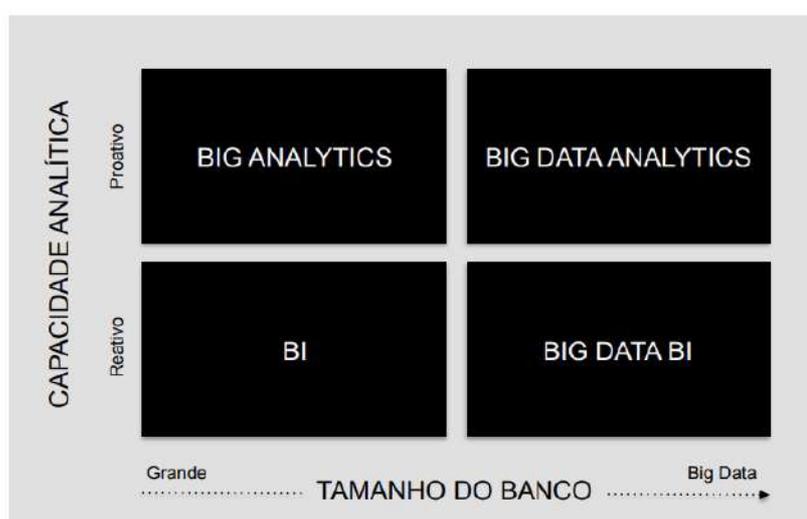


Para Schönberger e Cukier (2014), embora seja possível armazenar grandes quantidades de dados, o tempo necessário para processá-los se tornou excessivo. Foster *et al.* (2020) reforçam essa constatação, observando que a quantidade de dados gerados diariamente, combinada com dados acumulados de anos anteriores, resulta em análises demoradas quando se utiliza modelos tradicionais.

Isso ocorre porque os sistemas de armazenamento tradicionais não oferecem suporte para análises proativas. Quando a quantidade de dados atinge um ponto em que o tempo de processamento se torna inaceitável pelos meios tradicionais, está se lidando com situações de *Big Data* ou problemas analíticos (Sharda *et al.*, 2019). A Figura 18 ilustra técnicas para abordar diferentes questões analíticas, onde:

- Inteligência de Negócios (BI), do inglês *Business Intelligence*: lida com grande quantidade de dados e fornece relatórios aos usuários para análise de desempenho.
- *Big Data BI*: lida com fontes de dados externas ou grandes volumes de dados não estruturados, usando técnicas e ferramentas de *Big Data* para suportar soluções de BI.
- *Big Analytics*: lida com grandes volumes de dados e permite a tomada de decisão baseada em previsões por meio da correlação dos dados.
- *Big Data Analytics*: lida com uma massa enorme de dados e possibilita a tomada de decisão proativa baseada em previsões quase em tempo real, exigindo técnicas e ferramentas de *Big Data* para respostas rápidas.

Figura 18 - Dimensão do banco de dados e capacidade analítica

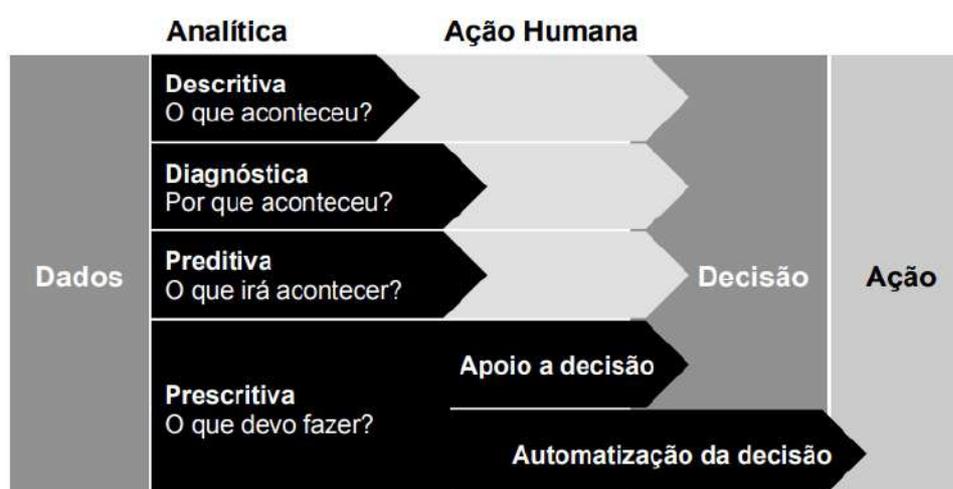


Fonte: Schönberger; Cukier (2014)

De acordo com Kleppmann (2017), a capacidade analítica pode ser classificada em quatro tipos: descritiva, diagnóstica, preditiva e prescritiva. A analítica descritiva abrange a coleta, tabulação, organização e apresentação dos dados e visa descrever os fatos, porém sem os explicar.

A analítica diagnóstica busca os motivos para a ocorrência de determinados eventos no processo de análise. A analítica preditiva emprega os dados do passado visando criar previsões sobre o futuro através da associação entre as variáveis e o cálculo da probabilidade dos fenômenos. A analítica prescritiva sugere ações a serem tomadas e, também, soluciona problemas de otimização. A Figura 19 mostra os tipos de capacidade analítica.

Figura 19 - Tipos de capacidade analítica



Fonte: Kleppmann (2017)

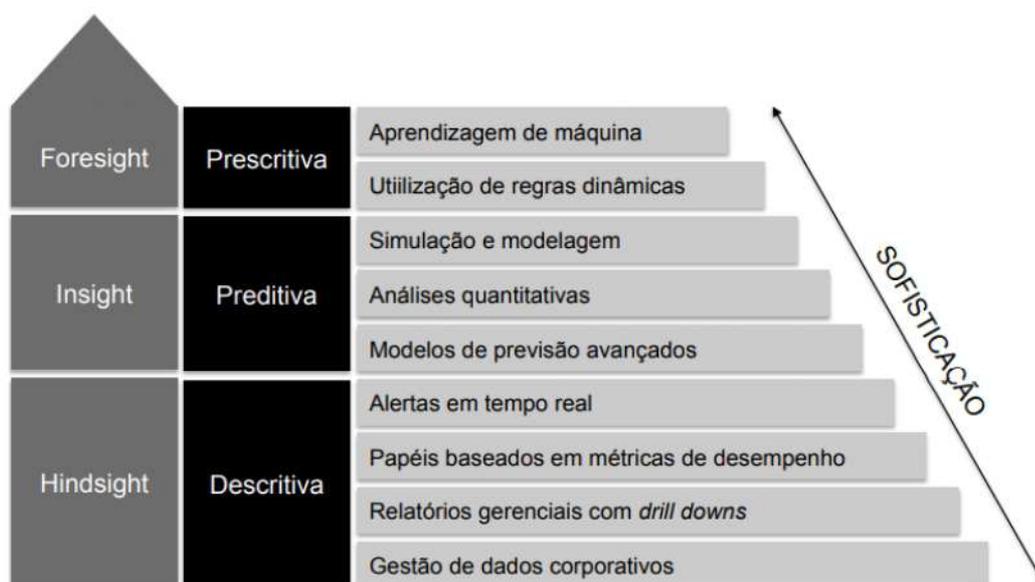
Conforme discutido por Gorelik (2019), a análise de dados abrange uma variedade de tipos, cada um dos quais emprega um conjunto distinto de ferramentas e técnicas para a interpretação e extração de *insights*.

O avanço contínuo das técnicas empregadas na análise de dados está intrinsecamente ligado aos objetivos analíticos subjacentes, que podem ser categorizados em retrospectivos (*hindsight*), interpretativos (*insight*) e prospectivos (*foresight*). Essas categorias refletem a natureza diversificada das abordagens analíticas, que variam desde a simples análise histórica até a previsão e planejamento para o futuro (Sharda *et al.*, 2019).

A Figura 20 ilustra de forma gráfica os diferentes tipos e estágios da analítica,

proporcionando uma visão abrangente das diversas facetas dessa disciplina em constante evolução.

Figura 20 - Estágios da Analítica



Fonte: Gorelik (2019)

- No *hindsight* estão as análises mais simples, onde se aborda uma análise descritiva e preocupa-se com a análise retrospectiva do objeto estudado. Nesse tipo de análise, limita-se a entender o que aconteceu e tem como característica a forte influência humana.
- No segundo tipo de analítica, o *insight*, observa-se uma maior preocupação com o que pode acontecer e uma maior sofisticação das técnicas e ferramentas aplicadas, embora a ação humana ainda seja bem representativa.
- O *foresight* é terceiro tipo de analítica e, também, é o estágio mais sofisticado e em por objetivo definir quais ações tomar no presente para se obter os melhores resultados futuros. Nesse tipo de analítica há pouca influência humana e as decisões são via ferramentas e técnicas de *Big Data*.

Neste contexto, Kleppmann (2017) destaca a exploração detalhada como uma técnica analítica que permite aos usuários explorar dados em diferentes níveis de detalhamento, transitando de uma visão geral para informações mais específicas e detalhadas. Essa abordagem é valiosa para identificar padrões e tendências nos dados, bem como para descobrir *insights* ou anomalias em diferentes granularidades.

Na prática da mineração de dados, a aplicação da exploração detalhada possibilita que os analistas iniciem sua análise examinando um conjunto de dados em um nível macro, como um resumo ou uma visão geral, e, posteriormente, aprofundem-se em subconjuntos de dados específicos.

Essa prática abrange a análise de diversas dimensões dos dados, tais como tempo, localização, categoria ou segmento de mercado, e a investigação das inter-relações entre esses fatores (Gorelik, 2019). Dessa forma, a integração entre as diversas abordagens analíticas e as técnicas específicas, como a exploração detalhada, promove uma compreensão mais ampla e minuciosa dos dados e dos fenômenos em estudo.

As técnicas de análise utilizadas no *Big Data* englobam uma série de áreas, tais como estatística, redes neurais, mineração de dados, aprendizagem de máquina, reconhecimento de padrões, métodos de otimização, entre outros (Escovedo; Koshiyama, 2020). Em consonância com Castro e Ferrari (2016), o *Big Data* quando aliado à analítica é denominado *Big Data Analytics* e, atualmente, é a técnica com menor interferência humana, mais sofisticada e que vem crescendo cada dia mais devido aos benefícios oferecidos.

O potencial do *Big Data Analytics* é maximizado quando o processo de tomada de decisão é empregado constantemente, ou seja, a cada nova iteração o processo se aperfeiçoa e torna-se mais preciso. Assim, Erl, Khattak e Buhler (2015) consideram que existem cinco fases principais no processo de tomada de decisão do *Big Data Analytics*, estando essas fases classificadas em dois grupos. A Figura 21 representa as principais fases do *Big Data Analytics*.

Figura 21 - Principais fases do *Big Data Analytics*

Data Management			Analytics	
Aquisição e Armazenamento	Extração e limpeza	Integração, agregação e carregamento	Modelagem e análise	Interpretação

Fonte: Erl; Khattak; Buhler (2015)

De acordo com Foster *et al.* (2020), *Data Management* envolve as etapas de aquisição e armazenamento das informações que anteciparão a transformação dos

dados, realizando a remoção de inconsistências e estruturando a base de dados para ser utilizada no desenvolvimento de modelos e análises. Segundo os mesmos autores, *Analytics* é definida como um conjunto de técnicas utilizadas para analisar os dados de forma a extrair *insights* para proporcionar tomadas de decisão de forma mais inteligente. Para Amaral (2018), *Big Data Analytics* deve ser uma prioridade para as instituições que desejam realizar tomadas de decisão mais rápidas, precisas e inteligentes, e, assim conseguir vantagens competitivas.

4.3. MINERAÇÃO DE DADOS

A mineração de dados, do inglês *Data Mining*, é a extração de conhecimento por meio da utilização de ferramentas computacionais, a partir das bases de dados que estão sendo coletadas e armazenadas atualmente. A obtenção desse conhecimento implícito tem sido útil, sobretudo, para as instituições tomarem decisões mais assertivas almejando potencializar a competitividade (Foster *et al.*, 2020). Esse conjunto de técnicas estatísticas e de IA visa estabelecer novas relações entre os dados, visando por fim, produzir novos conhecimentos (Barbieri, 2019).

Segundo Erl, Khattak e Buhler (2015), o conceito de descobrir conhecimento em bases de dados pode ser sintetizado como sendo o processo não trivial de identificar novos padrões válidos e, principalmente, potencialmente úteis. Sob a ótica de Webber e Zheng (2020), a mineração de dados é o processo de análise de um conjunto de dados que objetiva a descoberta de padrões que representem informações úteis. Por sua vez, um padrão pode ser definido como uma afirmação sobre uma distribuição probabilística, sendo que o mesmo, pode ser expresso na forma de regras, funções, fórmulas, entre outros.

O estudo de Kelleher e Tierney (2018), apresenta diversas aplicações de mineração de dados, desde trabalhos acadêmicos a implantações em empresas de diversos segmentos. Segundo os mesmos autores, a mineração de dados é um processo que culmina com a Descoberta de Conhecimento em Bases de Dados, do inglês *Knowledge Discovery in Databases*, e é em consonância com essa definição que esse termo será utilizado nesse estudo.

O crescente interesse relacionado à mineração de dados deve-se ao fato de as instituições estarem coletando e armazenando grandes volumes de dados em

decorrência do aumento e diversificação dos serviços de coleta e armazenagem de dados e à redução dos preços dos meios e serviços de armazenamento de dados (Gorelik, 2019).

A popularização do uso de Armazém de Dados, do inglês *Data Warehousing*, que são grandes bancos de dados criados visando à análise e ao suporte à decisão, tende a aumentar a quantidade de informações disponíveis. Os métodos tradicionais de análise de dados, como consultas e planilhas, já não são apropriados para tais quantidades de dados, uma vez que não conseguem analisar o conteúdo desses dados para a obtenção de conhecimentos (Barbieri, 2019). Para Kleppmann (2017), as instituições têm informações que possibilitam decisões, porém apenas algumas conseguem otimizar esse processo decisório.

As ferramentas de mineração de dados atuam com grandes bases de dados e devem retornar conhecimento novo e relevante; todavia, esse retorno deve ser avaliado criteriosamente, devido às inúmeras equações e relações geradas que podem inviabilizar o processamento dos dados (Sharda *et al.*, 2019).

De acordo com Kleppmann (2017), para gerar conhecimento a partir de uma base de dados com diferentes fontes e formatos, deve-se, primeiramente, organizar os dados. Após isso, análises estatísticas e de IA devem ser aplicados e, assim, novas relações poderão ser descobertas. Por isso inúmeras ferramentas têm sido utilizadas para análise, pois reconhece-se que há padrões, relacionamentos e regras ocultas nos dados que não podem ser encontrados utilizando métodos tradicionais.

Sob a ótica de Foster *et al.* (2020), há a necessidade de transformar o grande volume de dados armazenados em informações significativas e, conseqüentemente, conhecimento; porém a análise dessa quantidade de dados ainda é demorada, dispendiosa, pouco automatizada, sem precisão e sujeita a erros. Para o autor, a automatização dos processos de análise, com a utilização de ferramentas ligadas diretamente à base de dados é uma necessidade imprescindível.

A mineração de dados é fundamentalmente proveniente de três linhagens: a Estatística, a IA e o Aprendizado de Máquinas. A mais antiga é a Estatística e, sem a mesma, não seria possível a mineração de dados. A Estatística envolve conceitos como média, mediana, moda, variância, covariância, assimetria, correlação, análise de conjuntos, distribuição probabilística, intervalos de confiança, entre outros, e, todos são utilizados para estudar os dados e os seus respectivos relacionamentos (Deitel; Deitel, 2019).

A segunda linhagem é a IA, construída a partir dos fundamentos da heurística, e, que visa imitar o modo como os seres humanos resolvem os problemas estatísticos. Segundo Escovedo e Koshiyama (2020), em função dessa abordagem, a IA requer um alto poder de processamento, o que era inviável até alguns anos atrás devido ao elevado custo e às limitações computacionais. Todavia, hoje, o custo é mais acessível e os recursos computacionais apresentaram grande evolução e aperfeiçoamento.

A terceira linhagem é o Aprendizado de Máquina, do inglês *Machine Learning*, que é considerada como a junção das melhores práticas e técnicas da Estatística e IA. No Aprendizado de Máquina, os programas “aprendem” com os dados estudados a fim de tomarem decisões baseadas nas informações contidas nesses, usando combinações e algoritmos de heurística e análise estatística para alcançar esse objetivo (Castro; Ferrari, 2016).

De acordo com Kelleher e Tierney (2018), o grande objetivo da mineração de dados não é simplesmente encontrar padrões e tendências em meio a uma imensa quantidade de dados disponíveis em diversas bases, mas sim extrair conhecimento inteligível e utilizável para o apoio às decisões.

A origem dos dados utilizados é coletada em diferentes instantes e em diversas fontes, e, isso cria um esforço para a consolidação e agrupamento dos dados que serão a base para todo o processo. A compreensão do negócio no qual os dados estão inseridos é vital para a interpretação dos mesmos (Deitel; Deitel, 2019).

Em função da heterogeneidade dos dados, o pré-processamento é fundamental para a preparação dos dados, investigando e corrigindo as inconsistências de valores fora da normalidade (*outliers*) e observações errôneas a fim de se obter o conhecimento (Schönberger; Cukier, 2014).

A partir do pré-processamento, que convertem os dados em tratáveis e homogêneos, a mineração dos dados pode ser iniciada. Sendo assim, a interpretação, a compreensão e a aplicação dos resultados encontrados por meio de bases de dados são os fatores que possibilitam o conhecimento para apoiar às decisões (Amaral, 2018). Para Kleppmann (2017), há diferentes níveis de mineração de dados:

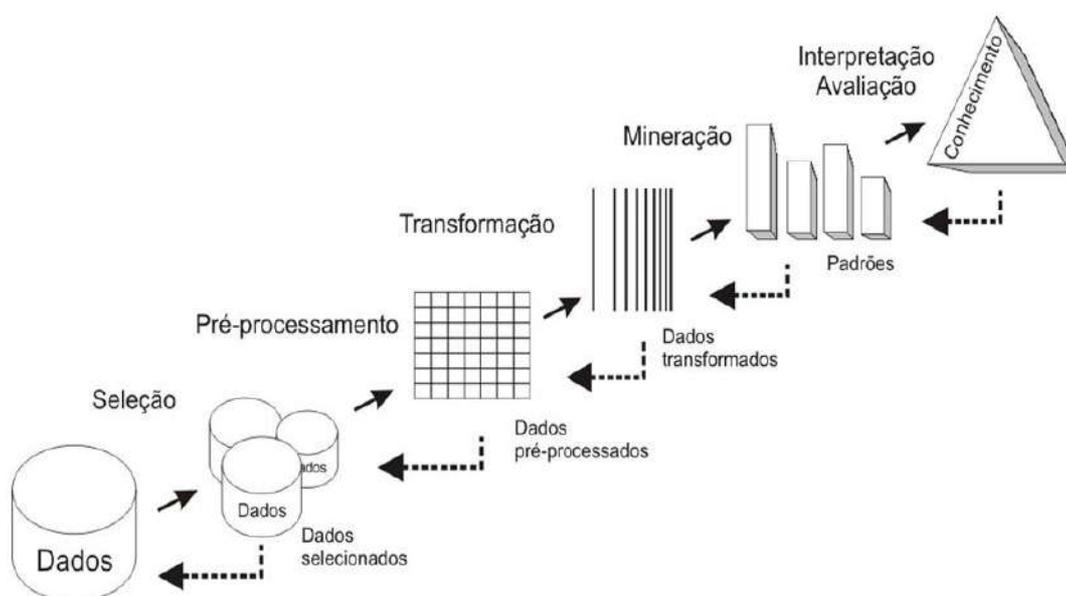
- Mineração de dados pura: neste modelo o problema é formulado genericamente, sem restrições e, com isso, não há indicações dos tipos de descobertas a serem retornadas. Os resultados podem apresentar padrões inesperados úteis ou também, há o risco de retorno de padrões já conhecidos

e que não trarão novas perspectivas ao processo.

- Mineração de dados direcionada: nesse modelo, o problema é mais específico e o retorno tende a trazer novas tendências ao processo.
- Teste de hipóteses e refinamento: neste modelo, a pesquisa é delineada e específica, fornecendo hipóteses e aguardando a validação do sistema e o refinamento das descobertas. Além disso, em casos de insucesso, as hipóteses são modificadas visando à obtenção de novas perspectivas ao processo.

De acordo com Fayyad *et al.* (1996), o ciclo de descoberta de conhecimento em bases de dados por meio do uso de diferentes hipóteses, é a fase principal da mineração de dados. Diferentes métodos foram criados para isso, todavia a inferência de hipóteses e verificação é apenas uma parte de todo o ciclo de descoberta de conhecimento. A Figura 22 mostra o ciclo de descoberta de conhecimento em bases de dados.

Figura 22 - Ciclo de descoberta de conhecimento em bases de dados



Fonte: Fayyad *et al.* (1996)

O presente estudo utilizará ciclo de descoberta de conhecimento em bases de dados referenciando nas obras Kleppmann (2017) e de Schönberger e Cukier (2014), cuja estrutura é definida pelas etapas:

1. Definição e análise do problema.
2. Entendimento e preparação dos dados.
3. Configuração da busca por conhecimento.
4. Busca de conhecimento.
5. Refinamento das descobertas.
6. Aplicação do conhecimento na resolução do problema.
7. Depuração e avaliação dos resultados.

A) *DEFINIÇÃO E ANÁLISE DO PROBLEMA*

O processo se inicia a partir da definição do problema a ser resolvido. O problema deve ser previamente definido e explícito para que os dados possam ser coletados e analisados posteriormente. A proposta desta etapa é dividir o problema em fatores importantes a fim de traçar metas para se prosseguir para as etapas seguintes. É vital observar se o que está sendo definido é um problema ou apenas um sintoma, que possui uma causa mais profunda (Kelleher; Tierney, 2018).

De acordo com Foster *et al.* (2020), é possível utilizar métodos, como a análise de causa raiz, para definir corretamente o problema e garantir que essa etapa seja bem-sucedida. A análise de causa raiz é um processo que visa descobrir a causa raiz de problemas. Essa análise parte do princípio de que é fundamental resolver os problemas subjacentes do que simplesmente tratar sintomas pontuais. Dentre as diversas ferramentas para se aplicar a análise de causa raiz, o Diagrama de Causa e Efeito é uma das práticas mais aplicadas e com melhor retorno.

O Diagrama de Causa e Efeito é uma ferramenta que possibilita levantar as causas-raízes de um problema, analisando os fatores envolvidos na execução do processo. Esse diagrama foi criado na década de 60 por Kaoru Ishikawa e considera todos os aspectos que podem ser a principal ocorrência do problema, assim, ao utilizá-lo, as chances de que algum detalhe seja esquecido e/ou omitido, diminuem consideravelmente.

Nessa ferramenta, todo problema possui causas específicas, e as mesmas, devem ser analisadas uma a uma para verificar qual delas é realmente o motivo do problema que se deseja eliminar. Em outras palavras, essa técnica propõe que excluindo as causas, o problema será suprimido (Webber; Zheng, 2020).

De acordo com Barbieri (2019), o Diagrama de Espinha de Peixe apresenta a relação entre o resultado indesejado de um efeito e as possíveis causas que podem contribuir para que esse resultado tenha ocorrido. Dessa forma, a sua relação com a imagem de uma espinha de peixe ocorre devido ao fato que é possível considerar suas espinhas como as causas dos problemas, que contribuirão para a descoberta de seu efeito. Para Marr (2020), é possível aplicar o Diagrama de Ishikawa em diversos contextos e de diferentes maneiras, tais como:

- Visualizar as causas principais e secundárias de um problema.
- Ampliar a visão das possíveis causas de um problema, observando-o de maneira mais abrangente.
- Identificar soluções, considerando os recursos disponíveis.
- Desenvolver melhorias nos processos.

B) ENTENDIMENTO DOS DADOS E PRÉ-PROCESSAMENTO

Os métodos do ciclo de descoberta de conhecimento são fundamentais para gerar *insights* e solucionar problemas, mas seu sucesso depende da disponibilidade e qualidade dos dados utilizados. De acordo com Erl, Khattak e Buhler (2015), é crucial considerar duas questões essenciais: a disponibilidade dos dados relevantes para o problema em questão e a capacidade de generalização desses dados para aplicação em questões objetivas.

No que diz respeito à disponibilidade dos dados, é necessário realizar uma cuidadosa seleção dos atributos relevantes para cada tabela do banco de dados. Os dados devem ser provenientes de diversas fontes e integrados em um único conjunto, visto que ferramentas de descoberta de conhecimento tendem a fornecer resultados mais precisos quando aplicadas a conjuntos de dados unificados (Escovedo; Koshiyama, 2020). Além disso, o modelo utilizado deve assegurar que o volume de dados seja suficiente para abranger uma variedade de resultados por meio das variáveis relevantes.

Quanto à capacidade de generalização dos dados, Castro e Ferrari (2016) enfatizam a importância de possuir dados disponíveis para contrastar inúmeras generalizações com grupos de controle apropriados. Se o objetivo é avaliar a eficácia de determinado tratamento em relação a uma necessidade específica, é imprescindível dispor de dados tanto dos elementos tratados quanto dos não tratados.

Dessa forma, a análise cuidadosa da disponibilidade e qualidade dos dados é essencial para o sucesso do ciclo de descoberta de conhecimento, garantindo resultados confiáveis e *insights* relevantes para a solução de problemas.

Após a definição do problema em questão, o próximo passo consiste na realização da coleta de dados pertinentes, a qual pode ser conduzida a partir de diversas fontes, incluindo bancos de dados e armazéns de dados. A partir desse processo, as informações são extraídas e armazenadas em formatos que facilitem o acesso e a interpretação por meio de algoritmos de mineração de dados. De acordo com Bahga e Madiseti (2016), em determinadas circunstâncias, a estratégia mais adequada pode ser a construção de uma nova base de dados, embora tal abordagem possa acarretar custos adicionais. Por exemplo, no contexto da retenção de clientes, é comum que não haja apenas uma única base de dados disponível.

Figura 23 - Exemplo de uma base de dados de retenção de clientes

Nome	Endereço	CEP	Tempo de contrato	Serviços	...	Retido?
Jones	Rua Elm, 51	94305	3,25	A		S
Davis	Rua Thayer, 14	82138	0,75	ACD		N
:						

Fonte: Amaral (2018)

A partir da Figura 23 é possível considerar que o fragmento de dados mostrado representa uma base de dados contendo informações de clientes, cujo objetivo da mineração de dados é evitar a perda dos dados. Segundo Amaral (2018), essa tabela é um exemplo da formação de uma nova base de dados a partir de outras fontes.

Neste exemplo, cada cliente apresenta o nome, o endereço, o CEP, o tempo de contrato, os serviços e uma variável indicando se o mesmo permanece como cliente ativo. Provavelmente o nome, o endereço e o CEP estão armazenados em uma tabela do sistema de cadastro de clientes da instituição. Para a mineração de dados, esses dados geográficos são importantes, pois possibilitam estabelecer outros clientes em potencial, uma vez que pessoas com características (idade, estado civil etc.) e estilos de vida (padrões de gastos, faixa salarial etc.) tendem a residir próximas. O tempo de contrato é uma informação obtida a partir de uma base de dados do sistema de faturamento. A partir disso, a mineração de dados pode relacionar informações adicionais, tais como se o cliente costuma pagar as faturas em dia, a

quantidade de vezes em que foi aberto um chamado de atendimento ou registro de reclamação, entre outros.

O procedimento previamente descrito é meramente ilustrativo, uma vez que não foram definidos os atributos que podem ser adquiridos dos dados reais e se esses atributos podem ser encontrados apenas com esse conjunto de dados disponível. Nesse cenário, poderia ser analisado se os pagamentos das contas foram feitos no prazo, especialmente nos últimos meses. Isso porque, assume-se que o pagamento de uma fatura em atraso tende a significar dificuldades financeiras, descontentamento do cliente com o serviço ou ambos.

O pré-processamento de dados no ciclo de descoberta de conhecimento é apresentado como uma fase importante, onde é necessário um bom domínio e conhecimento da área. Isso ocorre porque usualmente os dados coletados são de má qualidade, ou seja, apresentam incorreções, imprecisões, além de valores desconhecidos. Embora a maioria dos algoritmos aplicados nessa fase, que foram desenvolvidos para manipular dados em tais situações, observa-se que esses algoritmos apresentem melhores resultados quando os problemas previamente mencionados sejam corrigidos ou minimizados (Kelleher; Tierney, 2018).

De acordo com Schönberger e Cukier (2014), o pré-processamento tem por objetivo detectar e remover anomalias nos dados visando à melhoria de sua qualidade. Esse processo engloba a verificação da consistência das informações, a correção dos possíveis erros e o preenchimento e/ou a eliminação de valores nulos e/ou redundantes. Ou seja, os dados duplicados e/ou corrompidos são identificados e removidos. A execução do pré-processamento corrige a base de dados e, assim, elimina consultas desnecessárias que afetariam o desempenho. Normalmente esse processo é executado junto a um perito no negócio ao qual correspondem os dados, já que a detecção e correção de anomalias exigem conhecimento especializado.

Sob a ótica de Webber e Zheng (2020), de uma maneira geral, o pré-processamento de dados é um processo semiautomático, ou seja, essa etapa depende da capacidade do analista de dados em identificar os problemas e utilizar as melhores técnicas para solucioná-los. O autor também classifica as técnicas utilizadas no pré-processamento em dois grupos:

- Fortemente dependente de conhecimento de domínio: atividades que só podem ser alcançadas quando há conhecimento específico do domínio, ou seja, é possível implementar um método automático, desde que esse método

apresente um conhecimento específico fornecido previamente.

- Fracamente dependente de conhecimento de domínio: atividades que podem ser alcançadas por métodos que extraem as informações necessárias para tratar o problema dos próprios dados, ou seja, essas atividades podem ser implementadas por métodos que são mais automáticos do que os utilizados em atividades que dependem fortemente de conhecimento de domínio.

C) *TRANSFORMAÇÃO DOS DADOS*

O principal objetivo desta etapa é transformar a representação dos dados visando superar as limitações existentes nos algoritmos que serão utilizados para a extração dos padrões, ou seja, a decisão das transformações necessárias depende do algoritmo selecionado na etapa de mineração de dados. Esse processo visa atender às restrições impostas pelos algoritmos de mineração de dados e, certas ferramentas podem ser aplicadas somente em conjuntos de dados com atributos nominais, já outros algoritmos encontram padrões apenas sobre variáveis numéricas (Marr, 2020).

D) *NORMALIZAÇÃO*

É uma técnica que transforma os valores dos atributos de seus intervalos originais para um intervalo específico, como por exemplo $[0, 1]$ ou $[-1, 1]$. Essa técnica é fundamental para os métodos que calculam distâncias entre atributos, tal como o método “k-vizinhos mais próximos”, que tende a dar mais importância aos atributos que apresentam um intervalo maior de valores (Bahga; Madiseti, 2016).

De acordo com Marr (2020), a normalização dos dados desempenha um papel crucial na preparação e pré-processamento de conjuntos de dados para análise. Essa técnica busca reduzir a variabilidade entre os atributos de uma fonte de dados, garantindo que todas as variáveis residam dentro de um intervalo de amplitude definido. Essa homogeneização dos atributos é particularmente necessária quando os dados apresentam unidades diversas ou valores extremamente heterogêneos. A normalização, portanto, proporciona uma base uniforme para a análise subsequente, facilitando a comparação e interpretação dos dados. Ao aplicar essa prática, os pesquisadores podem mitigar distorções decorrentes de diferenças nas escalas ou

magnitudes dos atributos, promovendo uma análise mais precisa e robusta.

De acordo com Bahga e Madiseti (2016), duas das principais técnicas de normalização são a Normalização por Desvio Padrão, descrita na Equação 1, e a Normalização por Faixa de Variação, apresentada na Equação 2.

$$y = \frac{x - \mu}{\sigma} \quad (1)$$

$$y = \frac{x - \min}{\max - \min} \quad (2)$$

Na Equação 1 e na Equação 2, y representa o novo valor normalizado, enquanto x representa o valor atual. Na Equação 1, μ representa a média da variável e σ representa o desvio padrão da variável. Já na Equação 2, \max representa o valor máximo e \min representa o valor mínimo.

E) DISCRETIZAÇÃO DE ATRIBUTOS QUANTITATIVOS

De acordo com Foster *et al.* (2020), é notável que uma variedade de algoritmos apresenta restrições em sua capacidade de processamento, uma vez que se restringem ao manuseio exclusivo de atributos qualitativos. Entretanto, é comum encontrar conjuntos de dados que incluem tanto atributos quantitativos quanto qualitativos. Nesse cenário, a aplicação desses algoritmos requer a conversão dos atributos quantitativos em uma forma qualitativa, geralmente representada por faixas de valores discretos.

Essa abordagem visa aprimorar a interpretação e o processamento de dados pelos algoritmos, especialmente em situações em que a manipulação de informações quantitativas é desafiadora para os sistemas computacionais. O propósito fundamental é capacitar os algoritmos para lidar de maneira eficaz com conjuntos de dados que contenham atributos quantitativos, os quais, de outra forma, poderiam ser de difícil interpretação ou processamento direto.

A integração dessas técnicas tem um papel crucial em ampliar a aplicabilidade e a eficácia das análises realizadas pelos algoritmos em diversos contextos. Ao superar as limitações relacionadas à interpretação de dados quantitativos, essa

abordagem contribui significativamente para a utilização mais ampla e eficiente dessas técnicas analíticas em áreas como a ciência de dados, aprendizado de máquina e inteligência artificial.

F) *TRANSFORMAÇÃO DE ATRIBUTOS QUALITATIVOS EM QUANTITATIVOS*

Há algoritmos que não são capazes de manipular atributos qualitativos, assim é preciso aplicar determinadas técnicas para se efetuar essas transformações e isso depende das características e limitações de cada algoritmo. Assim, atributos qualitativos sem ordem inerente, tais como amarelo, verde e vermelho, podem ser mapeados como números. Já os atributos qualitativos com ordem, tais como pequeno, médio e grande, podem ser mapeados como valores numéricos a fim de manter a ordem dos valores (Castro; Ferrari, 2016).

G) *ATRIBUTOS DE TIPOS DE DADO COMPLEXOS*

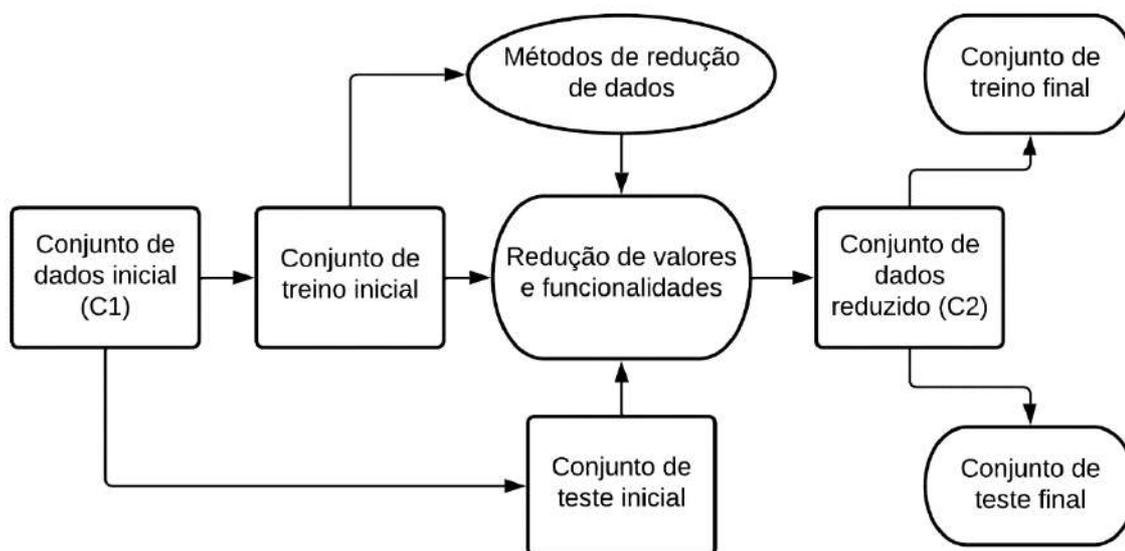
Normalmente os algoritmos aplicados para extrair padrões não conseguem trabalhar com tipos de dado complexos. Por exemplo, atributos do tipo data e hora não são analisados pela maioria dos algoritmos de mineração de dados. Assim, é preciso converter esses atributos para outros tipos de dados visando que esses algoritmos possam funcionar. Nesses atributos do tipo data e hora, por exemplo, a melhor escolha é a conversão para o tipo inteiro. Isso pode ser realizado através do cálculo da diferença entre os valores das datas do atributo em questão e uma data fixa (Escovedo; Koshiyama, 2020).

H) *REDUÇÃO DE DADOS*

As vantagens da utilização de grandes conjuntos de dados para treinamento e teste são inúmeras, porém esse conjunto pode tornar-se excessivamente grande e essas dimensões podem exceder a capacidade ou o tempo de processamento das ferramentas de predição e, assim, a produção de resultados pode apresentar um custo excessivo (Foster *et al.*, 2020). Tendo isso em vista, a técnica de redução de dados apresenta significativa importância uma vez que sua aplicação possibilita medir e

prever o custo de processamento sobre uma determinada base. A Figura 24 ilustra um modelo de redução de dados.

Figura 24 - Modelo de redução de dados.



Fonte: Foster *et al.* (2020)

Observa-se na Figura 24, que dado um determinado conjunto de dados inicial (C1), esses dados são reduzidos em quantidade de valores, atributos e/ou funcionalidades e, conseqüentemente, um novo conjunto de dados reduzido é produzido (C2). Enquanto as dimensões do conjunto de dados inicial estiverem dentro dos limites aceitáveis, não será necessário aplicar as técnicas de redução de dados. Uma vez que o conjunto inicial de dados for reduzido para um novo conjunto (C2), os dados podem ser divididos para os casos de treino e testes. Os casos de testes são fundamentais para a avaliação dos resultados. Por fim, observa-se que as escolhas e as interpretações não podem ser efetuadas sem estimativas apuradas da performance da predição (Sharda *et al.*, 2019).

Ressalta-se que quando se trata de simplificação dos dados, a redução da dimensão é fundamental, todavia esse descarte de dados não pode prejudicar a qualidade dos resultados. Em seus estudos, Kelleher e Tierney (2018), apresentam uma abordagem de redução de dados onde o uso de métodos de amostragem e características de populações são estimadas de uma forma eficiente, com mínima distorção. Para os autores, as principais técnicas de amostragem são:

- Amostragem Aleatória Simples: abordagem direta onde em um determinado conjunto de dados, todas as amostras possuem a mesma chance de serem selecionadas. Dessa forma, mesmo que seja possível obter um conjunto atípico, probabilisticamente quanto maior o tamanho da amostra, maior será a representatividade do conjunto.
- Amostragem Aleatória Estratificada: divide-se o conjunto de dados em grupos e aplica-se a técnica de amostragem aleatória simples em cada grupo, formando assim a amostragem final por meio dos agrupamentos dos subconjuntos.
- Amostragem em *Cluster*: é uma técnica em que os elementos da população são selecionados aleatoriamente individualmente e um de cada vez, de forma natural por grupos (*clusters*), sendo utilizada quando não é possível criar a amostragem de uma população alvo devido à dispersão e ao alto custo para recolher e organizar os dados.
- Amostragem Sistemática: quando indivíduos de um conjunto de dados são numerados e há a escolha aleatória de um membro da população entre 1 e n e, depois, é incluindo cada $n - \text{ésimo}$ membro em uma determinada amostra.
- Amostragem em Duas Fases: técnica utilizada quando necessita-se organizar a amostra com base em valores de uma ou mais variáveis, porém desconhece-se a variação dessas variáveis na população.

