



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Modelo de Apoio à Decisão para Avaliar Dados Governamentais Abertos do Setor Elétrico

Ingrid Palma Araújo

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Orientadora
Prof.a Dr.a Ana Carla Bittencourt Reis

Brasília

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

PI55m Palma Araújo, Ingrid
Modelo de Apoio à Decisão para Avaliar Dados
Governamentais Abertos do Setor Elétrico. / Ingrid Palma
Araújo; orientador Ana Carla Bittencourt Reis. -- Brasília,
2022.
154 p.

Dissertação (Mestrado - Mestrado Profissional em
Computação Aplicada) -- Universidade de Brasília, 2022.

1. Modelo de Avaliação de Dados Abertos. 2. AHP-TOPSIS
2N. 3. Dados Abertos Governamentais (Conectados). 4. Dados
Abertos do Setor Elétrico. 5. Modelo de Apoio à Decisão por
Multicritérios. I. Bittencourt Reis, Ana Carla , orient.
II. Título.



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Modelo de Apoio à Decisão para Avaliar Dados Governamentais Abertos do Setor Elétrico

Ingrid Palma Araújo

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Prof.a Dr.a Ana Carla Bittencourt Reis (Orientadora)
PPCA/UnB

Prof. Dr. Patricio Esteban Ramírez-Correa Prof. Dr. Ari Melo Mariano
Universidade Católica do Norte - Chile PPCA/UnB

Prof. Dr. Altino José Mentzingen de Moraes (Suplente)
Ministério da Economia

Prof. Dr. Marcelo Ladeira
Coordenador do Programa de Pós-graduação em Computação Aplicada

Brasília, agosto de 2022

Dedicatória

Dedico este trabalho à minha mãe Lúcia Palma (*in memoriam*), pelo exemplo de mulher à frente do seu tempo, incentivadora da arte e da ciência. Por me despertar para a educação com humildade, humor, honestidade e coragem.

Dedico, ainda, à Jade Palma, minha irmã e meu maior exemplo de generosidade e força de vontade.

Agradecimentos

A palavra obrigado vem do latim *obligare*, derivando de outro verbo latino (*ligare*), que significa: ligar, unir, atar. Dessa forma, esta dissertação me liga àqueles que, de diferentes formas, mas de igual importância, contribuíram para o desenvolvimento desta pesquisa. Decidi não citar nominalmente todas as pessoas que, em algum momento, tiveram Amor, paciência, zelo e energia durante as inúmeras discussões e incessantes revisões. A vocês, expresso a minha infinita gratidão.

À minha orientadora, Profa. Dra. Ana Carla, pela confiança, incentivo e competência na orquestração deste projeto, realizado em um momento extremamente crítico e de pandemia mundial causada pela Covid-19.

Aos professores que fizeram parte da banca: Dr. Ari Melo, Dr. Patrício Ramírez-Correa e Dr. Altino de Moraes, pela disponibilidade e presteza na avaliação deste estudo.

Agradeço também ao colegiado da banca de qualificação, Profa. Dra. Viviane Grubisic e Prof. Dr. Edgar Oliveira, pelas revisões que reforçaram o trabalho.

Aos colegas do MME e da UnB, meus agradecimentos pelo empenho, companheirismo e paciência em compartilhar o tempo, por colaborar com as pesquisas e estudos, pelas dicas preciosas, por ouvir as minhas dúvidas e debater acerca dos meus questionamentos e inquietações.

Agradeço também ao Prof. Dr. Marcelo Ladeira, pelo trabalho empreendido na Coordenação do PPCA/UnB, fomentando os meios necessários para a realização de projetos e publicações relevantes.

Os meus sinceros agradecimentos a todas e todos, não apenas por esta dissertação, mas principalmente pela indelével vivência pessoal e profissional.

Resumo

Tomar decisões assertivas e eficientes, frente à escassez dos recursos públicos, considerando todas as potenciais soluções, tornou-se um dos problemas mais comuns para gestores responsáveis por dados abertos governamentais. Ecossistemas de dados abertos estão adotando diretrizes com foco em energia elétrica, tendo em vista a crescente conscientização sobre crise hídrica, mudanças climáticas, recursos renováveis e iniciativas para aumentar a eficiência energética¹. Além disso, a subjetividade e imprecisão no processo de abertura de dados desse setor pode tornar essa tarefa ainda mais complexa, principalmente quando não se possuem medidas específicas para subsidiar a tomada de decisão. Assim, este estudo propõe simplificar esse processo, combinando dois métodos multicritério em um modelo capaz de avaliar e priorizar, à luz dos critérios de riscos do contexto *Open Government Data*, dados do setor elétrico, apresentando os resultados via *dashboards online* interativos desenvolvido em R Shiny. A metodologia combina os métodos AHP e TOPSIS-2N, criando um *ranking* dos *datasets* conforme o seu nível de abertura. O método AHP foi utilizado para avaliar a importância dos critérios definidos, considerando os aspectos de consistência da matriz de decisão. Passo seguinte foi aplicar o método TOPSIS-2N para ordenação desses *datasets*. O modelo proposto é útil não apenas para gestores responsáveis por decisões que envolvam aporte de recursos para aprimorar os *datasets* já disponibilizados, mas também para priorizar tópicos mais relevantes para a abertura de dados. Os resultados apresentam o desempenho de cada conjunto de dados, exibindo aqueles que devem ser aprimorados em relação aos respectivos critérios de risco e aos temas priorizados, a fim de tornar mais ágil e assertiva a tomada de decisão no gerenciamento dessas bases. Bases afetas ao planejamento (BD46 e BD45) e à tarifa do setor elétrico (BD47) se destacaram quanto ao seu nível de abertura, inferindo um potencial valor agregado relacionado aos benefícios do uso desses dados.

Palavras-chave: Modelo de Avaliação de Dados Abertos, AHP-TOPSIS-2N, Dados Abertos Governamentais (Conectados), Dados Abertos do Setor Elétrico, Modelo de Apoio à Decisão por Multicritérios.

¹<https://data.europa.eu/it/highlights/open-energy-data-european-data-portal>.

Abstract

Making assertive and efficient decisions in the face of scarce public resources while considering all potential alternatives has become one of the most common issues for managers responsible for open government data. Open data ecosystems around the world are assume guidelines and targets with a focus on electric power, given the growing awareness of topics such as the water crisis, climate change, renewable resources, and initiatives to increase energy efficiency². Moreover, the subjectivity and imprecision in the process of opening data from this sector can make this task complex, especially when there are no specific measures to support decision-making. Thus, this study proposes to simplify this process, combining two methods of decision support through multicriteria analysis in a model capable of assessing and prioritizing, in the light of the risk criteria of the *Open Data* context, open data energy from the power sector, presenting the results via interactive and online dashboards developed in R Shiny³. The methodology followed combines the AHP and TOPSIS-2N methods, creating a ranking of datasets according to the openness level of they datasets evaluated. The AHP technique was used to specify and normalize the importance of each criterion, considering the consistency aspects of the decision matrix. The next step was to apply the TOPSIS-2N method to sort and prioritize these datasets. The proposed model is useful not only for managers responsible for decisions involving the contribution of resources to improve the datasets already available, but also to prioritize the most relevant topics for data opening. The results present the performance of each dataset, displaying those that should be improved in relation to the respective risk criteria and prioritized topics to make the decision-making for the management of these bases more agile and assertive. Datasets related to planning (BD46 and BD45) and the electricity sector tariff (BD47) stood out in terms of their openness level, inferring a potential added value related to the benefits of using this data.

Keywords: Open Data Evaluation Model, AHP-TOPSIS-2N, (Linked) Open Government Data, Open Energy Data, Multicriteria Decision Support Model.

²<https://data.europa.eu/it/highlights/open-energy-data-european-data-portal>.

³<https://www.rstudio.com/>

Sumário

1	Introdução	1
1.1	Contextualização	1
1.1.1	Escopo	3
1.1.2	Delimitação	3
1.2	Descrição do Problema	4
1.3	Justificativa	6
1.4	Objetivos	8
1.5	Metodologia	9
1.5.1	Classificação e Fases da Pesquisa	9
1.5.2	Plano Metodológico	10
1.6	Estrutura da Dissertação	12
2	Base Conceitual e Revisão da Literatura	14
2.1	Panorama e Evolução Legal do Governo Aberto	14
2.1.1	Governo Aberto no Brasil	15
2.2	Dados Abertos (<i>Open Data</i>)	18
2.2.1	Dados, Recursos e Metadados	18
2.3	Dados Conectados (<i>Linked Data</i>)	22
2.4	Dados Abertos Conectados (<i>Linked Open Data</i> - LOD)	23
2.5	Dados Abertos Governamentais (Conectados)	28
2.6	Modelo CRISP-DM	32
2.7	Modelo de Apoio à Decisão	34
2.7.1	Critério de Risco	36
2.8	Modelo de Processo de Apoio à Decisão MCDA	36
2.8.1	Problema e Problemática de Decisão	37
2.9	Métodos de Apoio à Decisão sobre Dados Abertos	38
2.10	Métodos Multicritério sobre Dados Abertos	45
2.11	Métodos Decisórios sobre Dados Abertos do Setor Elétrico	47
2.12	Considerações do Capítulo	49

3	Estudo de Caso	50
3.1	Aplicação do Modelo CRISP-DM	50
3.1.1	Entendimento e Preparação dos Dados	50
3.1.2	Modelagem	54
3.1.3	Avaliação dos Temas	55
3.2	Aplicação do Modelo MCDA	58
3.2.1	Estruturação do Problema Decisório	58
3.2.2	Definição dos Critérios de Riscos	58
3.2.3	Decomposição Hierárquica do Problema	62
3.3	Construção do Modelo pelo AHP-TOPSIS-2N	65
3.3.1	Ponderação dos Critérios pelo método AHP	65
3.3.2	Aplicação do TOPSIS-2N para <i>Ranking</i> das Bases	69
3.4	Validação do Modelo	73
3.4.1	Análise de Sensibilidade	73
3.5	Considerações do Capítulo	76
4	Resultados e Análises	77
4.1	<i>Dashboards</i> das Bases e dos Critérios via <i>AppShiny</i>	77
5	Considerações Finais	85
5.1	Recomendações e Contribuições	87
5.1.1	Artigos Publicados e Aceitos	89
5.2	Sugestões para Trabalhos Futuros	90
	Referências	91
	Anexo	102
I	Considerações acerca da Dissertação - Autoridade de Monitoramento da LAI do MME	103
II	Impacto Potencial da Dissertação - Ouvidoria do MME	106
III	Artigo Publicado na <i>Association for Computing Machinery (ACM)</i>	109
IV	Artigo Aceito na <i>9th International Conference on Future Internet of Things and Cloud (FiCloud)</i>	111
V	Artigo Aceito na <i>6th World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)</i>	113

VI Código R do AHP-TOPSIS-2N	115
Apêndice	117
A Contexto de atuação do MME	118
B Bases Abertas para <i>Energia Elétrica</i> catalogadas no PBDA	120
C Catálogo dos Dados Abertos	121
D Necessidades e Benefícios da OGD	124
E Critérios de Risco definidos	125
F Conjunto de Matrizes dos Subcritérios	127
G Código R - Ponderação dos Critérios pelo método AHP	128

Lista de Figuras

1.1	Relacionamento Metodológico - Principais áreas da Pesquisa.	9
1.2	Fases e Etapas da Pesquisa.	11
2.1	Evolução do Governo Aberto no Brasil: principais marcos legais.	16
2.2	Abstração da Publicação de Dados na Web.	18
2.3	Dados Abertos no contexto governamental e princípios relacionados.	21
2.4	Princípios do padrão <i>Linked Open Data</i>	24
2.5	Princípios Arquiteturais <i>Linked Open Data</i> .1 Evolução da Web.	27
2.6	Justaposição dos conceitos (Linked) (Open) (Government) (Data).	30
2.7	Estágios do Modelo CRISP-DM.	32
2.8	Níveis de detalhamento do Estágio de Entendimento do Contexto.	34
2.9	Etapas do Processo de Apoio à Decisão.	36
3.1	Macro etapas do Processo de mineração de texto.	54
3.2	Tópicos identificados a partir da análise computacional do PNE e PDE.	55
3.3	Quantitativo de bases abertas distribuídos por tópico	57
3.4	Identificação dos <i>datasets</i> por tópico.	57
3.5	Decomposição Hierárquica do Problema.	64
3.6	Resultado da Análise de Sensibilidade sobre os Critérios CR4 e CR5.	75
4.1	Tela da Página inicial do <i>AppShiny</i>	78
4.2	<i>Dashboards</i> projetados pelo <i>AppShiny</i>	79
4.3	<i>Ranking</i> das 48 Bases.	80
4.4	Comparativo das Bases e Impacto dos Critérios de Risco.	81
4.5	Média de desempenho das bases.	82
4.6	Bases com desempenho abaixo da média.	82
4.7	Resultado das Bases DB46, DB26 e DB43.	83
5.1	Processo de Gerenciamento de Riscos.	88
A.1	Contexto de atuação do MME.	118
B.1	Número de bases abertas para <i>energia elétrica</i> catalogadas no PBDA.	120

Lista de Tabelas

3.1	Escopo da Análise de Conteúdo.	52
3.2	Escala de Importância do AHP	66
3.3	Matriz de Julgamentos.	67
3.4	Cálculo do Autovalor Máximo/ <i>Eigen</i> Principal (λ_{max})	68
3.5	Índice de Consistência Randômico	69
3.6	<i>Ranking</i> e Resultados da aplicação do AHP-TOPSIS-2N.	71
F.1	Matriz de Subcritérios em relação ao Critério Formatação.	127
F.2	Matriz de Subcritérios em relação ao Critério Padronização.	127
F.3	Matriz de Subcritérios em relação ao Critério Rejeição.	127

Lista de Quadros

1	Revisão da Literatura acerca do conceito OGD.	28
2	Métodos de Apoio à Decisão e Quantidade de Critérios Identificados.	39
3	Principais Critérios utilizados pelo método <i>Decision Tree</i>	40
4	Critérios identificados e mensurados por Redes Bayesianas.	42
5	Critério de Riscos identificados e avaliados por Redes Bayesianas.	43
6	Questões-Chave para mensurar a priorização da abertura de dados.	44
7	Critérios identificados e mensurados pelo AHP-TOPSIS.	46
8	Comparativo dos Métodos Identificados.	47
9	Dimensões de Riscos e Benefícios da Abertura de Dados de Energia.	48
10	Critério/Dimensão e Subcritérios.	59
11	Critérios de Risco, Subcritérios e Função Objetivo.	70
12	Catálogo das bases abertas utilizadas no Modelo.	122
13	Necessidades e Benefícios da OGD demandados pelos <i>Stakeholders</i>	124
14	Critérios de Risco identificados a partir dos Princípios <i>Open Data</i> , LOD e OGD.	126

(APIs)Tecnologias de Informação e Comunicação (TIC)

Lista de Abreviaturas e Siglas

AECI Assessoria Especial de Controle Interno.

AHP Analytic-Hierarchy Process.

ANEEL Agência Nacional de Energia Elétrica.

APF Administração Pública Federal.

API Application Programming Interface.

CGU Controladoria-Geral da União.

CR Critério de Risco.

CRISP-DM Cross Industry Standard for Data Mining.

e-SIC Sistema Eletrônico de Informações ao Cidadão.

EGD Estratégia de Governo Digital.

EPE Empresa de Pesquisa Energética.

FAHP Fuzzy Analytical Hierarchy Process.

GDOI Government Data Openness Index.

GR Gestão de Riscos.

Guia 73 ABNT ISO GUIA 73:2009.

HTTP Hypertext Transfer Protocol.

IA Inteligência Artificial.

IGA Índice Governo Aberto.

INDA Infraestrutura Nacional de Dados Abertos.

IoT Internet of Things.

LAI Lei de Acesso à Informação.

LD Linked Data.

LDA Latent Dirichlet Allocation.

LOD Linked Open Data.

LRF Lei de Responsabilidade Fiscal.

MCDA Multicriteria Decision Analysis.

MCDM Multiple Criteria Decision Making.

MME Ministério de Minas e Energia.

OCDE Organização para a Cooperação e Desenvolvimento Econômico.

ODS Open Document Spreadsheet.

OGD Open Government Data.

OGP Open Government Partnership.

OKF Open Knowledge Foundation.

OWL Ontology Web Language.

PBDA Portal Brasileiro de Dados Abertos.

PDA Plano de Dados Abertos.

PDE Plano Decenal de Expansão de Energia.

PERT Program Evaluation and Review Technique.

PNE Plano Nacional de Energia.

PO Pesquisa Operacional.

RDF Resource Description Framework.

RDFS RDF Schema.

REST Representational State Transfer.

RSL Revisão Sistemática da Literatura.

SEE Secretaria de Energia Elétrica.

SPARQL SPARQL Protocol and RDF Query Language.

SPE Secretaria de Planejamento e Desenvolvimento Energético.

TEMAC Teoria do Enfoque Meta Analítico Consolidado.

TIC Tecnologias de Informação e Comunicação.

TOPSIS Technique for Order Preference by Similarity to Ideal Solution.

URI Uniform Resource Identifier.

VGCE Vocabulário Controlado de Governo Eletrônico.

W3C World Wide Web Consortium.

Capítulo 1

Introdução

1.1 Contextualização

A disponibilização de *Dados Abertos* (*Open Data*), ou seja, dados que podem ser livremente acessados, modificados, reutilizados e redistribuídos sem quaisquer tipo de restrições (legais, tecnológicas, sociais, etc.), permite manipulações e mapeamentos semânticos de forma colaborativa e automatizada entre diferentes domínios e aplicações. Na extensão desse conceito, *Dados Abertos Governamentais* se caracterizam pela disponibilização de dados públicos, indexados na *internet* em um formato compreensível por máquina e protegidos por licença que permita, no mínimo, a sua reutilização por qualquer pessoa ou entidade [1].

Organizações e comunidades de todo o mundo, sobretudo indivíduos "letrados em dados" (*data literacy*¹), têm reunido esforços para fomentar a abertura de dados produzidos e/ou custodiados com recursos públicos, impulsionando o chamado movimento OGD (acrônimo em inglês para *Open Government Data*) [2, 3]. Este movimento intenta encorajar a abertura de dados pelas instituições públicas ou privadas, de forma integrada aos seus processos internos (*business as usual*) e ajustada proporcionalmente aos recursos disponíveis de cada organização [4, 5].

No Brasil, obrigações impostas por normativos legais e infralegais² têm induzido o Governo Federal a assumir um papel crítico frente à abertura de dados, principalmente nos órgãos ministeriais, os quais foram incumbidos de conduzir e implementar a Política de Dados Abertos, instituída pelo Decreto nº 8.777, de 11 de maio de 2016³ [6]. Para a implementação dessa Política, os órgãos da Administração Pública Federal (APF) devem executar uma série de ações, formalizadas por meio do Plano de Dados Abertos (PDA), o qual deve ser editado a cada biênio.

¹Capacidade e conhecimento para lidar com dados, como parte de um processo mais amplo, ante a tomada de decisão.

²<https://www.gov.br/governodigital/pt-br/legislacao/legislacao-governo-aberto>

³https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/D8777.htm

O PDA é o instrumento que disciplina e orienta a abertura de dados nos órgãos governamentais brasileiros, definindo os critérios mínimos para entendimento, reutilização e divulgação de informações públicas em formato de *dados abertos* legíveis por máquina. Tal Plano deve lidar com desafios inerentes à governança digital e à infraestrutura tecnológica, no sentido de buscar a oferta de dados em sua forma primária ou, se possível, com o menor nível de granularidade possível (maior nível de detalhe)⁴.

No que concerne à governança digital, auxiliar decisões para avaliar *dados abertos*, bem como os critérios de riscos atinentes, de forma prática e efetiva, vai ao encontro de um modelo orientado à análise multicritério de apoio à decisão - *Multicriteria Decision Analysis*. Impende ressaltar, todavia, que abordagens que utilizam análise orientadas à riscos representam um campo incipiente, pouco explorado pelas estratégias de governo.

Quanto à infraestrutura tecnológica, observa-se que as tecnologias disruptivas, que alcançam as arquiteturas dependentes de dados brutos, ainda carecem de maior aproveitamento de seu potencial no âmbito da APF. Dentre essas tecnologias, destacam-se a *Self-Service Data Preparation/Business Intelligence*, Inteligência Artificial (IA), *public blockchains*, *Internet of Things* (IoT), tecnologias *Web Semânticas*, *ontology based data access*, que fomentaram o aumento dos dados produzidos, além dos modelos preditivos e de análise de apoio à decisão, que visam diminuir incertezas e subsidiar a tomada de decisão em relação ao direcionamento de recursos para a disponibilização de *dados abertos governamentais*.

Diante então desse contexto, vislumbra-se um vasto campo a ser desenvolvido na APF, especificamente no Ministério de Minas e Energia (MME), Órgão que representa a União na elaboração e supervisão de políticas e diretrizes dos setores energético⁵ e mineral, tendo em vista que os *dados abertos* se tornam ativos relevantes e essenciais, tanto para a construção de soluções que permitam explorar cenários alternativos quanto para decisões que exijam formulação de políticas públicas, *due diligence*, investimentos/desinvestimentos, entre outras ações governamentais.

No exercício do papel estratégico de implementar o Plano de Transformação Digital, o MME busca alcançar, entre outros objetivos, a *"reformulação dos canais de transparência e dados abertos"*, expresso no Decreto nº 10.332, de 28 de abril de 2020, [8], que instituiu a Estratégia de Governo Digital (EGD).

⁴Para Inmon [7], granularidade indica o nível de detalhamento do dado: quanto maior o nível de detalhamento, menor o nível de granularidade do dado. Por outro lado, quanto mais alto o nível de granularidade, menor o nível de detalhe do dado.

⁵O setor energético inclui os atores envolvidos na geração, transmissão, distribuição, regulação, comercialização, gestão e planejamento de energia, abrangendo além do setor elétrico, os setores de petróleo, gás e biocombustíveis.

1.1.1 Escopo

Considerando a abrangência do contexto e escopo de atuação do MME, este estudo se delimita à construção de um modelo que apoie as decisões de gestores da Secretaria de Energia Elétrica (SEE) e da Secretaria de Planejamento e Desenvolvimento Energético (SPE) do MME quanto à avaliação de dados abertos, ou potenciais para abertura, produzidos/custodiados por aquelas Secretarias.

O setor elétrico compreende uma área de negócio do MME estratégica para o País, envolvendo importantes recursos naturais para o desenvolvimento econômico e socio-ambiental. Além disso, a diversidade de atos legais e regulatórios, *stakeholders* e ativos de dados pertinentes aos diferentes sistemas/agentes desse setor (Geração, Transmissão, Distribuição, Comercialização, Regulação/Fiscalização e Agregação) justificam a utilização das bases de dados desse setor como plano de fundo deste trabalho.

De parte desse escopo, este estudo de caso alcança as unidades pertencentes ao setor elétrico, internas e vinculadas ao MME, as quais estão destacadas na Figura A.1 do Apêndice A desta dissertação.

Em relação aos ministérios, a Controladoria-Geral da União (CGU), que é o Órgão central do Sistema de Controle Interno do Poder Executivo Federal, conta com o apoio da Assessoria Especial de Controle Interno (AECI) para fomentar e supervisionar assuntos afetos à governança, riscos, integridade, ética e transparência (incluindo o eixo *dados abertos*), devendo orientar e assessorar o respectivo Ministro de Estado e demais gestores públicos acerca desses assuntos⁶.

O MME, por ser Órgão supervisor dos segmentos de energia e mineração, tem relevante papel na condução de políticas públicas dessas áreas, incluindo a Política de Dados Abertos do setor elétrico. Para assegurar o cumprimento dessa Política, instituiu-se a "Autoridade de Monitoramento da Lei de Acesso à Informação (LAI)"⁷, função atribuída à AECI desse Ministério⁸. Essa Autoridade de Monitoramento, além de assegurar o cumprimento da LAI, deve orientar e produzir recomendações para a completa implementação e aperfeiçoamento do PDA.

1.1.2 Delimitação

Um dos maiores desafios da OGD é fomentar a descoberta de dados adequados e pertinentes para a tomada de decisão. O fácil acesso é uma das principais vertentes do conceito de dados abertos, não bastando apenas publicá-los na *internet*, mas vinculá-los a um diretório

⁶Artigo 13 do Decreto nº 3.591, de 06 de setembro de 2000.

⁷Artigo 40 da Lei nº 12.527, de 18 de novembro de 2011.

⁸<https://www.gov.br/mme/pt-br/acao-a-informacao/legislacao/portarias/2017/portaria-n-408-2017.pdf/view>

online que ofereça uma interface para a publicação dos conjuntos de dados e metadados correspondentes. Para isso, foram criadas plataformas digitais que reúnem dados de diferentes fontes e formatos, a exemplo do Portal Brasileiro de Dados Abertos (PBDA)⁹.

O PBDA é a página de referência do governo federal, disponibilizado pela Infraestrutura Nacional de Dados Abertos (INDA)¹⁰, que reúne e estrutura (como um catálogo *online*) dados abertos de diferentes versões (distribuições), de forma que os dados públicos, e respectivos metadados, possam ser facilmente encontrados, acessados e utilizados por diferentes atores ou ferramentas informatizadas, o que justifica delimitar esta pesquisa apenas aos conjuntos de dados publicados no referido Portal.

No que diz respeito a *dados abertos*, a LAI determina, em seu artigo 8º, que os órgãos públicos divulguem seus dados no PBDA, em formato aberto e não proprietário, de modo a facilitar a análise, a leitura e a coleta automatizada desses dados.

Atualmente esse Portal, que é gerenciado pela CGU, possui mais de 12 mil conjuntos de dados e aproximadamente 60 mil recursos catalogados por 211 organizações¹¹. Cada *dataset* deve possuir, no mínimo, um recurso de dados que pode ser acessado via um arquivo, uma planilha ou uma *Application Programming Interface (API)*. Apesar disso, constatou-se, conforme demonstrado na Figura B.1 do Apêndice B, uma pequena quantidade de conjuntos de dados (50) do setor elétrico catalogados no PBDA por órgãos governamentais desse segmento.

A baixa quantidade de dados do setor elétrico catalogadas no PBDA, conforme observado no *print* do portal PBDA (vide Figura B.1 do Apêndice B), remete à ANEEL, e principalmente ao MME, a realização de pesquisas para verificar incertezas e riscos relacionados a essas bases ou à identificação de tópicos potenciais para a abertura de novos dados.

1.2 Descrição do Problema

Na última década, nota-se a publicação de leis, decretos, entre outros normativos nacionais e internacionais que fomentam iniciativas e estratégias para a implementação e publicações de conjuntos de dados no formato aberto (vide Seção 2.1), tendo como exemplo a atual Política de Dados Aberto (Decreto nº 8.777/2016). No entanto, sua implantação mais ampla requer a identificação dos obstáculos que impedem os órgãos públicos de realizarem uma estratégia de Dados Abertos [9].

⁹<https://dados.gov.br/>.

¹⁰A INDA é um conjunto de princípios, padrões, tecnologias, procedimentos e ações de controle necessários à estruturação e compartilhamento de dados abertos produzidos ou custodiados pela APF.

¹¹Pesquisa atualizada em 28 de junho de 2022.

O PDA é um instrumento de governo criado para executar a Política de Dados Aberto em determinado órgão, que deve promover e definir critérios para a ampliação da quantidade de dados abertos disponibilizados no PBDA, bem como a correção e atualização dessas bases, conforme preconizado pela EGD e pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE)¹².

Apesar das ações implementadas pelo PDA, evidência-se, em recente estudo apresentado por Palma A., Ladeira e Reis [10], a crescente demanda por dados e informações afetos ao setor elétrico, inferindo-se, assim, que tais dados não foram encontrados ou publicados adequadamente no PBDA ou, se disponíveis, estão desatualizados, defasados ou errados. No citado artigo, os autores utilizaram a abordagem *topic modeling* e técnicas de mineração de texto para mapear os tópicos dos assuntos mais solicitados, pela sociedade e demais partes interessadas, via Ouvidoria e/ou Sistema Eletrônico de Informações ao Cidadão (e-SIC)¹³ no âmbito do MME. Os resultados apontaram que os temas mais frequentes estão relacionados aos assuntos de competência da ANEEL, o que ratifica o escopo mencionado na Subseção 1.1.2.

Além do exposto, salienta-se que o processo de abertura de dados¹⁴ é uma atividade complexa, podendo resultar em benefícios (oportunidades) ou em riscos (ameaças) potenciais para as partes interessadas [11]. Em relação aos riscos e incertezas que podem acarretar, por exemplo, em tomadas de decisão equivocadas ou fundamentadas em diagnósticos tendenciosos, a partir de dados inadequados, incorretos e/ou desatualizados, citam-se fatores como imprecisão, confidencialidade, violação de privacidade, inconsistência e uso indevido de dados abertos [12].

Se por um lado a OGD cria oportunidades para o desenvolvimento econômico [13], incentiva as produções científicas [14, 15, 16] e fomenta a participação da sociedade no governo, informando-se acerca das tomadas de decisões do gestor público [17], por outro, potencializa riscos a partir do uso indevido dos dados abertos inadequadamente, ou até mesmo o não uso dessas bases abertas, gerando retrabalho ou desperdício de recursos públicos. Esses riscos e incertezas podem resultar, inclusive, na aversão dos gestores em abrir suas bases e, conseqüentemente, na relutância das organizações em aderir à iniciativa OGD [18].

Os pontos mencionados suscitam uma lacuna de pesquisa frente à incipiência de modelos/métodos formais que subsidie a tomada de decisão para avaliar dados abertos e, conseqüentemente, aprimorar o correspondente processo de abertura e publicação dos *datasets* avaliados. Assim, no contexto governamental, apoiar a avaliação de dados do

¹²<https://www.oecd.org/digital/digital-government/open-government-data.htm>.

¹³O Sistema e-SIC está integrado à Plataforma Integrada de Ouvidoria e Acesso à Informação, ora denominada Fala.BR.

¹⁴<https://kit.dados.gov.br/Abertura-de-dados/>.

setor elétrico, no sentido de que as bases disponibilizadas no PBDA sejam acompanhadas de forma sistemática e menos subjetiva, ilustra o problema de pesquisa explorado neste trabalho, no âmbito do MME.

Em linha com o citado problema, Relly e Sabharwal [19] argumentam que a ausência de estruturas tecnológicas, institucionais e de governança, dificulta a implementação efetiva da abertura de dados no setor público [14, 1, 20, 4]. Este problema ainda se agrava com a falta de um modelo automatizado de gerenciamento de dados abertos publicados nos repositórios eletrônicos oficiais do governo.

Dessa forma, avaliar dados abertos do setor elétrico, de modo que as partes interessadas confiem na integridade, atualidade e fiabilidade dessas bases, evidencia um importante papel a ser desempenhado pelo MME, fomentando conceitos disruptivos como: Dados Abertos Governamentais Conectados (*Linked Open Government Data*), Web Semântica (Web 2.0), Indústria Conectada 4.0, Governo Eletrônico (*e-Government*), e-Participação, Governo Digital, *Open Energy*, entre outros.

Questões de Pesquisa

Para a estruturação do problema da pesquisa, em vista da abrangência da meta-análise com recursos e tempo limitados, recorreu-se à estratégia P.I.C.O.C. [21] (acrônimo para Problema, Interesse, Controles, Resultados/*Outcomes* e Contexto). Este protocolo norteou a formulação das seguintes questões¹⁵ de pesquisa, a fim de orientar a Revisão Sistemática da Literatura (RSL):

- QP1) Que critérios de risco são relevantes para a abertura de dados governamentais?
- QP2) Quais são os tópicos/temas do setor elétrico que se revelam potenciais para a abertura de dados?
- QP3) Como avaliar se os conjuntos de dados do setor elétrico (catalogados no PBDA) estão ou não aderentes aos princípios da *Open Government Data*?

1.3 Justificativa

É de amplo conhecimento a abrangência dos impactos positivos potencialmente trazidos pelo movimento OGD. Ao tempo que estimula a interação e a colaboração entre diversos atores da sociedade, promove transparência, *accountability*¹⁶, eficiência da gestão pública,

¹⁵Cabe ressaltar que QP1 e QP2 são questões de filtro para seleção e composição da amostra final (QP3).

¹⁶*Accountability* refere-se à prestação de contas e responsabilização pelos atos de gestão.

participação dos cidadãos nos processos decisórios do governo, bem como cria, a partir da conversão de dados brutos em valor agregado, oportunidades para otimizar as tomadas de decisão [22, 23, 24, 25]. Além disso, princípios éticos sugerem que qualquer dado ou informação produzido ou custeado com recursos do governo deve ser público, com exceção dos dados protegidos por sigilo¹⁷.

Segundo Molloy [15], quanto mais dados forem abertos e disponibilizados de maneira adequada e útil, maior será o nível de transparência e de reprodutibilidade e, consequentemente, mais eficiente se torna o progresso científico, em benefício da sociedade.

No contexto governamental, tem-se que as políticas públicas podem ser aprimoradas por meio da análise de um grande volume de dados. Nessa perspectiva, a abertura de dados possibilita *feedbacks* constantes, controle e monitoramento do uso de bens e recursos públicos, bem como participação nas decisões e estratégias de governo, contribuindo assim para um cenário de aumento da confiança pública, viabilização de negócios, serviços inovadores, estímulo à economia e competitividade comercial [5].

Diante da complexidade e mudanças recorrentes no setor elétrico, análises cada vez mais precisas são exigidas, no sentido de apoiar decisões que alcancem benefícios que possam atender às necessidades de diferentes *stakeholders* desse setor. O uso de dados abertos nesse cenário pode compreender desde a análise de investimentos para projeções de preços na realização de novos leilões, para o planejamento e expansão do setor, até o aprimoramento ou desenvolvimento de novas políticas, aplicativos e modelagens. Para isso, segundo Hirth [26], se faz necessária a abertura de um grande volume de dados que não apenas estejam disponíveis, mas que possuam qualidade em relação à conformidade aos princípios do padrão *Open Data*.

Conforme apresentado no Quadro 13, referenciado no Apêndice D, múltiplos *stakeholders* (sociedade civil, servidores públicos, pesquisadores/cientistas, jornalistas, entre outros) têm demandado cada vez mais a disponibilização de dados em formatos abertos, sobretudo, diante das necessidades e benefícios promovidos pela abertura de dados governamentais. Sob outra perspectiva, Chen et al. [27] e a OECD [28] ponderam que apenas abrir e disponibilizar dados e informações não é suficiente para a entrega de valor público. Nessa linha, Lourenço [17] aduz que disponibilizar uma quantidade grande de dados não implica em maior transparência, tampouco facilita a *accountability*. Essas ponderações vão ao encontro do conceito “*effective data use*” (uso efetivo de dados), abordado em [4].

No Brasil, a APF não apenas é responsável pela coordenação e execução da Política de Dados Abertos, como tem o dever de verificar o custo-benefício para a abertura de determinado conjunto de dados ou sistema. Incumbe aos órgãos e entidades governamentais

¹⁷Destaca-se que esta pesquisa não avaliou o conteúdo dos dados abertos do setor elétrico no âmbito da Lei Geral de Proteção de Dados Pessoal.

tomar decisões assertivas, considerando os principais critérios para a seleção e priorização das alternativas pertinentes ao movimento OGD, no sentido de buscar maior eficiência na aplicação de recursos públicos destinados ao processo de abertura de dados.

O governo, como principal indutor e produtor de dados e informações públicas, é instado a avaliar os critérios de riscos envolvidos nos projetos de abertura e catalogação de seus dados e metadados nos respectivos portais institucionais, bem como verificar os custos relacionados aos possíveis benefícios da OGD, a exemplo daqueles listados no Quadro 13 do Apêndice D.

Dados Abertos são uma ferramenta essencial para a disseminação dos princípios do *Open Government Data*. No entanto, sua implantação mais ampla requer a identificação de riscos e incertezas que impedem os avanços do Governo Digital na implementação estratégias de *Dados Abertos* ou de garantir sua sustentabilidade [9].

Dessa forma, para conduzir esta pesquisa, foram formulados os objetivos geral e específicos descritos a seguir.

1.4 Objetivos

Objetivo Geral

O objetivo geral deste trabalho é propor um modelo de apoio à decisão que permita, à luz dos critérios de riscos relacionados aos princípios *Linked Open Data* e *Open Government Data*, avaliar o nível de abertura das bases do setor elétrico catalogadas no Portal Brasileiro de Dados Abertos.

Objetivos Específicos

Os objetivos específicos deste trabalho são:

- a) Definir os critérios de risco relevantes para o contexto de Dados Governamentais Abertos, de modo geral e específico ao setor elétrico;
- b) Identificar o(s) método(s) de apoio à decisão aderente(s) à problemática estudada;
- c) Estabelecer os potenciais temas para a abertura de dados do setor elétrico;
- d) Construir o modelo, considerando os critérios de riscos definidos e o(s) método(s) identificado(s); e
- e) Validar o modelo proposto por meio da Análise de Sensibilidade, considerando os comentários dos gestores/especialistas.

1.5 Metodologia

A metodologia é uma instância de um modelo, ou de diferentes modelos, na qual se utiliza de técnicas e ferramentas para apoiar a realização das tarefas estabelecidas. Ou seja, se por um lado o modelo define *o quê fazer* por outro, a metodologia define *como fazer* [29].

Dessa forma, apresenta-se nesta Seção a estrutura metodológica utilizada para o alcance dos objetivos supracitados. Ancorada pela abordagem quali-quantitativa e etapas estabelecidas por Marconi e Lakatos [30] e Prodanov e De Freitas [31], esta pesquisa se classifica como descritiva e exploratória, tendo como finalidade a aplicação de um modelo de apoio à decisão para avaliar dados abertos governamentais do setor elétrico.

1.5.1 Classificação e Fases da Pesquisa

Segundo Cervo e Bervian [32], o projeto da pesquisa estabelece uma ordem imposta aos diversos processos que devem ser executados dentro de um cronograma definido, a fim de alcançar a resolução de um problema específico. Assim, a metodologia desta pesquisa se centra em macroprocessos que se inter-relacionam conforme representado na Figura 1.1.

Esta pesquisa, quanto a sua natureza, se classifica como **aplicada**, a fim de produzir conhecimento direcionado à resolução efetiva de um problema específico no Ministério de Minas e Energia, no sentido de contribuir, na prática, para a melhoria do processo de abertura de dados naquele Ministério.

