

CARLOS HENRIQUE TAVARES BRUMATTI

UM ARCABOUÇO COMPUTACIONAL PARA SUPORTE À TOMADA  
DE DECISÕES DE ESPECIALISTAS EM CENÁRIOS DE DESASTRES  
AMBIENTAIS UTILIZANDO DADOS PÚBLICOS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

Orientador: Julio Cesar Soares dos Reis

Coorientador: Jugurta Lisboa Filho

VIÇOSA - MINAS GERAIS

2023

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

B893a  
2023  
Brumatti, Carlos Henrique Tavares, 1996-  
Um arcabouço computacional para suporte à tomada de  
decisões de especialistas em cenários de desastres ambientais  
utilizando dados públicos / Carlos Henrique Tavares Brumatti. –  
Viçosa, MG, 2023.

1 dissertação eletrônica (75 f.): il. (algumas color.).

Inclui apêndices.

Orientador: Julio Cesar Soares dos Reis.

Dissertação (mestrado) - Universidade Federal de Viçosa,  
Departamento de Informática, 2023.

Referências bibliográficas: f. 65-68.

DOI: <https://doi.org/10.47328/ufvbbt.2024.211>

Modo de acesso: World Wide Web.

1. Mineração de dados (Computação). 2. Desastres  
ambientais - Sistemas de informação geográfica. I. Reis, Julio  
Cesar Soares dos, 1988-. II. Universidade Federal de Viçosa.  
Departamento de Informática. Programa de Pós-Graduação em  
Ciência da Computação. III. Título.

CDD 22. ed. 005.74


CARLOS HENRIQUE TAVARES BRUMATTI

**UM ARCABOUÇO COMPUTACIONAL PARA SUPORTE À  
TOMADA DE DECISÕES DE ESPECIALISTAS EM CENÁRIOS DE  
DESASTRES AMBIENTAIS UTILIZANDO DADOS PÚBLICOS**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.


APROVADA: 18 de dezembro de 2023.

Assentimento:

Documento assinado digitalmente  
 CARLOS HENRIQUE TAVARES BRUMATTI  
Data: 18/12/2023 16:26:10-0300  
Verifique em <https://validar.iti.gov.br>

---

Carlos Henrique Tavares Brumatti  
Autor

Documento assinado digitalmente  
 JULIO CESAR SOARES DOS REIS  
Data: 18/06/2024 18:07:44-0300  
verifique em <https://validar.iti.gov.br>

---

Julio Cesar Soares dos Reis  
Orientador

Dedico este trabalho a Deus e à minha mãe, por serem essenciais em minha vida.

# Agradecimentos

Gostaria de expressar minha sincera gratidão a todas as pessoas que contribuíram para a realização desta dissertação de mestrado. Este trabalho não teria sido possível sem o apoio, orientação e colaboração de muitas pessoas incríveis. Primeiramente, quero agradecer a Deus, por ter me dado saúde, força de vontade e persistência, para chegar até aqui. Desde o primeiro dia de aula da graduação até hoje, foram inúmeras as dificuldades enfrentadas, mas com muita fé foram superadas. Agradeço também ao meu orientador e ao meu coorientador, Julio Cesar Soares dos Reis e Jugurta Lisboa Filho, respectivamente. A orientação experiente de ambos, a paciência, o incentivo e o apoio contínuo ao longo deste processo foram essenciais para eu estar aqui hoje. Gostaria de agradecer ainda à minha mãe, Angela Maria Brumatti, que sempre reconheceu a importância dos estudos na vida de uma pessoa. Seu constante incentivo e apoio incondicional foram o meu alicerce durante este desafio acadêmico. Agradeço também ao meu amigo, Paulo Henrique de Barros Miguel, que na graduação e no mestrado, sempre ajudou em minha jornada educacional. Por fim, agradeço a todos que de alguma contribuíram com a minha jornada até o dia de hoje.

# Resumo

BRUMATTI, Carlos Henrique Tavares, M.Sc., Universidade Federal de Viçosa, junho de 2021. **Um Arcabouço Computacional para Suporte à Tomada de Decisões de Especialistas em Cenários de Desastres Ambientais utilizando Dados Públicos.** Orientador: Julio Cesar Soares dos Reis. Coorientador: Jugurta Lisboa Filho.

A exploração exagerada dos recursos naturais, aliada com as alterações climáticas dos últimos anos, vem causando um aumento na frequência da ocorrência de eventos climáticos extremos. Conseqüentemente, desastres ambientais, causados de forma natural ou humana, são cada vez mais comumente identificados. Esse aumento na ocorrência gera uma série de conseqüências significativas para as comunidades locais, gerando impactos que são refletidos em diversos setores da sociedade como economia, infraestrutura, saúde e educação. Dessa forma, a recuperação das localidades no pós-evento normalmente é bem custosa financeiramente para os setores públicos e privados. Sendo assim, existe ainda a necessidade de que tais valores sejam alocados da maneira mais eficiente possível. Baseado nessas premissas, este trabalho propõe uma metodologia voltada para suportar a recuperação dessas localidades atingidas, de forma clara e eficiente. Para isso, são utilizados dados públicos atrelados a conceitos de Mineração de Dados e Geoprocessamento. Por fim, foi implementado também um Sistema de Inteligência Geográfica, buscando assim facilitar o contato dos usuários com a abordagem desenvolvida, uma vez que o sistema é destinado aos usuários responsáveis pela alocação de recursos financeiros destinados à recuperação das localidades atingidas por desastres ambientais, e que não estão familiarizados muitas vezes com os conceitos aqui empregados (i.e., Mineração de Dados e Geoprocessamento). Em suma, os resultados destacam o potencial das ideias apresentadas para suportar a tomada de decisão nesse cenário e a viabilidade do sistema proposto. Espera-se que ele possa ser útil para direcionar os investimentos de recuperação pelos usuários.

**Palavras-chave:** Mineração de Dados. Sistema de Inteligência Geográfica. Desastres Ambientais.

# Abstract

BRUMATTI, Carlos Henrique Tavares, M.Sc., Universidade Federal de Viçosa, June, 2021. **A Computational Framework to Support Expert Decision Making in Environmental Disaster Scenarios Using Public Data.** Advisor: Julio Cesar Soares dos Reis. Co-advisor: Jugurta Lisboa Filho

The excessive exploration of natural resources, together with climate change in recent years, has caused an increase in the frequency of extreme weather events. Consequently, environmental disasters, whether caused naturally or humanly, are increasingly more commonly identified. This increase in occurrence generates a series of significant consequences for local communities, generating impacts that are reflected in different sectors of society such as economy, infrastructure, health and education. Therefore, post-event recovery of locations is normally very costly financially for the public and private sectors. Therefore, there is still a need for such values to be allocated in the most efficient way possible. Based on these premises, this work seeks to present a methodology aimed at supporting the recovery of these affected locations, in a clear and efficient way. To achieve this, it is planned to use public data linked to Data Mining and Geoprocessing concepts. Finally, a Geographic Intelligence System was also implemented, seeking to facilitate users' contact with the developed tool, since the system is aimed at users responsible for allocating financial resources intended for the recovery of locations affected by environmental disasters, and who are often not familiar with the concepts used here (i.e. Data Mining and Geoprocessing). In short, the results highlight the potential of the ideas presented to support decision-making in this scenario and the viability of the proposed system. It is expected that it can be useful in directing recovery investments by users.

**Keywords:** Data Mining. Geographic Intelligence System. Environmental Disasters.

# Lista de ilustrações

Figura 1	– Caminho percorrido pela lama, de Mariana(MG) até Linhares (ES). . . . .	15
Figura 2	– Etapas do KDD. . . . .	17
Figura 3	– Etapas do Pré-Processamento dos dados. . . . .	18
Figura 4	– Tarefas da Mineração de Dados. . . . .	19
Figura 5	– Relação entre as tarefas da MD com o ML. . . . .	20
Figura 6	– Quatro níveis da metodologia CRISP-DM. . . . .	20
Figura 7	– Fases do modelo CRISP-DM. . . . .	21
Figura 8	– Diagrama geral da metodologia de trabalho proposta. . . . .	25
Figura 9	– Detalhamento dos quatro níveis da metodologia CRISP-DM. . . . .	25
Figura 10	– Metodologia adotada para a execução deste trabalho. . . . .	29
Figura 11	– Gráfico obtido do Método <i>Elbow</i> considerando a reta traçada entre os pontos inicial e final. . . . .	30
Figura 12	– Exemplo dos valores de cada um dos centróides. . . . .	31
Figura 13	– Mapa de calor em relação a cada uma das propriedades do CAR e IBGE para cada um dos agrupamentos formados. . . . .	32
Figura 14	– Arquitetura geral da aplicação proposta. . . . .	33
Figura 15	– Sequência de telas da aplicação proposta. . . . .	34
Figura 16	– Tela inicial de <i>login</i> do sistema. . . . .	36
Figura 17	– Cadastro de um novo usuário no sistema. . . . .	36
Figura 18	– Inserção dos valores para aplicação da técnica de Agrupamento. . . . .	37
Figura 19	– Seleção dos atributos para serem considerados no Algoritmo <i>KMeans</i> . . . . .	37
Figura 20	– Sequência de passos realizados pelo <i>backend</i> do sistema. . . . .	37
Figura 21	– Interface final do sistema. . . . .	38
Figura 22	– Mapa de calor demonstrando a importância de cada atributo selecionado em cada um dos agrupamentos . . . . .	38
Figura 23	– Conjunto das funcionalidades desenvolvidas. . . . .	39
Figura 24	– Área do desastre de Brumadinho. . . . .	41
Figura 25	– Bacia Hidrográfica do Rio Paraopeba. . . . .	41
Figura 26	– Configuração inicial do sistema. . . . .	42
Figura 27	– Tela com todos os atributos do arquivo <i>csv</i> após <i>upload</i> . . . . .	43
Figura 28	– Seleção dos atributos de interesse pelo usuário. . . . .	44
Figura 29	– Disponibilização dos resultados após aplicação das técnicas pelo <i>backend</i> do sistema. . . . .	45
Figura 30	– Mapa de calor gerado. . . . .	45
Figura 31	– Definição manual do número de agrupamentos. . . . .	46
Figura 32	– Tela principal do sistema. . . . .	47
Figura 33	– Mapa de calor gerado para o exemplo definido. . . . .	47
Figura 34	– Gráfico gerado com as respostas do Pré-Teste . . . . .	49
Figura 35	– Minha interação com a ferramenta <i>TerrainIntelGIS</i> foi clara e compreensível. . . . .	50
Figura 36	– Interagir com a ferramenta <i>TerrainIntelGIS</i> não exige muito do meu esforço mental. . . . .	50
Figura 37	– Considero a ferramenta <i>TerrainIntelGIS</i> fácil de usar. . . . .	51
Figura 38	– Eu acho fácil conseguir que a ferramenta <i>TerrainIntelGIS</i> faça o que eu quero que ela faça. . . . .	51