I) ARMAZÉM DE DADOS

Atualmente os bancos de dados são imprescindíveis, independentemente do tipo de organização, porém, a análise e a compreensão dos dados armazenados sempre foi um desafio para os gestores uma vez que o volume de dados é grande, de tipos diferentes e não centralizados (Machado, 2013).

A abordagem de armazenamento de dados adotada pelas organizações é comumente tradicional, centrada na realização de transações diárias em diversos sistemas. Nestes casos, os bancos de dados frequentemente são concebidos apenas para a finalidade de armazenamento dos dados, carecendo de recursos inteligentes para análise e interpretação das informações contidas (Gorelik, 2019).

Por outro lado, existe uma segunda abordagem aplicada ao armazenamento de dados, conhecida como Armazém de Dados. Essa abordagem é concebida com o propósito de fornecer suporte para tomadas de decisões estratégicas, sendo construída sobre a infraestrutura de bancos de dados tradicionais das organizações. De acordo com Machado (2013), as características fundamentais que regem a estruturação dos Armazéns de Dados são:

- Dependência de Tempo: apresenta dados armazenadas ao longo do tempo, relacionando a informação com o período em que foi inserido. Esse aspecto está diretamente relacionado com mineração de dados, pois permite que a informação seja pesquisada e utilizada por períodos.
- Não Volátil: os dados não são atualizados e sim utilizados para consultas, ou seja, as informações são carregadas de outras fontes. Em geral, ao serem encaminhados para um Armazém de Dados, os dados operacionais são previamente processados para garantir que dados de diferentes fontes e formatos possuam as mesmas definições e obedeçam às mesmas regras.
- Orientação ao conteúdo: a organização é voltada para que as informações sejam agrupadas de acordo com indicadores como clientes, produtos etc.
- Integração: o conteúdo armazenado é definido a fim de que as informações possam ser utilizadas para outras aplicações da organização.

O Armazém de Dados, conforme descrito, representa uma estrutura crucial na gestão e análise de dados organizacionais. Ele facilita a extração, tratamento e agregação de dados provenientes de diversas fontes, muitas vezes apresentados em formato de cubo multidimensional, o que permite uma rápida agregação e detalhamento das análises. Além disso, oferece recursos visuais informativos que transcendem as capacidades dos bancos de dados convencionais. Um dos diferenciais desses modelos é sua capacidade de lidar com dados temporais, possibilitando a geração de informações históricas e integradas, o que auxilia significativamente o processo de tomada de decisões (Kimball; Ross, 2013).

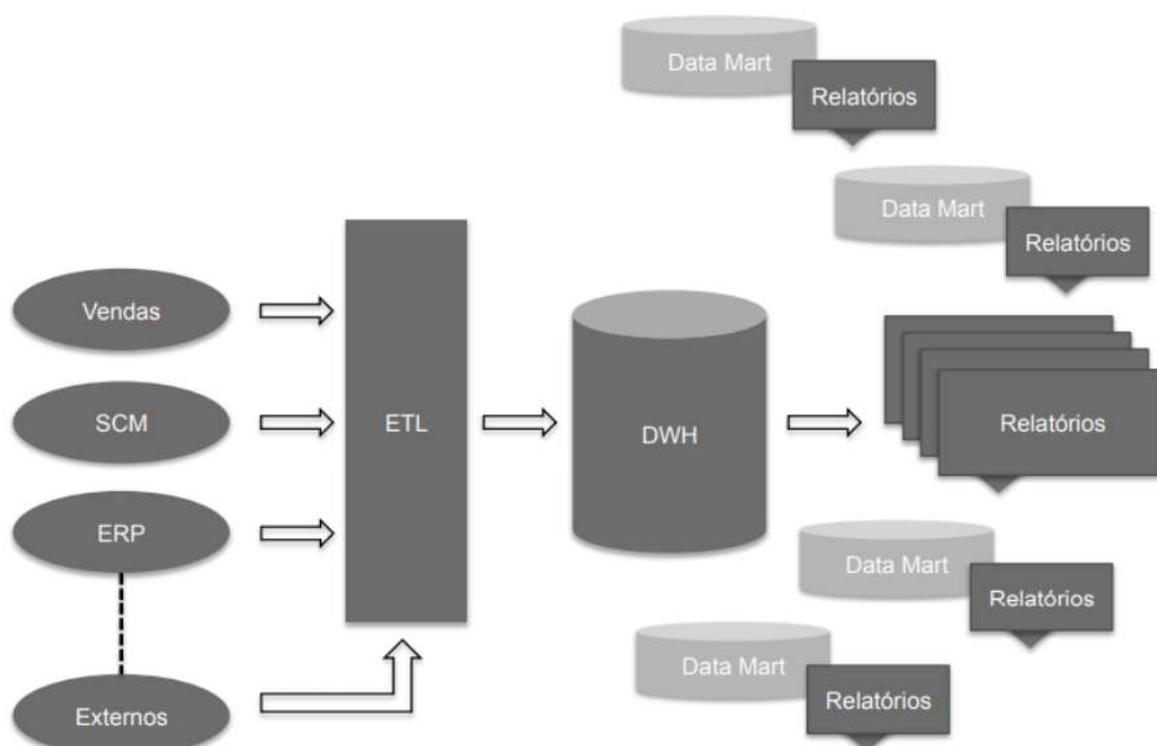
Em complemento, o *Data Mart* emerge como uma componente específica do Armazém de Dados, concentrando informações pertinentes a determinados departamentos, áreas de negócio ou temas dentro da organização. Como parte integrante do Armazém de Dados, ele representa um subconjunto direcionado e

segmentado dos dados globais da empresa. Enquanto o Armazém de Dados é utilizado para análises globais e estratégicas, o *Data Mart* possui um foco mais específico, atendendo às demandas de análise de um grupo particular de usuários ou de uma área de negócio específica (Gorelik, 2019).

Os *Data Marts* são desenvolvidos com o propósito de atender às necessidades analíticas de um conjunto específico de usuários, proporcionando acesso ágil e eficiente a informações relevantes para suas atividades e processos decisórios. Normalmente, são construídos utilizando estruturas dimensionais, como os esquemas estrela ou de floco de neve, que facilitam a navegação e compreensão dos dados (Machado, 2013). Essa integração entre o Armazém de Dados e os *Data Marts* representa um elemento fundamental na estratégia de gestão de dados e na capacidade de análise das organizações modernas.

A Figura 25 ilustra a Infraestrutura Genérica de um Armazém de Dados, abrangendo os *Data Marts*, que constituem subconjuntos do Armazém de Dados acessados por áreas específicas de negócios. Esta representação visual evidencia como um Armazém de Dados simplifica os sistemas de suporte à decisão e deve servir como o ponto de verdade único sobre os dados organizacionais (Machado, 2013).

Figura 25 - Infraestrutura genérica de um armazém de dados

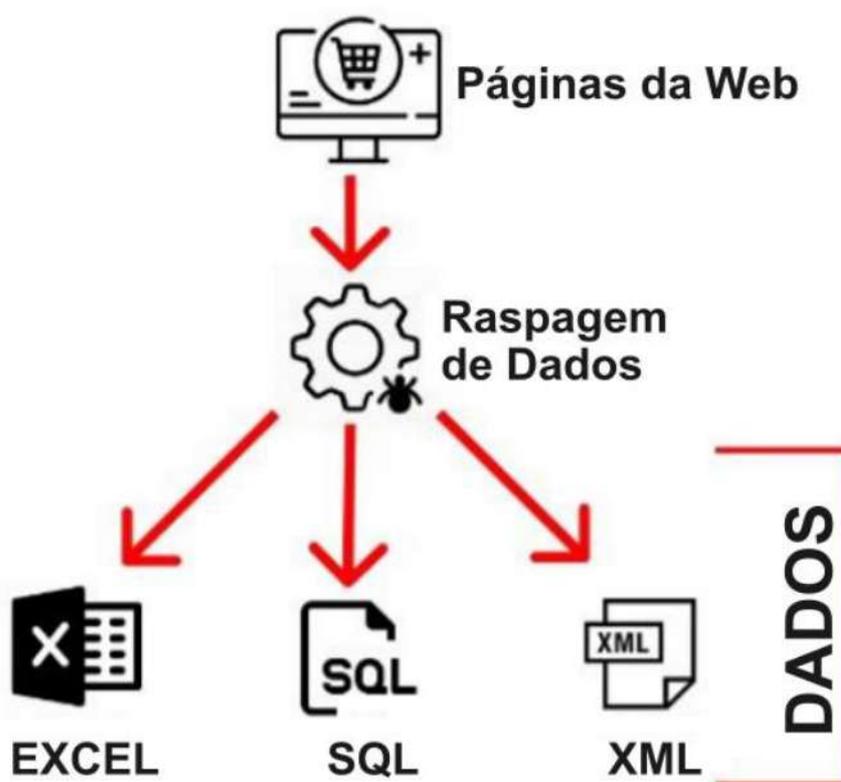


Fonte: Kimball; Ross (2013)

4.4. RASPAGEM DE DADOS

A raspagem de dados, uma subcategoria da mineração de dados, constitui uma técnica essencial para a extração de informações de páginas *web*, visando sua posterior armazenagem e análise (Kimball; Ross, 2013). Essa prática, variando desde métodos manuais até abordagens automatizadas, proporciona uma eficiente e poderosa capacidade de coletar grandes volumes de dados em curtos períodos. Destaca-se pela habilidade de extrair dados não estruturados de fontes *web* e convertê-los em formatos estruturados, aptos para armazenamento e análise posterior (Watson, 2009). A versatilidade e eficácia desta técnica são evidenciadas na Figura 26, que ilustra o processo de raspagem de dados em operação.

Figura 26 - Visão geral da raspagem de dados



Fonte: Watson (2009)

A extração de dados é um processo essencial na análise e manipulação de informações disponíveis na Internet. Este procedimento pode ser estruturado em distintos níveis, sendo dois deles particularmente relevantes: o nível de raspagem semântico e o nível de raspagem sintática. O nível semântico define o mapeamento

entre os dados da *web* e os recursos da *web* semântica e o nível sintático define as tecnologias necessárias para extrair dados reais das páginas da *web* (Vo, 2019).

As técnicas de raspagem de dados podem ser categorizadas em várias abordagens distintas, cada uma oferecendo uma maneira particular de extrair informações de fontes *online*. De acordo com Broucke e Baesens (2018), abaixo serão destacadas as principais formas utilizadas para efetuar a coleta de dados:

- Cópia e Cola Manual: representa o método mais rudimentar de extração de dados, sendo inteiramente conduzida por intervenção humana. Embora ofereça controle preciso sobre as informações a serem extraídas, sua eficácia é limitada a fontes de informação menores, tornando-se impraticável para grandes conjuntos de dados ou múltiplas fontes. Além disso, sua natureza manual resulta em um processo demorado e repetitivo, especialmente ao lidar com volumes significativos de dados. No entanto, essa abordagem pode ser útil em casos em que medidas de prevenção de raspagem de dados são implementadas em uma página da *web* específica.
- Busca de Texto e Expressão Regular: a técnica de busca de texto e expressão regular, do inglês *grep*, inspirada no comando UNIX 'grep', permite a extração de informações de uma variedade de fontes de texto, incluindo páginas da *web*. Esta abordagem baseia-se na pesquisa por expressões regulares dentro da fonte de dados, retornando linhas correspondentes aos padrões especificados. Geralmente realizada através de linha de comando, essa técnica pode ser automatizada e integrada a linguagens de programação como Python, R e Perl.
- Programação do Protocolo de Transferência de Hipertexto (HTTP): permite a extração de informações por meio do envio de solicitações GET. Ao solicitar uma página da *web*, é possível extrair seu conteúdo para posterior análise. Essa abordagem é adequada para páginas estáticas e dinâmicas, e pode ser implementada utilizando linguagens de programação como Java e Python.
- Análise de Marcação de Hipertexto (HTML): é especialmente concebida para páginas da *web* construídas utilizando essa linguagem de marcação. Esta abordagem explora o layout e a estrutura da página para extrair informações relevantes, concentrando-se principalmente em elementos HTML aninhados. A análise de HTML pode ser realizada utilizando linguagens de programação

como Python e JavaScript, além de ferramentas como Hyper Text Query Language (HTQL) e XQuery.

- Modelo de Objeto de Documento (DOM): permite a identificação e extração de informações de páginas da *web* através de sua representação estruturada. Linguagens como Python e XPath são frequentemente empregadas para esse fim, permitindo a navegação e seleção de nós contendo as informações desejadas. Embora exigindo maior consumo de memória em comparação com outras técnicas, a análise do DOM, do inglês *document object model*, tende a ser mais rápida e eficiente.
- Reconhecimento de Anotação Semântica: envolve a utilização de metadados para localizar e organizar trechos específicos de dados. Essas anotações são armazenadas separadamente das páginas da *web*, permitindo o acesso a um esquema de dados estruturado antes da raspagem. Essa técnica enriquece o conteúdo das páginas, estabelecendo uma conexão entre o texto e suas descrições semânticas.
- Analisadores Visuais de Páginas da *Web*: emprega técnicas avançadas de programação para extrair informações de páginas da *web* de forma semelhante à interpretação humana. Essa abordagem é especialmente útil para páginas não estruturadas, oferecendo uma maneira eficaz de extrair dados de forma automatizada.
- Interface de Programação de Aplicativos (API): as APIs, oferecidas pelos proprietários dos dados, fornecem acesso estruturado aos dados, destinado a ser consumido por programas de computador. No entanto, nem sempre uma API está disponível ou é preferível, dependendo das restrições impostas pela página da *web* em questão, como pagamento pelo uso ou restrições de dias e horários para a obtenção dos dados.

O emprego de técnicas automatizadas para a extração de dados demanda uma observação criteriosa da aplicação de novas metodologias e tendências emergentes. Nesse contexto, destaca-se a relevância do emprego conjunto de técnicas como o rastreamento da *web* junto a raspagem de dados automatizada (Mitchell, 2018).

O rastreamento da *web*, do inglês *Web Crawler*, representa uma abordagem amplamente adotada para a coleta de informações disponíveis na vastidão da Internet, implicando a navegação sistemática através de endereços da *web* mediante

a utilização de *hiperlinks*. Esse conceito implica na capacidade intrínseca de um programa ou algoritmo de percorrer autonomamente páginas da *web*, buscando e extraindo os dados pertinentes conforme critérios pré-estabelecidos (Broucke; Baesens, 2018).

Os rastreadores da *web* são ferramentas que podem ser empregadas com a finalidade de explorar conteúdos disponíveis em determinado endereço da *web*, embora também possam ser utilizados sem um objetivo ou propósito estritamente definido. Um exemplo de aplicação do rastreamento da *web* é a exploração de uma loja virtual em busca de produtos, onde o rastreador é capaz de operar dentro dos limites claros estabelecidos pelo domínio da loja e coletar informações contidas nas páginas dos produtos (Watson, 2009).

É importante ressaltar que, frequentemente, o rastreamento da *web* não envolve a extração direta de dados de um endereço da *web*, sendo esta atividade realizada posteriormente por meio do uso de técnicas de raspagem de dados. O objetivo principal do rastreamento da *web* é a exploração da vasta rede de informações disponíveis na Internet como um todo ou dentro de um domínio específico. Além disso, os rastreadores da *web* são fundamentais para o funcionamento dos motores de busca, os quais dependem do rastreamento para identificar e indexar as páginas da *web* disponíveis na Internet (Mitchell, 2018).

O emprego de rastreadores da *web* suscita duas questões de importância significativa: em primeiro lugar, a necessidade de desenvolver uma estratégia de rastreamento eficaz, a qual engloba não apenas a formulação de algoritmos adequados para a navegação por novos endereços da *web*, mas também a implementação de mecanismos inteligentes voltados para a otimização do processo de novos e repetitivos rastreamentos (Vo, 2019).

Em segundo lugar, é crucial otimizar o desempenho do rastreamento em consonância com o *hardware* disponível. Dada a natureza computacionalmente intensiva do rastreamento, é imprescindível que o sistema seja capaz de adaptar-se a diferentes cenários, incluindo falhas de *hardware*, problemas de servidor e erros de análise, ao mesmo tempo em que maximiza a eficiência operacional para garantir a melhor utilização dos recursos disponíveis, tais como memória, capacidade de processamento e largura de banda (Mitchell, 2018).

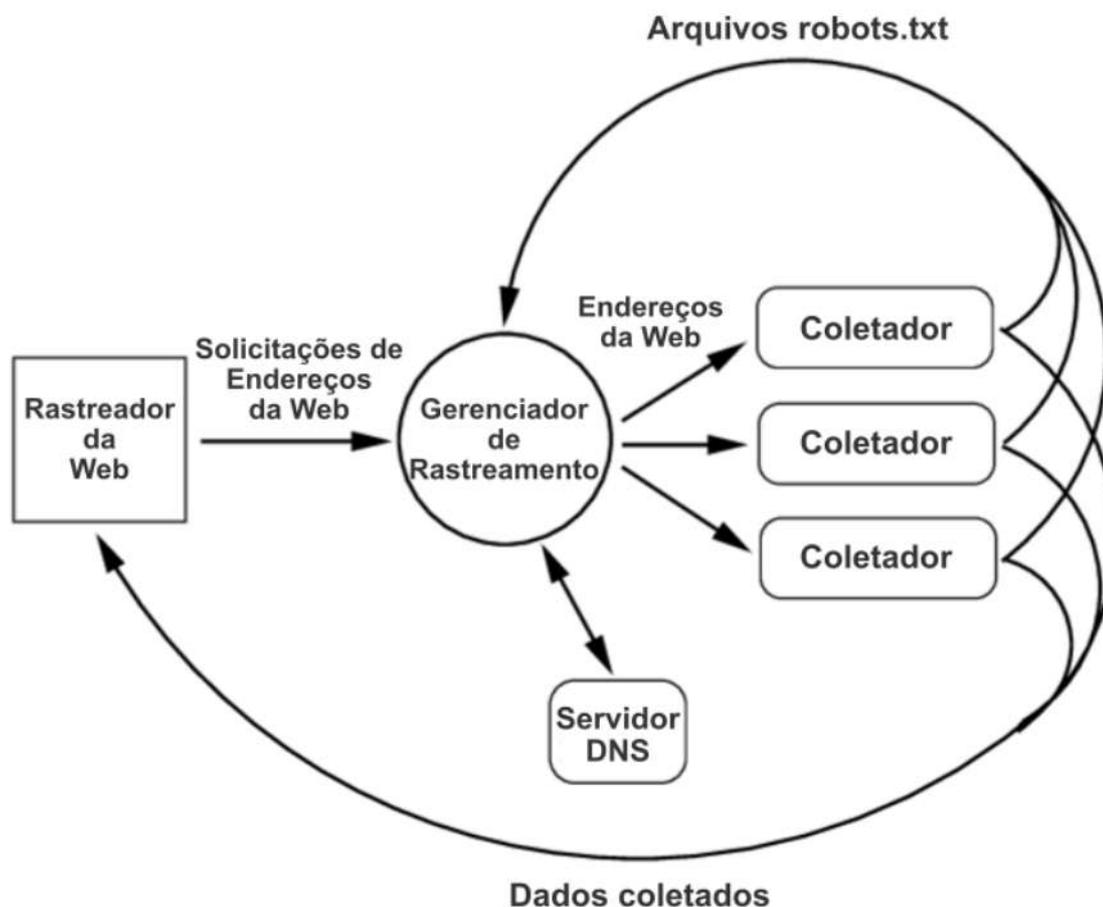
Um sistema de rastreamento da *web* é composto por componentes distintos, notadamente um gerenciador de rastreamento e coletadores. O gerenciador de

rastreamento desempenha um papel crucial ao direcionar solicitações de endereços da *web* para os coletadores designados (Watson, 2009).

Por sua vez, os coletadores assumem a responsabilidade de estabelecer conexões com uma variedade de servidores de *web* e monitorar ativamente essas conexões para a chegada de dados pertinentes. Além disso, os coletadores são incumbidos de recuperar o arquivo *robots.txt* associado a página da *web* em questão, o qual delinea as diretrizes a serem seguidas pelos agentes de rastreamento ao explorarem a estrutura da página da *web* (Mitchell, 2018).

Dessa maneira, o gerenciador de rastreamento também se encarrega de aderir às normativas estipuladas no arquivo *robots.txt*, assegurando uma operação de rastreamento consonante com as políticas estabelecidas pela página da *web* alvo (Shkapenyuk; Suel, 2002). A Figura 27 delinea um modelo de sistema de rastreamento da *web*.

Figura 27 - Modelo de um sistema de rastreamento da *web*



Fonte: Shkapenyuk; Suel (2002)

O presente estudo procedeu com a aquisição de dados por meio da aplicação de técnicas automatizadas de raspagem de dados em conjunto com o uso de rastreadores da *web*. Foi adotada uma abordagem que combina métodos automatizados, visando a maximização da eficiência temporal, dos recursos computacionais e dos custos associados. Posteriormente, os dados foram organizados em um formato de banco de dados estruturado, utilizando uma planilha no *software* Excel, com a assistência da linguagem de programação Python.

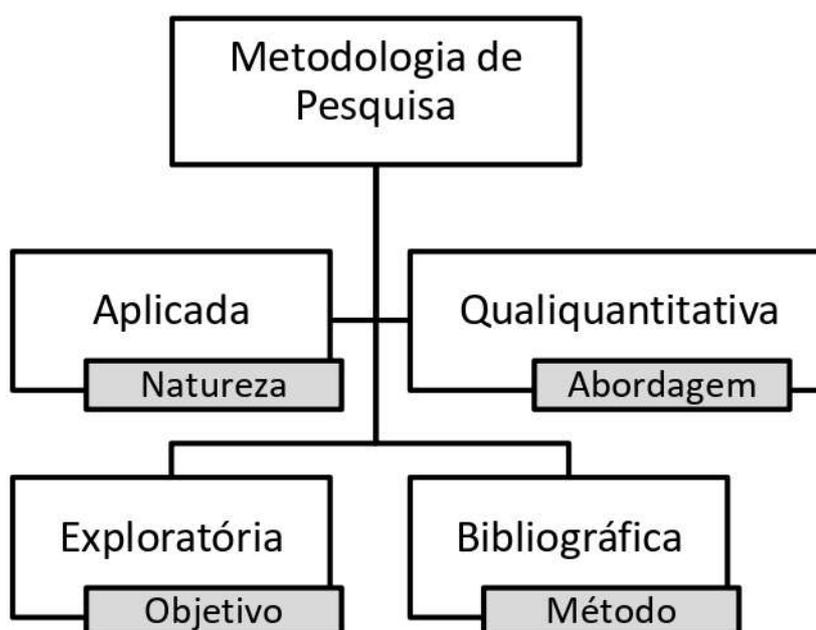
A escolha desta linguagem decorre de sua notória eficiência e velocidade na manipulação de dados, bem como da disponibilidade de uma variedade de bibliotecas poderosas voltadas para a coleta e o tratamento de dados. Cada uma dessas ferramentas apresenta vantagens distintas, que foram avaliadas com base em critérios como facilidade de aprendizado, desempenho, documentação disponível e possíveis restrições provenientes das páginas da *web*. Portanto, o processo de seleção, combinação, otimização e desenvolvimento de técnicas para a coleta, tratamento e verificação dos dados em Python exigiu uma análise criteriosa, considerando aspectos como estabilidade do *software*, frequência de atualizações da documentação e minimização de erros de execução.

5. METODOLOGIA

A metodologia aborda os procedimentos adotados para desenvolver uma pesquisa e validar todas as etapas traçadas a fim de atingir os objetivos propostos (Gerhardt; Silveira, 2009).

A metodologia da pesquisa do presente estudo está apresentada em forma de um organograma elaborado pelo autor e que é representada pela Figura 28. As suas etapas, assim como as suas características, são descritas na sequência.

Figura 28 - Organograma de metodologia de pesquisa



Fonte: Autor (2021)

Quanto ao ponto de vista da natureza desta pesquisa, a metodologia utilizada foi a pesquisa aplicada, que de acordo com Lakatos e Marconi (2003), tem como objetivo direcionar o conhecimento científico e teórico para a resolução de problemas práticos e a criação de aplicações úteis. Ela busca desenvolver soluções para desafios específicos em áreas como tecnologia, medicina, engenharia, entre outros. A pesquisa aplicada está voltada para a utilização do conhecimento existente para resolver problemas do mundo real, visando melhorar processos, produtos ou serviços. Seu foco principal é a transferência de conhecimento para o benefício da sociedade e a aplicação prática dos resultados da pesquisa.

Com relação ao ponto de vista da forma da abordagem do problema, este estudo foi uma pesquisa qualiquantitativa. Gerhardt e Silveira (2009) destacam que a pesquisa qualiquantitativa é uma abordagem que combina elementos da pesquisa qualitativa, centrada em aspectos subjetivos e exploratórios, com a pesquisa quantitativa, que utiliza dados numéricos e métodos estatísticos. Essa integração busca fornecer uma compreensão mais completa de um fenômeno, utilizando tanto análise qualitativa quanto quantitativa para obter *insights* abrangentes.

A pesquisa em questão segue uma abordagem exploratória, conforme descrita por Gil (2002), com o intuito de aprofundar a compreensão do problema em análise, visando esclarecê-lo e possibilitar a formulação de hipóteses. Destaca-se que o principal objetivo deste estudo reside no refinamento de conceitos e na descoberta de *insights*. Conforme mencionado por Andrade (2001), uma pesquisa exploratória envolve a realização de pesquisas bibliométricas, visando proporcionar uma visão ampla sobre o tema investigado ou servir como base para aprofundar os assuntos.

De acordo com a perspectiva dos métodos científicos, a presente investigação adotou uma abordagem bibliográfica, conforme delineado por Fonseca (2002). Tal abordagem implica na coleta e análise de referências teóricas previamente estudadas e disponibilizadas em formatos impressos ou eletrônicos, tais como livros, artigos científicos e páginas da *web*. Essa etapa é fundamental no início de qualquer trabalho científico, pois permite ao pesquisador familiarizar-se com o conhecimento já produzido sobre o assunto em questão. Lakatos e Marconi (2003) destacam que a pesquisa bibliográfica possibilita a análise de conceitos e publicações relacionados ao tema proposto, levando em consideração materiais elaborados por outros pesquisadores, como livros, artigos científicos, jornais e fontes eletrônicas. Essa análise desempenha um papel importante ao compreender os conceitos já divulgados (Barros; Lehfeld, 2008).

O procedimento metodológico para o desenvolvimento do estudo e obtenção dos resultados envolve a aplicação de dois instrumentos fundamentais: a mineração de dados por meio de raspagem de dados utilizando a linguagem de programação Python e o uso do Netlogo, uma linguagem de modelagem baseada em agentes para a criação de simulações. É importante destacar que tanto a raspagem de dados quanto o Netlogo são elementos essenciais para alcançar os objetivos propostos.

5.1. MATERIAIS

Nesta tese, foi empregado um sistema de *hardware* composto por um computador de 64 bits, equipado com um processador Intel Core i5-5250U de 1,60 GHz e 8,0 GB de RAM, utilizando o sistema operacional Windows 10.

Quanto aos *softwares* utilizados, a linguagem de programação Python na versão 3.9 foi adotada, juntamente com o ambiente computacional interativo Jupyter, para execução das tarefas.

A ferramenta Scrapy, versão 2.5, foi escolhida para realizar a extração eficiente de informações por meio da técnica de raspagem de dados. O Scrapy é uma biblioteca de código aberto integrada ao Python. O processo de raspagem de dados é realizado de forma não invasiva, no qual apenas as informações relevantes são extraídas das páginas da *web*, evitando sobrecarregar as páginas da *web*. Isso é possível através da implementação de código Python, que determina quais elementos são pertinentes na página de destino. Essa abordagem é desejável e evita impactos negativos nas páginas da *web*.

A etapa inicial do processo de raspagem de dados consiste na criação de um projeto Scrapy por meio do comando "`<scrapy startproject arquivo>`", onde "arquivo" é uma nomenclatura definida pelo usuário.

As páginas da *web* na Internet geralmente possuem uma estrutura composta por várias páginas, onde a página principal apresenta uma lista geral de assuntos. Durante a exploração de um assunto específico, o usuário é direcionado para páginas da *web* adicionais que oferecem informações mais detalhadas sobre o tema em questão. Portanto, ao utilizar a programação Scrapy em Python, é necessário adaptar-se às arquiteturas das páginas da *web*.

A navegação pelas páginas principais é realizada por meio de um *loop* que percorre os argumentos dos endereços da *web*. Em cada iteração desse *loop*, utiliza-se o comando `<rules>`, inserido dentro de uma classe chamada *CrawlSpider*. Esse sistema de leitura é capaz de extrair os endereços da *web* de todas as subpáginas que contêm informações mais detalhadas sobre cada assunto. Assim, as subpáginas são abertas uma por uma e as informações relevantes são coletadas.

Utilizando o comando `<xpath>`, é especificado o caminho onde as informações estão localizadas dentro de cada página HTML. Essas informações são direcionadas para campos de variáveis criadas no arquivo `Item.py`. Posteriormente, essas variáveis

são armazenadas em arquivos CSV para uso futuro.

No desenvolvimento do método de raspagem de dados, aplicado para gerar os dados de entrada das simulações, além das linguagens de programação Python e Scrapy, foram utilizadas as bibliotecas gratuitas BeautifulSoup, Python NLTK e Gensim. A biblioteca BeautifulSoup é empregada para analisar e extrair informações de documentos HTML e XML. A biblioteca Python NLTK desempenha um papel fundamental na análise de texto e processamento de linguagem natural. Ela oferece uma variedade de ferramentas para *tokenização*, *lematização* e outras técnicas de pré-processamento de texto. Por fim, a biblioteca Gensim é utilizada para modelagem de tópicos e processamento de linguagem natural avançado. Ela fornece algoritmos eficientes para criar modelos de tópicos a partir de grandes conjuntos de dados de texto.

A utilização conjunta das bibliotecas Python e Scrapy em processos de raspagem de dados representa um avanço significativo, proporcionando melhorias substanciais na eficácia e organização da coleta de dados. Cada uma dessas bibliotecas desempenha papéis distintos e complementares, o que resulta em uma sinergia que fortalece a consecução dos objetivos delineados. Enquanto o Python oferece uma ampla gama de funcionalidades de programação e manipulação de dados, o Scrapy destaca-se por sua capacidade de extrair informações de forma automatizada e estruturada a partir de diferentes fontes na *web*. A integração dessas ferramentas permite uma abordagem abrangente e eficiente na obtenção de dados.

No contexto da pesquisa em SMA, a escolha do NetLogo como ferramenta de desenvolvimento se destaca não apenas pela sua capacidade de modelar uma ampla variedade de fenômenos, mas também pela sua abordagem voltada para a representação e interação de agentes autônomos. Essa abordagem baseada em agentes é especialmente relevante para lidar com sistemas complexos, nos quais o comportamento emergente resulta das interações entre múltiplos agentes individuais (Salecker *et al.*, 2020).

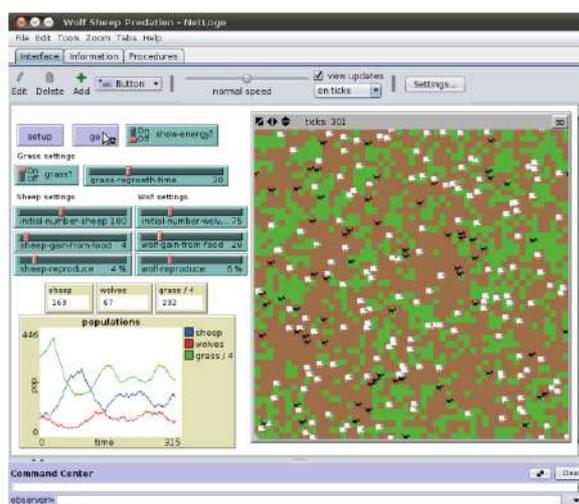
O NetLogo oferece uma série de recursos que facilitam a implementação de modelos de SMA. Sua interface gráfica intuitiva permite a criação e visualização de ambientes de simulação de forma interativa, o que facilita a compreensão do comportamento dos agentes e dos resultados da simulação. Além disso, a plataforma possui uma vasta biblioteca de modelos para uso, cobrindo uma ampla gama de domínios, desde ecologia e biologia até economia e ciências sociais, o que acelera o

processo de desenvolvimento (Musaeus; Musaeus, 2020).

Segundo Hjorth et al. (2020), outra vantagem do NetLogo é a sua linguagem de programação simplificada e de alto nível, que permite a expressão de conceitos complexos de forma clara e concisa. Além disso, a plataforma oferece suporte para testes ágeis de modelos, permitindo a rápida iteração sobre diferentes hipóteses e cenários, o que é crucial para o desenvolvimento de simulações realistas.

Em suma, o NetLogo se destaca como uma ferramenta poderosa para a implementação de modelos de SMA, fornecendo recursos avançados de modelagem, uma interface intuitiva e uma linguagem de programação amigável, que juntos contribuem significativamente para o avanço da pesquisa nesse campo em constante evolução. A interface do NetLogo é apresentada na Figura 29.

Figura 29 - Interface do NetLogo



Fonte: Salecker *et al.* (2020)

Conforme Musaeus e Musaeus (2020), essa linguagem é altamente adequada para a modelagem de sistemas complexos que evoluem ao longo do tempo, bem como para modelos que envolvem diversos tipos de indivíduos interagindo entre si e com o ambiente.

De acordo com Hjorth *et al.* (2020), o NetLogo oferece a oportunidade de explorar a relação entre as interações locais dos indivíduos e os padrões macro emergentes resultantes dessas interações. Segundo Nogare e Chitnis (2020), a interface do NetLogo:

- Permite experimentar simulações.

- Possibilita criar modelos para testar hipóteses sobre sistemas diversos.
- Apresenta uma grande biblioteca contendo simulações em ciências naturais e/ou sociais, que podem ser utilizadas e modificadas.

Embora a linguagem NetLogo seja conhecida por sua abordagem acessível, sua capacidade de modelagem não conhece limites. Ela permite a criação de modelos altamente sofisticados, especialmente voltados para sistemas que envolvem agentes autônomos interagindo entre si (Jaxa-Rozen; Kwakkel; Bloemendal, 2020).

Essa flexibilidade possibilita a modelagem de uma ampla variedade de fenômenos complexos, como o comportamento de agentes no mercado de ações, a previsão da propagação de epidemias, estratégias de evasão fiscal, evacuação de pessoas em emergências, o crescimento e declínio de civilizações, táticas de guerrilha, assistência a civis, entre muitos outros exemplos.

O uso do NetLogo se destaca pelo emprego de sistemas baseados em agentes, os quais representam um paradigma computacional de significativa relevância, amplamente empregado na condução de simulações em contextos complexos (Zappe, 2012). A Figura 30 representa um agente, suas características e interações com ambiente.



Os agentes interagem entre si e com o meio ambiente satisfazendo a um conjunto de regras previamente programadas. Os agentes aceitam novas

circunstâncias e apresentam a capacidade de aprenderem e adaptarem o seu comportamento baseado na experiência. Além disso, os agentes podem ainda, ter regras para mudar as suas próprias regras de comportamento (Musaeus; Musaeus, 2020). O NetLogo opera com o conceito de três tipos de agentes, denominados de acordo com os termos utilizados na plataforma:

- *Turtles*: estes são agentes que se deslocam pelo mundo, que é bidimensional e subdividido em *patches*.
- *Patches*: cada *patch* representa uma parcela de terreno onde as *turtles* podem se movimentar.
- *Observer*: este agente não possui uma localização específica e pode ser conceituado como aquele que observa o mundo das *turtles* e *patches*.

5.2. MÉTODOS

O desenvolvimento da técnica proposta nesta tese se inicia com a concepção de uma abordagem de raspagem de dados, destinada a extrair informações confiáveis de fontes *online* na Internet e organizá-las em tópicos específicos. A técnica visa gerar conjuntos de dados que servirão como entrada para a simulação de cenários complexos, baseada em agentes.

O procedimento delineado consiste em quatro etapas principais, a saber:

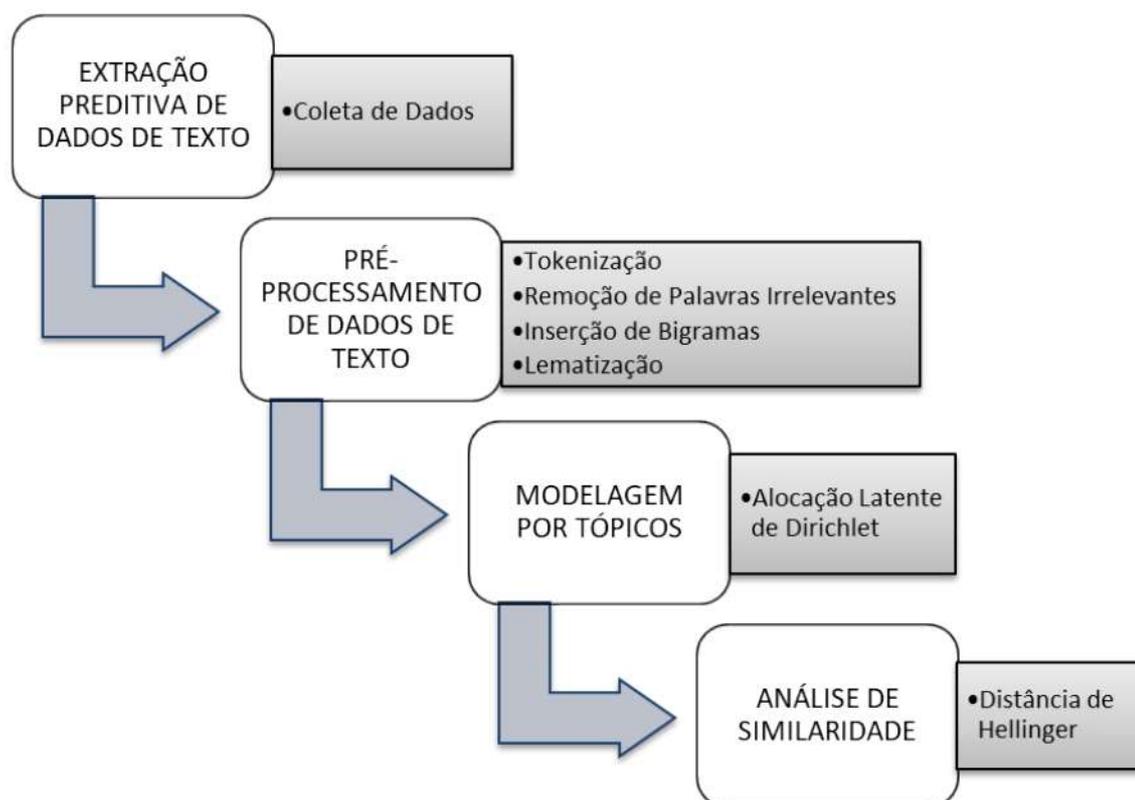
1. Extração preditiva de dados textuais: inicialmente ocorre a extração dos dados textuais relevantes, visando identificar padrões e tendências que possam ser úteis para previsões ou análises posteriores. Esta etapa envolve a coleta e o processamento dos dados textuais de fontes diversas.
2. Pré-processamento dos dados textuais: após a extração dos dados textuais, é realizada uma etapa de pré-processamento para limpar e preparar os dados para análise. Isso pode incluir a remoção de ruídos, a normalização de texto e a identificação de termos-chave.
3. Aplicação do método de modelagem por tópicos utilizando alocação latente de Dirichlet (LDA): Nesta fase, aplica-se o método de modelagem por tópicos utilizando a LDA para identificar padrões de associação de palavras nos dados textuais. A LDA é um modelo estatístico utilizado para analisar grandes

conjuntos de documentos e identificar tópicos latentes subjacentes.

4. Análise de similaridade entre os diferentes tópicos por meio da distância de Hellinger: Por fim, realiza-se uma análise de similaridade entre os tópicos identificados, utilizando a medida de distância de Hellinger. Essa análise permite avaliar quão semelhantes ou distintos são os diferentes tópicos identificados pelo modelo da LDA.