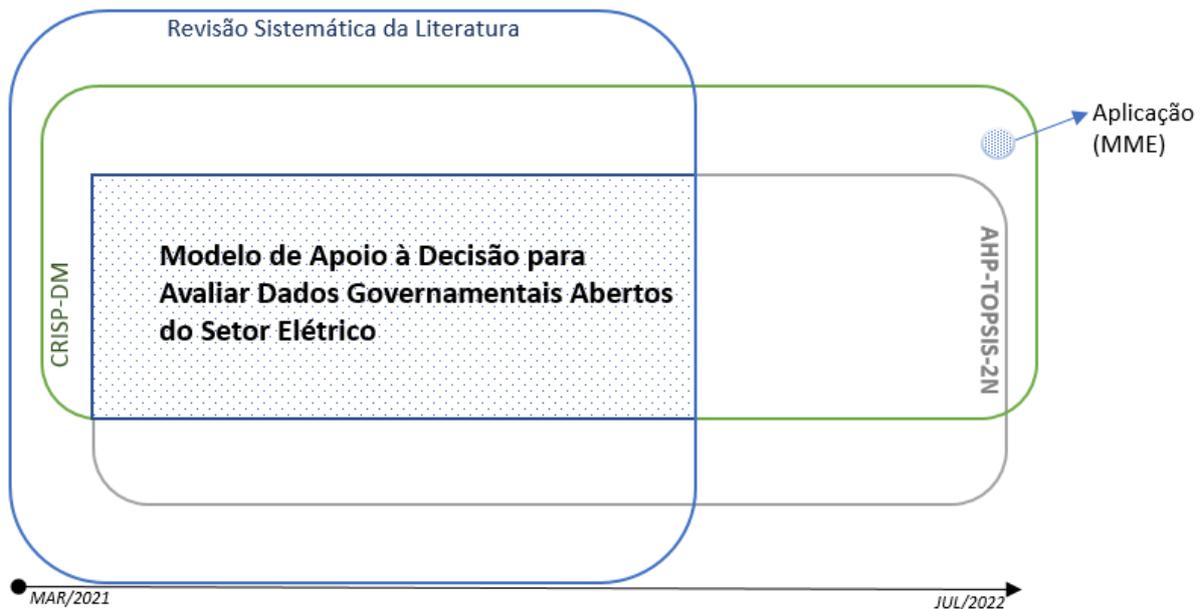


Figura 1.1: Relacionamento Metodológico - Principais áreas da Pesquisa.

Fonte: Adaptado de Triperina [33].

No que diz respeito aos objetivos desta pesquisa, esta se classifica como **exploratória**, porque, além de favorecer à condução de um estudo mais completo frente ao conhecimento prático produzido, buscando observar e analisar uma temática ainda incipiente na literatura acadêmica, que é o contexto de dados abertos do setor elétrico nacional e seus respectivos critérios de riscos. Também se classifica como pesquisa **descritiva**, tendo em vista a Revisão Sistemática da Literatura (RSL) realizada acerca dos principais métodos MCDA e critérios de riscos considerados na construção de um **estudo de caso**, para aplicação de um modelo que subsidie as decisões dos gestores responsáveis pela Política de Dados Abertos no MME.

1.5.2 Plano Metodológico

Com base no exposto, as Fases Planejamento, Execução, Estudo de Caso e Redacional sistematizam a sequência de procedimentos e métodos seguidos neste trabalho, conforme delineado no **Plano Metodológico** representado pela Figura 1.2. As mencionadas fases, e as correspondentes etapas, adaptadas ao contexto deste trabalho, assentam-se aos eixos/fases propostos por Gerhardt e Silveira [34] e Prodanov e De Freitas [31], respectivamente, a ver:

- a) Os três eixos da metodologia de pesquisa, segundo Gerhardt e Silveira [34]:
 - Ruptura: consiste em se sobrepor aos conceitos antecipados pelo senso comum e evidências não fundamentadas em dados.
 - Construção: constrói um sistema conceitual, formal e organizado, de forma que possa ser compreendido e replicado no plano de pesquisa a ser observado.
 - Constatação: relata a proposição para formar a produção científica à luz dos fatos verificados.

- b) As três fases da pesquisa, conforme Prodanov e De Freitas [31]:
 - Fase Decisória: refere-se à definição do tema e do problema de pesquisa;
 - Fase Construtiva: trata da execução do plano metodológico; e
 - Fase Redacional: apresenta a análise e avaliação dos resultados obtidos.

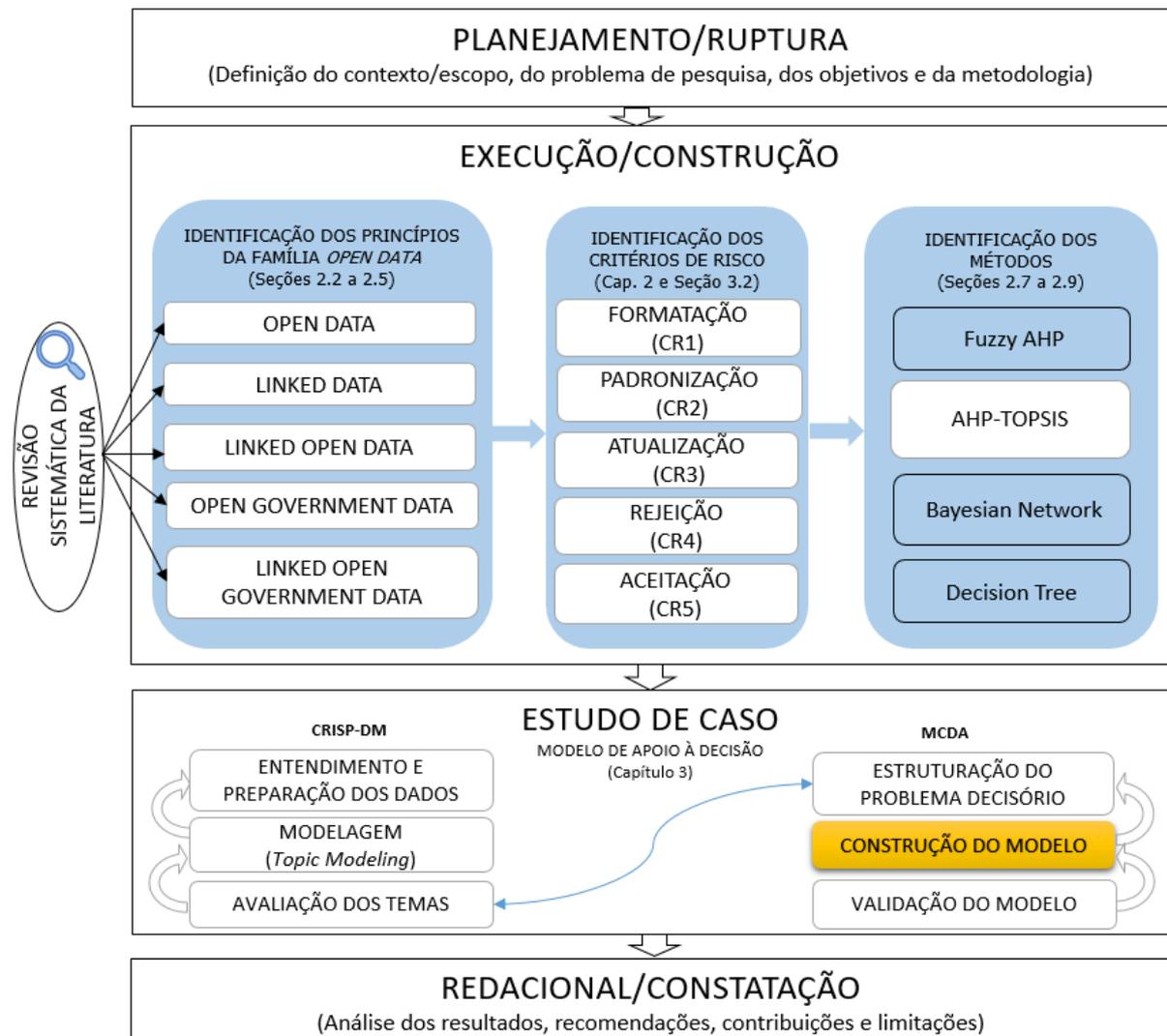


Figura 1.2: Fases e Etapas da Pesquisa.
Fonte: Elaboração própria.

Superada a Fase Planejamento/Ruptura, as Fases Execução e Estudo de Caso implementam os seguintes instrumentos e etapas, necessários para o alcance dos objetivos desta pesquisa, descritos a seguir:

- Revisão Sistemática da Literatura e aplicação do método Teoria do Enfoque Meta Analítico Consolidado (TEMAC)¹⁸ [35], que norteou a pesquisa bibliográfica para a **identificação dos critérios de risco e dos métodos** empregados nos trabalhos relacionados;
- o modelo de referência para execução das tarefas de mineração de dados/textos Cross Industry Standard for Data Mining (CRISP-DM) [36], para **Entendimento e Preparação dos Dados (3.1.1)**, **Modelagem (3.1.2)** e **Avaliação (3.1.3)**,

¹⁸<https://www.pesquisatemac.com/>.

necessários à identificação e **priorização dos temas potenciais para abertura de dados**;

- a técnica *web scraping*, para coletar e organizar metadados dos *datasets* do setor elétrico publicadas no PBDA, no sentido de encontrar e definir os valores adequados para a **definição e mensuração dos critérios identificados**;
- a abordagem Multicriteria Decision Analysis (MCDA) [37], para a **Estruturação do Problema Decisório e Construção do modelo de apoio à decisão** proposto neste trabalho;
- Análise de Sensibilidade e Entrevista não estruturada¹⁹ [38], realizada entre a pesquisadora e representantes do Núcleo de Dados Abertos da Controladoria-Geral da União²⁰, com o objetivo de apresentar o escopo da pesquisa e os critérios identificados com vistas à **Validação do Modelo**; e
- O pacote *Shiny* do Software R Studio²¹ para o desenvolvimento de uma Aplicação *Web*²², que permite apresentar, de forma interativa e dinâmica, os resultados encontrados para análise e avaliação dos especialistas.

Conforme destacadas na Figura 1.2, as etapas Entendimento e Preparação do Dados, Modelagem e Avaliação, que compõem o Modelo Cross Industry Standard for Data Mining (CRISP-DM), detalhadas na Seção 3.1, se constituem, em paralelo, das etapas do modelo do processo Multicriteria Decision Analysis (MCDA): Estruturação do Problema Decisório, Construção e Validação, descritas na Seção 3.2.

1.6 Estrutura da Dissertação

Esta dissertação está estruturada em cinco capítulos, a saber:

Capítulo 1. Introdução: a **Seção 1.1** contextualiza e situa conceitos-chave, atores e o problema de pesquisa abordado, além de delimitar o tema e o escopo deste trabalho. A **Seção 1.2** detalha o problema a ser estudado e os fatores motivacionais para compreensão e resolução das perguntas de pesquisa suscitadas para assim destacar, na **Seção 1.3**, os principais fatores que motivaram a definição dos objetivos descritos na **Seção 1.4**. Já a **Seção 1.5** se presta a identificar e estruturar as etapas relacionadas ao plano metodológico,

¹⁹Agenda *online*, via Microsoft Teams, ocorrida em 03 de dezembro 2021.

²⁰dadosabertos@cgu.gov.br.

²¹<https://www.rstudio.com/>.

²²<https://ingridpalma.shinyapps.io/ahptosis/>.

executado durante o desenvolvimento deste estudo de caso, utilizando-se de instrumentos e estratégias reconhecidos na literatura para o alcance dos objetivos definidos na Seção anterior.

Capítulo 2. Base Conceitual e Revisão da Literatura: permite compreender melhor os principais conceitos e termos que circundam a temática deste trabalho, por meio de uma análise exploratória de documentos produzidos pelo governo, bem como da revisão sistemática da literatura, revelando aspectos técnicos e legais que envolvem o padrão *Open Data*. Apresenta-se, nas **Seções 2.1 a 2.5**, um panorama dos aspectos legal e histórico do paradigma *Governo Aberto* e a evolução do conjunto de regras relacionados ao conceito de *Dados Abertos*, considerando o contexto técnico e os Princípios arquiteturais deste padrão, por meio de uma visão *top-down* da revisão da literatura, que fundamentam a identificação dos critérios relacionados aos princípios *Linked Open Data* e *Open Government Data*. Nas **Seções 2.6 a 2.8** são apresentadas as definições e conceitos essenciais para a melhor compreensão deste trabalho. A **Seção 2.9** trata dos principais trabalhos, relacionados a esta pesquisa, que utilizam métodos e técnicas multicritérios sobre dados abertos, inclusive, do setor elétrico.

Capítulo 3. Estudo de Caso: as fases do processo de desenvolvimento do Modelo CRISP-DM, as quais organizam e disciplinam o Entendimento, Preparação, Pré-Processamento e Modelagem dos Dados, estão descritas na **Seção 3.1**. A estruturação e construção do Modelo de Apoio à Decisão estão detalhadas na **Seção 3.2**, com destaque para a **Subseção 3.2.2**, que explora e define os critérios de risco relevantes para esse Modelo, etapa prévia à estruturação e decomposição hierárquica da problemática (**Subseção 3.2.3**) e aplicação das técnicas AHP e TOPSIS (**Subseção 3.3**). A **Subseção 3.4** versa acerca da Análise de Sensibilidade, bem como do teste de robustez realizado pelo algoritmo do TOPSIS-2N para validação do Modelo.

Capítulos 4 e 5. Resultados e Considerações Finais: discutem-se os resultados e as perspectivas deste trabalho quanto às contribuições tecnológicas e institucionais deste trabalho, aduzindo as implicações e relevância do Modelo para os publicadores de dados abertos do Ministério de Minas e Energia, entre outros agentes públicos. Também são apresentadas as contribuições acadêmicas, correspondentes aos artigos produzidos (publicados/aceitos) ao longo deste trabalho e, por fim, as reflexões desta pesquisa insculpidas nas considerações finais.

Capítulo 2

Base Conceitual e Revisão da Literatura

Apresenta-se, neste Capítulo, definições dos principais conceitos relacionados ao padrão *Open Data*, de modo a fornecer informações sobre seu arranjo arquitetural e os respectivos critérios de riscos que compreendem o conjunto de princípios desse ambiente. Previamente aos Trabalhos Relacionados (2.9), mostra-se necessário fornecer a definição de termos-chave utilizados neste trabalho, a fim de esclarecer a distinção desses conceitos entre diferentes disciplinas, tais como a Pesquisa Operacional (PO) e a Gestão de Riscos (GR).

Dessa forma, pretende-se, no final deste Capítulo, poder responder as seguintes questões: O que são dados abertos e quais são as suas variações tecnológicas e conceituais? Como e por quê surgiu essa temática no mundo e qual o arcabouço legal que trata da matéria no Brasil? De que forma o modelo proposto neste trabalho pode contribuir para o adequação de dados abertos do setor elétrico ao padrão *Open Data*?

2.1 Panorama e Evolução Legal do Governo Aberto

Conforme pesquisa realizada por Pinho e Silva [39], a ideia *Open Government* (Governo Aberto) começou a ser esboçada na Suécia, a partir do século XVIII, por meio da, ora denominada, *Lei de Liberdade de Imprensa*, datada de 1766.

Esse ato constitui o primeiro marco legal do movimento Governo Aberto no mundo. No entanto, somente após mais de 200 anos da promulgação dessa Lei, é que esse conceito começou a emergir também nos demais países.

Quanto à frase “*o direito de saber*”, aparentemente foi cunhada na década de 1940 por Kent Cooper, que era o diretor executivo da Associated Press. O New York Times

atribuiu-lhe a origem da frase em um editorial em 23 de janeiro de 1945¹.

Nos Estados Unidos, o termo Governo Aberto foi utilizado por Parks [40], em seu artigo intitulado *The open government principle: Applying the right to know under the constitution*, publicado no periódico *The George Washington Law Review*, em 1957. Esse marco legal e literário, segundo Jiménez [41] e Pinho e Silva [39], configurou a primeira aparição documentada do conceito **Governo Aberto** e inspirou a edição de normativos futuros que trataram do acesso à informação, da transparência (*disclosure*) e da prestação de contas (*accountability*).

Não impressiona, pelo contexto histórico em lide, que a Suécia lidere o *ranking* da Abertura de Dados Governamentais no mundo (*score* 0.81) - seguida pela Nova Zelândia e Noruega. Referido *ranking*, criado pela organização internacional *World Justice Project* (WJP)², tem por base o Índice Governo Aberto (IGA) e possui o objetivo de mensurar o nível de abertura de dados governamentais. Mediante o IGA, avaliou-se a efetiva abertura de dados governamentais em 125 países, considerando as seguintes perspectivas: Publicidade de Leis e Dados Governamentais, Direito à Informação, Participação Cívica e Mecanismos de Denúncia. Ressalta-se que o Brasil (*score*, 0.56) ocupa o 38º lugar no *ranking* *Governo Aberto*, criado pela WJP. Os EUA ocupam a 11ª posição dessa classificação, com um *score* de 0.73.

2.1.1 Governo Aberto no Brasil

No Brasil, do ponto de vista jurídico, o conceito *Open Government* começou a ser esboçado em 1988, por meio da Constituição Federal [42], consolidando-se duas décadas mais tarde, quando da publicação da Lei de Acesso à Informação (LAI) [43] e da Política de Dados Abertos [6], conforme ilustra a Figura 2.1. Todavia, somente com a entrada em vigor, no dia 16 de maio de 2012, que a Lei de Acesso à Informação efetivamente concretizou a garantia constitucional do acesso à informação no país.

Observa-se na Figura 2.1, que durante 20 anos (entre 1980 e 2000), registrou-se apenas o marco da Constituição de 1988 no que tange à abertura de dados/informação. Após isso, o assunto só retomou com a edição da Lei de Responsabilidade Fiscal (LRF), importante instrumento de transparência dos gastos públicos.

Esse movimento de abertura de dados e informações efetivamente começou a ter ênfase somente a partir de 2010, quando da criação do Portal da Transparência³, que representa um importante marco para a evolução do acesso a dados no Brasil. Trata-se de uma

¹Conforme observado por James S. Pope no Prefácio a “The People’s Right to Know” por Harold L. Cross, Columbia University Press, 1953, p. xi.

²<https://worldjusticeproject.org/open-government-around-world>.

³Disponível em <http://www.portaltransparencia.gov.br/>.

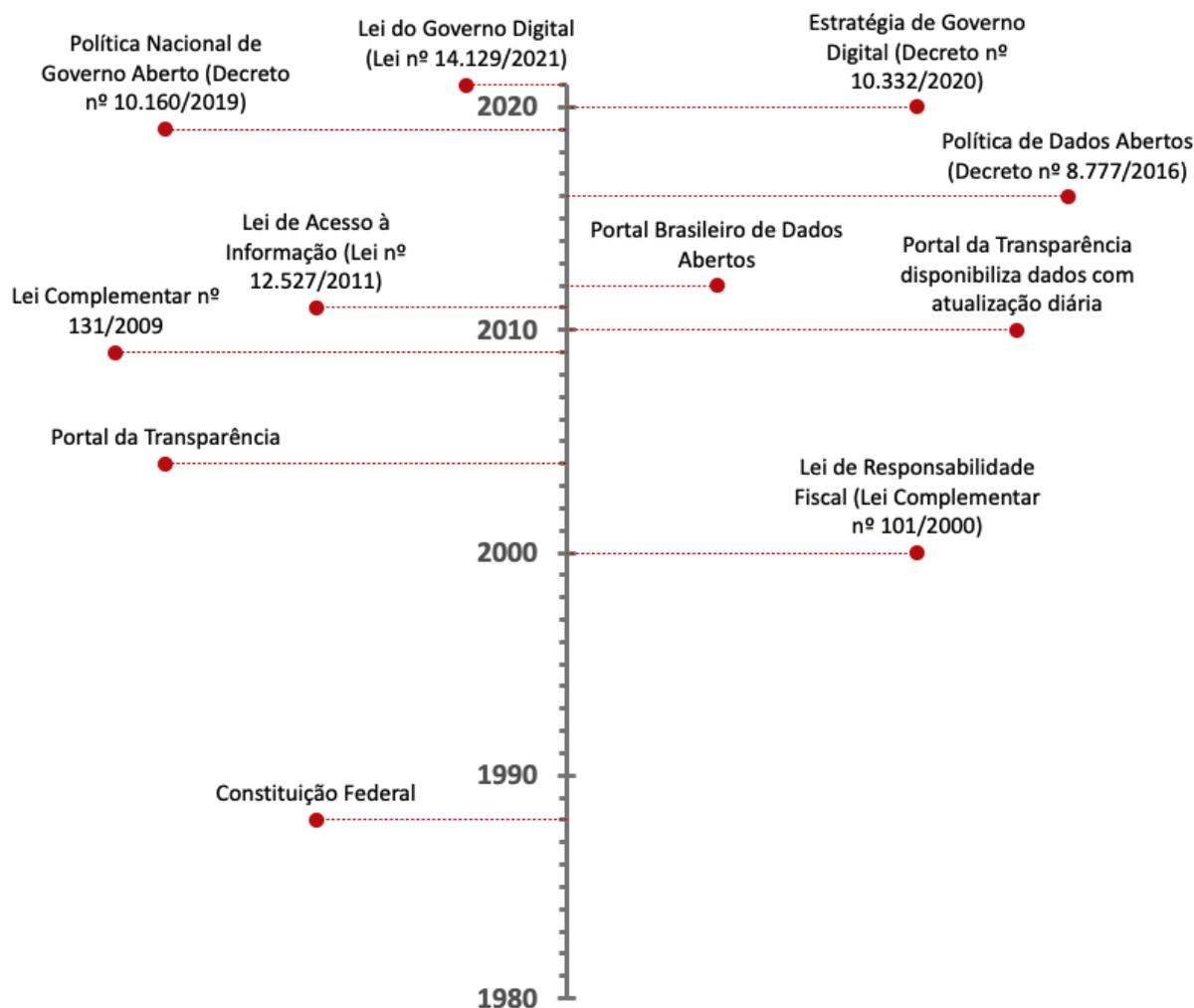


Figura 2.1: Evolução do Governo Aberto no Brasil: principais marcos legais.

Fonte: Elaboração Própria.

plataforma de integração e cruzamento (*mashup*) de dados de diferentes bases e sistemas estruturantes do Governo Federal.

Essa evolução de normativos, representada na Figura 2.1, assentou o arcabouço jurídico que materializou o conceito *Open Government Data* e demandou que entes públicos transformassem informações públicas em dados no formato aberto.

Em 2016, a Política de Dados Abertos (Decreto nº 8.777/2016) [6] decretou que órgãos da APF implementassem mecanismos de priorização de abertura de dados⁴, os quais devem obedecer os critérios estabelecidos pelo colegiado da INDA, publicados na Resolução nº 3, de 13 de outubro de 2017. Destarte, os critérios definidos pela INDA são:

- 1) Grau de relevância para a sociedade;
- 2) Dados que contribuem para o controle vertical (social);

⁴Vide inciso II do § 2º do art. 5º do Decreto nº 8.777/2016.

- 3) Obrigação legal ou compromisso assumido para disponibilização dos dados abertos;
- 4) Dados relacionados a projetos estratégicos do governo;
- 5) Relacionados a resultados diretos e efetivos do setor público;
- 6) Contribuição com o desenvolvimento sustentável e negócios na sociedade; e
- 7) Dados mais demandados pela sociedade, requeridos via LAI (transparência passiva).

Em atenção ao critério 1 dessa lista, destaca-se a metodologia percorrida na Subseção 3.1.2, que buscou encontrar os temas mais relevantes para a abertura de dados do setor elétrico. O critério 7 foi objeto do estudo publicado por Palma A., Ladeira e Reis [10]⁵, disposto no Anexo III deste trabalho, o qual apresentou os tópicos dos assuntos mais demandados, via LAI, no âmbito do MME.

Governo Aberto ganhou ainda mais robustez com a entrada em vigor da Lei nº 14.129, de 29 de março de 2021, a Lei do Governo Digital, que priorizou a disponibilização de dados e informações em uma única plataforma, de modo a incentivar a participação e colaboração no uso e compartilhamento de dados.

Destarte, órgãos e entidades da APF devem consignar, em seus respectivos Planos de Dados Abertos, as ações e os cronogramas de implementação do processo de abertura de dados, que precisam estar alinhados aos compromissos assumidos na *Open Government Partnership* (OGP) (Parceria para Governo Aberto)⁶

Na tese apresentada por Albano [44], observa-se nítida separação dos conceitos *Governo Aberto* (eixo estratégico) e *Dados Abertos* (eixo técnico). Enquanto Governo Aberto trata da "*disponibilização de informações em qualquer formato por parte dos governos e outras ações que visem promover maior transparência*", Dados Abertos, que é um termo mais recente do que o anterior, visa a "*disponibilização de informações, em alguns formatos (pré-estabelecidos), por governos, organizações privadas, com ou sem fins lucrativos, ou por outros atores de uma sociedade*".

Já para a OECD [28], Governo Aberto é "*uma cultura de governança que promove os princípios de transparência, integridade, responsabilidade e participação das partes interessadas em apoio à democracia e ao crescimento inclusivo*". Além de ser essencial para a disseminação desses princípios, o conceito de *Dados Abertos* converge aspectos legais e tecnológicos. Nesse sentido, apresenta-se na próxima Seção uma perspectiva quanto à evolução do padrão *Open Data*, ao encontro do movimento *Open Government Data*, considerando os respectivos marcos legais e tecnológicos.

⁵Disponível em: <https://dl.acm.org/doi/10.1145/3463677.3463715>.

⁶OGP é uma iniciativa internacional, criada em 2011, em que governo, sociedade civil, empresas e organizações sem fins lucrativos se comprometem a utilizar os recursos tecnológicos adequados para incentivar práticas de abertura de dados. No Brasil, a OGP é coordenada pela CGU.

2.2 Dados Abertos (*Open Data*)

O objetivo desta Seção é embasar e apresentar, por meio da Revisão Sistemática da Literatura (RSL), o princípio arquitetural relacionado aos conceitos do padrão *Open Data* e suas variações (*Linked Open Data* e *Open Government Data*), no sentido de sustentar a elaboração do modelo proposto à luz dos critérios de riscos identificados a partir dessa Revisão.

2.2.1 Dados, Recursos e Metadados

Para Date [45] e Ramez Elmasri [46], **dados** são "*fatos conhecidos que podem ser registrados e possuem significado implícito*", referindo-se a um banco de dados como "*uma coleção ou um conjunto de dados inter-relacionados*".

Os conjuntos de dados, conforme ilustrado pela Figura 2.2, são publicados em diversas distribuições, para acesso ou *download*, que contém um ou mais **recursos** representados por arquivos como JSON, XLS, CSV, ODT, CSV, HTML, API, etc. Em geral, e no bojo da Ciência da Computação, recurso é usado para qualquer coisa que possa ser identificada na *Web* de Dados por meio de um Uniform Resource Identifier (URI)⁷, entre outras tecnologias.

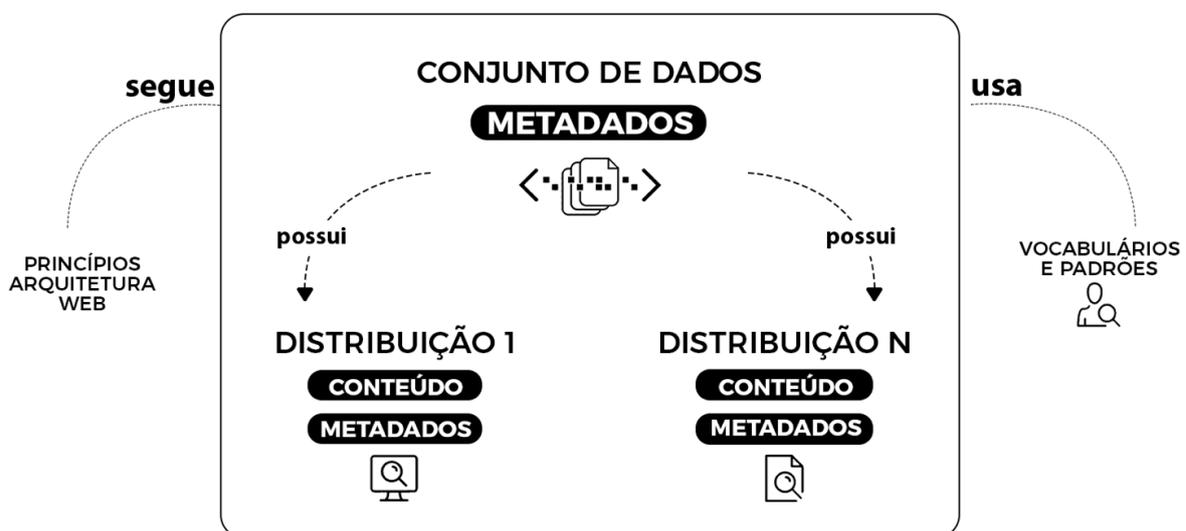


Figura 2.2: Abstração da Publicação de Dados na Web.

Fonte: Adaptado de Lóscio, Guimarães e Calegari [47].

⁷Uniform Resource Identifier (URI) permite identificar recursos físicos e/ou eletrônicos/digitais na *Web*, seja um livro, uma imagem, documentos, serviços ou mesmo outros conjuntos de dados, por exemplo.

Diante da heterogeneidade dessas distribuições, e do fato de que os produtores e consumidores de dados podem ser desconhecidos uns dos outros, é imprescindível fornecer algumas informações adicionais para a descoberta, exibição e interoperabilidade de recursos reconhecidos por URIs, a fim de que os dados publicados se tornem efetivamente reutilizáveis e acessíveis por todos. Dessa forma, se faz necessária a utilização de um modelo de dados que forneça uma estrutura "vinculável" (Resource Description Framework), permitindo, além de interoperabilidade, conexão semântica.

O uso de **metadados**, comumente conhecidos como "*dados sobre dados*", serve para descrever, padronizar e contextualizar o esquema e a estrutura de um determinado conjunto de dados, como por exemplo: índice, chaves, títulos, descrição, autor, geolocalização, método de acesso, termos/palavras-chave, periodicidade, data da modificação/revisões, fontes de dados, responsáveis, etc. Essas distribuições permitem que conjuntos de dados sejam usados e compartilhados em larga escala, independentemente da finalidade das partes interessadas [47].

Segundo Riley [48], metadados são informações que descrevem algum aspecto de um recurso ou conjunto de dados, classificando-se em três principais tipos: descritivos, administrativos e estruturais.

Informações que explicam, localizam ou tornam mais fácil a consulta ou recuperação de um recurso podem ser classificadas como *metadados descritivos*.

Na esfera das informações que servem para gerenciar um conjunto de dados ou que se relacionam com a sua criação ou licença (quando e como o dado foi gerado ou quem deve ter acesso, por exemplo) são classificados como *metadados administrativos*.

Metadados estruturais, por outro lado, descrevem os relacionamentos entre partes de recursos que possuem interesse em comum, permitindo localizar recursos que podem ser combinados entre si.

A publicação é condição *sine qua non* à abertura de dados. Assim, importa destacar a diferença entre *público* e *aberto* no bojo desta temática. Dados públicos, em geral, são dados fornecidos e custodiados pelo governo, quando disponíveis para acesso (gratuito), mas não necessariamente compreende os aspectos técnicos e legais vinculados ao conceito *Open Data* [26].

Listam-se as seguintes definições estabelecidas pela Política de Dados Abertos do Poder Executivo federal [6]:

- i) dado: sequência de símbolos ou valores, representados em qualquer meio, produzidos como resultado de um processo natural ou artificial;
- ii) dado acessível ao público: qualquer dado gerado ou acumulado pelo Governo que não esteja sob sigilo ou sob restrição de acesso nos termos da Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação - LAI. [43]);

- iii) dados abertos: dados acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento (*mashup*), limitando-se a creditar a autoria ou a fonte; e
- iv) formato aberto: formato de arquivo não proprietário, cuja especificação esteja documentada publicamente e seja de livre conhecimento e implementação, livre de patentes ou qualquer outra restrição legal quanto à sua utilização.

O conceito de **dados abertos** compreende não apenas disponibilizá-los para a livre utilização, mas torná-los aptos para reuso e redistribuição *online*, de modo que possibilitem a identificação e leitura por ferramentas informatizadas, sujeitando-se, no máximo, à exigência de atribuição da autoria à fonte que criou a base, para o compartilhamento dessas informações sob as mesmas condições ou licenças originalmente aplicadas [15, 1, 20, 49]. A reutilização desses dados depende do seu nível de vinculação a outros dados, evoluindo, assim, para o conceito de Linked Open Data (2.4).

Neste trabalho, o termo *dados abertos* será utilizado para se referir aos dados que foram criados, modificados e/ou coletados por órgãos públicos e/ou privados, para um propósito específico e que, quando disponíveis ao público, podem ser utilizados para qualquer finalidade.

No eixo governamental, Dados Abertos se relacionam com o princípio da Transparência Ativa, conforme ilustrado na Figura 2.3. Com efeito, destaca-se que há dois tipos de Transparência pública⁸, obrigatórios a qualquer organização:

- Transparência Ativa: o ente público disponibiliza dados/informações de forma proativa, ou seja, por iniciativa própria; e
- Transparência Passiva: o ente público disponibiliza dados/informações apenas quando provocado, geralmente, por meio de requisições/pedidos de cidadãos.

⁸Para maior detalhamento, vide os Capítulos II e III do Decreto nº 7.724, de 16 de maio de 2012.



Figura 2.3: Dados Abertos no contexto governamental e princípios relacionados.
Fonte: Elaboração Própria.

De acordo com a Fundação do Conhecimento Aberto (*Open Knowledge Foundation*⁹ - OKF), para que um conjunto de dados possa ser considerado *aberto*, deve possuir, no mínimo, os seguintes princípios:

- Disponibilidade e acesso em qualquer formato (PDF, PDF/A, CSV, JSON, XLS, XML, HTML, entre outros): os dados devem ser fáceis de encontrar, usar e combinar com outros dados e sob custo não maior que um custo razoável para a sua reprodução, preferencialmente disponíveis para *download* e possíveis de ser modificados;
- Reutilização e redistribuição: os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição, inclusive a combinação com outros conjuntos de dados; e
- Participação universal: livre acesso para usar, reutilizar e redistribuir - não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos. Por exemplo, restrições de uso ‘não-comercial’ que impediriam o uso ‘comercial’, ou restrições de uso para certos fins (ex.: somente educativos) excluem determinados dados do conceito de ‘abertos’.

⁹<https://opendefinition.org/>

2.3 Dados Conectados (*Linked Data*)

O crescimento exponencial da *big data*¹⁰ e a quantidade de dados estruturados disponíveis na *internet* promoveram o termo **Dados Conectados (Linked Data)**, o qual representa um conjunto de boas práticas e tecnologias direcionadas para facilitar a publicação, interconexão e compartilhamento de dados estruturados na *internet*, com base nos padrões da Web Semântica [51, 52]. Ou seja, enquanto os "dados sobre a *Web*" utilizam hipertexto semiestruturado (HTML) para criar conexões via *hiperlinks*, os Dados Conectados utilizam tecnologias que possibilitam interligação semântica (*Links RDF*), interconectando recursos a fontes de Dados Conectados.

Dados Conectados, conforme Berners-Lee [20] e Ubaldi [49], descrevem métodos de boas práticas de publicação de dados estruturados para que possam ser interconectados. O objetivo desses métodos é tornar os dados mais interoperáveis e úteis, com base em tecnologias que priorizam primeiro a legibilidade por máquinas e depois por humanos. Isso permite que dados de diferentes fontes sejam conectados, consultados e compartilhados de modo mais dinâmico e automatizado.

Dentre esse conjunto de boas práticas e suas tecnologias semânticas (por exemplo, RDF, RDFS, OWL¹¹ e SPARQL), inclui-se a adoção dos seguintes **princípios *Linked Data*** [20]:

- i) Utilizar URIs (Identificador Uniforme de Recurso) para nomear qualquer coisa:
 - Este princípio defende o uso de URIs para identificar e distinguir diferentes recursos publicados na *web* e não apenas documentos e conteúdo digital, mas também qualquer objeto do mundo real e/ou conceitos abstratos. Esta boa prática (URIs distintos para recursos distintos) busca evitar que um único recurso seja nomeado com identificadores diferentes, ou que recursos diferentes sejam identificados com o mesmo rótulo (*tag*) [50, 51].
- ii) Utilizar HTTP URIs para que esses nomes possam ser consultados:
 - Este princípio se importa em utilizar um modelo único de dados na Web, de modo a facilitar a sua escalabilidade, seguindo os mesmos princípios arquiteturais da Web de Documentos. URIs exploram fortemente métodos definidos pelo protocolo HTTP. URIs HTTP além de identificar recursos, também são um

¹⁰Termo utilizado para designar um enorme volume de dados. No entanto, trata-se de um conceito abstrato que compreende, além da quantidade de dados, aspectos como velocidade, variedade, veracidade, valor e complexidade [50].

¹¹RDFS e OWL fornecem vocabulários para descrever modelos conceituais em termos de classes e suas propriedades

meio de acessá-los por nomes exclusivos (identificadores únicos), globalmente acordados (reconhecidos) e com descrições (declarações) em formato comum à Web (RDF e HTML) [51].

iii) Pesquisar determinada URI fornecendo informações úteis usando os padrões RDF e SPARQL:

- A linguagem SPARQL é utilizada para realizar consultas a dados no padrão RDF, facilitando a interoperabilidade entre dados conectados, além da rápida localização de dados/informações na Web de Dados no contexto do nome identificado (significado). O uso de RDF é o ponto de partida para que o formato original dos dados não seja alterado, a fim de garantir o reuso e o *mashup* com outros dados. O enfoque do RDF é a leitura por máquinas e não por humanos (como é caso do HTML e do XML). Da mesma forma que as URIs são usadas para identificar os recursos, a linguagem SPARQL é utilizada para descrever e relacionar recursos pelo seu "significado", agregando valor aos resultados retornados nas consultas [50].

iv) Incluir *links* para outros URIs, permitindo a descoberta de mais recursos/*coisas*:

- Este princípio orienta o uso de *links* RDF para ligar/conectar dados a outros dados antes de publicá-los, ou seja, uma parte do dado/informação pode ser interconectada a outros recursos disponíveis na nuvem LOD [51].

2.4 Dados Abertos Conectados (*Linked Open Data - LOD*)

Segundo Berners-Lee [20] e Hyland et al. [53], **Dados Abertos Conectados** refletem a *Web Semântica* por meio da descoberta, interligação, armazenamento e consumo de conjuntos de dados disponíveis sob licenças abertas, convertendo-os em RDF de acordo com os princípios *Linked Data* supracitados.

No intuito de estimular a publicação de Dados Abertos Conectados, "*especialmente os proprietários de dados do governo*", Berners-Lee [20] propôs um sistema, ora denominado "Sistema de 5 estrelas"¹², para classificar os dados publicados no padrão aberto. Tal Sistema, ancorado pelo projeto *Linked Open Data (LOD)*, criado em 2007 pelo *World Wide Web Consortium* *World Wide Web Consortium (W3C)*, impulsionou a evolução

¹²<https://5stardata.info/pt-BR/>

da atual *Web orientada a documentos* para uma *Web de Dados Conectados* ou *Web Semântica* [51, 52].

Essa transformação, da *Web* tradicional (orientada a documentos) para *Web Semântica*, assentando a interligação e intercâmbio entre nuvens de dados heterogêneos, se apoia em APIs comuns à nuvem LOD¹³, destacando-se o estilo arquitetural *Representational State Transfer* (REST), implementado por serviços conhecidos como RESTful, e modelado por URIs que permitem operar em redes de dados que realizam mudanças constantes de seus valores.

O objetivo do projeto LOD é identificar diversos conjuntos de dados disponíveis sob licenças abertas (Licenças *Open Data*), convertendo-os em RDF conforme os princípios *Linked Data*, permitindo, assim, realizar a conexão automática de Dados Abertos, gerando um contexto informacional a partir da representação semântica dos dados interligados. Dessa forma, possibilita-se identificar e processar de forma automatizada, não apenas páginas sintaticamente conectadas e publicadas na *internet* mas, sobretudo, dados abertos de diferentes contextos e níveis de granularidade, conectados semanticamente [51].