Figura 39 – Utilizar a ferramenta <i>TerrainIntelGIS</i> pode melhorar o meu desempenho na atividade de agrupar cidades da região de interesse. . . . .	52
Figura 40 – Usar a ferramenta <i>TerrainIntelGIS</i> no meu trabalho pode melhorar a minha produtividade na atividade de agrupar cidades da região de interesse. . . . .	52
Figura 41 – Usar a ferramenta <i>TerrainIntelGIS</i> pode aumentar a minha eficácia no meu trabalho. . . . .	53
Figura 42 – Eu considero a ferramenta <i>TerrainIntelGIS</i> será útil para agrupar cidades de uma mesma região de interesse. . . . .	53
Figura 43 – Se eu tiver acesso à ferramenta <i>TerrainIntelGIS</i> , eu pretendo utilizá-la.	54
Figura 44 – Levando em conta que poderei escolher usar ou não a ferramenta <i>TerrainIntelGIS</i> , eu prevejo que irei utilizá-la. . . . .	54
Figura 45 – Eu penso que gostaria de usar esse sistema com frequência. . . . .	55
Figura 46 – Eu achei o sistema desnecessariamente complexo. . . . .	55
Figura 47 – Eu achei o sistema fácil de usar. . . . .	56
Figura 48 – Eu acho que precisaria de ajuda de uma pessoa com conhecimentos técnicos para usar o sistema. . . . .	56
Figura 49 – Eu achei que as várias funções do sistema estão bem integradas. . . . .	57
Figura 50 – Eu achei que existem inconsistências no sistema. . . . .	57
Figura 51 – Eu imagino que a maioria das pessoas aprenderiam a usar o sistema rapidamente. . . . .	58
Figura 52 – Eu achei o sistema desajeitado de usar. . . . .	58
Figura 53 – Eu me senti bastante confiante ao usar o sistema. . . . .	59
Figura 54 – Eu precisei aprender várias coisas novas antes de conseguir usar o sistema.	59

# Lista de tabelas

Tabela 1	–	Resumo do trabalho segundo a metodologia CRISP-DM. . . . .	26
Tabela 2	–	Relação dos dados coletados do IBGE. . . . .	27
Tabela 3	–	Relação dos dados coletados do CAR. . . . .	27
Tabela 4	–	Exploração inicial do conjunto de dados utilizado. . . . .	27
Tabela 5	–	Relação dos pontos do gráfico da Figura 11 com a distância calculada até a reta traçada entre os pontos inicial e final. . . . .	29
Tabela 6	–	Parâmetros utilizados no Agrupamento com o algoritmo <i>KMeans</i> . . . . .	31
Tabela 7	–	Exemplo do formato do arquivo <i>csv</i> necessário para <i>upload</i> . . . . .	34
Tabela 8	–	Relação dos atributos selecionados para o exemplo. . . . .	43
Tabela 9	–	Resultados obtidos após a realização do exemplo 01. . . . .	46
Tabela 10	–	Resultados obtidos após a realização do exemplo 02. . . . .	48
Tabela 11	–	Exemplo do questionário pré-teste aplicado. . . . .	58

# Lista de abreviaturas e siglas

AI	Aprendizado Indutivo
BHRP	Bacia Hidrográfica do Rio Paraopeba
CAR	Cadastro Ambiental Rural
CNN	<i>Convolutional Neural Networks</i>
CRISP-DM	<i>Cross-Industry Standard Process of Data Mining</i>
CSS	<i>Cascading Style Sheet</i>
CSV	<i>Comma Separated Values</i>
DIMS	<i>Disaster Information Management Systems</i>
DPS	Departamento de Solos
GIS	<i>Geographic Information System</i>
HTML	<i>HyperText Markup Language</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
KDD	<i>Knowledge Discovery in Database</i>
MD	Mineração de Dados
ML	<i>Machine Learning</i>
MPA	<i>Multiple Page Application</i>
MPMG	<i>Ministério Público de Minas Gerais</i>
ODS	Objetivos de Desenvolvimento Sustentável
OGC	<i>Open Geospatial Consortium</i>
ONU	Organização das Nações Unidas
PCA	<i>Principal Component Analysis</i>
PDF	<i>Portable Document Format</i>
QGIS	Quantum GIS
RMSD	<i>Root Mean Square Deviation</i>
SAS	<i>System Acceptance Scale</i>
SIG	Sistema de Inteligência Geográfica
SUS	<i>System Usability Scale</i>

SVM	<i>Support Vector Machine</i>
UFV	Universidade Federal de Viçosa
VGI	<i>Volunteered Geographic Information</i>
WFS	<i>Web Feature Service</i>
WMS	<i>Web Map Service</i>

# Sumário

<b>1</b>	<b>Introdução</b>	<b>14</b>
1.1	Objetivos	16
1.1.1	Objetivos Gerais	16
1.1.2	Objetivos Específicos	16
1.1.3	Organização do Trabalho	16
<b>2</b>	<b>Fundamentação Teórica</b>	<b>17</b>
2.1	Processo de Descoberta de Conhecimento em Bases de Dados	17
2.1.1	Seleção	17
2.1.2	Pré-Processamento	18
2.1.3	Transformação	18
2.1.4	Mineração de Dados	18
2.1.5	Avaliação	19
2.2	Mineração de Dados	19
2.3	Aprendizado de Máquina	19
2.4	Processo Padrão Intersetorial para Mineração de Dados	20
2.5	Algoritmo <i>KMeans</i>	21
2.6	Desastres Ambientais	22
<b>3</b>	<b>Trabalhos Relacionados</b>	<b>23</b>
3.1	Lacuna de Pesquisa	24
<b>4</b>	<b>Metodologia</b>	<b>25</b>
4.1	Etapa 1 - Compreensão do Negócio	26
4.2	Etapa 2 - Compreensão dos Dados	26
4.3	Etapa 3 - Preparação dos Dados	28
4.4	Etapa 4 - Modelagem	28
4.4.1	Método <i>Elbow</i>	28
4.4.1.1	Ponto de maior distância como valor ideal do número de agrupamentos K	28
4.4.2	<i>Principal Component Analysis (PCA)</i>	29
4.4.3	Método de Agrupamento	30
4.5	Etapa 5 - Avaliação	30
4.6	Etapa 6 - Implantação	31
4.6.1	Arquitetura Geral Proposta	32
4.6.2	Tecnologias Utilizadas	32
4.6.3	Modelo dos Arquivos como Fonte de Entrada	34
<b>5</b>	<b>Resultados</b>	<b>35</b>
5.1	<i>TerrainIntelGIS</i>	35
5.1.1	Plataforma <i>TerrainIntelGIS</i>	35
5.1.2	Funcionalidades Implementadas	38
5.1.2.1	Gerar Relatório no formato <i>Portable Document Format - PDF(1)</i>	38
5.1.2.2	Médias dos Atributos (2)	39
5.1.2.3	Cidades por Agrupamentos (3)	39
5.1.2.4	Delimitação da Região de Interesse (4)	39
5.1.2.5	Relação Cor e Agrupamento (5)	40
5.1.2.6	Funcionalidades Internas da Biblioteca de Mapas Interativos	40

5.2	Caso de Estudo: Configuração da Plataforma <i>TerrainIntelGIS</i> para a Região da BHRP, área afetada pelo desastre de Brumadinho . . . . .	40
5.2.1	Desastre de Brumadinho . . . . .	40
5.2.2	Área de Estudo . . . . .	41
5.2.3	Fonte dos Dados . . . . .	42
5.2.4	Configurando a plataforma <i>TerrainIntelGIS</i> . . . . .	42
5.2.4.1	Cenário 01: A plataforma definindo o número de agrupamentos de forma automatizada . . . . .	42
5.2.4.1.1	Conclusões do Exemplo 01 . . . . .	43
5.2.4.2	Cenário 02: O usuário informando o número de agrupamentos . . . . .	45
5.2.4.2.1	Conclusões do Exemplo 02 . . . . .	46
5.3	Avaliação Preliminar do Sistema . . . . .	48
5.3.1	Metodologia . . . . .	48
5.3.2	Resultados Obtidos . . . . .	49
5.3.2.1	Questionário Pré-Teste . . . . .	49
5.3.2.2	Questionário Pós-Teste . . . . .	49
5.3.2.2.1	Questionário SAS . . . . .	49
5.3.2.2.2	Questionário SUS . . . . .	49
5.3.3	Metodologia . . . . .	54
5.3.4	Resultados Obtidos . . . . .	57
5.3.4.1	Questionário Pré-Teste . . . . .	57
5.3.4.2	Questionário Pós-Teste . . . . .	57
5.3.4.2.1	Questionário SAS . . . . .	58
5.3.4.2.2	Questionário SUS . . . . .	59
<b>6</b>	<b>Conclusão . . . . .</b>	<b>62</b>
6.1	Contribuições Científicas . . . . .	62
6.2	Potenciais Impactos Sociais . . . . .	63
6.3	Trabalhos Futuros . . . . .	64
	<b>Referências . . . . .</b>	<b>65</b>
	<b>Apêndices . . . . .</b>	<b>69</b>
	<b>APÊNDICE A Termo de Consentimento . . . . .</b>	<b>70</b>
	<b>APÊNDICE B Tarefas solicitadas aos usuários ao longo do teste. . . . .</b>	<b>72</b>
	<b>APÊNDICE C Avaliação <i>System Acceptance Scale</i> (SAS) . . . . .</b>	<b>74</b>
	<b>APÊNDICE D Avaliação <i>System Usability Scale</i> (SUS) . . . . .</b>	<b>75</b>

# 1 Introdução

O ano de 2023 ficou marcado por uma série de desastres ambientais que ecoaram pelo mundo, despertando preocupações e chamando a atenção para a urgência das ações em prol da preservação do meio ambiente. Desde eventos climáticos extremos, como tempestades<sup>1</sup> e ciclones<sup>2</sup>, até crises de poluição, causadas por rejeitos de mineração<sup>3</sup> ou acidentes ferroviários<sup>4</sup>, tais ocorrências deixaram claro que as mudanças climáticas e a degradação ambiental são desafios cada vez mais urgentes e globais.

É sabido que desastres ambientais, apesar de terem origens diferentes, podem impactar da mesma forma na sociedade como um todo em diferentes proporções. Estimativas do Banco Mundial (MARRAKESH, 2016) apontam que anualmente há um prejuízo de cerca de 520 bilhões de dólares associados aos desastres ambientais. Além disso, atualmente existem cerca de 13,297 áreas de risco no país (Serviço Geológico do Brasil, 2023)<sup>5</sup>. Para se ter uma ideia, a recuperação das áreas afetadas pelo rompimento da barragem de Fundão, em Mariana, considerado o maior desastre ambiental do Brasil, irá custar cerca de 150 milhões às prefeituras envolvidas, além da criação de um fundo de 20 bilhões de dólares pela empresa envolvida (BARBOSA et al., 2015).

Tal desastre ambiental foi causado pelo rompimento da barragem de Fundão, em Mariana (MG), no dia 05 de novembro de 2015, evidenciou a íntima relação existente entre o meio biótico e o meio antrópico. Estima-se que tal desastre atingiu 39 municípios, em dois estados distintos, afetando cerca de aproximadamente 1,2 milhões de pessoas<sup>6</sup>. Segundo GODOY; DIAS (2021), o impacto causado pelo referido desastre provocou sérias consequências não somente na região de ocorrência, mas em todo o percurso dos rejeitos até a sua chegada ao mar, como pode ser visto pela Figura 1. Ademais, estas consequências não se restringem somente às questões ambientais, pois suas ocorrências também acarretam desequilíbrio político, econômico e social das cidades envolvidas.

Dessa forma, é necessário a mobilização de todos os setores da sociedade para o processo de recuperação da área, em especial do governo. Com o ocorrido, o Ministério Público de Minas Gerais (MPMG) buscou diversas formas de reparar ou compensar os danos causados (Superintendência de Comunicação Integrada do Ministério Público de Minas Gerais, 2024). Uma dessas formas, por exemplo, é a realização de acordos extrajudiciais e judiciais, buscando assim a economia de tempo e recursos.