Essas etapas proporcionam um arcabouço metodológico abrangente para a análise de dados textuais, permitindo a identificação de padrões e tendências significativas nos dados que podem subsidiar análises e tomadas de decisão fundamentadas. Após a validação e confirmação da funcionalidade do processo desenvolvido, procedeu-se a aplicação da simulação em estudos de caso, tendo o Brasil como cenário. O primeiro estudo de caso explorou os determinantes que influenciam a adoção de sistemas solares, enquanto o segundo analisou as principais causas subjacentes à redução do consumo de energia elétrica. A Figura 31 ilustra o processo desenvolvido para coletar informações.

Figura 31 - Processo da coleta de dados



Fonte: Autor (2021)

Esse processo representa uma significativa evolução em relação às práticas convencionais de coleta de dados por diversos motivos. Em primeiro lugar, destaca-se a capacidade de realizar a coleta rápida de dados atuais e não estruturados em páginas da *web* específicas, priorizando aquelas que oferecem informações confiáveis e relevantes. Em seguida, é enfatizada a etapa de aprimoramento, limpeza e refinamento dos dados coletados, por meio da aplicação de técnicas avançadas de pré-processamento de texto, resultando na obtenção de dados estruturados que refletem de forma precisa a realidade e a atualidade dos temas abordados.

Na terceira fase do processo, ocorre a modelagem por tópicos e a atribuição de pesos às palavras-chave, o que permite identificar a relevância de cada termo em relação ao assunto tratado. As palavras-chave mais frequentes indicam sua importância no contexto, enquanto a distribuição de pesos irá proporcionar uma representação mais precisa e detalhada do cenário abordado. Por fim, a quarta etapa envolve uma análise dos tópicos gerados, a relevância das palavras-chave em cada um e a avaliação da coerência e inter-relação entre os termos dentro de cada tópico.

Essa técnica, ao combinar essas quatro tarefas, promove uma otimização abrangente nos processos de coleta, preparação e análise de dados, resultando em uma eficiência aprimorada, redução no uso de recursos computacionais, tempo e custos. Além disso, ao utilizar dados reais e atualizados, as simulações resultantes são mais precisas e menos propensas a erros, tornando-se aplicáveis a diversas áreas do conhecimento humano.

Portanto, ao integrar essas etapas em um processo coeso de otimização, essa abordagem se destaca como uma alternativa inovadora e diferenciada em relação aos métodos de coleta de dados tradicionais.

A) *EXTRAÇÃO PREDITIVA DE DADOS DE TEXTO*

O presente estudo adota a extração preditiva de um conjunto de dados de texto, que consiste em analisar uma base de dados para verificar ou refutar noções, ideias ou relacionamentos pré-concebidos, ou seja, realizar testes de hipóteses. Para esse fim, serão coletadas informações provenientes de páginas da *web* governamentais, associações e redes sociais na Internet. Os dados passarão por um processo de seleção, descartando-se informações curtas. As *tags*, que servem como etiquetas

para organizar informações e agrupar conteúdos relacionados, serão mantidas para refletir tendências mais abrangentes sobre o tema em questão.

A coleta de dados em larga escala na Internet será conduzida com o objetivo de obter conjuntos de dados abrangentes, mesmo que ocasionalmente incluam informações irrelevantes. Esses conjuntos de dados serão utilizados para modelagem de tópicos, a fim de reduzir a complexidade e identificar claramente as principais tendências abordadas nos estudos de caso.

No caso das páginas da *web* disponíveis na Internet, as informações são codificadas individualmente seguindo o modelo HTML comum. A estrutura desses documentos HTML, como as *tags*, será examinada para implementar um programa em Python baseado no BeautifulSoup, que navegará nesses modelos e extrairá as informações textuais. Os dados recuperados serão armazenados em arquivos CSV.

B) PRÉ-PROCESSAMENTO DE DADOS DE TEXTO

Nesta fase, serão realizados os preparativos dos dados, que envolvem uma adequada organização dos dados para permitir uma análise eficiente de texto. Essa preparação inclui a eliminação de caracteres desnecessários, como caracteres não alfabéticos e espaços em branco, bem como a correção de imprecisões que possam ter surgido durante o processo de extração de dados da Internet.

Para ser mais específico, a preparação dos dados de texto abrange a normalização do texto, incluindo etapas como *tokenização*, remoção de palavras irrelevantes, inserção de bigramas e *lematização*. No processamento de linguagem natural, a unidade de análise textual é o *token*, que representa uma sequência de caracteres que compõem o texto. A *tokenização* consiste em segmentar os dados de texto em palavras e pontuação.

Para reduzir o ruído nos dados textuais, será utilizada a biblioteca Python NLTK para remover a pontuação e as palavras contidas em uma lista previamente compilada de centenas de palavras irrelevantes, como "o", "de", "com", entre outras.

Em seguida, será conduzida a fase de identificação de bigramas, os quais constituem sequências de dois elementos adjacentes em uma série de *tokens*. Normalmente, esses elementos compreendem letras, sílabas ou palavras. Ademais, será elaborada uma função destinada a reconhecer bigramas que denotam pares de

palavras frequentemente correlacionadas em publicações.

Posteriormente, será realizada a *lematização*, que consiste em reduzir as palavras às suas formas de raiz correspondentes. A *lematização* é útil para analisar palavras em contextos sem levar em consideração suas flexões. Isso evita a necessidade de pesquisar todas as formas de uma palavra para encontrá-la em um texto. Por exemplo, em uma busca por palavras-chave, como "gato", espera-se que o resultado inclua todas as formas relacionadas, como "gata", "gatos" e "gatas", e não apenas a forma masculina singular.

A fim de preparar os dados de entrada para a próxima fase, será também aplicada a biblioteca Gensim do Python. Ela será utilizada para criar um dicionário que mapeia palavras distintas em IDs numéricos únicos e uma lista que representa cada publicação como uma combinação de IDs de palavras e suas respectivas frequências.

A lista de "word_id" é uma estrutura de dados que contém os IDs numéricos únicos atribuídos a palavras distintas em um determinado contexto ou conjunto de documentos. Cada palavra é mapeada para um ID específico, permitindo uma representação numérica das palavras em um modelo ou análise.

Por outro lado, a lista de "word_frequency" representa a contagem de frequência de cada palavra em um texto ou conjunto de documentos. Essa contagem indica quantas vezes uma palavra específica ocorre em um determinado contexto. A frequência das palavras pode ser usada como indicador de importância ou relevância em uma análise de texto ou modelagem de tópicos.

Essas duas listas, "word_id" e "word_frequency", são usadas em conjunto para representar o documento como uma combinação de IDs numéricos únicos e suas respectivas contagens de frequência para cada palavra presente no texto.

C) *MODELAGEM POR TÓPICOS*

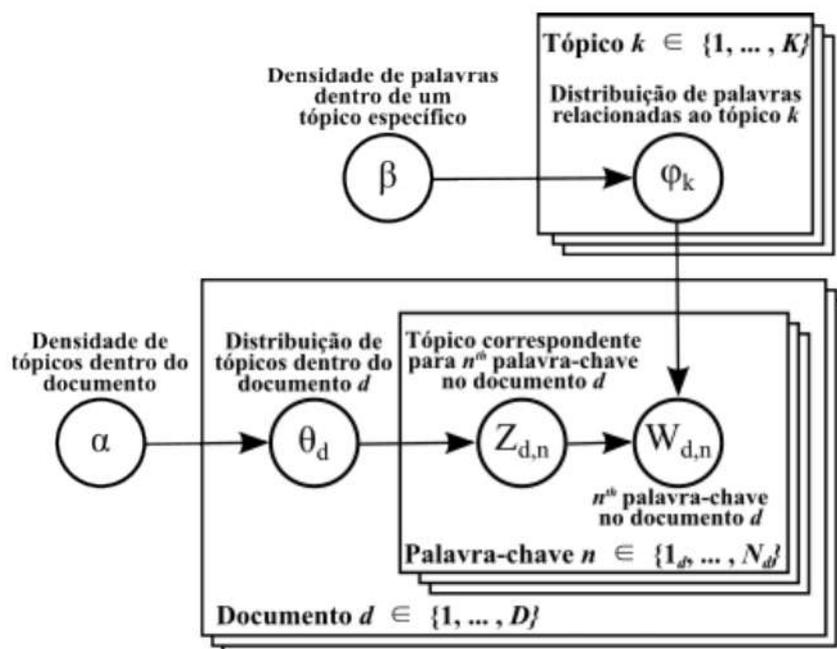
A linguagem constitui um corpo de informações não estruturadas, criadas por seres humanos com a finalidade de serem compreendidas por outros indivíduos. Os dados textuais não são gerados de forma aleatória; eles seguem propriedades linguísticas que facilitam sua compreensão por humanos e processamento por computadores. Nesse contexto, a modelagem de tópicos oferece uma abordagem robusta para extrair e resumir o conteúdo de grandes volumes de texto. Esses

modelos são construídos com base na análise semântica latente, proporcionando uma visão significativa dos dados textuais.

Ao contrário dos métodos convencionais de extração de palavras-chave, que oferecem apenas uma cobertura parcial do texto devido a conjuntos predefinidos, a modelagem de tópicos automatiza a análise de conteúdo e reduz as dimensões do texto sem comprometer a descoberta de tópicos. Isso é possível graças às propriedades sensíveis ao contexto incorporadas nesse método.

Nos últimos anos, a técnica conhecida como LDA tem ganhado popularidade no campo de gerenciamento de projetos e pesquisa em engenharia. Essa abordagem, introduzida por Blei, Nig e Jordan (2003), consiste em um método de aprendizado de máquina não supervisionado capaz de identificar os principais tópicos em um conjunto de documentos não rotulados. A Figura 32 mostra uma representação gráfica do método da LDA.

Figura 32 - Representação gráfica do método de LDA

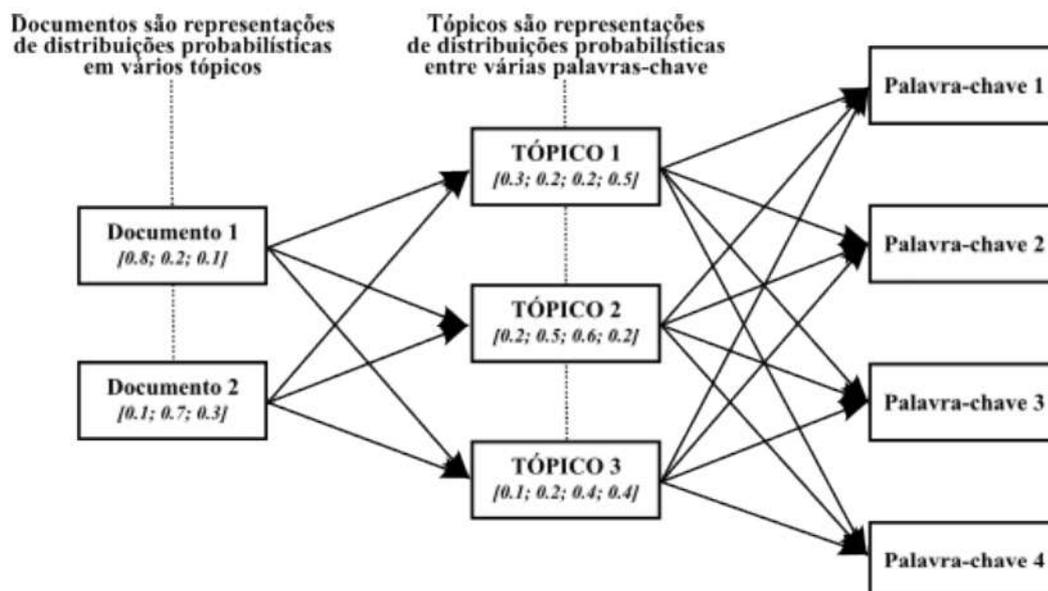


Fonte: Blei; Nig; Jordan (2003)

Na LDA, cada documento d , é considerado como uma distribuição de probabilidade θ_d , sobre um conjunto de tópicos K e cada tópico, $k \in \{1, \dots, K\}$ é, por sua vez, representado como uma distribuição de probabilidade, φ_k , sobre palavras-chave do vocabulário (Blei; Nig; Jordan, 2003). A notação matemática usada neste artigo é indicada na Figura 32; por exemplo, θ denota uma matriz com linhas definidas

por documentos e colunas definidas por tópicos, e $\theta_{d,k}$ representa a probabilidade do tópico k ocorrer no documento d . Da mesma forma, φ é uma matriz com linhas definidas por tópicos e colunas definidas por palavras-chave. Uma representação simplificada do método de LDA é mostrada na Figura 33.

Figura 33 - Representação simplificada do LDA



Fonte: Blei; Nig; Jordan (2003)

No modelo LDA, os documentos e as palavras-chave são variáveis observáveis, enquanto os tópicos são variáveis latentes que são inferidas estatisticamente para reduzir as dimensões do problema. Em teoria, o algoritmo LDA conecta cada documento d às suas N palavras-chave constitutivas $w_{d,n}$ por um encadeamento e agrupa probabilisticamente encadeamentos (ou seja, palavras-chave que ocorrem repetidamente nos mesmos documentos) em tópicos, usando parâmetros α e β e a composta Dirichlet multinomial. Assim, a distribuição de probabilidade $\varphi_{k,w}$ para uma dada palavra-chave w cair no tópico k é calculada junto com a distribuição $\theta_{d,k}$ para um tópico k estar em um documento d . Com base nessas distribuições, o algoritmo pode introduzir uma camada oculta de tópicos entre palavras-chave e documentos. A qualidade da inferência estatística induzida por K tópicos, pode ser quantificada por meio de medidas de coerência de tópicos pela biblioteca Gensim.

A biblioteca Gensim, desenvolvida para ser utilizada em Python, foi usada para

implementar o algoritmo LDA em conjuntos de documentos de texto previamente processados. Após a implementação, os tópicos principais foram extraídos automaticamente. Neste estudo, dois indicadores de coerência são considerados: C_{UMass} e C_V . Para medir a coerência interna de um determinado tópico t , $C_{UMass}(t; W^{(t)})$ calcula a média da soma da função logarítmica de pontuação emparelhada, conforme descrito na Equação 3.

$$C_{UMass}(k; W^{(k)}) = \frac{2}{N(N-1)} \sum_{i < j} \log \left(\frac{D(w_i^{(k)}, w_j^{(k)})}{D(w_i^{(k)})} + \varepsilon \right) \quad (3)$$

Na Equação 3, $D(w_i)$ é a contagem de documentos que contêm a palavra-chave w_i ; $D(w_i, w_j)$ é a contagem de documentos que contêm as palavras-chave w_i e w_j ; e $W^{(k)} = (w_1^{(k)}, \dots, w_N^{(k)})$ é a lista das N palavras-chave mais prováveis do tópico k . Além disso, $\varepsilon = 1$ é usado para evitar o logaritmo de 0. Devido à função logarítmica, C_{UMass} retorna valores negativos, com valores mais próximos de 0 referindo-se a tópicos mais compreensíveis e interpretáveis pelos seres humanos.

A C_{UMass} do modelo de tópico calcula a média das pontuações de coerência para todos os tópicos na Equação 4.

$$C_{UMass} = \frac{1}{K} \sum_{k=1}^K (C_{UMass}(k; W^{(k)}) + \varepsilon) \quad (4)$$

A Equação 4 oferece um cálculo da métrica C_{UMass} para o modelo de tópico, refletindo a coerência geral do modelo. Ela se baseia na C_{UMass} individual de cada tópico, derivada na Equação 3, e proporciona uma avaliação global da coesão e consistência dos tópicos identificados pelo modelo, onde K representa o número total de tópicos identificados pelo modelo, $\sum_{k=1}^K$ é a somatória que computa a soma das C_{UMass} individuais em $(k; W^{(k)})$ para todos os tópicos, de $k = 1$ a K , por sua vez, $C_{UMass}(k; W^{(k)}) + \varepsilon$ refere-se ao valor da coerência individual do tópico k , como calculado na Equação 3, acrescido de ε para evitar problemas de logaritmo com zero e, por fim, $\frac{1}{K}$ é o divisor que normaliza a soma total das coerências individuais, gerando uma medida média para o modelo como um todo.

Com relação aos resultados, os valores próximos de 0 indicam que o modelo

de tópico identificou tópicos coesos e consistentes, sugerindo uma representação eficaz da estrutura semântica dos documentos. Por sua vez, valores distantes de 0 sugerem que o modelo de tópico identificou tópicos com menor coesão interna ou inconsistências temáticas, indicando a necessidade de ajustes ou refinamentos.

Em suma, a Equação 4 complementa a Equação 3, oferecendo uma visão abrangente da coerência do modelo de tópico, além de avaliar a qualidade e eficácia da representação semântica dos documentos. Essa medida global é essencial para uma análise completa da performance do modelo.

Por sua vez, para medir a coerência C_V , Röder, Both e Hinneburg (2015), propuseram uma forma que consiste em quatro partes principais: (1) segmentação de tópicos em pares de palavras-chave, (2) estimativa de probabilidades de uma palavra-chave ou pares de palavras-chave por janela deslizante booleana, (3) cálculo de medidas de confirmação de tópico que combinam as informações mútuas pontuais normalizadas com a similaridade de cosseno e, finalmente, (4) agregação por média aritmética de todas as medidas de confirmação na coerência final. Esse indicador de coerência retorna uma saída com valores mais altos referindo-se a modelos de tópicos mais interpretáveis. Embora seja computacionalmente caro, descobriu-se que o C_V supera as medidas existentes, incluindo o C_{UMass} , porém sendo mais lenta.

A técnica de otimização de busca é implementada para encontrar a combinação de hiper parâmetros do modelo (ou seja, o número de tópicos K ; densidade de tópicos de documentos α ; e densidade de tópicos de palavras-chave β) que produz o mais coerente modelo. E, por fim, para determinar o número ideal de tópicos (K), as métricas de coerência, C_V e C_{UMass} , devem ser calculadas após o treinamento dos modelos de linha de base dos valores de K para α e β .

Após o ajuste dos modelos LDA visando a otimização das métricas de coerência, o pyLDAvis, uma ferramenta de visualização implementada em Python, foi utilizada para examinar e validar a compreensão dos tópicos extraídos pelos usuários. Para fornecer um contexto mais detalhado, é importante abordar o método de redução da dimensionalidade empregado, conhecido como $t-SNE$, que é uma técnica de redução da dimensionalidade amplamente empregada em análise exploratória de dados e visualização. Essa abordagem permite representar dados complexos de alta dimensionalidade em um espaço bidimensional ou tridimensional, preservando as relações de proximidade entre os pontos originais.

No contexto da visualização de tópicos LDA, o $t - SNE$ é utilizado para projetar os tópicos em um plano bidimensional de forma a destacar padrões e estruturas relevantes. No pyLDAvis, os tópicos do LDA são representados como círculos, e a área de cada círculo é proporcional à frequência geral do tópico, facilitando a interpretação visual dos resultados obtidos.

A métrica de relevância de palavras-chave em análise de tópicos, implementada no pyLDAvis, fornece um indicador valioso de similaridade entre os tópicos em um plano bidimensional (2D). Essa ferramenta permite aos usuários ajustar a medida de relevância para classificar as palavras-chave associadas aos tópicos. A relevância de uma palavra-chave w para um tópico k é calculada pela Equação 5.

$$r(w, k|\lambda) = \lambda \log \phi_{kw} + (1 - \lambda) \log \frac{\phi_{kw}}{p_w} \quad (5)$$

Na Equação 5, ϕ_{kw} é a probabilidade de ocorrência da palavra-chave w para o tópico k , p_w é a probabilidade de encontrar a palavra-chave w no *corpus* e λ é um parâmetro de peso.

O termo *corpus* refere-se ao conjunto de documentos ou textos utilizados em uma análise de tópicos. É a matéria-prima que alimenta o processo de mineração de tópicos, fornecendo o conteúdo textual que será explorado para identificar padrões, temas e relacionamentos latentes. Em relação à métrica de relevância de palavras-chave discutida anteriormente na Equação 5, o *corpus* é essencial, pois as probabilidades de ocorrência das palavras-chave nos tópicos são calculadas com base nos documentos presentes no *corpus*. Portanto, a qualidade e a representatividade do *corpus* são fundamentais para garantir resultados precisos e significativos na análise de tópicos.

De acordo com Röder, Both e Hinneburg (2015), a interpretação dos tópicos pode ser desafiadora quando $\lambda = 1$, pois neste caso as palavras são classificadas exclusivamente por sua probabilidade, resultando na exibição das palavras-chave mais frequentes e possivelmente menos significativas. Por outro lado, quando $\lambda = 0$, são listadas as palavras-chave mais específicas do tópico, o que pode dificultar o estabelecimento de relações semânticas entre elas.

Um exemplo prático de aplicação é a análise de um conjunto de documentos

sobre futebol, onde a métrica de relevância pode ser utilizada para identificar as palavras-chave mais importantes para cada tópico. Por exemplo, um tópico sobre "táticas de jogo" pode ter palavras-chave como "posse de bola", "formação" e "estratégia", enquanto um tópico sobre "notícias de transferências" pode ter palavras-chave como "jogador", "contrato" e "clube".

A visualização dos tópicos no pyLDAvis é enriquecida pela métrica de relevância, refletida no tamanho dos círculos. O tamanho de cada círculo é proporcional à probabilidade geral do tópico, e as palavras-chave mais relevantes são posicionadas mais próximas do centro do círculo.

Além disso, é relevante mencionar o método de redução de dimensionalidade $t - SNE$, pois o mesmo preserva as relações de vizinhança entre os pontos de dados, facilitando a interpretação visual dos relacionamentos entre os tópicos e as palavras-chave associadas.

D) ANÁLISE DE SIMILARIDADE

Na etapa subsequente à extração e categorização dos tópicos, é essencial proceder à quantificação das distâncias e identificação das possíveis inter-relações entre os tópicos delineados. Este estágio reveste-se de significância crítica, uma vez que representa o momento em que se almeja uma compreensão mais aprofundada da estrutura subjacente dos dados coletados, bem como a exploração das relações entre os distintos temas identificados.

A fim de atingir tal desiderato, optou-se por empregar a métrica de distância de Hellinger, uma medida fundamentada no modelo LDA. Essa métrica oferece uma abordagem robusta para avaliar a similaridade ou dissimilaridade entre as distribuições de tópicos (Röder; Both; Hinneburg, 2015). A aplicação da distância de Hellinger objetiva capturar de maneira eficaz a essência dos tópicos identificados, revelando nuances em sua relação e organização.

Cumprе salientar que a escolha da métrica de distância de Hellinger não é arbitrária; ao contrário, foi selecionada em virtude de sua comprovada capacidade de capturar as nuances das distribuições de tópicos de modo preciso e significativo. Ao explorar as relações entre os tópicos por meio dessa métrica, almeja-se obter insights valiosos que permitam avançar na compreensão dos dados e, por conseguinte, gerar

contribuições significativas para o campo da pesquisa (Röder; Both; Hinneburg, 2015).

A escolha pela distância de Hellinger deriva de sua propriedade de sempre produzir um valor finito ao ser aplicada em distribuições com suportes distintos, tornando-a uma opção comum em estudos relacionados à LDA. Para avaliar o grau de similaridade entre duas distribuições de probabilidade p e q , a distância de Hellinger é calculada conforme a Equação 6.

$$HD(p, q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^T (\sqrt{p_i} - \sqrt{q_i})^2} \quad (6)$$

Na Equação 6 $HD(p, q)$ representa a Distância de Hellinger entre as distribuições de probabilidade p e q , T representa o número total de elementos nas distribuições, p_i e q_i representam as probabilidades do i -ésimo elemento nas distribuições p e q , respectivamente.

A escolha da Distância de Hellinger para avaliar a similaridade entre tópicos na LDA se justifica por suas propriedades vantajosas:

- Propriedade de Não-Negatividade: garante que a diferença entre as distribuições seja sempre maior ou igual a zero.
- Propriedade de Trivialidade: indica máxima similaridade quando as distribuições p e q são idênticas.
- Propriedade de Simetria: é independente da ordem das distribuições.
- Propriedade de Triângulo: garante que a distância entre duas distribuições não exceda a soma das distâncias entre cada uma delas e uma terceira distribuição.

Na LDA, os tópicos são representados como distribuições de probabilidade de palavras-chave e a Distância de Hellinger é aplicada para calcular a similaridade entre elas, quantificando a diferença entre suas distribuições de palavras-chave.

A interpretação da Distância de Hellinger ocorre em um intervalo de $[0,1]$, onde valores mais baixos indicam maior similaridade entre os tópicos e valores mais altos indicam maior divergência. Essa medida permite comparar a similaridade entre pares de tópicos, auxiliando na interpretação dos resultados da LDA e na identificação de tópicos relacionados.

6. ESTUDOS DE CASO E DISCUSSÕES

As pesquisas de estudos de caso visam à validação da metodologia adotada, desempenhando papel crucial na resposta ao problema apresentado na introdução desta tese. A Tabela 13 proporciona uma síntese dos estudos de caso, reservando-se a exposição pormenorizada de todos os elementos para as seções subsequentes.

Tabela 13 - Síntese dos Estudos de Caso

	Estudo de Caso 1	Estudo de Caso 2
Objetivo	O estudo aborda a investigação dos fatores determinantes para a decisão de adotar sistemas solares em residências no contexto brasileiro, abrangendo aspectos demográficos, técnicos, financeiros e de opinião.	O estudo visa investigar os principais fatores relacionados à diminuição do consumo de energia elétrica em domicílios no Brasil, considerando aspectos comportamentais, climáticos e as características da infraestrutura.
Páginas da web para a Extração Preditiva de Dados	<ol style="list-style-type: none"> 1. Agência Nacional de Energia Elétrica 2. Instituto Brasileiro de Geografia e Estatística 3. Banco Nacional de Desenvolvimento Econômico e Social 4. Associação Brasileira de Energia Solar Fotovoltaica 5. Facebook 	<ol style="list-style-type: none"> 1. Caixa Econômica Federal 2. Instituto de Pesquisa Econômica Aplicada 3. Instituto Nacional de Pesquisas Espaciais 4. Empresa de Pesquisa Energética 5. WWF Brasil 6. Ministério do Desenvolvimento Regional 7. Instituto Brasileiro de Geografia e Estatística 8. Instituto Nacional de Meteorologia 9. Agência Nacional de Energia Elétrica 10. Greenpeace Brasil 11. Facebook
Período da Coleta dos Dados	Outubro de 2020 a Dezembro de 2020	Abril de 2023 a Junho de 2023
Páginas da Web verificadas durante a Extração Preditiva	10.908	13.098
Palavras Relevantes resultantes do Pré-Processamento	1.933.096	1.711.200
Quantidade de Tópicos Modelados	4	3
Tópicos agrupados por Análise de Similaridade	Demográfico, Técnico, Financeiro e Opinião	Habitante, Habitação e Clima
Configuração das Características dos Agentes Autônomos	<ul style="list-style-type: none"> • Autonomia: Social • Mobilidade: Estática • Reatividade: Contextual • Proatividade: Alta • Comunicação: Adaptativa • Habilidade Social: Alta • Cooperação: Ativa • Aprendizagem: Social 	<ul style="list-style-type: none"> • Autonomia: Hierárquica, Social e Execução • Mobilidade: Estática • Reatividade: Contextual • Proatividade: Alta • Comunicação: Adaptativa • Habilidade Social: Alta • Cooperação: Ativa • Aprendizagem: Contextualizada
Resultados	O aumento na adoção desses sistemas está relacionado com a percepção dos benefícios financeiros e com a proximidade com pessoas que já adotaram essa tecnologia.	A notável redução no consumo de energia elétrica está associada à conscientização das pessoas, ao uso da ventilação natural e à implementação de sistemas de controle térmico nas residências.

Fonte: Autor (2024)

6.1. ESTUDO DE CASO 1

A aceitação da sociedade é necessária para uma difusão bem-sucedida de uma nova tecnologia e com os sistemas de microgeração fotovoltaica não é diferente. A disposição à mudança do comportamento humano em instalar um sistema de microgeração de energia fotovoltaica é importante para uma implantação bem-sucedida de tais tecnologias.

Diversos fatores influenciam o processo de implantação da microgeração de energia fotovoltaica. Os consumidores são motivados por diferentes razões, como a busca por autonomia energética, interesse em novas tecnologias, preocupações ambientais e benefícios econômicos. Além desses fatores, há também a necessidade imediata de fontes energéticas confiáveis e ininterruptas, pois as pessoas dependem da eletricidade, independentemente de sua origem. Nesse sentido, as tecnologias de microgeração fotovoltaica se apresentam como opções acessíveis e disponíveis para suprir essa demanda.

No entanto, a adoção desses sistemas solares domésticos é impactada por desafios como o conhecimento técnico necessário, os custos elevados e a dificuldade de realizar testes práticos. Esses obstáculos contribuem para a disseminação e implantação lenta desses sistemas.

Para promover a adoção da microgeração fotovoltaica, é crucial aumentar o conhecimento e a familiaridade das pessoas com essa tecnologia. No caso dos sistemas solares domésticos, as informações disponíveis muitas vezes são excessivamente técnicas, o que dificulta a compreensão por parte dos consumidores. Além disso, os custos envolvidos também representam um desafio, tornando necessária a busca por soluções que viabilizem a redução desses custos.

Superar esses obstáculos é essencial para acelerar a disseminação e implantação dos sistemas de microgeração fotovoltaica, permitindo que mais pessoas possam se beneficiar dessa fonte de energia sustentável e renovável.

A teoria da difusão de Rogers (2010), categoriza os adotantes com base no tempo. De acordo com essa teoria, o avanço da implantação de uma nova tecnologia em uma determinada área depende de sua vantagem relativa, compatibilidade, facilidade de uso e aceitação social. O pesquisador explica que os consumidores inovadores, ou seja, aqueles que primeiro instalam a nova tecnologia, influenciam futuros adotantes em sua vizinhança. Tendo isso em vista, o modelo desenvolvido por

Bass (1969), estima sobre a influência que os consumidores inovadores tendem a exercer sobre os futuros adotantes. De acordo com esse modelo, a probabilidade de uma família decidir pela instalação de energia fotovoltaica solar em um determinado momento t é determinada por um coeficiente de inovação p , associado ao adotante inicial, e um coeficiente de imitação q , associado à influência da comunidade, conforme indicado na Equação 7.

$$(p + qF(t)) \quad (7)$$

Na Equação 7, $F(t)$ é a proporção de adotantes no momento t . Sem a influência da vizinhança, $p > 0, q = 0$, enquanto sem adotantes precoces $p = 0, q > 0$. A função de probabilidade para uma casa que está decidindo sobre a instalação de energia fotovoltaica solar em um momento t é dada pela Equação 8.

$$f(t) = (p + q(F(t)))(1 - F(t)) \quad (8)$$

E a função de densidade cumulativa correspondente é dada pela Equação 9.

$$F(t) = \frac{1 - \exp(-(p+q)t)}{1 + \frac{q}{p} \exp(-(p+q)t)} \quad (9)$$

Dado um fator potencial de mercado m , a adoção cumulativa de energia fotovoltaica solar em um momento t é dada por $F(t) \times m$. Os coeficientes p e q , e o fator de mercado m são considerados variáveis para explicar o ambiente em mudança dentro do qual a difusão de uma nova tecnologia ocorre. As decisões iniciais e independentes de adoção são principalmente influenciadas por uma variedade de fatores, tais como benefícios econômicos, pressões sociais, conscientização da importância e entusiasmo em relação à nova tecnologia, além de preocupações ambientais no caso específico da energia fotovoltaica solar. Esses fatores são capturados no coeficiente de inovação p . Os benefícios percebidos e comentados pelos consumidores inovadores em suas redes de contato, inclusive redes sociais, da nova tecnologia são capturados no coeficiente de imitação q . Fatores geográficos como localização e demografia determinam os níveis de saturação do mercado que

são então capturados no fator potencial de mercado m .

Tanto a teoria de Rogers (2010) quanto o modelo Bass (1969), ressaltam a influência da vizinhança como um fator importante na aceitação social e, portanto, na difusão de uma nova tecnologia em uma determinada área. Bollinger e Gillingham (2012), argumentam que a influência da vizinhança desempenha um papel importante, pois os consumidores iniciais ao aumentar a visibilidade da nova tecnologia junto aos comentários falados e/ou escritos levam à adoção em um raio de influência.

Muitas famílias em determinadas áreas necessitam de eletricidade para o uso, uma vez que há ausência e/ou precariedade no fornecimento da rede elétrica tradicional. Pesquisas indicam que, para essas famílias, existe um interesse em adquirir sistemas de microgeração elétrica fundamentados em fontes de energia renovável, devido aos benefícios socioeconômicos globais proporcionados por tais sistemas; além disso, os benefícios adicionais da microgeração incluem o incremento da autossuficiência energética, uma melhor percepção de prestígio na comunidade e a disponibilidade de energia elétrica que estimula o aumento das atividades socioeconômicas durante a noite. Além disso, as infraestruturas de microgeração fornecem novas fontes de qualificação e de emprego.

É difícil modelar os impactos de diferentes aspectos não quantitativos na adoção de uma nova tecnologia. No entanto, o raio de influência, que é um parâmetro mensurável, evidencia como a adoção de uma tecnologia por uma família influencia outras a fazer o mesmo. Com isso, é possível modelar esse cenário complexo para observar os impactos desse parâmetro na adoção de uma nova tecnologia.

6.1.1. COLETA DE DADOS

Uma pesquisa utilizando a mineração de dados foi desenvolvida e realizada para coletar dados sobre as famílias e os domicílios com instalações de sistemas solares domésticos em áreas do Brasil. A pesquisa foi dividida em quatro tópicos principais: (1) demográfico, (2) técnico, (3) financeiro e (4) opinião.

O tópico com informações demográficas visou identificar dados relacionados às características das famílias, incluindo o tamanho do domicílio e o nível de escolaridade dos membros familiares. Por sua vez, o tópico com informações técnicas teve como objetivo identificar o dimensionamento dos sistemas solares domésticos

instalados, bem como as motivações por trás de sua implementação, eventuais problemas enfrentados durante a instalação e quaisquer manutenções ou substituições realizadas até o momento. No que diz respeito ao tópico com informações financeiras, a ênfase foi dada à coleta de dados referentes ao nível de renda das famílias e ao processo de aquisição dos sistemas solares domésticos. Adicionalmente, tópico com informações das opiniões buscou elucidar o impacto da influência social, incluindo a pressão social e a influência da vizinhança, sobre a decisão de adotar sistemas solares domésticos. Especificamente, este tópico examinou como observações, comentários verbais ou escritos e propagandas influenciaram as decisões de instalação desses sistemas.

É importante destacar que a obtenção de dados foi conduzida por meio de investigações realizadas em páginas da *web* disponíveis na Internet, no período compreendido entre outubro e dezembro de 2020.

A área de pesquisa foi dividida em cinco regiões do Brasil, com o objetivo de assegurar uma distribuição equilibrada das amostras e uma maior proximidade com a realidade do país. Foram selecionadas amostras que correspondiam a residências com sistemas solares instalados. A Tabela 14 mostra a população, a área, a densidade demográfica e as porcentagens das amostras de domicílios coletadas.

Tabela 14 - Dados regionais do Brasil

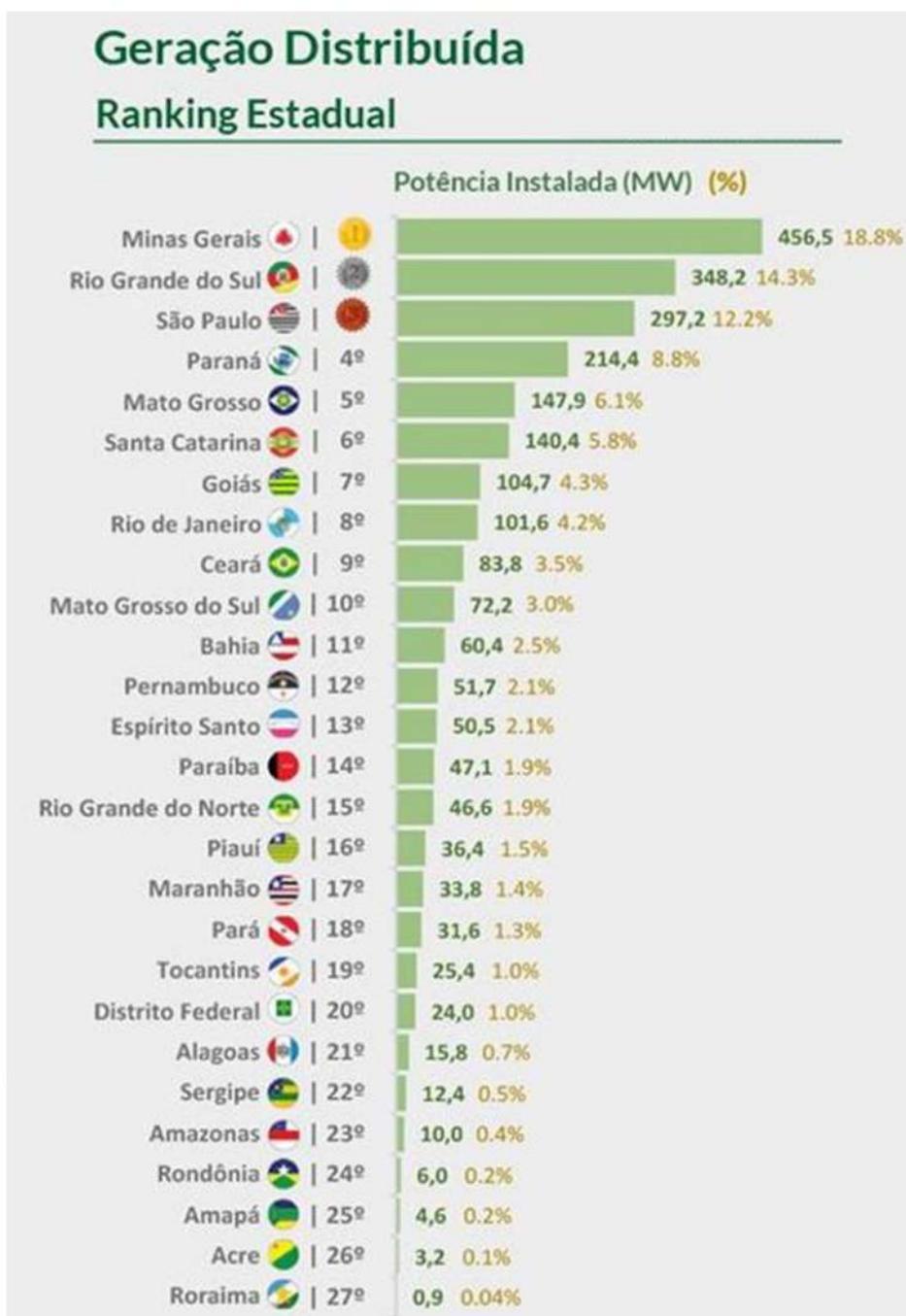
Região	População	Área	Densidade Demográfica	Amostras (%)
Sudeste	87.711.946	924.620	94,86	41,1
Sul	29.016.114	576.774	50,31	24,9
Nordeste	56.560.081	1.544.291	36,63	16,4
Centro Oeste	16.001.102	1.606.403	9,96	14,7
Norte	18.182.253	3.853.676	4,72	2,9

Fonte: Autor (2021)

Com base nos resultados apresentados na Tabela 14, observa-se uma estreita concordância entre as proporções das amostras coletadas por meio da raspagem de dados e os dados oficiais divulgados pela Associação Brasileira de Energia Solar Fotovoltaica em 2020. As estatísticas, segmentadas por estados, indicam que a distribuição da energia solar fotovoltaica no Brasil segue a seguinte distribuição regional: Sudeste 38,4%, Sul 25,5%, Nordeste 17,1%, Centro-Oeste 15,8% e Norte 3,2% (Associação Brasileira de Energia Solar Fotovoltaica, 2020). Este padrão de

distribuição é representado graficamente na Figura 34.