Os princípios Linked Open Data, representados na Figura 2.4, são descritos a seguir:

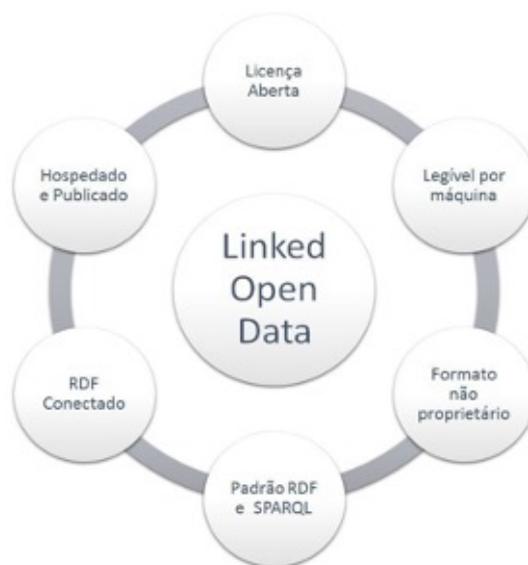


Figura 2.4: Princípios do padrão *Linked Open Data*.
Fonte: Adaptado de Berners-Lee [20].

P1) Licença Aberta: a licença é o principal metadado para que o recurso disponível seja considerado como dado aberto. O recurso deve ser de domínio público¹⁴ ou fornecido sob uma licença aberta, que se refere às condições legais, restrições (razoáveis),

¹³Um diagrama de nuvem LOD pode ser visualizado pelo seguinte endereço: <http://lod-cloud.net>. Esse diagrama é gerado a partir dos metadados disponíveis em <https://ckan.org/>.

¹⁴Segundo o art. 41 da Lei nº 9.610, de 19 de fevereiro de 1998 (Lei do Direito Autoral) os direitos patrimoniais terminam após 70 anos do falecimento do autor.

conformidades, cláusulas e princípios relacionados ao conceito *open data*, sob os quais os dados estão sujeitos. Trata-se da concessão de dados para amplo acesso e reutilização e/ou redistribuição, exigindo-se, no máximo, creditar a autoria à fonte original e o compartilhamento pela mesma licença. Entre as mais conhecidas citam-se: *Creative Commons* e *Open Data Commons*¹⁵.

Sendo o primeiro princípio a ser abordado por Berners-Lee [20], a Licença Aberta confirma a sua relevância quando Máchová e Lněnička [54] a relaciona diretamente aos domínios da aplicação, atribuição, interligação, armazenamento e compartilhamento dos dados, tendo em vista que os dados podem representar riscos em relação à confiabilidade, qualidade e formato [11].

Dessa forma, a licença é a forma que o proprietário ou gestor de dados possui para autorizar o livre acesso, re(uso) e redistribuição desses dados para qualquer finalidade (mormente para fins comerciais, devendo respeitar a proteção de dados pessoais). Além disso, respalda a organização contra os eventuais riscos supramencionados, detalhados por Martin [11].

- P2) Legível por máquina: conversão da base de dados em um conjunto de instruções processáveis por máquina (por exemplo: CSV, XML, JSON, GeoJSON, XLS, entre outros)¹⁶;
- P3) Formato não proprietário: formato livre de utilização sem qualquer restrição ou patentes, por exemplo: uso de arquivos ODS ou CSV (não proprietário) ao invés XLS (proprietário). Cada recurso se refere a uma forma específica disponível a partir de determinado conjunto de dados, podendo estar em formato proprietário ou aberto (CSV, HTML, ODS, JSON, XML, RDF, entre outros).
- P4) Padrão RDF e Consulta SPARQL: permite a realização de consultas, via linguagem SPARQL, às bases de dados RDF, que é um *framework* para representar informações na *web* [55]. Mecanismos de busca de dados conectados permitem pesquisar na rede de dados. A busca convencional pode apresentar informações derivadas de dados conectados.
- P5) Resource Description Framework (RDF) conectado (bases RDF conforme os princípios *Linked Data* ou JSON-LD, que permite o uso de ontologias¹⁷ para a descrição e descobertas de dados): Resource Description Framework (RDF) é uma estrutura ou

¹⁵A OKF apresenta uma lista de licenças em: <http://opendefinition.org/licenses/>.

¹⁶Uma lista mais completa de formatos legíveis por máquina está disponível em: https://github.com/ingridpalma/mestrado/blob/main/machine_readable

¹⁷Ontologias são vocabulários que registram as relações lógicas entre seus termos

modelo de dados, recomendado pela W3C¹⁸, utilizado para representar e serializar metadados de qualquer informação e/ou recursos referenciados na *internet*.

P6) Hospedado e publicado: refere-se aos aspectos técnicos recomendados para catalogação, hospedagem e publicação de dados na *Web*. Para Bauer e Kaltenböck [56], os dados conectados podem ser publicados em arquivos simples em um servidor web, em bancos de dados ou em “armazenamentos triplos” especializados, construídos para armazenar e compartilhar dados conectados.

Ainda, no que concerne ao conceito de Dados Abertos Conectados¹⁹ e às tecnologias relacionadas à Web Semântica, Triperina [33] aduz que o principal objetivo dos princípios LOD é a identificação e a conectividade semântica entre dados, viabilizando a descoberta e o consumo dos dados abertos de maneira rápida e automatizada, por meio de URIs representadas pela estrutura de linguagem RDF.

Para Hitzler, Krotzsch e Rudolph [50], o principal objetivo do RDF é representar dados estruturados, preservando o seu significado original, a fim de priorizar a leitura e processamento automatizados. Registra-se que, para consultas de recursos (dados) em RDF, utiliza-se o protocolo SPARQL, permitindo conectar fontes de dados entre si para criar uma base de conhecimento combinada que aumenta a aplicabilidade e a utilidade dos dados consumidos.

Bauer e Kaltenböck [56] reúnem os elementos comuns e essenciais dos Dados Abertos Conectados (LOD), destacados na Figura 2.5, os quais foram compilados em um manual de boas práticas que esquematiza e descreve os principais serviços arquiteturais da *Web Semântica* para apoiar a publicação e consumo de LOD.

¹⁸<https://www.w3.org/TR/rdf-primer/>.

¹⁹Uma lista de definições do ecossistema de Dados Abertos pode ser consultada no seguinte endereço eletrônico: <https://www.w3.org/TR/ld-glossary/#vocabulary>.

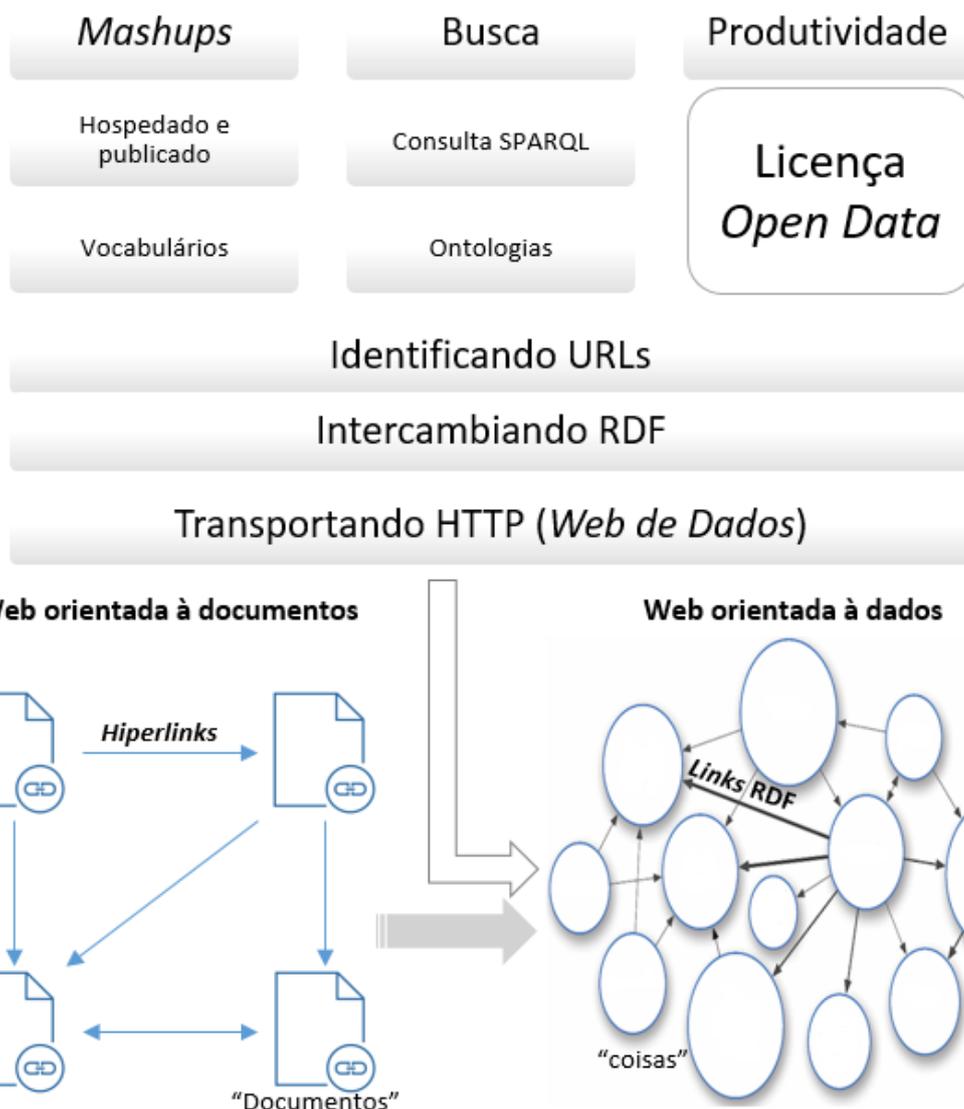


Figura 2.5: Princípios Arquiteturais *Linked Open Data*.1 Evolução da Web.
Fonte: Adaptado de Bauer e Kaltenböck [56].

Os *mashups* cruzam dados de diferentes origens, produzindo análise e resultados comparativos a partir da combinação desses recursos, no intuito de criar um serviço, uma visualização ou simplesmente uma informação útil.

Diferente dos *hiperlinks* que são utilizados na Web de Documentos, *Links RDF* são usados para conectar qualquer coisa no contexto da *Web de Dados* do setor elétrico, por exemplo, um transformador se conectaria aos seus medidores e esse sistema se ligaria aos conjunto de dados coletados por esses medidores.

Dados conectados facilitam a integração de dados para inteligência de negócios ou pesquisa, fomentando a *produtividade* em diferentes contextos.

Os *vocabulários* fornecem listas e definições de termos comuns que podem ser usados para descrever *coisas* e relacionamentos em um conjunto de dados.

Identificando URLs significa o uso de HTTP URL para que dados pesquisados na *internet*, via “*namespaces*”, sejam gerenciados por meio de DNS (*Domain Name System* – Sistema de nome de domínio).

Os dados são hospedados em servidores que podem comunicar o protocolo de transferência de hipertexto (HTTP) entre si (*Web de Dados*) e com navegadores para trocar dados pela *internet*.

2.5 Dados Abertos Governamentais (Conectados)

No trabalho apresentado por Haini et al. [57], os autores identificaram 16 fatores de influência à decisão de adoção da Open Government Data (OGD) pelas organizações públicas. Tais fatores foram categorizados em três dimensões (Tecnologia, Organizacional e Ambiental).

A revisão da literatura permitiu organizar, por meio do Quadro 1, algumas definições apresentadas para Open Government Data.

Quadro 1: Revisão da Literatura acerca do conceito OGD.

Fonte: Elaboração Própria.

Definição para <i>Open Government Data</i>	Referencial Teórico
Refere-se à publicação e disseminação de uma grande quantidade de informações públicas em formato bruto, aberto e compreensíveis logicamente, permitindo sua reutilização e <i>mashups</i> de dados.	Vieira [58]
São dados do setor público disponíveis livremente em formatos abertos.	Kalampokis, Tambouris e Tarabanis [59]
É uma filosofia - e cada vez mais um conjunto de políticas - que promove a transparência, a responsabilidade e a criação de valor ao produzir, custodiar e disponibilizar um grande volume de dados e informações do governo para todos.	Ubaldi [49]
Conteúdo publicado na Web pública por autoridades governamentais em uma variedade de formatos não proprietários.	Hyland et al. [53]
Representa a disponibilização de dados por parte dos governos, nos mesmos formatos estabelecidos para dados abertos.	Albano [44]

Quadro 2: (continuação)

Definição para <i>Open Government Data</i>	Referencial Teórico
São os dados gerados nos processos governamentais disponibilizados para a sociedade.	Barbalho [60]
Dados abertos produzidos ou custodiados pelo governo ou outros operadores em nome do governo, disponíveis livremente para uso e redistribuição.	Yannoukakou e Araka [61]
É uma filosofia de trabalho que visa capacitar os cidadãos e servidores, no sentido de prover conhecimento e habilidades para acessar os dados produzidos ou custodiados pela administração públicas, de modo que eles podem usar, armazenar em bases de dados próprias, redistribuir e integrá-los com outras fontes de dados.	Solar et al. [62]

Verifica-se a abrangência de definições que cercam o conceito de *Dados Abertos Governamentais*, o qual se relaciona, sobretudo, com os conceitos de Transparência, Prestação de Contas, Controle/Participação Social, Inovação, Política Pública, Tecnologia, entre outras ferramentas, normas, padrões e técnicas que fomentam o movimento OGD.

É consenso que Dados Governamentais são produzidos e/ou coletados por meio da prestação de bens ou serviços públicos. *Dados Abertos Governamentais Conectados* (em inglês, *Linked Open Government Data*), efetivamente disponibilizados seguindo os princípios LOD, representado na Figura 2.4, devem permitir a integração com outras bases, sem quaisquer restrições de direitos autorais [14, 53]. O padrão e a qualidade dessas bases facilitam a identificação e o consumo automatizado dos dados. Repisando que os órgãos da APF devem realizar a devida divulgação de seus dados nos seus portais institucionais, referenciando-os no PBDA [6].

A Figura 2.6 representa o relacionamento entre os conceitos (*Linked*) (*Open*) (*Government*) (*Data*) e sintetiza a justaposição dos termos Dados Abertos, Dados Conectados (e possivelmente "fechados") e Dados Governamentais. Dados Abertos Governamentais Conectados (LOGD) e Dados Abertos Conectados (LOD) estão estritamente relacionados às tecnologias da *Web Semântica*.

Existem várias dimensões que podem ser consideradas na avaliação dos *Dados Abertos Governamentais*, tais como a Transparência e a Prestação de Contas. Considerando essas dimensões, Murillo [64] formulou um Índice de Abertura de Dados Governamentais (*Government Data Openness Index* - GDOI), representado pela seguinte equação:

$$GDOI = \frac{1}{m} \sum_{T=1}^n [\beta_r DOI_r] \quad (2.1)$$

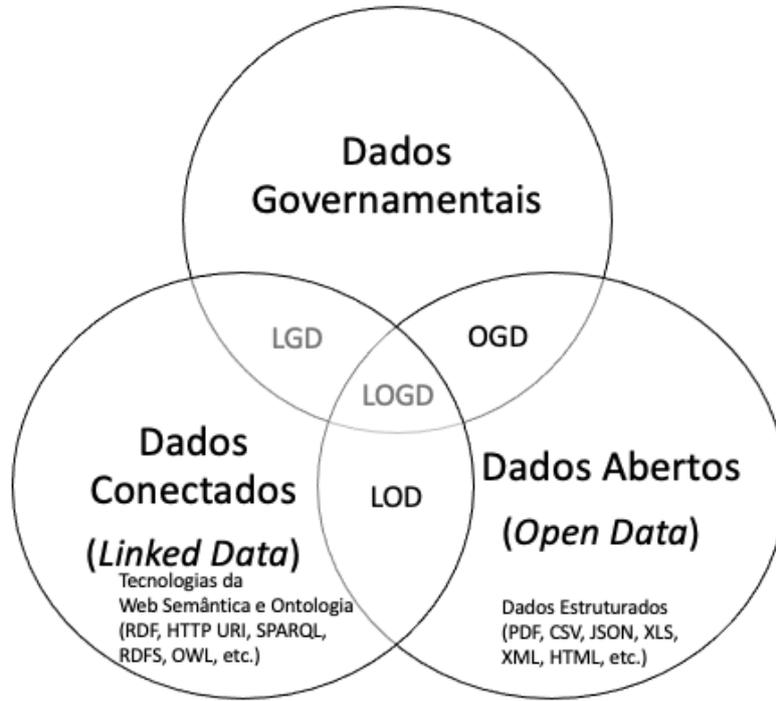


Figura 2.6: Justaposição dos conceitos (Linked) (Open) (Government) (Data).
Fonte: Adaptado de Hitz-Gamper, Neumann e Stürmer [63].

Dessa forma, β é o peso designado a determinado recurso ou tema (T), e m representa o total de recursos avaliados por esse modelo. Devido à subjetividade na escolha do peso β , este valor pode ser definido por especialistas ou normativos técnicos pertinentes ao modelo.

Para calcular o GDOI, deve-se fazer uma avaliação quantitativa dos princípios da OGD abordados por Tauberer e Lessig [65] e Ubaldi [49], listados a seguir, que medem o grau de abertura do conjunto de dados (T) frente às dimensões avaliadas. Esse cálculo é realizado pela seguinte equação:

$$DOI_T = \frac{1}{p} [\alpha_1 S_{T,1} + \alpha_2 S_{T,2} + \dots + \alpha_n S_{T,p}] \quad (2.2)$$

Conforme demonstrado na equação 2.3, α_v é o inverso da pontuação máxima atribuída a $(S_{a,v})_{max}$, assumida pela variável v .

$$\alpha_v = \frac{1}{(S_{a,v})_{max}} \quad (2.3)$$

A pontuação média de cada tema ou recurso (T) é valorada entre 0 e 1 e define o respectivo grau de abertura de dados. N representa o número dos princípios OGD contemplados pelo conjunto de dados e assume o valor máximo $N = 8$, caso contemple todos os princípios OGD.

Os oito princípios da *Open Government Data*, descritos a seguir, foram definidos durante um encontro²⁰, realizado em 2007, na Califórnia (EUA), que reuniu cientistas, pesquisadores, entre outros representantes e estudiosos do movimento OGD. Tais princípios orientam o processo de abertura de dados, considerando que dados governamentais abertos devem ser:

1. Completos: todos os dados públicos estão disponíveis. Dado público é o dado que não está sujeito a limitações válidas de privacidade, segurança ou controle de acesso;
2. Primários: os dados devem ser apresentados tais como os coletados na fonte ou, no mínimo, com o menor nível de granularidade possível, ou seja, sem agregações e/ou modificações;
3. Atuais: os dados são disponibilizados tão rapidamente quanto necessário à preservação do seu valor;
4. Acessíveis: os dados são disponibilizados para o maior alcance possível de usuários e para o maior conjunto possível de finalidades;
5. Compreensíveis por máquinas: os dados são razoavelmente estruturados de modo a possibilitar processamento automatizado;
6. Não discriminatório: os dados são disponíveis para todos, sem exigência de requerimento ou cadastro;
7. Não proprietários: os dados são disponíveis em formato sobre o qual nenhuma entidade detenha controle exclusivo; e
8. Licença Livre: os dados não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial. Restrições sensatas relacionadas à privacidade, segurança e privilégios de acesso são permitidas.

²⁰https://public.resource.org/open_government_meeting.html.

2.6 Modelo CRISP-DM

Conforme apresentado na Seção 1.5.2 (Plano Metodológico), as etapas do processo MCDA (vide Figura 2.9) se adequam aos desenvolvimentos do Modelo Cross Industry Standard for Data Mining (CRISP-DM) [36], denotando estreita relação entre ambos. Neste trabalho, os resultados do processo de mineração de dados são utilizados na aplicação da abordagem MCDA.

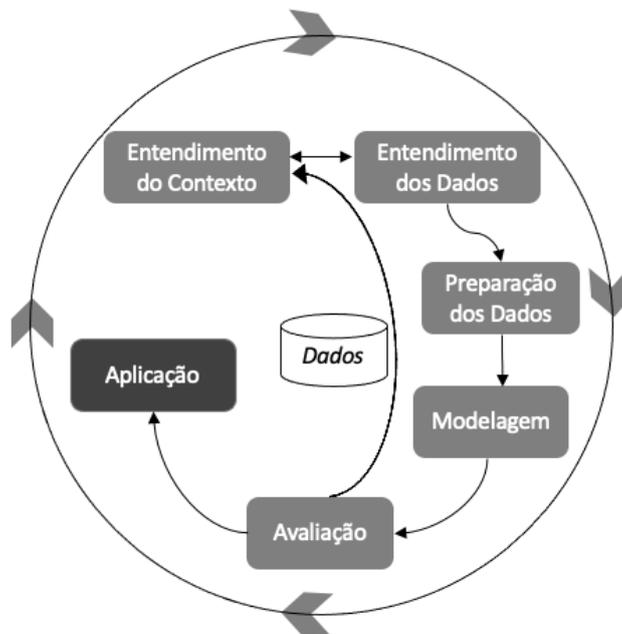


Figura 2.7: Estágios do Modelo CRISP-DM.

Fonte: Adaptada de Chapman et al. [36].

O Modelo CRISP-DM é a referência mais utilizada para o desenvolvimento de projetos relacionados a problemas de *text* e *data mining* [29]. Além disso, e para uma melhor compreensão deste projeto, objetivou-se a identificação dos potenciais temas para a abertura de dados do setor elétrico, sistematizada pelos estágios iniciais deste Modelo.

O Modelo iterativo e incremental CRISP-DM, representado pela Figura 2.7, fornece uma visão geral do ciclo de vida de um projeto de mineração, constituindo-se dos seis estágios descritos a seguir:

1. Entendimento do Contexto/Negócio: este estágio coincide com a Fase do Planejamento/Ruptura, supramencionada na Subseção 1.5.2 e envolve, inicialmente, entender o domínio do negócio, os objetivos do projeto e seus principais critérios/requisitos para, na sequência, definir o problema.
2. Entendimento dos Dados: o objetivo desse estágio é obter uma compreensão inicial dos dados. Para tal, iniciou-se a coleta, a descrição, a exploração e a verificação da

qualidade dos dados, permitindo a descoberta das primeiras percepções do modelo e uma melhor compreensão acerca da amostra analisada. Para otimizar a coleta dos inúmeros dados e metadados suportados pelo modelo, utilizou-se, conforme mencionado, técnicas de mineração (*web scraping*), o que facilitou a coleta não apenas de dados estruturados, mas também de não-estruturados e semi-estruturados.

3. Pré-Processamento (Preparação) dos Dados: este estágio se concentra na preparação dos dados para a construção do banco de dados final que irá alimentar, possivelmente em vários momentos, as ferramentas de modelagem selecionadas para a limpeza, integração, transformação e redução dos dados.
4. Modelagem: neste quarto estágio, várias técnicas de modelagem de mineração de dados podem ser aplicadas para o alcance de um determinado objetivo. No caso, foram seguidas técnicas de agrupamento e classificação para a identificação dos temas mais relevantes para a abertura de dados governamentais do setor elétrico.
5. Avaliação dos resultados com base nos objetivos do Modelo Decisório: neste estágio do projeto, avalia-se o resultado da modelagem construída a partir de técnicas conhecidas do CRISP-DM, considerando, ainda, o ponto de vista de especialistas e os objetivos definidos durante a fase do Planejamento da pesquisa.
6. Aplicação do Modelo: é sabido que o projeto não termina com a implementação e implantação do modelo. No caso, os temas encontrados serão úteis para a construção do Apoio à Decisão constituído a partir das etapas estabelecidas por Belton e Stewart [37] e Marttunen et al. [66], especificamente para a melhor compreensão das bases de dados separadas por *grupos temáticos*, etiquetas e/ou *tags*, as quais podem ser adotadas nos respectivos recursos catalogados no PBDA.

O modelo CRISP-DM fornece uma visão geral do ciclo de vida de um projeto de mineração de dados. Ele contém os estágios correspondentes de um projeto, suas respectivas tarefas e os relacionamentos entre essas tarefas [36, 29]. Cada estágio do CRISP-DM implementa diversas tarefas, genéricas e específicas, compreendendo, respectivamente, o segundo e terceiro níveis de um modelo, conforme exemplificado na Figura 2.8, que detalha o Estágio de Entendimento do Negócio/Contexto desta pesquisa. As tarefas executadas durante esse processo resultam em ações e resultados, que são as instâncias desse processo.

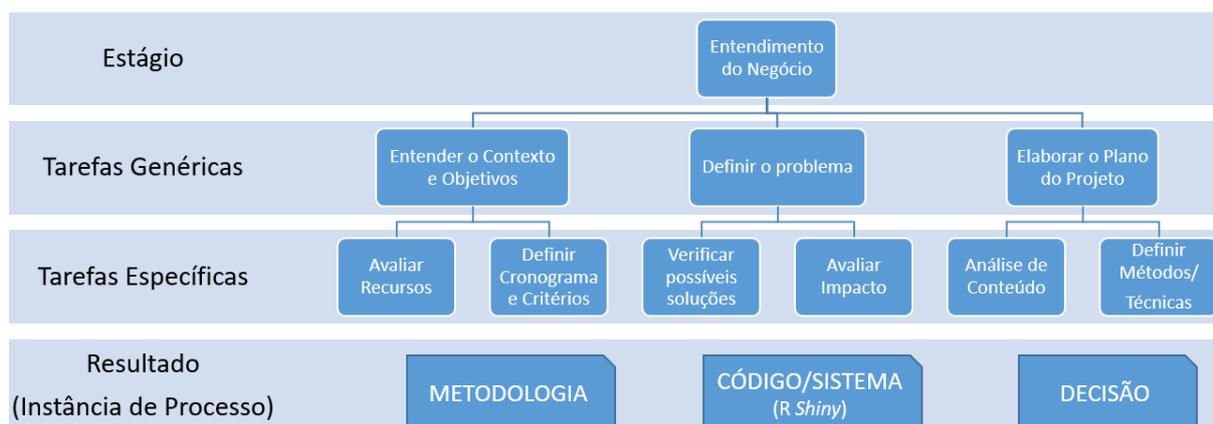


Figura 2.8: Níveis de detalhamento do Estágio de Entendimento do Contexto.

Fonte: Adaptada de Chapman et al. [36].

2.7 Modelo de Apoio à Decisão

Antes de abordar acerca do *Modelo de Apoio à Decisão* proposto, deve-se ter entendimento claro sobre a definição de *modelo*. Em primeiro, vale ressaltar que não há nada de novo no uso de *modelos* para auxiliar a tomada de decisão.

A *modelagem* é utilizada em diferentes contextos (militar, computacional, matemático, administração, sócio-econômico, psicologia), sobretudo, no contexto da Pesquisa Operacional (PO)²¹, a qual engloba um conjunto de métodos multicritério e técnicas de otimização matemática, ciência de dados, aprendizado de máquina, entre outros métodos direcionados a problemas complexos, a partir dos quais se escolhe a decisão mais adequada dada as características do modelo (natureza, objetivos, critérios, restrições, incertezas, alternativas, partes interessadas, entre outros aspectos) [67].

Nos trabalhos intitulados "*Who is a Modeler?*", "*Models in Science*" e "*Design and Natural Science Research on Information Technology*", Weisberg [68], Frigg e Hartmann [69] e March e Smith [70] apresentam diferentes aspectos e definições de *modelo* (tipicamente matemáticos e/ou computacionais) reconhecidos na literatura. No entanto, para fins deste estudo, entende-se **modelo** como "*uma representação simplificada de um problema real, devendo ser suficientemente detalhada para captar os elementos essenciais do problema, mas suficientemente tratável por métodos de resolução*" [71].

Para March e Smith [70], *modelos* são avaliados conforme a sua completude, robustez, consistência e fidelidade com os fenômenos do mundo real. Dessa forma, o *modelo* se torna um meio de representar e estudar um fenômeno do mundo real, seja este fenômeno um problema, um processo, uma operação ou um evento qualquer. A solução do modelo

²¹<https://www.ifors.org/what-is-or/>.

deve **apoiar** o processo de tomada de decisão [71, 72], daí o termo *Modelo de Apoio à Decisão*.

Para Forrester [73, p. 3],

todas as nossas decisões são tomadas com base em modelos. A questão não é usar ou ignorar os modelos. A questão é realizar uma escolha entre diferentes modelos alternativos.

A teoria do processo decisório foi marcada pelos clássicos estudos de Herbet Simon [74] que, após observar o comportamento de decisores em diferentes organizações, apresentou o conceito da *racionalidade limitada*, afirmando que a tomada de decisão era comprometida pela busca da solução mais próxima e "satisfatória", afastando-se de resultados otimizados por meio da análise comparativa entre todas as alternativas possíveis. O conceito da *racionalidade limitada* sobre a tomada de decisão alicerçou diversos modelos de apoio à decisão [75, 76, 77].

O trabalho *Decision-Aid and Decision-Making* [78] apresenta as principais diferenças entre a *tomada de decisão* e o *apoio à tomada de decisão*. No referido trabalho, *tomada de decisão* deve definir um conjunto de alternativas (soluções) viáveis para a estruturação de um modelo ajustado às preferências dos decisores e à problemática formulada. Já o **apoio à decisão** corresponde a uma atividade que se apoia em modelos formais e específicos para aumentar o grau de conformidade e coerência da tomada de decisão, oferecendo respostas às questões levantadas dentro do processo decisório.

Não obstante, Bouyssou et al. [76] afirmam que o *auxílio à decisão* não se resume apenas a um modelo formal, mas a um processo complexo que envolve, além desse modelo: i) a formulação do problema decisório; ii) a validação do modelo e iii) um plano de recomendações.

Na linha de Bouyssou et al. [76], a concepção do *Modelo de Apoio à Decisão* para a resolução de um problema, seja para avaliar dados abertos governamentais de modo geral ou do setor elétrico, compreende as fases de Identificação e Estruturação da Problemática, Construção e Validação do Modelo e, por fim, a Implementação do Plano de Ação [37, 79]. Alguns exemplos de modelos, no âmbito da Pesquisa Operacional, são: rede Program Evaluation and Review Technique (PERT); Modelo de Markov, Modelo de Decisão Sequencial, Programação Dinâmica, entre outros.

Por abstrair e representar, de forma simplificada e mais completa possível, aspectos de um problema e/ou fenômeno do mundo real, *Modelos de Apoio à Decisão* utilizam métodos multicritério que auxiliam a análise de riscos e incertezas [79]. Além disso, esses modelos padronizam a descrição e a linguagem para a discussão e o melhor entendimento de um problema identificado [80].

O Modelo de Apoio à Decisão, nesta pesquisa, é uma abordagem que requer uma coleta de dados relevantes para a sua construção e validação. Para isso, deve-se definir critérios que, segundo Roy [81] e Bouyssou [82], permitem realizar comparações entre as alternativas identificadas com base na preferência dos decisores.

2.7.1 Critério de Risco

Para a ABNT ISO GUIA 73:2009 (Guia 73), o Critério de Risco (CR) serve como um termo de referência (padrão, medida ou expectativa) utilizado para fazer um julgamento ou uma decisão sobre a importância de um risco [83]. Tais critérios podem abranger variáveis como custos, benefícios, requisitos legais, preferências ou prioridades das partes interessadas, aspectos socioeconômicos, ambientais, entre outros [83, 84].

2.8 Modelo de Processo de Apoio à Decisão MCDA

Processo de Apoio à Decisão corresponde à construção de um sistema de preferências em que as alternativas são avaliadas em relação aos critérios selecionados para o modelo em desenvolvimento, a fim de que as possíveis soluções sejam recomendadas e validadas ante a consecução da etapa seguinte, conforme se apresenta na Figura 2.9.

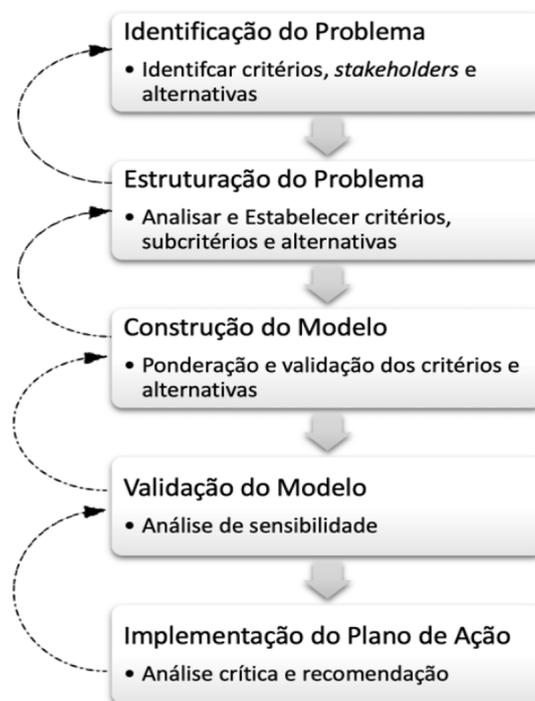


Figura 2.9: Etapas do Processo de Apoio à Decisão.
Fonte: Adaptada de Belton e Stewart [37] e Marttunen et al. [66].

Multicriteria Decision Analysis (MCDA) (ou seja, Análise Multicritério de Apoio à Decisão) fornece um *framework* para a estruturação e resolução de inúmeros problemas de decisão, explorando critérios relevantes a um contexto específico, os quais são ponderados e avaliados por meio de um sistema de preferências, que utilizam métodos e técnicas reconhecidos pelas abordagens MCDA e/ou Tomada de Decisão por Multicritérios (Multiple Criteria Decision Making - MCDM)²² [37].

Principais Elementos do Modelo MCDA

Conforme Bana e Costa e Vincke [85], Roy [81] e Garcia [86], os principais elementos de um modelo multicritério de apoio à decisão são:

- Alternativas: representam um conjunto finito ($A_\alpha = A_1, A_2, A_3, \dots, A_n$) de ações potenciais aplicáveis a determinado problema, a partir das quais o tomador de decisão deve decidir; e
- Critérios: representa um conjunto finito de critérios ($C_\alpha = C_1, C_2, \dots, C_n$), construídos com base em indicadores e preferências das partes interessadas, a fim de realizar comparações entre as alternativas aplicáveis ao modelo. Para isso, cada critério é utilizado para mensurar o desempenho de determinada alternativa ($C_{j(A_i)}$).

Hierarquias mais complexas podem ser construídas adicionando **subcritérios** à estruturação do modelo, permitindo, assim, mais detalhamento dos critérios.

Segundo Roy [81] e Bouyssou [82], *critérios de riscos* permitem realizar comparações entre as alternativas existentes a partir da preferência dos decisores. Nessa via, *critério de risco* não avalia a assertividade de determinado elemento ou variável, mas serve como parâmetro para auxiliar as etapas do processo de Gerenciamento de Riscos (análise e avaliação), as quais irão determinar a significância de um risco (se o risco residual é tolerável frente aos objetivos pertinentes), subsidiando a tomada de decisão.

2.8.1 Problema e Problemática de Decisão

Cabe destacar a diferença conceitual entre *Problema de Decisão* e *Problemática de Decisão*. A compreensão do *problema de decisão* busca, segundo Marttunen et al. [66], "*harmonia entre o caos*". Ou seja, o objetivo é organizar os elementos essenciais a todo problema de decisão, que são o conjunto de alternativas, critérios e a estrutura de preferências, no intuito de organizar a derivação desses critérios²³, os quais permitirão aos *stakeholders* ponderar preferências entre as potenciais alternativas.

²²Considera-se, neste trabalho, sinônimas as abordagens MCDA e MCDM.

²³Neste aspecto, Costa [87] cita a condição *ceteris paribus*, ou seja, a derivação de um critério independe de mudanças no contexto ou alterações em algum elemento da matriz de preferência.

já as *problemáticas de decisão* refletem a forma como a avaliação das alternativas deve ser representada ao final do processo decisório. Os principais tipos de problemáticas de decisão são Escolha, Ordenação, Classificação e Descrição [81], a ver:

- Escolha: seleciona uma alternativa a partir de um conjunto predefinido de alternativas.
- Ordenação (*ranking*): fornece alternativas a partir de uma ordem crescente ou decrescente.
- Classificação (*clustering*): as alternativas são organizadas em grupos pré-definidos e homogêneos.
- Descrição: explica as alternativas em relação às suas especificidades e características.

Para cada problemática existe um resultado específico. Os problemas de ordenação e classificação se apoiam em julgamentos relativos para obter a matriz de decisão.

Diferentes abordagens MCDA podem ser construídas e aplicadas a inúmeros problemas que, em geral, envolvem uma situação específica e única ou mais ampla e repetitiva. A estruturação da problemática permite distinguir e selecionar os métodos decisórios apropriados e convenientes ao modelo em comento [85, 66].

2.9 Métodos de Apoio à Decisão sobre Dados Abertos

Esta Seção apresenta uma Revisão Sistemática da Literatura (RSL) acerca dos principais e mais recentes trabalhos relacionados a modelos decisórios para a abertura de dados, considerando, sobretudo, os critérios de riscos abordados.

Conforme se observa em Gutjahr e Nolz [88], ainda são poucos os trabalhos que tratam de modelos decisórios para a seleção e priorização de temas potenciais para a abertura de dados governamentais. Tal constatação evidencia a existência de uma lacuna teórica na utilização desses modelos em relação aos principais conjuntos de dados produzidos ou custodiados pelas entidades governamentais.

Em Gurstein [4] e Worthy [24], observa-se a preocupação com a ausência de medidas específicas quanto à abertura e catalogação de dados públicos, apresentando modelos estritamente qualitativos quanto à definição e avaliação de potenciais critérios.

A revisão sistemática da literatura realizada por Hossain, Dwivedi e Rana [89] demonstrou que apenas 11% das produções científicas envolvendo técnicas de avaliação de *dados abertos* aplicaram métodos quantitativos, representando uma grande diferença entre o número de produções científicas qualitativas (47%) ou meramente interpretativas (38%) em relação às abordagens menos subjetivas.

Nada obstante, a partir da pesquisa bibliográfica realizada, identificou-se os principais trabalhos que abordam métodos de apoio à tomada de decisão frente à complexidade de avaliar dados abertos governamentais. Tais métodos e o quantitativo de critérios utilizados nos trabalhos analisados estão relacionados no Quadro 2.

Quadro 2: Métodos de Apoio à Decisão e Quantidade de Critérios Identificados.

Fonte: Adaptado de Luthfi, Janssen e Cromptvoets [90] e Máchová e Lněnička [54].

Referencial Teórico	Nº de Critérios de Risco	Método
Luthfi, Janssen e Cromptvoets [90]	12 critérios	Árvore de Decisão
Luthfi, Janssen e Cromptvoets [18]	5 critérios e 15 subcritérios	Redes Bayesianas
Luthfi e Janssen [91]	7 critérios e 13 subcritérios	
Botchkarev [92]	4 critérios	Trade-off de decisão para abertura de dados em relação aos riscos associados
Zuiderwijk e Janssen [93]	7 critérios	
Manfren et al. [94]	5 critérios	Framework Qualitativo
Buda et al. [95]		
Luthfi, Janssen e Cromptvoets [96]		
Lyimo et al. [97]	12 critérios	Fuzzy AHP
Máchová e Lněnička [54]	3 critérios e 14 subcritérios	
Luthfi et al. [98]	4 critérios e 8 subcritérios	
Kubler et al. [99]	5 critérios e 9 subcritérios	AHP-TOPSIS
Shaikh et al. [100]	4 critérios	
Parung et al. [101]	5 critérios e 27 subcritérios	Fuzzy AHP-TOPSIS
Rafi et al. [102]	5 critérios	Fuzzy TOPSIS

Árvore de Decisão e Técnica *Delphi*

Luthfi, Janssen e Cromptvoets [90] desenvolveram um Modelo, ora denominado *Decision Tree Analysis for Open Data* (DTAOD), que se utilizou da abordagem de Árvore de Decisão para estimar os custos e os benefícios relacionados ao processo de abertura de dados, apresentando as seguintes alternativas: Abrir (A1), Limitar Acesso (A2) e Fechar (A3). Os valores estimados foram sustentados pelo julgamento de apenas 4 especialistas. Apesar de não ser citada no aludido estudo, a técnica estruturada para obter a opinião dos especialistas muito se assemelha à Técnica *Delphi* que, conforme a ISO 31000:2018 [84], é fortemente aplicada na etapa de Identificação de Riscos.

