

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

Ingrid Palma Araújo, Ana C. B. Reis (orientadora)

Departamento de Ciência da Computação  
Universidade de Brasília (UnB)  
Brasília – DF – Brasil

ingrid.palma@unb.org, anacarla@unb.br

**Abstract.** *This study proposes 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 Shiny. 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 results present the performance of each dataset, displaying those that should be improved in relation to the respective risk criteria and prioritized topics.*

**Resumo.** *Este estudo propõe 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. Os resultados apresentam o desempenho de cada conjunto de dados, exibindo aqueles que devem ser aprimorados em relação ao critérios de risco e aos temas priorizados.*

## 1. Contextualização e Relevância da Pesquisa

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 [Eaves 2009].

O Plano de Dados Abertos (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 Portal Brasileiro de Dados Abertos (PBDA), bem como a correção e atualização dessas bases, conforme preconizado pela Estratégia de Governo Digita (EGD) e pela OCDE<sup>1</sup>.

Apesar das ações implementadas pelo PDA, evidência-se, em recente estudo apresentado por [Palma et al. 2021], a crescente demanda por dados e informações afetos

<sup>1</sup><https://www.oecd.org/digital/digital-government/open-government-data.htm>.

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)<sup>2</sup>.

Além do exposto, salienta-se que o processo de abertura de dados é uma atividade complexa, podendo resultar em benefícios (oportunidades) ou em riscos (ameaças) potenciais para as partes interessadas [Martin 2013]. 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 [Ali-Eldin et al. 2017]. 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 *Open Government Data* (OGD) [Luthfi et al. 2018a].

É consenso que OGD são produzidos e/ou coletados por meio da prestação de bens ou serviços públicos. Tal iniciativa 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. 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.

O objetivo geral deste trabalho é propor um modelo de apoio à decisão que permita, à luz dos critérios de riscos relacionados aos princípios (LOD) e *Open Government Data* (OGD), avaliar o nível de abertura das bases do setor elétrico catalogadas no PBDA. Para isso foi preciso definir os critérios de risco atinentes e relevantes, de modo geral e específico ao setor elétrico, identificar os métodos de apoio à decisão aderentes à problemática estudada para, assim, construir o modelo proposto.

## 2. Metodologia

A metodologia desta pesquisa se divide em quatro Fases (Planejamento, Execução, Estudo de Caso e Redacional), que sistematizam a sequência de procedimentos e métodos seguidos neste trabalho, conforme delineado no **Plano Metodológico** representado pela Figura 1. As mencionadas fases e as correspondentes etapas, adaptadas ao contexto deste trabalho, assentam-se aos eixos propostos por [Gerhardt and Silveira 2009] e [Prodanov and De Freitas 2013].

Os seguintes instrumentos e etapas necessários para o alcance dos objetivos desta pesquisa estão descritos a seguir:

- Revisão Sistemática da Literatura (RSL) e aplicação do método Teoria do Enfoque Meta Analítico Consolidado (TEMAC)<sup>3</sup> [Mariano and Santos 2017],

---

<sup>2</sup>O Sistema e-SIC está integrado à Plataforma Integrada de Ouvidoria e Acesso à Informação, ora denominada Fala.BR.