Figura 34 - Geração de energia solar fotovoltaica no Brasil



Fonte: Associação Brasileira de Energia Solar Fotovoltaica (2020)

EXTRAÇÃO PREDITIVA DE DADOS DE TEXTO

Nesta etapa da pesquisa, foi realizada a coleta de um conjunto abrangente de dados textuais por meio de técnicas de extração preditiva. As fontes utilizadas

abrangeram páginas da *web* governamentais, associações e redes sociais *online*, com o intuito de retratar a estratégia de utilização da energia solar fotovoltaica no contexto nacional. Os dados foram submetidos a uma análise criteriosa e uma cuidadosa seleção, eliminando informações curtas e duplicadas, ao mesmo tempo em que todas as *tags* relevantes foram mantidas, visando à captura de tendências mais amplas. Essa abordagem de coleta em larga escala foi adotada com o propósito de obter conjuntos de dados, que foram posteriormente empregados na modelagem de tópicos, para identificar as principais tendências relacionadas aos sistemas solares.

Portanto, o presente estudo utilizou técnicas de extração preditiva de dados textuais provenientes de fontes *online* pertinentes, permitindo a análise e modelagem de tópicos para a identificação das principais tendências relacionadas aos sistemas solares residenciais. Ao final desta etapa, um total de 9.702 informações foram coletadas, consolidando assim uma base de dados substancial para as etapas subsequentes da pesquisa. A Tabela 15 mostra a descrição dos dados coletados.

Tabela 15 - Descrição dos dados de texto coletados

Páginas da web	URL	Número de Informações	Período da Coleta		Frequência Mensal das Publicações
			Início	Fim	
Agência Nacional de Energia Elétrica	https://www.aneel.gov.br	2.206	Out./2020	Dez./2020	17,49
Instituto Brasileiro de Geografia e Estatística	https://www.ibge.gov.br	1.550	Out./2020	Dez./2020	22,92
Banco Nacional de Desenvolvimento Econômico e Social	https://www.bndes.gov.br	2.757	Out./2020	Dez./2020	20,74
Associação Brasileira de Energia Solar Fotovoltaica	https://www.absolar.org.br	1.206	Out./2020	Dez./2020	10,49
Facebook	https://www.facebook.com	3.189	Out./2020	Dez./2020	44,52

Fonte: Autor (2021)

PRÉ-PROCESSAMENTO DE DADOS DE TEXTO

Neste estudo, foi realizado um levantamento de 3.710.862 *tokens* provenientes de 9.702 conjuntos de informações. Para aprimorar a qualidade dos dados textuais, foi aplicada uma técnica de redução de ruído que envolveu a remoção de pontuação e de palavras presentes na lista de bibliotecas Python, em especial a biblioteca NLTK. Essas medidas resultaram em uma redução de aproximadamente 17% do número total de palavras presentes no *corpus* original. A Figura 35 mostra a nuvem de palavras coletada, bem como a frequência das mesmas.

Figura 35 - Nuvem de palavras coletadas



Fonte: Autor (2021)

Após isso, foram identificados os bigramas que representam pares de palavras frequentemente associados nas publicações, como por exemplo, "painel solar" e "placa fotovoltaica". Ao fim dessa fase foram aplicadas técnicas de pré-processamento de texto. A Tabela 16 ilustra a decomposição dos dados.

Tabela 16 - Decomposição dos dados de texto coletados

Páginas da web	Publicações	Sentenças	Tokens	Palavras	Palavras Relevantes	% Palavras Não Relevantes	Vocabulário
Agência Nacional de Energia Elétrica	2.206	27.904	605.079	376.977	311.376	17,40	16.369
Instituto Brasileiro de Geografia e Estatística	1.550	30.080	528.220	320.053	268.802	16,01	17.960
Banco Nacional de Desenvolvimento Econômico e Social	2.757	30.793	719.664	447.852	365.678	18,35	16.770
Agência Nacional de Energia Elétrica	1.206	20.040	417.551	297.459	243.857	18,02	15.084
Facebook	3.189	70.488	1.440.348	908.278	743.383	17,60	30.662

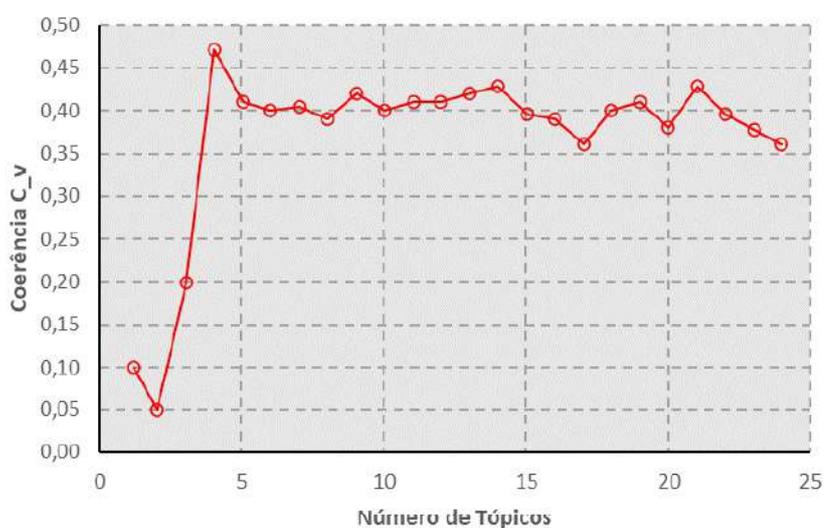
Fonte: Autor (2021)

MODELAGEM POR TÓPICOS E ANÁLISE DE SIMILARIDADE

Aqui a LDA foi aplicada para determinar o número ideal de tópicos (K) a partir das métricas de coerência, C_V e C_{UMass} , que foram calculadas após o treinamento dos modelos onde $\alpha = 0,01$ e $\beta = 0,11$. Nesse cenário, como o C_V superou as medidas de

C_{UMass} , a faixa de K com as medidas de C_V mais altas. No entanto, uma vez que as pontuações C_V tendem a continuar aumentando com o número de tópicos, os valores K ideais foram determinados encontrando um ponto de onde retornos crescentes não valeria mais o custo adicional da complexidade do modelo. Essas determinações heurísticas foram realizadas mantendo um olhar crítico sobre as flutuações de C_{UMass} . A Figura 36 mostra a determinação do número ótimo de tópicos K usando as medidas de coerência C_V e C_{UMass} .

Figura 36 - Número de tópicos. (a) Métrica: C_V . (b) Métrica: C_{UMass}



(a)



(b)

Fonte: Autor (2021)

A Figura 36 (a) mostra que a métrica C_V atinge seu pico localmente em 4, enquanto na Figura 36 (b), a métrica C_{UMass} atinge um platô de pontuação a partir desse ponto. Isso confirma que os tópicos modelados são quatro: (1) demográfico, (2) técnico, (3) financeiro e (4) opinião. Páginas da web com maior conteúdo tendem a ter maior densidade documento-tópico (α). A Tabela 17 indica ajustes no modelo visando melhorar o desempenho nas medidas de coerência.

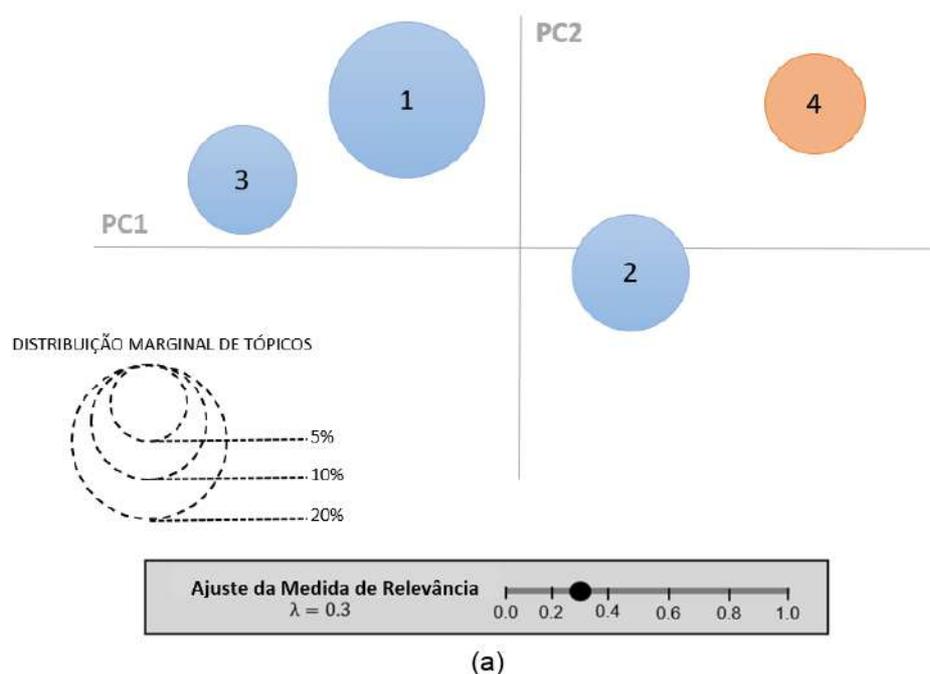
Tabela 17 - Ajuste e melhoria das medidas de coerência

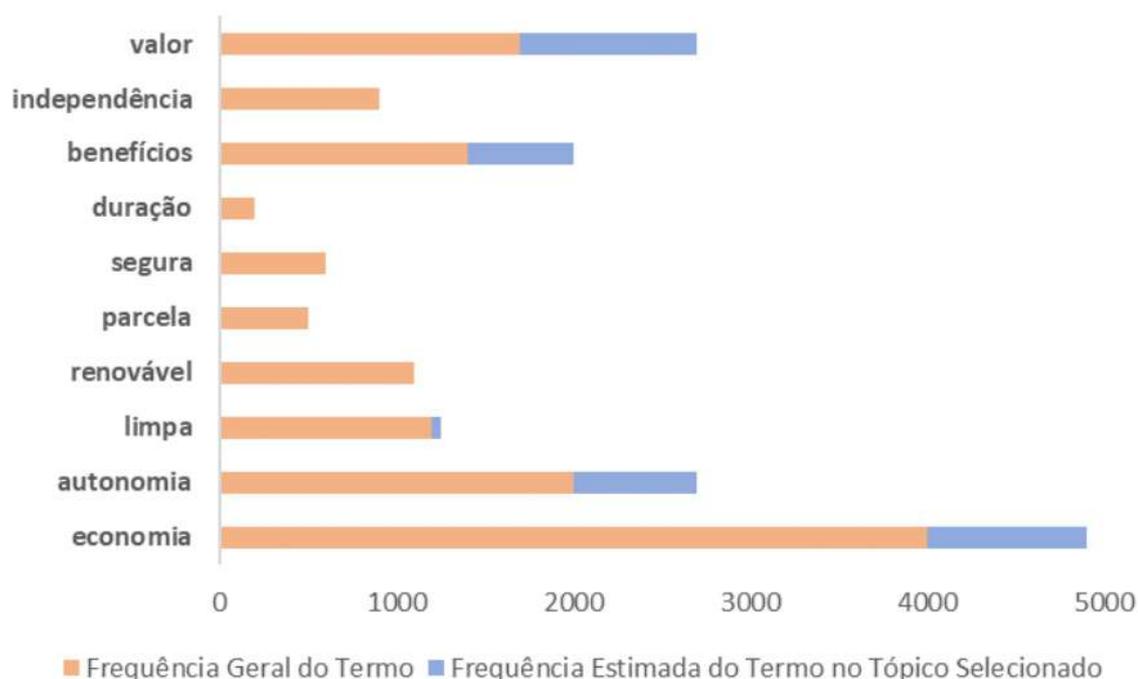
Dados	Valores
Número de tópicos (K)	4
Densidade de tópico de documento (α)	0,81
Densidade de tópico de palavras (β)	0,81
C_{UMass} (modelo original)	-1.269
C_{UMass} (modelo ajustado)	-1.047
Melhoria C_{UMass} (%)	+17%
C_V (modelo original)	0,472
C_V (modelo ajustado)	0,529
Melhoria C_V (%)	+12%

Fonte: Autor (2021)

Com o objetivo de obter informações relevantes, os tópicos do modelo LDA foram semanticamente categorizados, como ilustrado na Figura 37.

Figura 37 - Layout pyLDAvis (a) Distância entre tópicos (b) Palavras por tópico





(b)

Fonte: Autor (2021)

Após isso, a métrica de distância de Hellinger foi empregada para quantificar as distâncias e as inter-relações entre os tópicos extraídos.

Com base na análise das métricas de relevância, pode-se observar na Figura 37 (a), que há uma interpretação clara e concisa sobre a divisão dos principais tópicos. Essa constatação foi destacada e exemplificada na Figura 37 (b), onde são fornecidas informações relacionadas à frequência do termo geral ocorrido no tópico 4, que é relativa à opinião das pessoas.

Este processo de análise visa fornecer uma compreensão mais aprofundada dos fatores que moldam a simulação do estudo de caso, contribuindo assim para uma investigação mais robusta e abrangente na área de agentes inteligentes.

Ao término desta etapa, as 10 palavras mais relevantes dos tópicos modelados contribuíram para a interpretação aprimorada dos quatro tópicos, conforme apresentado na Tabela 18.

Tabela 18 - Modelagem de tópicos

No.	Tópico	Freq. (%)	C _v	C _{UMass}	As 10 palavras mais relevantes ($\lambda = 0,3$)
1	Demográfico	27,41	0,43	- 0,58	{norte, sul, nordeste, centro-oeste, casa, sudeste, região, residência, brasil, domicílio}
2	Técnico	15,51	0,41	- 1,02	{solar, energia, painel, fotovoltaico, sistema, placa, instalação, potência, autonomia, elétrica}
3	Financeiro	13,73	0,44	- 0,98	{banco, financiamento, vista, prazo, parcela, empréstimo, dinheiro, cartão, dinheiro, pagamento}
4	Opinião	10,36	0,51	- 0,76	{economia, autonomia, limpa, renovável, limpa, segura, duração, benefícios, independência, valor}

Fonte: Autor (2021)

6.1.2. MODELAGEM DO CENÁRIO UTILIZANDO SMA

No Estudo de Caso 1, um modelo foi desenvolvido no NetLogo, como um instrumento para modelar os impactos da influência da vizinhança na aceitação social e na difusão dos sistemas solares domésticos em áreas do Brasil. Esse modelo considera a distribuição populacional em uma área específica, o potencial de geração de energia solar nessa área, as limitações das tecnologias de geração de energia solar e as decisões dos indivíduos com base nos custos, influência da vizinhança e opções de eletrificação disponíveis. O modelo simula as interações entre esses fatores para entender os efeitos globais resultantes de diferentes decisões individuais.

Os dados provenientes da etapa de raspagem de dados foram utilizados para criar indicadores e alimentar um modelo. Esses dados incluem informações sobre a distribuição da população da região, o número total de sistemas solares domésticos instalados, os tamanhos dos sistemas solares domésticos instalados, os motivos para a instalação desses sistemas, a influência da vizinhança nas decisões de instalação, o impacto dos custos nas decisões de instalação e as opiniões sobre sistemas solares domésticos. Esses dados são usados para simular a difusão ao longo do tempo dos sistemas solares domésticos na área. Os seguintes agentes foram criados no modelo:

- uma representação do meio ambiente e do potencial solar.
- as populações que requerem sistemas solares domésticos.
- sistemas solares domésticos que produzem eletricidade.
- um observador central determina as estratégias e preferências para as difusões de energia fotovoltaica solar.

As características dos agentes autônomos em relação ao estudo de caso fornecido foram modeladas de acordo com as seguintes dimensões dentro do ambiente do NetLogo:

- **Autonomia:**
 - Social: os agentes são capazes de considerar a opinião de outros agentes e o contexto social ao decidirem sobre a adoção de sistemas solares em residências.
- **Mobilidade:**
 - Estática: os agentes permanecem em uma localização fixa e não requerem mobilidade física para a simulação, uma vez que a decisão de adotar sistemas solares é feita em relação à residência de cada agente.
- **Reatividade:**
 - Contextual: os agentes respondem a mudanças no ambiente, como variações nos custos ou na percepção dos benefícios dos sistemas solares, adaptando suas decisões de acordo.
- **Proatividade:**
 - Alta: os agentes podem antecipar tendências de adoção de sistemas solares com base em padrões históricos e tomar medidas proativas para promover a adoção, como campanhas de conscientização.
- **Comunicação:**
 - Adaptativa: os agentes ajustam sua forma de comunicação com base nas características demográficas e na opinião dos outros agentes, adaptando suas estratégias de persuasão.
- **Habilidade Social:**
 - Alta: os agentes podem interagir com os habitantes e outros agentes, trocando informações sobre os benefícios dos sistemas solares e coordenando esforços para promover sua adoção.
- **Cooperação:**
 - Ativa: os agentes colaboram entre si e com os habitantes para promover a adoção de sistemas solares, coordenando campanhas de conscientização e compartilhando recursos.
- **Aprendizagem:**
 - Social: os agentes aprendem não apenas com sua própria experiência,

mas também com a experiência compartilhada por outros agentes em interações sociais, adaptando suas estratégias de adoção com base no sucesso observado em outros contextos demográficos ou regionais.

Os dados da coleta de dados foram usados para criar indicadores e alimentar o modelo de simulação ao longo do tempo. Esses dados simulam a difusão dos sistemas, com parâmetros detalhados na Tabela 19.

Tabela 19 - Parâmetros usados no modelo NetLogo

Parâmetro	Descrição
$LUCE_{PV}$	Custo unitário nivelado da eletricidade fornecida
$C_{A/kWh}$	Custo evitado
$EHFS$	Horas equivalentes de sol pleno por dia
CUF	Fator de utilização da capacidade
$ALCC_{PV}$	Custo do ciclo de vida anualizado
C_{0PV}	Custo de capital do módulo fotovoltaico
C_{0batt}	Custo de capital da bateria
C_{0cc}	Custo de capital do controlador de carga
C_{0appl}	Custo de capital de eletrodomésticos
CRF	Fator de recuperação de capital
$C_{O\&M}$	Custo de operação e manutenção
i	Taxa de desconto
T_{IR}	Limiar da vizinhança
p	Coefficiente de inovação
q	Coefficiente de imitação
$F(t)$	Proporção de adotantes no tempo t
m	Fator potencial de mercado
$H_{PV/IR}$	Famílias com PV dentro de um determinado raio de influência
$H_{Total/IR}$	Total de domicílios dentro do mesmo raio de influência

Fonte: Autor (2021)

A Equação 10 representa os parâmetros a serem considerados ao avaliar a viabilidade da instalação de um sistema solar doméstico em residências desprovidas desse sistema.

$$LUCE_{PV} < C_{A/kWh} \quad (10)$$

Na Equação 10, $C_{A/kWh}$ representa o custo evitado por quilowatt-hora kWh , indicando o custo da eletricidade da rede nacional predominante por kWh , enquanto $LUCE_{PV}$ denota o custo unitário nivelado da eletricidade fornecida, conforme expresso na Equação 11.

$$LUCE_{PV} = \frac{ALCC_{PV}}{W_p \times EHFS \times 365 \times CUF} \quad (11)$$

Na Equação 11, W_p representa a capacidade nominal de pico em watts do sistema de painéis solares domésticos, e é determinada com base no perfil de atividades de uma família e na demanda energética. $EHFS$ é o equivalente a horas de luz solar plena por dia, enquanto CUF denota o fator de utilização da capacidade, o qual incorpora a subutilização e interrupções do sistema devido a diversas razões. Por fim, $ALCC_{PV}$ representa o custo do ciclo de vida anualizado, calculado pela soma dos custos de todos os componentes individuais do sistema - incluindo painéis, baterias, controladores de carga e dispositivos - multiplicados pelos respectivos fatores de recuperação de capital, além dos custos de operação e manutenção. Essa relação é expressa pela Equação 12.

$$ALCC_{PV} = (C_{0PV} \times CRF_{PV}) + (C_{0batt} \times CRF_{batt}) + (C_{0cc} \times CRF_{cc}) + (C_{0appl} \times CRF_{appl}) + C_{O\&M} \quad (12)$$

A Equação 12 representa o cálculo do custo total do sistema de energia solar residencial, onde os custos de capital de diferentes componentes, como o painel solar C_{0PV} , a bateria C_{0batt} , o controlador de carga C_{0cc} e os aparelhos C_{0appl} , são ponderados pelos fatores de recuperação de capital correspondentes CRF_{PV} , CRF_{batt} , CRF_{cc} e CRF_{appl} , além de incluir o custo de operação e manutenção $C_{O\&M}$. O fator de recuperação de capital CRF é determinado pela Equação 13.

$$CRF = \frac{i(1+i)^n}{(1+i)^n - 1} \quad (13)$$

Na Equação 13, onde a taxa de desconto i e o tempo de vida útil do

componente específico n são os principais parâmetros considerados.

A Equação 14 descreve a condição para a instalação real de sistemas de energia solar residencial por famílias.

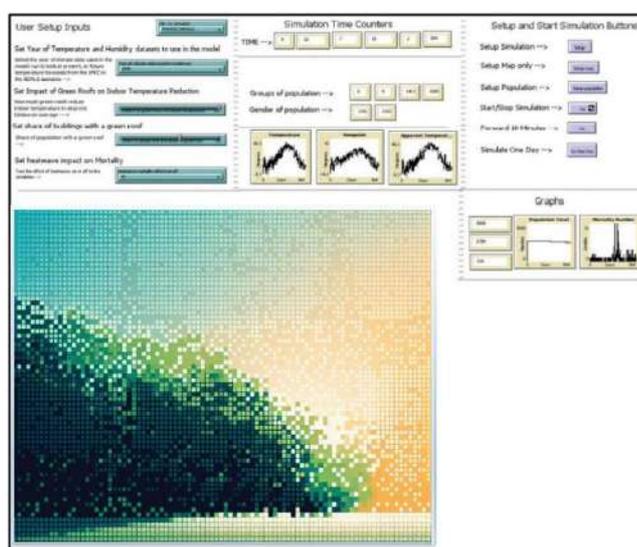
$$\frac{H_{PV/IR}}{H_{Total/IR}} \times 100 > T_{IR} \quad (14)$$

Na Equação 14, $H_{PV/IR}$ é o número de domicílios com energia fotovoltaica solar dentro de um determinado raio de influência (IR), $H_{Total/IR}$ é o número total de domicílios dentro do mesmo raio de influência e T_{IR} é o limite da vizinhança.

Os dados da raspagem de dados são usados para alimentar o modelo baseado em agentes autônomos a partir do qual o raio de influência e o limiar da vizinhança são usados para simular os impactos da influência de amigos, parentes e vizinhos, e da pressão social exercida pelos meios de comunicação na implantação dos sistemas solares domésticos considerando o Brasil.

A Figura 38 mostra a simulação em análise, cuja programação foi estabelecida para abranger um período de 25 anos. A paisagem é representada por tons de verde, enquanto os pontos negros representam as residências desprovidas de sistemas solares domésticos. Por outro lado, os pontos brancos denotam as residências que optaram por instalar esses sistemas, e os pontos amarelos indicam aquelas que já possuem sistemas solares domésticos em funcionamento.

Figura 38 - Visão durante as simulações



Fonte: Autor (2021)

6.1.3. RESULTADOS E DISCUSSÕES

Os resultados das amostras de domicílios foram definidos, em porcentagens, para que a coleta dos dados elicitasse os indicadores mais realistas a fim de alimentar a simulação, uma vez que, os dados acerca dos indicadores sobre os sistemas solares domésticos são, em sua maioria, distribuídos, heterogêneos e não-estruturados.

De acordo com dados coletados, o principal motivo para a instalação de sistemas solares domésticos é a economia a médio e longo prazo em relação aos custos de energia. Em segundo lugar, está o desejo de obter maior autonomia, com a possibilidade de ter uma fonte de energia própria. Em terceiro lugar está a falta ou ineficiência de acesso à rede elétrica em um número considerável de residências, o que leva as pessoas a buscar uma fonte de energia alternativa. A preocupação com o uso de fontes de energia renováveis ocupa o quarto lugar, seguida pelo interesse na tecnologia fotovoltaica. A Tabela 20 resume essas informações.

Tabela 20 - Principais razões para instalação

Razões	Porcentagem (%)
Economia	44,33
Autonomia	36,00
Acesso a uma fonte de energia	9,00
Preocupações ambientais	7,00
Novas tecnologia	3,67

Fonte: Autor (2021)

Conforme as amostras coletadas, verificou-se que em 11% dos domicílios a instalação de sistemas solares domésticos ocorreu de forma independente, sem influência da vizinhança ou pressão social, mas sim por interesse dos moradores.

Os demais domicílios foram classificados como consumidores secundários, pois instalaram os sistemas solares domésticos influenciados por outros fatores, como observar vizinhos, amigos e/ou parentes com sistemas já instalados, receber comentários (por escrito ou verbalmente) de vizinhos, amigos e/ou parentes sobre os benefícios e serem impactados por propagandas na TV, Internet ou outros meios.

Os resultados indicam que a influência da vizinhança teve um papel mais significativo nas decisões de instalação do sistema solar doméstico do que as propagandas. No entanto, ressalta-se que os anúncios também desempenharam um

papel relevante nas escolhas dos consumidores ao optarem pela instalação dos sistemas solares em suas residências. A Tabela 21 resume essas informações.

Tabela 21 - Maiores influências para instalação

Influência	Porcentagem (%)
Influência (vizinhos, amigos, parentes)	54,00
Propagandas (TV, Internet)	35,00
Primeiros instaladores	11,00

Fonte: Autor (2021)

Segundo os 162 domicílios analisados, que foram influenciados por vizinhos, amigos e/ou parentes por meio de redes sociais, 111 decidiram instalar um sistema solar doméstico após receber influência de comentários (falados e/ou escritos) diretos que destacavam os potenciais benefícios desses sistemas. Outros 26 domicílios foram influenciados após verem sistemas solares domésticos instalados nas residências de pessoas próximas.

Para 105 domicílios que instalaram o sistema solar doméstico após propaganda, em média, 79 famílias adotaram essa decisão após ver propagandas em redes sociais, enquanto outras 51 residências foram influenciadas por propagandas em outros meios, como televisão.

Portanto, a influência de vizinhos, amigos e/ou parentes que já possuíam sistemas solares domésticos, seja por meio de comentários diretos ou de visualização, desempenhou um papel mais significativo em suas decisões de instalação do que as propagandas.

Em média, 54% dos domicílios optaram por instalar um sistema solar doméstico após receberem comentários (verbalizados e/ou escritos), enquanto aproximadamente 35% decidiram pela instalação após serem influenciados por propagandas em diferentes mídias, incluindo Internet, televisão e outros canais de comunicação. Esse resultado sugere que a influência de pessoas próximas que já possuem sistemas solares domésticos tem uma probabilidade 54% maior de levar à adoção desses sistemas em comparação com os 35% das famílias que decidem pela instalação após serem influenciadas por propagandas.

Essas descobertas ressaltam a importância do aspecto social e da influência interpessoal na adoção de sistemas solares domésticos. Além disso, evidenciam a eficácia da comunicação direta entre pares no processo de tomada de decisão

relacionado à tecnologia sustentável, em contraste com estratégias de marketing mais tradicionais. Esses dados estão sumarizados na Tabela 22.

Tabela 22 - Comparação entre maiores influências para instalação

Influência	Porcentagem (%)
Influência (vizinhos, amigos, parentes) em redes sociais	54,01
Propaganda (Internet)	21,32
Propaganda (TV ou outros meios de comunicação)	13,67
Primeiros instaladores	11

Fonte: Autor (2021)

Com relação ao método de pagamento, a coleta dos dados revelou que mais de 65% das famílias, em média, escolhem adquirir sistemas solares domésticos por meio de parcelamento direto com as empresas que vendem e instalam esses sistemas.

Por outro lado, aproximadamente 22% das residências optam por recorrer a empréstimos bancários para efetuar a compra desses sistemas. Além disso, cerca de 12% dos domicílios optam por pagar à vista pela aquisição e instalação dos sistemas solares domésticos.

Esse último valor é bastante semelhante ao número de consumidores iniciais, o que sugere que essas famílias se preparam mais adequadamente para essa aquisição e, possivelmente, conseguem descontos mais vantajosos. A Tabela 23 condensa essa comparação.

Tabela 23 - Comparação dos meios de pagamento

Pagamento	Domicílios	Porcentagem (%)
Parcelamento	196	65,33
Empréstimo	68	22,67
A vista	36	12,00

Fonte: Autor (2021)

A Figura 39 ilustra um gráfico que representa o crescimento do número de domicílios com sistemas solares domésticos ao longo de um período de 25 anos. No ano inicial, 2021, estima-se que tenha havido aproximadamente 350 residências com tais sistemas instalados, considerando uma média hipotética para uma cidade no

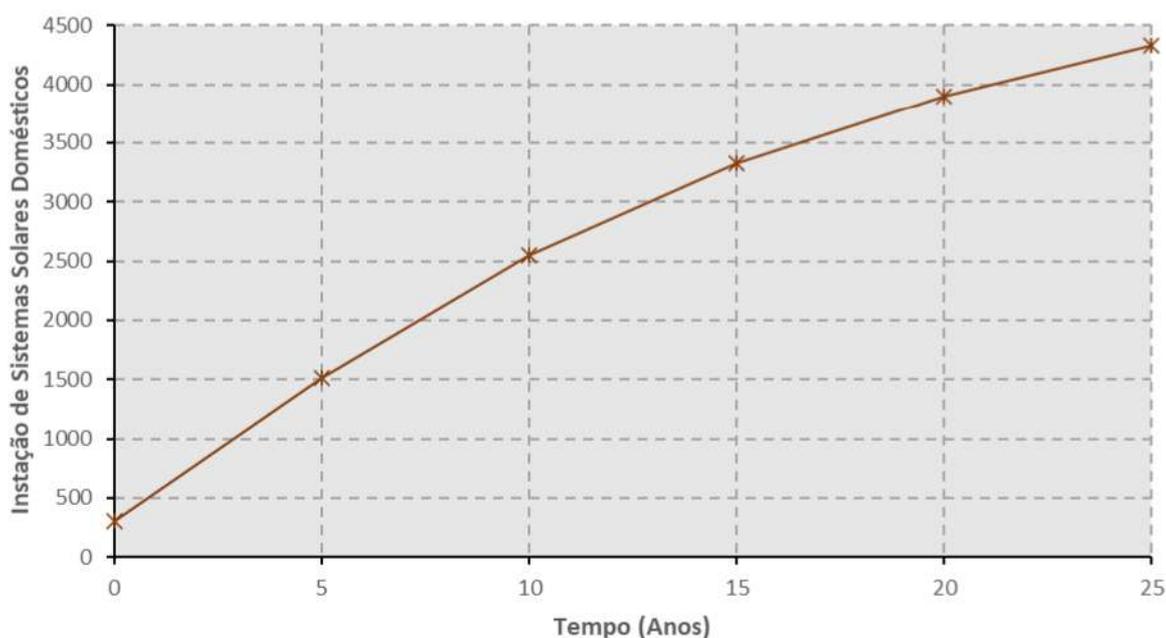
Brasil.

Os resultados obtidos por meio de simulação revelam que, após 25 anos, esse número aumentará para cerca de 4.325 domicílios, representando um crescimento de 1235,71% nas instalações.

Essa taxa de crescimento exponencial pode ser atribuída a diversos fatores, destacando-se, principalmente, o aumento da influência da vizinhança, a maior disponibilidade de crédito para aquisição desses sistemas, o aumento da pressão social e a crescente conscientização acerca dos benefícios socioeconômicos associados aos sistemas solares domésticos.

Outros fatores que contribuem, em menor escala, incluem a redução dos custos da energia fotovoltaica solar e o aprimoramento da eficiência dos sistemas de energia fotovoltaica solar.

Figura 39 - Instalações sistemas solares domésticos em 25 anos



Fonte: Autor (2021)

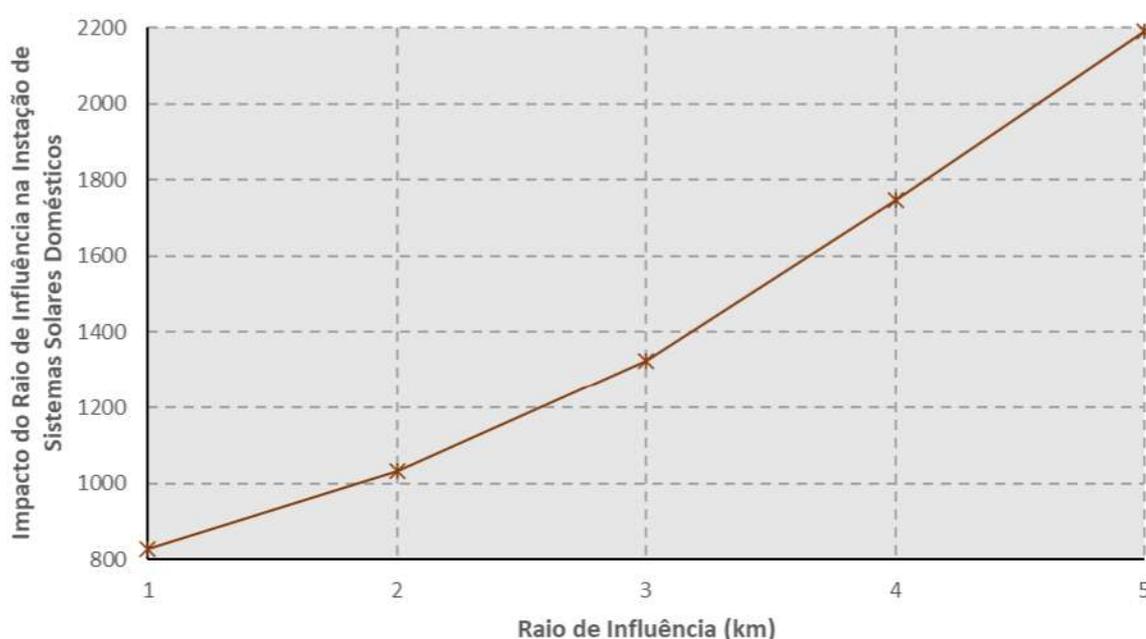
O raio de influência é o raio dentro do qual uma família influencia ou é influenciada por seus vizinhos. O raio padrão é definido em 1 km com base nos estudos de Rogers (2010) e Bass (1969).

O modelo simula como o aumento do raio de influência de uma família afeta a decisão de instalação de sistemas solares domésticos.

Conforme mostrado na Figura 40 com um raio de influência padrão de 1 km, 828 domicílios já teriam instalado sistemas solares domésticos, considerando que 350 domicílios são de consumidores iniciais e os 478 domicílios restantes são de consumidores secundários que já foram influenciados.

A simulação mostra que esse número aumenta exponencialmente com o aumento do raio de influência, mostrando que 2.189 domicílios teriam instalado os sistemas solares domésticos se o raio de influência fosse definido em 5 km.

Figura 40 - Impacto do raio de influência na instalação



Fonte: Autor (2021)

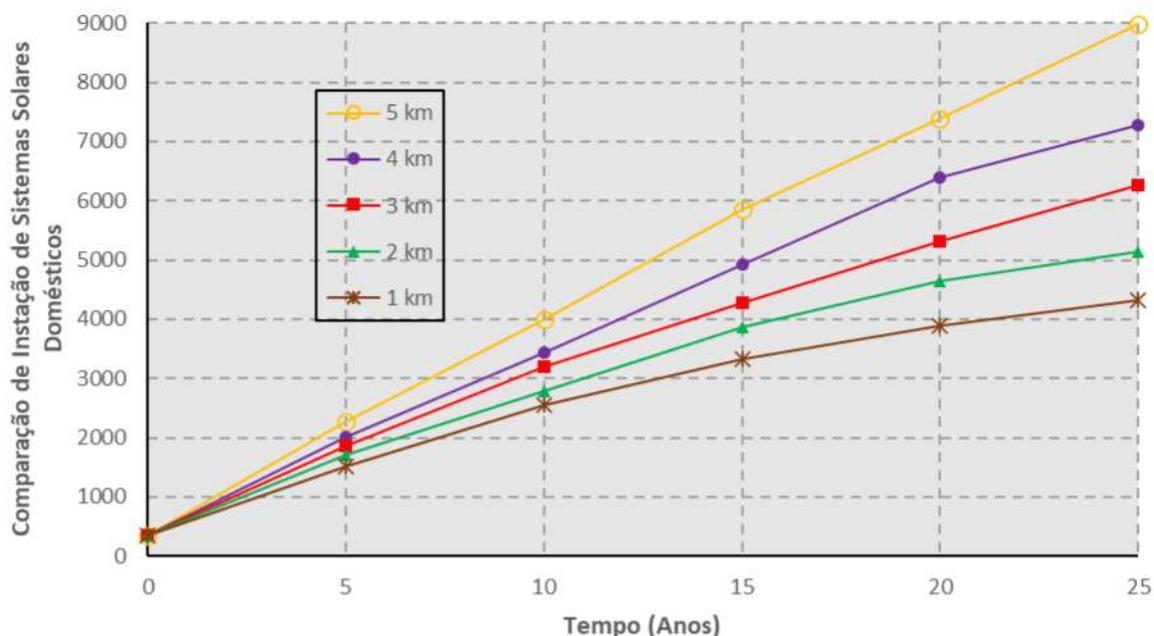
De acordo com a Figura 41, a qual apresenta uma comparação dos impactos de diferentes raios de influência ao longo do tempo para a instalação de sistemas solares domésticos no Brasil, observa-se que um raio de influência padrão de 1 km resultaria na instalação de aproximadamente 4.325 sistemas solares domésticos após 25 anos.

No entanto, constata-se um aumento significativo desse número à medida que o raio de influência é ampliado.

Em particular, com um raio de influência de 5 km, prevê-se que ocorrerão 8.984 instalações de sistemas solares domésticos após o mesmo período de 25 anos, representando mais do que o dobro do valor obtido com um raio de 1 km.

É importante ressaltar que o aumento do raio de influência da vizinhança resulta em crescimentos exponenciais nas instalações de sistemas solares domésticos.

Figura 41 - Comparação de instalações segundo o raio de influência



Fonte: Autor (2021)

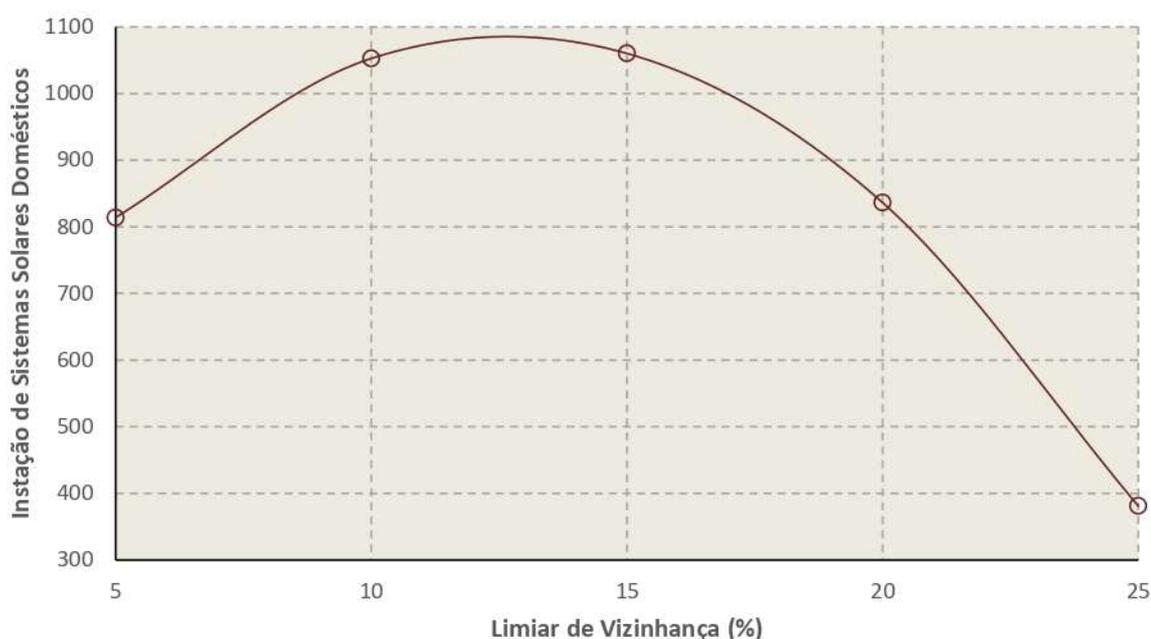
O limite de vizinhança é a porcentagem mínima de vizinhos em um determinado raio de influência que devem ter instalado os sistemas solares domésticos para que uma família considere fazer o mesmo. É uma medida de pressão social e como isso leva as famílias a instalar os sistemas solares domésticos conforme um número cada vez maior de vizinhos o faz. Ele também mostra o ponto de inflexão, acima do qual as instalações dos sistemas solares domésticos começam a cair.