Nesse contexto, um dos problemas identificados, por exemplo, está relacionado à alocação de recursos financeiros, de modo adequado e eficiente, para a recuperação das áreas afetadas, uma vez que são necessários abundantes quantidades de dinheiro para a restauração do local. Segundo BARBOSA et al. (2015), após o desastre, estima-se que as prefeituras das áreas envolvidas terão que gastar cerca de R\$150 milhões, além de que

<sup>1</sup> <<https://www.aljazeera.com/gallery/2023/9/6/fierce-storm-in-southern-brazil-kills-at-least-27>>

<sup>2</sup> <<https://www.theguardian.com/world/2023/feb/13/cyclone-gabrielle-new-zealand-declares-national-state-of-emergency>>

<sup>3</sup> <<https://www.brasildefato.com.br/2023/06/24/agua-toxica-cinco-bacias-brasileiras-contaminadas-pela-mineracao>>

<sup>4</sup> <<https://www.pbs.org/newshour/show/ohio-town-searches-for-answers-months-after-toxic-train-derailment>>

<sup>5</sup> <<http://www.cprm.gov.br/>>

<sup>6</sup> <<https://www.todamateria.com.br/desastre-de-mariana/>>



Figura 1 – Caminho percorrido pela lama, de Mariana(MG) até Linhares (ES).

Disponível em: <<https://www.todamateria.com.br/desastre-de-mariana/>>.

Acesso em: 01 de jul. de 2023.

há uma proposta da criação de um fundo de US\$20 bilhões, ao longo de 10 anos, pelas empresas envolvidas.

Pensando em formas de auxiliar nesse objetivo, buscou-se, através do processo de Descoberta de Conhecimento em Bases de Dados (do inglês, *Knowledge Discovery in Database* KDD) (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996), na etapa de Mineração de Dados (MD), utilizar conceitos específicos dessa área como forma de abordar tal problema. Segundo CAMILO; SILVA (2009), tais técnicas possibilitam a resolução de tarefas como Descrição, Classificação, Estimção, Predição, Agrupamento e Associação. Cada uma das técnicas permite resolver um determinado tipo de problema, não existindo assim uma melhor ou pior técnica (NONATO, 2010), apenas técnicas mais adequadas ou não para a resolução.

Dessa forma, este trabalho propõe o desenvolvimento de uma metodologia baseada em MD, que explora técnicas de Agrupamento, como o *KMeans*, para aglomerar cidades que, de alguma forma, são semelhantes, por exemplo, em relação a aspectos socioeconômicos, ambientais etc. Acredita-se que dessa forma seja capaz de auxiliar o usuário responsável pelo processo de tomada de decisão em relação à alocação de recursos financeiros em cenários de desastres ambientais. Por fim, a metodologia proposta é implementada e refletida em um *software Web*<sup>7</sup>, permitindo assim ao usuário interagir com todo esse processo metodológico proposto, mesmo não o conhecendo a fundo, e, obtendo no final, uma visão ampla dos dados mediante um painel descritivo disponível. Espera-se que o arcabouço proposto, ou seja, a metodologia juntamente com o desenvolvimento de uma aplicação que a abarque, possa ser útil para suportar a tomada de decisões de especialistas neste contexto, como, por exemplo, no processo de decisão de cidades (ou áreas) atingidas

<sup>7</sup> O link para a versão inicial disponibilizada para testes está disponível em: <<http://www.webgisbras.com.br/cases-sucesso.html>>.

por desastres ambientais e que, em detrimento de alguma característica (em comum), cuja recuperação deve ser priorizada.

## 1.1 Objetivos

### 1.1.1 Objetivos Gerais

Propor um arcabouço computacional buscando auxiliar um usuário responsável pelo processo de tomada de decisão em relação à alocação de recursos financeiros em cenários de desastres ambientais.

### 1.1.2 Objetivos Específicos

- Investigar e compreender o funcionamento dos principais algoritmos de Agrupamento, em especial do *KMeans*;
- Identificar e coletar dados que consigam refletir os principais aspectos socioeconômicos, ambientais e políticos de uma dada área de interesse;
- Propor uma metodologia para agrupamento das cidades em caso de desastres ambientais, que utilize para isso conceitos de Mineração de Dados e Geoprocessamento;
- Desenvolver um *software* que reflita toda a metodologia elaborada;
- Realizar um caso de teste com *software* proposto para validação da metodologia;
- Avaliar o *software* proposto, com os possíveis usuários especialistas da área de Solos.

### 1.1.3 Organização do Trabalho

O trabalho foi dividido da seguinte forma: o Capítulo 2 apresenta conceitos que fundamentam este trabalho e que são necessários para uma melhor compreensão do mesmo. Já o Capítulo 3 inclui um levantamento dos trabalhos realizados na área e que apresentam certa interseção com o que é proposto aqui. No Capítulo 4 é apresentada a metodologia elaborada, assim como todos os passos realizados desde a coleta dos dados, o processamento, a aplicação dos algoritmos escolhidos e a interpretação dos resultados, assim como todas as etapas do desenvolvimento do *software*. O Capítulo 5 traz uma apresentação do sistema final desenvolvido, assim como a avaliação dele juntamente com o usuário final e sua aplicação em um caso de estudo real. Por fim, o Capítulo 6 apresenta as conclusões obtidas, assim como o avanço científico proporcionado pelo trabalho, além de algumas propostas de Trabalhos Futuros que podem ser realizados.

## 2 Fundamentação Teórica

Esse capítulo busca explicar os principais conceitos utilizados nesse trabalho. Para isso, buscou-se por trabalhos na literatura a fim de fundamentar tais conceitos. A Seção 2.1 define, de maneira geral, o processo de KDD e cada uma das suas etapas, enquanto que na Seção 2.2 o foco é somente na etapa de Mineração de Dados. Já na Seção 2.3 é apresentado sobre o Aprendizado de Máquina e os aprendizados supervisionado e não-supervisionado. Na sequência, na Seção 2.4, é falado sobre o Processo Padrão Intersectorial para Mineração de Dados. Adiante, na Seção 2.5, é apresentado o Algoritmo KMeans e sua ideia de funcionamento. Por fim, na Seção 2.6, é definido o que são Desastres Ambientais e como podem ser classificados.

### 2.1 Processo de Descoberta de Conhecimento em Bases de Dados

O processo de KDD é uma sequência contínua e linear de cinco etapas que permitem a obtenção de algum conhecimento oculto *a priori* em um determinado conjunto de dados. Originalmente, tal processo foi proposto para substituir a análise manual dos dados, uma vez que era lenta, custosa e subjetiva (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996). Essa substituição era necessária porque o volume de informação produzida estava aumentando, desde os anos 1980, com a popularização dos computadores pessoais e com a popularização da Internet, nos anos 1990.

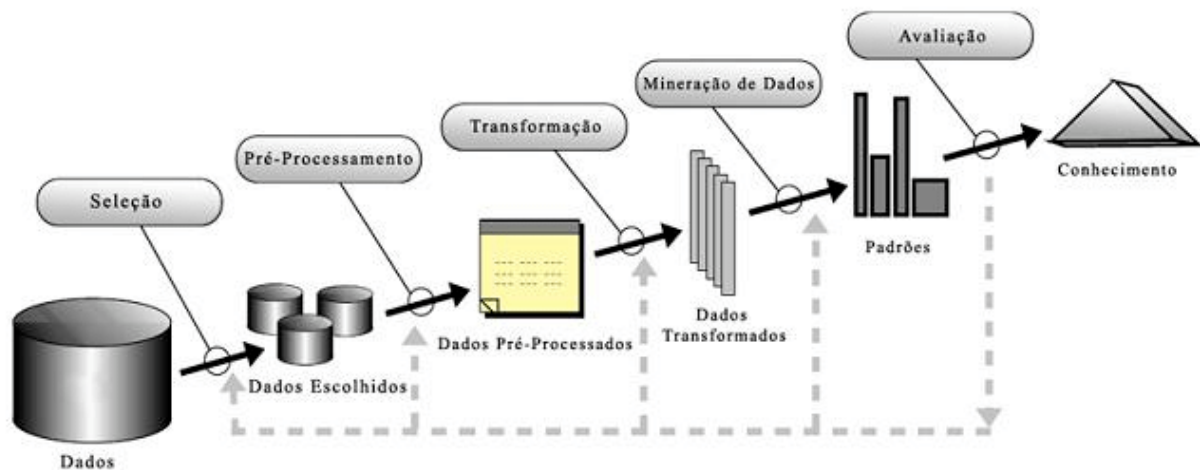


Figura 2 – Etapas do KDD.

Fonte: (LIRA et al., 2016).

Como pode ser visto na Figura 2, as cinco etapas que compõem o processo são: Seleção, Pré-Processamento, Transformação, Mineração de Dados e Avaliação, sendo cada uma delas são detalhadas a seguir.

#### 2.1.1 Seleção

Nessa primeira etapa, são selecionadas as várias fontes de dados que serão utilizadas ao longo do processo e que tenham alguma relação com o contexto do usuário.

Espera-se muitos dados, logo existe a probabilidade de que eles estejam em formatos diferentes. O que pode ser desafiador.

### 2.1.2 Pré-Processamento

Com o volume total de dados já selecionado, nessa etapa é realizado o tratamento deles. Segundo [HAN; PEI; TONG \(2022\)](#), é necessário realizar a limpeza dos dados, verificando se há dados faltantes, assim como a necessidade de normalização e a remoção dos valores discrepantes (*outliers*). Ainda nessa etapa, é necessário garantir também a integridade dos dados, buscando evitar dados redundantes, dependentes ou valores conflitantes. Este processo é ilustrado na Figura 3.

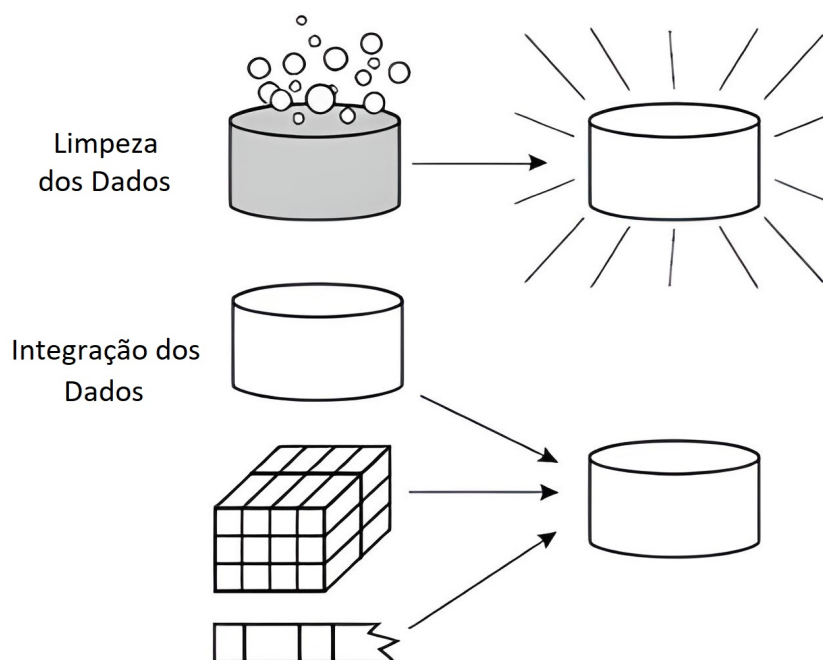


Figura 3 – Etapas do Pré-Processamento dos dados.

Fonte: Adaptada pelo autor a partir de ([CAMILO; SILVA, 2009](#)).

### 2.1.3 Transformação

Uma vez que seja garantido a integridade dos dados, é necessário converter todos eles para um formato mais adequado ao tipo de entrada do algoritmo selecionado, armazenando-os em um único lugar. Por exemplo, para facilitar a aplicação do algoritmo *KMeans*, é interessante que todos os dados estejam em um único formato e em um único arquivo centralizado.