A escolha do Modelo DTAOD se mostra interessante quando se objetiva prever as possíveis consequências e antecipar o *payoff*²⁴ referente aos caminhos alternativos que envolvem decisões complexas. Por outro, há de se ressaltar que qualquer alteração nos valores estimados pode conduzir o modelo para caminhos e alternativas menos assertivos, impactando na confiabilidade desse Modelo [37]. Destacam-se, no Quadro 3, os critérios de riscos mais relevantes identificados a partir desta revisão.

Quadro 3: Principais Critérios utilizados pelo método *Decision Tree*.

Fonte: Adaptado de Luthfi, Janssen e Cromptvoets [90].

Critério de Risco	Descrição
Alto custo na produção ou manutenção dos dados abertos	Verifica os mecanismos de produção e coleta do conjunto de dados sobre as variáveis de interesse utilizando técnicas específicas
Alto custo para a divulgação e apresentação dos dados	Verifica tecnologias utilizadas na publicação, acesso e visualização dos dados, em uma interface interativa e amigável.
Baixa ou ausência da participação social na elaboração do PDA	Verifica a realização das consultas públicas e o nível de participação e interação social para a priorização dos dados abertos.
Nível de Manutenção	Verifica a adequação do processo da gestão dos dados.
Classificação e Privacidade	Verifica se os dados pessoais, sensíveis, sigilosos e/ou confidenciais estão armazenados e protegidos de maneira adequada.
Aversão ao risco	Verifica as soluções/alternativas de abertura de dados que oferecem menos risco.

²⁴A grosso modo, *Payoff* = (Ganho/Prejuízo), ou o resultado apresentado após a execução de um conjunto de ações.

Redes Bayesianas

Redes Bayesianas são usadas para inferir os parâmetros do modelo decisório, tendo em vista os dados históricos disponíveis. Nesse método, as possíveis alternativas deixam de ter respostas simplesmente binárias (abrir/fechar ou falso/verdadeira) e permitem indicar o nível de preferência relativa, estruturadas pelo teorema de Bayes, que se aplicam à maioria dos problemas de decisão para chegar na *probabilidade a priori*, a partir dos dados examinados na *probabilidade posteriori* [103].

O teorema de Bayes, adotando-se a convenção de Skilling e Gull [104], é usualmente descrito pela probabilidade condicional representada pela seguinte equação:

$$P(\theta|\mathbf{D}) = P(\theta) \frac{P(\mathbf{D}|\theta)}{P(\mathbf{D})} \quad ||I, \text{ Onde :} \quad (2.4)$$

- $P(\theta)$ é a *probabilidade a priori* da alternativa-hipótese θ , inferindo a crença em θ antes de conhecer e analisar os dados representados por \mathbf{D} ;
- $P(\mathbf{D})$ é a *probabilidade a priori* dos dados de treinamento (\mathbf{D});
- $P(\theta|\mathbf{D})$ é a *probabilidade posterior*, inferindo nossa crença em θ após a análise dos dados (\mathbf{D}); e
- $P(\mathbf{D}|\theta)$ é a probabilidade de \mathbf{D} (verossimilhança ou, em inglês, *Likelihood*²⁵), considerando θ como verdade.

Modelos de apoio à decisão, parametrizados em Redes Bayesianas, foram propostos por Luthfi e Janssen [91], Luthfi, Janssen e Crompvoets [18], ponderando potenciais riscos e benefícios associados à abertura de dados históricos da saúde de pacientes.

Após revisão da literatura, Luthfi e Janssen [91] propuseram um modelo relacional de abertura de dados com base em 13 subcritérios de riscos, os quais foram distribuídos em 07 categorias (critérios): Imprecisão, Mau Uso, Sensibilidade, Privacidade, Violação, Consistência e Integração.

O Quadro 4 apresenta os 5 dos 7 critérios levantados por Luthfi e Janssen [91], os quais configuram os primeiros critérios considerados relevantes para este projeto, pinçados a partir da citada pesquisa.

²⁵A verossimilhança faz a comparação da hipótese com os dados observados.

Quadro 4: Critérios identificados e mensurados por Redes Bayesianas.

Fonte: Adaptado de Luthfi e Janssen [91].

Critério de Risco
Dados (ou metadados) imprecisos, vulneráveis ou com erros identificados.
Dados privados e/ou pessoais não anonimizados.
Dados sensíveis, confidenciais, sigilosos, restritos e/ou provenientes de bases restritas.
Dados complexos (diferentes fontes/origens/curadores) e estruturalmente heterogêneos.
Dados duplicados e/ou desatualizados.

Destacam-se as etapas para a construção do modelo de apoio à decisão proposto por Luthfi, Janssen e Crompvoets [18], que envolve a definição das cinco categorias de critérios de risco listados no Quadro 5, a saber:

- Definir critérios e subcritérios de riscos;
- Construir uma estrutura de rede para mapeamento do problema de decisão;
- Validar o modelo por meio da Análise de Sensibilidade; e
- Formalizar e apresentar o modelo, por meio plataforma informatizada, a fim de facilitar a compreensão e a execução do modelo por parte dos *stakeholders*, sobretudo, em termos de análise e avaliação dos riscos relacionados ao processo de abertura de dados.

Quadro 5: Critério de Riscos identificados e avaliados por Redes Bayesianas.

Fonte: Adaptado de Luthfi, Janssen e Cromptvoets [18].

Critério de Risco	Descrição
Imprecisão	Verifica se os dados abertos possuem erros identificados e/ou dados nulo ou faltantes (<i>missing</i>).
Uso indevido	Verifica se houve alguma ocorrência, incidência ou risco de imagem ou de vazamento de dados pessoais ou sensíveis, por exemplo, associados à publicação de determinado <i>dataset</i> .
Sensibilidade	Verifica ocorrência da divulgação de atributos sensíveis e/ou classificados (confidenciais).
Dados incompletos	Verifica a existência de metadados obrigatórios e desejáveis.
Má interpretação	Verifica fatores de riscos que podem causa uma má interpretação dos dados, tais como omissão de variáveis relevantes para o conjunto de dados, apresentação ou visualização inadequada, erro de correlação entre os atributos, entre outros fatores.

As limitações insculpidas no mencionado trabalho são: i) utilizar somente a revisão da literatura para fundamentar o peso dos critérios de risco e ii) não mensurar a probabilidade dos fatores/critérios de risco relacionados.

Modelo *Trade-off* Quali-quantitativo

Análise qualitativa de dados de entrevistas semiestruturadas foi realizada por Buda et al. [95] e Luthfi, Janssen e Cromptvoets [96] na proposição de estruturas conceituais, *Frameworks* Qualitativos, de suporte à decisão para abertura de dados, oferecendo, respectivamente, recomendações e estratégias de abertura de bases sob as Dimensões *Custo* e *Ecossistema* (fechado, restrito ou aberto) e para comparar políticas de dados abertos.

A Técnica *Trade-off*, comumente utilizada para a compensação dos pesos dos critérios e potenciais riscos relacionados à decisão de abertura de dados (avaliando os prós e contras dessa decisão), foi abordada por Zuiderwijk e Janssen [93]. Nesse trabalho, não se identificou a análise quantitativa.

Uma proposta qualiquantitativa para a priorização de bases de dados abertos foi desenvolvida por Botchkarev [92]. Ela se centra na avaliação do melhor custo-benefício ajustado aos seguintes critérios: valor, custo, risco e urgência. Sendo assim, as questões elencadas no Quadro 6 foram suscitadas pelo autor para quantificar subcritérios e obter o nível de priorização para a abertura de bases de dados governamentais.

Quadro 6: Questões-Chave para mensurar a priorização da abertura de dados.

Fonte: Adaptado de Botchkarev [92].

Dimensão	Questão-Chave
Valor	A coleta ou produção de dados foi exigida por um ato legal ou estatuto? O conjunto de dados abrange interesses de vários setores da economia?
Custo	Qual é o custo e/ou esforço geral estimado para a preparação de dados? Qual é o retorno sobre o investimento (ROI) sobre um conjunto de dados abertos?
Risco	Há algum risco potencial acerca da credibilidade ou integridade dos dados? Existem riscos de os dados serem mal interpretados?
Urgência	Existem processos ou atividades que necessitam utilizar dados e/ou informações governamentais de forma estruturada e aberta? Existem prazos iminentes para essas atividades? A divulgação de dados trará valor imediato ao público?

O nível de prioridade de um conjunto de dados foi calculado utilizando a expressão: $Prioridade = (Valor - Custo) \times [(1 - Risco) + Urgência]$. Para isso, a pontuação de cada resposta obtida foi registrada considerando os seus limites mínimo e máximo, respectivamente, $VR_n = 1$ e $VR_n = 5$, tendo em vista a escala Escala de Likert de cinco pontos utilizada pelo autor [105].

Dessa forma, o Critério $Valor_i$ foi mensurado a partir da soma da pontuação dada a cada resposta, considerando o peso dos critérios ($Criteria_w$) e N como o número das respectivas questões (subcritérios do Critério Valor), obedecendo à seguinte fórmula:

$$Valor_i = \sum_{n=1}^N [VR_n \times Criteria_w] \quad (2.5)$$

Entende-se que a abordagem apresentada por Botchkarev [92], apesar de simplificar o cálculo do valor de prioridade dos *datasets* analisados, não resolve os desafios de imprecisão para a mensuração dos pesos dos subcritérios analisados, coletados via um questionário

escalonado semanticamente em 5 categorias, implicando maior subjetividade na avaliação dos critérios definidos.

2.10 Métodos Multicritério sobre Dados Abertos

Fuzzy Analytical Hierarchy Process (FAHP)

Fuzzy Analytical Hierarchy Process (FAHP), método que faz parte da abordagem MCDA, foi utilizado por Luthfi et al. [98] para analisar os riscos e os benefícios associados à abertura de bases com históricos da saúde de pacientes. Nesse trabalho, foram empregados 4 critérios referentes às dimensões *Riscos* e *Benefícios*, além dos respectivos subcritérios, definidos a partir da revisão da literatura e do julgamento de especialistas (Técnica *Delphi*). Apoiados pela técnica FAHP, foram propostas as seguintes alternativas referentes ao processo de abertura de dados: Eliminar Registros (A1), Fornecer Acesso Limitado (A2), Manter Fechada (A3) e Abrir (A4).

Um modelo para a avaliação e priorização de sistemas de gerenciamento de dados governamentais abertos (CKAN, DKAN, Junar, OGPL, Prognoz e Socrata) foi proposto por Máchová e Lněnička [54]. A problemática de decisão foi estruturada em uma hierarquia de quatro níveis (objetivo do modelo, 3 dimensões/critérios, 14 subcritérios e as 6 alternativas), fornecendo uma base para a aplicação do método FAHP e *ranking* dos sistemas avaliados. O resultado desse estudo apresentou o CKAN como o sistema mais adequado. Para validação do modelo, foi realizada uma análise de sensibilidade envolvendo cinco especialistas.

Nessa linha, Lyimo et al. [97] realizou o cruzamento de dados abertos geoespaciais com outras fontes de dados governamentais, para estimar, com maior precisão e a partir de mapas de decisão orientado a riscos representados graficamente, a distribuição populacional da região de Dodoma, capital da República Unida da Tanzânia. Essa estimativa foi realizada pelo método Fuzzy AHP, que serviu para modelar a imprecisão apresentadas nos conjuntos de dados coletados pelos autores.

AHP-TOPSIS

Kubler et al. [99], após identificarem a ausência de modelos de medição que avaliassem os critérios de qualidade de dados e metadados publicados em portais de dados abertos, aplicaram as técnicas do Processo de Análise Hierárquica (AHP) e a Ordem de Preferência por Similaridade a uma Solução Ideal (TOPSIS) para avaliar a qualidade dessas publicações em 144 portais, comparando os metadados atinentes em relação às preferências e necessidades dos usuários.

A proposta apresentada no citado estudo [99], que muito se aproxima com o escopo desta dissertação, utiliza o AHP para a ponderação dos critérios de preferências, subsidiando o TOPSIS no ranqueamento final das alternativas (144 portais). Os autores utilizam 5 dimensões (Uso, Completude, Abertura, Endereçamento e Recuperação) que, reunidas, parametrizaram e mensuraram a qualidade daqueles portais em relação ao conteúdo disponibilizado. O Quadro 7 apresenta a descrição de cada um desses critérios.

Quadro 7: Critérios identificados e mensurados pelo AHP-TOPSIS.

Fonte: Adaptado de Kubler et al. [99].

Critério	Descrição
Uso	Quantifica os tipos de metadados usados por um conjunto de dados disponíveis em um determinado portal.
Completude	Verifica o nível de completude dos metadados utilizados em um determinado conjunto de dados abertos.
Abertura	Verifica o nível de conformidade dos recursos em relação ao formato e à licença
Endereçamento	Verifica a existência de metadados relativos aos canais de contato do produtor ou mantenedor do conjunto de dados abertos disponibilizados.
Recuperação	Verifica se o recurso está disponível por meio da URL informada, conferindo se o acesso via o método HTTP GET possui alguma restrição e/ou erro.

Utilizando a mesma abordagem híbrida do AHP-TOPSIS, Shaikh et al. [100] apresentam uma proposta que permitiu verificar a localização ideal para a instalação de um posto comercial. O referido estudo envolveu a análise de um conjunto de dados abertos geoespaciais, considerando os três seguintes critérios: concorrência do local, condições de tráfego e popularidade do lugar.

O estudo proposto por Parung et al. [101] adota o método *Fuzzy* AHP-TOPSIS para priorizar estratégias e ações da OGD, envolvendo 27 subcritérios distribuídos em 5 dimensões/critérios (Legal, Cultural, Social, Técnico e Econômico). No total, dez estratégias foram identificadas como prioritárias para a adoção nas organizações públicas. Seguindo a mesma metodologia, Rafi et al. [102] identificaram os principais fatores de riscos que impactam negativamente o processo de avaliação da qualidade de dados no ambiente DevOps²⁶.

²⁶A palavra DevOps envolve os termos “*development*” e “*operations*” para oferecer

O trabalho desenvolvido por Luthfi e Janssen [106], intitulado *A Comparative Study of Methods for Deciding to Open Data*, apresenta uma análise comparativa entre os métodos *Decision Tree* (Árvore de Decisão), *Bayesian Network* (Redes Bayesianas) e *Fuzzy MCDM* (Lógica Difusa de Decisão por Multicritérios), ressaltando os pontos fortes e fracos de cada um desses métodos no processo decisório para a abertura de dados.

O Quadro 8 apresenta uma síntese desse comparativo, o qual foi adaptado ao acrescentar, ao referido Quadro, o método híbrido AHP-TOPSIS.

Quadro 8: Comparativo dos Métodos Identificados.

Fonte: Adaptado de Luthfi e Janssen [106].

Método	Vantagens	Desvantagens
AHP Fuzzy	Sólidas construções de modelos decisórios, especificamente voltados para problemas de classificação.	Complexidade moderada e exige rígida análise para construir a matriz de comparação par a par
Redes Bayesianas	Regras bem definidas para resolução de problemas probabilísticos com alto grau de incerteza	Complexidade alta e dependência de um agente externo (observação heurística).
AHP-TOPSIS e Árvore de Decisão	Fácil compreensão e interpretação do modelo.	Alterações das variáveis durante o processo de análise podem exigir redesenho do modelo.

2.11 Métodos Decisórios sobre Dados Abertos do Setor Elétrico

As razões pelas quais os dados do setor de energia devem ou não ser abertos são avaliadas por Pfenninger et al. [107], que apresentaram um modelo qualitativo (orientativo-conceitual), discutindo os benefícios dessa abertura e os motivos pelos quais estes devem permanecer fechados. Os principais critérios pertencentes aos domínios *Benefícios* e *Riscos*, selecionados e transcritos no Quadro 9, sustentaram o referido trabalho e inspiraram a construção do modelo proposto.

Quadro 9: Dimensões de Riscos e Benefícios da Abertura de Dados de Energia.

Fonte: Adaptado de Pfenninger et al. [107].

Benefícios	Riscos/Custos
Inovação e Qualidade dos Serviços	Problemas Éticos e de Segurança
<i>Feedbacks</i> e colaboração mais eficaz	Exposição Indesejada
Compartilhamento colaborativo de encargos do setor	Carga de trabalho adicional
Aumento da produtividade	Inércia institucional ou pessoal

Ainda sobre o estudo apresentado por Pfenninger et al. [107], verifica-se as seguintes recomendações voltadas para um projeto *Open Energy*, tais como:

- Considerar a utilização de uma interface documentada, de modo a permitir a redistribuição fiável da fonte. A escolha da distribuição (por exemplo, por meio de um site institucional ou uma plataforma estabelecida para compartilhamento de dados abertos do setor elétrico) pode influenciar a utilização desses dados pelos *stakeholders*.
- Decidir acerca da base que deve ser aberta e o respectivo volume de dados. Os gestores não devem se sentir forçados a tomar uma decisão binária de abrir ou não abrir. Publicar partes de um conjunto de dados pode ser melhor do que não publicar nada, e pode ser visto como parte de um processo incremental para decidir o quê e quanto deve abrir, dependendo dos objetivos e contextos analisados.
- Utilizar uma licença adequada (como a licença MIT para código ou a licença CC-BY para dados) ou uma licença *copyleft* ou compartilhada (como GNU GPL ou CC-BY-SA). Embora isso limite a escolha da licença (por exemplo, a licença GPL é tecnicamente incompatível com um modelo que depende de um ambiente de código-fonte fechado para ser executado), os modelos construídos em tais ferramentas ainda podem ser licenciados abertamente e disponibilizados. Pode ser consultada a página da *Open Definition*²⁷, que fornece uma lista detalhada com mais de 100 licenças diferentes.

Buscando melhorar a eficiência dos recursos públicos, Palevičius et al. [108] combinaram vários métodos MCDM para realizar comparações de infraestruturas públicas, de 18 cidades da Lituânia, direcionadas para veículos elétricos. Um grupo de gestores públicos e pesquisadores, essencialmente dos setores de energia, transporte e meio ambiente, foi

²⁷<https://opendefinition.org/licenses/>.

consultado para a elucidação ou definição dos principais fatores e dimensões relacionados ao cenário abordado.

Têm-se, ainda, os modelos apresentados por Manfren et al. [94], os quais destacaram dados do setor energético considerados relevantes para os *stakeholders* desse setor, os quais, segundo os autores, devem ser abertos e compartilhados entre ciência-política-mercado, no sentido de criar modelos mais eficazes na identificação e resolução de problemas técnicos críticos do setor (tais como: eficiência energética, processos de conversão de energia, operação de infraestruturas de energia, etc.). A aludida pesquisa fomenta o conceito de *Open Energy*, indicando informações essenciais para o monitoramento, análise e rastreabilidade de dados relativos ao desempenho de modelos, processos, projetos/operações de sistemas energéticos.

2.12 Considerações do Capítulo

Até aqui, foram discutidos os diversos parâmetros arquiteturais do padrão *Open Data* e suas variações - *Linked Open Data (LOD)* e *Open Government Data (OGD)* - que compreendem um conjunto de princípios direcionados para o compartilhamento de informações *online* (*readiness*) legíveis por máquina, de modo a permitir o seu reuso por qualquer pessoa, organização ou *software*. Nesse sentido, buscou-se demonstrar a evolução da arquitetura *Open Data*, fundamentada pela Revisão Sistemática da Literatura, a fim de elencar e relacionar os princípios arquiteturais do LOD e OGD aos correspondentes subcritérios de riscos identificados.

O conteúdo revisado, complementado pela análise dos trabalhos relacionados, também proporcionou identificar técnicas/métodos implementados em modelos multicritérios, os quais foram utilizados para avaliar dados abertos em diferentes contextos, sobretudo, no setor elétrico. Dessa forma, a Revisão Sistemática da Literatura (RSL) com base nos trabalhos relacionados permitiu encontrar 78 critérios de risco para posterior definição e análise apresentada na Seção 4. Também foi possível elencar diferentes abordagens de métodos decisórios pertinentes ao tema, destacando o método híbrido AHP-TOPSIS, utilizado por Kubler et al. [99] e Parung et al. [101].

A RSL dos trabalhos relacionados apontam para a necessidade de um *framework* capaz de avaliar critérios relevantes ao contexto de *dados abertos* do setor elétrico. A partir do estudo secundário desses trabalhos, pode-se inferir, ainda, a inexistência da aplicação de métodos multicritérios (MCDA/MCDM), principalmente, em dados relacionados a esse setor.

Capítulo 3

Estudo de Caso

Este Capítulo apresenta o detalhamento das etapas sustentadas pela Fase **Estudo de Caso**, conforme apresentado no Plano Metodológico (Seção 1.5.2) traçado para o desenvolvimento desta pesquisa. As etapas foram agregadas em dois ciclos distintos, ciclo do Modelo CRISP-DM e ciclo do Modelo MCDA.

As etapas que compõem o ciclo do Modelo CRISP-DM são:

- Entendimento e Preparação do Dados (3.1.1);
- Modelagem (3.1.2); e
- Avaliação (3.1.3).

Em paralelo, foram empregadas e detalhadas as seguintes etapas do Modelo MCDA proposto nesta pesquisa:

- Estruturação do Problema Decisório (3.2.1);
 - Definição dos Critérios de Risco (3.2.2); e
 - Decomposição Hierárquica do Problema (3.2.3).
- Construção do Modelo (3.3); e
- Validação (3.4).

3.1 Aplicação do Modelo CRISP-DM

3.1.1 Entendimento e Preparação dos Dados

Esta etapa envolve as seguintes tarefas genéricas e específicas, as quais facilitaram o entendimento e a preparação dos dados, especificamente no processo de descoberta de temas/tópicos potenciais para a abertura de dados.

- a) Tarefas genéricas: determinar objetivos do contexto/negócio, analisar a situação atual, realizar a coleta, exploração e verificação da qualidade dos dados.
- b) Tarefas específicas: avaliar recursos e técnicas para a coleta inicial, definir origem dos dados para treinamento, selecionar ferramentas e técnicas da mineração de dados/texto.

Após o entendimento do contexto e dos dados, ocorre a preparação do ambiente para a realização do pré-processamento dos dados na construção do *corpus* a ser utilizado na modelagem de mineração. Nesta etapa, técnicas e ferramentas de limpeza e transformação de dados (*Bag of Words*, *Stemming*, eventual anonimização de dados confidenciais, sensíveis e/ou pessoais, *Lemmatization*, remoção de *stop words*, entre outras) foram aplicadas para encontrar os termos e palavras mais comuns em dicionários criados pelo processo de *tokenização*.

Foram filtrados os *tokens* que aparecem em mais de 50% da fração total do *corpus*. A amostra restante foi separada em um *Bag of Words (BoW)* para a seleção de *n-Grams*. Considerou-se *n-Grams* com até quatro palavras.

Análise de Conteúdo

A primeira macro etapa deste trabalho compreendeu em delimitar, ante à análise de conteúdo, os dados não bibliométricos (estruturados, semiestruturados e não estruturados), os quais correspondem aos relatórios de gestão e recentes planos energéticos e de dados abertos produzidos pelos órgãos e entidades avaliados, conforme o resumo apresentado na Tabela 3.1.

A coleta e a análise de conteúdo, submetido ou não ao tratamento estatístico, são tarefas características da pesquisa exploratória [32, 38], configurando uma importante etapa do processo metodológico empregado neste trabalho.

Revisão Sistemática da Literatura

A partir do estudo secundário da literatura, objetivando alcançar a maior variedade de publicações relevantes, foram consultadas as seguintes bases bibliográficas eletrônicas: *IEEEExplore*, *Web of Science*, *Scopus*, *ACM Digital Library*, *Google Scholar*, *ScienceDirect* e *Scielo*, além dos conteúdos governamentais concernentes à temática proposta para esta pesquisa (leis, decretos, portarias, planos, relatórios, entre outros documentos).

No total, 504 publicações atinentes foram identificadas nas bases de dados supracitadas, as quais foram reduzidas para 114 artigos relevantes, a partir do refinamento da busca pela análise do extrato bibliográfico e do mapeamento da literatura. Considerou-se priorizar

os trabalhos publicados em periódicos e conferências internacionais indexados às bases *IEEEExplore*, *Web of Science*, *Scopus* e *ACM Digital Library*.

Os títulos pesquisados têm enfoque nos termos (“OGD (Linked)” OR “Linked Open Data” OR “Open Energy”) AND “Risk Opening Data” OR “Risk Analysis” OR “risk criteria”).

Seleção e Coleta dos Conjuntos de Dados

Conforme o resumo apresentado na Tabela 3.1, além das fontes primárias de pesquisa (relatórios e planos do setor elétrico), foram encontrados 50 conjuntos de dados abertos que serviram de horizonte para a análise desta pesquisa. Tais bases foram catalogadas no PBDA pelo MME e pela ANEEL. Desse total de bases (50), verificou-se que duas delas não faziam parte do escopo do setor elétrico e foram excluídas dessa análise.

Tabela 3.1: Escopo da Análise de Conteúdo.
Fonte: Elaboração Própria.

Órgão/ Entidade	Corpus Empírico		Datasets encontrados no PBDA
	Documentos	PDA	
MME	0	1	3
EPE	2 ^a	0	0
ANEEL	0	1	47
Eletrobras	15 ^b	N/A ^c	0

^aPlano Decenal de Expansão de Energia 2030 e Plano Nacional de Energia 2050.

^bRelatórios de Gestão referente aos exercícios de 2005 a 2019.

^cN/A = Não se Aplica. A Eletrobras, apesar de vinculada ao MME, é uma empresa de economia mista e por isso não é obrigada à publicar seus dados no PBDA. [6].

A Eletrobras não aparece na pesquisa apresentada pela Figura B.1, tendo em vista a sua natureza jurídica-econômica, visto que, conforme o próprio Decreto nº 8.777/2016 (inciso I do artigo 1º), apenas os órgãos e entidades da APF direta, autárquica e fundacional devem promover a abertura de seus dados, catalogando-os no PBDA.

À EPE também não se aplica a Política de Dados Abertos, visto que a Empresa pública possui personalidade jurídica de direito privado. A esse respeito, não obstante e em prol dos benefícios fomentados pela conversão desses dados em valor agregado, considera-se boa prática que, até mesmo as entidades pertencentes à administração pública indireta, divulguem seus dados, sobretudo em formatos abertos (a exemplo do Banco Nacional de Desenvolvimento Econômico e Social, que possui 23 conjuntos de dados publicados no PBDA). A divulgação de Dados Abertos Governamentais não deve se restringir apenas aos órgãos e entidades da administração direta da APF.

Análise Documental - Fontes PBDA, PDA e Planos setoriais¹

Foram selecionados apenas os dados dos órgãos e entidades vinculados ao MME, os quais fazem parte do escopo deste trabalho. Dessa forma, verifica-se no PBDA, 50 conjuntos de dados pertinentes ao setor elétrico, conforme esquematizado na Tabela 3.1.

Para coleta, análise e avaliação dos dados do setor elétrico, foram seguidas as etapas sugeridas por Böhms, Rieswijk e Lijster [109], que são:

1. Pesquisar os conjuntos de dados e metadados de energia elétrica publicados no PBDA;
2. Priorizar a coleta automatizada via API e uso da técnica de *Web Scraping* para captura dos conjuntos de dados e respectivos metadados;
3. Reunir os conjuntos de dados relevantes para avaliação; e
4. Interpretar os metadados dos *datasets* selecionados.

O Plano Nacional de Energia (PNE) 2050 é um relatório que auxilia os *stakeholders* a construir consensos em torno de grandes elementos norteadores para o setor energético, a partir de uma leitura de **custos e benefícios** de longo prazo para o País.

O Plano Decenal de Expansão de Energia (PDE) 2030² é um documento informativo, elaborado anualmente pela EPE sob as diretrizes e o apoio do MME. O principal objetivo do PDE é indicar as perspectivas, sob a ótica do governo, da expansão do setor de energia no horizonte de dez anos, extraíndo importantes elementos para o planejamento do setor de energia.

Os referidos Planos estão disponibilizados em formato PDF, arquivo de leitura humana, ou seja, para ser processado por máquina foi necessário a preparação e manipulação do documento. Dessa forma foram analisados os documentos PNE e PDE, para extração dos *corpora* e criação de uma base não estruturada.

Os *corpora* gerados permitem a análise automatizada em relação ao conteúdo linguísticas do texto, descobertas de padrões e inferências. Alguns exemplos disso são as técnicas *tokenization*, lematização (*lemmatization*), marcação de classes gramaticais e relações sintáticas (*syntactic relation*).

Ainda nesta etapa foram realizadas a limpeza e a integração dos dados para posterior análise. O Software *RStudio* foi utilizado para a leitura, preparação e transformação dos

¹Releva mencionar que planos setoriais analisados não se confundem aos instrumentos transversais de governo, tais como o Plano Plurianual.

²Disponível em: https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-490/PDE%202030_RevisaoPosCP_rv2.pdf

dados não estruturados e semiestruturados para um formato estruturado. O produto final desta etapa de preparação é o *corpus*.

Foi realizado *merge* dos *datasets* oriundos dos Planos e documentos selecionados, e aplicação do método de distribuição Holdout, orientado por [110], para separação das amostras em 70% para treinamento e 30% para teste e validação.

3.1.2 Modelagem

Esta etapa, que objetiva a identificação dos temas potenciais para a abertura, corresponde à definição do modelo a ser utilizado para mineração de texto que, em termos práticos, envolve a escolha das tarefas específicas do problema de mineração: agrupamento e classificação, com aprendizado não supervisionado, para análise de desempenho preditivo e identificação de informações relacionadas às bases de dados do setor elétrico catalogadas no PBDA.



Figura 3.1: Macro etapas do Processo de mineração de texto.

Fonte: Elaboração própria.

Para se obter uma representação computacional do conteúdo analisado, foi aplicada a técnica TF-IDF (em inglês *Term Frequency-Inverse Document Frequency*). Seguindo a abordagem do trabalho apresentando por Palma A., Ladeira e Reis [10], foram atribuídos pesos para cada termo (frequência dos termos), para assim, medir a relação entre as palavras e um termo específico dentro do *corpus*, enfatizando a sua raridade (frequência inversa do documento), antes de iniciar o processamento de similaridade e classificação.

A Figura 3.1 ilustra as macro etapas do Processo de Mineração de Dados realizada sobre o *corpora* gerado a partir da mencionada análise documental realizada para a definição dos

O resultado apresentado na Figura 3.2 ratifica o enfoque desta pesquisa, apontando para subtópicos afetos ao contexto do setor de energia elétrica (empreendimentos, capacidade instalada, consumo, agentes, transmissão, sistema interligado nacional, geração, tarifa, entre outros), que devem abarcar o conjunto de dados analisados. Dessa forma, infere-se que:

- a) *Energia Elétrica* teve um papel proeminente nas projeções energéticas para 2030 e 2050.
- b) Os principais termos encontrados revelam subtemas potenciais para compor a matriz de priorização e análise quanto aos critérios de risco selecionados, tais como:
 - Empreendimentos: dados/informações sobre licenciamento ambiental, leilões, outorgas, concessões e afins;
 - Indústria: informações, por exemplo, sobre as etapas do fornecimento de energia (geração, transmissão, distribuição e comercialização), do uso de energia elétrica para acionar novas linhas de produção ou do consumo de energia elétrica no processo industrial; e
 - Consumo de Energia: informações sobre tarifa social, bandeiras tarifárias, sobre índices de contaminação ambiental *versus* consumo de energia, consumo inteligente, entre outros.

O grupo temático das 48 bases de dados analisadas, conforme destacam as Figuras 3.3 e 3.4, corresponde aos seguintes tópicos/temas identificados: Geração, Transmissão, Distribuição, Tarifa, Pesquisa e Desenvolvimento, Fiscalização, Políticas Públicas, Eficiência Energética, Qualidade de Energia, Distribuição, Expansão da Oferta/Consumo e outras informações relacionadas à transparência, ouvidoria e dados abertos.

A partir da análise estatística realizada sobre os Planos PNE e PDE, verifica-se uma maior atenção aos empreendimentos direcionados para o setor de energia elétrica. No entanto, menos de 5% dos dados abertos, relacionados aos temas identificados, foram disponibilizados ou custodiados pelo MME.

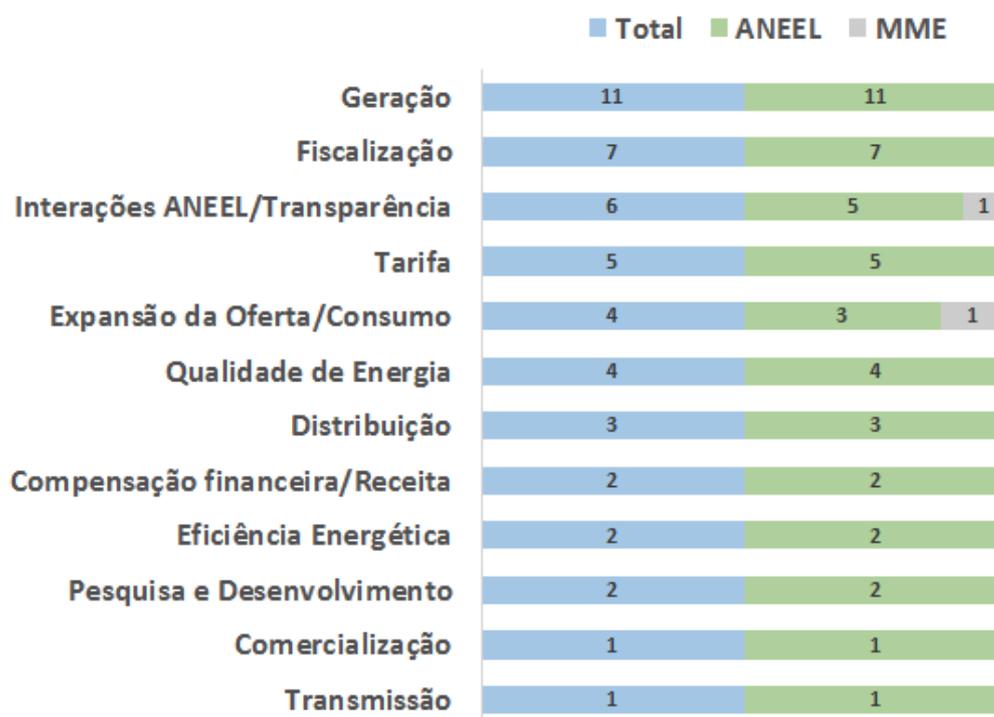


Figura 3.3: Quantitativo de bases abertas distribuídos por tópico
Fonte: Elaboração própria.

A Figura 3.4 apresenta, além do quantitativo, o código (ID) das bases relacionadas a cada tema/tópico identificado. As bases avaliadas estão listadas no Quadro 12, contida no Apêndice C desta trabalho, com os respectivos *links* para acesso.

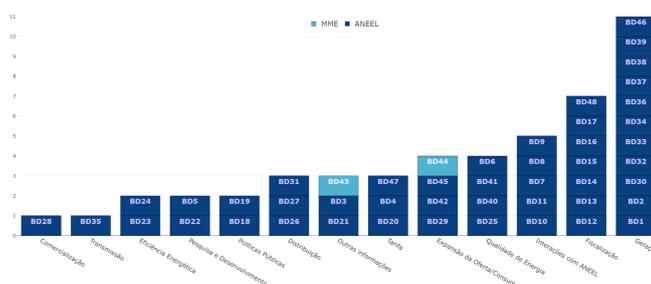


Figura 3.4: Identificação dos *datasets* por tópico.
Fonte: Elaboração própria.

Os 25 temas emergidos dessa análise, distribuídos em 48 bases abertas, demonstram a pulverização desses assuntos em diferentes *datasets* avaliados. Tais temas, encontrados após a análise estatística dos principais Planos do setor elétrico (PDE e PNE), foram confrontados com o grupo de tópicos relacionados às bases catalogadas no PBDA, permitindo, após coleta de metadados correspondentes (*tags dos recursos*), mensurar o critério de risco relacionado aos dados que possuem "baixa relevância para a sociedade ou *stakeholders*",

refletindo a expectativa desses atores em consumir dados abertos relacionados aos temas revelados.

3.2 Aplicação do Modelo MCDA

Para se chegar no objetivo do problema decisório (*Ranking* das Alternativas), esta Seção implementa as seguintes etapas: a) Estruturação do Problema Decisório, que trata da definição dos critérios de risco identificados; b) Construção do Modelo, amparada pela aplicação dos métodos AHP e TOPSIS; e c) Validação do Modelo, fundamentada pela Análise de Sensibilidade que reforça a robustez do modelo.

3.2.1 Estruturação do Problema Decisório

A Etapa de Estruturação do Problema Decisório envolve entender e identificar o problema de decisão, com vistas a sua estruturação (identificação das possíveis soluções/alternativas e o exame da sua viabilidade) de acordo com o contexto a ser analisado³ [66, 33].

A construção do modelo proposto neste trabalho está em linha com o problema de ordenação e posterior classificação das alternativas (*ranking* dos *datasets* avaliados por similaridade em relação à distância da solução ideal e anti-ideal). Para tal, se faz necessário analisar e definir as alternativas, os critérios e subcritérios de riscos.

As alternativas e os critérios definidos estão listados, respectivamente, nos Apêndices C e E deste trabalho.

3.2.2 Definição dos Critérios de Riscos

Em Saaty [112, 113] são sugeridas a utilização das Técnicas de *brainstorming* e de *Delphi*, além da análise bibliográfica, para a eleição de critérios relevantes ao modelo de decisão. O Quadro 14, disponível no Apêndice E, apresenta a lista dos Critérios de Risco, e respectivos subcritérios, possibilitando uma rápida consulta a esses elementos. Foram definidos 12 (doze) subcritérios, consolidados em 05 (cinco) critérios, definidos a partir dos Princípios arquiteturais do padrão *Open Data*, LOD e OGD, além de fundamentados pela análise de conteúdo e revisão da literatura.

A definição dos critérios corresponde a um dos objetivos específicos estabelecidos na Subseção 1.4. No entanto, os critérios de risco e respectivos subcritérios definidos (vide Quadro 10), que representam os principais *drivers* para o modelo decisório, seguem o encadeamento do processo de apoio a decisão (Figura 2.9) proposto por Belton e Stewart

³O contexto específico deste estudo de caso está descrito no subitem 1.1, na Introdução deste trabalho.

[37] e Marttunen et al. [66], os quais precedem à decomposição do problema e aplicação dos métodos multicritério.

Quadro 10: Critério/Dimensão e Subcritérios.

Fonte: Elaboração própria.

Critério/Dimensão	ID Subcritérios
CR1. Formatação	C ₁₁ , C ₁₂ , C ₁₃ e C ₁₄
CR2. Padronização	C ₂₁ , C ₂₂ , C ₂₃ , C ₂₄ e C ₂₅
CR3. Atualização	Não aplicados
CR4. Rejeição	C ₄₁ , C ₄₂ e C ₄₃
CR5. Aceitação	Não aplicados

CR1 - Formatação

Formatação configura os aspectos mensuráveis relacionados ao tipo, formato, ontologia e semântica do dado. Este critério verifica se o conjunto de dados abertos possui URI/URL identificados, considerando os subcritérios elencados na Tabela F.1 no Apêndice desta dissertação, fundamentados pelos Princípios *Open Data*, LOD e OGD.