<sup>3</sup><https://www.pesquisatemac.com/>

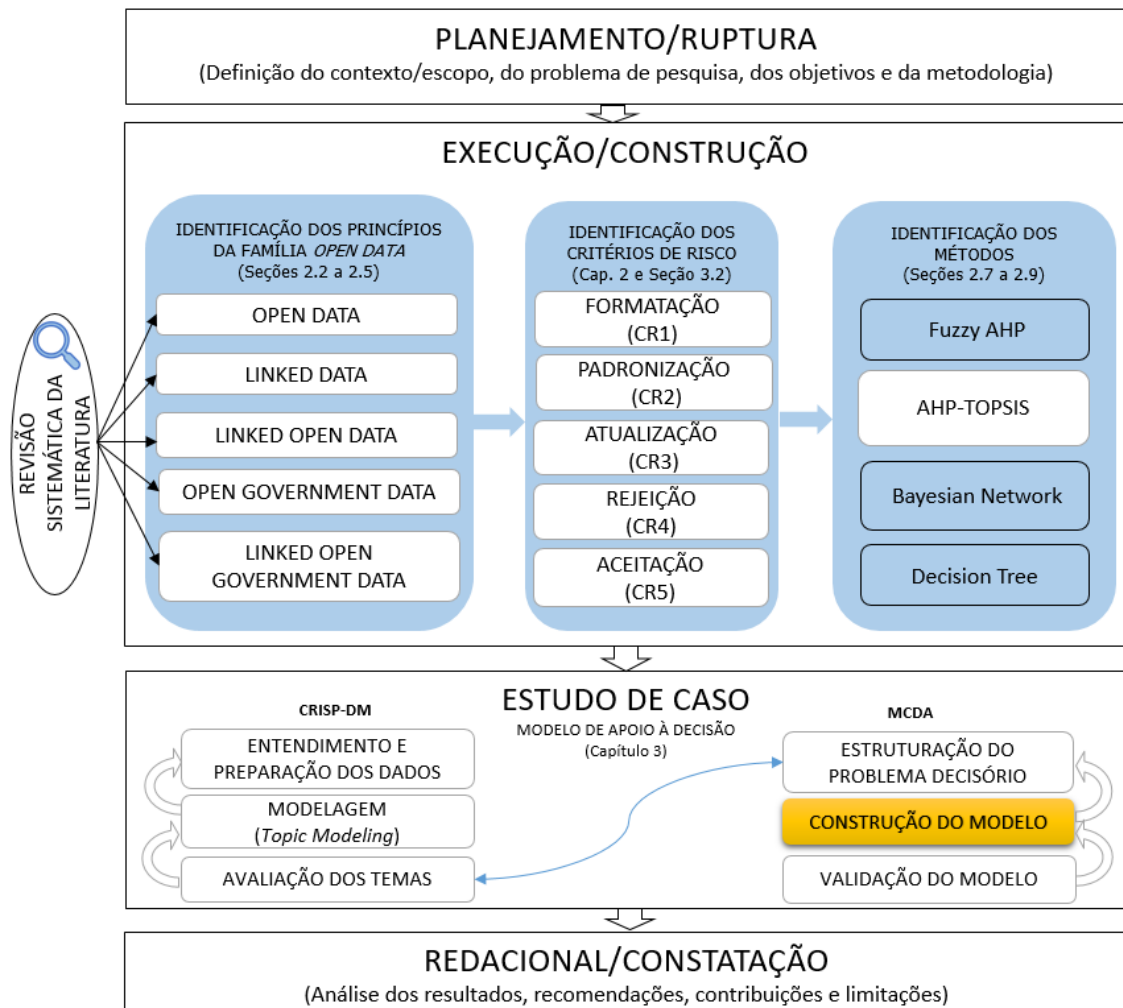


Figura 1. Fases e Etapas da Pesquisa.

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) [Chapman et al. 2000], para **Entendimento e Preparação dos Dados, Modelagem e Avaliação**, 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) [Belton and Stewart 2002], para a **Estruturação do Problema Decisório e Construção do modelo de apoio à decisão** proposto neste trabalho;

- O pacote *Shiny* do Software R Studio<sup>4</sup> para o desenvolvimento de uma Aplicação *Web*<sup>5</sup>, que permite apresentar, de forma interativa e dinâmica, os resultados encontrados para análise e avaliação dos especialistas.

### 3. Trabalhos Relacionados

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

#### 3.1. Métodos de Apoio à Decisão sobre Dados Abertos

Conforme se observa em [Gutjahr and Nolz ], 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 [Worthy 2015], 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 et al. 2016] 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. Tais métodos e o quantitativo de critérios utilizados nos trabalhos relacionados estão descritos no estudo apresentado por [Araújo 2022].

#### Árvore de Decisão e Técnica *Delphi*

[Luthfi et al. 2019] 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, 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*<sup>6</sup> 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 [Belton and Stewart 2002].

---

<sup>4</sup><https://www.rstudio.com/>.

<sup>5</sup><https://ingridpalma.shinyapps.io/ahptosis/>.

<sup>6</sup>A grosso modo, *Payoff* = (Ganho/Prejuízo), ou o resultado apresentado após a execução de um conjunto de ações.

## Redes Bayesianas

Modelos de apoio à decisão, parametrizados em Redes Bayesianas, foram propostos por [Luthfi and Janssen 2017], [Luthfi et al. 2018a], ponderando potenciais riscos e benefícios associados à abertura de dados históricos da saúde de pacientes.