### 2.1.4 Mineração de Dados

Já nessa etapa, é implementado e aplicado o algoritmo em si, dado como base o objetivo final do usuário. De forma resumida, o objetivo desta etapa é realizar tarefas de Associação, Classificação, Agrupamento, Regressão e Predição, usando para isso diversos algoritmos disponíveis, como Árvore de Decisão, *KMeans*, dentre outros. Maiores detalhes são apresentados na Seção a seguir.

### 2.1.5 Avaliação

Uma vez que o usuário possui o resultado da aplicação do algoritmo de Mineração de Dados, cabe a ele interpretar e traduzir o que foi obtido para o contexto do problema em que ele se encontra. Para isso, podem ser utilizados gráficos e tabelas para auxiliar nessa visualização.

## 2.2 Mineração de Dados

Essa é a etapa mais importante de todo o processo de KDD, sendo inclusive sinônimo do mesmo. Há diversas tarefas que podem ser resolvidas com os métodos e ferramentas aqui disponíveis, sendo as principais, segundo CAMILO; SILVA (2009), as representadas na Figura 4.

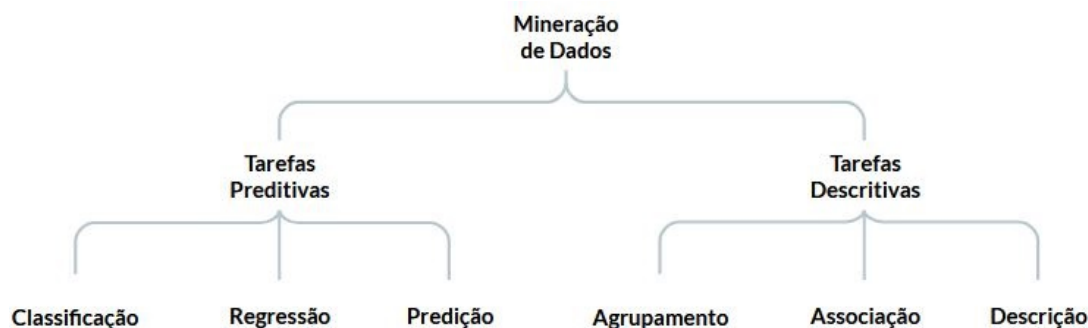


Figura 4 – Tarefas da Mineração de Dados.

Fonte: Autor

Dessa forma, tem-se que a MD abrange tanto tarefas preditivas quanto descritivas, visando dessa forma prever e descrever valores, respectivamente, com base em um conjunto de dados. Em linhas gerais, o conceito adotado nesse trabalho é o mesmo definido em (WITTEN; FRANK, 2002), onde MD objetiva unicamente buscar por um conhecimento oculto no conjunto de dados trabalhados.

## 2.3 Aprendizado de Máquina

O Aprendizado de Máquina (do inglês, *Machine Learning* - ML), assim como em MD, também busca extrair conhecimento de um conjunto de dados, porém objetiva criar modelos que ensinem o computador a fazer isso de forma automatizada (HAN; PEI; TONG, 2022), permitindo dessa forma a previsão de valores (WITTEN; FRANK, 2002), dentre outras tarefas. Assim, uma releitura da Figura 4 pode ser vista na Figura 5, sob a ótica agora do ML.

Em relação à forma que as tarefas de MD podem ocorrer, (MONARD; BARANAUSKAS, 2003) definem como Aprendizado Indutivo (AI) o processo de utilizar inferência lógica para obter conclusões genéricas dado um conjunto de dados. Dessa forma, o AI pode se dividir em supervisionado e não-supervisionado, resolvendo assim as mesmas tarefas preditivas e descritivas, respectivamente da MD (CARVALHO, 2014). Em linhas gerais, o aprendizado supervisionado ocorre fornecendo um conjunto de exemplos para treinamento, enquanto o aprendizado não-supervisionado considera de imediato o conjunto total de dados.



Figura 5 – Relação entre as tarefas da MD com o ML.  
 Fonte: Autor.

## 2.4 Processo Padrão Intersetorial para Mineração de Dados

O Processo Padrão Intersetorial para Mineração de Dados (do inglês, *Cross-Industry Standard Process of Data Mining - CRISP-DM*) pode ser entendido como uma metodologia genérica e um processo cíclico que permite aplicar a MD em um determinado contexto. Sua criação e desenvolvimento partiu do interesse da indústria, uma vez que não havia até então nenhuma metodologia que desse esse tipo de suporte para eles (MEIRA, 2012). Seu modelo hierárquico pode ser visto na Figura 6 e seu funcionamento pode ser visto na Figura 7, onde cada retângulo representa uma fase ou processo da metodologia.

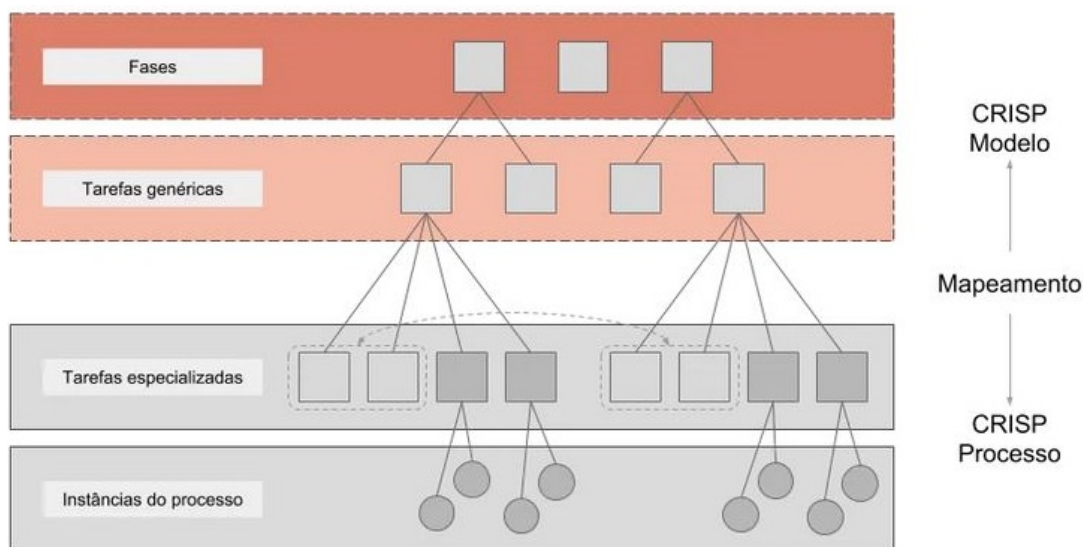


Figura 6 – Quatro níveis da metodologia CRISP-DM.  
 Disponível em: <<https://medium.com/@kvmoura/crisp-dm-79580b0d3ac4>>.  
 Acesso em: 31 de jul. de 2023.

No nível mais alto da hierarquia, “Fases”, se encontra o KDD. Nele, cada uma das etapas é dividida em uma fase. Cada fase pode apresentar, ou não, uma ou mais tarefas genéricas associadas, resolvidas por uma ou mais tarefas especializadas. Para cada tarefa especializada, é possível existir uma instância de processo, que representa um resultado alcançado no fim daquela fase. A aplicação do CRISP-DM na metodologia aqui.

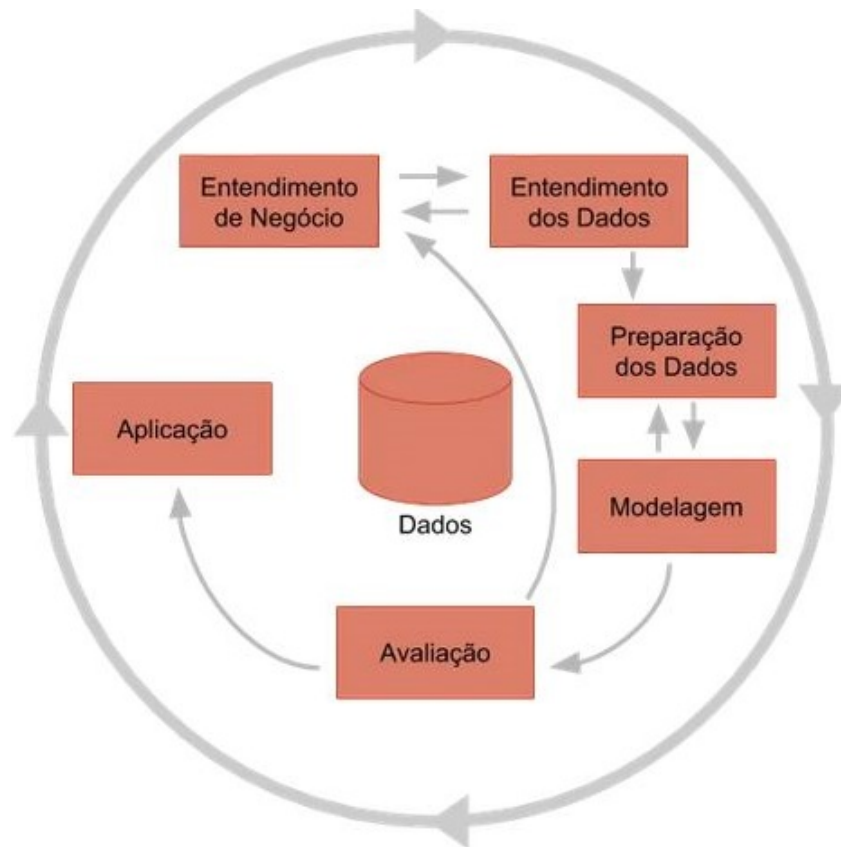


Figura 7 – Fases do modelo CRISP-DM.

Disponível em: <<https://medium.com/@kvmoura/crisp-dm-79580b0d3ac4>>.

Acesso em: 31 de jul. de 2023.

## 2.5 Algoritmo *KMeans*

O *KMeans* é um algoritmo clássico de ML bastante explorado na literatura, uma vez que existem diversas pesquisas (ZHANG et al., 2015) buscando aprimorar a determinação do número de agrupamentos (LAI; LIU, 2008), na seleção dos centróides (LAI; LIAW, 2008) e na forma de realizar o agrupamento em geral (ZHANG; ZHANG; LIU, 2011). Ele foi proposto em 1967 e por ser de fácil implementação e baixa complexidade, é comumente empregado para a geração de grupos (MACQUEEN et al., 1967). O pseudocódigo do algoritmo é apresentado a seguir.

---

### Algorithm 1 Algoritmo K-Means

---

Especifique o número  $K$  de agrupamentos para gerar

Inicialize os  $K$  centróides de forma randômica

**while**  $A$  posição do centróide não mudar **do**

    Atribua cada ponto ao centróide mais próximo

    Calcule o novo valor de centróide (média) de cada agrupamento

Retorne os agrupamentos

---

Dessa forma, o algoritmo em si pode ser dividido em duas fases (YADAV; SHARMA, 2013):

1. Os  $K$  centróides são escolhidos de maneira aleatória, sendo  $K$  um número real definido como  $2 \leq K \leq \alpha$ , com  $\alpha$  sendo o número total de elementos do conjunto de

dados.

2. Cada um dos elementos é associado ao conjunto do centroide mais próximo. Essa distância pode ser determinada através dos métodos Euclidiano, Manhattan ou Minkowski. É sabido que o método Euclidiano é o mais comumente empregado nesse tipo de abordagem (THANT; AYE; MANDALAY, 2020).

## 2.6 Desastres Ambientais

Um desastre pode ser definido como sendo um evento que ocorre de maneira inesperada e que impacta de forma profunda na vida de uma comunidade<sup>8</sup>. Os desastres podem ser classificados como naturais ou humanos<sup>9</sup>. A principal diferença entre ambos é a origem, pois desastres humanos ocorrem através da intervenção ou omissão humana como fator inicial. Ambos os tipos de desastres possuem a capacidade de gerar grandes danos ao meio ambiente. Quando esses danos e prejuízos são incalculáveis e de difícil restituição, tem-se um desastre ambiental.