- **CR₁₁ Recursos disponíveis em formatos estruturados:** identifica o quantitativo de recursos publicados em formatos estruturados e legíveis por máquinas.
- **CR₁₂ Recurso disponível em formato não proprietário:** verifica a quantidade de recursos publicados em formatos não proprietários.

A Primeira Lei da Computação de Jack Schofield⁴ afirma que não se deve depositar dados em um *software* sem poder garantir, posteriormente, a sua retirada do domínio proprietário. Os formatos proprietários são controlados por uma empresa ou entidade de *software* específicas. Nesse caso, existe o risco dessas entidades não suportarem mais o formato de arquivo proprietário anteriormente disponibilizado. Além disso, ferramentas não proprietárias geralmente possuem documentação mais completa, com códigos documentados publicamente, o que aumenta a probabilidade de obter maior suporte por tecnologias direcionadas ao padrão aberto.

- **CR₁₃ Quantidade de recursos disponíveis e variedade de formatos:** verifica a quantidade de versões/formatos de diferentes distribuições (CSV, JSON, XML,

⁴<https://www.theguardian.com/technology/2003/jul/24/onlinesupplement.columnists>.

RDF, HTML, etc.) de um mesmo conjunto de dados e o total de recursos catalogados no Portal Brasileiro de Dados Abertos.

- **CR₁₄ Dados em formatos RDF (permitem serialização):** verifica *datasets* que possibilitam interligação semântica (*Links* RDF) para conjuntos de dados associados. Este Critério pode impactar diretamente o esforço e o tempo gastos para descobertas de bases de dados relacionados, ou seja, que fazem parte de um mesmo tema ou tópico, mas que não estejam conectadas. Neste caso, recomenda-se avaliar a necessidade de implementar *links* RDF, verificando os *datasets* que são frequentemente acessados em conjunto, por exemplo.

CR2 - Padronização

Esse critério analisa o grau de observância dos dados em relação à formatação intrínseca de cada base disponibilizada em relação ao padrão arquitetural recomendado pelos Princípios elencados no Capítulo 3.1.1.

- **CR₂₁ Coerência entre os títulos das bases no PBDA em relação ao PDA.**

Conforme o artigo 8º da Resolução nº 3/2017, as bases de dados relacionadas nos Plano de Dados Abertos (PDA) deverão ser catalogadas no PBDA, devendo possuir a mesma nomenclatura utilizada no PDA.

- **CR₂₂ Licença adequada ao conjunto de dados:** verifica o número de conjuntos de dados associados a uma licença livre.

Conforme a Fundação do Conhecimento Aberto (OKF), uma licença pode ser considerada *aberta*, *fechada* ou *desconhecida*. Nesta situação, identificar determinado recurso sob uma licença *desconhecida* aumenta o nível de incerteza da respectiva base [114].

A obtenção e uso de uma licença adequada (ou em acordo) com os conjuntos de dados abertos, quando os dados são colocados sob o domínio público, é o critério inicial básico e mínimo para a concepção do conceito *Open Data* [20]. A Licença aberta, inclusive, é uma previsão legal, regulamentada pela Política de Dados Abertos do Poder Executivo federal [6].

- **CR₂₃ Recurso possui dicionário de dados:** verifica se o conjunto de dados publicados possui dicionário de dados.
- **CR₂₄ Nível de completude de metadados obrigatórios:** verifica a quantidade de metadados obrigatórios associados ao recurso publicado no PBDA.

Checklist dos metadados obrigatórios indicados pela INDA:

- Título do conjunto de dados.
- Descrição: Uma breve explicação sobre os dados.
- Catálogo origem: Página URL do órgão onde está publicado o conjunto de dados.
- Órgão responsável: Nome e sigla do órgão ou entidade responsável pela publicação do conjunto de dados.
- Categorias no VCGE⁵
- Identificador: URL persistente que aponta para o recurso na Web.
- Título do recurso.
- Formato: Formato do recurso. Ex.: XML, JSON, CSV, entre outros.

CR₂₅ Nível de completude de metadados desejáveis: verifica a quantidade de metadados desejáveis associados ao recurso publicado no PBDA.

Checklist dos metadados desejáveis:

- Etiquetas: Lista de palavras-chave relacionadas ao conjunto de dados, as quais são úteis na classificação e busca.
- Autoria
- Cobertura geográfica
- Cobertura temporal: Data ou período à que se referem os dados.
- Granularidade geográfica: Precisão geográfica da cobertura geográfica (Exemplo: federal, estadual, municipal).
- Granularidade temporal: Precisão temporal da cobertura temporal (Ex.: dia, mês, ano).
- Referências e *tags*: Relacionamento com outros conjuntos de dados.
- Metodologia: Processo de criação dos dados.
- Vocabulário/ontologia: Documentos estruturados com metadados específicos do conjunto de dados.

⁵Vocabulário Controlado de Governo Eletrônico (VGCE) é uma lista hierarquizada de assuntos do governo que utiliza termos comuns e é voltada para a sociedade. A lista de categorias do VCGE está disponível em <https://vocab.e.gov.br/2011/03/vcge>.

CR3 - Atualização

Atualização (CR3): refere-se ao tempo decorrido (em dias) a partir da última atualização do conjunto de dados. Este critério deve observar, ainda, a granularidade da periodicidade definida para *update* dos dados/metadados.

CR4 - Rejeição

Este critério consiste em um campo binário, disponível no PBDA, que avalia aspectos negativos do *dataset* sob o ponto de vista do usuário. Neste estudo, somam-se, ainda, para cada ação de *rejeição*, os seguintes subcritérios:

- **CR₄₁ Link do *dataset* indisponível ou descontinuado:** verifica se o conjunto de dados contém recursos indisponíveis ou com algum erro que inviabilize o acesso (*broken links*).

Sobre este subcritério, ressalta-se que o PBDA é o catálogo oficial do governo que reúne dados abertos hospedados nos *sites* dos órgãos publicadores da APF. Sendo assim, é importante que os *links* (URLs) fornecidos estejam operacionais e que, se um usuário clicar no recurso, ele possa realizar o *download* desses dados ou ter acesso direto via API.

- **CR₄₂ Baixa relevância para a sociedade ou *stakeholders*:** verifica se conjunto de dados abertos publicados possui alguma relação com os temas apresentados na Subseção anterior (3.4), encontrados após a análise estatística dos principais Planos do setor elétrico. Esse critério deve considerar, por exemplo, metadados descritivos e *tags* associados a outros recursos do mesmo grupo temático.
- **CR₄₃ Número de *feedbacks* negativos:** quantifica as bases que apresentam erro ou algum evento que inviabilize o acesso a esses dados (*broken link*, por exemplo), bem como o número de *feedbacks* negativos atribuídos àquela base.

CR5 - Aceitação

Este critério verifica o campo binário disponível no PBDA, mensurando os aspectos positivos do *dataset* sob o ponto de vista do usuário.

3.2.3 Decomposição Hierárquica do Problema

Segundo Saaty [115] e Harker e Vargas [116], a Estruturação do Problema Decisório exige, ainda, a decomposição dos elementos básicos da abordagem MCDA, os quais foram apresentados no tópico anterior.

O princípio da decomposição é o primeiro passo do Processo de Análise Hierárquica (em inglês, *Analytic Hierarchy Process* - AHP), técnica desenvolvida por Saaty [113, 117], que consiste na modelagem de um sistema hierárquico, ou seja, na construção de uma hierarquia de decisão, permitindo analisar o problema em partes e sob diferentes aspectos.

Conforme Wernke e Bornia [118], a Técnica AHP facilita a tomada de decisão analisando a distribuição e os impactos causados por multivariáveis relevantes a um contexto específico, as quais são transformadas em critérios e subcritérios, e organizadas hierarquicamente em relação aos objetivos para a solução de um problema complexo.

O problema de decisão desta pesquisa foi organizado em quatro níveis que separam os principais elementos do modelo (critérios, subcritérios e alternativas), a ver:

- Nível 1: Ordenar as bases de dados do setor elétrico catalogadas no PBDA;
- Níveis 2 e 3: Dimensões, Critérios e subcritérios de Riscos elencados a partir da revisão bibliográfica, referenciados na Tabela 10;
- Nível 4: *Ranking* das alternativas, no caso, conjunto de dados abertos do setor elétrico ordenados e classificados por níveis proximidade ao padrão *Open Data*.

Dessa forma, para que o objetivo do modelo decisório fosse alcançado (primeiro nível), foi necessário identificar e definir os critérios e subcritérios de riscos que compõem a segunda e terceira camadas dessa estrutura hierárquica representada na Figura 3.5, imprimindo o resultado das alternativas existentes.

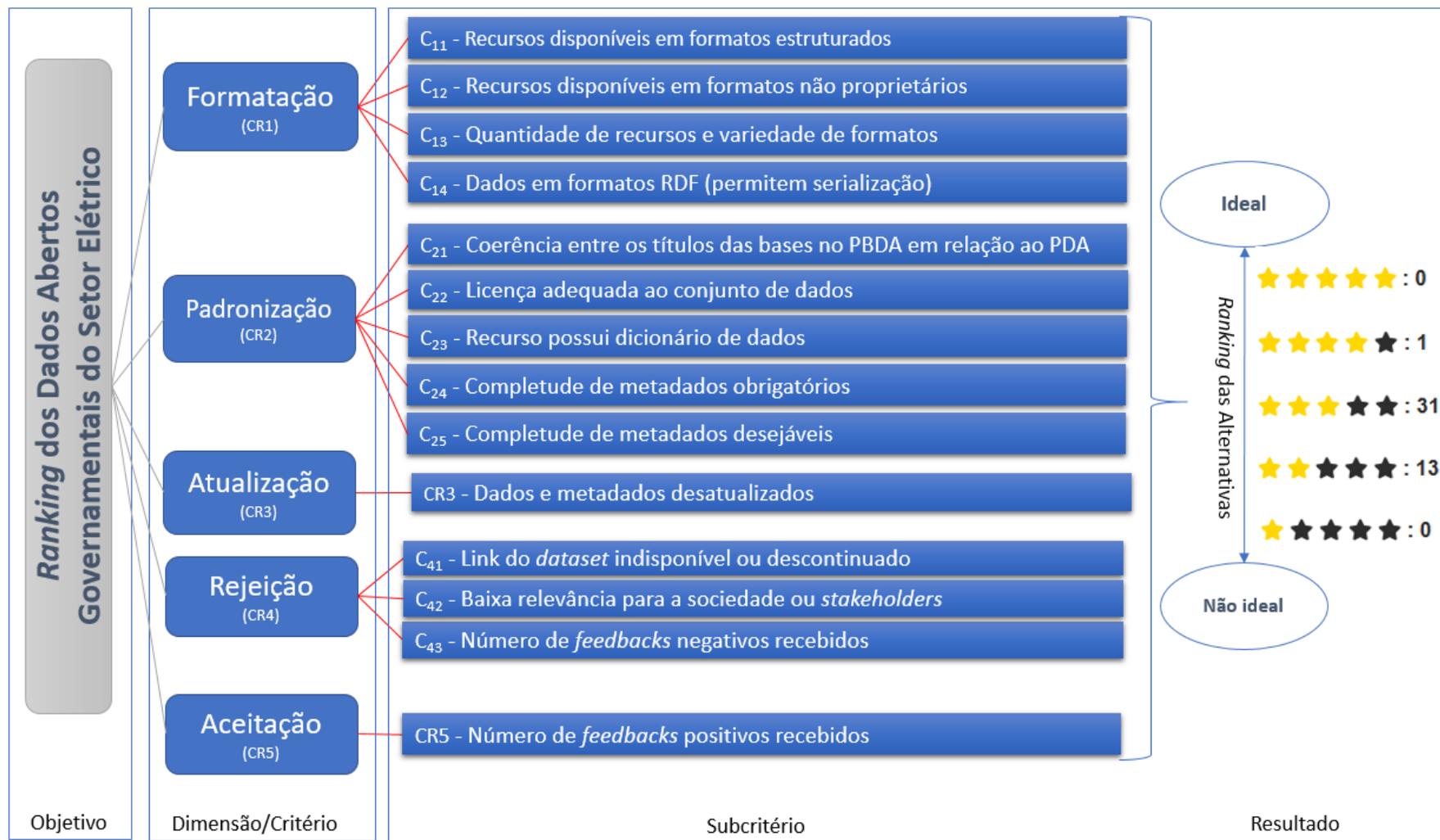


Figura 3.5: Decomposição Hierárquica do Problema.
Fonte: Elaboração própria.

3.3 Construção do Modelo pelo AHP-TOPSIS-2N

O método para Ordem de Preferência por Similaridade a uma Solução Ideal (*Technique for Order Preference by Similarity to Ideal Solution* - TOPSIS), desenvolvido por Hwang e Yoon [119], busca avaliar a proximidade (similaridade) de cada alternativa em relação aos vetores virtuais gerados pelo seu algoritmo, que simulam duas alternativas extremas, uma ideal, que representa a solução positiva (A^+) e outra não-ideal, representando a solução negativa (A^-).

O objetivo do TOPSIS é avaliar a performance de cada alternativa, estimando aquela mais próxima da solução ideal e, concomitantemente, mais distante da não-ideal. A versão do TOPSIS-2N, que realiza duas normalizações para cada *ranking* gerado, faz uso do método Analytic-Hierarchy Process (AHP) para calcular os pesos dos critérios.

Dessa forma, o método híbrido AHP-TOPSIS-2N envolve duas fases, a Ponderação dos Critérios pelo método AHP e o *Ranking* das Bases pelo método TOPSIS-2N, as quais admitem as seguintes atividades:

- Para a Ponderação dos Critérios (AHP):
 1. Construir a Matriz de Julgamentos para comparação par a par;
 2. Calcular o Vetor de Prioridades; e
 3. Verificação da Consistência: calcular o Índice de Consistência.

- Para o *Ranking* das Bases (TOPSIS-2N):
 4. Calcular o Desempenho dos Critérios: determinar a solução ideal (S_i^+) e a solução ideal negativa, ou não ideal (S_i^-);
 5. Calcular a distância euclidiana (medida de separação das alternativas): determina a distância de cada alternativa em relação à solução ideal (S_i^+) e não ideal (S_i^-);
 6. Calcular a proximidade relativa da solução ideal (S_i^+); e
 7. *Ranking* das alternativas.

3.3.1 Ponderação dos Critérios pelo método AHP

O método AHP foi aplicado seguindo a metodologia apresentada por Saaty [120] e Reis e Schramm [121], a fim de determinar os pesos dos critérios/subcritérios de riscos definidos.

Análise de Prioridades

Este passo envolve obter a opinião de especialistas para a ponderação dos critérios/sub-critérios de risco, bem como a organização dos dados em uma matriz de julgamentos, no sentido de encontrar a importância relativa desses elementos. Para isso, é preciso coletar valores numéricos que possibilitem a comparação par a par entre os critérios e subcritérios decompostos, considerando a importância atribuída entre esses elementos [112, 122, 113].

Para realização das comparações par a par, utilizou-se a escala fundamentada por Saaty [112], representada na Tabela 3.2, como forma de embasar as ponderações compensatórias (*tradeoffs*), considerando o seguinte intervalo $0 < peso \leq 9$, onde 1 se refere à *igualmente importante* e 9 à *extremamente importante*.

Tabela 3.2: Escala de Importância do AHP

Fonte: Adaptada de Saaty [122].

Escala Verbal	Valor (Julgamento)	Recíproco ^a
Igualmente importante	1	1
Levemente mais importante	3	1/3
Mais importante	5	1/5
Muito mais importante	7	1/7
Extremamente importante	9	1/9
Valores intermediários	2, 4, 6, 8	1/2, 1/4, 1/6, 1,8

^aValor positivo recíproco ao julgamento, por exemplo: $\alpha_{ij} = 1/\alpha_{ji}$, onde α é o valor do julgamento. Essa relação, que é a base do AHP, explica que se um elemento é considerado cinco vezes mais importante em comparação a outro, então esse outro deve necessariamente ter um quinto do valor de importância do primeiro.

Construção da Matriz de Julgamentos

Esta atividade envolve a construção da matriz de julgamentos para a comparação par a par dos critérios, bem como a avaliação do nível de concordância acerca dos julgamentos iniciais, a fim de obter os pesos dos critérios, necessários para posterior aplicação do método TOPSIS-2N.

A mensuração dos pesos é o principal passo na geração da matriz de preferências, construída, em regra, com a ajuda de especialistas. Atribuir um peso para determinado critério é definir o nível de importância em relação a outros critérios.

Uma forma de obter um resultado aproximado das prioridades dos critérios, segundo Saaty [120], é simplificar e organizar os julgamentos dos decisores seguindo uma escala linear, embora as respostas ainda sejam aproximadas. A fim de diminuir essa subjetividade,

a Escala de Importância do AHP adota cinco valores absolutos (1, 3, 5, 7 e 9) e quatro intermediários (2, 4, 6 e 8).

A técnica AHP não se preocupa com a precisão da medição desses pesos, mas sim com as proporções entre eles. Dessa forma, os valores numéricos da Tabela 3.2 correspondem às opções de julgamento apresentadas nos encontros realizados entre esta pesquisadora e os atores diretamente envolvidos no Plano de Dados Abertos do MME, entre eles a Autoridade de Monitoramento da LAI do MME, que por meio do processo de hierarquia analítica e da técnica de *brainstorming*, definiram o grau de importância e a intensidade de preferências para cada critério avaliado.

Os julgamentos iniciais foram organizados na matriz representada na Tabela 3.3. A partir dessa matriz resultará o vetor de prioridades (pesos dos critérios), necessário para indicar o grau de contribuição dos critérios em relação ao objetivo mapeado na etapa de Estruturação do Problema Decisório (3.2.3).

Tabela 3.3: Matriz de Julgamentos.

Fonte: Elaboração própria.

	Formatação	Padronização	Atualização	Rejeição	Aceitação
Formatação	1	3	5	1	2
Padronização	0,33	1	2	0,5	0,5
Atualização	0,2	0,5	1	0,5	0,5
Rejeição	1	2	2	1	2
Aceitação	0,5	2	2	0,5	1

Para a construção da matriz representada na Tabela 3.3, emprega-se o cálculo de julgamento da importância dos critérios, comparando-os a todos os respectivos pares, utilizando a seguinte fórmula:

$$c_{ji} = \frac{1}{c_{ij}}, c_{ii} = 1 \quad (3.1)$$

Cálculo do Vetor de Prioridades

É possível encontrar na literatura diferentes algoritmos utilizados no cálculo do vetor de prioridades, a exemplo da média geométrica, média aritmética, cálculo dos mínimos quadrados, potenciação de matrizes, entre outros [120, 121]. No entanto, quando se faz necessário lidar com modelos da ordem $n \geq 3$, sendo n o número de critérios/alternativas, e considerando o julgamento de vários decisores para se chegar a um consenso e satisfazer a propriedade recíproca, Reis e Schramm [121] recomendam utilizar o cálculo da matriz limite (potenciação de matrizes), explicitado por Saaty [120], com vistas a encontrar uma condição altamente consistente e adequada ao modelo.

Para obter o Vetor de Prioridades (pesos dos critérios), somam-se as linhas da Matriz de Julgamentos normalizada, dividindo esse resultado pela soma total da matriz de potenciação. Buscando evitar inconsistências nos julgamentos dos critérios, Saaty [120] propôs o uso do cálculo de λ_{max} , que pode ser determinado pelo produto da Matriz de Julgamentos e pelo Vetor de Prioridades, seguido da divisão desse novo vetor (Soma Ponderada) pelo Vetor de Prioridades. O λ_{max} resulta da média desse resultado, que é 5,1254.

Tabela 3.4: Cálculo do Autovalor Máximo/*Eigen* Principal (λ_{max})

Fonte: Adaptado de Reis e Schramm [121].

	Matriz de Julgamentos					Vetor de prioridades	Vetor (soma ponderada)	Vetor de prioridades	Resultado
	CR1	CR2	CR3	CR4	CR5				
CR1	1	3	5	1	2	0,3445	1,7659	0,3445	5,1254
CR2	0,33	1	2	0,50	0,50	0,1235	0,6329	0,1235	5,1254
CR3	0,20	0,50	1	0,50	0,50	× 0,0858	= 0,4395	÷ 0,0858	= 5,1254
CR4	1	2	2	1	2	0,2703	1,3852	0,2703	5,1254
CR5	0,50	2	2	0,50	1	0,1760	0,9018	0,1760	5,1254

Os demais conjuntos de matrizes resultantes dessa atividade, envolvendo a ponderação dos subcritérios em relação aos Critérios Formatação, Padronização e Rejeição, estão dispostos, respectivamente, nas Tabelas F.1, F.2 e F.3 do Apêndice F desta dissertação.

Verificação da Consistência

A consistência ideal condiz com a igualdade de valores entre o autovalor máximo (λ_{max}) e o número inteiro que representa o total de critérios da matriz, ou seja $\lambda_{max} = n$. No entanto, se $\lambda \geq n$. Dessa forma, Saaty [120] recomenda utilizar o cálculo do Índice de Consistência (IC), que visa diminuir a subjetividade inerente dos julgamentos orientados pela Escala de Importância apresentada na Tabela 3.2, e determinada pela seguinte equação:

$$IC = (\lambda_{max} - n)/(n - 1) \quad (3.2)$$

Se λ é o autovalor de uma matriz de decisão, λ_{max} corresponde ao autovalor máximo (*Eigen* Principal) e n à ordem dessa matriz (total de critérios considerados). Conhecido o valor de $\lambda_{max} = 5,1254$, obtém-se $IC = 0,03135$.

O método AHP fornece um limite de tolerância para inconsistência dos julgamentos. Para verificar a consistência do cálculo das prioridades, deve-se respeitar o limite da Razão de Consistência (RC), que deve ser menor ou igual a 10%. Este valor resulta do cálculo da razão entre o IC e o Índice Randômico (IR). A Tabela 3.5 fornece o valor de IR para matrizes de diferentes ordens, no caso, atribuiu-se o valor do $IR = 1,12$.

Tabela 3.5: Índice de Consistência Randômico
Fonte: Saaty [120]

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
IR	0,00	0,00	0,58	0,90	1,12	1,24	1,32	1,41	1,45	1,49	1,51	1,48	1,56	1,57	1,59

$$\frac{IC}{IR} = RC \leq 0,1 \quad (3.3)$$

O resultado desse cálculo é a ponderação do grau de dominância (prioridades) entre os critérios e subcritérios avaliados, de forma a influenciar a tomada de decisão balizada pelos pesos atribuído a cada um dos critérios, conforme se verifica na Tabela 3.3.

As comparações binárias (entre os pares de todos os critérios) resultaram em uma $RC = 2\%$, não sendo necessária a revisão dos julgamentos, tendo em vista que a consistência está adequada à aplicação do TOPSIS-2N na etapa seguinte.

3.3.2 Aplicação do TOPSIS-2N para *Ranking* das Bases

As soluções ideais positivas (A^+) e ideais negativas (A^-) são definidas de acordo com a matriz de decisão ponderada na etapa anterior (3.3.1). Duas variáveis são utilizadas para avaliar as performances dos *datasets*, são elas:

- Custo e/ou risco negativo (que objetiva ser mínimo, ou seja, quanto menor mais próximo do não-ideal) e
- Benefício e/ou oportunidade (objetiva maximizar o seu valor, ou seja, quanto maior, mais próximo do ideal).

Cálculo de Desempenho dos Critérios

O cálculo que determina a solução ideal e a ideal-negativa é dado por:

$$A^+ = \{v_1^+, \dots, v_n^+\} = \{(max_i v_{ij} \mid j \in C_b) \cdot (min_i v_{ij} \mid j \in C_c)\} \quad (3.4)$$

$$A^- = \{v_1^-, \dots, v_n^-\} = \{(min_i v_{ij} \mid j \in C_b) \cdot (max_i v_{ij} \mid j \in C_c)\} \quad (3.5)$$

Verifica-se que C_b denota o conjunto de critérios de oportunidades (benefícios) e C_c os critérios de risco (custo).

O Quadro 11 apresenta o relacionamento entre os critérios de riscos, subcritérios e funções objetivo correspondentes. Dessa forma, os principais *drivers* para aplicação do TOPSIS são: maximizar a Formatação (CR1), maximizar a Padronização (CR2),

minimizar a Atualização (CR3), minimizar a Rejeição (CR4) e maximizar a Aceitação (CR5).

Quadro 11: Critérios de Risco, Subcritérios e Função Objetivo.

Fonte: Elaboração própria.

Critério/Dimensão	ID Subcritérios	Função Objetivo⁶
CR1. Formatação	C ₁₁ , C ₁₂ , C ₁₃ e C ₁₄	Maximizar (C _b)
CR2. Padronização	C ₂₁ , C ₂₂ , C ₂₃ , C ₂₄ e C ₂₅	Maximizar (C _b)
CR3. Atualização	Não aplicados	Minimizar (C _c)
CR4. Rejeição	C ₄₁ , C ₄₂ e C ₄₃	Minimizar (C _c)
CR5. Aceitação	Não aplicados	Maximizar (C _b)

Verifica-se que as *funções objetivo* de benefício (C_b) devem ser maximizadas e, conseqüentemente, as de custo (C_c) minimizadas. Ou seja, para obter o *valor ótimo*, mais próximo possível da *solução ideal* do modelo, três critérios (CR1, CR2 e CR5) devem derivar o maior valor à função objetivo equivalente, e dois critérios (CR3 e CR4) o menor valor [71].

Calcular a medida de separação das alternativas

A base do TOPSIS é o cálculo da medida de separação da distância de cada alternativa em relação à solução ideal (S_i⁺) e não-ideal (S_i⁻). Para encontrar essa medida é realizado o cálculo da distância euclidiana (valor de separação) das alternativas, expresso por:

$$S_i^+ = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^+)^2}, \forall i \quad (3.6)$$

$$S_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^-)^2}, \forall i \quad (3.7)$$

Calcular a proximidade relativa

Por fim, a localização de cada alternativa é calculada pela sua proximidade relativa em relação ao vetor hipotético criado da solução positiva (S_i⁺). Para tanto, calcula-se a pontuação geral de preferência obtida pela i-ésima alternativa em relação ao vetor ideal

(Rc_i^+), que assume valores entre 0 e 1. Quanto maior o valor de Rc_i^+ , mais próximo está da solução ideal.

$$Rc_i^+ = \frac{S_i^-}{S_i^+ + S_i^-}, \forall i \quad (3.8)$$

Dessa forma, o algoritmo seleciona os melhores desempenhos de cada alternativa, criando posições em vetores mais altos para aqueles com maior eficiência (ou seja, benefícios) e posições inferiores para aquelas alternativas que estão mais distantes da solução ideal solução ideal (A^+) e próximas da alternativas não-ideal (A^-). O cálculo da proximidade relativa entre determinada alternativa (A_i) em relação à solução ideal e não-ideal é determinado pela equação 3.9.

Ranking das Alternativas

As alternativas, os *datasets* avaliados pelo AHP-TOPSIS-2N, foram ranqueadas por ordem de preferência conforme os valores apresentados na Tabela 3.6. O *ranking* foi gerado em ordem decrescente do valor de C_i (coeficiente de proximidade).

$$C_i = \frac{S_i^-}{S_i^- + S_i^+} \quad (3.9)$$

Tabela 3.6: *Ranking* e Resultados da aplicação do AHP-TOPSIS-2N.

ID	S_{i+}	S_i^-	C_i^+	Ranking N_1	C_j^+	Ranking N_2
DB46	0.034974383	0.171522542	0.830630008	1	0.66312	1
DB45	0.127296229	0.071434232	0.359452856	4	0.567833952	2
DB9	0.168202206	0.054383165	0.244324972	11	0.53866	3
DB47	0.128691866	0.059665665	0.316768142	5	0.492841116	4
DB10	0.168489086	0.051651716	0.234630363	13	0.48411	5
DB37	0.149578861	0.045721465	0.234108492	14	0.481687214	6
DB1	0.135255493	0.057587227	0.298622769	6	0.46520	7
DB48	0.148970526	0.047519891	0.241843299	12	0.462225864	8
DB39	0.090269041	0.106128769	0.540376538	2	0.44339	9
DB31	0.171550706	0.044965169	0.207676082	16	0.440105058	10
DB29	0.135548898	0.048504475	0.263534833	8	0.41021	11
DB30	0.135552606	0.048500121	0.263512103	9	0.410127821	12
DB40	0.174760791	0.043166333	0.198076918	27	0.40769	13
DB27	0.174763668	0.043161441	0.198056302	28	0.407631506	14
DB2	0.137023052	0.053729432	0.281670943	7	0.40560	15
DB23	0.174778937	0.040980122	0.189934652	32	0.401549101	16

ID	S_{i+}	S_i^-	C_i^+	Ranking N_1	C_j^+	Ranking N_2
DB44	0.116636391	0.067950848	0.368123215	3	0.40090	17
DB15	0.155344449	0.040533443	0.206932199	17	0.400688011	18
DB8	0.155347207	0.040528603	0.206909691	18	0.40062	19
DB13	0.15537495	0.040481251	0.206688635	19	0.399963555	20
DB14	0.15537495	0.040481251	0.206688635	20	0.39996	21
DB24	0.155390156	0.040456295	0.206571497	21	0.399610077	22
DB28	0.155390156	0.040456295	0.206571497	22	0.39961	23
DB26	0.155393391	0.040451075	0.206546938	23	0.39953553	24
DB41	0.174833364	0.038834185	0.181750504	33	0.39537	25
DB18	0.155573913	0.038773412	0.199505765	24	0.393487501	26
DB19	0.155573913	0.038773412	0.199505765	25	0.39348	27
DB42	0.155573913	0.038773412	0.199505765	26	0.393487501	28
DB3	0.137561866	0.048977143	0.262557109	10	0.39196	29
DB17	0.155798022	0.037186748	0.192692655	29	0.387492513	30
DB25	0.155798022	0.037186748	0.192692655	30	0.38749	31
DB35	0.155801249	0.037181069	0.192665678	31	0.38741629	32
DB22	0.154932217	0.03245521	0.17319844	36	0.38707	33
DB20	0.156367014	0.03434915	0.180106128	34	0.376075934	34
DB21	0.156711457	0.033126986	0.174500937	35	0.37075	35
DB16	0.156545192	0.030665479	0.163801982	37	0.368980851	36
DB4	0.176023579	0.031949712	0.15362411	45	0.36365	37
DB6	0.176211441	0.025812839	0.127770972	46	0.355316377	38
DB12	0.158467538	0.029940009	0.15891088	38	0.35354	39
DB32	0.15848562	0.029899197	0.15871341	39	0.353100736	40
DB33	0.15848562	0.029899197	0.15871341	40	0.35310	41
DB36	0.15848562	0.029899197	0.15871341	41	0.353100736	42
DB5	0.176553095	0.024423784	0.121525341	47	0.35053	43
DB11	0.176553095	0.024423784	0.121525341	48	0.350539273	44
DB7	0.159007502	0.029620147	0.157029722	42	0.35024	45
DB34	0.159025522	0.029578893	0.156830333	43	0.349808569	46
DB38	0.159025522	0.029578893	0.156830333	44	0.34980	47
DB43	0.152314621	0.041626876	0.214636253	15	0.347138527	48

3.4 Validação do Modelo

Para Qureshi, Harrison e Wegener [123], três abordagens asseguram se um modelo MCDA foi desenvolvido de maneira aceitável e de acordo com a metodologia aplicada, são elas: Verificação, Análise de Sensibilidade e Validação. No caso da primeira, o modelo proposto nesta dissertação fez uso da Verificação da Consistência (3.3.1), seguindo os cálculos sugeridos por Saaty [120]. O resultado dessa verificação aponta para uma consistência alta ($RC = 2\%$), tendo em vista o limite de até 10% considerado para esse índice.

Nessa linha, Mukherjee [124] apresentou um levantamento bibliográfico sobre a estrutura de um modelo MCDA, concebido a partir dos métodos mais citados (AHP e TOPSIS) para destacar as seguintes abordagens de validação utilizadas nos trabalhos relacionados, a ver:

1. Validar do modelo por meio de comentários dos especialistas;
2. Análise de Sensibilidade alterando o julgamento dos critérios;
3. Perguntas, em grupo ou individual, direcionadas aos especialistas/decisores; e
4. Validação e Verificação, usadas juntamente com a análise de sensibilidade.

A Autoridade de Monitoramento da LAI e a Ouvidoria-Geral do MME foram consultadas para a realização dos itens 1 e 2 supra, com a finalidade de reforçar a validação do modelo adotado nesta pesquisa, bem como confirmar a utilidade dessa validação. Os citados especialistas formalizaram considerações acerca do impacto do modelo proposto, pelos quais demonstraram que o modelo é uma representação suficientemente adequada em relação ao propósito para o qual foi projetado (vide Anexos I e II).

3.4.1 Análise de Sensibilidade

A Análise de Sensibilidade, que pode ser aplicada de forma individual e em grupo, é utilizada para testar a estabilidade e a consistência de um modelo quando realizadas alterações em seus parâmetros [123, 37]. Conforme Triantaphyllou e Sánchez [125], essa Análise deve ser realizada para determinar como a priorização das bases de dados pode variar conforme a alteração dos pesos dos critérios. À medida que o peso de um dos critérios aumenta, o nível de importância dos critérios restantes diminui de forma proporcional. Dessa forma, as prioridades globais das alternativas são recalculadas.

Seguindo Máchová e Lněnička [54], a sensibilidade do modelo foi testada em relação às mudanças na importância dos critérios. A Figura 3.6 mostra que ocorreu uma leve alteração no *ranking* das bases após alteração nos pesos dos Critérios Rejeição (CR4) e

Aceitação (CR5). A escolha desses critérios se deve à relativa volatilidade dos seus valores, apresentados pelos usuários e consumidores das respectivas bases, implicando a inexorável alteração do *ranking*.

Para garantir a validação do modelo proposto nesta dissertação, constituído a partir das etapas estabelecidas por Belton e Stewart [37] e Marttunen et al. [66], optou-se por desenvolver uma plataforma computacional para compilação e comparação contínua das bases. O aplicativo, desenvolvido em linguagem R, pode ser acessado via *web* de forma prática e adaptável à volatilidade de dados, permitindo alterações nos parâmetros (critérios e alternativas), conforme mudanças no cenário e preferência do decisor.

Além do exposto, cabe destacar que, para reforçar a robustez do modelo, o método TOPSIS-2N permite gerar dois *rankings* com normalizações distintas (vide Tabela 3.6), o que possibilita realizar uma comparação das bases ordenadas.

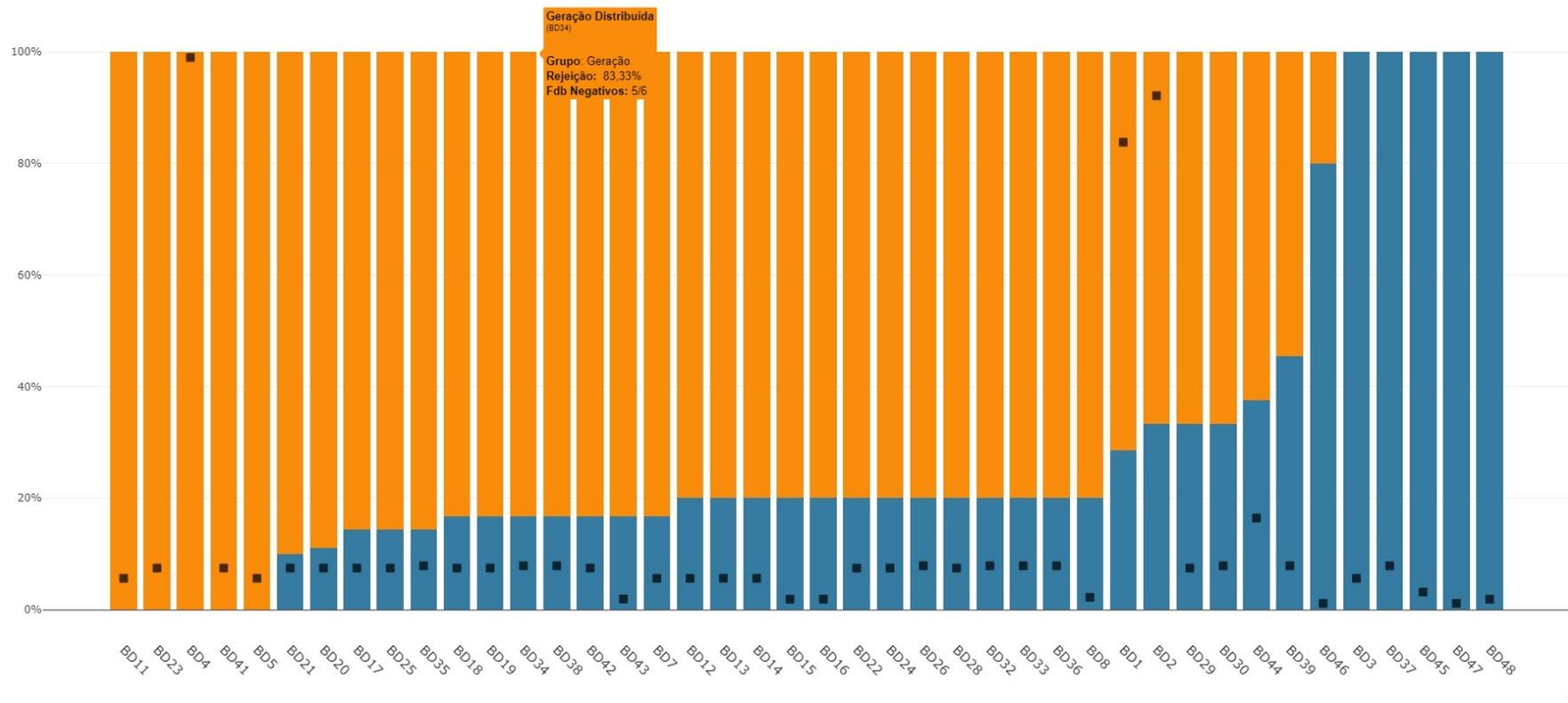


Figura 3.6: Resultado da Análise de Sensibilidade sobre os Critérios CR4 e CR5.
Fonte: Elaboração própria.

3.5 Considerações do Capítulo

A partir dos subsídios apresentados pela Revisão Sistemática da Literatura (identificação dos critérios e métodos afetos a este trabalho), foi possível executar o estudo de caso para a construção do modelo proposto nesta pesquisa, sustentado pelas etapas dos *frameworks* CRISP-DM e MCDA. A utilização do CRISP-DM decorreu frente ao desafio de encontrar as bases de dados relevantes para a sociedade, considerando eixos estratégicos do setor elétrico. Inicialmente, na Seção 3.1, foram apresentadas as atividades necessárias para o projeto de mineração que revelou, na Subseção 3.1.3, temas potenciais para a abertura de dados, relevantes para verificar a adequação entre a oferta e a demanda de dados abertos relacionados a esses temas.

A partir da consecução das etapas do processo MCDA (dissertada na Seção 3.2), que envolveu a Estruturação do Problema (Subseção 1.2), a Definição dos Critérios de Riscos (Subseção 3.2.2) e a Decomposição dos elementos basilares para ordenação dos *datasets* avaliados, foi possível realizar a aplicação dos métodos multicritério AHP e TOPSIS (Seção 3.3). Completando a Fase do Estudo de Caso, foi apresentado, na Seção 3.4, o resultado da Análise de Sensibilidade e dos comentários de gestores/especialistas do setor, que pautaram a Validação do Modelo.