Após revisão da literatura, [Luthfi and Janssen 2017] 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.

## Modelo *Trade-off* Quali-quantitativo

Análise qualitativa de dados de entrevistas semiestruturadas foi realizada por [Buda et al. 2015] e [Luthfi et al. 2018b] 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 and Janssen 2015]. Nesse trabalho, não se identificou a análise quantitativa.

Uma proposta quali-quantitativa para a priorização de bases de dados abertos foi desenvolvida por [Botchkarev 2018]. Ela se centra na avaliação do melhor custo-benefício ajustado aos seguintes critérios: valor, custo, risco e urgência.

### 3.2. Métodos Multicritério sobre Dados Abertos

O trabalho desenvolvido por [Luthfi and Janssen 2019], 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.

### Fuzzy Analytical Hierarchy Process (FAHP)

FAHP, método que faz parte da abordagem MCDA, foi utilizado por [Luthfi et al. 2018c] 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 and Lněnička 2019]. 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.

Nessa linha, [Lyimo et al. 2020] 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. 2016], 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 [Kubler et al. 2016], que muito se aproxima com o escopo desta pesquisa, 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.

Utilizando a mesma abordagem híbrida do AHP-TOPSIS, [Shaikh et al. 2020] 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. 2018] 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. 2020] identificaram os principais fatores de riscos que impactam negativamente o processo de avaliação da qualidade de dados no ambiente DevOps.

### **3.3. 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. 2017], que apresentaram um modelo qualitativo (orientativo-conceitual), discutindo os benefícios dessa abertura e os motivos pelos quais estes devem permanecer fechados.

## **4. Construção do Modelo**

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 2, imprimindo o resultado das alternativas existentes.

#### 4.1. Decomposição Hierárquica do Problema

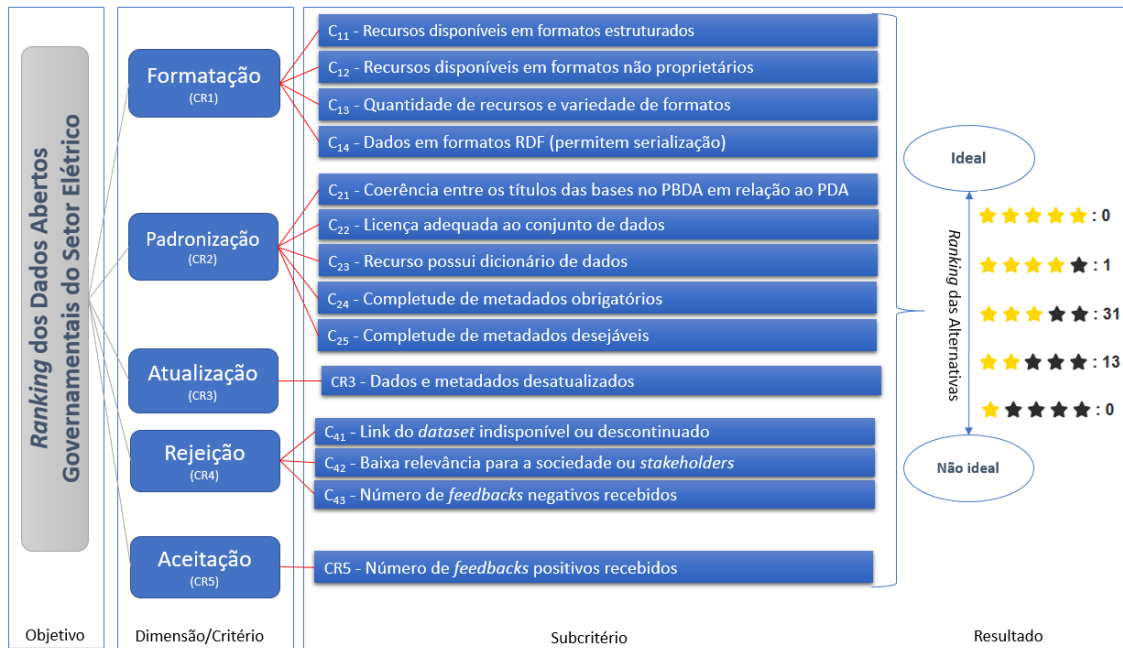


Figura 2. Decomposição Hierárquica do Problema.