---

<sup>8</sup> <[https://www.prefeitura.sp.gov.br/cidade/secretarias/seguranca\\_urbana/defesa\\_civil/terminologia\\_desastres/\index.php?p=179948](https://www.prefeitura.sp.gov.br/cidade/secretarias/seguranca_urbana/defesa_civil/terminologia_desastres/\index.php?p=179948)>

<sup>9</sup> <<https://www.defesacivil.pr.gov.br/Pagina/O-que-e-um-desastre>>

### 3 Trabalhos Relacionados

Há diferentes formas de desenvolver sistemas que podem auxiliar tanto antes quanto após a ocorrência de algum fenômeno. Em (BRUMATTI et al., 2021), são explorados conceitos de Informação Geográfica Voluntária (do inglês, *Volunteered Geographic Information* - VGI) e uso de padrões no compartilhamento de dados e da usabilidade da ferramenta, para o desenvolvimento de um *framework* para o desenvolvimento de Sistemas de Informação de Desastres (do inglês, *Disaster Information Management Systems* - DIMS). Outra forma de abordagem é a possibilidade de utilizar conceitos de MD de forma associada. Em (HUANG et al., 2019) é usado tanto VGI, proposto por meio do uso das redes sociais, quanto MD, por meio de Algoritmos de *Machine Learning* (ML), como Floresta Randômica (do inglês, *Random Forest*), Máquinas de Vetores de Suporte (do inglês, *Support Vector Machines* - SVM) e Redes Neurais Convolucionais (do inglês, *Convolutional Neural Networks* - CNNs) como forma de identificação da ocorrência de desastres. Os autores buscam descrever características textuais e visuais, dos conteúdos das redes sociais. Após essa descrição, ambos os dados são utilizados em conjunto para representação do desastre ambiental. Os resultados indicam uma acurácia de 92,92% e 93,45%, respectivamente, em relação a uma descrição exata do evento.

O uso de CNNs também é visto em (CHEN; OUYANG; AGAM, 2019) sendo aplicado em imagens de satélite para comparação entre o antes e o depois de áreas atingidas por desastres. O mesmo emprego de CNNs ocorre em (PENG et al., 2019), mas dessa vez com o objetivo específico de realizar a comparação em áreas urbanas atingidas por inundações.

O uso de Algoritmos de Agrupamento, considerados também como de ML, uma vez que utilizam do Aprendizado Não-Supervisionado como estratégia para compreender os dados disponibilizados e, a partir deles, agrupar dados semelhantes de alguma forma também é visto na literatura. Em (SUPRIYADI et al., 2018), onde a localização da Indonésia no chamado “Anel de Fogo” faz com que já exista um histórico da ocorrência de desastres<sup>10</sup>. Dessa forma, o autor propõe o uso do algoritmo *KMeans* em uma base de dados pública, obtida do *Central Bureau of Statistics*<sup>11</sup>. Tal base de dados relaciona uma determinada província com os números de enchentes, terremotos e deslizamentos, ocorridos entre os anos de 2008 e 2014. Com isso, ele objetiva agrupar o maior número de províncias em torno de 3 agrupamentos possíveis: número baixo, médio e alto de ocorrência de desastres. Para isso, ele utilizou o algoritmo disponível através *software RapidMiner*. Os resultados obtidos revelam que há 4 províncias com altos índices de ocorrência de desastres, 14 províncias com risco médio e 16 províncias com risco baixo.

Outro trabalho semelhante foi realizado por (ANNAS; RAIS, 2020), em que o autor também utiliza o *KMeans* como técnica para realizar agrupamento dos dados públicos obtidos a partir do *National Disaster Mitigation Agency of Indonesia*<sup>12</sup>. Diferente do trabalho anterior, que utilizou a distância em relação ao centroide e o Índice de Davies-Boulding, agora é utilizado a técnica de *Root Mean Square Deviation* (RMSD) (PHILLIPS; COLVIN; NEWSAM, 2011) para determinar o valor do número de agrupamentos a serem

<sup>10</sup> <<https://revistaplaneta.com.br/indonesia-e-o-pais-que-mais-sofre-efeitos-do-anel-de-fogo/>>

<sup>11</sup> <<https://nsonepal.gov.np/>>

<sup>12</sup> <<https://www.bnpb.go.id/>>

gerados pelo algoritmo. Nesse caso, o valor definido foi igual a 7. Assim, para cada um dos agrupamentos gerados, foi analisado a média dos valores em relação aos seguintes tipos de desastres: enchentes, enchentes com deslizamentos, erosão causada por fortes ondas, terremotos, incêndios florestais, erupções vulcânicas, tornados e deslizamentos de terra. Um diferencial desse trabalho, porém, foi a plotagem do resultado em um mapa, associando cada uma das províncias com o agrupamento a qual pertence, utilizando dessa forma, conceitos associados da área de Geoprocessamento para isso.

A visualização permitida pelo uso do mapa no exemplo anterior, é uma ferramenta de grande importância para o usuário. Em (DALLU, 2020) é proposto uma ferramenta dedicada ao gerenciamento e a governança de riscos nas organizações. Um dos pontos destacados é a capacidade que esse tipo de ferramenta possui em caracterizar quantitativamente e comunicar os riscos de um determinado evento. Sendo assim, essas são características interessantes para serem exploradas.

### 3.1 Lacuna de Pesquisa

Diferente dos trabalhos anteriores, aqui propõe-se a aplicação direta do *KMeans*, não sendo usado para isso ferramentas de terceiros como o *RapidMiner*<sup>13</sup> ou o *Weka*<sup>14</sup>. Isso é interessante por permitir analisar e comparar os diferentes parâmetros de inicialização do algoritmo. Além disso, é proposto o emprego de diferentes tipos de dados como forma de caracterização das áreas, permitindo assim uma melhor descrição do local. Dessa forma, a ferramenta se torna genérica para o usuário utilizar quaisquer dados relacionados a região de interesse. Por fim, o desenvolvimento de uma ferramenta de Inteligência Geográfica, isto é, uma ferramenta que considera variáveis de natureza espacial ou geográfica para explicar e traduzir fenômenos<sup>15</sup>, permitindo o usuário visualizar e interagir com toda essa metodologia, de forma simples e direta, representa um aspecto positivo na proposta desse trabalho, uma vez que a integração de dados distintos, o monitoramento em tempo real, aliado com uma visualização clara e objetiva da informação, permite uma tomada de decisão mais consciente por parte dele.

---

<sup>13</sup> <<https://rapidminer.com/>>

<sup>14</sup> <<https://www.cs.waikato.ac.nz/ml/weka/>>

<sup>15</sup> <<https://observatoriodeeducacao.institutounibanco.org.br/em-debate/inteligencia-geografica>>

# 4 Metodologia

Este capítulo descreve a metodologia proposta neste trabalho, com base no processo do KDD como um todo, porém com foco na etapa de MD. De maneira geral, a Figura 8 ilustra todas as etapas abordadas ao longo desse trabalho. Em cada etapa, destaca-se também as fontes de dados utilizadas, cada uma das atividades realizadas, assim como os produtos esperados. Por fim, é exibido também o processo adotado para implementação do sistema proposto. Todo esse processo pode ser visto também conforme a metodologia CRISP-DM, de acordo com a Tabela 1 e a Figura 9.

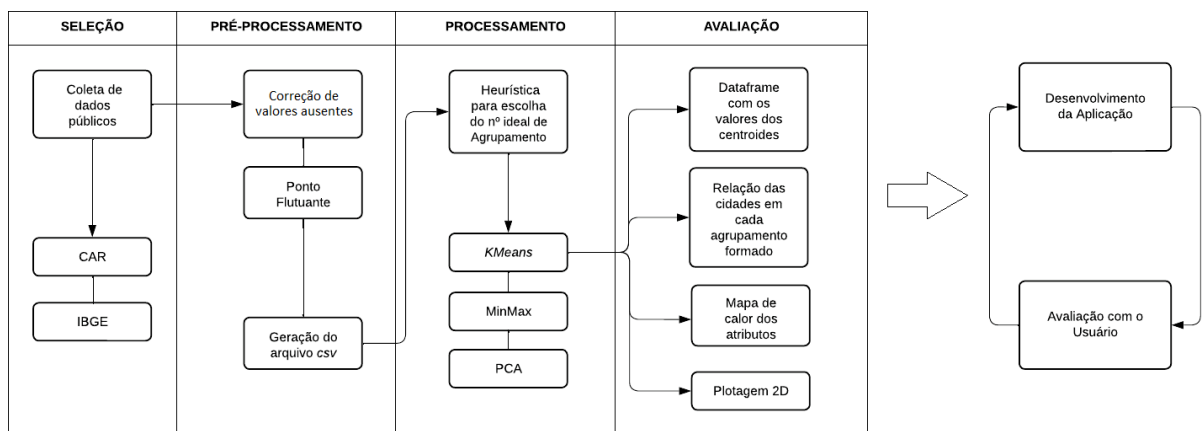


Figura 8 – Diagrama geral da metodologia de trabalho proposta.

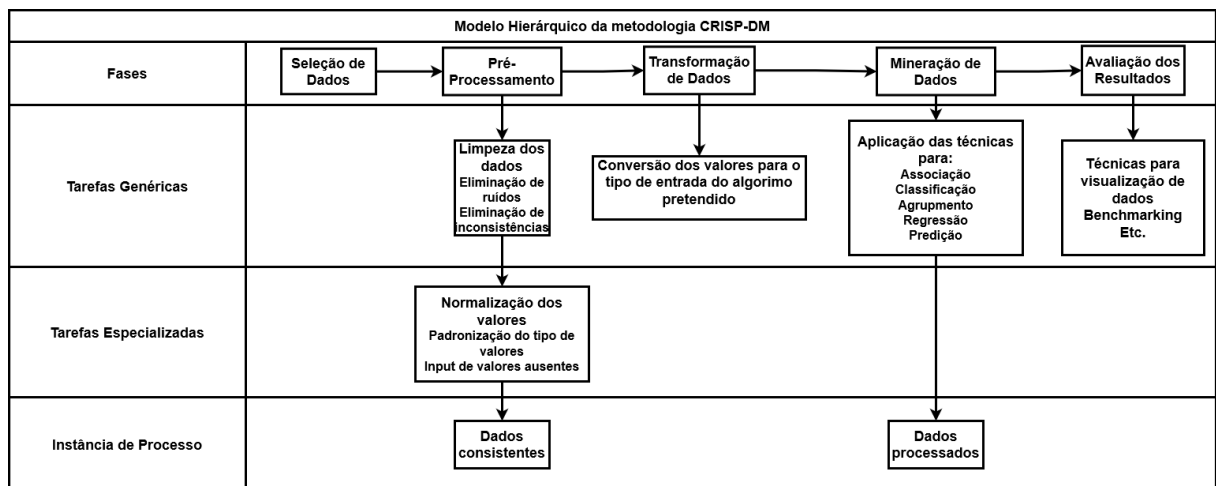


Figura 9 – Detalhamento dos quatro níveis da metodologia CRISP-DM.

Fonte: Autor.

Sendo assim, esse capítulo foi dividido da seguinte forma: na Etapa 1 é descrito como foi feito a compreensão do negócio, na Etapa 2 é descrito sobre o domínio dos dados, na Etapa 3 ocorre a preparação inicial dos dados para as etapas subsequentes, na Etapa 4 é realizado a modelagem dos dados já consolidados, na Etapa 5 ocorre a avaliação dos resultados obtidos e, por fim, na Etapa 6, tem-se uma breve descrição da metodologia implementada através de uma aplicação *Web*.