A Figura 42 ilustra a relação entre o aumento do limiar da vizinhança e a instalação de sistemas solares domésticos. Ao adotar um limiar padrão de 5%, já foram registradas 828 instalações em domicílios. Dentre essas instalações, 350 foram realizadas por consumidores iniciais, enquanto os 478 domicílios restantes pertencem a consumidores secundários que foram influenciados a adotar essa tecnologia.

Identifica-se uma tendência de crescimento exponencial no número de instalações, totalizando 1.089 quando o limiar da vizinhança é ajustado para variar entre 12,5% e 15%. Todavia, nota-se uma significativa redução no número de instalações à medida que o limiar da vizinhança ultrapassa o patamar de 20%. Esse

fenômeno sugere a presença de um ponto crítico ou de saturação, onde o aumento no limiar da vizinhança resulta em uma resposta não linear na quantidade de instalações.

Figura 42 - Impacto do limiar de vizinhança



Fonte: Autor (2021)

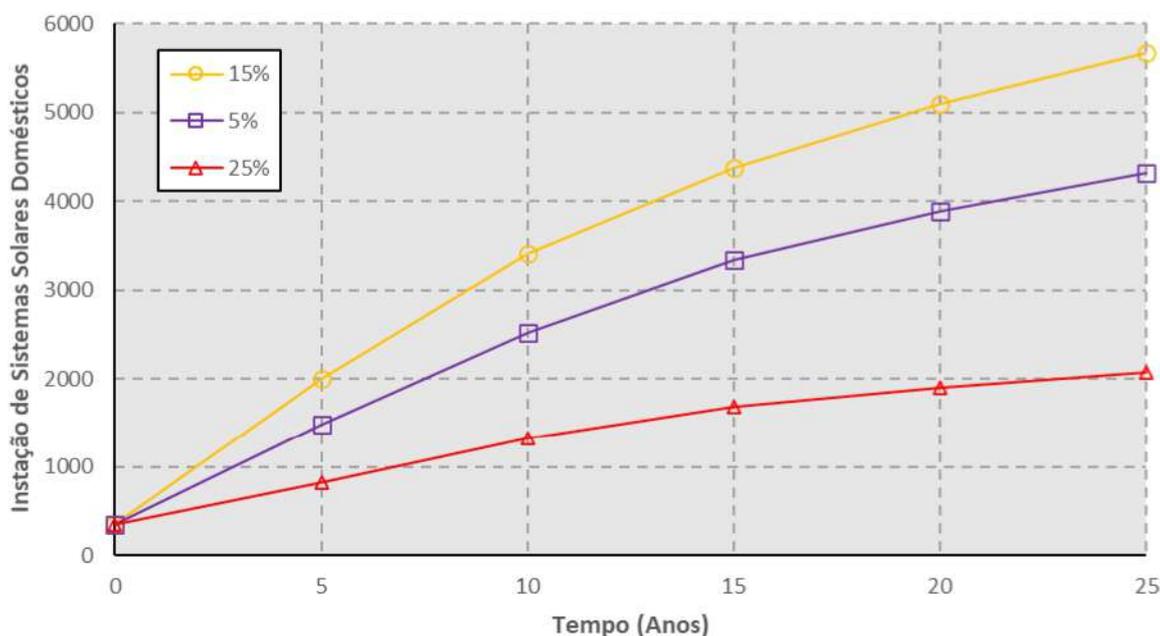
A Figura 43 apresenta uma comparação dos efeitos de diferentes limiares de vizinhança na implantação de sistemas solares domésticos no Brasil ao longo do tempo. Considerando um limiar padrão de 5%, estima-se que após 25 anos, 4.325 famílias tenham instalado esses sistemas.

No entanto, observa-se que o limiar ideal para a adoção de sistemas solares domésticos situa-se em uma faixa entre 12% e 15%, o que resultaria em um total de 5.666 domicílios com tais sistemas após o período analisado. Em contraste, a implementação de limiares superiores a 20% está correlacionada a uma redução na taxa de instalação, sendo que um limiar de 25% resultaria em apenas 2.076 instalações ao longo de um período de 25 anos. Este achado destaca a importância da definição criteriosa do limiar de adoção na promoção da expansão sustentável da energia solar residencial.

Considerando que o limiar de vizinhança é um fator determinante nas decisões de instalação de sistemas solares domésticos, pode-se argumentar que os valores

compreendidos entre 12,5% e 15% são considerados ideais. Esta faixa representa um equilíbrio ótimo entre a influência da vizinhança e a autonomia na tomada de decisão dos indivíduos em relação à adoção de sistemas solares. Dessa forma, valores dentro deste intervalo podem ser interpretados como indicativos de uma forte influência do comportamento dos vizinhos na decisão de um indivíduo, ao mesmo tempo em que ainda permite certa autonomia na decisão.

Figura 43 - Comparação de instalações segundo o limiar de vizinhança



Fonte: Autor (2021)

Em suma, a influência da vizinhança, especialmente por meio de amigos, vizinhos e/ou parentes que já instalaram esses sistemas, estimula a adoção dos sistemas solares domésticos por famílias que residem no mesmo raio de influência. Essa influência se dá pelo compartilhamento dos benefícios dos sistemas, por meio de comentários verbais e/ou escritos em redes sociais, que despertam o interesse dessas famílias em adquirir tais benefícios. Observou-se indiretamente que a posse de um sistema solar doméstico é associada a um sentimento de status e estabilidade financeira, o que influencia outras famílias a também adotarem essa tecnologia.

À medida que os comentários verbais e/ou escritos em redes sociais aumentam, juntamente com a visibilidade dos sistemas solares domésticos, as famílias são cada vez mais inclinadas a considerar a adoção dos mesmos. Isso

resultou em um aumento de instalações desses sistemas dentro de uma área específica, devido a influência mais intensa exercida pela vizinhança. Entre as estratégias para aumentar o desejo da vizinhança em adotar esses sistemas, inclui-se o uso de propagandas em diferentes meios de comunicação.

Os principais benefícios que levam as famílias a adotarem sistemas solares domésticos são a economia nas contas de energia, a autonomia na geração de energia, o acesso a uma fonte energética, as preocupações ambientais e o desejo de adquirir novas tecnologias.

Especificamente, os resultados mostram que o aumento do raio de influência e do limiar da vizinhança está diretamente associado ao aumento na adoção de sistemas solares domésticos. Isso ocorre porque, à medida que mais famílias instalam esses sistemas dentro de um determinado raio, chega-se a um ponto em que uma família começa a perceber os benefícios e decide adotá-los também.

Por fim, em relação ao método de pagamento, o estudo de caso revelou que a maioria das famílias opta por adquirir os sistemas solares domésticos por meio de parcelamentos diretos com as empresas fornecedoras e instaladoras. Por outro lado, é possível inferir que os primeiros instaladores se preparam para a aquisição e instalação desses sistemas, o que pode levá-los a adquiri-los à vista e possivelmente com um desconto maior.

6.2. ESTUDO DE CASO 2

As cidades enfrentam o desafio premente de confrontar as mudanças climáticas, cuja urgência transcende as fronteiras nacionais e se insere no contexto internacional. Para abordar eficazmente essa questão premente, torna-se imperativo empreender uma transformação substancial tanto nas infraestruturas de habitação já existentes quanto no comportamento dos habitantes urbanos. Entretanto, até o presente momento, as políticas governamentais têm se restringido principalmente à concessão de incentivos fiscais para a aquisição de materiais destinados à implementação de tecnologias mais sustentáveis. Observa-se, contudo, que tais medidas têm alcançado um êxito limitado e, conseqüentemente, não têm desempenhado um papel significativo no cumprimento das metas relacionadas às mudanças climáticas.

Em resposta a essa limitação, instituições e autoridades reconhecem a premente necessidade de desenvolver e apoiar planos abrangentes visando a redução do consumo de energia. Tais planos englobam estratégias que incluem a conscientização pública, a melhoria das infraestruturas habitacionais e a exploração de novas fontes energéticas. A implementação bem-sucedida desses projetos de transformação deve ser precedida por um planejamento meticuloso, sendo altamente recomendável começar por áreas residenciais, onde as melhorias possam ser efetivamente implementadas dentro de prazos viáveis. No entanto, ressalta-se que a implementação de um sistema direcionado à redução do consumo de energia exige uma compreensão abrangente das medidas adotadas, incluindo a promoção da conscientização energética em termos de eficiência no consumo de energia.

É também digno de nota que, atualmente, subsistem lacunas significativas no conhecimento específico relacionado aos agentes que exercem a maior influência na redução do consumo de energia, bem como em relação aos prazos de implementação e custos associados, o que complica ainda mais a tarefa de encontrar uma combinação adequada de medidas a serem adotadas. Além disso, os sistemas de apoio à tomada de decisão em eficiência energética existentes não têm priorizado as áreas residenciais.

Nesse contexto, a utilização da simulação de cenários complexos, por meio da abordagem conhecida como SMA, em conjunto com os princípios da análise multicritério, conforme explorado por Campisi, Gitto e Morea (2018), desempenha um

papel crucial no preenchimento dessas lacunas existentes, ao buscar combinações otimizadas para a seleção dos fatores que exercem a maior influência na redução do consumo de energia. Dentro desse arcabouço analítico, faz-se uso de um indicador denominado Índice de Viabilidade Ajustado (IVA), o qual representa uma métrica destinada a avaliar a viabilidade de projetos específicos. No contexto do presente estudo de caso, esse indicador é empregado com o objetivo específico de otimizar a redução do consumo de energia.

O modelo, desenvolvido por Campisi, Gitto e Morea (2018), pressupõe que a métrica em questão compreende seis fatores relacionados à redução do consumo de energia, conforme indicados na Equação 15.

$$IVA = \frac{\alpha \times \beta \times \gamma \times \delta}{C \times T} \quad (15)$$

Na Equação 15, α é empregado para denotar o impacto do projeto na diminuição do consumo energético, β quantifica a disponibilidade dos recursos necessários para a implementação do projeto, γ representa a complexidade inerente à implementação, δ mensura o risco associado à possibilidade de falha do projeto, C denota o custo financeiro associado à implementação do projeto e, finalmente, T simboliza o período requerido para a completa realização do projeto.

Assim, o presente estudo de caso tem como objetivo desenvolver um modelo abrangente e programável que simule o uso de energia e o potencial de redução energética. Esse modelo levará em consideração fatores climáticos, técnicos e comportamentais, simulando a busca pela diminuição do consumo energético. Tal abordagem permitirá a avaliação dos efeitos de medidas comportamentais e técnicas potenciais destinadas a alcançar esse objetivo.

6.2.1. COLETA DE DADOS

Uma pesquisa utilizando a raspagem de dados foi desenvolvida e realizada para coletar dados abrangentes que englobam as principais entidades envolvidas na redução do consumo de energia de uma área residencial. Essa técnica gerou três tópicos principais: (1) habitante, (2) habitação e (3) clima.

No que se refere aos habitantes, verifica-se que sua idade, profissão e comportamento exercem influência direta sobre a utilização de energia ao longo do tempo, sendo que o estilo de vida adotado por esses indivíduos, que engloba uma variedade de atividades que demandam energia por meio da iluminação e dos eletrodomésticos, tem impacto no consumo energético. Por sua vez, as características das habitações evidenciam desempenhar um papel crucial no consumo de energia, uma vez que sua forma e os materiais utilizados na construção e acabamento podem ampliar ou reduzir a demanda energética. Além disso, o clima desempenha um papel fundamental ao estabelecer as condições tanto para o consumo quanto para a produção de energia.

A partir do modelo conceitual, verifica-se que o consumo de energia das residências é mantido por meio da interação entre as características das moradias e seus habitantes, além das influências climáticas. Ressalta-se que a coleta de dados foi realizada em páginas da web disponíveis na Internet durante o período compreendido entre janeiro de 2023 e junho de 2023.

EXTRAÇÃO PREDITIVA DE DADOS DE TEXTO

Nesta fase da pesquisa, conduziu-se à coleta de um extenso conjunto de dados textuais por meio da aplicação de técnicas de extração preditiva. Foram utilizadas fontes de dados como páginas da *web* governamentais, associações e redes sociais *online*, com o propósito de retratar o consumo energético em áreas residenciais em um contexto nacional. Os dados foram submetidos à análise e seleção criteriosas, eliminando-se informações curtas e duplicadas, enquanto todas as *tags* relevantes foram preservadas, visando à identificação de tendências mais amplas. Essa abordagem de coleta em larga escala foi adotada para a obtenção de conjuntos de dados que, posteriormente, foram empregados na modelagem de tópicos, com o objetivo de identificar as principais tendências relacionadas à redução do consumo de energia.

Portanto, o presente estudo empregou técnicas de extração preditiva de dados textuais provenientes de fontes *online* relevantes, permitindo a análise e modelagem de tópicos para identificação das principais tendências relacionadas ao consumo de energia em áreas residenciais. Ao término desta fase, foi coletado um total de 13.098 registros de dados, consolidando assim uma base de dados substancial para as

etapas subsequentes da pesquisa. A Tabela 24 ilustra a descrição dos dados.

Tabela 24 - Descrição dos dados de texto coletados

Páginas da web	URL	Número de Informações	Período da Coleta		Frequência Mensal das Publicações
			Início	Fim	
Caixa Econômica Federal	https://www.caixa.gov.br	2.538	Abr./2023	Jun./2023	23,5
Instituto de Pesquisa Econômica Aplicada	https://www.ipea.gov.br	957	Abr./2023	Jun./2023	8,9
Instituto Nacional de Pesquisas Espaciais	http://www.inpe.br	510	Abr./2023	Jun./2023	4,7
Empresa de Pesquisa Energética	https://www.epe.gov.br	1.385	Abr./2023	Jun./2023	12,8
WWF Brasil	https://www.wwf.org.br	123	Abr./2023	Jun./2023	1,1
Ministério do Desenvolvimento Regional	https://www.gov.br/mdr/pt-br	940	Abr./2023	Jun./2023	8,7
Instituto Brasileiro de Geografia e Estatística	https://www.ibge.gov.br	1.252	Abr./2023	Jun./2023	11,6
Instituto Nacional de Meteorologia	https://portal.inmet.gov.br	840	Abr./2023	Jun./2023	7,8
Agência Nacional de Energia Elétrica	http://www.aneel.gov.br	1.037	Abr./2023	Jun./2023	9,6
Greenpeace Brasil	https://www.greenpeace.org/brasil	302	Abr./2023	Jun./2023	2,8
Facebook	https://www.facebook.com	3.214	Abr./2023	Jun./2023	29,7

Fonte: Autor (2023)

PRÉ-PROCESSAMENTO DE DADOS DE TEXTO

Neste estudo, foi realizada uma análise abrangendo um total de 4.124.837 *tokens* extraídos de 13.098 conjuntos de informações. Com o objetivo de aprimorar a qualidade dos dados textuais, foi empregada uma técnica de redução de ruído que consistiu na remoção de pontuação e de palavras pertencentes à lista de bibliotecas Python, com destaque para a biblioteca NLTK. Essas medidas resultaram em uma redução de aproximadamente 15,86% do número total de palavras presentes nos conjuntos de dados originais.

A Figura 44 exibe uma representação visual da nuvem de palavras coletadas, acompanhada da respectiva frequência de cada termo. Esse recurso gráfico possibilita identificar quais palavras se destacaram em termos de ocorrência nos dados que foram submetidos à análise. A utilidade dessa representação visual reside na sua capacidade de simplificar a interpretação de tendências e padrões de palavras recorrentes nos dados em questão. A análise de frequência é valiosa para a compreensão do conteúdo e das informações essenciais presentes nos dados coletados no presente estudo.

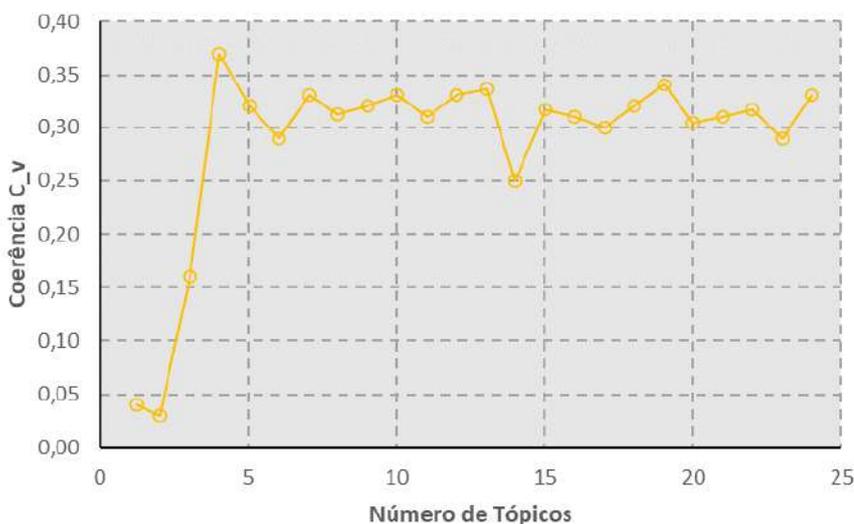
MODELAGEM POR TÓPICOS E ANÁLISE DE SIMILARIDADE

No estudo em questão, foi empregada a técnica de LDA com o objetivo de determinar o número ideal de tópicos (K) com base em métricas de coerência, nomeadamente C_V e C_{UMass} . Essas métricas foram calculadas após o treinamento dos modelos utilizando valores de $\alpha = 0,01$ e $\beta = 0,13$.

Nesse contexto, observou-se que a métrica C_V apresentou resultados superiores às medidas de C_{UMass} , o que indicou uma faixa de valores de K com maior coerência de acordo com C_V . No entanto, considerando que as pontuações de C_V tendem a aumentar à medida que o número de tópicos cresce, foram determinados os valores ideais de K a partir da identificação do ponto em que os retornos adicionais não compensam mais a complexidade adicional do modelo.

Essas determinações foram realizadas utilizando uma abordagem heurística e levando em consideração as flutuações de C_{UMass} . A Figura 45 ilustra o processo de determinação do número ótimo de tópicos K usando as medidas de coerência C_V e C_{UMass} .

Figura 45 - Número de tópicos (a) Métrica: C_V (b) Métrica: C_{UMass}



(a)



(b)

Fonte: Autor (2023)

Observa-se na Figura 45 (a) que a métrica C_V atingiu um pico local em 3, enquanto na Figura 45 (b) C_{UMass} apresentou um platô de pontuação a partir desse ponto. Esses resultados confirmam que o modelo identificou três tópicos principais: (1) habitante, (2) habitação e (3) clima. Além disso, verificou-se que as páginas da *web* com conteúdo mais extensos tendiam a apresentar uma maior densidade documento-tópico (α). As melhorias realizadas no ajuste do modelo, visando aprimorar o desempenho nas medidas de coerência, são apresentadas na Tabela 26.

Tabela 26 - Ajuste e melhoria das medidas de coerência

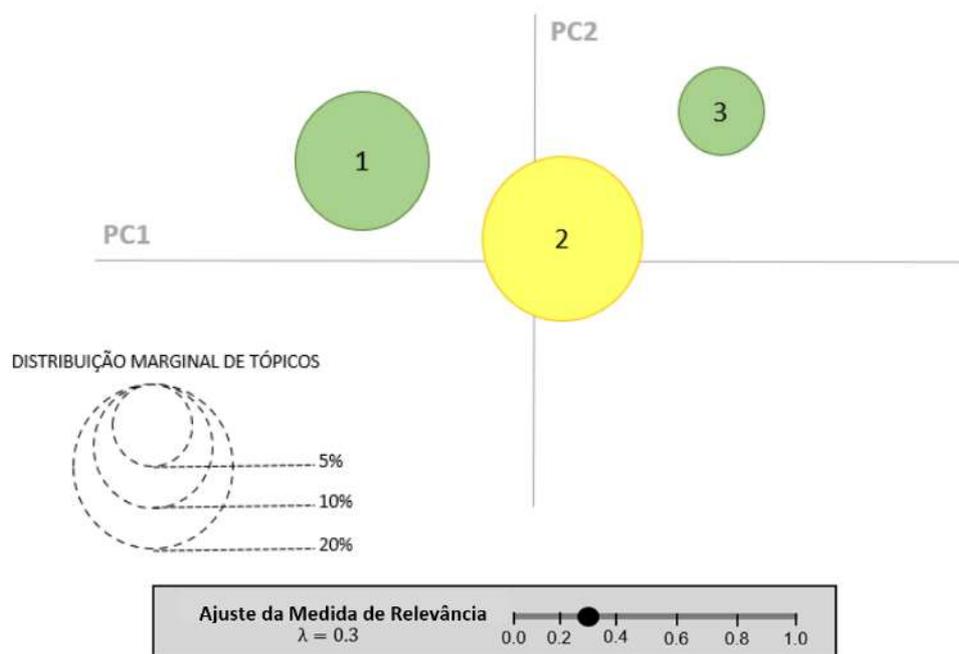
Dados	Valores
Número de tópicos (K)	3
Densidade de tópico de documento (α)	0,83
Densidade de tópico de palavras (β)	0,82
C_{UMass} (modelo original)	-1.312
C_{UMass} (modelo ajustado)	-1.151
Melhoria C_{UMass} (%)	+14,9%
C_V (modelo original)	0,346
C_V (modelo ajustado)	0,319
Melhoria C_V (%)	+15,7%

Fonte: Autor (2023)

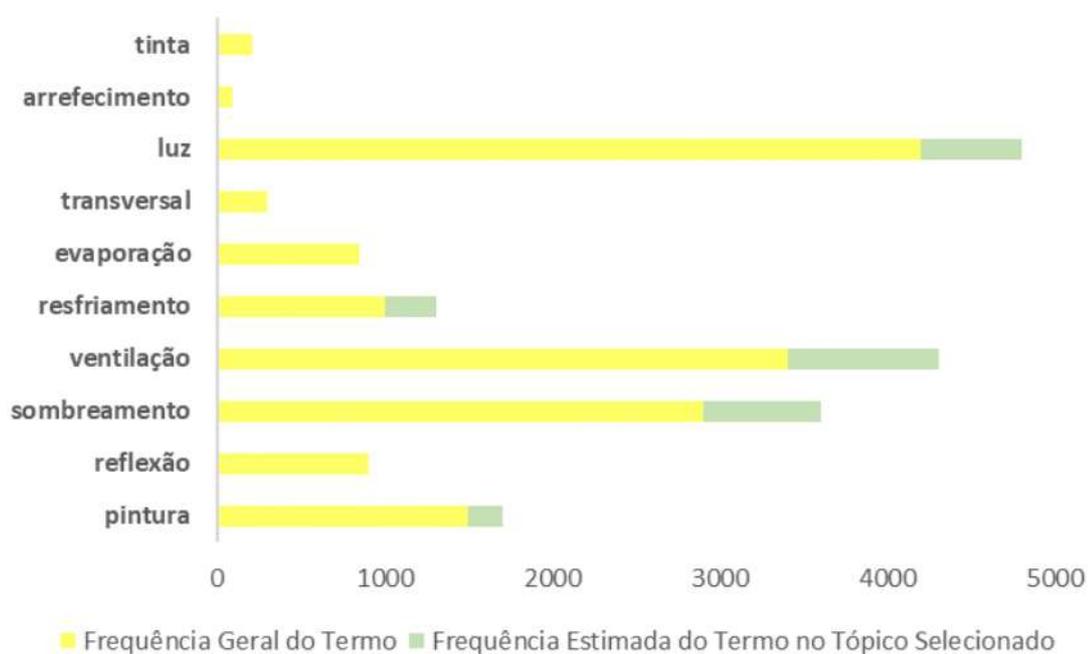
Visando obter informações relevantes e evitar detalhes específicos, os tópicos do modelo LDA foram categorizados semanticamente em três tópicos principais: (1)

habitante, (2) habitação e (3) clima, conforme ilustrado na Figura 46.

Figura 46 - Layout pyLDAvis (a) Distância entre tópicos. (b) Palavras por tópico



(a)



(b)

Fonte: Autor (2023)

Utilizando as métricas de relevância, busca-se obter *insights* sobre como cada uma das principais categorias impactam no estudo de caso. Posteriormente, a métrica de distância de Hellinger foi empregada para quantificar as distâncias e as inter-relações entre os tópicos extraídos. Com base na análise das métricas de relevância, pode-se observar na Figura 46 (a), que há uma interpretação clara e concisa sobre a divisão dos principais tópicos. Essa constatação foi destacada e exemplificada na Figura 46 (b), onde são fornecidas informações relacionadas à frequência do termo geral ocorrido na categoria 2, que é relativa às habitações. Essa abordagem busca compreender quais as principais categorias que podem influenciar na simulação do estudo de caso.

Ao término dessa fase, as 10 palavras mais relevantes dos tópicos modelados aprimoraram a interpretação dos três tópicos, conforme apresentado na Tabela 27.

Tabela 27 - Modelagem de tópicos

No.	Tópico	Freq. (%)	C_V	C_{UMass}	As 10 palavras mais relevantes ($\lambda = 0,3$)
1	Habitante	37,82	0,55	- 0,95	{idade, aposentado, conscientização, desempregado, estudante, vida, trabalho, natural, integral, empregado}
2	Habitação	29,51	0,49	- 1,08	{pintura, reflexão, sombreamento, ventilação, resfriamento, evaporação, transversal, luz, arrefecimento, tinta}
3	Clima	14,77	0,41	- 0,88	{clima, verão, sol, chuva, frio, temperatura, raio, umidade, poluição, inverno}

Fonte: Autor (2023)

A partir dos dados coletados, uma ampla gama de informações foi obtida, as quais serão detalhadas, agrupadas e analisadas a seguir.

Com base nos dados referentes às habitações em áreas residenciais, verificou-se que as casas térreas, que abrangem entre 3 e 5 cômodos e possuem área variando de 100 a 300 metros quadrados, ocupam a primeira posição. Na sequência, encontram-se os apartamentos, contendo entre 2 e 4 cômodos, com tamanho entre 40 e 150 metros quadrados. Em terceiro lugar estão as quitinetes, caracterizadas pela presença de 1 a 2 cômodos, ou seja, cômodos integrados, e com tamanho entre 20 e 40 metros quadrados. Outros tipos de moradia constituem menos de 2% das opções disponíveis. E por fim foram estimados os valores médios para cada tipo de habitação. Para informações mais detalhadas, ver a Tabela 28.

Tabela 28 - Tipos de Habitação

Tipo	Quantidade	Nº Cômodos	Cômodos Frequentes	Tamanho (m ²)	Valor (R\$)
Casa	58%	3 a 5	Quarto, Sala, Cozinha, Banheiro	100 a 300	200.000 a 800.000
Apartamento	25%	3 a 4	Quarto, Sala, Cozinha, Banheiro	40 a 150	150.000 a 600.000
Quitinete	3%	1 a 2	Quarto, Banheiro	20 a 40	80.000 a 200.000

Fonte: Autor (2023)

No que se refere aos sistemas de iluminação e dispositivos elétricos encontrados nas habitações, foi constatado que 99% delas utilizam lâmpadas fluorescentes, tendo cada cômodo ao menos um ponto de luz. No que se refere aos dispositivos elétricos, o conjunto mínimo encontrado em cada residência inclui televisão, geladeira e chuveiro. Essas informações são detalhadas na Tabela 29.

Tabela 29 - Informações sobre Iluminação e Aparelhos Elétricos

Dispositivos	Presença nas Habitações	Localização nos Cômodos	Potência (W) por unidade
Lâmpada Fluorescente	99%	Todos	60
Televisão	98%	Sala	250
Geladeira	96%	Cozinha	300
Chuveiro	96%	Banheiro	5.500
Ventilador	79%	Sala, Quartos	80
Computador	72%	Sala, Quartos	300
Ar-condicionado	70%	Sala, Quartos	1.500
Aquecedor	51%	Sala, Quartos	2.000

Fonte: Autor (2023)

Com relação aos habitantes, os dados revelaram que a maioria das habitações abriga famílias compostas por três membros, correspondendo a 31% da população. Em relação à faixa etária, destaca-se que a maior proporção das famílias é composta por indivíduos entre 25 e 54 anos, abrangendo 40% do total. Quanto ao estilo de vida, a soma dos indivíduos empregados em regime integral e parcial representa 43% da população. Por outro lado, a parcela de aposentados corresponde a 24% do total. Essas informações estão detalhadas na Tabela 30.

Tabela 30 - Informações sobre os Habitantes

Composição das Famílias		Faixa Etária		Estilo de Vida	
Nº Pessoas	Quantidade	Idades	Quantidade	Tipos	Quantidade
1 pessoa	5%	0-17 anos	27%	Tempo Integral	28%
2 pessoas	24%	18-24 anos	10%	Tempo Parcial	15%
3 pessoas	31%	25-54 anos	40%	Estudante	12%
4 pessoas	26%	55-64 anos	15%	Aposentado	24%
5 pessoas ou mais	14%	65+ anos	8%	Desempregado	21%

Fonte: Autor (2023)

Os dados obtidos durante a etapa de coleta de informações climáticas no Brasil indicam que a temperatura média varia em 8°C entre a estação mais fria e a estação mais quente. Em relação à incidência dos raios solares, a estação do verão registrou, em média, 7 horas diárias de exposição solar, representando o período com maior número de horas de sol entre todas as estações. A média de nebulosidade ao longo do ano situou-se em torno de 50%. Quanto à umidade, observou-se uma variação considerável entre as estações, atingindo 80% no verão e 25% no inverno. Com relação à poluição atmosférica, observou-se que apresentou níveis mais baixos no verão. Para uma análise mais detalhada, recomenda-se consultar a Tabela 31.

Tabela 31 - Informações sobre o Clima

Estação do Ano	Média da Temperatura	Incidência dos Raios Solares	Média da Radiação Solar	Média de Nebulosidade	Média de Umidade	Média de Poluição Atmosférica
Verão	28°C	7 horas de sol por dia	5,2 kWh/m ²	45%	80%	14 ppm
Outono	25°C	6 horas de sol por dia	4,5 kWh/m ²	50%	45%	26 ppm
Inverno	20°C	5 horas de sol por dia	3,5 kWh/m ²	60%	25%	33 ppm
Primavera	26°C	6 horas de sol por dia	4,8 kWh/m ²	50%	75%	21 ppm

Fonte: Autor (2023)

6.2.2. MODELAGEM DO CENÁRIO UTILIZANDO SMA

Neste estudo de caso, um modelo foi desenvolvido utilizando o *software* NetLogo, com o objetivo de se tornar uma ferramenta para a modelagem dos principais elementos envolvidos na análise da obtenção de uma redução energética em um bairro residencial já estabelecido. O modelo considera as características dos residentes, seus comportamentos em relação ao uso de energia, as características das residências existentes (levando em consideração apenas as moradias já construídas), as condições climáticas locais e o potencial de geração de energia limpa na área em questão. Por meio da simulação das interações entre esses fatores, busca-se compreender os efeitos resultantes de diferentes possibilidades.

Os dados obtidos na etapa de coleta de dados foram utilizados para criar indicadores e fornecer dados para o modelo. Esses dados são empregados na simulação dos fatores mais eficazes para a redução do consumo energético, assim como combinações desses fatores visando alcançar a redução energética desejada.

No modelo, foram criados os seguintes agentes:

- a representação do ambiente da simulação contendo as habitações, incluindo suas características e o clima da região.
- os aparelhos elétricos e a iluminação, incluindo suas características.
- os habitantes, incluindo suas características e comportamentos.
- um observador central que determina as melhores estratégias para alcançar a redução do consumo de energia.

As características dos agentes autônomos, no contexto do estudo de caso apresentado, foram modeladas no ambiente do NetLogo conforme detalhado a seguir:

- **Autonomia:**
 - Hierárquica: os agentes operam em diferentes níveis hierárquicos, com capacidade para tomar decisões autônomas em níveis inferiores (por exemplo, ajustando o uso de energia em residências individuais) e requerer supervisão em níveis superiores (por exemplo, na formulação de políticas públicas relacionadas à gestão de energia).
 - Social: os agentes incorporam regras de interação social, como a consideração das preferências de consumo de energia dos habitantes sobre seu próprio comportamento.
 - Execução: os agentes possuem autonomia para realizar ações específicas relacionadas à gestão de energia, como ajustar o uso de dispositivos eletrônicos em residências.
- **Mobilidade:**
 - Estática: os agentes permanecem em posições fixas e não requerem mobilidade física para a simulação.
- **Reatividade:**
 - Contextual: os agentes respondem a mudanças ambientais, como variações climáticas ou alterações nos padrões de comportamento dos habitantes, ajustando seu consumo de energia em conformidade.
- **Proatividade:**
 - Alta: os agentes têm capacidade para antecipar tendências de consumo de energia com base em históricos de dados e tomar iniciativas proativas

para reduzir o consumo, como sugerir práticas de eficiência energética aos habitantes.

- Comunicação:
 - Adaptativa: Os agentes adaptam sua comunicação de acordo com as condições do ambiente, incluindo a capacidade de compartilhar informações sobre padrões de consumo de energia entre os habitantes.
- Habilidade Social:
 - Alta: os agentes são capazes de interagir tanto com os habitantes quanto com outros agentes, compartilhando informações sobre práticas de consumo de energia e coordenando esforços para reduzir o consumo.
- Cooperação:
 - Ativa: os agentes colaboram entre si e com os habitantes para alcançar objetivos compartilhados relacionados à gestão de energia, como coordenar o uso de eletricidade durante os horários de pico de demanda.
- Aprendizagem:
 - Contextualizada: os agentes aprendem com experiências passadas e adaptam seu comportamento de acordo com as condições ambientais específicas, ajustando estratégias de economia de energia em resposta a mudanças climáticas ou variações nos padrões de consumo dos habitantes.

O modelo desenvolvido está situado em um ambiente que considera as moradias como já construídas e não onde haverá a construção a partir do zero. Além disso, o ambiente apresenta o clima dessa região, considerando as condições climáticas ocorridas nas diferentes estações do ano.

A modelagem também possibilita aos habitantes a realização de uma atividade por vez, inclui as atividades que estão relacionadas ao uso de dispositivos que utilizam energia elétrica e distingue as diferentes atividades realizadas. Os habitantes e os aparelhos elétricos atuam em um ambiente residencial composto por cômodos. Esses cômodos afetam a mobilidade e o número de locais onde os habitantes podem estar e, conseqüentemente, o consumo de energia elétrica.

Os aparelhos elétricos foram modelados como agentes estáticos controlados por agentes habitantes, que respondem às ações desses habitantes para alterar seu

estado de ligado para desligado ou vice-versa. Em cada momento, cada aparelho elétrico registra a quantidade de energia elétrica consumida com base em seu estado.

A modelagem do ambiente de simulação foi definida como T, H, Ed , onde:

- T é o tempo da simulação em 1 ano, que é definido por t, d, l, av, pa . Onde t é um intervalo de tempo de 10 minutos em cada dia, d é o dia composto por 24 horas, l é a quantidade de luz natural do dia, av representa áreas verdes e pa a poluição atmosférica.
- H é o tipo de habitação, que é definida por c, m, v, a, p, Ac, Oc . Onde c é conjunto de cômodos, m é a metragem do cômodo, v é o conjunto de janelas e o sistema de vedação das mesmas, a representa o conjunto de acabamento (pintura, piso, telhado), p é o valor médio da residência, Ac é o conjunto de aparelhos do cômodo e Oc é o conjunto de habitantes que estão no cômodo.
- Ed é o conjunto de eletrodomésticos, que é definido pelo conjunto e, p, Oa, Ea . Onde e é a quantidade de energia usada quando o dispositivo está ligado, p é o cômodo em que o aparelho está, Oa é o conjunto de habitantes que usam o eletrodoméstico e Ea é a matriz de consumo do eletrodoméstico ao longo de um ano inteiro.

A modelagem dos habitantes inclui diferentes perfis de comportamento com relação ao consumo de energia elétrica para assim verificar a diminuição ou aumento gerado e, com isso, detectar hábitos que acarretam um desperdício de energia elétrica nas residências. Os atributos que modelam essa consciência de consumo de energia elétrica, juntamente com as atividades dos habitantes e a hora, são usados para controlar quando os habitantes ligam ou desligam os aparelhos elétricos e as luzes. As idades dos habitantes e os tipos de emprego são fornecidos como dados de entrada para o modelo. A cada etapa, os habitantes mudam o estado do ambiente, mudam a sua localização e usam os eletrodomésticos.

Para efetuar a modelagem dos habitantes foram analisadas e realizadas as seguintes considerações:

- Comportamento do Habitante: antes de simular a ocupação, as rotinas de trabalho e os padrões de ocupação são atribuídos a cada habitante. A cada momento, o habitante seleciona um novo estado de ocupação. O agente primeiro seleciona um novo estado em função de seu padrão de ocupação,

estado anterior e hora do dia. A simulação distingue entre atividades que podem ser realizadas por um habitante de cada vez e atividades que podem ser realizadas por todos os moradores e podem ser compartilhadas. Quando o residente está ativo, ele pode selecionar iniciar uma atividade pessoal/grupo ou diminuir a duração de uma atividade em andamento. Essa forma de modelagem permite que o morador execute mais de uma atividade por vez. A decisão de realizar uma atividade é baseada na idade do habitante, tipo de emprego, hora do dia e o dia da semana. A cada nova atividade que for selecionada para ser realizada, o agente seleciona a duração da atividade.