#### 4.2. Aplicação do 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 and Yoon 1981], 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 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.

#### 4.3. Ponderação dos Critérios pelo método AHP

O método AHP foi aplicado seguindo a metodologia apresentada por [Saaty 1987, Reis and Schramm 2022], 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/subcritérios 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 [Saaty 1987].

Para realização das comparações par a par, utilizou-se a escala fundamentada por [Saaty 1987] 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*.

### 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 1987], 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).

Os julgamentos iniciais foram organizados na matriz representada na Tabela 1. A partir dessa matriz resultará o vetor de prioridades (pesos dos critérios), necessário para indicar o grau de contribuição de cada critério.

	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

**Tabela 1.** Matriz de Julgamentos.

### 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 [Saaty 1987, Reis and Schramm 2022]. 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 and Schramm 2022] recomendam utilizar o cálculo da matriz limite (potenciação de matrizes), explicitado por [Saaty 1987], 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 1987] propôs o uso do cálculo de  $\lambda_{max}$ , 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  $\lambda_{max}$  resulta da média desse resultado, que é 5,1254.

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

**Tabela 2.** Cálculo do Autovalor Máximo/*Eigen* Principal ( $\lambda_{max}$ )

**Fonte:** Adaptado de [Reis and Schramm 2022].

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 3, 4 e 5 a seguir.

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
<b>Razão de Consistência (RC)</b>			<b>0,06084886</b>		

**Tabela 3.** Matriz de Subcritérios em relação ao Critério Formatação.

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
<b>Razão de Consistência (RC)</b>			<b>0,0782119</b>			

**Tabela 4.** Matriz de Subcritérios em relação ao Critério Padronização.

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
<b>Razão de Consistência (RC)</b>			<b>0,05593757</b>	

**Tabela 5.** Matriz de Subcritérios em relação ao Critério Rejeição.

## 5. Resultados e Discussões

[De Souza et al. 2018] desenvolveram *scripts*<sup>7</sup> do R para aplicação do método híbrido AHP-TOPSIS-2N, implementando 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, 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/>.

Em uma versão adaptada e inspirada no Sistema de 5 estrelas<sup>8</sup>, a tela inicial do *AppShiny* classifica as bases conforme o seu nível de abertura (*openess*) e à luz do padrão *Open Data* e do movimento *OGD*.

A Figura 3 a seguir apresenta os *dashboards* 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>.

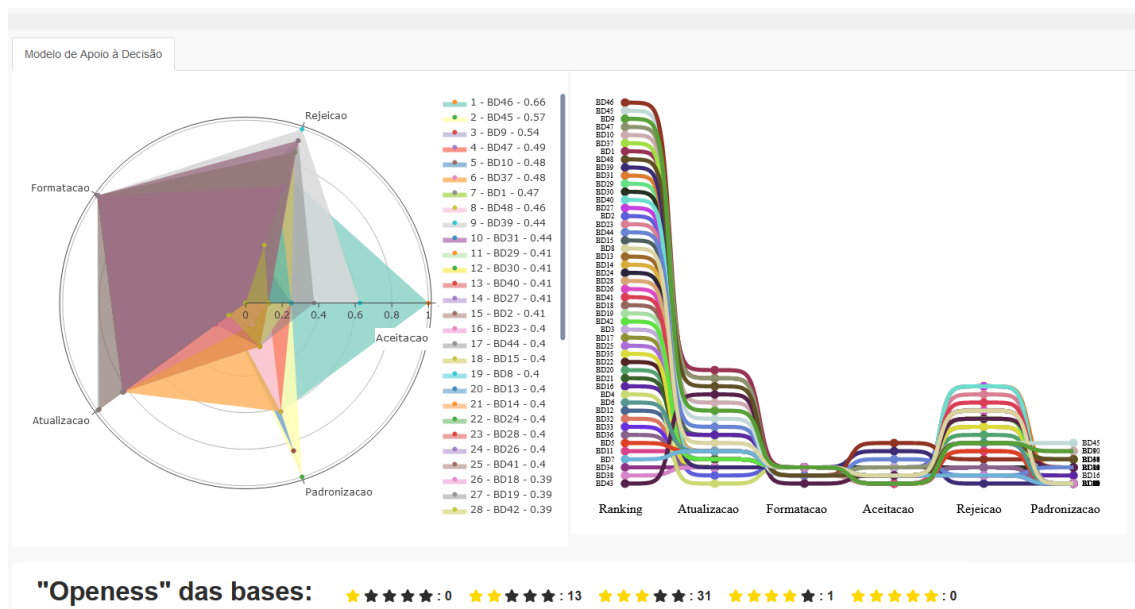


Figura 3. Dashboards projetados pelo *AppShiny*.