Tabela 1 – Resumo do trabalho segundo a metodologia CRISP-DM.

Etapa	Descrição
(1) Compreensão do Negócio	Compreensão da área através de uma revisão de literatura. Definição do problema: Uso de técnicas de Agrupamento de Dados como ferramenta de apoio ao usuário responsável pela tomada de decisão em relação à alocação de recursos financeiros em cenários de desastres ambientais.
(2) Compreensão dos Dados	Coleta dos dados públicos - CAR e IBGE Exploração inicial dos dados - análise descritiva
(3) Preparação dos Dados	Limpeza dos dados Seleção dos dados Transformação dos dados Normalização dos valores
(4) Modelagem	Aplicação dos algoritmos iniciais: Método <i>Elbow</i> , PCA Aplicação do algoritmo KMeans
(5) Avaliação	Plotagem 2D dos agrupamentos formados Visualização do valor dos centroides em cada um dos agrupamentos Visualização das cidades de cada agrupamento Mapa de calor
(6) Implantação	Desenvolvimento do sistema web, refletindo totalmente os passos 2, 3 e 4, e parcialmente o 5

## 4.1 Etapa 1 - Compreensão do Negócio

Nessa primeira fase, almejou-se definir os objetivos que deveriam ser alcançados neste trabalho. Uma vez definido, foi necessário elencar de qual forma isso seria feito e qual era o escopo pretendido. Em seguida, por meio de uma revisão da literatura, buscou-se o entendimento do contexto do problema, assim como a coleta de informações necessárias sobre técnicas e algoritmos que seriam interessantes para a resolução pretendida. Em paralelo a isso, verificou-se com os possíveis usuários a viabilidade do projeto.

## 4.2 Etapa 2 - Compreensão dos Dados

Para a elaboração do estudo, o foco se deu na utilização de dados públicos com potencial de descrever os aspectos demográficos, geográficos, econômicos e ambientais, nos seus vários aspectos, da região de interesse a ser considerada. Para isso, utilizou-se dados provenientes do Cadastro Ambiental Rural (CAR<sup>16</sup>) e do Instituto Brasileiro de Geografia e Estatística (IBGE<sup>17</sup>). A relação dos dados coletados pode ser vista nas Tabelas 2 e 3.

Pode-se observar assim que existem dados do tipo numérico, em sua maioria. Ainda é possível dizer que eles podem ser divididos em categorias, representando principalmente os aspectos demográficos, econômicos, ambientais e geográficos da região.

Os dados foram todos coletados de fontes públicas e de forma manual para cada cidade, isto é, para cada cidade, foram anotados os seus respectivos valores. Somente os dados do CAR apresentam a particularidade de não serem valores brutos de áreas ou distâncias, mas sim o número de vezes que determinado atributo aparece em uma cidade. Por exemplo, o valor de APP não representa a área média de uma APP em um dado conjunto de municípios, mas sim o número médio de APPs, por cidade, dentro do agrupamento em questão.

<sup>16</sup> <<https://www.car.gov.br/>>:

<sup>17</sup> <<https://www.ibge.gov.br/>>

Tabela 2 – Relação dos dados coletados do IBGE.

Demográfico	Econômico
População Estimada (2021)	Salário médio mensal (2019)
População do último censo (2010)	Percentual da população com renda até meio salário mínimo (2010)
Densidade demográfica (2010)	Produto Interno Bruto Per Capita (2019)
População ocupada (2019)	Percentual de receitas provenientes de fontes externas (2015)
Taxa de escolarização entre 6 e 14 anos (2010)	Índice de Desenvolvimento Humano (2010)
IDEB dos anos finais do ensino fundamental (2019)	Total de receitas realizadas (2017)
IDEB dos anos iniciais do ensino fundamental (2019)	Total de receitas empenhadas (2017)
Número de matrículas no ensino fundamental (2020)	Pessoal ocupado (2019)
Número de matrículas no ensino médio (2020)	
Número de docentes no ensino fundamental (2020)	
Número de docentes no ensino médio (2020)	
Número de escolas de ensino fundamental (2010)	
Número de escolas de ensino médio (2020)	
Mortalidade infantil (2019)	
Internações por diarreia (2016)	
Número de estabelecimentos SUS (2009)	
Área territorial (2021)	
Percentual Esgotamento adequado (2010)	
Percentual arborização das vias públicas (2010)	
Percentual de urbanização das vias públicas (2010)	

Tabela 3 – Relação dos dados coletados do CAR.

Ambiental	Geográfico	Demográfico
Número de Áreas de Proteção Permanente	Número de áreas com altitude superior a 1800 m	Número de imóveis rurais
Número de áreas consolidadas	Número de áreas com declividade superior a 45 graus	
Número de áreas banhadas	Número de áreas de pouso	
Número de unidades de recursos hídricos	Número de topos de morros	
Número de nascentes	Número de bordas de chapada	
Número de reservas legais	Número de manguezais	
Número de unidades com vegetação nativa	Número de áreas de restinga	
Número de unidades do tipo vereda	Número de áreas de servidão administrativa	
	Número de áreas do tipo restrito	

A exploração inicial dos dados foi realizada visando compreender os valores coletados e como eles se organizam. Essa exploração consistiu na obtenção de métricas tanto de tendência central (média, moda, mediana) quanto de dispersão (valores de mínimo, máximo e desvio-padrão) para cada um dos atributos coletados em relação a cada uma das cidades. A Tabela 4 apresenta alguns dos resultados obtidos.

Tabela 4 – Exploração inicial do conjunto de dados utilizado.

Nome do Atributo	Valor Mínimo	Valor Máximo	Média	Moda	Mediana	Desvio-Padrão
APP	262.0	15142.0	3903.25	262.0	3184.0	3125.686964
Altitude Superior 1800 m	0.0	1.0	0.020833	0.0	0.0	0.144338
Consolidada	49.0	2235.0	727.166667	49.0	648.0	556.621279
Declividade Maior 45	0.0	17.0	3.25	0.0	2.0	4.354015
Imóveis	61.0	2064.0	843.520833	61.0	738.5	582.231401

De maneira geral, nota-se que os dados apresentam uma grande amplitude entre si, o que indica uma grande dispersão entre os valores de um mesmo atributo. Essa dispersão pode ser comprovada ainda pelos valores indicados de desvio-padrão, os quais são muito elevados para alguns atributos como APP e Imóveis, por exemplo.

### 4.3 Etapa 3 - Preparação dos Dados

Após a exploração inicial dos dados, realizou-se a limpeza. Para isso, foi primeiramente feita a padronização da representação dos valores de ponto flutuante, modificando quando necessário o caractere de separação dos valores. Em seguida, constatou-se que não havia valores faltantes para serem acrescentados. Por fim, foi feita a normalização dos valores, uma vez que a amplitude dos atributos, isto é, a diferença entre o valor máximo e mínimo, era grande, havendo assim a necessidade de colocar todos os valores em uma mesma escala. Neste trabalho, optou-se pela normalização *MinMax* uma vez que ela é realizada anterior a aplicação do algoritmo *KMeans*, de acordo com (VIRMANI; TANEJA; MALHOTRA, 2015). Além disso, entende-se também que é mais fácil compreender o peso de um atributo quando esse se encontra em um intervalo bem definido, como no caso entre 0 e 1.

### 4.4 Etapa 4 - Modelagem

A metodologia proposta para essa etapa pode ser vista na Figura 10. Nela, é visto que o *KMeans* é aplicado tanto para os dados brutos, obtendo assim os valores reais dos centroides, quanto para os dados normalizados, colocando todos os valores em uma mesma escala e obtendo o mapa de calor e a relação das cidades em cada Agrupamento.

#### 4.4.1 Método *Elbow*

Após a etapa de preparação dos dados, é aplicado uma heurística para auxiliar na escolha do valor do número inicial de agrupamentos a ser gerado pelo algoritmo *KMeans*. Esse valor é aqui representado por  $K$ . Dessa forma, foi utilizado o Método *Elbow* (THORNDIKE, 1953). Baseando-se no cálculo da inércia, isto é, na soma da distância quadrada do centroide de cada um dos agrupamentos até o conjunto mais próximo, tal método utiliza vários valores de  $K$ , visando identificar entre quais valores ocorre a maior diferença da inércia. Esse valor também pode ser indicado como sendo o ponto mais distante da curva formada pelo gráfico bidimensional do número de agrupamentos pelo valor de inércia calculado. Como a inicialização do algoritmo *KMeans* é aleatório, tem-se que para cada tentativa de descoberta do valor  $K$ , o valor sugerido pode variar. Para contornar esse problema, tal método é executado várias vezes, a fim de escolher a moda dentre todos os valores encontrados, isto é, o valor que mais aparece.

##### 4.4.1.1 Ponto de maior distância como valor ideal do número de agrupamentos $K$

O ponto de inflexão da curvatura do gráfico gerado pelo método *Elbow* representa o valor ideal do número de agrupamentos a serem formados, isto é, o valor de  $K$ . De acordo com (SANTANA et al., 2020), tal ponto deve ser o mais distante quando comparado a uma reta traçada entre o primeiro e último ponto. A Figura 11 representa o gráfico traçado a partir da execução do Método *Elbow* para os dados carregados. A partir do gráfico, foi determinado a reta entre os pontos  $K = 2$  e  $K = 9$ .

Sendo assim, para cada um dos pontos do intervalo, calculou-se a distância dele à reta. Os valores obtidos estão representados na Tabela 5. Como o ponto  $K = 3$  é o mais distante, tem-se que ele representa o valor ideal de agrupamentos.

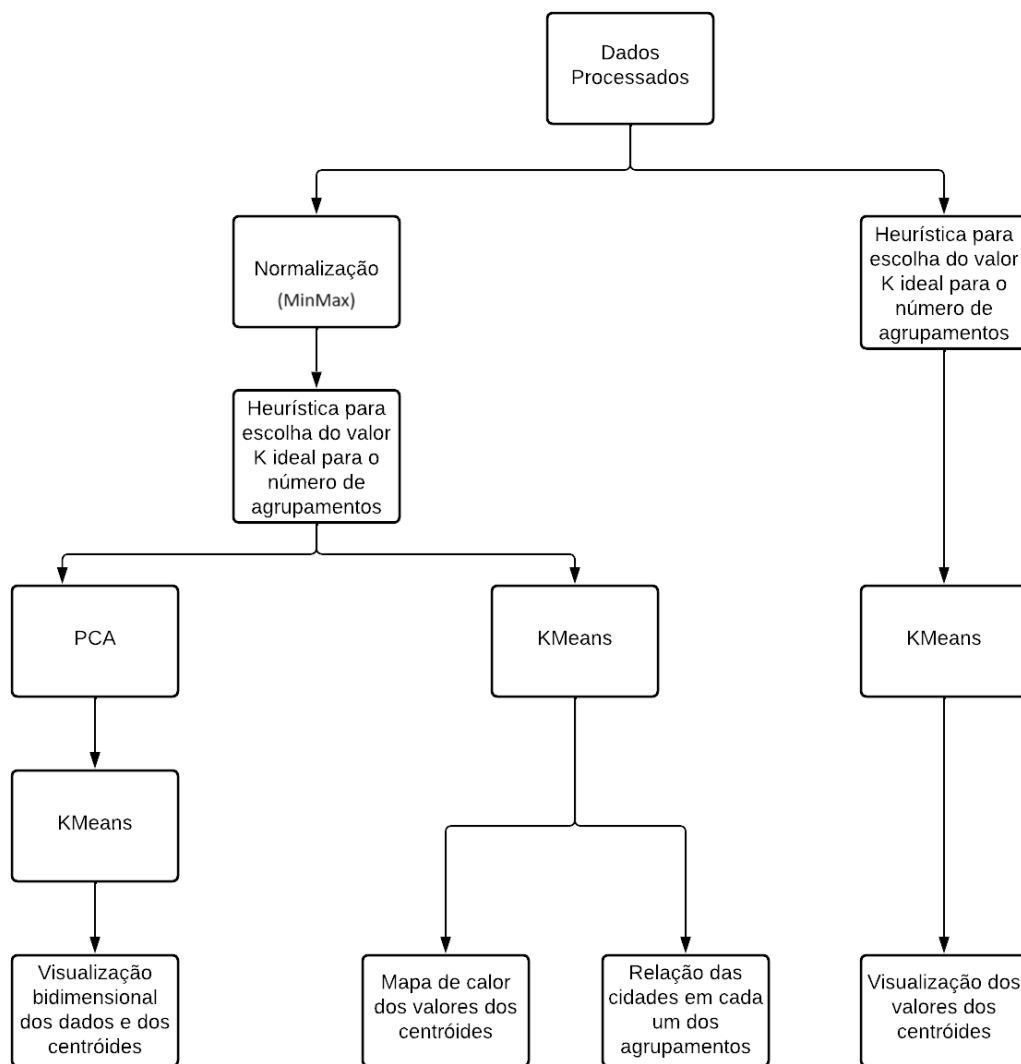


Figura 10 – Metodologia adotada para a execução deste trabalho.