- **Localização do Habitante:** No contexto da simulação, a localização do habitante é determinada com base em seu estado atual, que pode ser Ativo ou Dormindo. Estes estados indicam se o habitante está dentro de casa e ocupado em atividades cotidianas ou em repouso. A atribuição de cômodos específicos da residência para cada atividade é uma prática adotada. Quando um habitante realiza múltiplas atividades simultaneamente, a sua localização pode ser alterada dinamicamente entre os cômodos associados a cada atividade. Esse processo de alternância de localização ocorre em intervalos de tempo determinados, refletindo a movimentação do habitante dentro da casa durante suas atividades diárias.
- **Consciência do Consumo de Energia Elétrica do Habitante:** a consciência do consumo de energia elétrica por parte do habitante é determinada com base em seu perfil de consumidor, atribuindo-se um valor entre 0 e 100 a esse atributo. Esse valor é utilizado para calcular a probabilidade de o morador adotar medidas de economia de energia elétrica, tais como desligar dispositivos quando não estão em uso. O cálculo é realizado por meio de uma distribuição normal específica para cada tipo de consumidor. Quando o habitante inicia uma atividade, ele liga o aparelho elétrico associado a essa atividade. Ao término da atividade e levando em consideração a consciência do consumo de energia elétrica do morador, ele pode optar por desligar o aparelho ou mantê-lo ligado. Além disso, o habitante pode se comunicar com outros moradores que possam estar utilizando o mesmo aparelho elétrico simultaneamente para decidir se o desliga. A ação de desligar os aparelhos elétricos também é realizada sempre que um residente adentra um cômodo e encontra aparelhos ligados, porém não utilizados. Por outro lado, o controle de ligar/desligar as luzes difere do uso dos

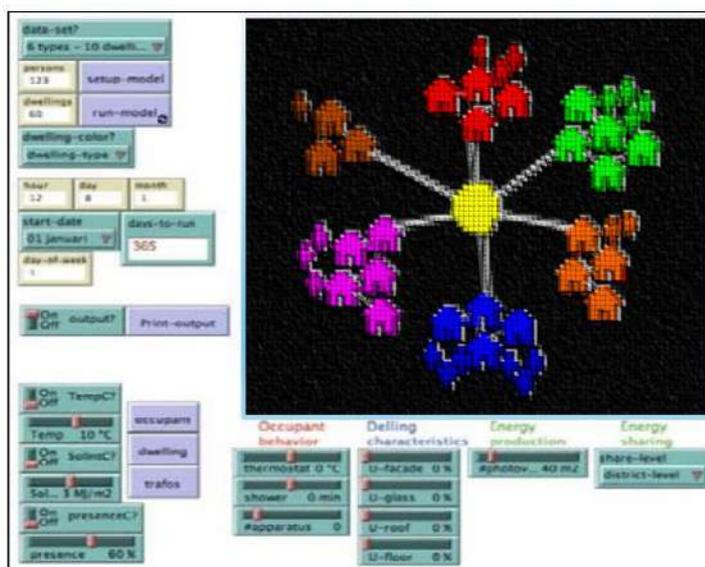
aparelhos elétricos, pois depende da quantidade de luz natural disponível durante o dia e da localização dos habitantes. Cada vez que o morador se encontra em um ambiente, ele pode decidir acender a luz com base na iluminação natural presente naquele cômodo. Quando um residente deixa um ambiente, ele verifica a presença de outros ocupantes e, com base em sua conscientização sobre o consumo de energia elétrica, toma a decisão de desligar ou não a luz.

A partir dessas considerações, o agente habitante O é modelado e definido pelo conjunto $i, o, aa, ap, dr, ce, Tpc, Lc$ onde:

- i representa a idade do habitante.
- o define a ocupação do habitante e pode ser: empregado, desempregado, estudante, aposentado.
- aa fornece a atividade atual que o habitante está executando em relação ao(s) eletrodoméstico(s).
- ap indica a atividade prévia que o habitante estava realizando em relação ao(s) eletrodoméstico(s).
- dr informa a duração que o habitante executou no eletrodoméstico.
- ce representa o comportamento energético do habitante, que pode ser gastador, normal ou econômico.
- Tpc fornece o conjunto de tarefas pessoais contínuas e possíveis que o habitante pode executar.
- Lc indica a localização do cômodo em que o habitante se encontra.

Através da condução de simulações utilizando um modelo baseado em agentes autônomos, que incorporou os elementos de habitações, habitantes e condições climáticas, foi possível obter uma ampla gama de informações relacionadas ao consumo energético em uma área residencial preexistente. Estas informações foram então submetidas a um processo de detalhamento, agrupamento e análise, visando uma compreensão mais aprofundada dos padrões e comportamentos observados. Esse procedimento permitiu a identificação de tendências significativas, relações causais e possíveis estratégias de otimização para a gestão eficiente da energia nesse contexto residencial.

Figura 47 - Visão do ambiente de simulação



Fonte: Autor (2023)

A Figura 47 mostra a simulação em execução, na qual a programação foi configurada para abranger um período de 1 ano. Os diversos agentes associados às habitações, habitantes e dispositivos elétricos foram representados em tons coloridos, de forma proporcional aos resultados obtidos na etapa anterior. Ao fundo, em tom escuro, encontra-se o ambiente no qual o clima interage com os demais agentes.

Inicialmente, a simulação conduzida neste estudo de caso teve como objetivo primordial determinar o consumo específico de energia por pessoa, a fim de estabelecer uma base sólida para a subsequente análise dos impactos dos principais grupos influentes no contexto da redução do consumo de energia. Essa fase inicial desempenhou um papel fundamental na aplicação de uma formulação específica, denominada Equação 16, fundamentada em dados confiáveis, com o propósito de calcular o consumo anual de energia.

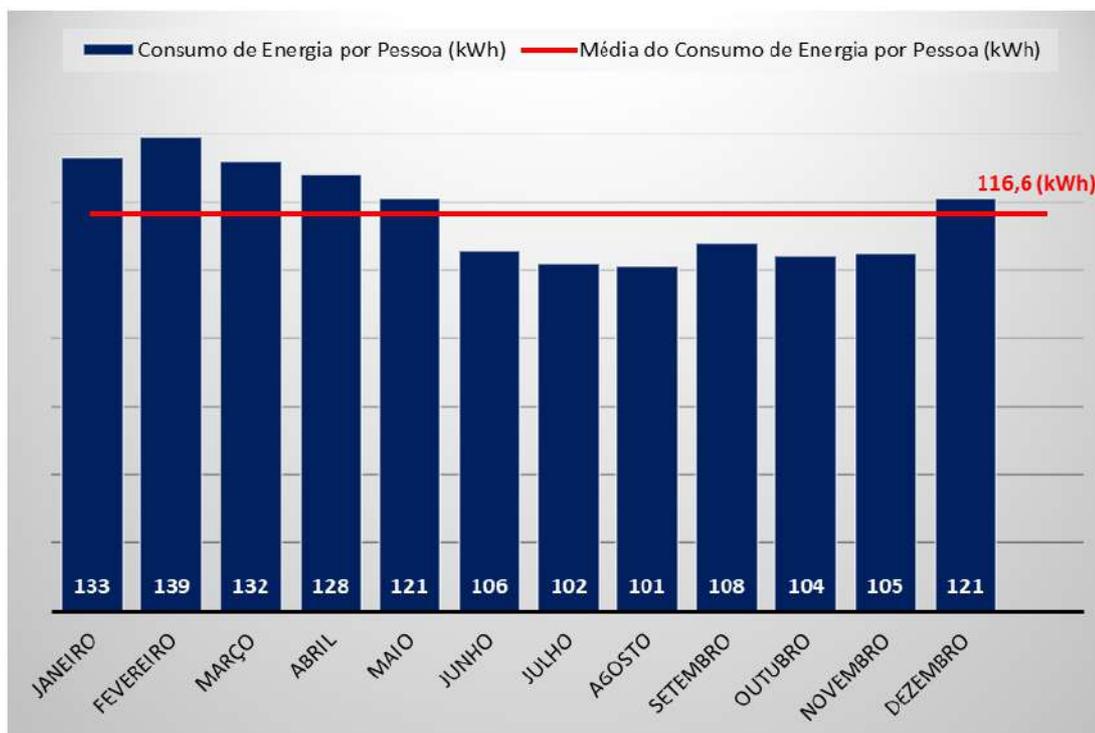
$$G_{MEP} = \frac{C_{TE}}{N_P} \quad (16)$$

Na Equação 16, G_{MEP} representa o gasto médio de energia por pessoa em kWh por mês, C_{TE} é o consumo total de energia em kWh e N_P indica o número de pessoas.

A análise cuidadosa dos dados coletados permitiu chegar a um resultado robusto, revelando a magnitude do consumo energético individual ao longo do ano. Com base na Figura 48, que apresenta de forma gráfica o perfil do consumo de

energia ao longo dos meses, foi possível traçar uma linha representativa da média de energia consumida nesse período, que se estabeleceu em 116,6 kWh/ano/pessoa.

Figura 48 - Consumo médio de energia por pessoa anualmente



Fonte: Autor (2023)

Com base nos dados previamente coletados, o estudo de caso teve a execução de novas simulações com a finalidade de investigar minuciosamente os fatores predominantes que exerceram influência no consumo energético em cada um dos três amplos grupos categorizados, a saber: habitação, habitantes e clima. O intuito dessa análise detalhada foi obter um entendimento mais aprofundado sobre as variáveis que desempenharam papéis significativos no consumo de energia dentro de cada grupo.

Nesse sentido, primeiro considerou-se a influência do consumo de energia e os custos aproximados associados à sua implementação em habitações, destacando a influência de diversos fatores na redução do consumo energético residencial, bem como os custos relativos à implementação de cada um.

Em primeiro lugar, percebeu-se que o fator de isolamento térmico atingiu o pico mais elevado de influência, atingindo um decréscimo de 9% no consumo de energia. Este resultado sinaliza que o isolamento térmico exerceu um papel crítico na promoção da eficiência energética, contribuindo de maneira substantiva para a

redução do consumo.

Em seguida, identificou-se o fator ventilação natural como o segundo mais influente, com um pico de 8%. A ventilação natural é amplamente reconhecida por sua capacidade de proporcionar um ambiente mais confortável e, simultaneamente, promover a economia de energia, e essa constatação reiterou a sua importância neste contexto.

Não obstante, vale ressaltar que o fator sombreamento também atingiu um pico de 8%, o que denota a sua relevância na mitigação do consumo de energia. Importa frisar que, embora a magnitude percentual seja idêntica à da ventilação natural, o período de implementação do sombreamento revelou-se substancialmente mais protraído (47 meses em contraposição aos 7 meses da ventilação natural).

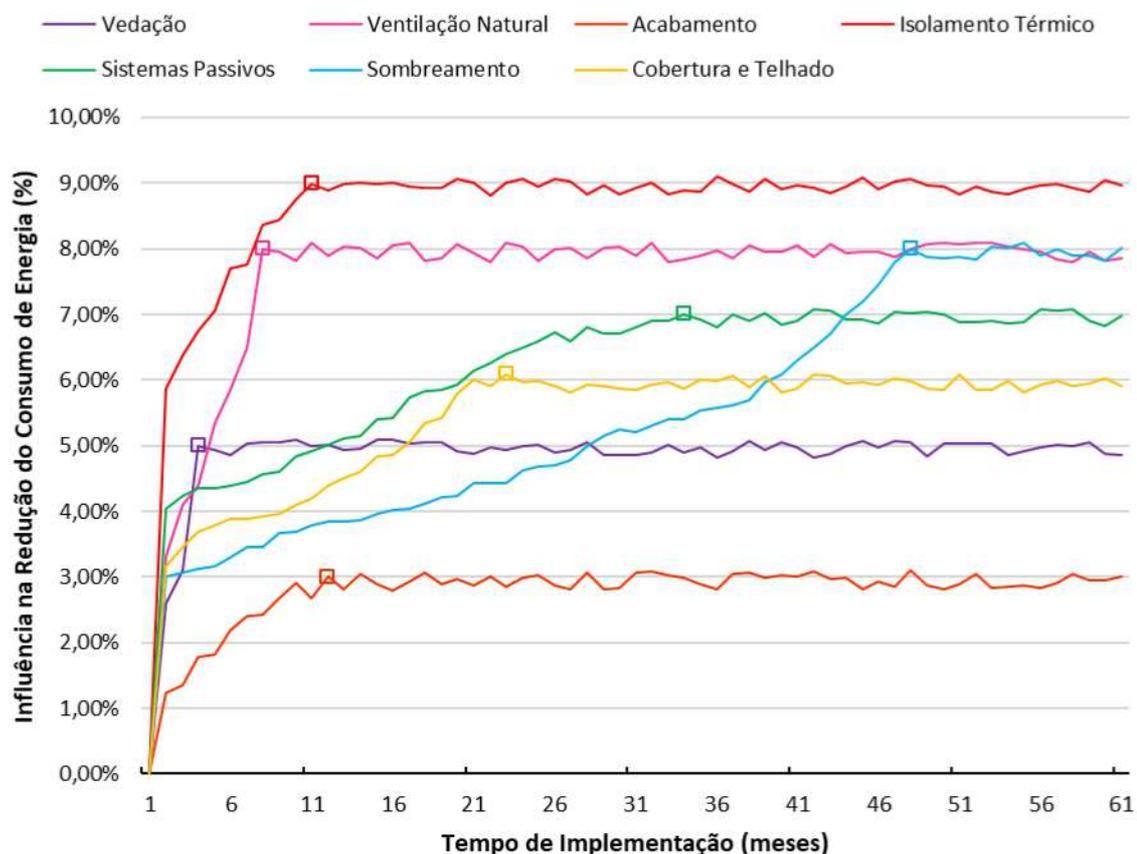
No tocante à celeridade de implementação, os fatores que atingiram o apogeu mais rapidamente foram a vedação (5% em 3 meses) e o acabamento (3% em 11 meses). Tal observação sugere que essas medidas puderam ser instauradas de maneira relativamente célere, proporcionando benefícios imediatos em termos de economia de energia.

Por outro ângulo, o fator sistemas passivos alcançou um pico de 7%, porém demandou um período de implementação notavelmente mais longo, totalizando 33 meses. Isso sugere que a instauração de sistemas passivos se caracteriza por uma complexidade e duração consideráveis, não obstante o seu impacto substancial na eficiência energética.

Por último, vale ressaltar que o fator referente à cobertura e ao telhado alcançou um pico de 6% em sua eficácia, demandando um período de implementação de 20 meses. Embora não tenha se destacado pela rapidez na implementação ou pela magnitude máxima de impacto, sua contribuição foi significativa na mitigação do consumo de energia. Este aspecto evidencia a importância relativa de cada componente no contexto do sistema avaliado, enfatizando a necessidade de considerar múltiplos fatores ao projetar intervenções para otimização do consumo energético.

A Figura 49 exibe um gráfico mais detalhado, ilustrando a importância de considerar esses fatores ao renovar uma habitação para aumentar sua eficiência energética e reduzir o consumo de energia.

Figura 49 - Fatores de influência do consumo relacionados a habitação



Fonte: Autor (2023)

Na esfera do estudo relacionado à habitação, procedeu-se à simulação e subsequente análise dos custos aproximados. O ponto culminante da porcentagem do custo de implementação, um marco essencial na análise, delineia o ponto no qual os custos atingiram sua magnitude máxima em relação ao valor da habitação. No presente conjunto de dados, constatou-se que o fator acabamento alcançou o ápice mais proeminente, atingindo a proporção de 8% em relação ao valor total da habitação. Isso denota que o aporte financeiro destinado aos acabamentos representou a parcela mais substancial dos recursos alocados no contexto do processo construtivo, considerando o montante global do preço da habitação.

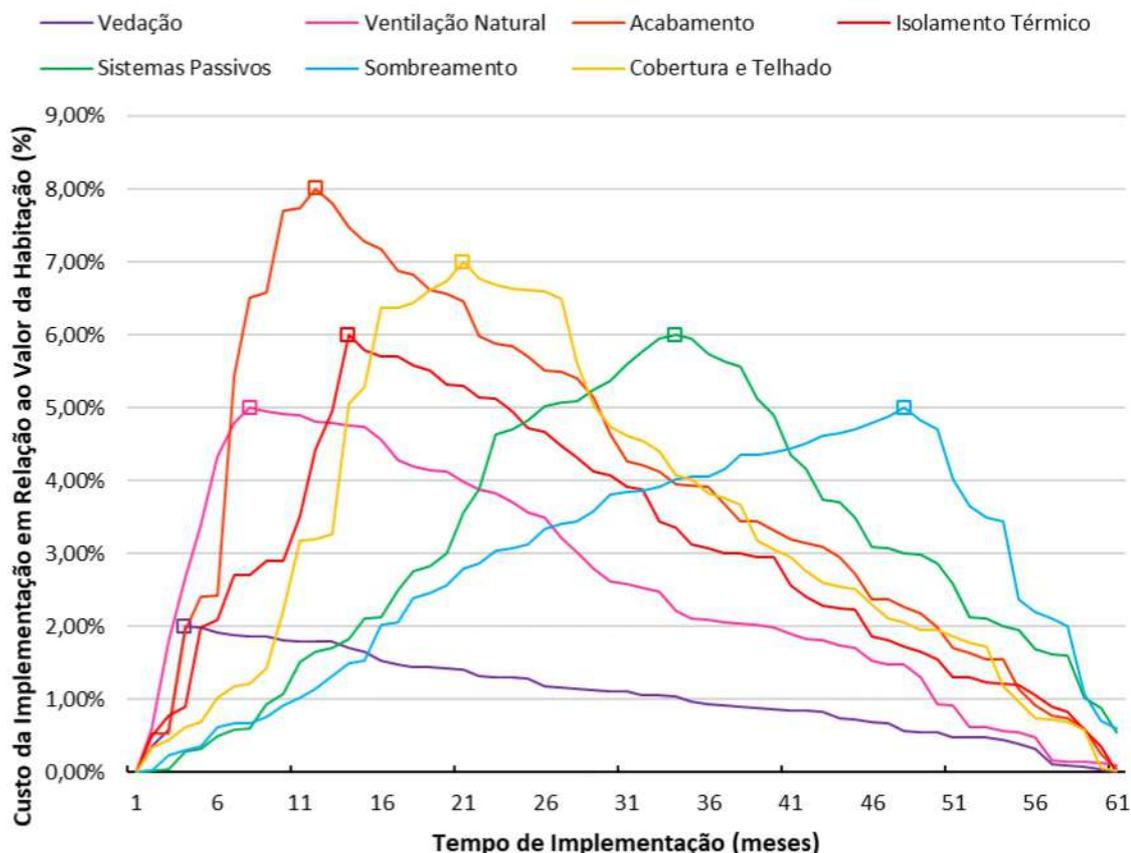
No tocante ao período de tempo no qual os custos manifestaram tendência à estabilização, observou-se que diferentes fatores exibiram comportamentos distintos. A título de exemplo, os fatores vedação e ventilação natural alcançaram seus picos de custo em 3 e 7 meses, respectivamente, antes de revelar sinais de estabilização. Este fenômeno aponta para a implementação relativamente célere desses elementos,

com impacto significativo nos custos durante as fases iniciais do projeto.

Por outro lado, os fatores sistemas passivos e sombreamento evidenciaram períodos de implementação mais prolongados, totalizando 33 e 47 meses, respectivamente. Essa observação sugere que a implementação desses elementos ocorreu de maneira gradual ao longo de um período extenso, possivelmente em decorrência da complexidade intrínseca do processo de implementação.

Em resumo, a análise do gráfico mostrou que o fator acabamento teve o maior impacto nos custos em relação ao preço da habitação, atingindo 8%. Além disso, diferentes fatores apresentaram diferentes tempos de implementação, com vedação e ventilação natural sendo mais rápidos, enquanto sistemas passivos e sombreamento foram mais demorados. Essas descobertas são relevantes para a tomada de decisões na construção de habitações. A Figura 50 expõe um gráfico mais minucioso, ilustrando os custos aproximados associados à implementação dos fatores voltados para o aumento da eficiência energética em uma habitação.

Figura 50 - Custos da implementação em relação ao valor da habitação



Fonte: Autor (2023)

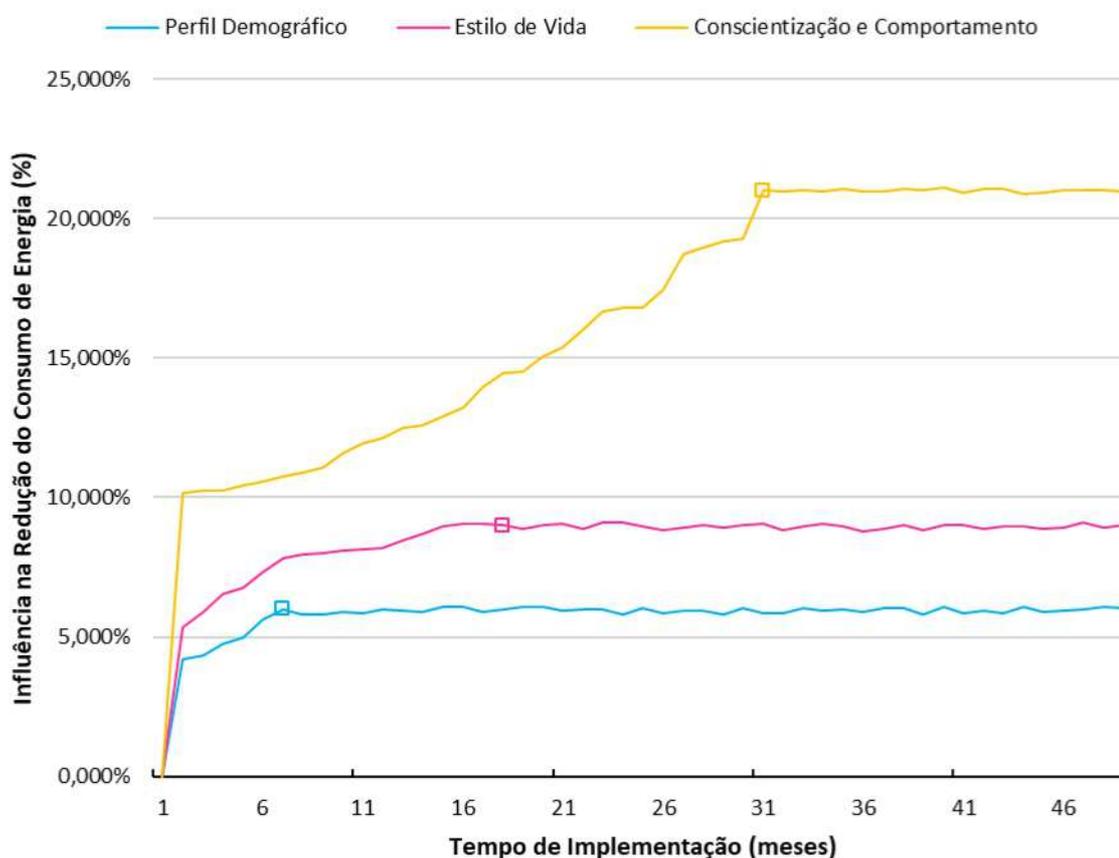
Em seguida, foram realizadas investigações sobre os fatores influenciadores do consumo de energia e os custos associados à sua aprendizagem pelos moradores, abrangendo programas educacionais e campanhas de conscientização.

Primeiramente, a partir do fator de perfil demográfico, observou-se que contribuiu com 6% para a redução do consumo de energia. O pico da influência foi atingido em 6 meses após a implementação das mudanças relacionadas a esse perfil.

O estilo de vida exerceu uma influência significativa, correspondendo a 9% na redução do consumo de energia, com um período de 17 meses para atingir seu pico de influência. Isso indica que as mudanças no estilo de vida demandaram um tempo mais longo para refletir no consumo de energia em comparação ao perfil demográfico.

A conscientização e comportamento foi o mais influente, contribuindo com 21% para a redução do consumo de energia, alcançando seu pico após um período de 30 meses. Isso indica um impacto significativo das mudanças na conscientização e comportamento ao longo do tempo, conforme mostrado na Figura 51.

Figura 51 - Fatores de influência do consumo relacionados aos habitantes

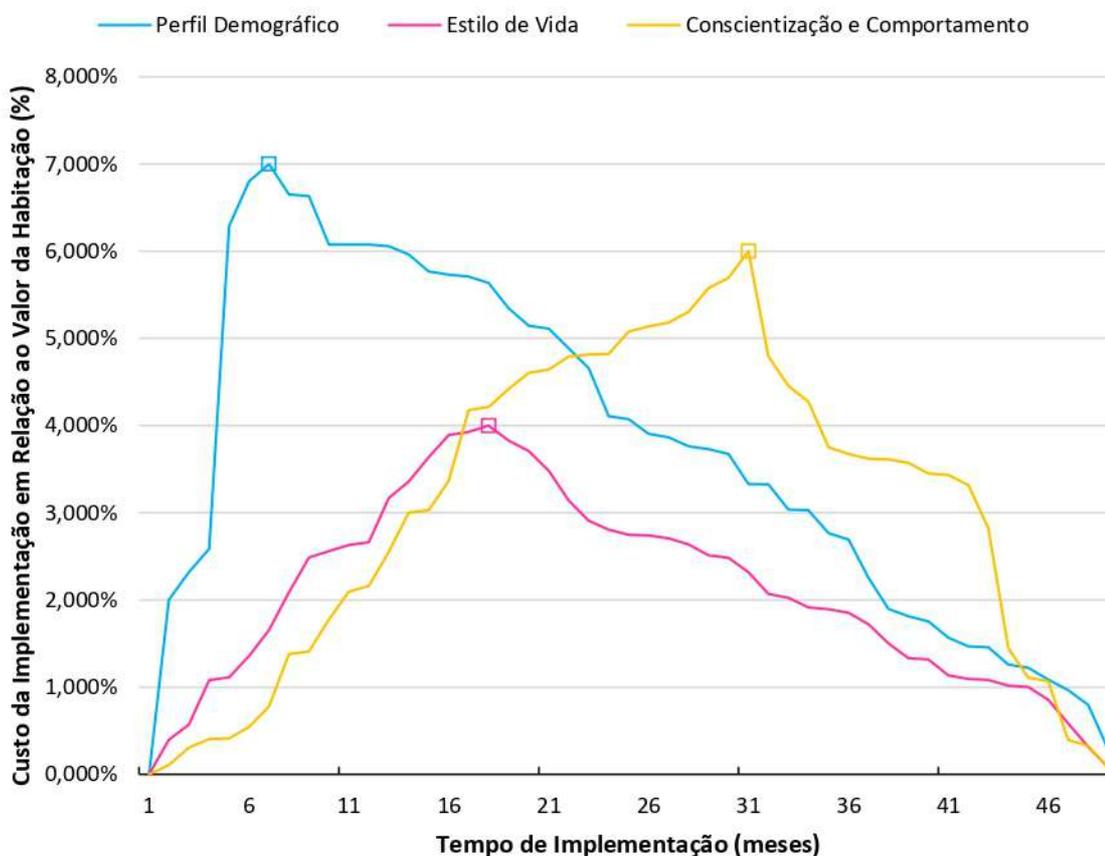


Fonte: Autor (2023)

Após análise da simulação, o perfil demográfico revelou o menor custo de aprendizado em relação ao preço da habitação, representando apenas 7%. Isso sugere uma adaptação e aprendizado mais rápidos das características demográficas em comparação com outros fatores.

O estilo de vida revelou-se menos impactante que o perfil demográfico, com um custo de aprendizado de 4%, alcançando seu pico após 17 meses. Isso sugere mudanças graduais, exigindo mais tempo para maturação. Por fim, o fator de conscientização e comportamento representou um custo de aprendizado de 6%, situando-se entre os dois fatores anteriores. Surpreendentemente, exigiu um período de aprendizado mais prolongado, totalizando 30 meses para atingir seu ponto máximo. Isso sugere que a conscientização e as mudanças comportamentais podem envolver um processo de aprendizado mais complexo e gradual. A Figura 52 fornece um gráfico detalhado dos custos aproximados associados à eficiência energética em residências.

Figura 52 - Custos da aprendizagem em relação ao valor da habitação



Fonte: Autor (2023)

Posteriormente, a influência do clima foi considerada, e observou-se que diversos elementos exerceram impacto sobre o consumo de energia em uma área residencial. As porcentagens médias de influência no consumo foram analisadas em conjunto com os custos aproximados em relação ao valor da habitação para a devida implementação de medidas para maximizar a redução energética.

Primeiramente, constatou-se que a influência predominante no consumo de energia se fez sentir por meio da variável temperatura, a qual evidenciou um impacto significativo, correspondente a 11% do total. Esse resultado indica que as flutuações na temperatura influenciaram de maneira substancial a demanda de energia. Esse fenômeno pode ser justificado pelo amplo emprego de sistemas de aquecimento e resfriamento em ambientes residenciais e comerciais. A implementação de estratégias visando a regulação desse fator demandou aproximadamente 12 meses, sinalizando a complexidade subjacente às soluções requeridas.

Por outro lado, percebe-se que a umidade e a poluição do ar exerceram influências relativamente menores, com porcentagens de 3% e 2%, respectivamente. Isso sugere que a relação entre esses fatores e o consumo de energia foi menos acentuada. Importante destacar que o período necessário para a implementação de medidas mitigatórias foi consideravelmente extenso, totalizando 32 e 40 meses, respectivamente. Essa demora é atribuída à imperatividade de desenvolver e executar políticas de longo prazo voltadas para o aprimoramento da qualidade do ar e o controle da umidade em ambientes fechados.

Os fatores que atingiram o seu pico de implementação de forma mais expedita foram as correntes de ar, que requereram 8 meses, e a precipitação, com um período de apenas 5 meses. Este cenário denota que medidas destinadas a otimizar o impacto desses fatores puderam ser adotadas de maneira relativamente célere em comparação a outros elementos. A rapidez na implementação pode ser creditada à viabilidade de incorporar tecnologias e estratégias de curto prazo para a gestão eficaz desses fatores, como a instalação de sistemas de ventilação eficientes ou a adoção de técnicas de captação de água da chuva.

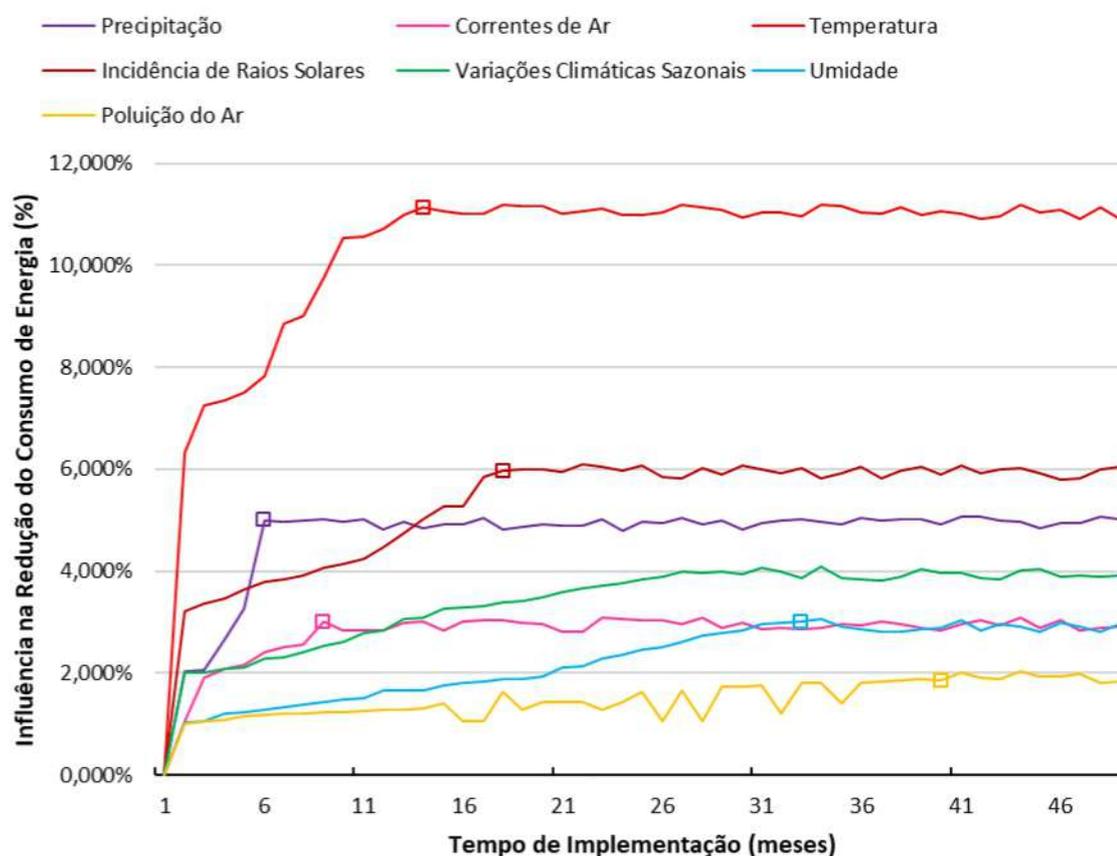
Por fim, observou-se que a incidência de raios solares e as variações climáticas sazonais demandaram um tempo de implementação mais prologado, correspondendo a 19 e 26 meses, respectivamente. Esses fatores, embora tenham impactos moderados no consumo de energia, requereram um período mais estendido para a concretização de soluções significativas. Isso pode ser atribuído à necessidade

premente de desenvolver infraestrutura e políticas de longo prazo voltadas para a geração sustentável de energia e a administração de flutuações sazonais na demanda energética.

Em síntese, a análise dos dados conduz à conclusão de que a temperatura representou o fator de maior magnitude no consumo de energia, seguida pela incidência de raios solares. A umidade e a poluição do ar exibiram impactos inferiores, embora tenham demandado prazos substanciais para a implementação de soluções. Em contrapartida, as correntes de ar e a precipitação ilustraram uma implementação ágil, enquanto as variações climáticas sazonais requereram um intervalo temporal intermediário. Essas informações revestem-se de relevância substancial para nortear a elaboração de políticas e estratégias voltadas para a otimização do consumo de energia em distintos contextos ambientais.

Na Figura 53, apresentada a seguir, é exibido um gráfico que ilustra de maneira visual a importância dos fatores climáticos na redução energética.

Figura 53 - Fatores de influência do consumo e custos relacionados ao clima



Fonte: Autor (2023)

Na esfera do estudo relacionado ao clima, procedeu-se à simulação e subsequente análise dos custos aproximados.

Primeiramente, ressalta-se que os custos de implementação variaram consideravelmente entre os diferentes fatores ambientais. Observou-se que a temperatura apresentava o custo de implementação mais elevado, representando 6% do preço da habitação. Isso sinaliza que o controle da temperatura em habitações constituía um investimento substancial em comparação com outros fatores ambientais relacionados. Portanto, tornava-se essencial ponderar cuidadosamente os benefícios e a necessidade dessa intervenção em projetos habitacionais.

Contudo, ao se avaliar o tempo requerido para a implementação desses fatores ambientais, emergiu uma tendência notável. Apesar de o custo associado à temperatura ser significativo, a sua implementação se consumava em um período de 12 meses. Isso indicava que, embora custoso, a materialização desse fator poderia ser conduzida em um espaço de tempo relativamente curto.

Contrastando com isso, fatores como variações climáticas sazonais ostentava custos de implementação menores, correspondentes a 3% do preço da habitação. No entanto, demandava um período substancialmente mais extenso para ser completamente implementado, sendo necessários 26 meses. Essa realidade sublinhava a complexidade e os desafios à incorporação desse fator em projetos habitacionais, mesmo com um custo inicial reduzido.

Além disso, era digno de nota que os fatores que atingiram o auge mais rapidamente, em termos de porcentagem do preço da habitação, foram precipitação, correntes de ar e umidade, todos apresentando custos de implementação de 2%. Entretanto, divergiam no que se referia ao período de implementação. A precipitação e correntes de ar atingiam o seu auge em 5 e 8 meses, respectivamente, ao passo que a umidade demandava 32 meses. Tal fato sugeria que, no que diz respeito ao custo, precipitação e correntes de ar eram os fatores que podiam ser implementados com maior celeridade e eficácia, representando, portanto, uma solução mais imediata para questões ambientais em habitações.

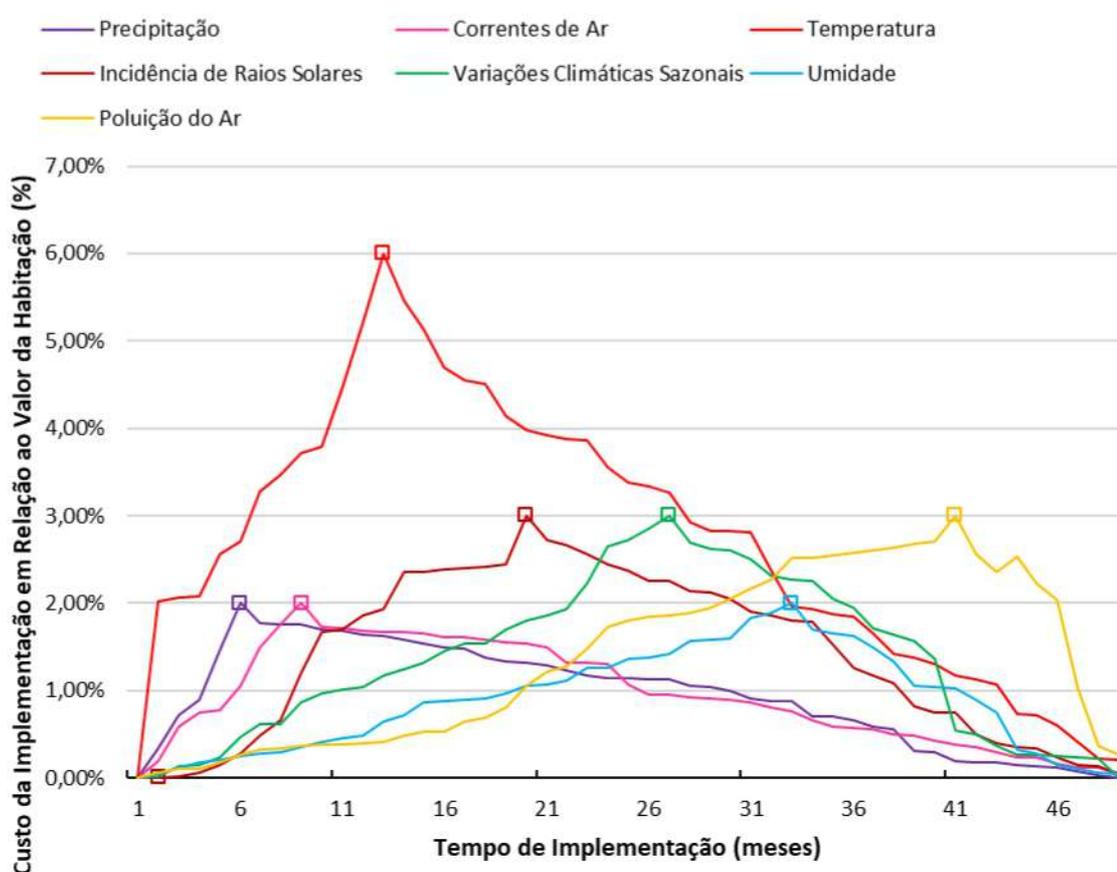
Por fim, a incidência de raios solares e a poluição do ar ocupavam posições intermediárias em termos de custo de implementação, correspondendo a 3% do preço da habitação cada uma. No entanto, demandavam 19 e 40 meses, respectivamente, para atingir o pico de implementação. Isso indicava que esses fatores poderiam se revelar mais desafiadores de implementar, mesmo diante de um custo relativamente

moderado.

Resumidamente, a análise dos dados evidenciava a variação significativa dos custos de implementação e do tempo necessário para concretizar fatores ambientais em projetos habitacionais. Enquanto a temperatura destacava-se como o fator mais dispendioso, podendo ser implementada em um espaço de tempo relativamente breve, a precipitação e as correntes de ar emergiam como os mais ágeis em termos de implementação, representando, assim, uma solução imediata. Tais informações assumiam importância crítica na tomada de decisões em projetos habitacionais, pautando-se no equilíbrio entre custo e tempo de implementação, a fim de atender às necessidades ambientais e de conforto dos ocupantes.

A Figura 54, por sua vez, proporciona uma visão mais detalhada, ilustrando os custos aproximados relacionados à implementação dos fatores direcionados ao aprimoramento da eficiência energética relacionados a fatores climáticos.

Figura 54 - Custos dos fatores climáticos em relação ao valor da habitação



Fonte: Autor (2023)

Após realizada a simulação do modelo, os resultados obtidos serão analisados e discutidos na próxima seção.

6.2.3. RESULTADOS E DISCUSSÕES

A análise dos resultados conduziu, em seu estágio inicial, à avaliação da correlação existente entre os fatores preeminentes que incidiram sobre o consumo energético e o tempo requerido para a efetiva implementação das medidas correspondentes. Este procedimento analítico representa um marco substancial na compreensão das dinâmicas inerentes à redução do consumo de energia em virtude do tempo mínimo necessário para a implementação. Tal análise é imperativa, considerando o complexo e mutável ambiente em que se insere, caracterizado pela intrincada interação de diversos elementos.

Dentre os fatores listados, aqueles que mais se destacaram por possuírem os menores tempos de implementação são: a vedação, a precipitação, o perfil demográfico, a ventilação natural e as correntes de ar. Estes fatores, quando submetidos a uma análise mais aprofundada, revelam sua importância estratégica na busca por um uso mais eficiente de energia.