<sup>7</sup>[http://cran.rstudio.org/src/contrib/Archive/ahptopsis2n/ahptopsis2n\\_0.1.0.tar.gz](http://cran.rstudio.org/src/contrib/Archive/ahptopsis2n/ahptopsis2n_0.1.0.tar.gz)

<sup>8</sup><https://5stardata.info/pt-BR/>

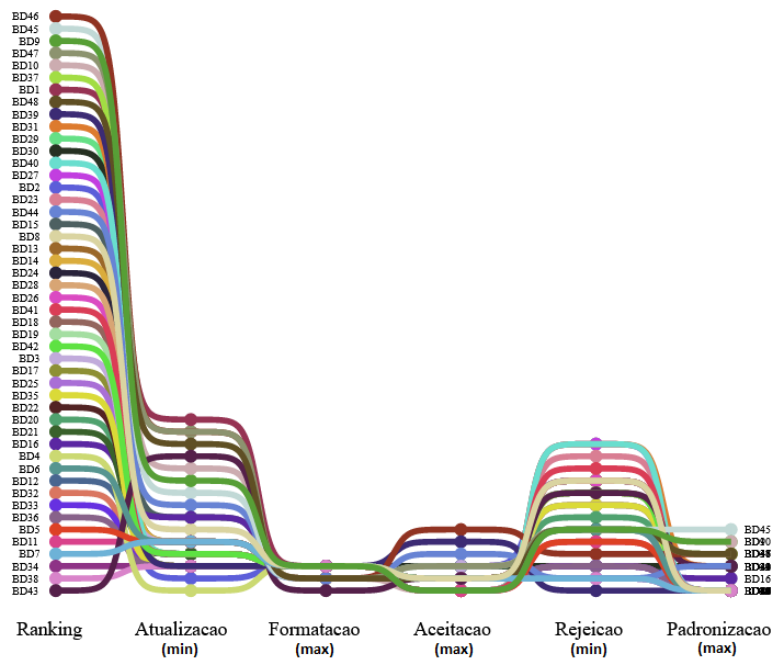


Figura 4. Ranking das 48 Bases.

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.

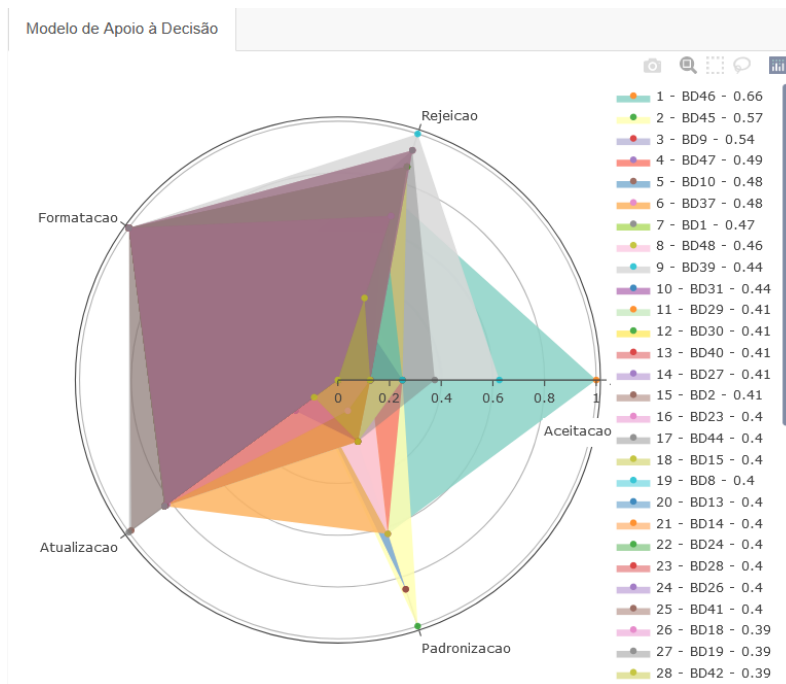


Figura 5. Comparativo das Bases e Impacto dos Critérios de Risco.