Capítulo 4

Resultados e Análises

4.1 *Dashboards* das Bases e dos Critérios via *AppShiny*

De Souza, Gomes e De Barros [126] desenvolveram *scripts*¹ do R para aplicação do método híbrido AHP-TOPSIS-2N, desenvolvendo um modelo decisório mais fácil e dinâmico para quem está familiarizados com esse tipo de linguagem. Por outro lado, pode ser muito oneroso e difícil para usuários que desconhecem códigos e plataformas computacionais, sobretudo, se o processo de avaliação e/ou a abertura de dados demandar uma análise repetitiva para diferentes critérios e cenários ou que dependam de grande quantidade de alternativas.

Frente à questão supra, foi implementado um aplicativo *Web*, de interface amigável (*user-friendly*), utilizando o pacote *Shiny* do R (*AppShiny*), que fornece a visualização dos resultados, via *dashboards online*, de forma interativa e dinâmica a cada mudança nos parâmetros do Modelo. Esse aplicativo pode ser acessado pelo seguinte endereço eletrônico: <https://ingridpalma.shinyapps.io/ahptosis/>.

A Figura 4.1 apresenta a tela da página inicial do *AppShiny* e a Figura 4.2 os *dashboards* projetados para análise dos resultados, permitindo, ainda, selecionar apenas as bases de interesse para o gestor/decisor, em um exame mais pontual.

O *script* adaptado, o conjunto de dados, entre outros descritores e algoritmos utilizados para o desenvolvimento deste trabalho, estão disponíveis no repositório: <https://github.com/ingridpalma/mestrado>.

¹http://cran.rstudio.org/src/contrib/Archive/ahptopsis2n/ahptopsis2n_0.1.0.tar.gz

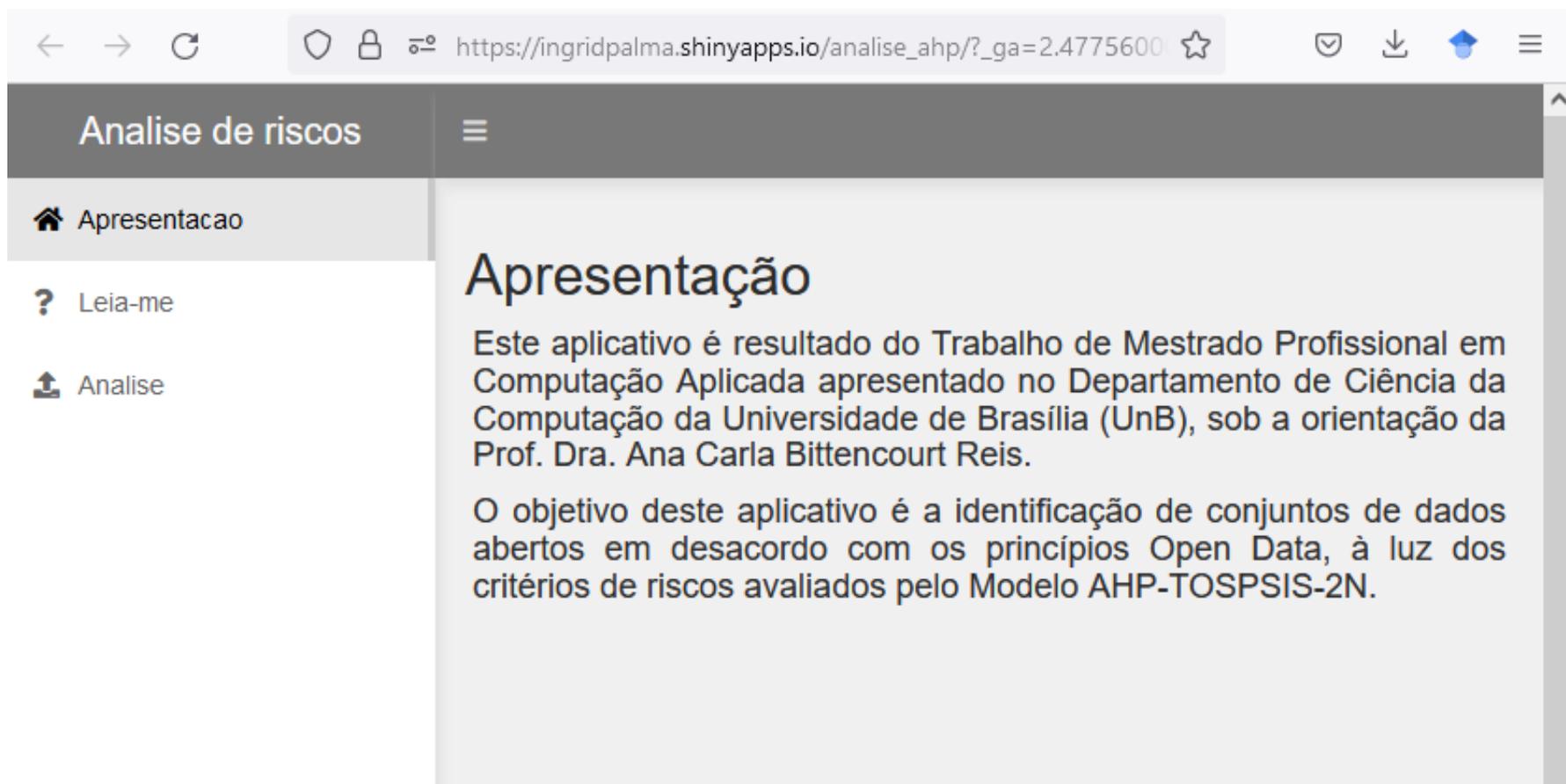
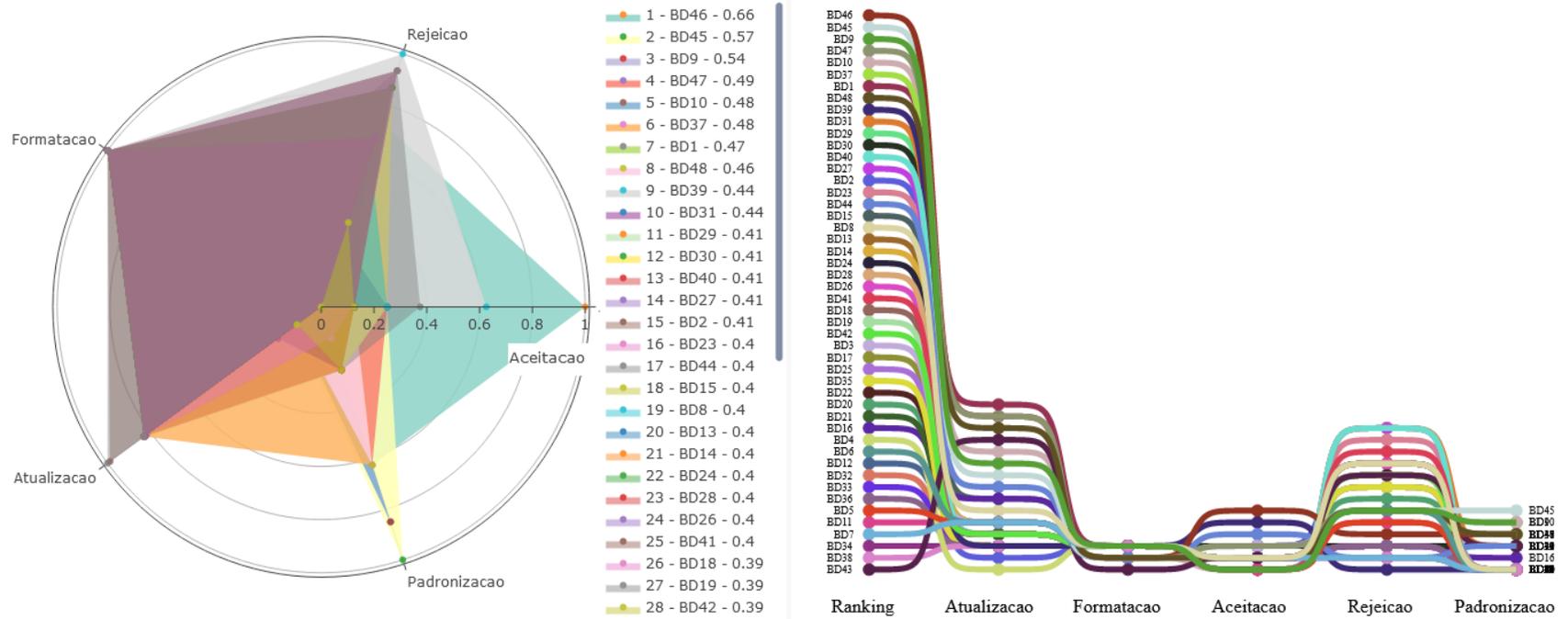


Figura 4.1: Tela da Página inicial do *AppShiny*.
Fonte: Elaboração própria.

Modelo de Apoio à Decisão



"Openess" das bases:

★★★★★ : 0 ★★★★★ : 13 ★★★★★★ : 31 ★★★★★★ : 1 ★★★★★★ : 0

Figura 4.2: Dashboards projetados pelo AppShiny.

Fonte: Elaboração própria.

Em uma versão adaptada e inspirada no Sistema de 5 estrelas² [20], a tela inicial do *AppShiny* (Figura 4.1) classifica as bases conforme o seu nível de abertura (*openness*) e à luz do padrão *Open Data* e do movimento *OGD*.

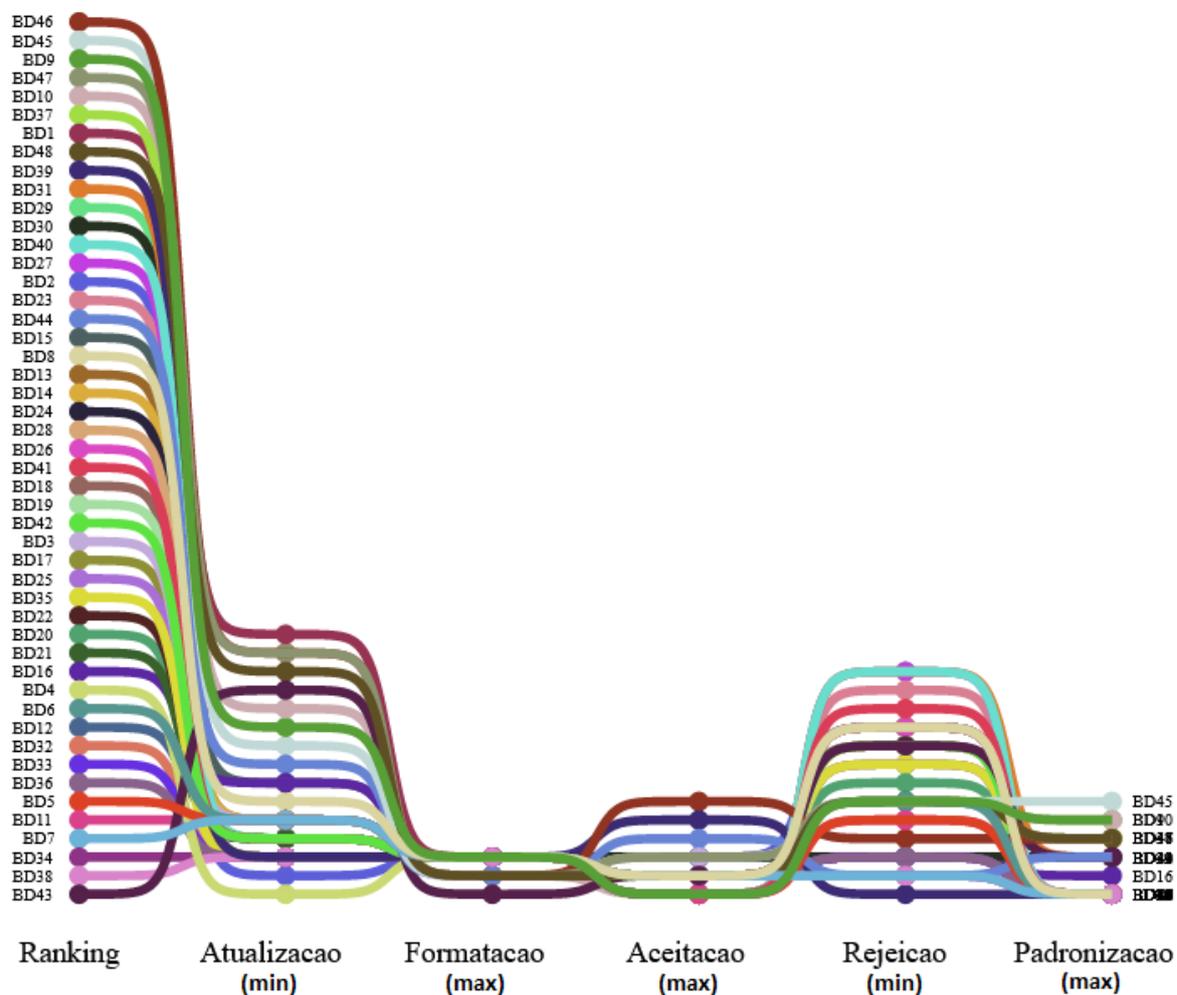


Figura 4.3: *Ranking* das 48 Bases.
Fonte: Elaboração própria.

²<https://5stardata.info/pt-BR/>

O gráfico de *Bump*, conhecido também como gráfico de *ranking* ou de classificação, é uma evolução do gráfico de linhas usado para comparar dimensões entre si, sendo útil para explorar alterações na posição e no desempenho de cada conjunto. A partir disso, e visando fornecer uma visualização mais robusta, decidiu-se utilizar esse tipo de gráfico para projetar o *ranking* das 48 bases analisadas, conforme mostra a Figura 4.3. Destacam-se, em consonância com a Tabela 3.3, os critérios de maior impacto para as Bases analisadas: Formatação, seguido do Critério Rejeição.

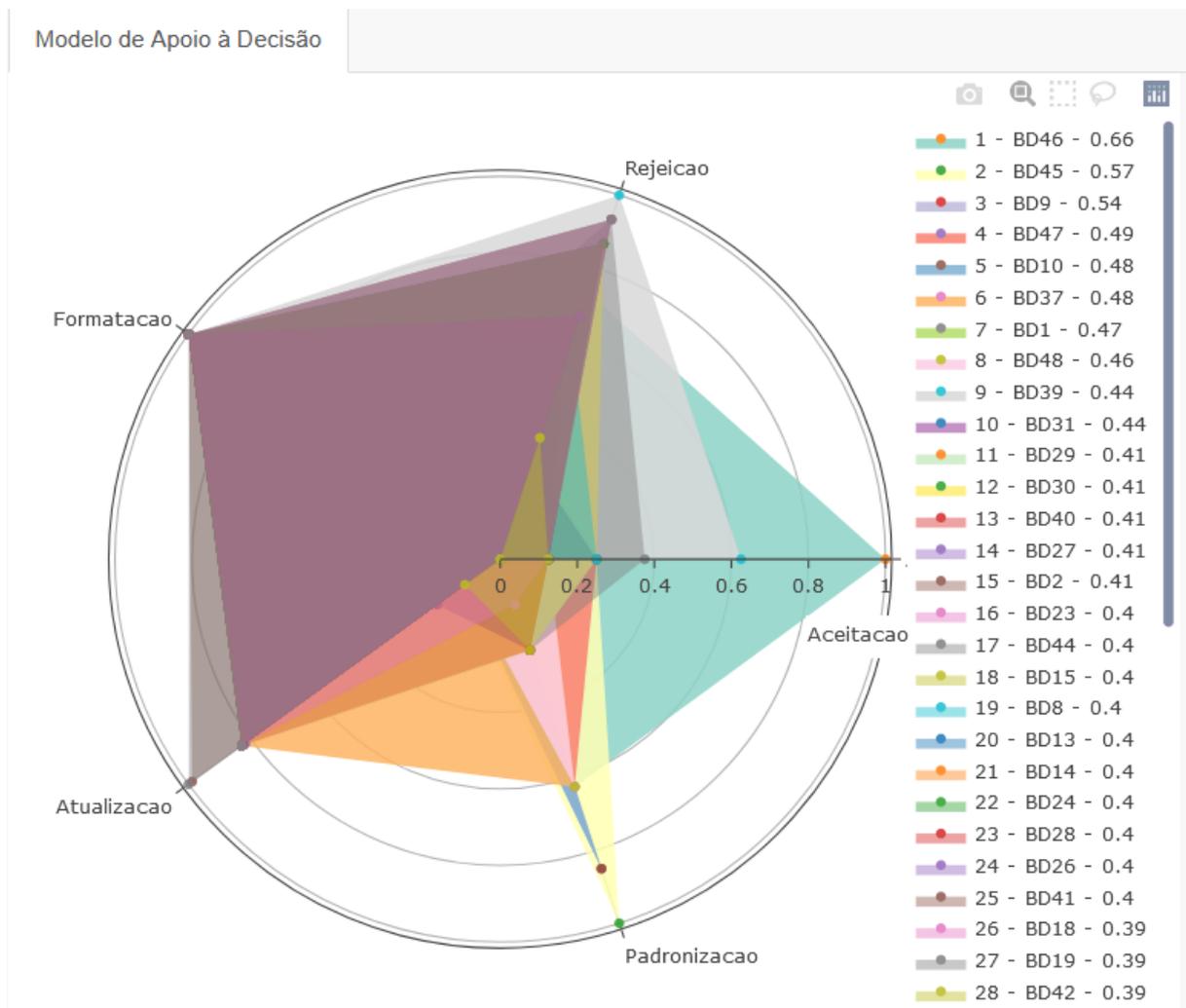


Figura 4.4: Comparativo das Bases e Impacto dos Critérios de Risco.

Fonte: Elaboração própria.

O gráfico de radar da Figura 4.4 apresenta uma outra forma de visualizar e analisar o mesmo *ranking* projetado na Figura 4.3, sendo possível avaliar os *datasets* em relação ao desempenho médio das alternativas.

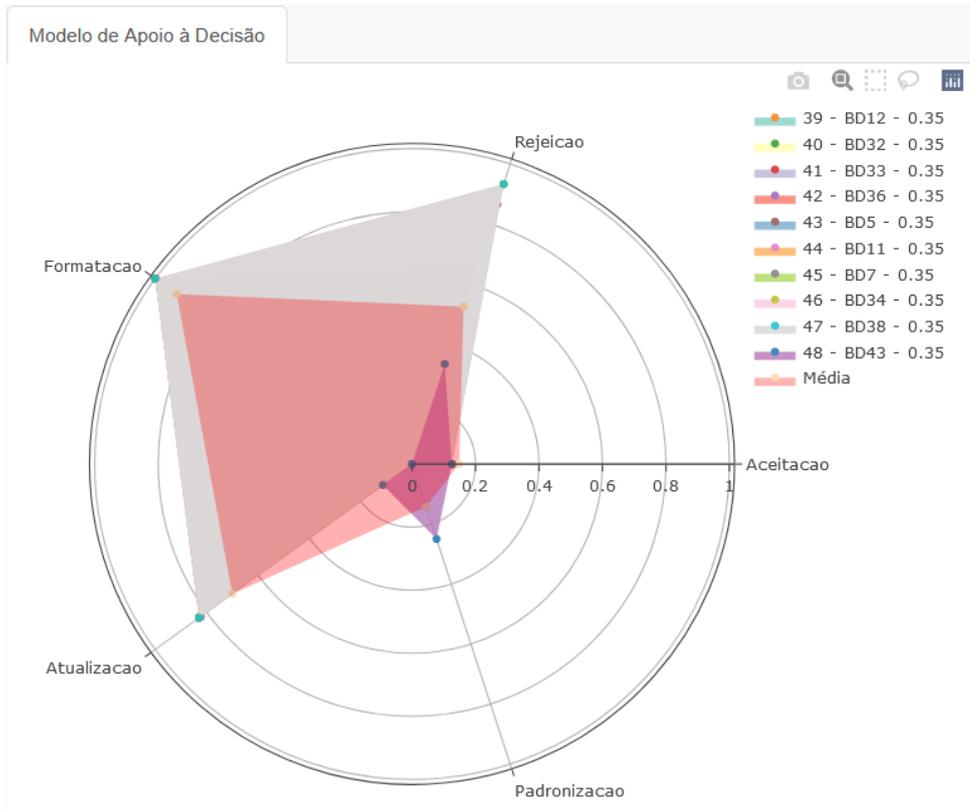


Figura 4.5: Média de desempenho das bases.
Fonte: Elaboração própria.

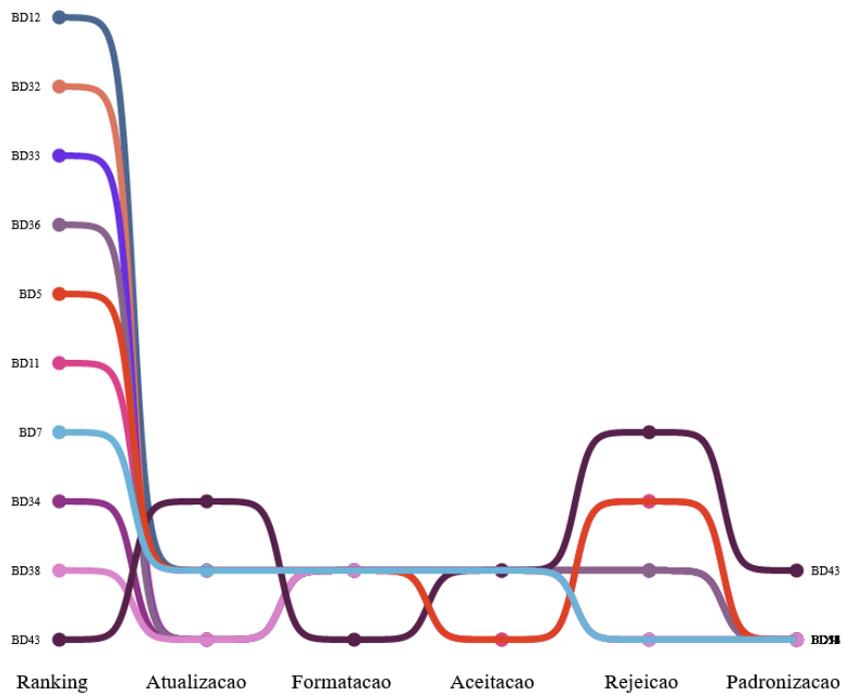


Figura 4.6: Bases com desempenho abaixo da média.
Fonte: Elaboração própria.

Os gráficos das Figuras 4.5 e 4.6 representam, respectivamente, a média do nível de abertura das bases e o desempenho daquelas mais distantes do padrão *Open Data*, classificadas como crítica (inferiores à média calculada). Essas bases podem representar custos não previstos durante o processo de abertura ou de gerenciamento.

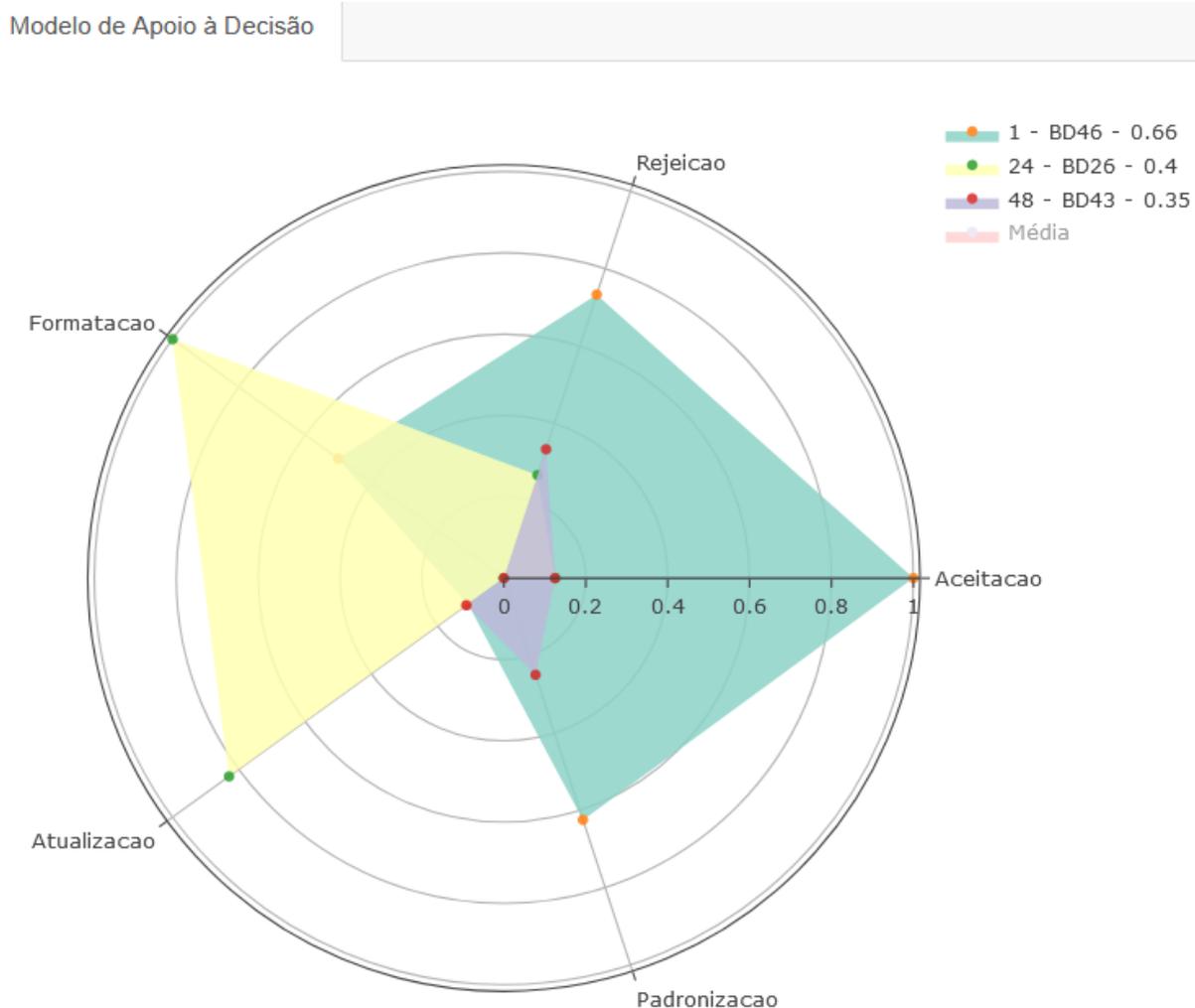


Figura 4.7: Resultado das Bases DB46, DB26 e DB43.

Fonte: Elaboração própria.

Os resultados foram organizados via *dashboards* interativos, em uma plataforma *Web*³ desenvolvida em R Shiny, de modo que o conhecimento adquirido possa efetivamente ser entendido e compartilhado entre as partes interessadas.

Recorremos ao *radar chart* para uma análise mais detalhada, e comparativa, conforme plotado na Figura 4.7. Para isso, foram selecionadas as seguintes bases: BD46, (Sistema de Informações de Geração da ANEEL), BD26 (Os Dez Maiores Agentes por unidade de consumidores) e BD43 (Sistemas de Informações Energética do Brasil).

³<https://ingridpalma.shinyapps.io/ahptosis/>

Foram elegidas as bases localizadas no extremos do vetor virtual do TOPSIS, referente à primeira e última posições do *ranking*, considerando também a base DB26, com valor intermediário entre aquelas duas, e mais próxima da média, permitindo verificar o critério de maior impacto em cada uma dessas bases.

Capítulo 5

Considerações Finais

Esta dissertação versa acerca da proposição de um modelo de apoio à decisão direcionado para a avaliação de dados do setor elétrico catalogados no Portal Brasileiro de Dados Abertos, pelo Ministério de Minas e Energia e pelas respectivas unidades vinculadas, considerando o nível de abertura desses dados em relação aos Princípios *Open Data* e *Open Government Data* (OGD).

Para o alcance do objetivo geral em tela foi necessário identificar, selecionar e mensurar (definir e ponderar) critérios de riscos relevantes para o contexto da OGD. Para isso, foi realizada uma Revisão Sistemática da Literatura, que também serviu para a identificação e definição dos métodos multicritério atinentes: AHP e TOPSIS. A lista dos critérios e subcritérios especificados está disposta do Apêndice E deste trabalho. Diante disso, foram considerados cumpridos os dois primeiros objetivos específicos desta pesquisa.

A baixa quantidade de publicação de dados abertos pelo MME remeteu ao problema de identificação de tópicos potenciais para a abertura de dados, que trata do terceiro objetivo específico desta dissertação, abordado na Subseção 3.1.3. Além disso, a "*baixa relevância para a sociedade ou stakeholders*" (CR42) consubstanciou um critério de risco essencial no âmbito da Política de Dados Abertos do MME. Diante disso, se fez necessária a utilização de técnicas de mineração de dados, guiadas pelo modelo CRISP-DM e implementadas no *Software* RStudio, que permitiram mapear estatisticamente os principais planos energéticos, produzidos pelos agentes desse setor e, no âmbito do MME, os temas e assuntos mais demandados pela sociedade e demais partes interessadas.

Com os temas mapeados, foi possível identificar o grupo de tópicos relacionados às bases correspondentes catalogadas no PBDA, bem como mensurar os critérios de riscos pertinentes, a fim de refletir a expectativa dos usuários frente à oferta dos dados relacionados aos temas revelados. O implemento desse objetivo específico, inclusive, resultou na produção do artigo [10], publicado na base ACM Digital Library (vide Anexo III).

Superados os mencionados objetivos específicos, foi possível, enfim, concluir a Construção e a Validação do modelo proposto, por meio da aplicação dos métodos AHP e TOPSIS-2N e das abordagens de Análise de Sensibilidade, que consideraram os comentários dos gestores acerca da proposta apresentada neste trabalho.

Na ocasião, foi formalizado, por meio dos Ofícios referenciados nos Anexos I e II) desta dissertação, que o modelo em lide tem potencial para atender às atuais necessidades do MME, entre outras considerações acerca do impacto do modelo proposto. A esse respeito, releva destacar o contido no documento elaborado por um dos especialistas consultados, que expressou a intenção de encaminhar esta pesquisa para apreciação do Comitê de Governança Digital daquele Ministério, *com vistas à avaliação da incorporação da solução, no que for aplicável, por ocasião do processo de revisão e elaboração do novo PDA do MME* (grifo nosso).

Outro especialista, diretamente relacionado a esta pesquisa expressou a relevância deste estudo, **tendo em vista que a abertura de dados abertos minimiza as atividades do seu setor e os custos de processamento e gerenciamento de pedidos individuais.**

A versão *web* do modelo, oferecido pelo pacote Shiny do R, facilita a sua compreensão fornecendo *dashboards* das bases avaliadas e ordenadas conforme o desempenho apresentado em relação aos critérios, permitindo realizar *insights* (percepções) para análise e avaliação a ser realizada pelos gestores responsáveis pelas bases. É importante formalizar o modelo para que este se faça compreensível e replicável pelos agentes responsáveis por customizar, produzir e/ou ou gerenciar dados governamentais abertos do setor elétrico.

Com os objetivos geral e específicos atingidos, foi possível direcionar o modelo ao encontro dos aspectos técnicos e legais exigidos. Por oportuno, vislumbre-se que esta pesquisa pode ser considerada um importante instrumento metodológico e de inovação, especificamente, no setor público, fomentando mudanças positivas na cultura de transparência e nas demais ações afetas ao aprimoramento do Plano de Dados Abertos do MME.

Por fim, entende-se que não adianta bons índices e aspectos técnicos contemplados em modelos informatizados pensando apenas na eficiência para cumprir os compromissos assumidos pelo Governo Aberto/Digital. Os problemas da abertura de dados requerem uma habilidade técnica profunda do tema. A *Open Energy Data*, ou melhor, a *Open Government Data* voltada para o setor elétrico, deve ser realizada com o aporte de uma estrutura mínima de Governança de Dados, responsável não apenas por fomentar a cultura de dados com ética e responsabilidade, mas por traçar estratégias adequadas para os riscos inerentes às inovações tecnológicas e de políticas de dados abertos.

5.1 Recomendações e Contribuições

O processo de tomada de decisão na abertura de dados pode ser demorado e exigir muitos recursos tecnológicos, financeiros e de pessoal. Para entender melhor as consequências de cada resultado possível, os especialistas/decisores (no caso, gestores e publicadores de dados) precisam simplificar o processo de abertura e gerenciamento de dados.

O aplicativo *AppShiny*, implementado sobre o método híbrido AHP-TOPSIS-2N, apresenta uma interface amigável do Modelo de Apoio à Decisão proposto neste trabalho, permitindo extrair diferentes análises a partir da configuração efetuada ao aplicativo, tais como análises de custo-benefício (riscos-oportunidades) e da maturidade do nível de abertura e gestão dos dados.

O modelo possibilita, ainda, avaliar a similaridades dos *datasets* em relação aos critérios de risco de maior impacto. Tais critérios foram identificados e definidos a partir do levantamento bibliográfico dos princípios da família *Open Data*. Para a estruturação e execução da metodologia, foram selecionados dados abertos do setor elétrico catalogados no PBDA, disponibilizados pelo MME e ANEEL.

O modelo proposto, considerando as técnicas aplicadas, pode ser de simples interpretação e baixo custo, contribuindo, sobretudo, para a diminuição da subjetividade das decisões que envolvem o processo de abertura de dados e seus respectivos benefícios (oportunidades) e custos (riscos negativos). Em termos técnicos, a implementação do modelo, no âmbito governamental, se justifica, também, pelas seguintes razões:

- a) Possibilita avaliar um grande número de alternativas simultaneamente;
- b) Critérios e alternativas são facilmente estruturados em um modelo interativo e dinâmico;
- c) Baixa complexidade computacional e do esforço humano exigido, particularmente no que tange à construção das matrizes para comparação binária;
- d) Os critérios podem ser altamente correlacionados entre si. Adicionalmente, por meio da abordagem de *trade-off*, o modelo afasta a possibilidade de não considerar critérios com menor desempenho [127, 37]; e
- e) Integra de forma eficaz os critérios de riscos às preferências dos decisores.

Belton e Stewart [37] dizem que, após a análise, deve-se aplicar os resultados, isto é, traduzir a análise em planos de ação específicos. Conforme corroborado por Mariscal, Marban e Fernandez [29], a implementação do Plano de Ação do modelo dependerá do gestor do órgão, que realizará as etapas metodológicas em adequação ao respectivo *modus*

operandi e aos recursos disponíveis. Dessa forma, se faz necessário que o gestor entenda quais ações precisam ser realizadas para realmente fazer uso do modelo proposto, o qual apresenta uma metodologia que pode ser aplicada ao contexto governamental.

Especificamente ao MME, e a partir deste piloto e da metodologia apresentada, pode-se confrontar a análise acerca dos critérios de riscos identificados (Subseção 3.2.2) com as informações disponíveis no vigente Plano de Dados Abertos do MME, para propor a adoção dos critérios que possam aprimorar e melhor definir a Política de Dados Abertos daquele Ministério. Essa formulação poderá ocorrer quando da revisão geral das bases de dados catalogadas no Portal Brasileiro de Dados Abertos.

Apenas as etapas iniciais (destacadas na Figura 5.1) do Processo de Gerenciamento de Riscos da ISO 31000:2018, foram necessárias para a identificação e seleção dos Critérios de Risco. No entanto, ações de contingência e implementações de controles podem resultar de uma avaliação mais completa dos critérios de risco definidos neste trabalho. Para isso, sugere-se utilizar o mesmo processo de gerenciamento de riscos, apoiado às técnicas apresentadas pela ABNT NBR ISO 31000:2018 [84].

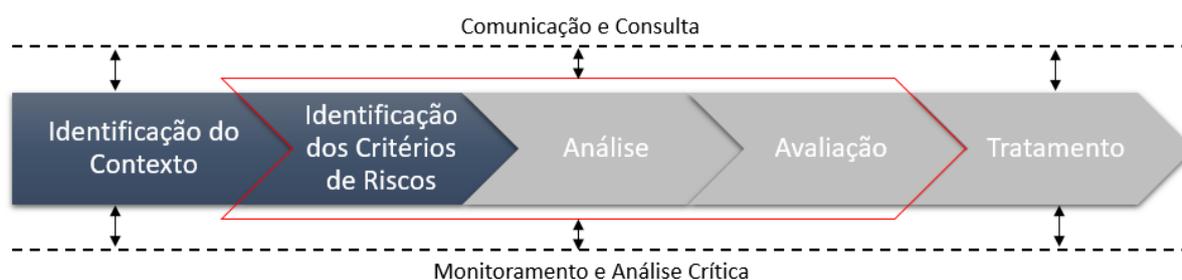


Figura 5.1: Processo de Gerenciamento de Riscos.
Fonte: Adaptada da ABNT NBR ISO 31000:2018 [84].

Conforme a ABNT NBR ISO 31000:2018 [84], o “*propósito da identificação de riscos é encontrar, reconhecer e descrever riscos que possam ajudar ou impedir que uma organização alcance seus objetivos*”.

Para a ponderação dos critérios, recomenda-se utilizar o cálculo da matriz limite, sobretudo em modelos MCDA da ordem de $n > 3$ [120, 121]. Assim, um *Script* em R foi desenvolvido e disponibilizado no Apêndice G desta dissertação, no sentido de garantir a mais alta consistência do modelo.

Por meio desta pesquisa, órgãos e entidades alcançados pela Política de Dados Abertos, ou seja, aqueles obrigados a publicar e atualizar dados publicados no PBDA, podem, conhecido o contexto e a metodologia, considerar outros critérios além daqueles definidos pela INDA por meio da Resolução nº 3/2017¹.

¹<https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes/resolucao-n-o-3-de-13-de-outubro-de-2017>

5.1.1 Artigos Publicados e Aceitos

Machine Learning Predictive Model for the Passive Transparency at the Brazilian Ministry of Mines and Energy - Anexo III

Os assuntos mais relevantes para a sociedade, no bojo do MME, foram mapeados por Palma A., Ladeira e Reis [10], publicado na base *Association for Computing Machinery* (ACM), em julho de 2021. Nesse trabalho, para gerar o modelo de tópicos mais frequentes, foram utilizados algoritmos de Processamento de Linguagem Natural (NLP) sobre os requerimentos de informações direcionados àquele Ministério.

A estabilidade do modelo foi testada a partir da análise comparativa entre cinco conhecidos algoritmos de classificação, entre eles o XGBoost, que apresentou melhor desempenho e precisão nos resultados com multiclases. Os resultados constataram que a frequência de palavras como "energia" e "ANEEL" estiveram mais próximas dos tópicos mapeados, o que infere interesse nos assuntos de regulação/fiscalização de energia. Outras observações podem ser feitas com o citado trabalho, por exemplo: as palavras "horário" e "valor" sugeriram a preocupação da sociedade/*stakeholders*, no biênio 2018-2019, em relação às políticas tarifárias de energia elétrica (bandeiras tarifárias) e ao horário de verão.

Design and Evaluation of IoT Gateway for Data Prioritization based on Van Emde Boas Tree - Anexo IV

A proposta deste artigo é, por meio de análise das variações estruturais de Dados Heap (Fibonacci e Binominal), realizar a priorização de dados oriundos de dispositivos IoT, considerados críticos em sistemas que exigem uma ação imediata ou em tempo ideal.

Design and Application of the AHP-TOPSIS-2N to Evaluate (Linked) Open Government Data from the Electricity Datasets - Anexo V

Esse artigo apresenta resultados intermediários desta dissertação e uma versão simplificada do Modelo de Apoio à Decisão, o qual envolveu, inicialmente, 4 critérios de riscos e 23 bases de dados do setor elétrico.