Tabela 5 – Relação dos pontos do gráfico da Figura 11 com a distância calculada até a reta traçada entre os pontos inicial e final.

Quantidade de agrupamentos	Valor da distância do ponto a reta (x1,000,000,000)
3	93,42252022061319
4	86,02646623075537
5	60,88322275336499
6	32,795689949925415
7	17,192214404739647

#### 4.4.2 Principal Component Analysis (PCA)

Após a aplicação do algoritmo *KMeans*, é necessário converter os diversos atributos existentes, para cada uma das cidades, em um conjunto bidimensional, buscando dessa forma realizar a plotagem desses valores em um gráfico, para avaliar assim a distribuição e densidade dos dados. Assim, foi utilizada a técnica de PCA (BRO; SMILDE, 2014) para isso. A implementação aqui utilizada foi a versão padrão da biblioteca *scikit-learn*, sendo o único parâmetro alterado o  $n\_components = 2$ , que é fixo e define assim o caráter bidi-

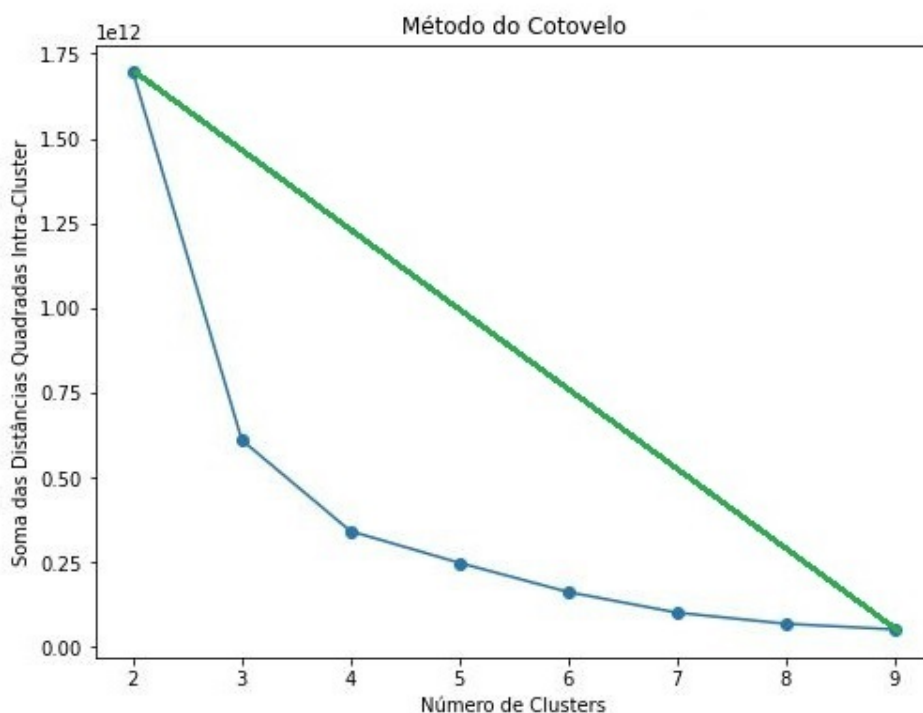


Figura 11 – Gráfico obtido do Método *Elbow* considerando a reta traçada entre os pontos inicial e final.

mensional pretendido. Em resumo, o PCA simplifica a complexidade dos dados, reduzindo a dimensão deles e mantendo a variabilidade da informação.

#### 4.4.3 Método de Agrupamento

Para a implementação dessa metodologia, utilizou o algoritmo *KMeans* como algoritmo de Agrupamento. Como dito anteriormente no Capítulo 2, tal algoritmo foi escolhido por ser bastante explorado na literatura (BRUMATTI et al., 2022). Nessa etapa, utilizou-se a versão disponível na biblioteca *scikit-learn*<sup>18</sup> da linguagem de programação *Python*. Os parâmetros utilizados para isso e uma breve descrição do significado de cada um podem ser visto na Tabela 6.

### 4.5 Etapa 5 - Avaliação

Uma vez aplicado o algoritmo de Agrupamento, foi realizado a redução da dimensionalidade dos dados, isto é, todos os 46 atributos escolhidos foram comprimidos e reduzidos para apenas 2 valores, para cada uma das cidades. Além dos valores coletados, todos os centróides obtidos também tiveram suas dimensões reduzidas. Para essa compressão, utilizou-se o PCA, como dito anteriormente. Com os novos valores obtidos, foi realizada a plotagem bidimensional, através de um gráfico de dispersão, no sistema de coordenadas cartesiano.

Com a normalização dos dados, cada um dos centróides obtidos apresenta os seus atributos também normalizados. Assim, fica impossível a caracterização de cada agrupamento formado, em relação ao atributo original. Para contornar esse problema, foi

<sup>18</sup> <<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>>

Tabela 6 – Parâmetros utilizados no Agrupamento com o algoritmo *KMeans*.

Parâmetro	Significado	Valor adotado
n_cluster	Número de agrupamentos a ser gerado.	K
init	Método de inicialização dos centróides.	k-means++
tol	Tolerância para a convergência do algoritmo. Quando a variação entre as iterações subsequentes for menor que esse valor, o algoritmo será considerado como tendo convergido.	0.0001
verbose	Nível de verbosidade durante o ajuste do modelo.	0
random_state	Semente para a geração de números aleatórios, que afeta a seleção dos centróides iniciais.	0
max_iter	Número máximo de iterações para cada execução do algoritmo.	100
copy_x	Se os dados originais devem ser copiados antes do ajuste. Ajuda na economia de memória.	True
n_init	Número de vezes que o algoritmo será executado com diferentes centróides iniciais. Os resultados finais serão escolhidos como o melhor entre as execuções.	10
algorithm	Algoritmo a ser utilizado para o cálculo dos centróides.	elkan

realizada uma segunda aplicação do algoritmo *KMeans*, utilizando agora os dados brutos, isto é, que não passaram pelo processo de normalização.

Após essa segunda aplicação da técnica de Agrupamento e contando agora com os valores de centróides sem modificações, foi realizada a plotagem desses valores por meio de um *dataframe*, conforme exemplificado na Figura 12. O objetivo com isso é caracterizar os respectivos agrupamentos com base no centróide, uma vez que é entendido que ele é o dado que melhor representa o agrupamento (K-MEANS, 2019).

	APP	Altitude Superior 1800	Consolidada	Declividade Maior 45	Imoveis	Pousio	Topo de Morro	Banhado	Borda Chapada	Hidrografia	...	IDHM(2010)	Total receitas realizadas(2017)	Total receitas empenhadas(2017)	in
0	8180.67	0.00	1494.25	6.08	1619.00	45.92	6.33	10.67	0.08	1384.67	...	0.70	131925.30	121586.80	
1	2524.73	0.03	473.12	2.24	578.58	19.94	1.73	3.55	0.30	481.97	...	0.68	57804.72	50533.48	
2	1957.33	0.00	453.33	3.00	656.00	13.33	2.00	1.67	0.00	384.00	...	0.76	1380818.10	1199108.92	

3 rows × 46 columns

Figura 12 – Exemplo dos valores de cada um dos centróides.

Uma vez que os centróides estão caracterizando cada um dos agrupamentos, foi necessário saber quais as cidades que de fato fazem parte do agrupamento. Para isso, implementou-se uma função responsável por definir cada uma dessas cidades. Por fim, foi obtido o mapa de calor, destacando a importância de cada um dos atributos dentro de cada um dos agrupamentos formados. Na Figura 13 pode ser visto o mapa de calor assim como a escala.

## 4.6 Etapa 6 - Implantação

Buscando refletir as Etapas 2, 3, 4 e 5 aqui descritas nessa seção, foi implementado um sistema *Web*, denominado *TerrainIntelGIS*. Tal sistema será melhor descrito na Seção



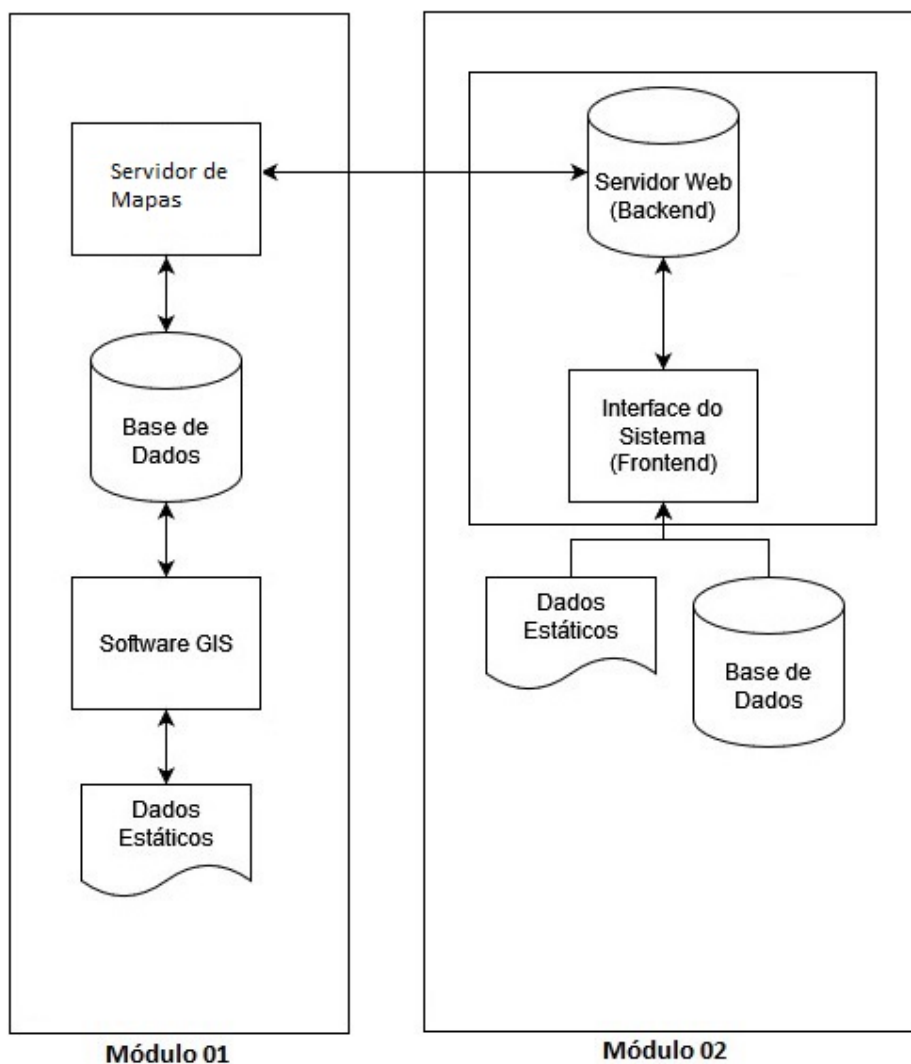


Figura 14 – Arquitetura geral da aplicação proposta.

são armazenados em um sistema de banco de dados geográficos, como o PostgreSQL<sup>22</sup> com a extensão espacial *PostGIS*<sup>23</sup>, o que permite o armazenamento estruturado da informação e a realização de consultas simples ou complexas, relacionando as camadas entre si (FILHO; IOCHPE, 1999). Todas as camadas armazenadas no PostgreSQL são disponibilizadas no servidor de mapas. Para esse trabalho, optou-se pelo *GeoServer*<sup>24</sup>, devido principalmente ao seu uso mais difundido.