O gráfico de radar da Figura 5 apresenta uma outra forma de visualizar e analisar o mesmo *ranking* projetado na Figura 4, sendo possível avaliar os *datasets* em relação ao desempenho médio das alternativas.

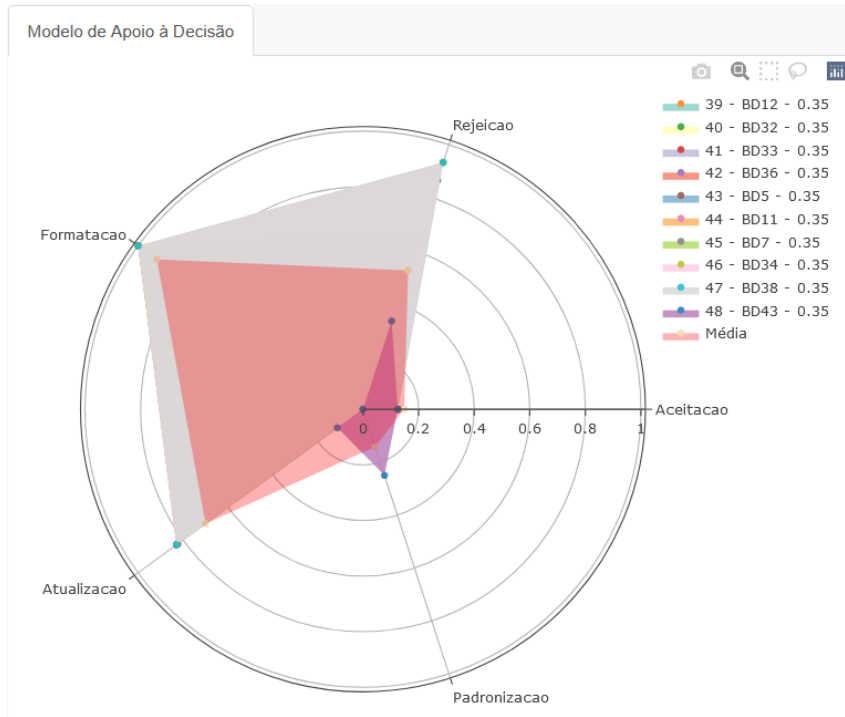


Figura 6. Média de desempenho das bases.

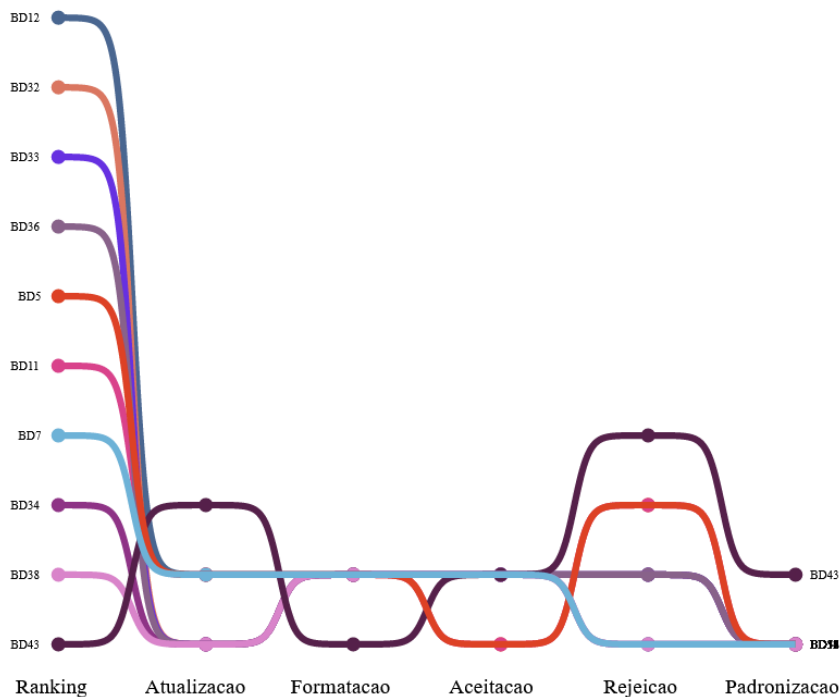
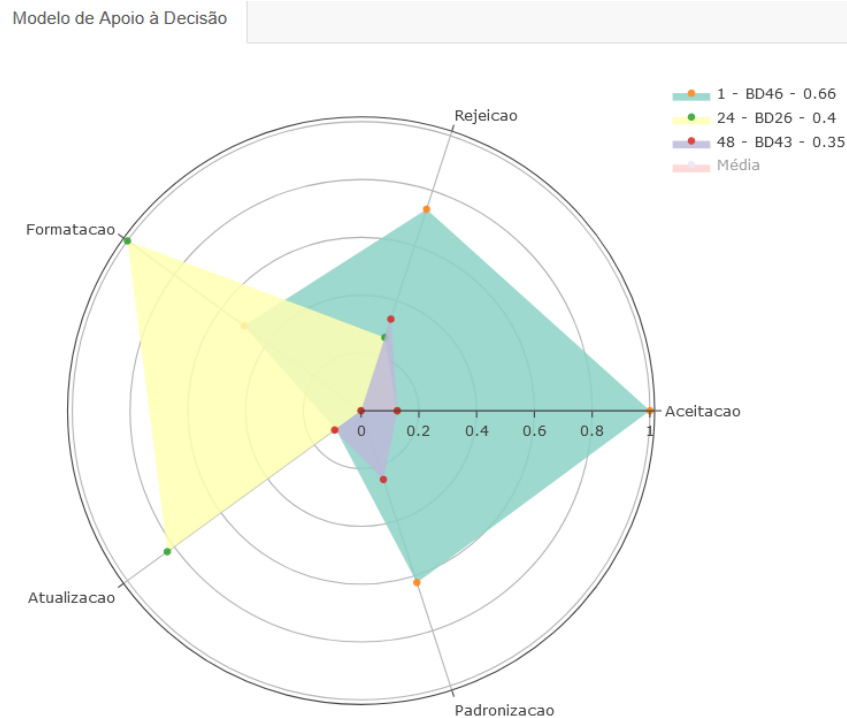