5.2 Sugestões para Trabalhos Futuros

Para trabalhos futuros, poder-se-á desenvolver um sistema de pontuação para indicar o nível de abertura de cada conjunto de dados, por meio de um índice automatizado e apresentado em tempo real de atualização, a partir dos respectivos metadados avaliados.

A aplicação do método híbrido AHP-TOPSIS-2N poderá ser ampliada com a inclusão de novas métricas e critérios/dimensões embasados por meio de uma revisão mais completa e atualizada da literatura, especificamente na parametrização e estudos acerca da alfabetização em dados abertos (*data literacy*), *Web Semântica* e *Ontologia*. Quanto ao baixo índice de pessoas letradas em dados abertos, os vieses são diversos, o que se sugere considerar, em termos de perspectivas futuras, as seguintes proposições iniciais:

- a) Abrir dados que podem efetivamente gerar valor para as partes interessadas; e
- b) Identificar e criar condições para dar vazão aos dados abertos, conforme os padrões e as boas práticas estabelecidas por tecnologias da *web* semântica.

Bibliografia

- [1] David Eaves. *The three laws of open government data*. 30 de set. de 2009. URL: <https://eaves.ca/2009/09/30/three-law-of-open-government-data/> (acesso em 16/02/2021).
- [2] Catherine D’Ignazio e Rahul Bhargava. “DataBasic: Design principles, tools and activities for data literacy learners”. Inglês. Em: *The Journal of Community Informatics* 12.3 (2016).
- [3] Mariel Garcia Montes e Dirk Slater. “Data literacy”. Em: *The State of Open Data* (2019), p. 274. URL: <https://www.stateofopendata.od4d.net/chapters/issues/data-literacy.html#fn:9>.
- [4] Michael B. Gurstein. “Open data: Empowering the empowered or effective data use for everyone?” Em: *First Monday* 16.2 (2011). DOI: <https://doi.org/10.5210/fm.v16i2.3316>. URL: <https://firstmonday.org/ojs/index.php/fm/article/view/3316> (acesso em 16/02/2021).
- [5] Marijn Janssen, Yannis Charalabidis e Anneke Zuiderwijk. “Benefits, adoption barriers and myths of open data and open government”. Inglês. Em: *Information systems management* 29.4 (2012), pp. 258–268.
- [6] Ato Institucional N° BRASIL. “Lei nº 8.777, de 11 de Maio de 2016.” Em: *Institui a Política de Dados Abertos do Poder Executivo federal*. (2016).
- [7] William H Inmon. *Building the data warehouse*. John wiley & sons, 2005.
- [8] BRASIL. “Decreto nº 10.332, de 28 de abril de 2020.” brazil. Em: *Institui a Estratégia de Governo Digital para o período de 2020 a 2022, no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional e dá outras providências*. (2020).
- [9] Sébastien Martin et al. “Open data: Barriers, risks and opportunities”. Em: *Proceedings of the 13th European Conference on eGovernment: ECEG*. 2013, pp. 301–309.

- [10] Ingrid Palma A., Marcelo Ladeira e Ana Carla Bittencourt Reis. “Machine Learning Predictive Model for the Passive Transparency at the Brazilian Ministry of Mines and Energy”. Em: *DG.O’21: The 22nd Annual International Conference on Digital Government Research, Omaha, NE, USA, June 9-11, 2021*. 2021, pp. 76–81. DOI: 10.1145/3463677.3463715. URL: <https://doi.org/10.1145/3463677.3463715>.
- [11] Sébastien Martin. “Risk analysis to overcome barriers to open data”. Inglês. Em: *Electronic Journal of e-Government* 11.2 (2013), pp348–359.
- [12] AMT Ali-Eldin, AMG Zuiderwijk-van Eijk e MFWHA Janssen. “Opening more data: A new privacy risk scoring model for open data”. Em: *Proceedings of the 7th International Symposium on Business Modeling and Software Design 2017*. 2017.
- [13] María Verónica Alderete. “The mediating role of ICT in the development of open government”. Em: *Journal of Global Information Technology Management* 21.3 (2018), pp. 172–187. DOI: 10.1080/1097198X.2018.1498273. eprint: <https://doi.org/10.1080/1097198X.2018.1498273>. URL: <https://doi.org/10.1080/1097198X.2018.1498273>.
- [14] Nigel Shadbolt et al. “Linked open government data: Lessons from Data.gov.uk”. Em: *IEEE Intelligent Systems* 27.3 (18 de mai. de 2012), pp. 16–24. URL: <https://eprints.soton.ac.uk/340564/> (acesso em 16/02/2021).
- [15] Jennifer C. Molloy. “The open knowledge foundation: Open data means better science”. Em: *PLoS biology* 9.12 (6 de dez. de 2011), pp. 1–4. DOI: <https://doi.org/10.1371/journal.pbio.1001195>. URL: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001195> (acesso em 16/02/2021).
- [16] Jan Kucera e Dusan Chlapek. “Benefits and risks of open government data”. Em: *Journal of Systems Integration* 5.1 (2014), pp. 30–41. ISSN: 1804-2724. DOI: <http://dx.doi.org/10.20470/jsi.v5i1.185>. URL: <http://si-journal.org/index.php/JSI/article/viewFile/185/254> (acesso em 16/02/2021).
- [17] Rui Pedro Lourenço. “Open government portals assessment: A transparency for accountability perspective”. Em: *International Conference on Electronic Government*. Springer. 2013, pp. 62–74. ISBN: 978-3-642-40357-6. DOI: https://doi.org/10.1007/978-3-642-40358-3_6. URL: https://link.springer.com/chapter/10.1007/978-3-642-40358-3_6 (acesso em 16/02/2021).
- [18] Ahmad Luthfi, Marijn Janssen e Joep Cromptvoets. “A Causal Explanatory Model of Bayesian-belief Networks for Analysing the Risks of Opening Data”. Em: Boris Shhishkov. *International Symposium on Business Modeling and Software Design*.

- Springer Verlag, 2018, pp. 289–297. ISBN: 978-3-319-94213-1. DOI: 10.1007/978-3-319-94214-8_20. URL: https://link.springer.com/chapter/10.1007/978-3-319-94214-8_20 (acesso em 16/02/2021).
- [19] Jeannine E Relly e Meghna Sabharwal. “Perceptions of transparency of government policymaking: A cross-national study”. Em: *Government Information Quarterly* 26.1 (2009), pp. 148–157.
- [20] Tim Berners-Lee. *Linked data. Design issues: 2006*. URL: <https://www.w3.org/DesignIssues/LinkedData.html> (acesso em 16/02/2022).
- [21] Mark Petticrew e Helen Roberts. *Systematic reviews in the social sciences: A practical guide*. John Wiley & Sons, 2008.
- [22] Evangelos Kalampokis, Michael Hausenblas e Konstantinos Tarabanis. “Combining Social and Government Open Data for Participatory Decision-Making”. Em: *Electronic participation*. Springer, 2011, pp. 36–47. ISBN: 978-3-642-23332-6. DOI: https://doi.org/10.1007/978-3-642-23333-3_4. URL: https://link.springer.com/chapter/10.1007/978-3-642-23333-3_4 (acesso em 16/02/2021).
- [23] Thorhildur Jetzek, Michel Avital e Niels Bjorn-Andersen. “The Value of Open Government Data: A Strategic Analysis Framework”. Em: *SIG eGovernment pre-ICIS Workshop*. Orlando, 2012, pp. 1–12. URL: https://www.researchgate.net/publication/260740425_The_Value_of_Open_Government_Data_A_Strategic_Analysis_Framework (acesso em 16/02/2021).
- [24] Ben Worthy. “The impact of open data in the UK: Complex, unpredictable, and political”. Em: *Public Administration* 93.3 (2015), pp. 788–805. DOI: <https://doi.org/10.1111/padm.12166>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/padm.12166> (acesso em 16/02/2021).
- [25] Thorhildur Jetzek. “Managing complexity across multiple dimensions of liquid open data: The case of the Danish basic data program”. Em: *Government Information Quarterly* 33.1 (2016), pp. 89–104. DOI: <https://doi.org/10.1016/j.giq.2015.11.003>. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0740624X15300186> (acesso em 16/02/2021).
- [26] Lion Hirth. “Open data for electricity modeling: Legal aspects”. Em: *Energy Strategy Reviews* 27 (2020), p. 100433.
- [27] Xiaoli Chen et al. “Open is not enough”. Inglês. Em: *Nature Physics* 15.2 (2019), pp. 113–119. DOI: 10.1038/s41567-018-0342-2. URL: <https://doi.org/10.1038/s41567-018-0342-2>.

- [28] OECD. “The OECD Digital Government Policy Framework”. Inglês. Em: 02 (2020). DOI: <https://doi.org/https://doi.org/10.1787/f64fed2a-en>. URL: <https://www.oecd-ilibrary.org/content/paper/f64fed2a-en>.
- [29] Gonzalo Mariscal, Oscar Marban e Covadonga Fernandez. “A survey of data mining and knowledge discovery process models and methodologies”. Em: *The Knowledge Engineering Review* 25.2 (2010), pp. 137–166.
- [30] Marina de Andrade Marconi e Eva Maria Lakatos. *Fundamentos de metodologia científica*. 5. ed.-São Paulo: Atlas, 2003.
- [31] Cleber Cristiano Prodanov e Ernani Cesar De Freitas. “Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico - 2ª Edição”. Em: (2013).
- [32] AL Cervo e PA Bervian. *Metodologia científica*. 5aed. 2002.
- [33] Evangelia Triperina. “Visual interactive knowledge management for multicriteria decision making and ranking in linked open data environments”. Tese de dout. Limoges, 2020.
- [34] Tatiana Engel Gerhardt e Denise Tolfo Silveira. *Métodos de pesquisa*. Plageder, 2009.
- [35] Ari Melo Mariano e Maíra Santos. “Revisão da literatura: apresentação de uma abordagem integradora”. Em: 18 (2017), pp. 427–442.
- [36] Pete Chapman et al. “CRISP-DM 1.0: Step-by-step data mining guide”. Em: *SPSS inc* 9 (2000), p. 13.
- [37] Valerie Belton e Theodor J. Stewart. *Multiple Criteria Decision Analysis: an Integrated Approach*. Boston, MA: Springer US, 2002. 391 pp. ISBN: 978-1-4615-1495-4. DOI: <https://doi.org/10.1007/978-1-4615-1495-4>. URL: <https://link.springer.com/book/10.1007/978-1-4615-1495-4> (acesso em 16/02/2021).
- [38] Antonio Carlos Gil. “Métodos e técnicas de pesquisa social. 8 reimpr”. Em: *São Paulo: Atlas* 201 (2010).
- [39] Maria Dominguez Costa Pinho e Maria Paula Almada Silva. “Governo aberto e dados abertos governamentais: um mapeamento e sistematização da produção acadêmica”. Em: *Comunicação & Inovação* 20.43 (2019).
- [40] Wallace Parks. “The Open Government Principle: Applying the Right to Know Under the Constitution, 26 Geo”. Em: *Wash. L. Rev* 1.8-9 (1957), p. 12.
- [41] Carlos Jiménez. “Sobre el término Open Government y su aparición documentada”. Em: *Blog de Estratic* (2013).

- [42] Senado Federal, ed. *Constituição da República Federativa do Brasil: texto constitucional promulgado em 5 de outubro de 1988, com as alterações determinadas pelas Emendas Constitucionais de Revisão nos 1 a 6/94, pelas Emendas Constitucionais nos 1/92 a 91/2016 e pelo Decreto Legislativo no 186/2008*. Coordenação de Edições Técnicas, 1988. 498 pp.
- [43] Ato Institucional N^o BRASIL. “Lei n^o 12.527, de 18 de Novembro de 2011.” Em: *Regula o acesso à informações previsto no inciso XXXIII do art. 5^o, no inciso II do § 3^o do art. 37 e no § 2^o do art. 216 da Constituição Federal*. 37 (2011).
- [44] Cláudio Sonaglio Albano. “Dados governamentais abertos: proposta de um modelo de produção e utilização de informações sob a ótica conceitual da cadeia de valor”. Tese de dout. Universidade de São Paulo, 2014.
- [45] C. J. Date. *Introdução a Sistemas de Banco de Dados*. 7^a ed. Campus, 1989. ISBN: 85-7001-392-2. URL: <http://gen.lib.rus.ec/book/index.php?md5=D67319E9E2A397591DFF6C8BA15A3142>.
- [46] Shamkant B. Navathe Ramez Elmasri. *Sistemas de Banco de Dados*. 6^a edição. Pearson Education do Brasil, 2011. ISBN: 9788543013817. URL: <http://gen.lib.rus.ec/book/index.php?md5=27CE680D7EB8DB893CB7C63333B015F9>.
- [47] Bernadette Lóscio, Caroline Guimarães e Newton Calegari. “Data on the web best practices: Challenges and benefits”. Em: *Open Data Reserach Symposium (ODRS 2016)*. 2016.
- [48] Jenn Riley. “Understanding metadata”. Em: *Washington DC, United States: National Information Standards Organization (<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>)* 23 (2017).
- [49] Barbara Ubaldi. *Open Government Data. Towards Empirical Analysis of Open Government Data Initiatives*. OECD Working Papers on Public Governance 22. OECD, 2013. 61 pp. DOI: <https://doi.org/https://doi.org/10.1787/5k46bj4f03s7-en>. URL: <https://www.oecd-ilibrary.org/content/paper/5k46bj4f03s7-en>.
- [50] Pascal Hitzler, Markus Krotzsch e Sebastian Rudolph. *Foundations of semantic web technologies*. CRC press, 2009.
- [51] Tom Heath e Christian Bizer. “Linked data: Evolving the web into a global data space”. Em: *Synthesis lectures on the semantic web: theory and technology* 1.1 (2011), pp. 1–136.
- [52] Sören Auer, Volha Bryl e Sebastian Tramp. *Linked Open Data--Creating Knowledge Out of Interlinked Data: Results of the LOD2 Project*. Vol. 8661. Springer, 2014.

- [53] Bernadette Hyland et al. “Linked data glossary”. Em: *W3C Government Linked Data Working Group*. <http://www.w3.org/TR/ld-glossary> (2013).
- [54] Renáta Máchová e Martin Lněnička. “A multi-criteria decision making model for the selection of open data management systems”. Em: *Electronic Government, an International Journal* 15.4 (2019), pp. 372–391.
- [55] Graham Klyne. “Resource description framework (RDF): Concepts and abstract syntax”. Em: <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/> (2004).
- [56] Florian Bauer e Martin Kaltenböck. “Linked open data: The essentials”. Em: *Edition mono/monochrom, Vienna* 710 (2011).
- [57] Siti Isnaine Haini et al. “Factors Influencing the Adoption of Open Government Data in The Public Sector: A Systematic Literature Review”. Em: *Int. J. Adv. Sci. Eng. Inf. Technol* 10 (2020), pp. 611–617.
- [58] Augusto Cesar Gadelha Vieira. “Melhorando o Acesso ao Governo com o Melhor Uso da Web”. Em: *São Paulo: W3C Brasil, São Paulo* (2009).
- [59] Evangelos Kalampokis, Efthimios Tambouris e Konstantinos Tarabanis. “A classification scheme for open government data: Towards linking decentralised data”. Em: *International Journal of Web Engineering and Technology* 6.3 (2011), pp. 266–285. DOI: <https://doi.org/10.1504/IJWET.2011.040725>. URL: <https://www.inderscienceonline.com/doi/abs/10.1504/IJWET.2011.040725> (acesso em 16/02/2021).
- [60] Fernando Almeida Barbalho. “Emergência de um campo de ação estratégica: o caso de política pública sobre dados abertos”. Em: (2014).
- [61] Aikaterini Yannoukakou e Iliana Araka. “Access to government information: Right to information and open government data synergy”. Em: *Procedia-Social and Behavioral Sciences* 147 (2014), pp. 332–340.
- [62] Mauricio Solar et al. “A Model to Guide the Open Government Data Implementation in Public Agencies.” Em: *J. UCS* 20.11 (2014), pp. 1564–1582.
- [63] Benedikt Simon Hitz-Gamper, Oliver Neumann e Matthias Stürmer. “Balancing control, usability and visibility of linked open government data to create public value”. Em: *International journal of public sector management* (2019).
- [64] Martin J Murillo. “Position paper: Including all audiences in the government loop: From transparency to empowerment through open government data”. Em: *The European Commission’s Albert Borschette Conference Center* (2012).
- [65] Joshua Tauberer e Larry Lessig. “The 8 principles of open government data”. Em: *Obtido de <http://www.opengovdata.org/home/8principles>* 28 (2007).

- [66] Mika Marttunen et al. “Methods to inform the development of concise objectives hierarchies in multi-criteria decision analysis”. Em: *European Journal of Operational Research* 277.2 (2019), pp. 604–620.
- [67] Frederick S Hillier e Gerald J Lieberman. *Introdução à pesquisa operacional*. McGraw Hill Brasil, 2013.
- [68] Michael Weisberg. “Who is a Modeler?” Em: *The British journal for the philosophy of science* 58.2 (2007), pp. 207–233.
- [69] Roman Frigg e Stephan Hartmann. “Models in science”. Em: (2006).
- [70] Salvatore T March e Gerald F Smith. “Design and natural science research on information technology”. Em: *Decision support systems* 15.4 (1995), pp. 251–266.
- [71] Marcos ARENALES et al. *Pesquisa Operacional [recurso eletrônico]*. 2011.
- [72] Heloise Acco Tives Leão et al. “Research Project Selection and Classification using MCDA Methods”. Em: *Proceedings of the XIV Brazilian Symposium on Information Systems*. 2018, pp. 1–9.
- [73] Jay W Forrester. “Counterintuitive behavior of social systems”. Em: *Theory and decision* 2.2 (1971), pp. 109–140.
- [74] Herbert A Simon. *Administrative behavior*. Simon e Schuster, 2013.
- [75] Herbert A Simon. “The new science of management decision.” Em: (1960).
- [76] Denis Bouyssou et al. *Evaluation and decision models with multiple criteria: Stepping stones for the analyst*. Vol. 86. Springer Science & Business Media, 2006.
- [77] Michael Mintrom. “Herbert A. Simon, Administrative behavior: A study of decision-making processes in administrative organization”. Em: *The Oxford Handbook of Classics in Public Policy and Administration* (2015), pp. 12–21.
- [78] Bernard Roy. “Decision-aid and decision-making”. Em: *European Journal of Operational Research* 45.2-3 (1990), pp. 324–331.
- [79] Michael W Carter, Camille C Price e Ghaith Rabadi. *Operations research: a practical introduction*. Chapman e Hall/CRC, 2018.
- [80] Thomas Finne. “Information systems risk management: key concepts and business processes”. Em: *Computers & Security* 19.3 (2000), pp. 234–242.
- [81] Bernard Roy. *Multicriteria Methodology for Decision Aiding*. Boston, MA: Springer US Imprint Springer, 1996. 315 pp. ISBN: 978-1-4419-4761-1. URL: <https://link.springer.com/book/10.1007/978-1-4757-2500-1> (acesso em 16/02/2021).

- [82] Denis Bouyssou. “Building criteria: a prerequisite for MCDA”. Em: *Readings in multiple criteria decision aid*. Springer, 1990, pp. 58–80.
- [83] ISO GUIA ABNT. “73: 2009-Gestão de riscos--Vocabulário”. Em: *Rio de Janeiro: Associação Brasileira de Normas Técnicas* (2009).
- [84] IEC ISO. “ABNT NBR ISO/IEC 31000: 2018 Gestão de riscos - Diretrizes”. Em: *Associação Brasileira de Normas Técnicas* (2018).
- [85] Carlos A. Bana e Costa e Philippe Vincke. “Multiple criteria decision aid: An overview”. Em: *Readings in multiple criteria decision aid*. Springer, 1990, pp. 3–14. ISBN: 978-3-642-75937-6. DOI: https://doi.org/10.1007/978-3-642-75935-2_1. URL: https://link.springer.com/chapter/10.1007/978-3-642-75935-2_1 (acesso em 16/02/2021).
- [86] Miriam Martínez Garcia. “Enhancing the electre decision support method with semantic data”. Tese de dout. Universitat Rovira i Virgili, 2019.
- [87] Carlos A Bana e Costa. *Readings in multiple criteria decision aid*. Springer Science & Business Media, 2012.
- [88] Walter J. Gutjahr e Pamela C. Nolz. “Multicriteria optimization in humanitarian aid”. Em: *European Journal of Operational Research* 252.2 (16 de jul. de 2016), pp. 351–366. DOI: <https://doi.org/10.1016/j.ejor.2015.12.035>. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0377221715011741> (acesso em 16/02/2021).
- [89] Mohammad Alamgir Hossain, Yogesh K Dwivedi e Nripendra P Rana. “State-of-the-art in open data research: Insights from existing literature and a research agenda”. Em: *Journal of organizational computing and electronic commerce* 26.1-2 (2016), pp. 14–40.
- [90] Ahmad Luthfi, Marijn Janssen e Joep Cromptvoets. “Decision Tree Analysis for Estimating the Costs and Benefits of Disclosing Data”. Em: *Conference on e-Business, e-Services and e-Society*. Springer. 2019, pp. 205–217.
- [91] Ahmad Luthfi e Marijn Janssen. “A conceptual model of decision-making support for opening data”. Em: *International Conference on e-Democracy*. Springer Verlag, 2017, pp. 95–105. ISBN: 978-3-319-71116-4. DOI: https://doi.org/10.1007/978-3-319-71117-1_7. URL: https://link.springer.com/chapter/10.1007/978-3-319-71117-1_7 (acesso em 16/02/2021).
- [92] Alexei Botchkarev. “Towards a Methodology of Multi-criteria Prioritization of Open Data for Public Release”. 17 de set. de 2018.

- [93] Anneke Zuiderwijk e Marijn Janssen. “Towards decision support for disclosing data: Closed or open data?” Inglês. Em: *Information Polity* 20.2, 3 (2015), pp. 103–117.
- [94] Massimiliano Manfren et al. “Open data and energy analytics-An analysis of essential information for energy system planning, design and operation”. Em: *Energy* 213 (2020), p. 118803.
- [95] Anamaria Buda et al. “Decision support framework for opening business data”. Em: (2015).
- [96] Ahmad Luthfi, Marijn Janssen e Joep Crompvoets. “A Framework for Analyzing How Governments Open Their Data: Institution, Technology, and Process Aspects Influencing Decision-Making”. Em: *EGOV-CeDEM-ePart 2018* (2018), p. 163.
- [97] Neema Nicodemus Lyimo et al. “A Fuzzy Logic-Based Approach for Modelling Uncertainty in Open Geospatial Data on Landfill Suitability Analysis”. Em: *ISPRS International Journal of Geo-Information* 9.12 (2020), p. 737.
- [98] Ahmad Luthfi et al. “A fuzzy multi-criteria decision making approach for analyzing the risks and benefits of opening data”. Em: *Conference on e-Business, e-Services and e-Society*. Springer Verlag, 2018, pp. 397–412. ISBN: 978-3-030-02130-6. DOI: https://doi.org/10.1007/978-3-030-02131-3_36. URL: https://link.springer.com/chapter/10.1007/978-3-030-02131-3_36 (acesso em 16/02/2021).
- [99] Sylvain Kubler et al. “Open data portal quality comparison using AHP”. Em: *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*. ACM. Association for Computing Machinery, 2016, pp. 397–407. DOI: <https://doi.org/10.1145/2912160.2912167>. URL: <https://dl.acm.org/doi/abs/10.1145/2912160.2912167> (acesso em 16/02/2021).
- [100] Salman Ahmed Shaikh et al. “An AHP/TOPSIS-Based Approach for an Optimal Site Selection of a Commercial Opening Utilizing GeoSpatial Data”. Em: *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*. 2020, pp. 295–302. DOI: 10.1109/BigComp48618.2020.00-58.
- [101] Gary Alexander Parung et al. “Barriers and strategies of open government data adoption using fuzzy AHP-TOPSIS: A case of Indonesia”. Em: *Transforming Government: People, Process and Policy* (2018).
- [102] Saima Rafi et al. “Multicriteria based decision making of DevOps data quality assessment challenges using fuzzy TOPSIS”. Em: *IEEE Access* 8 (2020), pp. 46958–46980.

- [103] Christian P Robert, Nicolas Chopin, Judith Rousseau et al. “Harold Jeffreys’s theory of probability revisited”. Em: *Statistical Science* 24.2 (2009), pp. 141–172.
- [104] John Skilling e Stephen F Gull. “Bayesian maximum entropy image reconstruction”. Em: *Lecture Notes-Monograph Series* (1991), pp. 341–367.
- [105] Rensis Likert. “A technique for the measurement of attitudes.” Em: *Archives of psychology* (1932).
- [106] Ahmad Luthfi e Marijn Janssen. “A Comparative Study of Methods for Deciding to Open Data”. Em: *International Symposium on Business Modeling and Software Design*. Springer. 2019, pp. 213–220.
- [107] Stefan Pfenninger et al. “The importance of open data and software: Is energy research lagging behind?” Em: *Energy Policy* 101 (2017), pp. 211–215.
- [108] Vytautas Palevičius et al. “Decision-Aiding Evaluation of Public Infrastructure for Electric Vehicles in Cities and Resorts of Lithuania”. Em: *Sustainability* 10.4 (2018), p. 904. DOI: <https://doi.org/10.3390/su10040904>. URL: <https://www.mdpi.com/2071-1050/10/4/904> (acesso em 16/02/2021).
- [109] M. Böhms, T. Rieswijk e E. Lijster. “Linked energy data: Enabling monitoring and decision support for improved energy management”. Em: *2015 IEEE International Conference on Engineering, Technology and Innovation/ International Technology Management Conference (ICE/ITMC)*. 2015, pp. 1–8. DOI: 10.1109/ICE.2015.7438671.
- [110] Thomas Hofmann. “Unsupervised learning by probabilistic latent semantic analysis”. Em: *Machine learning* 42.1 (2001), pp. 177–196.
- [111] John Stasko, Carsten Görg e Zhicheng Liu. “Jigsaw: supporting investigative analysis through interactive visualization”. Em: *Information visualization* 7.2 (2008), pp. 118–132.
- [112] Thomas L. Saaty. “A scaling method for priorities in hierarchical structures”. Em: *Journal of Mathematical Psychology* 15.3 (1977), pp. 234–281. DOI: [https://doi.org/10.1016/0022-2496\(77\)90033-5](https://doi.org/10.1016/0022-2496(77)90033-5). URL: <https://www.sciencedirect.com/science/article/abs/pii/0022249677900335> (acesso em 16/02/2021).
- [113] Thomas L. Saaty. “How to make a decision: The analytic hierarchy process”. Em: *European Journal of Operational Research* 48.1 (1990). Decision making by the analytic hierarchy process: Theory and applications, pp. 9–26. ISSN: 0377-2217. DOI: [https://doi.org/10.1016/0377-2217\(90\)90057-I](https://doi.org/10.1016/0377-2217(90)90057-I). URL: <http://www.sciencedirect.com/science/article/pii/037722179090057I> (acesso em 16/02/2021).

- [114] Sylvain Kubler et al. “A state-of-the-art survey & testbed of fuzzy AHP (FAHP) applications”. Em: *Expert Systems with Applications* 65 (2016), pp. 398–422.
- [115] Thomas L Saaty. “Some mathematical concepts of the analytic hierarchy process”. Em: *Behaviormetrika* 18.29 (1991), pp. 1–9.
- [116] Patrick T Harker e Luis G Vargas. “The theory of ratio scale estimation: Saaty’s analytic hierarchy process”. Em: *Management science* 33.11 (1987), pp. 1383–1403.
- [117] Thomas L. Saaty. *Theory and Applications of the Analytic Network Process: Decision Making with Benefits, Opportunities, Costs, and Risks*. 2^a ed. RWS publications, 2005. 352 pp. ISBN: 978-1-8886030-6-4. URL: <http://www.rwspublications.com/books/index.php?section=anp&book=4> (acesso em 16/02/2021).
- [118] Rodney Wernke e Antonio Cezar Bornia. “A contabilidade gerencial e os métodos multicriteriais”. Em: *Revista Contabilidade & Finanças* 12.25 (2001), pp. 60–71. ISSN: 1808-057X. DOI: <https://doi.org/10.1590/S1519-70772001000100004>. URL: http://old.scielo.br/scielo.php?pid=S1519-70772001000100004&script=sci_abstract&tlng=pt (acesso em 16/02/2021).
- [119] CL Hwang e K Yoon. “Multiple Attributes Decision Making Methods and Applications, spring”. Em: *New York* (1981).
- [120] Roseanna W Saaty. “The analytic hierarchy process—what it is and how it is used”. Em: *Mathematical modelling* 9.3-5 (1987), pp. 161–176.
- [121] A. C. B. Reis e V. B. Schramm. “Guia para Aplicação da Análise Multicritério em Análise de Impacto Regulatório (AIR) no Inmetro”. Em: *Projeto de Melhoria da Qualidade Regulatória - PN 15.2099.8-019.00* (2022), pp. 21–30.
- [122] Thomas L. Saaty. “Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process”. Em: *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. A. Matematicas* 102.2 (2008), pp. 251–318. DOI: <https://doi.org/10.1007/BF03191825>. URL: <https://link.springer.com/article/10.1007/BF03191825> (acesso em 16/02/2020).
- [123] ME Qureshi, Steve R Harrison e MK Wegener. “Validation of multicriteria analysis models”. Em: *Agricultural Systems* 62.2 (1999), pp. 105–116.
- [124] Krishnendu Mukherjee. “Analytic hierarchy process and technique for order preference by similarity to ideal solution: a bibliometric analysis from past, present and future of AHP and TOPSIS”. Em: (2014).

- [125] Evangelos Triantaphyllou e Alfonso Sánchez. “A sensitivity analysis approach for some deterministic multi-criteria decision-making methods”. Em: *Decision sciences* 28.1 (1997), pp. 151–194. DOI: <https://doi.org/10.1111/j.1540-5915.1997.tb01306.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5915.1997.tb01306.x> (acesso em 16/02/2021).
- [126] Leandro Peçanha De Souza, Carlos Francisco Simões Gomes e Alexandre Pinheiro De Barros. “Implementation of new hybrid AHP-TOPSIS-2N method in sorting and prioritizing of an it CAPEX project portfolio”. Em: *International Journal of Information Technology & Decision Making* 17.04 (2018), pp. 977–1005.
- [127] Constantin Zopounidis e Michael Doumpos. *Multiple Criteria Decision Making: Applications in Management and Engineering*. Cham: Springer, 2017. 223 pp. ISBN: 978-3-319-39292-9. DOI: <https://doi.org/10.1007/978-3-319-39292-9>.
- [128] Judie Attard et al. “A systematic review of open government data initiatives”. Em: *Government Information Quarterly* 32.4 (2015), pp. 399–418. ISSN: 0740-624X. DOI: <https://doi.org/10.1016/j.giq.2015.07.006>. URL: <http://www.sciencedirect.com/science/article/pii/S0740624X1500091X> (acesso em 16/02/2021).

Anexo I

Considerações acerca da Dissertação - Autoridade de Monitoramento da LAI do MME



MINISTÉRIO DE MINAS E ENERGIA
Assessoria Especial de Controle Interno
Esplanada dos Ministérios - Bloco U, Brasília/DF, CEP 70065-900
Telefone: (61) 2032-5028/5364 / aeci@mme.gov.br

Ofício nº 19/2022/AECI-MME

Ao Senhor Prof. Dr.

MARCELO LADEIRA

Coordenador do Programa de Pós-Graduação em Computação Aplicada (PPCA)
Departamento de Ciência da Computação da Universidade de Brasília (CIC/UnB)
ICC Centro - Módulo 14 - Campus Universitário Darcy Ribeiro
CEP 70910-900 - Brasília - DF

Assunto: **Considerações acerca da Dissertação de Mestrado da servidora Ingrid Palma Araújo.**

Senhor Coordenador,

1. Cumprimentando-o cordialmente, participo que a dissertação intitulada "Modelo de Apoio à Decisão para Avaliar Dados Governamentais Abertos do Setor Elétrico", elaborada pela servidora Ingrid Palma Araújo, com a orientação da Profª Drª Ana Carla Bittencourt Reis, no âmbito do Mestrado Profissional em Computação Aplicada da UnB, apresenta contribuições importantes para o processo de aprimoramento da Política de Dados Abertos do Ministério de Minas e Energia (MME), contribuindo para o incremento da exigida Transparência Ativa na Administração Pública Federal, assunto de significativa para a sociedade civil, acadêmica, gestores e demais atores; e indo ao encontro do interesse desta Autoridade de Monitoramento da Lei de Acesso à Informação (Lei nº 12.527, de 18 de novembro de 2011).
2. Cabe registrar que o Decreto nº 10.332, de 28 de abril de 2020 (Estratégia de Governo Digital), define uma série de objetivos a serem alcançados pela Administração Pública, com destaque para o "*acesso amplo à informação e aos dados abertos governamentais, para possibilitar o exercício da cidadania e a inovação em tecnologias digitais*". Nesse sentido, a referida pesquisa, que se apoiou em dados fornecidos pela Ouvidoria deste Ministério e em relatórios do setor energético, além de dados abertos do MME e da Agência Nacional de Energia Elétrica, desenvolveu um modelo que pode ser útil não apenas para priorizar temas mais relevantes para a abertura de dados, à luz dos critérios de riscos definidos na dissertação e da aplicação das técnicas e métodos descritos, mas também para os gestores responsáveis por decisões que envolvam direcionamento de recursos a políticas públicas voltadas ao atendimento de demandas de maior interesse da Sociedade.
3. Ressalta-se, ainda, que a referida servidora, Analista em Tecnologia da Informação em exercício neste Ministério, com experiência de 14 anos nas ações relacionadas ao tema em comento, vem empreendendo esforços para propor mudança positivas na cultura de transparência e nas demais ações afetas ao aprimoramento do Plano de Dados Abertos (PDA) deste Órgão. Assim, participo ser intenção encaminhar, oportunamente, a referida pesquisa para apreciação do Comitê

de Governança Digital (CGD/MME), com vistas à avaliação da incorporação da solução, no que for aplicável, por ocasião do processo de revisão e elaboração do novo PDA do MME.

4. Por fim, agradecendo a oportunidade, apresento meu apreço pelo esmerado trabalho desenvolvido pelos integrantes desse valoroso Programa, certo de sua contribuição para o contínuo crescimento de todos os setores da Sociedade e para a formação de profissionais cada vez mais qualificados.

Atenciosamente,



Documento assinado eletronicamente por **Sergio Henrique Lopes de Sousa**, **Autoridade de Monitoramento da Lei de Acesso à Informação**, em 21/07/2022, às 15:01, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site http://www.mme.gov.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0650059** e o código CRC **7703F1CD**.

Referência: Caso responda este Ofício, indicar expressamente o Processo nº 48300.001007/2022-12

SEI nº 0650059

Anexo II

Impacto Potencial da Dissertação - Ouvidoria do MME



MINISTÉRIO DE MINAS E ENERGIA
Assessoria Especial de Controle Interno
Esplanada dos Ministérios - Bloco U, Brasília/DF, CEP 70065-900
Telefone: (61) 2032-5028/5364 / aeci@mme.gov.br

Ofício nº 22/2022/AECI-MME

Ao Senhor Prof. Dr.

MARCELO LADEIRA

Coordenador do Programa de Pós-Graduação em Computação Aplicada (PPCA)
Departamento de Ciência da Computação da Universidade de Brasília (CIC/UnB)
ICC Centro - Módulo 14 - Campus Universitário Darcy Ribeiro
CEP 70910-900 - Brasília - DF

Assunto: **Impacto Potencial da Dissertação de Mestrado da servidora Ingrid Palma.**

Senhor Coordenador,

1. Com meus cordiais cumprimentos, participo que a dissertação intitulada "Modelo de Apoio à Decisão para Avaliar Dados Governamentais Abertos do Setor Elétrico", elaborada pela servidora Ingrid Palma Araújo, com a orientação da Prof^a Dr^a Ana Carla Bittencourt Reis, no âmbito do Mestrado Profissional em Computação Aplicada da UnB, apresenta potenciais contribuições para o efetivo aprimoramento da Política de Dados Abertos do Ministério de Minas e Energia (MME), com efeito, sobretudo, para a projeção da Transparência Ativa no Poder Executivo, assunto disposto na Lei nº 12.527, de 18 de novembro de 2011 (Lei de Acesso à Informação), que prevê como uma de suas diretrizes a divulgação de informações de interesse público, independentemente de solicitações, dando-se preferência na divulgação através dos meios de comunicação viabilizados pela tecnologia da informação.
2. Assim, em relação aos trabalhos desenvolvidos nesta Ouvidoria-Geral cabe destacar a relevância do estudo apresentado pela mencionada servidora, tendo em vista que a abertura de dados abertos minimiza as atividades do setor e os custos de processamento e gerenciamento de pedidos individuais, além de ser fonte de divulgação ativa de informações de interesse público, pois quanto mais os dados são abertos, maior o acesso às informações e maior a transparência dessas informações.
3. Sobre a pesquisa em tela, pontua-se que a coleta de dados envolveu informações fornecidas por esta Ouvidoria-Geral e os dados abertos do MME e da Agência Nacional de Energia Elétrica. A partir disso, e com base na revisão sistemática da literatura, foi possível identificar critérios de riscos relacionados à abertura de dados e a consecutiva construção do modelo proposto na dissertação.
4. Os resultados apresentam um *ranking* do conjunto de dados abertos governamentais do setor elétrico, indicando o desempenho de cada base, a fim de

subsidiar a avaliação e a tomada de decisão para o gerenciamento e o adequado aprimoramento dessas bases.

5. Por fim, ressalta-se o incentivo e suporte oferecidos por esse Programa para a formação técnica e profissional da aludida servidora, inclusive para a publicação do artigo [Machine Learning Predictive Model for the Passive Transparency at the Brazilian Ministry of Mines and Energia](#), apresentado na 22ª Conferência Internacional Anual sobre Pesquisa em Governo Digital, evento de referência mundial sobre essa temática.

Atenciosamente,

Grayce Martins da Silva Gonçalves
Ouvidora



Documento assinado eletronicamente por **Grayce Martins da Silva Gonçalves, Ouvidor(a)-Geral**, em 21/07/2022, às 17:19, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site http://www.mme.gov.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0650608** e o código CRC **582B507E**.

Referência: Caso responda este Ofício, indicar expressamente o Processo nº 48300.001007/2022-12

SEI nº 0650608

Anexo III

Artigo Publicado na *Association for
Computing Machinery (ACM)*



Machine Learning Predictive Model for the Passive Transparency at the Brazilian Ministry of Mines and Energy

Ingrid Palma, Department of Computer Science, University of Brasilia (UnB), Brazil,
ingrid.palma@uiot.org

Marcelo Ladeira, Department of Computer Science, University of Brasilia (UnB), Brazil,
mladeira@unb.br

Ana Carla Bittencourt Reis, Department of Computer Science, University of Brasilia (UnB), Brazil,
anacarlabr@unb.br

DOI: <https://doi.org/10.1145/3463677.3463715>

DG.O'21: [DG.O2021: The 22nd Annual International Conference on Digital Government Research](#),
Omaha, NE, USA, June 2021

This paper presents a case study based on the CRISP-DM Model and the use of Text Mining tools and techniques to automate the Passive Transparency process at the Brazilian Ministry of Mines and Energy. Thus, a Machine Learning Model is proposed to predict the class of the technical unit responsible for the data/information requested by citizens. Through the application of the algorithm LDA and TF-IDF it was possible to map the topics of the most relevant subjects for society. The stability of the model was tested from the comparative analysis between 5 known classification algorithms (Random Forest, Multinomial NB, Linear SVC, Logistic Regression, XGBoost and Gradient Boosting). XGBoost presented better performance and precision in multiclass learning outcomes.