Em primeiro lugar, a eficácia da vedação reside na capacidade de reduzir as perdas de energia térmica, mantendo uma temperatura interna mais estável. Essa implementação rápida, realizada em aproximadamente 3 meses, assume relevância especial, principalmente em climas extremos, onde a adequada vedação pode minimizar a demanda por aquecimento ou refrigeração, resultando em economia energética significativa.

Em seguida, a gestão da precipitação, com um prazo de implementação de 5 meses, possui o potencial de economizar energia de diversas maneiras. A captação e armazenamento de água da chuva, por exemplo, podem ser implementados em um período curto, reduzindo a dependência de fontes convencionais de água potável, o que, por sua vez, economiza energia associada ao seu tratamento e distribuição.

O estudo do perfil demográfico da região é fundamental para um planejamento energético eficaz. A rápida implementação de estratégias, cujo tempo de implementação gira em torno de 6 meses, para atender às necessidades específicas da população pode resultar em uma alocação mais eficiente de recursos energéticos

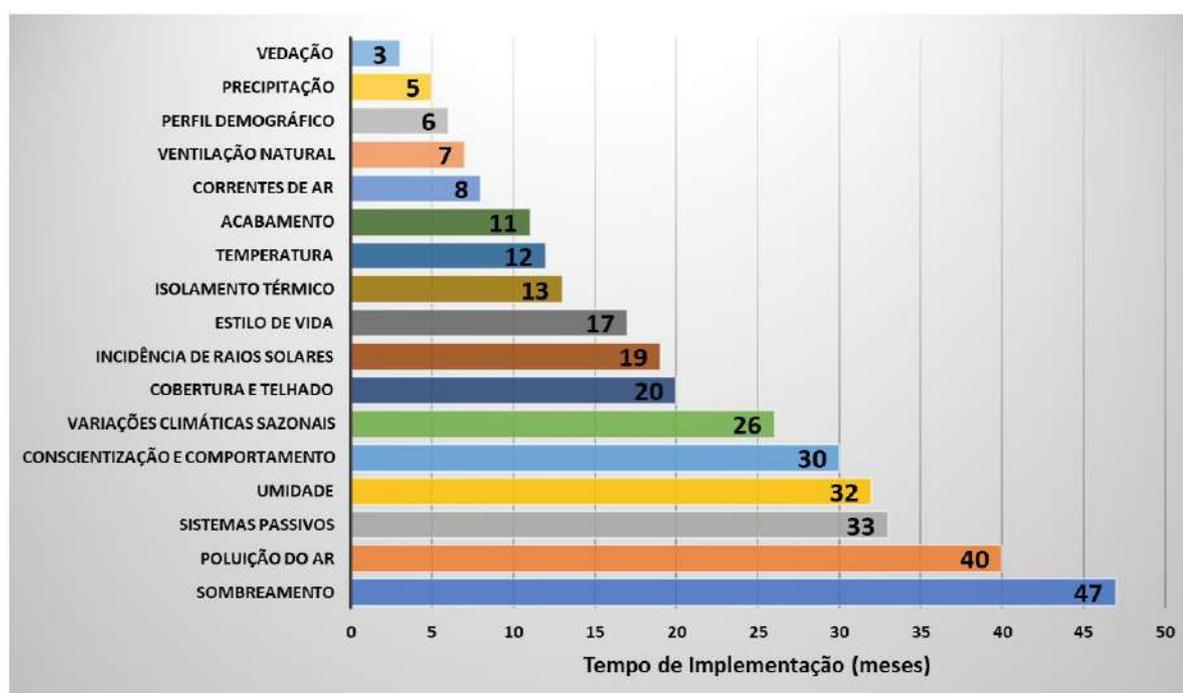
e, conseqüentemente, na redução do consumo de energia.

Em seguida, a implementação de sistemas de ventilação natural pode ser realizada em um período relativamente curto, aproximadamente 7 meses, e oferece uma solução ambientalmente amigável para a gestão da temperatura interna. Isso reduz a necessidade de sistemas de climatização elétrica, conhecidos por seu alto consumo de energia.

Finalmente, a otimização das correntes de ar internas por meio de estratégias de design e tecnologias como cortinas de ar pode ser efetuada em um prazo relativamente curto, em torno de 8 meses. Isso permite uma melhor regulação da temperatura ambiente, reduzindo a necessidade de aquecimento ou resfriamento, com impacto direto na redução do consumo de energia.

A representação gráfica, que considera a relação entre o período de implementação e o ganho resultante, é ilustrada na Figura 55. Essa abordagem proporciona uma melhor compreensão de como esses fatores influenciaram o consumo energético, considerando o prazo de aplicação e os benefícios adquiridos.

Figura 55 - Resultados considerando prazo x benefício



Fonte: Autor (2023)

A análise seguir, avaliou a correlação existente entre os fatores de influência e o custo de implementação das medidas destinadas à redução do consumo energético.

Esse custo foi expresso em porcentagens do custo total de cada habitação, independente das particularidades de cada uma. Além disso, foram considerados os custos relacionados a campanhas de conscientização sobre a importância do uso eficiente de energia.

A análise resultante desta simulação representa um progresso notável no entendimento das dinâmicas envolvidas na redução do consumo de energia em habitações. O foco deste estudo concentrou-se na identificação dos fatores que desempenham um papel crucial nesse processo, levando em consideração o menor custo de implementação, expresso em porcentagens do custo total de cada habitação.

Dentre os fatores mencionados, aqueles que mais se destacaram por apresentarem os menores custos de implementação foram a umidade, as correntes de ar, a precipitação, a vedação e a poluição do ar. Apesar de suas vantagens econômicas, é importante notar que esses fatores desempenham um papel menos significativo na redução do consumo de energia.

Primeiramente, a gestão da umidade, com um custo de implementação de 2%, sobressai-se como uma medida de baixo custo que impacta diretamente no conforto térmico. Isso possibilita que as habitações mantenham temperaturas internas mais moderadas, diminuindo, assim, a necessidade de aquecimento ou refrigeração e, conseqüentemente, reduzindo o consumo de energia.

Seguidamente, o controle das correntes de ar internas, com um custo de implementação de 2%, representa uma estratégia acessível para manter um ambiente com temperatura agradável. Essa abordagem reduz a dependência de sistemas de climatização elétrica, que são conhecidos por seu alto consumo de energia.

Posteriormente, a gestão da água da chuva, com um custo de implementação relativamente baixo de 2%, apresenta-se como uma estratégia econômica para a economia de energia.

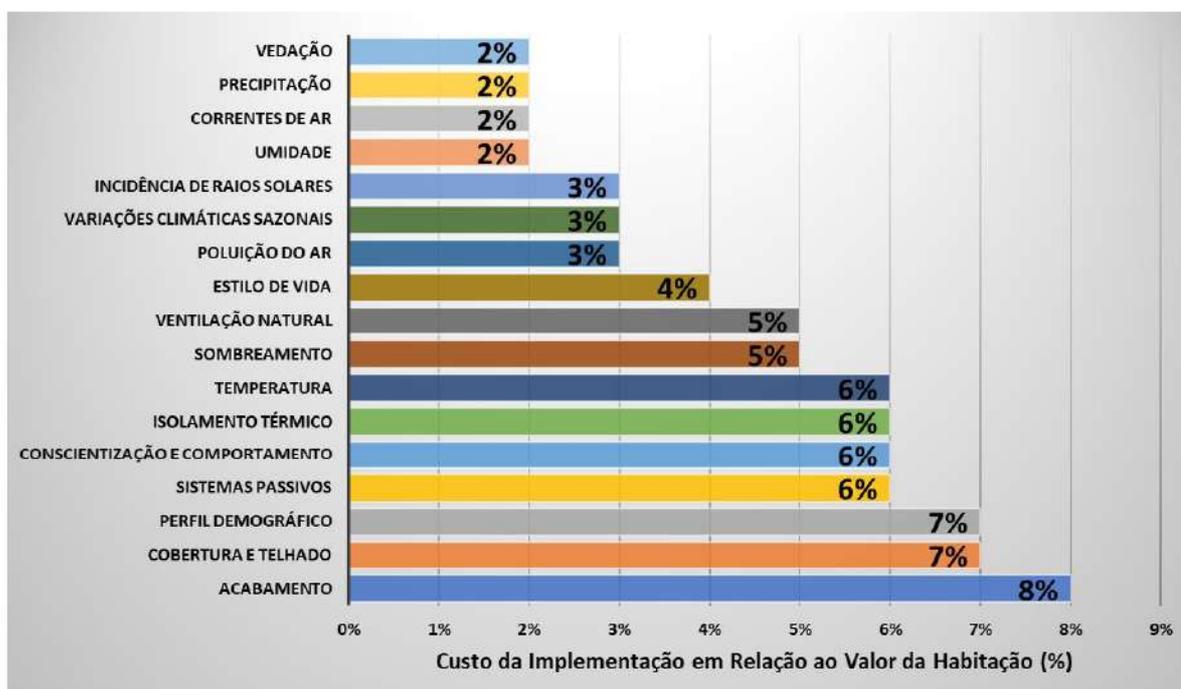
Em seguida, a vedação adequada das habitações, com um custo de implementação de 2%, destaca-se como uma medida eficaz para minimizar as perdas de energia térmica. O isolamento apropriado de portas e janelas evita a entrada de ar indesejado, resultando em uma temperatura interna estável e, conseqüentemente, em menor necessidade de aquecimento ou refrigeração.

Por fim, a mitigação da poluição do ar, embora tenha um custo ligeiramente mais elevado em relação aos fatores anteriores, representa 3% do custo de implementação. A melhoria da qualidade do ar interno, por meio de sistemas de

filtragem e ventilação adequados, contribui não apenas para um ambiente mais saudável, mas também pode reduzir a necessidade de sistemas de climatização elétrica, resultando em economia de energia.

A Figura 56 ilustra graficamente a relação entre o custo de implementação e os fatores de influência na redução do consumo energético, proporcionando uma representação visual dessa análise.

Figura 56 - Resultados considerando custo x benefício



Fonte: Autor (2023)

A próxima análise das simulações avaliou os fatores que mais influenciam na redução do consumo energético. Neste contexto, a compreensão dos resultados se reveste de um significado profundo, uma vez que a eficiência energética desempenha um papel central na busca por um uso mais sustentável dos recursos naturais e na mitigação dos impactos ambientais associados ao consumo de energia. Portanto, identificar os fatores que podem proporcionar a maior redução no consumo de energia é fundamental para o desenvolvimento de estratégias que promovam um uso mais eficiente e consciente da energia em ambientes habitacionais.

Dentre os fatores analisados, os que mais se destacaram por apresentarem as maiores reduções no consumo de energia foram a conscientização e comportamento, a temperatura, o isolamento térmico, o estilo de vida e o sombreamento.

A conscientização e comportamento dos habitantes apresentou a maior redução no consumo de energia, representada por um expressivo percentual de 21%. A mudança de hábitos e práticas cotidianas, como a adoção de medidas para evitar o desperdício de energia, desligar equipamentos não utilizados e otimizar o uso de iluminação e aparelhos elétricos, desempenha um papel fundamental na redução do consumo de energia. Este resultado sublinha a importância de educar e conscientizar as pessoas sobre a relevância da economia de energia.

A temperatura no interior das habitações é um fator de grande influência no consumo de energia e a obtenção de uma redução de 11% nesse consumo mostra a relevância do controle adequado da temperatura. Isso pode ser alcançado por meio da utilização de equipamentos que consomem menos energia, bem como pelo uso de materiais mais eficientes em paredes, telhados e pisos, resultando em menor necessidade para manter o aquecimento ou resfriamento dentro das habitações após atingir a temperatura desejada pelos moradores.

O isolamento térmico, com uma redução de 9% no consumo de energia, desempenhou um papel de suma relevância. A utilização de materiais isolantes de elevada qualidade em elementos estruturais proporcionou o eficaz isolamento do ambiente interno das habitações em relação às flutuações de temperatura externa, resultando em uma notável diminuição no consumo energético.

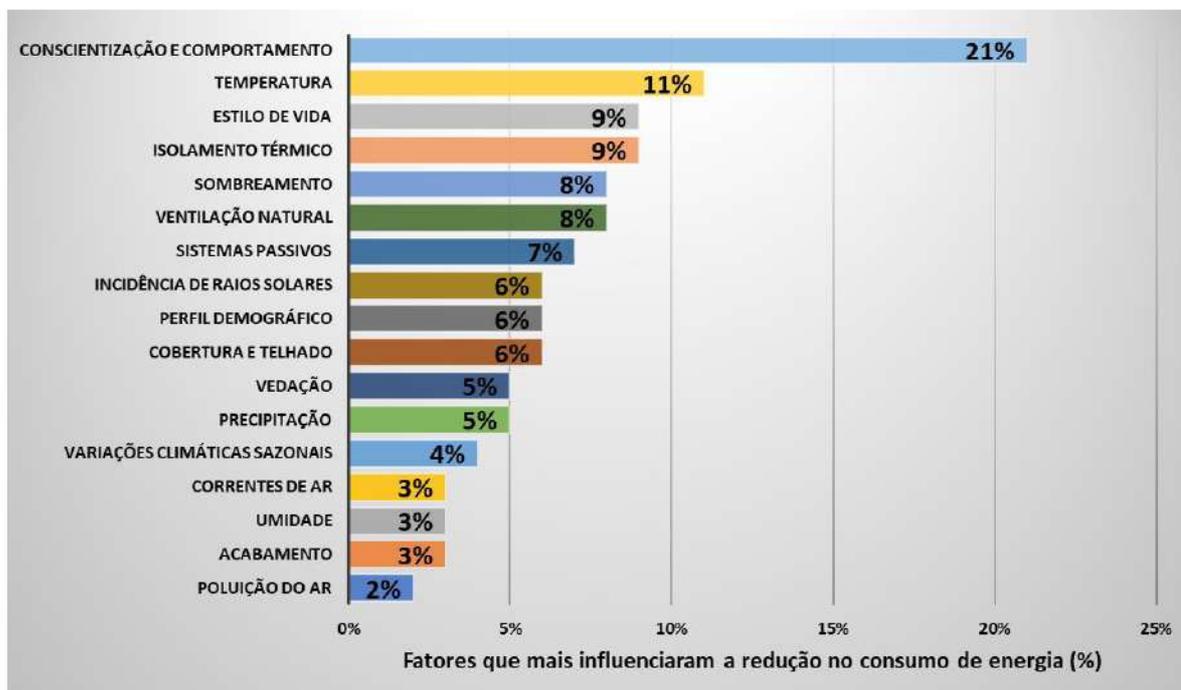
O estilo de vida adotado pelos habitantes também exerce uma influência substancial na redução do consumo de energia, com uma redução de 9%. A implementação de práticas mais sustentáveis, como a redução do desperdício de água e energia, bem como a escolha de eletrodomésticos eficientes, contribui para uma maior eficiência energética.

A estratégia de sombreamento, cujo objetivo é regular a exposição direta ao sol em ambientes internos, revela-se eficaz ao proporcionar uma redução substancial no consumo de energia. Essa redução, estimada em 8%, é alcançada mediante a implementação de elementos como persianas, cortinas e toldos, os quais atuam bloqueando o excesso de calor proveniente da luz solar. Essa ação resulta na diminuição da demanda por sistemas de resfriamento, contribuindo significativamente para a eficiência energética do ambiente. Além disso, a estratégia de sombreamento não apenas reduz os custos operacionais associados ao consumo de energia, mas também promove um ambiente interno mais confortável e sustentável.

Nesse contexto, a Figura 57 ilustra graficamente os fatores que mais

influenciam na redução no consumo de energia.

Figura 57 - Resultados considerando redução x benefício



Fonte: Autor (2023)

A análise subsequente dedicou a avaliar qual dos fatores estabeleceu a relação mais favorável entre os aspectos considerados, abrangendo o tempo necessário para implementação, os custos associados à implantação e as taxas mais substanciais de redução do consumo de energia. Nesse contexto, foi empregado o IVA como indicador para verificar os melhores fatores que influenciam na redução do consumo de energia considerando diversos aspectos, sendo importante ressaltar quanto maior é o valor desse indicador, melhor é a adequação do projeto para a redução do consumo de energia. O IVA baseia-se nos seguintes índices: influência no consumo de energia, disponibilidade de recursos, complexidade da implementação, risco de falha, custo de implantação e tempo de implementação. Onde:

- **Influência no Consumo de Energia:** Os valores são expressos em percentagem. Por exemplo, um valor de 5% significa que o fator de influência tem um impacto de 5% no consumo de energia.
- **Custo da Implementação:** Os valores são expressos em percentagem do preço da habitação. Por exemplo, um valor de 2% significa que o custo de implementação do fator de influência é de 2% do preço da habitação.

- Tempo de Implementação: Os valores são expressos em meses.
- Disponibilidade de Recursos: Os valores são expressos em uma escala de 1 a 5, sendo 1 o valor mais baixo e 5 o valor mais alto.
- Complexidade da Implementação: Os valores são expressos em uma escala de 1 a 3, sendo 1 o valor mais baixo e 3 o valor mais alto.
- Risco de Falha: Os valores são expressos em uma escala de 1 a 2, sendo 1 o valor mais baixo e 2 o valor mais alto.

A Tabela 32 apresenta os resultados dos índices obtidos.

Tabela 32 - Indicadores do IVA

Fatores de Influência	Influência (%)	Custo (%)	Tempo (meses)	Disponibilidade	Complexidade	Risco de Falha
Precipitação	5	2	5	3	2	1
Correntes de Ar	3	2	8	2	2	1
Temperatura	11	6	12	5	3	2
Incidência de Raios Solares	6	3	19	4	2	1
Variações Climáticas Sazonais	4	3	26	3	2	1
Umidade	3	2	32	2	2	1
Poluição do Ar	2	3	40	2	2	1
Perfil Demográfico	6	7	6	5	3	2
Estilo de Vida	9	4	17	4	2	1
Conscientização e Comportamento	21	6	30	5	3	2
Vedação	5	2	3	3	2	1
Ventilação Natural	8	5	7	4	2	1
Acabamento	3	8	11	2	2	1
Isolamento Térmico	9	6	13	5	3	2
Sistemas Passivos	7	6	33	5	3	2
Sombreamento	8	5	47	4	2	1
Cobertura e Telhado	6	7	20	4	2	1

Fonte: Autor (2023)

Após a obtenção dos índices, foi realizado a simulação e a avaliação dos fatores que apresentam o melhor IVA para a determinação da viabilidade e eficácia das medidas voltadas para a redução do consumo energético. Dentre os fatores listados, os que mais se destacaram por apresentarem os maiores valores de IVA, o que indica sua maior viabilidade para implementação foram a conscientização e comportamento, a ventilação natural, os sistemas passivos, o isolamento térmico e o sombreamento.

A conscientização e a alteração de comportamento dos indivíduos desempenham papéis fundamentais no contexto da mitigação do consumo energético. O notável IVA de 0,045%, sugere que a alocação de recursos para a

implementação de programas de sensibilização e promoção de práticas sustentáveis pode conduzir a resultados altamente benéficos, mantendo custos relativamente baixos. Esse fenômeno se deve à natureza intrínseca das mudanças comportamentais, que frequentemente não requerem investimentos substanciais em infraestrutura, mas, mesmo assim, podem exercer um impacto considerável na redução do consumo de energia.

A utilização de ventilação natural representa uma estratégia altamente eficaz para gerenciar a temperatura interna de um ambiente, sem a necessidade de recorrer a sistemas de climatização elétrica. Isso é particularmente relevante, considerando o IVA de 0,028%, o qual ressalta a viabilidade intrínseca desse método. A implementação de medidas voltadas para favorecer a ventilação natural pode ser realizada com custos notavelmente acessíveis, quando comparada a outras soluções tecnológicas disponíveis.

Por outro lado, sistemas de controle térmico passivos, tais como um projeto arquitetônico apropriado, possuem um potencial considerável para proporcionar eficiência energética, sem exigir investimentos excessivamente elevados. O IVA de 0,024% revela que tais medidas são igualmente viáveis e têm a capacidade de resultar em reduções substanciais no consumo de energia, sem sobrecarregar significativamente os recursos disponíveis.

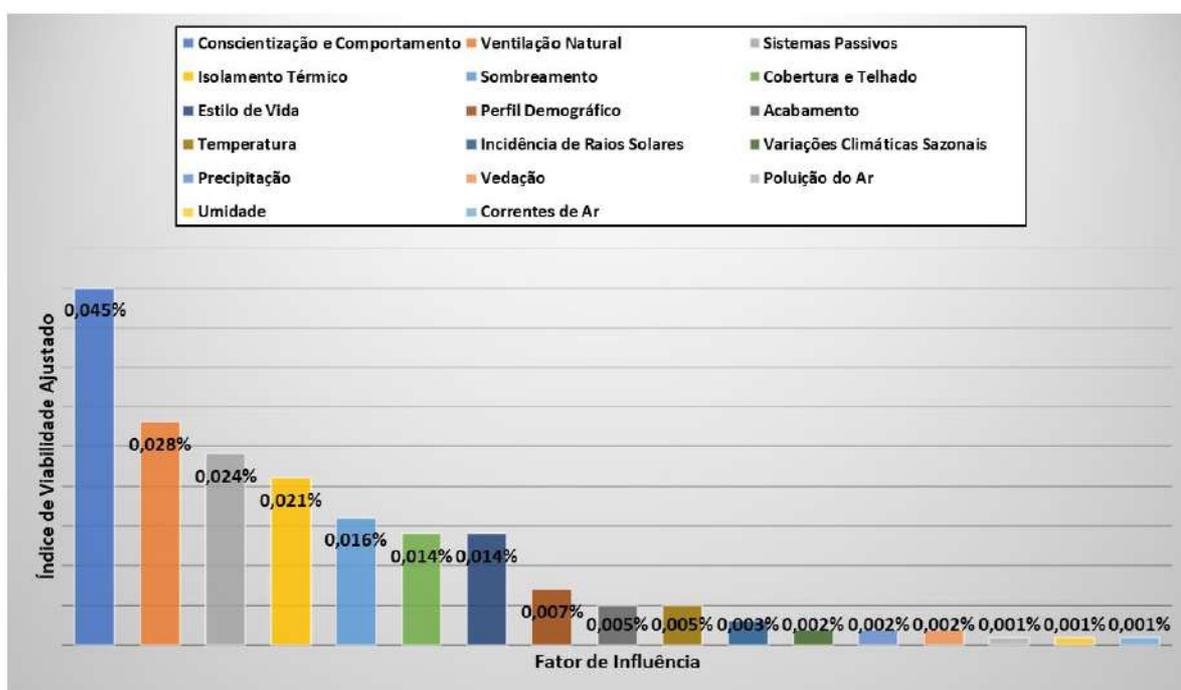
No que concerne ao isolamento térmico, uma técnica estabelecida e comprovada para mitigar as perdas de energia térmica, observou-se um IVA de 0,021%. Tal valor indica claramente que o investimento em aprimoramentos no isolamento térmico representa uma abordagem economicamente viável para a conservação de energia. Isso se justifica pelo fato de que os custos inerentes à implementação dessa medida podem ser plenamente respaldados pelos benefícios resultantes, notadamente a redução da demanda por sistemas de aquecimento ou resfriamento. Dessa forma, fica evidente que o isolamento térmico emerge como uma estratégia sensata e sustentável para otimizar a eficiência energética e, conseqüentemente, reduzir os custos operacionais associados ao consumo de energia térmica.

O emprego adequado de técnicas de sombreamento desempenha um papel crucial no gerenciamento das condições térmicas internas de uma residência, resultando na atenuação da carga térmica durante os períodos de clima quente. Notavelmente, a análise econômica revela um IVA de 0,016%, o que destaca a sua

atratividade em termos de custo-benefício. Esta conclusão se justifica pelo fato de que as estratégias de sombreamento podem ser implementadas com despesas relativamente baixas, contudo, elas conferem benefícios substanciais em termos de eficiência energética e, por conseguinte, representam uma opção econômica e eficaz para mitigar o calor excessivo em edifícios residenciais.

A Figura 58 representa graficamente o IVA dos principais fatores que exercem influência significativa na diminuição do consumo de energia.

Figura 58 - Índice de Viabilidade Ajustado



Fonte: Autor (2023)

Neste estudo de caso, realizou-se a simulação de um cenário complexo com o objetivo de reproduzir de forma precisa a dinâmica das interações entre os moradores, as unidades habitacionais e as condições climáticas em um bairro residencial situado em uma cidade brasileira. O propósito principal dessa simulação foi identificar os principais fatores que exerceram influência significativa na redução do consumo de energia em tal contexto específico.

A busca incessante pela eficiência energética e pela redução do consumo de energia desempenhou um papel crucial na promoção da sustentabilidade e na mitigação dos impactos ambientais. No âmbito da gestão energética em residências, a incorporação de variáveis de influência desempenhou um papel fundamental. Essas

variáveis compreenderam estratégias que variaram desde a conscientização e a modificação de comportamento dos habitantes até a aplicação de tecnologias e práticas de construção sustentável. Dentro desse contexto, foram realizadas as análises em relação a como cada um desses fatores contribuiu para a economia de energia e a redução dos custos operacionais em ambientes residenciais.

Antes da implementação dos mencionados fatores, a simulação do consumo de energia, realizada por meio de SMA, indicou um valor de 116,6 kWh por pessoa por ano. Após a aplicação dos fatores de influência, com foco naqueles identificados como mais relevantes pelo IVA, foram observadas mudanças significativas no cenário de consumo energético.

A conscientização e a mudança de comportamento desempenharam um papel vital na redução do consumo de energia. Os esforços para educar os residentes sobre práticas de consumo eficiente de energia e incentivá-los a adotar essas práticas resultaram em uma notável economia de 24,546 kWh/ano por pessoa. Isso equivale a uma redução significativa de 21% em relação ao consumo original de energia por pessoa. Esse êxito não apenas mostra a importância da educação e da sensibilização para questões de eficiência energética, mas também ressalta como pequenas mudanças comportamentais podem ter um impacto substancial na redução do consumo de energia. A alocação de recursos para programas de conscientização e campanhas educacionais revela-se um investimento fundamental para promover a sustentabilidade e reduzir os custos a longo prazo.

A adoção de estratégias de ventilação natural ofereceu uma abordagem eficaz para reduzir o consumo de energia relacionado à climatização. A implementação dessas práticas resultou em uma economia de 9,328 kWh/ano por pessoa, o que representa uma economia notável de 8% em relação ao consumo original. A eficiência desse método é evidente, pois utiliza recursos naturais para controlar a temperatura interna, reduzindo assim a dependência de sistemas de climatização que consomem energia. Além disso, os custos de implementação permanecem acessíveis, representando 5% do valor da habitação. Isso mostra a eficácia das soluções de baixo custo para melhorar a eficiência energética.

A implementação de sistemas passivos, projetados para maximizar a eficiência energética, resultou em uma economia de 8,162 kWh/ano por pessoa, o que equivale a uma redução de 7% em relação ao consumo original. Estes sistemas incorporam princípios de design que otimizam o aproveitamento da luz natural e o controle da

temperatura, reduzindo assim a demanda por energia. A vantagem desses sistemas está em sua capacidade de operar de forma autônoma, sem a necessidade de consumo constante de energia. Isso não apenas contribui para a redução de custos a longo prazo, mas também reduz a pegada de carbono, tornando-se uma solução ambientalmente amigável.

A aplicação de isolamento térmico foi fundamental para reduzir a perda de calor e, conseqüentemente, o consumo de energia relacionado ao aquecimento. Essa medida resultou em uma economia considerável de 10,494 kWh/ano por pessoa, correspondendo a uma redução de 9% em relação ao consumo original. O isolamento térmico, além de reduzir o consumo de energia, proporciona maior conforto, mantendo as temperaturas internas estáveis e agradáveis durante as diferentes estações do ano. Os custos de implementação, que representam 6% do valor da habitação, são justificados pela economia significativa de energia ao longo do tempo.

A adoção de estratégias de sombreamento mostrou ser uma solução eficaz para reduzir o consumo de energia relacionado à climatização. Essa medida resultou em uma economia de 9,328 kWh/ano por pessoa, equivalente a uma redução de 8% em relação ao consumo original. O uso de sombreamento não apenas reduz a carga térmica sobre o edifício, mas também contribui para um ambiente interno mais confortável. Os custos de implementação, representando 5% do valor da habitação, são proporcionais aos benefícios gerados pela redução da necessidade de ar-condicionado e resfriamento.

Somando todas essas alterações, alcançou-se uma economia global de 61,858 kWh/ano por indivíduo. Esse êxito representa uma redução substancial no consumo energético, contribuindo não apenas para a fomentação da sustentabilidade energética, mas também para o alívio dos encargos associados ao consumo de energia nas habitações.

Em relação aos gastos totais de implementação, que variaram de 5% a 6% do valor do imóvel, de acordo com o fator de influência em consideração, destaca-se a importância de salientar que esses dispêndios se mostram plenamente justificados. Tais aportes constituem um investimento direcionado à eficiência energética e à promoção da sustentabilidade, o que se reflete em economias de longo prazo.

No que tange ao intervalo de tempo requerido para a efetivação das medidas, variando entre 7 e 47 meses, é imperativo enfatizar que esses prazos se alinham de maneira sensata com a execução das melhorias que redundam em economia

substancial de energia. As intervenções são planejadas de forma gradual e meticulosa, assegurando um impacto positivo e duradouro.

Em síntese, a avaliação do consumo energético se faz necessária para compreender os impactos potenciais desses fatores no consumo individual de energia ao longo do ano, onde a aplicação dos fatores de influência simulados neste estudo resultou em expressiva possibilidade de economia de energia, com custos de implementação devidamente justificados e prazos de execução perfeitamente razoáveis. Tal êxito evidencia que as técnicas desenvolvidas aliando a raspagem de dados e os SMA promovem uma imensa relevância na promoção de medidas para a redução do consumo energético em residências, promovendo, assim, a sustentabilidade a longo prazo. O êxito alcançado não somente beneficia os moradores, mas também contribui para a preservação do ambiente e a construção de um horizonte energético mais sustentável.

7. CONSIDERAÇÕES FINAIS

A presente pesquisa representou um avanço significativo no campo das simulações multiagentes, destacando-se pela introdução e validação de uma abordagem inovadora para a coleta de dados destinada à modelagem de cenários complexos. O desdobramento dessa investigação evidenciou a meticulosa conquista dos objetivos inicialmente delineados, estabelecendo um precedente significativo para futuros avanços na compreensão e aplicação de SMA em ambientes dinâmicos e heterogêneos.

A análise dos resultados obtidos confirmou a validação da hipótese inicial de que o desenvolvimento de uma técnica pioneira de mineração de dados, fundamentada na raspagem de dados, resultaria em uma abordagem única e eficaz para a simulação de cenários de alta complexidade em SMA. Ao longo do desenvolvimento desta pesquisa, a combinação dessas técnicas desempenhou um papel crucial no aprimoramento do processo de simulação, proporcionando uma compreensão aprimorada de fenômenos complexos em diversas esferas de aplicação.

A técnica de raspagem de dados, inicialmente utilizada para extrair informações de fontes *online*, emergiu como um componente-chave no sucesso do desenvolvimento proposto. A sua integração singular às simulações multiagentes ofereceu uma abordagem única e altamente relevante na coleta de dados em tempo real, dinâmicos e diversificados. Esse método inovador, ao possibilitar a geração de dados de entrada em simulações, superou as limitações tradicionais, contribuindo substancialmente para o aprimoramento do realismo e da atualidade das informações coletadas.

Os objetivos gerais e específicos estabelecidos no início da pesquisa foram plenamente alcançados. A revisão abrangente da literatura sobre simulações baseadas em SMA proporcionou uma base teórica sólida e robusta, conferindo à abordagem desenvolvida uma fundamentação aprofundada. A integração de conhecimentos de linguagens de programação, bibliotecas avançadas e plataformas de simulação não apenas estabeleceu um arcabouço técnico consistente, mas também delineou padrões de excelência para pesquisas futuras neste campo em constante evolução.

Os estudos de caso, conduzidos em diferentes cenários complexos, não apenas validaram a eficácia e aplicabilidade da abordagem proposta, mas também evidenciaram sua capacidade de melhorar processos, reduzir o tempo para a coleta de dados, diminuir os custos e aumentar a eficiência das simulações multiagentes. A análise crítica das áreas de aplicação mais promissoras, considerando suas contribuições potenciais para a tomada de decisões em diferentes contextos, proporcionou *insights* valiosos para futuras pesquisas.

A justificativa para a realização desta pesquisa encontra-se fundamentada na crescente demanda por avanços significativos no domínio das simulações multiagentes, especialmente na modelagem de cenários complexos. A necessidade premente de aprimorar a capacidade de tomada de decisões em ambientes complexos motivou a busca por estratégias inovadoras, culminando na criação de uma técnica pioneira de raspagem de dados para simulações multiagentes. Esta abordagem não só otimizou processos decisórios, mas também contribuiu para a identificação eficaz de tempo, custos e esforços, redução de riscos e erros, e aprimoramento do conhecimento sobre cenários e possibilidades.

A incorporação da raspagem de dados, ao oferecer flexibilidade na coleta de informações, permitiu o acesso e a síntese de dados que, de outra forma, seriam de difícil aquisição ou inacessíveis. Esta capacidade de transcender limitações e integrar uma variedade mais ampla de dados enriqueceu substancialmente as simulações, fornecendo uma base de dados robusta e alinhada com a complexidade do ambiente simulado. Desta forma, a pesquisa não apenas respondeu às demandas atuais por avanços em simulações multiagentes, mas também abriu caminho para um entendimento mais profundo e eficaz da complexidade inerente a cenários complexos.

A motivação que impulsionou esta pesquisa, derivada da necessidade de interpretar e direcionar o vasto mar de informações característico da Era da Informação, foi plenamente atendida. A pesquisa contribuiu para um entendimento mais equilibrado e consciente, utilizando a informação como ferramenta-chave para auxiliar o bem-estar humano. Ao adaptar a raspagem de dados para geração de dados de entrada em simulações baseadas em agentes para cenários complexos, a tese propôs uma técnica inovadora que promete avanços significativos na coleta, processamento e integração de dados em simulações multiagentes.

A relevância desta pesquisa transcende seus benefícios para a tomada de decisões assertivas. A técnica proposta e desenvolvida não apenas otimizou o tempo dedicado a processos decisórios, mas também contribuiu para a identificação eficaz de custos e esforços, redução de riscos e erros, aprimoramento do conhecimento sobre cenários e possibilidades, além de proporcionar um aumento significativo no controle das ações a serem determinadas.

Além das vantagens multifacetadas advindas da implementação da técnica de raspagem de dados na simulação de cenários complexos, como a coleta rápida e eficaz de dados atualizados e a capacidade de extrair dados de fontes não estruturadas, esta pesquisa destaca-se pelo emprego de linguagens de programação avançadas e plataformas de simulação. Ao desenvolver uma técnica inovadora que molda a raspagem de dados ao serviço da simulação de cenários complexos, estabelece-se um novo padrão de excelência para a pesquisa e aplicação prática neste campo.

Em suma, a presente tese não apenas respondeu à demanda crescente por avanços no campo das simulações multiagentes, mas estabeleceu um novo padrão de excelência ao propor uma abordagem para a coleta e integração de dados. A técnica criada não só otimizou processos decisórios, mas também contribuiu para a identificação eficaz de custos e esforços, redução de riscos e erros, e aprimoramento do conhecimento sobre as possibilidades. Este trabalho, ao propor uma abordagem única, estabelece-se como um marco no avanço da pesquisa acadêmica e aplicada nesta área em constante evolução.

7.1. RECOMENDAÇÕES PARA PESQUISAS FUTURAS

A presente tese proporcionou avanços no campo das simulações multiagentes, consolidando-se pela introdução e validação de uma abordagem singular para a coleta de dados em cenários complexos. Esta seção se dedica a explorar as possíveis direções e áreas de pesquisa que emergem desse trabalho pioneiro, delineando caminhos promissores para futuros estudos.

A) APROFUNDAMENTO NAS TÉCNICAS DE MINERAÇÃO DE DADOS

O êxito da abordagem desenvolvida com foco na raspagem de dados em SMA, sugere a necessidade de um aprofundamento nesse domínio. Pesquisas futuras podem explorar modificações e aprimoramentos nesse processo, considerando novas fontes de dados e algoritmos mais avançados para otimizar a eficácia da coleta de informações em tempo real.

B) EXTENSÃO PARA NOVOS AMBIENTES E ESFERAS DE APLICAÇÃO

A abordagem desenvolvida nesta pesquisa mostrou-se eficaz em diversos cenários complexos. No entanto, futuras investigações podem se concentrar na abordagem desenvolvida para ambientes específicos, como saúde, finanças ou logística, explorando como a técnica de raspagem de dados pode ser ajustada para atender às demandas exclusivas dessas áreas.

C) DESENVOLVIMENTO DE FERRAMENTAS ESPECIALIZADAS

Considerando a integração de linguagens de programação, bibliotecas avançadas e plataformas de simulação, uma linha promissora de pesquisa envolve o desenvolvimento de ferramentas e plataformas especializadas. Tais recursos poderiam facilitar a implementação da abordagem em diferentes contextos, tornando-a mais acessível para pesquisadores e profissionais interessados em explorar simulações multiagentes.

D) AVALIAÇÃO DE IMPACTO EM PROCESSOS DECISÓRIOS REAIS

Embora a pesquisa tenha validado a eficiência da abordagem, é crucial expandir essa validação para situações do mundo real. Estudos futuros podem focar na aplicação prática da técnica de raspagem de dados em processos decisórios reais, avaliando seu impacto em tomadas de decisões complexas em diversas organizações e setores.

E) *EXPLORAÇÃO DE TENDÊNCIAS EMERGENTES EM TECNOLOGIA*

À medida que a tecnologia evolui, é fundamental manter-se atualizado com as tendências emergentes. Pesquisas futuras podem explorar como avanços em inteligência artificial, aprendizado de máquina e processamento de linguagem natural podem ser integrados à abordagem existente, aprimorando ainda mais sua capacidade de lidar com a complexidade crescente dos dados.

F) *CONSIDERAÇÕES ÉTICAS E GOVERNANÇA DE DADOS*

A crescente importância da ética na pesquisa e no uso de dados exige uma análise aprofundada das implicações éticas associadas à técnica de raspagem de dados. Investigar práticas éticas e estabelecer diretrizes para a governança de dados em simulações multiagentes torna-se crucial para garantir a integridade e responsabilidade na pesquisa.

G) *EDUCAÇÃO E DISSEMINAÇÃO DE CONHECIMENTO*

A disseminação eficaz de conhecimento é essencial para impulsionar avanços no campo. Pesquisas futuras podem se concentrar no desenvolvimento de programas educacionais, *workshops* e materiais de treinamento para capacitar estudantes e profissionais na aplicação da abordagem em simulações multiagentes.

Em síntese, esta tese procurou responder a demandas contemporâneas e estabeleceu um ponto de partida para investigações futuras. Ao delinear estas perspectivas, almeja-se inspirar novos pesquisadores a contribuir para o constante avanço e inovação no campo das simulações multiagentes.

REFERÊNCIAS BIBLIOGRÁFICAS

AHLBRECHT, T.; DIX, J.; FIEKAS, N. **The Multi-Agent Programming Contest**. 1. ed. Berlim: Springer, 2020.

AKHTAR, N.; REHMAN, A.; HUSSNAIN, M.; ROHAIL, S. **Hierarchical Coloured Petri-Net Based Multi-Agent System for Flood Monitoring, Prediction, and Rescue**. IEEE Access, vol. 7, 2019. Disponível em: <https://ieeexplore.ieee.org/abstract/document/8926354>. Acesso em: 18 nov. 2020.

ALMAGOR, J.; MARTIN, A.; McCORRIE, P.; MITCHELL, R. **How Can an Agent-Based Model Explore the Impact of Interventions on Children's Physical Activity in an Urban Environment?** Disponível em: <https://www.sciencedirect.com/science/article/pii/S1353829221001842>. Acesso em: 29 jan. 2022.