Figura 7. Bases com desempenho abaixo da média.

Os gráficos das Figuras 6 e 7 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 8.** Resultado das Bases DB46, DB26 e DB43.

Recorremos ao *radar chart* para uma análise mais detalhada, e comparativa, conforme plotado na Figura 8. 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).

Foram elegidas as bases localizadas no extremos do vetor virtual do TOPSIS, referente à primeira e última posições do *ranking*, considerando também a base BD26, com valor intermediário entre aquelas duas, e mais próxima da média, expondo o critério de maior impacto em cada uma dessas bases.

## 6. Considerações Finais

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 e *insights* (percepções) a partir da configuração efetuada ao aplicativo, tais como análises de custo-benefício (risco-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 [Zopounidis and Doumpos 2017]; e
- e) Integra de forma eficaz os critérios de riscos às preferências dos decisores.

### Referências

- Ali-Eldin, A., Zuiderwijk-van Eijk, A., and Janssen, M. (2017). Opening more data: A new privacy risk scoring model for open data. In *Proceedings of the 7th International Symposium on Business Modeling and Software Design 2017*.
- Araújo, I. P. (2022). Modelo de apoio à decisão para avaliar dados governamentais abertos do setor elétrico.
- Belton, V. and Stewart, T. J. (2002). *Multiple Criteria Decision Analysis*. Springer US.
- Botchkarev, A. (2018). Towards a methodology of multi-criteria prioritization of open data for public release.
- Buda, A., Ubacht, J., Janssen, M., and Sips, R.-J. (2015). Decision support framework for opening business data.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., et al. (2000). Crisp-dm 1.0: Step-by-step data mining guide. *SPSS inc*, 9:13.
- De Souza, L. P., Gomes, C. F. S., and De Barros, A. P. (2018). Implementation of new hybrid ahp-topsis-2n method in sorting and prioritizing of an it capex project portfolio. *International Journal of Information Technology & Decision Making*, 17(04):977–1005.
- Eaves, D. (2009). The three laws of open government data.
- Gerhardt, T. E. and Silveira, D. T. (2009). *Métodos de pesquisa*. Plageder.
- Gutjahr, W. J. and Nolz, P. C. Multicriteria optimization in humanitarian aid. 252(2):351–366.