CCS Concepts: • **Theory of computation** → **Unsupervised learning and clustering**; • **Applied computing** → **Multicriteria optimization and Decision-Making**; • **Theory of computation**~**Design and analysis of algorithms**;

Keywords: Machine Learning Algorithms, Passive Transparency, Multicriteria Decision Making, Topic Modeling, Predictive Analysis and XGBoost.

ACM Reference Format:

Ingrid Palma, Marcelo Ladeira, and Ana Carla Bittencourt Reis. 2021. Machine Learning Predictive Model for the Passive Transparency at the Brazilian Ministry of Mines and Energy. In *DG.O2021: The 22nd Annual International Conference on Digital Government Research (DG.O'21)*, June 9–11, 2021, Omaha, NE, USA. ACM, New York, NY, USA, 6 pages.

Anexo IV

*Artigo Aceito na 9th International
Conference on Future Internet of
Things and Cloud (FiCloud)*

Design and Evaluation of IoT Gateway for Data Prioritization based on Van Emde Boas Tree

Ingrid Palma Araújo, Ronaldo C. R. de Araújo, Michel Alves Ribeiro,
Edison Ishikawa, Ana C. B. Reis and Christiano A. M. Rodopoulos

Department of Computer Science

University of Brasilia (UnB)

Brasilia, Brazil

[ingridpalma, ronald.ecomp, eng.maribeiro, christianorodopoulos]@gmail.com, [edisonishikawa, anacarlbr]@unb.br

Abstract—Critical data can require actions immediate (rate appropriate to the context) as well as priority in routing the device to the Cloud. In this sense, one of the contributions of this work is to show that the prioritization of packets originating from IoT devices, which travel between Local Gateway and Cloud Gateway, via HTTPS protocols, using data structures such as Heap (Fibonacci and Binominal) and Van Emde Ems (vEB tree) are more effective than commonly used FIFO queues. Another contribution of this work is the adaptation of the mentioned data structures to the IoT. The results showed the huge advantage of vEB tree over Heaps implemented primarily in a context with the large quantity of data.

Index Terms—vEB tree, Heap Structure, IoT Data Prioritization, Largescale Data optimization, IoT Networks.

I. INTRODUCTION

The interconnection between different devices, *smart objects* and things in real and virtual world, made possible by the diversity of communication *low data rate*, *low power* and *low cost* protocols and technologies that sets the paradigm of the *Internet of Things* (IoT), which one gain even more expression with the arrival of 5G network.

In addition, the notion of **information aging** is important for making decisions based on information provided through IoT. In this case, IoT is also used to provide of the situation context. Old data can lead to wrong decisions. On the other hand, the more current is data, the better can be decision-making. Therefore, lowering latency with low cost influences two metrics: *Quality of Context* and *Cost of Context*. Factors relevant to *Smart Cities* ecosystems.

Several studies have been developed in this context, whether in the hardware part with high power MIMO (Multiple-Input Multiple-Output) antennas, or in the logical part with increasingly robust algorithms.

In the midst a lot of innovative technologies, some traditional routing algorithms, such as the ones that will be presented in this work, have been shown to be able to satisfactorily deal with the challenges of packet prioritization, enabling high performance and contributing for required Quality of Service (QoS) by most recent technologies.

Thus, problems such as congestion caused by unclassified data transmission and consequent delays, packet loss, wasted resources such as: memory, energy, *bandwidth*, processing,

among others, are minimized, generating value to IoT Networks. As an example, we can mention some cases that require prioritization in IoT Networks: i) measurement sensors, such as bio sensors; and ii) tasks involving authentication and synchronous connections.

The problem addressed in this study is related to the adequate and efficient dynamics bandwidth control and packet prioritization band on every packet which traverses in an IoT Gateway, namely, a large volume of network data. Therefore, an analysis of the speed of prioritization data using *Heaps Trees* (Fibonacci and Binominal) [1] and the Van Emde Boas (vEB) [2] structure is proposed in this paper, draw on a *Local Gateway* as a facilitating agent against the *throughput* generated by the IoT networks.

A. Local Gateway - vEB and Heaps trees Schedulers

IoT devices, usually represented by sensors and actuators, have a little computational resource due to *hardware* restrictions such as power, memory, processing and bandwidth [3], which justifies the use of a *Local Gateway* IoT, in order to support the prioritization algorithms promoted by for ementioned Heaps structures and *vEB tree*. To facilitate this understanding, the architecture of this strategy is represented by Figure 1.

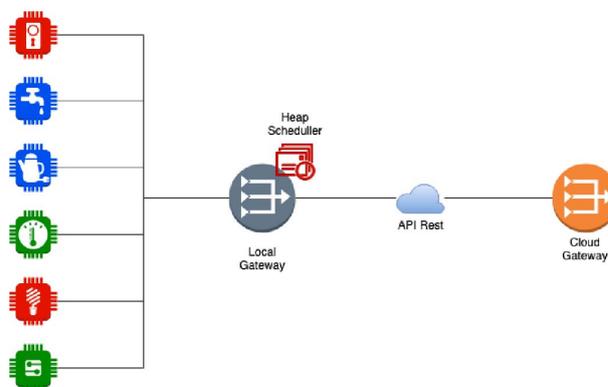


Figure 1. IoT Architecture with Heap Data Structure: An Overview

Anexo V

*Artigo Aceito na 6th World
Conference on Smart Trends in
Systems, Security and Sustainability
(WorldS4)*

Design and Application of the AHP-TOPSIS-2N to Evaluate (Linked) Open Government Data from the Electricity Datasets

ABSTRACT

This study proposes to simplify and automate this process, combining two different methods of decision support through multicriteria analysis in a model capable of judging and prioritizing risk criteria in the context of Open Data, presenting the results via iterative online dashboards developed in R. The methodology followed combines the AHP and TOPSIS-2N methods, creating a ranking of the open governmental dataset of the electricity sector in light of the risk criteria evaluated by the proposed model. The AHP technique was used to specify and normalize the importance of each criterion, considering the consistency aspects of the decision matrix. The next step was to apply the TOPSIS-2N method to sort and prioritize these datasets. The results present the datasets that should be improved concerning the respective metadata and the prioritized themes to make the decision-making for the management of the respective bases more agile and assertive.

KEYWORDS

Open Data Evaluation Model, AHP-TOPSIS-2N, (Linked) Open Government Data, Electricity Sector Open Data, Hybrid Multicriteria Decision Support Model

1 INTRODUCTION

In the last decade, it is noted the publication of laws, decrees, among other national and international regulations that foster initiatives and strategies for the implementation and publications of datasets in open format, having as an example the current Open Data Policy (Decree No. 8.777/2016). However, its wider implementation requires the identification of obstacles that prevent public agencies from implementing an Open Data strategy or ensuring its sustainability[38].

The Open Data Plan (PDA) is a government instrument created to execute the Open Data Policy in a certain body and must promote and define criteria for the expansion of the amount of open data available on the Brazilian Open Data Portal (PBDA)¹, as well as the correction and updating of this base, as recommended by the Digital Government Strategy² and the Organization for Economic Cooperation and Development.

The PBDA is the federal government's official website, which gathers data considered open from different versions (distributions). Structured as a web catalog, the Portal organizes datasets that can be easily found, accessed, and used by any person or computerized tool. Currently, this portal

has more than 10,000 datasets and approximately 51,000 cataloged resources.

Despite the actions implemented by the PDA, it is evident, in a recent study presented by [43], the growing demand for data and information related to the electricity sector, inferring that the required data were not found or published properly in the PBDA or, if available, have outdated or wrong values. In the aforementioned article, the authors used the topic modeling approach and text mining techniques to map the topics of the issues most demanded by society and other stakeholders within the Ministry of Mines and Energy (MME). The results pointed out that the most frequent themes are related to issues under the National Electric Energy Agency (ANEEL), which ratifies the scope mentioned in Subsection 1.4.

Allied to the problem at hand, it was found a small number of datasets cataloged in the PBDA, held by government agencies of the electric power segment. There are 44 electricity datasets available on dados.gov.br (see Table 1). The low quantity of data from the electric power sector cataloged in the PBDA leads, specifically to those responsible for implementing the Open Data Policy of this segment (MME and ANEEL), to uncertainties and risks in decision making related to the adequacy of these databases or in the identification of potential topics for data opening.

1.1 Context and Issues

Although the Open Data paradigm creates opportunities for economic development [1], encourages scientific production [26, 40, 50], and increases civil society's participation in government by informing itself about the public manager's decision-making [27], it also generates problems such as the opening of inaccurate/inconsistent, sensitive, and/or private data, potential risks of inappropriate interpretation and misuse of the data, or even the non-use of these open bases, which may result in organizations' aversion or reluctance to adhere to the OGD initiative [30], as well as the reluctance of managers to open their bases.

The aforementioned works point to the need for a *framework* (model) capable of assessing criteria that are relevant to the context of *open data*. On the other hand, it is also verified the inexistence of multicriteria methods applied in the referenced works, especially those that deal with specific data from the electricity sector.

The systematic literature review conducted by [19] showed that only 11% of the scientific productions involving *open data* evaluation techniques applied quantitative methods, representing a large difference between the number of qualitative (47%) or purely interpretive (38%) scientific productions compared to less subjective approaches.

¹<https://dados.gov.br/>.

²<https://www.in.gov.br/web/dou/-/decreto-n-10.332-de-28-de-abril-de-2020-254430358>

Anexo VI

Código R do AHP-TOPSIS-2N

Fonte: De Souza, Gomes e De Barros [126].

https://cran.r-project.org/web/packages/ahptopsis2n/vignettes/AHP-TOPSIS-2N_Example.html

```
1 ahptopsis2n<-function(decision, criteria, minmax){
2
3
4   if(! is.matrix(decision))
5     stop("decision must be a matrix with values of alternatives for each
6       criterion")
7   if(ncol(decision)>15)
8     stop("15 is the maximum number for criteria")
9   if(ncol(decision)<3)
10    stop("3 is the minimum number for criteria")
11
12  if(! is.matrix(criteria))
13    stop("criteria must be a matrix with pairwise comparisons of
14      criteria")
15  if(ncol(criteria)!=nrow(criteria))
16    stop("criteria must be a square matrix")
17  if(ncol(criteria)!=ncol(decision))
18    stop("criteria must be the same on criteria and decision arguments")
19  if(!all(diag(criteria ==1)))
20    stop("criteria must be a matrix with diagonal entries equal to one")
21  # if(!all(criteria == 1/t(criteria)))
22  # stop("criteria entries must be revised")
23
24  if(!is.character(minmax))
25    stop("minmax must be a character vector")
26  if(!all(minmax=="max" | minmax=="min"))
```

```

27     stop("minmax must be filled only with the characters 'min' or 'max'"
28     )
29     if(length(minmax)!= ncol(decision))
30     stop("minmax must have the same length of decision argument columns"
31     )
32
33
34     normcriteria <- sweep(criteria, 2, colSums(criteria), "/")
35     gmean <- apply(normcriteria, 1, function(x){prod(x)^(1/length(x))})
36     weights<-gmean/sum(gmean)
37     maxlambda<- colSums(criteria)%*% weights
38     consistencyindex<-(maxlambda-ncol(criteria))/(ncol(criteria)-1)
39     randomindices<-list("3"= 0.58, "4"= 0.9 , "5"= 1.12, "6"= 1.24, "7"=
40     1.32, "8"= 1.41, "9"= 1.45, "10"= 1.49, "11"=1.51, "12"=1.48, "13"
41     =1.56, "14"=1.57, "15"=1.59)
42     ncriteria<- as.character(ncol(criteria))
43     consistencyratio<-consistencyindex/randomindices[[ncriteria]]
44
45     if (consistencyratio > 0.10)
46     stop("consistency ratio is greater than 10%, you must revise the
47     judgments on criteria matrix")
48
49
50     normdecision1<- sweep(decision,2, apply(decision, 2, function(x) sqrt(
51     sum(x^2))),"/")
52     weidecision1<- sweep(normdecision1, 2, weights, "*")
53     possolution1<-ifelse(minmax=="max", apply(weidecision1, 2, max), apply
54     (weidecision1, 2, min))
55     negsolution1<-ifelse(minmax=="max", apply(weidecision1, 2, min), apply
56     (weidecision1, 2, max))
57
58     distpos1<- apply(weidecision1, 1, function(x) sqrt(sum((possolution1-x
59     )^2)))
60     distneg1<- apply(weidecision1, 1, function(x) sqrt(sum((negsolution1-x
61     )^2)))
62
63     result1<- data.frame(distpos1, distneg1)
64     result1$values<- distneg1/(distpos1+distneg1)
65     result1$ranking<-rank(-result1$values, ties.method = "first")
66     result1$distpos1 <- NULL

```

```

62 result1$distneg1 <- NULL
63
64
65
66
67 colmin<-apply(decision, 2, min)
68 colmax<-apply(decision, 2, max)
69 normdecision2<-sweep(sweep(decision, 2,colmin, "-"), 2, colmax-colmin,
70 "/")
71 weidecision2<- sweep(normdecision2, 2, weights, "*")
72 possolution2<-ifelse(minmax=="max", apply(weidecision2, 2, max), apply
73 (weidecision2, 2, min))
74 negsolution2<-ifelse(minmax=="max", apply(weidecision2, 2, min), apply
75 (weidecision2, 2, max))
76
77
78 distpos2<- apply(weidecision2, 1, function(x) sqrt(sum((x-possolution2
79 )^2)))
80 distneg2<- apply(weidecision2, 1, function(x) sqrt(sum((x-negsolution2
81 )^2)))
82
83
84
85 result2<- data.frame(distpos2, distneg2)
86 result2$values<- distneg2/(distpos2+distneg2)
87 result2$ranking<-rank(-result2$values, ties.method = "first")
88 result2$distpos2 <- NULL
89 result2$distneg2 <- NULL
90
91
92 return(list(consistencyratio, result1, result2))
93
94 }

```

Apêndice A

Contexto de atuação do MME

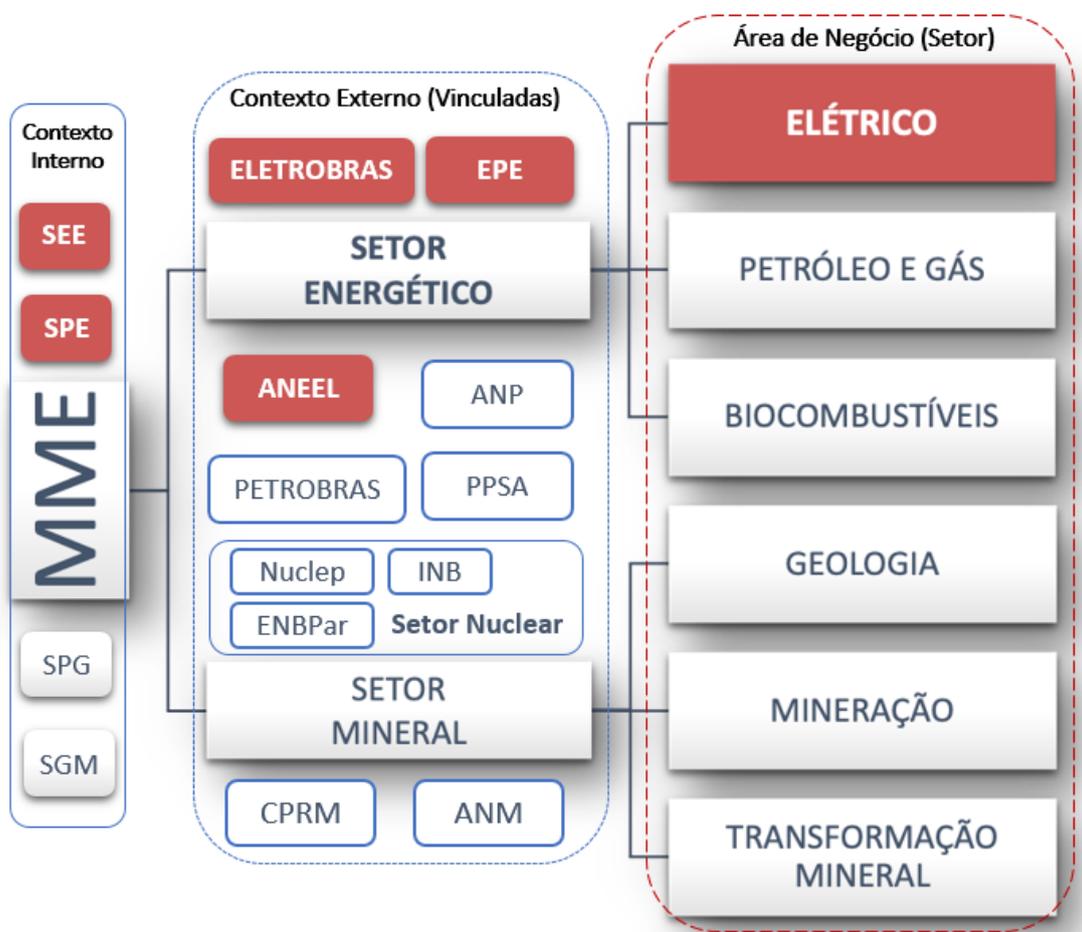


Figura A.1: Contexto de atuação do MME.

Fonte: Elaboração própria.

- Unidades internas
 - SPE: elabora políticas, diretrizes e os planos do setor energético¹, além de coordenar os sistemas de informações energéticas²; e
 - SEE: executa e gerencia as políticas e diretrizes do setor de energia elétrica.

- Unidades externas (vinculadas)
 - Empresa de Pesquisa Energética (EPE): empresa pública que realiza estudos para subsidiar o planejamento do setor energético;
 - Agência Nacional de Energia Elétrica (ANEEL): autarquia especial que, em suma, regulamenta e fiscaliza as políticas do setor elétrico; e
 - Centrais Elétricas Brasileiras S.A. (Eletrobras): empresa brasileira de capital aberto (desestatizada em junho de 2022), que atua como *holding* coordenando empresas subsidiárias de geração, transmissão e distribuição de energia elétrica.

¹Plano Decenal de Expansão de Energia, Plano Nacional de Energia e Balanço Energético Nacional.

²<https://www.mme.gov.br/SIEBRASIL/>.

Apêndice B

Bases Abertas para *Energia Elétrica* catalogadas no PBDA



Figura B.1: Número de bases abertas para *energia elétrica* catalogadas no PBDA.

Fonte: Portal Brasileiro de Dados Abertos^a.

^aPesquisa atualizada em 28 de junho de 2022.

Apêndice C

Catálogo dos Dados Abertos

Quadro 12: Catálogo das bases abertas utilizadas no Modelo.

Fonte: Portal Brasileiro de Dados Abertos - Adaptado.

ID	Conjuntos de Dados Abertos catalogados no PBDA
BD1	Geração Distribuída - Relação de Empreendimentos
BD2	Relação de Empreendimentos de Geração
BD3	Compensação Financeira / Royalties
BD4	Tarifas Residenciais de Energia Elétrica
BD5	Projetos de P&D em Energia Elétrica – temas estratégicos
BD6	Índice ANEEL de Satisfação do Consumidor – Iasc
BD7	Indicadores Quantitativos do Portal da ANEEL
BD8	Reclamações mais registradas pela ANEEL
BD9	Reuniões Públicas da Diretoria da ANEEL
BD10	Reclamações no 1º e 2º nível da Distribuidora
BD11	Audiências e Consultas Públicas
BD12	Autos de infração cadastrados pelas áreas de fiscalização da ANEEL
BD13	Indicadores Quantitativos de Fiscalização da Energia Elétrica
BD14	Fiscalização Econômica e Financeira
BD15	Geração
BD16	Transmissão e Distribuição
BD17	Programa Luz para todos
BD18	Tarifa Social de Energia Elétrica – Custeio
BD19	Tarifa Social de Energia Elétrica – Beneficiários
BD20	Tarifas médias de energia elétrica por classe de consumo e região
BD21	Receita e Consumo – Mercado Cativo
BD22	Projetos Res n. 316/2008, 219/2006 e anteriores
BD23	Projetos por tipologia
BD24	Projetos, Retornos e Investimentos
BD25	Desempenho das concessionárias de transmissão
BD26	Os Dez Maiores Agentes por unidade de consumidores
BD27	Os Dez Maiores Agentes por consumo MWh
BD28	Agentes do mercado
BD29	Estudos de Inventário e projetos aprovados de usinas hidrelétricas
BD30	Outorgas de Geração
BD31	Os Dez Maiores Agentes por receita de faturamento
BD32	Usinas Termelétricas por tipo
BD33	Capacidade Instalada por Unidade da Federação
BD34	Geração Distribuída
BD35	Linhas de transmissão concedidas
BD36	Os Dez Maiores Agentes por Capacidade Instalada
BD37	Acréscimo da potência instalada
BD38	Usinas Termelétricas por tipo
BD39	Geração por fonte
BD40	Compensações pagas aos consumidores
BD41	DEC e FEC Apurado

BD42	Previsão de Entrada em Operação de Novas Usinas
BD43	Sistemas de Informações Energéticas do Brasil
BD44	Programas Luz para Todos e Mais Luz para a Amazônia
BD45	Empreendimento hidrelétrico em estudo
BD46	Sistema de Informações de Geração da ANEEL
BD47	Tarifas de aplicação das distribuidoras de energia elétrica
BD48	Termo de Notificação

Apêndice D

Necessidades e Benefícios da OGD

Quadro 13: Necessidades e Benefícios da OGD demandados pelos *Stakeholders*.

Fonte: Elaboração própria.

<i>Stakeholders</i>					
Necessidades/ Benefícios	Sociedade Civil	Órgãos de Controle	<i>Ombudsman</i> (Ouvidoria)	Cientistas/ Especialistas	Referencial Teórico
Transparência	✓	✓	✓	✓	Molloy [15] e Lourenço [17], Kucera e Chlapek [16]
Controle Social	✓				Lourenço [17]
<i>Accountability</i>		✓			Molloy [15], Kucera e Chlapek [16]
Confiança Pública	-	✓		✓	Kucera e Chlapek [16]
Serviços Públicos	-	✓		✓	Molloy [15] e Kucera e Chlapek [16]
Inovações				✓	Molloy [15], Janssen, Charalabidis e Zuiderwijk [5], Shadbolt et al. [14] e Kucera e Chlapek [16], Attard et al. [128]
Análise Estatísticas	-			✓	Janssen, Charalabidis e Zuiderwijk [5], Kubler et al. [99], Luthfi e Janssen [91], Luthfi et al. [98] e Luthfi, Janssen e Crompvoets [18]

Apêndice E

Critérios de Risco definidos

Quadro 14: Critérios de Risco identificados a partir dos Princípios *Open Data*, LOD e OGD.

Fonte: Elaboração própria.

ID	Critério/Subcritério	Princípio/Fundamentação
Formatação (CR1)		
CR ₁₁	Recursos disponíveis em formato estruturado.	Reutilização e redistribuição. Legível por máquina (<i>machine-readable</i>).
CR ₁₂	Recursos disponíveis em formatos não proprietários.	Participação universal. Dados não proprietário.
CR ₁₃	Quantidade de recursos disponíveis e baixa variedade de formatos.	Participação universal. Dados/-Metadados disponíveis. Disponibilidade e acesso em qualquer formato.
CR ₁₄	Dados em formatos RDF (permitem serialização).	RDF conectado para a descoberta de outros recursos LOD.
Padronização (CR2)		
CR ₂₁	Coerência entre os títulos das bases no PBDA em relação ao PDA.	Obrigação legal: artigo 8º da Resolução n º3/2017.
CR ₂₂	Licença adequada ao conjunto de dados.	Participação universal. Licença Livre/ <i>Open Data</i> .
CR ₂₃	Recurso possui dicionário de dados.	Hospedado e publicado. Metadados Descritivos/Vocabulários.
CR ₂₄	Nível de completude de metadados obrigatórios.	Completos (Metadados obrigatórios disponíveis). Padrão de metadados definidos pela INDA ¹ .
CR ₂₅	Nível de completude de metadados desejáveis.	Completos (Metadados desejáveis disponíveis). Padrão de metadados definidos pela INDA
Atualização (CR3)		
CR3	Dados/Metadados desatualizados.	Atuais. Dados atualizados, pelo menos, dentro da periodicidade definida.
Rejeição (CR4)		
CR ₄₁	<i>Link</i> do <i>dataset</i> indisponível ou descontinuado.	Hospedado e publicado. Disponibilidade e acesso em qualquer formato.
CR ₄₂	Baixa relevância para a sociedade ou <i>stakeholders</i> .	Obrigação legal: Inciso I do artigo 1º da Resolução n º3/2017
CR ₄₃	Número de <i>feedbacks</i> negativos recebidos.	Obrigação legal: parágrafo 1º do artigo 1º da Resolução n º3/2017.
Critério: Aceitação (CR5)		
CR5	Número de <i>feedbacks</i> positivos recebidos.	Obrigação legal: parágrafo 1º do artigo 1º da Resolução n º3/2017.

Apêndice F

Conjunto de Matrizes dos Subcritérios

Tabela F.1: Matriz de Subcritérios em relação ao Critério Formatação.

Formatação	CR11	CR12	CR13	CR14	Pesos
CR11	1	3	7	7	0.5943
CR12	0,33	1	5	3	0,2552
CR13	0,14	0,2	1	2	0,0834
CR14	0,14	0,33	0.5	1	0.0670
Razão de Consistência (RC)					0,06084886

Tabela F.2: Matriz de Subcritérios em relação ao Critério Padronização.

Padronização	CR21	CR22	CR23	CR24	CR25	Pesos
CR21	1	5	3	7	9	0.5223
CR22	0,2	1	2	3	7	0.2031
CR23	0,33	0,5	1	3	6	0,1653
CR24	0.14	0.33	0.33	1	5	0,0799
CR25	0,11	0,14	0.1667	0.2	1	0.0293
Razão de Consistência (RC)					0,0782119	

Tabela F.3: Matriz de Subcritérios em relação ao Critério Rejeição.

Rejeição	CR41	CR42	CR43	Pesos
CR41	1	3	7	0,6491
CR42	0,33	1	5	0,2790
CR43	0,14	0,2	1	0,0719
Razão de Consistência (RC)				0,05593757

Apêndice G

Código R - Ponderação dos Critérios pelo método AHP

```
1 print("Teste Geral:")
2 cat("\n")
3 cat("\n")
4
5 CR1_CR2 =3
6 CR1_CR3 =5
7 CR1_CR4 =1
8 CR1_CR5 =2
9 CR2_CR3 =2
10 CR2_CR4 =0.5
11 CR2_CR5 =0.5
12 CR3_CR4 =0.5
13 CR3_CR5 =0.5
14 CR4_CR5 =2
15
16 RI = 1.12
17
18
19 Criteria_Comparison_Matrix = matrix(
20
21
22 c(1, CR1_CR2, CR1_CR3, CR1_CR4, CR1_CR5,
23    1/CR1_CR2, 1, CR2_CR3, CR2_CR4, CR2_CR5,
24    1/CR1_CR3, 1/CR2_CR3, 1, CR3_CR4, CR3_CR5,
25    1/CR1_CR4, 1/CR2_CR4, 1/CR3_CR4, 1, CR4_CR5,
26    1/CR1_CR5, 1/CR2_CR5, 1/CR3_CR5, 1/CR4_CR5, 1),
27
28
29 nrow = 5,
```

```

30
31     ncol = 5,
32
33     byrow = TRUE
34 )
35
36
37 rownames(Criteria_Comparison_Matrix) = c("Formatacao", "Padronizacao", "
    Atualizacao", "Rejeicao", "Aceitacao")
38 colnames(Criteria_Comparison_Matrix) = c("Formatacao", "Padronizacao", "
    Atualizacao", "Rejeicao", "Aceitacao")
39
40 print("A matriz de comparacao e:")
41 print(Criteria_Comparison_Matrix)
42
43 Matriz_powered = Criteria_Comparison_Matrix%^%10
44
45 cat("\n")
46 cat("\n")
47 print("Resultado da potenciacao da matriz:")
48 print(Matriz_powered)
49
50
51 cat("\n")
52 cat("\n")
53 print("Resultado da soma dos elementos da matriz:")
54 Soma = sum(Matriz_powered)
55 print(Soma)
56
57
58 Criteria_Weights = rowSums(Matriz_powered)/Soma
59 cat("\n")
60 cat("\n")
61 print("Resultado dos pesos dos criterios (vetor de prioridades):")
62 print(Criteria_Weights)
63
64
65 Criterio_Maior_Index = which.max(Criteria_Weights)
66 cat("\n")
67 cat("\n")
68 print("O criterio com maior peso e:")
69 print(Criteria_Weights[Criterio_Maior_Index])
70
71
72 Ws = Criteria_Comparison_Matrix%*%Criteria_Weights

```

```

73 cat("\n")
74 cat("\n")
75 print("O vetor de soma ponderada e:")
76 print(Ws)
77
78 Consis =((1/Criteria_Weights)*%Ws)
79 cat("\n")
80 cat("\n")
81 print("Resultado da soma da divisao do vetor de soma pelo vetor
      prioridade:")
82 print(Consis)
83
84 Lambda = Consis/5
85 cat("\n")
86 cat("\n")
87 print("Logo o lambda e:")
88 print(Lambda)
89
90 CI=(Lambda-5)/(5-1)
91 cat("\n")
92 cat("\n")
93 print("O indice de consistencia e:")
94 print(CI)
95
96 cat("\n")
97 cat("\n")
98 print("A razao de consistencia e")
99 CR = CI/RI
100 print(CR)
101
102 cat("\n")
103 cat("\n")
104 print("Teste Geral:")
105
106 if (CR<0.1) {
107
108     print("Os pesos atribuidos aos criterios sao consistentes")
109
110 } else {
111
112     print("Os pesos atribuidos aos criterios nao sao consistentes")
113
114 }
115
116

```

```

117
118 cat("\n")
119 cat("\n")
120 print("Teste para os subcriterios de CR1:")
121
122
123
124
125 CR11_CR12 =3
126 CR11_CR13 =7
127 CR11_CR14 =7
128 CR12_CR13 =5
129 CR12_CR14 =3
130 CR13_CR14 =2
131
132 RI = 0.9
133
134 Criteria_Comparison_Matrix = matrix(
135
136   c(1, CR11_CR12, CR11_CR13, CR11_CR14,
137     1/CR11_CR12, 1,CR12_CR13, CR12_CR14,
138     1/CR11_CR13,1/CR12_CR13,1,CR13_CR14,
139     1/CR11_CR14,1/CR12_CR14,1/CR13_CR14,1),
140
141
142   nrow = 4,
143
144   ncol = 4,
145
146   byrow = TRUE
147 )
148
149
150 rownames(Criteria_Comparison_Matrix) = c("CR11", "CR12", "CR13", "CR14")
151 colnames(Criteria_Comparison_Matrix) = c("CR11", "CR12", "CR13", "CR14")
152 cat("\n")
153 cat("\n")
154 print("A matriz de comparacao para os subcriterios de CR1 e:")
155 print(Criteria_Comparison_Matrix)
156
157 Matriz_powered = Criteria_Comparison_Matrix%^%10
158
159 cat("\n")
160 cat("\n")
161 print("Resultado da potenciacao da matriz:")

```

```

162 print(Matriz_powered)
163
164
165 cat("\n")
166 cat("\n")
167 print("Resultado da soma dos elementos da matriz:")
168 Soma = sum(Matriz_powered)
169 print(Soma)
170
171
172 Criteria_Weights = rowSums(Matriz_powered)/Soma
173 cat("\n")
174 cat("\n")
175 print("Resultado dos pesos dos criterios:")
176 print(Criteria_Weights)
177
178 Criterio_Maior_Index = which.max(Criteria_Weights)
179 cat("\n")
180 cat("\n")
181 print("O criterio com maior peso e:")
182 print(Criteria_Weights[Criterio_Maior_Index])
183
184 Ws = Criteria_Comparison_Matrix**Criteria_Weights
185 cat("\n")
186 cat("\n")
187 print("O vetor de soma ponderada e:")
188 print(Ws)
189
190 Consis =((1/Criteria_Weights)**Ws)
191 cat("\n")
192 cat("\n")
193 print("Resultado da soma da divisao do vetor de soma pelo vetor
    prioridade:")
194 print(Consis)
195
196 Lambda = Consis/4
197 cat("\n")
198 cat("\n")
199 print("Logo o lambda e:")
200 print(Lambda)
201
202
203 CI=(Lambda-4)/(4-1)
204 cat("\n")
205 cat("\n")

```

```

206 print("O indice de consistencia e:")
207 print(CI)
208
209 cat("\n")
210 cat("\n")
211 print("A razao de consistencia e")
212 CR = CI/RI
213 print(CR)
214
215 cat("\n")
216 cat("\n")
217 print("Teste de Consistencia Subcriterios CR1:")
218
219 if (CR<0.1) {
220
221     print("Os pesos atribuidos aos criterios sao consistentes")
222
223 } else {
224
225     print("Os pesos atribuidos aos criterios nao sao consistentes")
226
227 }
228
229
230
231 cat("\n")
232 cat("\n")
233 print("Teste para os subcriterios de CR2:")
234
235
236 CR21_CR22 =5
237 CR21_CR23 =3
238 CR21_CR24 =7
239 CR21_CR25 =9
240 CR22_CR23 =2
241 CR22_CR24 =3
242 CR22_CR25 =7
243 CR23_CR24 =3
244 CR23_CR25 =6
245 CR24_CR25 =5
246
247 RI = 1.12
248
249 Criteria_Comparison_Matrix = matrix(
250

```

```

251 # Sequencia de elementos da Matriz
252 c(1, CR21_CR22, CR21_CR23, CR21_CR24, CR21_CR25,
253     1/CR21_CR22, 1,CR22_CR23, CR22_CR24, CR22_CR25,
254     1/CR21_CR23,1/CR22_CR23,1,CR23_CR24,CR23_CR25,
255     1/CR21_CR24,1/CR22_CR24,1/CR23_CR24,1,CR24_CR25,
256     1/CR21_CR25,1/CR22_CR25,1/CR23_CR25,1/CR24_CR25,1),
257
258 nrow = 5,
259
260 ncol = 5,
261
262 byrow = TRUE
263 )
264
265 rownames(Criteria_Comparison_Matrix) = c("CR21", "CR22", "CR23", "CR24",
      "CR25")
266 colnames(Criteria_Comparison_Matrix) = c("CR21", "CR22", "CR23", "CR24",
      "CR25")
267
268 cat("\n")
269 cat("\n")
270 print("A matriz de comparacao para os subcriterios de CR2 e:")
271 print(Criteria_Comparison_Matrix)
272
273 Matriz_powered = Criteria_Comparison_Matrix%%10
274
275 cat("\n")
276 cat("\n")
277 print("Resultado da potenciacao da matriz:")
278 print(Matriz_powered)
279
280
281 cat("\n")
282 cat("\n")
283 print("Resultado da soma dos elementos da matriz:")
284 Soma = sum(Matriz_powered)
285 print(Soma)
286
287 Criteria_Weights = rowSums(Matriz_powered)/Soma
288 cat("\n")
289 cat("\n")
290 print("Resultado dos pesos dos criterios:")
291 print(Criteria_Weights)
292
293

```

```

294 Criterio_Maior_Index = which.max(Criteria_Weights)
295 cat("\n")
296 cat("\n")
297 print("O criterio com maior peso e:")
298 print(Criteria_Weights[Criterio_Maior_Index])
299
300 Ws = Criteria_Comparison_Matrix%*%Criteria_Weights
301 cat("\n")
302 cat("\n")
303 print("O vetor de soma ponderada e:")
304 print(Ws)
305 # COnsistency Vector
306 Consis =((1/Criteria_Weights)%*%Ws)
307 cat("\n")
308 cat("\n")
309 print("Resultado da soma da divisao do vetor de soma pelo vetor
      prioridade:")
310 print(Consis)
311
312 Lambda = Consis/5
313 cat("\n")
314 cat("\n")
315 print("Logo o lambda e:")
316 print(Lambda)
317
318 CI=(Lambda-5)/(5-1)
319 cat("\n")
320 cat("\n")
321 print("O indice de consistencia e:")
322 print(CI)
323
324 cat("\n")
325 cat("\n")
326 print("A razao de consistencia e")
327 CR = CI/RI
328 print(CR)
329
330 cat("\n")
331 cat("\n")
332 print("Teste de Consistencia Subcriterios CR2:")
333
334 if (CR<0.1) {
335
336     print("Os pesos atribuidos aos criterios sao consistentes")
337

```

```

338 } else {
339
340     print("Os pesos atribuidos aos criterios nao sao consistentes")
341
342 }
343
344
345
346
347
348
349
350 cat("\n")
351 cat("\n")
352 print("Teste para os subcriterios de CR4:")
353
354 CR31_CR32 =3
355 CR31_CR33 =7
356 CR32_CR33 =5
357
358 RI = 0.58
359
360 Criteria_Comparison_Matrix = matrix(
361
362
363     c(1,CR31_CR32, CR31_CR33,
364       1/CR31_CR32, 1, CR32_CR33,
365       1/CR31_CR33,1/CR32_CR33,1),
366
367
368     nrow = 3,
369
370
371     ncol = 3,
372
373     byrow = TRUE
374 )
375
376
377
378 rownames(Criteria_Comparison_Matrix) = c("CR31", "CR32", "CR33")
379 colnames(Criteria_Comparison_Matrix) = c("CR31", "CR32", "CR33")
380
381 cat("\n")
382 cat("\n")

```

```

383 print("A matriz de comparacao para os subcriterios de CR4 e:")
384 print(Criteria_Comparison_Matrix)
385
386 Matriz_powered = Criteria_Comparison_Matrix%^10
387
388 cat("\n")
389 cat("\n")
390 print("Resultado da potenciacao da matriz:")
391 print(Matriz_powered)
392
393
394 cat("\n")
395 cat("\n")
396 print("Resultado da soma dos elementos da matriz:")
397 Soma = sum(Matriz_powered)
398 print(Soma)
399
400
401 Criteria_Weights = rowSums(Matriz_powered)/Soma
402 cat("\n")
403 cat("\n")
404 print("Resultado dos pesos dos criterios:")
405 print(Criteria_Weights)
406
407
408 Criterio_Maior_Index = which.max(Criteria_Weights)
409 cat("\n")
410 cat("\n")
411 print("O criterio com maior peso e:")
412 print(Criteria_Weights[Criterio_Maior_Index])
413
414 Ws = Criteria_Comparison_Matrix%*Criteria_Weights
415 cat("\n")
416 cat("\n")
417 print("O vetor de soma ponderada e:")
418 print(Ws)
419
420 Consis = ((1/Criteria_Weights)%*Ws)
421 cat("\n")
422 cat("\n")
423 print("Resultado da soma da divisao do vetor de soma pelo vetor
      prioridade:")
424 print(Consis)
425
426 Lambda = Consis/3

```

```

427 cat("\n")
428 cat("\n")
429 print("Logo o lambda e:")
430 print(Lambda)
431
432 CI=(Lambda-3)/(3-1)
433 cat("\n")
434 cat("\n")
435 print("O indice de consistencia e:")
436 print(CI)
437
438 cat("\n")
439 cat("\n")
440 print("A razao de consistencia e")
441 CR = CI/RI
442 print(CR)
443
444 cat("\n")
445 cat("\n")
446 print("Teste de Consistencia Subcriterios CR4:")
447
448
449 if (CR<0.1) {
450
451     print("Os pesos atribuidos aos criterios sao consistentes")
452
453 } else {
454
455     print("Os pesos atribuidos aos criterios nao sao consistentes")
456
457 }
458
459 sink()

```