AMARAL, F. **Introdução à Ciência de Dados: versus**
e Big Data. 1. ed. Rio de Janeiro: Alta Books, 2018.

ANDRADE, M. M. de. **Introdução e Metodologia do Trabalho Científico**. 5. ed. São Paulo: Atlas, 2001.

ASSOCIAÇÃO BRASILEIRA DE ENERGIA SOLAR FOTOVOLTAICA. **Geração Distribuída no Brasil de Energia Solar Fotovoltaica**. 2020. Disponível em: <http://www.aneel.gov.br>. Acesso em: 8 jan. 2021.

BADHAM, J.; CHATTOE-BROWN, E.; GILBERT, N.; CHALABI, Z.; KEE, F.; HUNTER, R. F. **Developing Agent-Based Models of Complex Health Behaviour**. Health & Place, v. 54, p. 170-177, 2018. Disponível em: <https://doi.org/10.1016/j.healthplace.2018.08.022>. Acesso em: 4 fev. 2019.

BAHGA, A.; MADISETTI, V. **Big Data Science & Analytics: A Hands-On Approach**. 1. ed. Nova Iorque: VPT, 2016.

BAI, Q.; REN, F.; FUJITA, K. **Multi-agent and Complex Systems**. 1. ed. Berlim: Springer, 2017.

BARBIERI, C. **Governança de Dados: Práticas, Conceitos e Novos Caminhos**. 1. ed. Rio de Janeiro: Alta Books, 2019.

BARNETT-NEEFS, C.; WIEDMANN, M.; IVANEK, R. **Examining Patterns of Persistent Listeria Contamination in Packinghouses Using Agent-Based Models**. Journal of Food Protection, v. 85, n. 12, p. 1824-1841, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0362028X22110744>. Acesso em: 6 mar. 2023.

BARROS, A. J. S.; LEHFELD, N. A. de S. **Fundamentos de Metodologia**

Científica. 3. ed. São Paulo: Person Education, 2008.

BASS, F. **A New Product Growth for Model Consumer Durables**. Management Science, Catonsville, v. 15, p. 2015-227, 1969.

BAZZAN, A. L. C.; KLUGL, F. **Multi-Agent Systems for Traffic and Transportation Engineering**. Hershey: Information Science Reference, 2011.

BEER, M.; FASLI, M.; RICHARDS, D. **Multi-Agent Systems for Education and Interactive Entertainment: Design, Use and Experience**. Hershey: Information Science Reference, 2011.

BENSON, T.; BIE, J.; GASKELL, J.; VEZZA, P.; KERR, J. R.; LUMBROSO, D.; OWEN, M. R.; KEMP, P. S. **Agent-Based Modelling of Juvenile Migration via Selective Tidal Stream Transport**. Ecological Modelling, v. 453, 2021. Disponível em: <https://doi.org/10.1016/j.ecolmodel.2021.109448>. Acesso em: 18 mar. 2022.

BLEI, D. M.; NIG, A. Y.; JORDAN, M. I. **Latent Dirichlet allocation**. Journal of Machine Learning Research. New York, v. 3, p. 993-1022, 2003.

BOISSIER, O.; BORDINI, R. H.; HUBNER, J.; RICCI, A. **Multi-Agent Oriented Programming**. Cambridge: The MIT Press, 2020.

BOLLINGER, B.; GILLINGHAM, K. **Peer Effects in the Diffusion of Solar Photovoltaic Panels**. Marketing Science, Catonsville, v. 31, p. 800–812, 2012.

BORDINI, R. H.; DASTANI, M.; DIX, J.; SEGHROUCHNI, A. E. F. **Multi-Agent Programming: Languages, Platforms and Applications**. 1. ed. Berlim: Springer, 2006.

BORTOLETTO, R. C.; MOITINHO, V. S.; WALDMAN, H. **Utilização de Sistemas Multiagentes para a Simulação do Comportamento de Usuários em Redes de Telecomunicações**. In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS (SBRC), 38, 2020. Porto Alegre: Sociedade Brasileira de Computação, 2020. Disponível em: <https://sol.sbc.org.br/index.php/sbrc/article/view/12313>. Acesso em: 8 jan. 2021.

BROUCKE, S.; BAESENS, B. **Practical Web Scraping for Data Science: Best Practices and Examples with Python**. 1. ed. Nova Iorque: Apress, 2018.

BURG, V.; TROITZSCH, K. G.; AKYOL, D.; BAIER, U.; HELLWEG, S.; THEES, O. **Farmer's Willingness to Adopt Private and Collective Biogas Facilities: An Agent-Based Modeling Approach**. 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0921344921000070>. Acesso em: 18 mar. 2022.

BUTTS, D. J.; THOMPSON, N. E.; CHRISTENSEN, S. A.; WILLIAMS, D. M.; MURILLO, M. S. **Data-Driven Agent-Based Model Building for Animal Movement Through Exploratory Data Analysis**. Ecological Modelling, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0304380022001132>. Acesso em:

27 fev. 2023.

CAMPISI, D.; GITTO, S.; MOREA, D. **An Evaluation of Energy and Economic Efficiency in Residential Buildings Sector**: A multi-criteria analysis on an Italian case study. *International Journal of Energy Economics and Policy*. 2018. Disponível em: <https://doaj.org/article/9c346a833dac48e5943fad4d314f1215>. Acesso em: 30 jun. 2019.

CAMPOS, R. F. A.; CUNHA, D. A.; BUENO, N. P. **Disseminação de Informações em Sistemas Socioecológicos**: Análise de um Modelo Híbrido de Dinâmica de Sistemas e Modelagem Baseada em Agentes. 2020. Disponível em: https://www.scielo.br/scielo.php?pid=S0103-63512020000100257&script=sci_arttext&lng=pt. Acesso em: 12 mai. 2020.

CASTRO, L. N.; FERRARI, D. G. **Introdução à Mineração de Dados**: Conceitos Básicos, Algoritmos e Aplicações. 1. ed. São Paulo: Saraiva, 2016.

CASTRO, L. S.; MANOEL, F. C. P. B.; JESUS, V. S.; PANTOJA, C. E.; BORGES, A. P.; ALVES, G. V. **Integrando Sistemas Multi-Agentes Embarcados, Simulação Urbana e Aplicações de IoT**. Disponível em: https://www.researchgate.net/publication/344827393_Integrando_Sistemas_Multi-Agentes_Embarcados_Simulacao_Urbana_e_Aplicacoes_de_IoT. Acesso em: 8 abr. 2020.

CHERSONI, G.; DELLAVALLE, N.; FONTANA, M. **Modelling Thermal Insulation Investment Choice in the EU via the Behaviourally Informed Agent-Based Model**. *Energy Policy*, v. 12, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0301421522000489>. Acesso em: 27 fev. 2023.

CHWIF, L. **Modelagem e Simulação de Eventos Discretos**: Teoria e Aplicações. 4. ed. Barueri: GEN LTC, 2014.

COELHO, H. **Modelação Computacional Baseada em Agentes**: Enfrentar a Complexidade. 2018. Disponível em: <https://rcc.dcet.uab.pt/index.php/rcc/article/view/57>. Acesso em: 22 set. 2020.

DE BOSSCHER, B. C. D.; MOHAMMADI ZIABARI, S. S.; SHARPANSKYKH, A. A. **A Comprehensive Study of Agent-Based Airport Terminal Operations Using Surrogate Modeling and Simulation**. *Simulation Modelling Practice and Theory*, v. 128, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1569190X23000886>. Acesso em: 5 dez. 2023.

DEITEL, P.; DEITEL, H. **Intro to Python for Computer Science and Data Science**: Learning to Program with AI, Big Data and The Cloud. 1. ed. Londres: Pearson, 2019.

DENG, C.; YANG, G. **Distributed Adaptive Fault-Tolerant Control Approach to Cooperative Output Regulation for Linear Multi-Agent Systems**. *Automatica*,

Volume 103, 2019. Disponível em: <https://doi.org/10.1016/j.automatica.2019.01.013>. Acesso em: 2 out. 2020.

DERKENBAEVA, E.; HOFSTEDE, G. J.; VAN LEEUWEN, E.; HALLECK VEGA, S. **Simulating Households' Energy Transition in Amsterdam: An Agent-Based Modeling Approach**. *Energy Conversion and Management*, v. 294, 2023. Disponível em: <https://doi.org/10.1016/j.enconman.2023.117566>. Acesso em: 23 nov. 2023.

DIXON, A. **Multi-Agent Systems: Design, Synthesis and Analysis**. Nova Iorque: Clarye International, 2019.

ERL, T.; KHATTAK, W.; BUHLER, P. **Big Data Fundamentals: Concepts, Drivers & Techniques**. 1. ed. Londres: Pearson, 2015.

ESCOVEDO, T.; KOSHIYAMA, A. **Introdução a Data Science: Algoritmos de Machine Learning e Métodos de Análise**. 1. ed. São Paulo: Casa do Código, 2020.

FAN, J.; CAI, M.; MENG, Y.; BAO, C. **An Agent-Based Model for Public Travel Decision-Making Considering Risk Perception on the Pandemic**. *Procedia Computer Science*, v. 214, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050922020294>. Acesso em: 26 fev. 2023.

FARIAS, G. P.; LEITZKE, B. S.; BORN M. B.; AGUIAR, M. S.; ADAMATTI, D. F. **Modelagem Baseada em Agentes para Análise de Recursos Hídricos**. 2019. Disponível em: https://gsigma.ufsc.br/wesaac2019/paper/WESAAC_2019_paper_10.pdf. Acesso em: 8 ago. 2020.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P.; UTHURUSAMY, R. **Advances in Knowledge Discovery and Data Mining**. Cambridge: The MIT Press, 1996.

FONSECA, J. J. S. **Metodologia da Pesquisa Científica**. Fortaleza: UEC, 2002.

FONTES, B.; SOUZA, J. **Pesquisa Científica: métodos e técnicas**. São Paulo: Saraiva Educação, 2021.

FOSTER, I.; GHANI, R.; JARMIN, R. S.; KREUTER, F.; LANE, J. **Big Data and Social Science: Data Science Methods and Tools for Research and Practice**. 2. ed. Londres: Chapman and Hall/CRC, 2020.

GARCIA, E.; GIRET, A.; BOTTI, V. **Regulated Open Multi-Agent Systems (ROMAS): A Multi-Agent Approach for Designing Normative Open Systems**. 1. ed. Berlim: Springer, 2015.

GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de Pesquisa**. Porto Alegre: Plageder, 2009.

GHAFFARIAN, S.; ROY, D.; FILATOVA, T.; KERLE, N. **Agent-Based Modelling of**

Post-disaster Recovery with Remote Sensing Data. Disponível em:
<https://doi.org/10.1016/j.ijdr.2021.102285>. Acesso em: 20 abr. 2022.

GIL, A. C. **Como Elaborar Projetos de Pesquisa.** 4. ed. São Paulo: Atlas, 2002.

GOLDBARG, E. **Otimização Combinatória e Meta-heurísticas:** Algoritmos e Aplicações. 1. ed. Barueri: GEN LTC, 2021.

GÓMEZ-MARÍN, C. G.; ARANGO-SERNA, M. D.; SERNA-URÁN, C. A. **Agent-Based Microsimulation Conceptual Model for Urban Freight Distribution.** *Transportation Research Procedia*, v. 33, p. 155-162, 2018. Disponível em:
<https://www.sciencedirect.com/science/article/pii/S2352146518302436>. Acesso em: 4 fev. 2019.

GONG, S.; DONG, X.; WANG, K.; LEI, B.; JIA, Z.; QIN, J.; ROADKNIGHT, C.; LIU, Y.; CAO, R. **Agent-Based Modelling with Geographically Weighted Calibration for Intra-Urban Activities Simulation Using Taxi GPS Trajectories.** *International Journal of Applied Earth Observation and Geoinformation*, 2023. Disponível em:
<https://www.sciencedirect.com/science/article/pii/S1569843223001929>. Acesso em: 2 dez. 2023.

GORELIK, A. **The Enterprise Big Data Lake:** Delivering the Promise of Big Data and Data Science. 1. ed. Sebastopol: O'Reilly Media, 2019.

GRACIA-LÁZARO, C.; DERCOLE, F.; MORENO, Y. **Dynamics of Economic Unions:** An Agent-Based Model to Investigate the Economic and Social Drivers of Withdrawals. *Chaos, Solitons & Fractals*, v. 162, 2022. Disponível em:
<https://www.sciencedirect.com/science/article/pii/S0960077922004337>. Acesso em: 26 fev. 2023.

HJORTH, A.; HEAD, B.; BRADY, C.; WILENSKY, U. **LevelSpace:** A NetLogo Extension for Multi-Level Agent-Based Modeling. *Journal of Artificial Societies & Social Simulation*, 2020. Disponível em:
https://www.researchgate.net/publication/338962164_LevelSpace_A_NetLogo_Extension_for_Multi-Level_Agent-Based_Modeling. Acesso em: 8 jan. 2021.

HUNTER, E.; SAHA, S.; KUMAWAT, J.; CARROLL, C.; KELLEHER, J. D.; BUCKLEY, C.; McALOON, C.; KEARNEY, P.; GILBERT, M.; MARTIN, G. **Assessing the Impact of Contact Tracing with an Agent-Based Model for Simulating the Spread of Covid-19:** The Irish Experience. *Healthcare Analytics*, v. 4, 2023. Disponível em:
<https://www.sciencedirect.com/science/article/pii/S2772442523000965>. Acesso em: 2 dez. 2023.

IBM. **IBM's Intelligent Agent Strategy.** 1997. Disponível em:
<http://activist.gpl.ibm.com:81/WhitePaper/ptc2.htm>. Acesso em: 6 fev. 2019.

JAMES, C. L.; BRADSHAW, K. L. **Agent-Based Model Development of a Complex Socio-Ecological System:** Methods for Overcoming Data and Domain Limitations. *Ecological Informatics*, v. 77, 2023. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S1574954123002534>. Acesso em: 3 dez. 2023.

JAXA-ROZEN, M.; KWAKKEL, J. H.; BLOEMENDAL, M. **A Coupled Simulation Architecture for Agent-Based Geohydrological Modelling with NetLogo and MODFLOW**. *Environmental Modelling & Software*, Volume 115, 2019. Disponível em: <https://doi.org/10.1016/j.envsoft.2019.01.020>. Acesso em: 19 abr. 2020.

JEZIC, G.; CHEN-BURGER, J.; KUSEK, M. **Agents and Multi-Agent Systems: Technologies and Applications**. 1. ed. Berlim: Springer, 2020.

KAZIYEVA, D.; STUTZ, P.; WALLENTIN, G.; LOIDL, M. **Large-Scale Agent-Based Simulation Model of Pedestrian Traffic Flows**. *Computers, Environment and Urban Systems*, v. 105, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0198971523000844>. Acesso em: 28 nov. 2023.

KEANE, C.; EGAN, J. E.; HAWK, M. **Effects of Naloxone Distribution to Likely Bystanders: Results of an Agent-based Model**. 2018. Disponível em: <https://doi.org/10.1016/j.drugpo.2018.02.008>. Acesso em: 22 fev. 2019.

KELLEHER, J. D.; TIERNEY, B. **Data Science**. 1. ed. Cambridge: The MIT Press, 2018.

KIMBALL, R.; ROSS, M. **The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling**. 3. ed. Nova Jérsei: John Wiley & Sons, 2013.

KJÆR, L. J.; SCHAUBER, E. M. **The Effect of Landscape, Transmission Mode and Social Behavior on Disease Transmission: Simulating the Transmission of Chronic Wasting Disease in White-Tailed Deer (*Odocoileus virginianus*) Populations Using a Spatially Explicit Agent-Based Model**. *Ecological Modelling*, v. 472, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0304380022002162>. Acesso em: 4 fev. 2023.

KLEPPMANN, M. **Designing Data Intensive Applications: The Big Ideas Behind Reliable, Scalable and Maintainable Systems**. 1. ed. Sebastopol: O'Reilly Media, 2017.

KOU, L.; WANG, X.; LI, Y.; GUO, X.; ZHANG, H. **A Multi-Scale Agent-Based Model of Infectious Disease Transmission to Assess the Impact of Vaccination and Non-Pharmaceutical Interventions: The COVID-19 Case**. *Journal of Nonlinear Systems and Applications*. Disponível em: <https://doi.org/10.1016/j.jnlssr.2021.08.005>. Acesso em: 23 jan. 2022.

KRUPNIK, O.; MORDATCH, I.; TAMAR, A. **Multi-Agent Reinforcement Learning with Multi-Step Generative Models**. *Proceedings of Machine Learning Research*, Volume 100, 2020. Disponível em: <http://proceedings.mlr.press/v100/krupnik20a.html>. Acesso em: 1 fev. 2021.

KUDENKO, D.; KAZAKOV, D.; ALONSO, E. **Adaptive Agents and Multi-Agent Systems**. Berlim: Springer, 2005.

LAKATOS, E. M.; MARCONI, M. A. **Técnicas de Pesquisa: Planejamento e Execução de Pesquisas, Amostragem e Técnicas de Pesquisas, Elaboração, Análise e Interpretação de Dados**. 5. ed. São Paulo: Atlas, 2003.

LANGE, K. P. H.; KOREVAAR, G.; OSKAM, I. F.; NIKOLIC, I.; HERDER, P. M. **Agent-Based Modelling and Simulation for Circular Business Model Experimentation**. *Resource Conservation & Recycling*, v. 10, n. 2, p. 1-14, 2021.

Disponível em:

<https://www.sciencedirect.com/science/article/pii/S2667378921000055>. Acesso em: 29 jan. 2022.

LAW, A. M.; KELTON, W. D. **Simulation Modeling and Analysis**. Boston: McGraw-Hill, 2000.

LI, Z.; DUAN, Z. **Cooperative Control of Multi-Agent Systems: A Consensus Region Approach**. 1. ed. Boca Raton: CRC Press, 2014.

LIU, J.; ZHANG, M.; XIA, Y.; ZHENG, H.; CHEN, C. **Using Agent-Based Modeling to Assess Multiple Strategy Options and Trade-Offs for the Sustainable Urbanization of Cultural Landscapes: A case in Nansha, China**. *Landscape and Urban Planning*, v. 228, 2022. Disponível em:

<https://www.sciencedirect.com/science/article/pii/S0169204622002043>. Acesso em: 16 mar. 2023.

LONG, M.; SU, H.; LIU, B. **Second-Order Controllability of Two-Time-Scale Multi-Agent Systems**. *Applied Mathematics and Computation*, Volume 343, 2019.

Disponível em: <https://doi.org/10.1016/j.amc.2018.09.033>. Acesso em: 22 jul. 2020.

MACHADO, F. N. R. **Tecnologia e projeto de Data Warehouse**. 6. ed. São Paulo: Érica, 2013.

MANLEY, E.; CHENG, T. **Exploring the Role of Spatial Cognition in Predicting Urban Traffic Flow Through Agent-Based Modelling**. 2018. Disponível em:

<https://doi.org/10.1016/j.tra.2018.01.020>. Acesso em: 28 jan. 2019.

MARR, B. **Data Strategy: How to Profit from a World of Big Data, Analytics and the Internet of Things**. 1. ed. Londres: Kogan Page, 2020.

MEHDIZADEH, M.; NORDFJAERN, T.; KLOCKNER, C. A. **A Systematic Review of the Agent-Based Modelling/Simulation Paradigm in Mobility Transition**.

Technological Forecasting & Social Change, v. 184, 2022. Disponível em:

<https://doi.org/10.1016/j.techfore.2022.122011>. Acesso em: 6 mar. 2023.

MINARSCH, D.; FAVORITO, M.; HOSSEINI, A.; WARD, J. **Trading Agent Competition with Autonomous Economic Agents**. AAMAS, 2020. Disponível em:

<http://www.ifaamas.org/Proceedings/aamas2020/pdfs/p2107.pdf>. Acesso em: 7 jan. 2021.

- MITCHELL, R. **Web Scraping with Python: Collecting More Data**. 2. ed. Sebastopol: O'Reilly Media, 2018.
- MOYA, I.; CHICA, M.; SÁEZ-LOZANO, J. L.; CORDÓN, O. **Simulating the Influence of Terror Management Strategies on the Voter Ideological Distance Using Agent-Based Modeling**. *Telematics and Informatics*, v. 101, n. 656, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0736585321000952>. Acesso em: 25 fev. 2022.
- MUSAEUS, L. H.; MUSAEUS, P. **Computational Thinking in the Danish High School: Learning Coding, Modeling, and Content Knowledge with NetLogo**. 2019. Disponível em: <https://doi.org/10.1145/3287324.3287452>. Acesso em: 18 fev. 2020.
- MUSSAWAR, O.; URS, R. R.; MAYYAS, A.; AZAR, E. **Performance and Prospects of Urban Energy Communities Conditioned by the Built Form and Function: A Systematic Investigation Using Agent-Based Modeling**. *Sustainable Cities and Society*, v. 99, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2210670723005681>. Acesso em: 23 nov. 2023.
- NAGOEV, Z.; PSHENOKOVA, I.; GURTUEVA, I.; BZHIKHATLOV, K. **A Simulation Model for the Cognitive Function of Static Objects Recognition Based on Machine-Learning Multi-agent Architectures**. In: Samsonovich A. (eds) *Biologically Inspired Cognitive Architectures 2019. BICA 2019. Advances in Intelligent Systems and Computing*, vol 948. Springer, Cham. Disponível em: https://doi.org/10.1007/978-3-030-25719-4_48. Acesso em: 1 jan. 2020.
- NEDIĆ, A.; PANG, J.; SCUTARI, G. **Multi-agent Optimization**. 1. ed. Berlim: Springer, 2018.
- NING, B.; HAN, Q.; ZUO, Z. **Practical Fixed-Time Consensus for Integrator-Type Multi-Agent Systems: A Time Base Generator Approach**. *Automatica*, Volume 105, 2019. Disponível em: <https://doi.org/10.1016/j.automatica.2019.04.013>. Acesso em: 5 mar. 2020.
- NOGARE, D. D.; CHITNIS, A. B. **NetLogo agent-based models as Tools for Understanding the self-organization of Cell Fate, Morphogenesis and Collective Migration of the Zebrafish Posterior Lateral Line Primordium**. *Seminars in Cell & Developmental Biology*, 2020. Disponível em: <https://doi.org/10.1016/j.semcdb.2019.12.015>. Acesso em: 13 jan. 2021.
- PARVIERO, R.; HELLTON, K.H.; HAUG, O.; ENGØ-MONSEN, K.; ROGNEBAKKE, H.; CANRIGHT, G.; FRIGESSI, A.; SCHEEL, I. **An Agent-Based Model with Social Interactions for Scalable Probabilistic Prediction of Performance of a New Product**. *International Journal of Information Management Data Insights*, v. 2, 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2667096822000702>. Acesso em: 16 mar. 2023.

PEREIRA, G.; CARVALHO, J. R. **Campos Potenciais Artificiais e Consenso Aplicados ao Problema de Navegação de Sistemas Multiagente Heterogêneo**. In: ARTIGOS COMPLETOS - SIMPÓSIO BRASILEIRO DE ENGENHARIA DE SISTEMAS COMPUTACIONAIS (SBESC), 10, 2020. Porto Alegre: Sociedade Brasileira de Computação, 2020. Disponível em: https://sol.sbc.org.br/index.php/sbesc_estendido/article/view/13086. Acesso em: 17 jan. 2021.

QIAO, Q.; CHEUNG, C.; YUNUSA-KALTUNGO, A.; MANU, P.; CAO, R.; YUAN, Z. **An Interactive Agent-Based Modelling Framework for Assessing Covid-19 Transmission Risk on Construction Site**. *Safety Science*, [s.l.], v. 168, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0925753523002540>. Acesso em: 5 dez. 2023.

RÖDER, M.; BOTH, A.; HINNEBURG, A. **Exploring the space of topic coherence measures**. *Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, p. 399-408, 2015.

ROGERS, E. **Diffusion of Innovations**. 5. ed. Nova Iorque: Free Press, 2010.

RUSSEL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. Nova Jérsei: Prentice Hall, 1995.

SALECKER, J.; SCIAINI, M.; MEYER, K. M.; WIEGAND, K. **The NLRX R Package: A Next Generation Framework for Reproducible NetLogo Model Analyses**. 2019. Disponível em: <https://doi.org/10.1111/2041-210X.13286>. Acesso em: 18 nov. 2020.

SANTOS FILHO, R. H.; MATTOS, D. M. F.; MEDEIROS, D. S. V. **Agentes Inteligentes baseados em Aprendizado por Reforço para Alocação Dinâmica de Tráfego em Nuvens**. In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS (SBRC), 38, 2020. Porto Alegre: Sociedade Brasileira de Computação, 2020. Disponível em: <https://sol.sbc.org.br/index.php/sbrc/article/view/12279>. Acesso em: 9 jan. 2021.

SCHÖNBERGER, V. M.; CUKIER, K. **Big Data: A Revolution That Will Transform How We Live, Work and Think**. 2. ed. Londres: Eamon Dolan/Mariner Books, 2014.

SCHWARTZ, H. M. **Multi-Agent Machine Learning: A Reinforcement Approach**. 1. ed. Nova Jersey: Wiley, 2014.

SEBEM, R.; LEAL, A. B. **Proposta de Arquitetura para Controle Distribuído de Sistemas a Eventos Discretos Multiagentes Autônomos**. In: ARTIGOS COMPLETOS - SIMPÓSIO BRASILEIRO DE AUTOMAÇÃO INTELIGENTE (SBAI), 14, 2019. Ouro Preto: Sociedade Brasileira de Automática, 2019. Disponível em: https://sol.sbc.org.br/index.php/sbesc_estendido/article/view/13086. Acesso em: 7 jan. 2021.

SHARDA, R.; DELEN, D.; TURBAN, E.; BRODBECK, A. **Business Intelligence e Análise de Dados para Gestão do Negócio**. 4. ed. Porto Alegre: Bookman, 2019.

SHKAPENYUK, V.; SUEL, T. **Design and Implementation of a High-Performance Distributed Web Crawler**. Proceedings 18th International Conference on Data Engineering. 2002. Disponível em: <https://ieeexplore.ieee.org/document/994750>. Acesso em: 25 nov. 2020.

SLAVKOVIK, M. **Multi-Agent Systems**. 1. ed. Berlim: Springer, 2019.

SRIKRISHNAN, V.; KELLER, K. **Small Increases in Agent-Based Model Complexity can Result in Large Increases in Required Calibration Data**. Environmental Modelling & Software, v. 144, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1364815221000219>. Acesso em: 23 jan. 2022.

STEVENS, M.; CORREIA, G. H. A.; SCHELTES, A.; AREM, B. V. **An Agent-Based Model for Assessing the Financial Viability of Autonomous Mobility On-Demand Systems used as First and Last-Mile of Public Transport Trips: A Case Study in Rotterdam, the Netherlands**. Research in Transportation Business & Management, v. 45, 2022. Disponível em: <https://doi.org/10.1016/j.rtbm.2022.100875>. Acesso em: 4 fev. 2023.

TAVARES, C.; BILLA, C.; ADAMATTI, D. **Uma Extensão ao Modelo de Emoções e de Personalidades Baseado em Sistemas Multiagentes Utilizando Distribuição Normal, Lógica Fuzzy e Matemática Intervalar**. Revista Brasileira de Computação Aplicada, v. 12, n. 1, p. 54-70, 21 mar. 2020. Disponível em: <http://seer.upf.br/index.php/rbca/article/view/9326>. Acesso em: 20 jul. 2020.

UHRMACHER, A. M.; WEYNS, D. **Multi-Agent Systems: Simulation and Applications**. 1. ed. Boca Raton: CRC Press, 2009.

VO, L. T. **Mining Social Media: Finding Stories in Internet Data**. 1. ed. São Francisco: No Starch Press, 2019.

WATSON, M. **Scripting Intelligence: Web 3.0 Information Gathering and Processing**. 1. ed. California: Apress Berkeley, 2009.

WANG, D.; WANG, W. **Necessary and Sufficient Conditions for Containment Control of Multi-Agent Systems with Time Delay**. Automatica, Volume 103, 2019. Disponível em: <https://doi.org/10.1016/j.automatica.2018.12.029>. Acesso em: 24 jun. 2020.

WASLANDER, S. L. **Multi-Agent System Design**. Saarbrücken: VDM Verlag, 2008.

WEBBER, K. L.; ZHENG, H. Y. **Big Data on Campus: Data Analytics and Decision Making in Higher Education**. Baltimore: Johns Hopkins University Press, 2020.

WISE, S.; MILUSHEVA, S.; AYLING, S.; SMITH, R. M. **Scale Matters: Variations in Spatial and Temporal Patterns of Epidemic Outbreaks in Agent-Based Models**. Journal of Computational Science, v. 69, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877750323000595>. Acesso em: 3 dez. 2023.

WOLTMANN, S.; ZARTE, M.; KITTEL, J.; PECHMAN, A. **Agent-Based Simulation Model of Virtual Power Plants for Greener Manufacturing**. *Procedia CIRP*, v. 69, p. 377-382, 2018. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2212827117308272>. Acesso em: 28 jan. 2019.

WOOLDRIDGE, M. J. **An Introduction to MultiAgent Systems**. 1. ed. Nova Jérsei: John Wiley & Sons, 2009.

XIE, H.; YOU, L.; DILE, Y. T.; WORQLUL, A. W.; BIZIMANA, J. C.; SRINIVASAN, R.; RICHARDSON, J. W.; GERIK, T.; CLARK, N. **Mapping Development Potential of Dry-Season Small-Scale Irrigation in Sub-Saharan African Countries Under Joint Biophysical and Economic Constraints: An Agent-Based Modeling Approach with an Application to Ethiopia**. *Agricultural Systems*, v. 187. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0308521X20308489>. Acesso em: 25 fev. 2022.

YANG, R.; ZHANG, H.; FENG, G.; YAN, H.; WANG, Z. **Robust Cooperative Output Regulation of Multi-Agent Systems via Adaptive Event-Triggered Control**. *Automatica*, v. 102, 2019. Disponível em: <https://doi.org/10.1016/j.automatica.2019.01.001>. Acesso em: 4 set. 2020.

YAZAN, D. M.; FRACCASCIA, L.; MES, M.; ZIJM, H. **Cooperation in Manure-Based Biogas Production Networks: An Agent-based Modeling Approach**. Disponível em: <https://doi.org/10.1016/j.apenergy.2017.12.074>. Acesso em: 22 fev. 2019.

YOON, J.; WAN, H.; DANIEL, B.; SRIKRISHNAN, V.; JUDI, D. **Structural Model Choices Regularly Overshadow Parametric Uncertainty in Agent-Based Simulations of Household Flood Risk Outcomes**. *Computers, Environment and Urban Systems*, v. 103, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S019897152300042X>. Acesso em: 28 nov. 2023.

ZAPPE, S. C. A. C. **Análise de Dados, Otimização e Modelagem de Simulação**. 4. ed. Nashville: South Western, 2012.

ZHANG, J. **Multi-Agent-Based Production Planning and Control**. 1. ed. Nova Jérsei: Wiley, 2017.

ZHANG, K.; KAKADE, S. M.; BAŞAR, T.; YANG, L. F. **Model-Based Multi-Agent RL in Zero-Sum Markov Games with Near-Optimal Sample Complexity**. Disponível em: <https://arxiv.org/abs/2007.07461>. Acesso em: 1 fev. 2021.

APÊNDICES

**APÊNDICE A - CÓDIGO BASE DA TÉCNICA DE RASPAGEM DE DADOS
DESENVOLVIDO NA LINGUAGEM DE PROGRAMAÇÃO PYTHON**

```
import os

# Instalação automática das bibliotecas necessárias
os.system("pip install nltk openpyxl gensim pyLDAvis
wordcloud matplotlib requests scrapy beautifulsoup4
pyarrow")

import time
import time
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from gensim.corpora import Dictionary
from gensim.models.ldamodel import LdaModel
from gensim.matutils import hellinger
import pyLDAvis.gensim_models
import pandas as pd
import numpy as np
from wordcloud import WordCloud
import matplotlib.pyplot as plt
import requests
from bs4 import BeautifulSoup
import openpyxl
import datetime
nltk.download('stopwords')
```

```
nlTK.download('punkt')
nlTK.download('wordnet')

# Fase 1: EXTRAÇÃO PREDITIVA DE DADOS DE TEXTO
def extract_keywords(urls, username=None,
password=None, scraping_time_days=1):
    words_list = []
    collection_times = []
    site_words_list = []
    updates_list = []

    url_updates = {}

    # Adiciona a opção para verificar quantas vezes a
URL foi atualizada
    check_url_updates(urls, username, password,
max_checks=5)

    for url in urls:
        start_time = time.time()
        extracted_words =
extract_keywords_single(url, username, password,
scraping_time_days)
        end_time = time.time()
        collection_time =
str(datetime.timedelta(seconds=(end_time -
start_time)))
```

```
word_count = len(extracted_words)
updates = url_updates.get(url, 0)

# Adiciona os dados à lista
words_list.append(word_count)
collection_times.append(collection_time)
site_words_list.append(extracted_words)
updates_list.append(updates)

# Geração de nuvem de palavras
generate_wordcloud([word for extracted_words in
site_words_list for word in extracted_words])

# Salva os dados na planilha XLS
save_to_excel(words_list, urls, updates_list,
"Extraction", collection_times, site_words_list,
"extraction_output.xlsx")

def extract_keywords_single(url, username=None,
password=None, scraping_time_days=1):
    words = set()
    start_time = time.time()
    scraping_time_seconds = scraping_time_days * 24 *
60 * 60 # Convertendo dias para segundos

    while time.time() - start_time <
scraping_time_seconds:
```

```
        # Adapte conforme necessário para lidar com
autenticação, timeout, etc.
        if not url.startswith(('http://',
'https://')):
            url = 'http://' + url # Adiciona http://
se o esquema não estiver presente

        # Aguarda um intervalo de 1 segundo antes de
fazer a próxima verificação
        time.sleep(1)

        # Adiciona autenticação se fornecida
        auth = (username, password) if username and
password else None

        response = requests.get(url, auth=auth)

        if response.status_code == 200:
            html = response.text
            soup = BeautifulSoup(html, 'html.parser')

            # Coleta todas as palavras no conteúdo da
página
            all_text = ' '.join([p.get_text() for p
in soup.find_all('p')])

            words.update(word_tokenize(all_text.lower()))
```

```
        else:
            print(f"Failed to fetch data from {url},
Status Code: {response.status_code}")

        return list(words)

# Fase 2: PRÉ-PROCESSAMENTO DE DADOS DE TEXTO
def preprocess_text(text):
    stop_words = set(stopwords.words('english'))
    lemmatizer = WordNetLemmatizer()

    tokens = word_tokenize(text)
    tokens = [lemmatizer.lemmatize(token) for token
in tokens if token.isalpha()]
    tokens = [token for token in tokens if token not
in stop_words]

    return tokens

# Fase 3: MODELAGEM POR TÓPICOS
def topic_modeling(corpus, dictionary, num_topics=5):
    lda_model = LdaModel(corpus,
num_topics=num_topics, id2word=dictionary)
    return lda_model

# Fase 4: ANÁLISE DE SIMILARIDADE
def calculate_hellinger_distance(topic1, topic2):
```

```
    distance = np.sqrt(0.5 * np.sum((np.sqrt(topic1)
- np.sqrt(topic2)) ** 2))
    return distance

# Fase 5: EXPORTAÇÃO PARA NETLOGO
def export_to_netlogo(data, filename):
    with open(filename, 'w') as file:
        for entry in data:
            file.write(f"set-distance topic{entry[0]}
topic{entry[1]} {entry[2]}\n")

# Função para verificar quantas vezes a URL foi
atualizada
def check_url_updates(urls, username=None,
password=None, check_interval_seconds=10,
max_checks=6):
    global url_updates

    url_updates = {}

    for url in urls:
        if not url.startswith(('http://',
'https://')):
            url = 'http://' + url
            auth = (username, password) if username and
password else None
```

```
last_etag = None

for _ in range(max_checks):
    response = requests.get(url, auth=auth)

    if response.status_code == 200:
        current_etag =
response.headers.get('etag')

        if current_etag != last_etag:
            url_updates[url] =
url_updates.get(url, 0) + 1
            last_etag = current_etag

        time.sleep(check_interval_seconds)

print("Número de atualizações por URL:")
for url, updates in url_updates.items():
    print(f"{url}: {updates} vezes")

# Função para salvar as palavras em uma planilha XLS
def save_to_excel(words, urls, updates, subject,
collection_times, site_words_list, filename):
    wb = openpyxl.Workbook()
    ws = wb.create_sheet(title="Keywords")
    headers = ["URL", "Contagem de Palavras", "Tempo
de Coleta", "Palavras Encontradas", "Atualizações"]
```

```

    for col_num, header in enumerate(headers,
start=1):
        ws.cell(row=1, column=col_num, value=header)

    for i, (url, word_count, collection_time,
site_words, update_count) in enumerate(
        zip(urls, words, collection_times,
site_words_list, updates), start=2):
        ws[f"A{i}"] = url
        ws[f"B{i}"] = word_count
        ws[f"C{i}"] = collection_time
        ws[f"D{i}"] = ", ".join(site_words)
        ws[f"E{i}"] = update_count

wb.save(filename)

# Função para gerar uma nuvem de palavras e salvar
como imagem PNG
def generate_wordcloud(words):
    if not words:
        print("Nenhuma palavra disponível para gerar
a nuvem de palavras.")
        return

    wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(' '.join(words))
    plt.figure(figsize=(10, 5))

```

```
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.savefig("wordcloud.png", format="png",
dpi=300)

# Função para limpar a tela do console
def clear_screen():
    os.system('cls' if os.name == 'nt' else 'clear')

# Função principal do programa
def main():
    clear_screen()
    subject = input("Digite um assunto para pesquisar
nas URLs: ")
    urls = input("Digite a URL alvo: ").split()
    username = input("Digite o nome de usuário (se
necessário): ")
    password = input("Digite a senha (se necessário):
")
    scraping_time_days = int(input("Digite o tempo de
raspagem desejado em dias: "))

    # Fase 1: Extração Preditiva de Dados de Texto
    extract_keywords(urls, username, password,
scraping_time_days)

    # Fase 2: Pré-processamento de Dados de Texto
```

```
words_list = []
collection_times = []
site_words_list = []
updates_list = []

url_updates = {}

check_url_updates(urls, username, password,
max_checks=5)

for url in urls:
    start_time = time.time()
    extracted_words =
extract_keywords_single(url, username, password,
scraping_time_seconds)
    end_time = time.time()

    collection_time =
str(datetime.timedelta(seconds=(end_time -
start_time)))
    word_count = len(extracted_words)
    updates = url_updates.get(url, 0)

words_list.append(word_count)
collection_times.append(collection_time)
site_words_list.append(extracted_words)
updates_list.append(updates)
```

```

preprocessed_data = preprocess_text('
'.join([word for extracted_words in site_words_list
for word in extracted_words]))

# Fase 3: Modelagem por Tópicos
dictionary = Dictionary([preprocessed_data])

if not dictionary:
    print("Nenhum termo disponível para a
modelagem de tópicos.")
    return

corpus = [dictionary.doc2bow(preprocessed_data)]
lda_model = topic_modeling(corpus, dictionary)

# Fase 4: Análise de Similaridade
similarity_matrix = []
for i in range(len(lda_model.get_topics())):
    for j in range(i + 1,
len(lda_model.get_topics())):
        distance =
calculate_hellinger_distance(lda_model.get_topics()[i
], lda_model.get_topics()[j])
        similarity_matrix.append((i, j,
distance))

# Fase 5: Exportação para NETLOGO

```

```
    export_to_netlogo(similarity_matrix,
"netlogo_input.txt")

    # Salva os dados na planilha
    save_to_excel(words_list, urls, updates_list,
subject, collection_times, site_words_list,
"data_output.xlsx")

if __name__ == "__main__":
    main()
```