- Hossain, M. A., Dwivedi, Y. K., and Rana, N. P. (2016). State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of organizational computing and electronic commerce*, 26(1-2):14–40.
- Hwang, C. and Yoon, K. (1981). Multiple attributes decision making methods and applications, spring. *New York*.
- Kubler, S., Robert, J., Le Traon, Y., Umbrich, J. r., and Neumaier, S. (2016). Open data portal quality comparison using ahp. In *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*, pages 397–407. ACM, Association for Computing Machinery.
- Luthfi, A. and Janssen, M. (2017). A conceptual model of decision-making support for opening data. In *International Conference on e-Democracy*, pages 95–105. Springer Verlag.
- Luthfi, A. and Janssen, M. (2019). A comparative study of methods for deciding to open data. In *International Symposium on Business Modeling and Software Design*, pages 213–220. Springer.
- Luthfi, A., Janssen, M., and Cromptvoets, J. (2018a). *A Causal Explanatory Model of Bayesian-belief Networks for Analysing the Risks of Opening Data*, pages 289–297. Springer Verlag.
- Luthfi, A., Janssen, M., and Cromptvoets, J. (2018b). A framework for analyzing how governments open their data: Institution, technology, and process aspects influencing decision-making. *EGOV-CeDEM-ePart 2018*, page 163.
- Luthfi, A., Janssen, M., and Cromptvoets, J. (2019). Decision tree analysis for estimating the costs and benefits of disclosing data. In *Conference on e-Business, e-Services and e-Society*, pages 205–217. Springer.
- Luthfi, A., Rehena, Z., Janssen, M., and Cromptvoets, J. (2018c). A fuzzy multi-criteria decision making approach for analyzing the risks and benefits of opening data. In *Conference on e-Business, e-Services and e-Society*, pages 397–412. Springer Verlag.
- Lyimo, N. N., Shao, Z., Ally, A. M., Twumasi, N. Y. D., Altan, O., and Sanga, C. A. (2020). A fuzzy logic-based approach for modelling uncertainty in open geospatial data on landfill suitability analysis. *ISPRS International Journal of Geo-Information*, 9(12):737.
- Máchová, R. and Lněnička, M. (2019). A multi-criteria decision making model for the selection of open data management systems. *Electronic Government, an International Journal*, 15(4):372–391.

- Mariano, A. M. and Santos, M. (2017). Revisão da literatura: apresentação de uma abordagem integradora. 18:427–442.
- Martin, S. (2013). Risk analysis to overcome barriers to open data. *Electronic Journal of e-Government*, 11(2):pp348–359.
- Palma, I., Ladeira, M., and Reis, A. C. B. (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*, pages 76–81.
- Parung, G. A., Hidayanto, A. N., Sandhyaduhita, P. I., Ulo, K. L. M., and Phusavat, K. (2018). Barriers and strategies of open government data adoption using fuzzy ahp-topsis: A case of indonesia. *Transforming Government: People, Process and Policy*.
- Pfenninger, S., DeCarolis, J., Hirth, L., Quoilin, S., and Staffell, I. (2017). The importance of open data and software: Is energy research lagging behind? *Energy Policy*, 101:211–215.
- Prodanov, C. C. and De Freitas, E. C. (2013). Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico - 2<sup>a</sup> edição.
- Rafi, S., Yu, W., Akbar, M. A., Alsanad, A., and Gumaei, A. (2020). Multicriteria based decision making of devops data quality assessment challenges using fuzzy tophis. *IEEE Access*, 8:46958–46980.
- Reis, A. C. B. and Schramm, V. B. (2022). Guia para aplicação da análise multicritério em análise de impacto regulatório (air) no inmetro. *Projeto de Melhoria da Qualidade Regulatória - PN 15.2099.8-019.00*, pages 21–30.
- Saaty, R. W. (1987). The analytic hierarchy process—what it is and how it is used. *Mathematical modelling*, 9(3-5):161–176.
- Shaikh, S. A., Memon, M. A., Prokop, M., and Kim, K.-s. (2020). An ahp/topsis-based approach for an optimal site selection of a commercial opening utilizing geospatial data. In *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 295–302.
- Worthy, B. (2015). The impact of open data in the uk: Complex, unpredictable, and political. 93(3):788–805.
- Zopounidis, C. and Doumpos, M. (2017). *Multiple Criteria Decision Making*. Springer.
- Zuiderwijk, A. and Janssen, M. (2015). Towards decision support for disclosing data: Closed or open data? *Information Polity*, 20(2, 3):103–117.