Já em relação ao Módulo 02, é importante destacar o retângulo representado por 1. Ele simboliza que tanto o *backend* quanto o *frontend* da aplicação foram construídos utilizando o conceito de containerização possibilitado através da tecnologia *Docker*<sup>25</sup>. Dessa forma, a distribuição e implantação da aplicação ocorre mais facilmente. Em relação ao *backend*, ele foi construído utilizando a linguagem de programação Python, versão 3.6. Vale destacar que os algoritmos de MD utilizados foram os da versão disponível na biblioteca *scikit-learn*<sup>26</sup>. No *frontend*, foi empregado um fluxo linear de páginas, do tipo

<sup>22</sup> <<https://www.postgresql.org/>>

<sup>23</sup> <<http://postgis.net/>>

<sup>24</sup> <<https://geoserver.org/>>

<sup>25</sup> <<https://www.docker.com/>>

<sup>26</sup> <<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>>

Aplicação de Múltiplas Páginas (do inglês *Multiple Page Application*<sup>27</sup> - MPA), conforme apresentado na Figura 15, utilizando para isso a Linguagem de Marcação de Hipertexto (do inglês *HyperText Markup Language* HTML), a linguagem de estilo Folhas de Estilo em Cascata (do inglês *Cascading Style Sheet* CSS), e a linguagem de programação Javascript.

Em relação aos dados estáticos, a aplicação desenvolvida permite o *upload* de um arquivo do tipo *Comma Separated Values*<sup>28</sup> (csv). Esse arquivo é utilizado como fonte de dados para os algoritmos de MD, conforme explicado na Etapa 4.4 da Seção 4. Por fim, um segundo banco de dados é utilizado para gestão de usuários e eventuais dados que necessitem serem armazenados, como os relacionados a estratégias de otimização do processo de agrupamento. No contexto do sistema aqui desenvolvido, optou-se por utilizar o banco de dados relacional MySQL<sup>29</sup>.

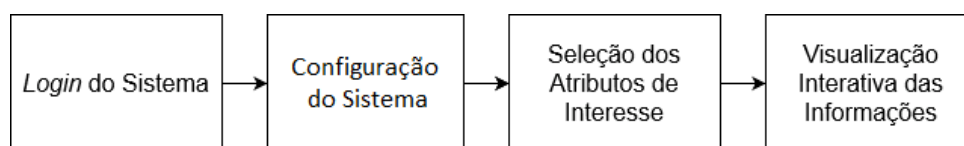


Figura 15 – Sequência de telas da aplicação proposta.

### 4.6.3 Modelo dos Arquivos como Fonte de Entrada

O sistema descrito foi pensado para realizar a leitura de arquivos no formato *csv*, através do *upload* feito pelo usuário. Dessa forma, os dados contidos nesse arquivo podem ter origem pública, privada ou ser uma mescla de ambos. Tal arquivo deve ser estruturado da seguinte forma: a primeira coluna deve conter o nome da cidade e a seguir deve constar os atributos. A Tabela 7 exemplifica um dos arquivos de teste criados.

Tabela 7 – Exemplo do formato do arquivo *csv* necessário para *upload*.

Nome do Município	IDH 2010	IDHM Renda 2010	IDHM Longevidade 2010	IDHM Educação 2010
Belo Vale	0.655	0.672	0.814	0.514
Betim	0.749	0.709	0.864	0.684
Bonfim	0.637	0.669	0.834	0.464
Brumadinho	0.747	0.761	0.857	0.639

A metodologia aqui proposta foi implementada em sistema real que será apresentado no próximo capítulo.

<sup>27</sup> <<https://imasters.com.br/front-end/por-que-utilizamos-single-page-applications-spa>>

<sup>28</sup> <<https://www.loc.gov/preservation/digital/formats/fdd/fdd000323.shtml>>

<sup>29</sup> <<https://www.mysql.com/>>

## 5 Resultados

Este capítulo busca apresentar os resultados obtidos a partir da metodologia elaborada na Seção 4. Para isso, ele é dividido da seguinte forma: na Seção 5.1 é apresentado o sistema desenvolvido e que reflete toda a metodologia anterior elaborada. Para validar a metodologia, na Seção 5.2 é realizado um caso de estudo em um cenário de desastre real com o sistema. Por fim, na Seção 5.3 são apresentados os resultados da avaliação do sistema obtidas juntamente com os possíveis usuários especialistas em solos.

### 5.1 *TerrainIntelGIS*

O sistema foi implementado visando auxiliar o usuário no processo de tomada de decisão de forma analítica, isto é, baseada em dados. Dessa forma, permite a aplicação e o uso de técnicas de Agrupamento em conjuntos de dados públicos ou privados, mesmo se quem utiliza a ferramenta desconhece a fundo tais técnicas. As seções seguintes apresentam a plataforma como um todo.

#### 5.1.1 Plataforma *TerrainIntelGIS*

Ao entrar no *link* do sistema, o usuário é redirecionado para a tela de acordo com a Figura 16. Nela, o usuário consegue realizar o *login* no sistema inserindo o seu e-mail e senha cadastrados. Caso o usuário esqueça a senha definida, é possível trocar ela clicando no botão "Esqueceu a senha?"(1). Além disso, a tela de cadastro no sistema também é acessada nessa interface, clicando no botão "Novo Usuário"(2).

Conforme a Figura 17, para realizar o cadastro no sistema, o usuário deve informar o seu e-mail, uma senha e o nome completo. Vale destacar que é com o e-mail informado que ele irá utilizar para fazer o *login* na aplicação.

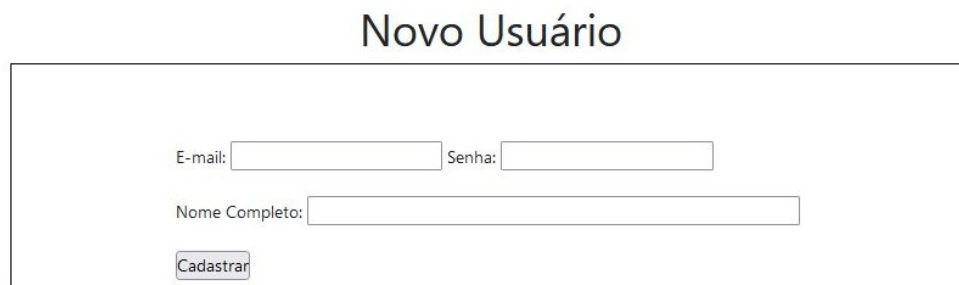
A Figura 18 exibe os parâmetros considerados obrigatórios para a aplicação da técnica de Agrupamento. Primeiramente, o usuário deve escolher se ele pretende inserir manualmente o valor dos agrupamentos a serem gerados (K) ou se esse número será definido com base no Método *Elbow* implementado no *backend* da aplicação. Caso ele opte por informar o valor de K, ele deverá escolher a opção "Sim"(1) para a pergunta "Informar o número de agrupamentos?". Com essa escolha, o campo *input* "Digite o valor"(2) irá surgir para preenchimento do usuário. Vale ressaltar que se a opção (1) for "Não", o campo (2) continuará oculto no sistema. Na sequência, basta o usuário selecionar o arquivo *csv* para *upload*, no padrão descrito na Seção 4.6.3, clicando em "Procurar"(3). Por fim, deve-se clicar no botão "Fazer a consulta", para dar continuidade ao processo.

Nessa próxima etapa, é exibido para o usuário cada um dos atributos contidos no arquivo *csv* que ele realizou *upload* na etapa anterior. Agora, basta ele indicar, marcando o *checkbox* dos atributos de interesse a serem considerados na aplicação do Algoritmo *KMeans*, como é mostrado na Figura 19. Ao clicar sobre o botão "Submit", grande parte da metodologia descrita na Seção 4 e implementada aqui é executada, conforme é exibido na Figura 20.

No primeiro momento, o Método *Elbow* (1) é executado, sendo definido uma estru-



A tela inicial de login do sistema TerrainIntelGIS. No topo, há um ícone de um globo terrestre com o nome "TerrainIntelGIS" logo abaixo. Abaixo do ícone, há dois campos de entrada: "E-mail" e "Senha". Na base da tela, há um botão "Entrar" e dois links: "Esqueceu a senha?" (marcado com o número 1) e "Novo usuário" (marcado com o número 2).

Figura 16 – Tela inicial de *login* do sistema.

A tela de cadastro de novo usuário, intitulada "Novo Usuário". Ela contém três campos de entrada: "E-mail:", "Senha:" e "Nome Completo:". Abaixo dos campos, há um botão "Cadastrar".

Figura 17 – Cadastro de um novo usuário no sistema.

tura de dados do tipo *array*, contendo em cada posição o número do agrupamento testado (valores entre 2 e 10) e o valor de inércia obtido. Com o *array* totalmente preenchido, é determinado o ponto de inflexão calculando a distância entre cada um dos valores de K até a reta formada entre os pontos inicial e final, isto é, o primeiro e último valor de K.

Com a definição do valor do número K ideal, o *KMeans* é agora de fato executado. Nesse momento, ocorre tanto a execução do Algoritmo considerando os dados com valores brutos quanto os dados sendo normalizados através do *MinMax*. Assim, é obtido os valores de todos os centroides em ambas execuções.

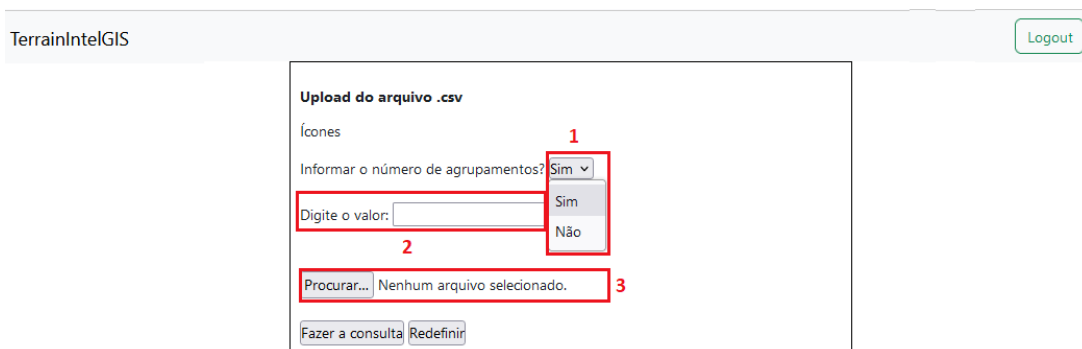


Figura 18 – Inserção dos valores para aplicação da técnica de Agrupamento.

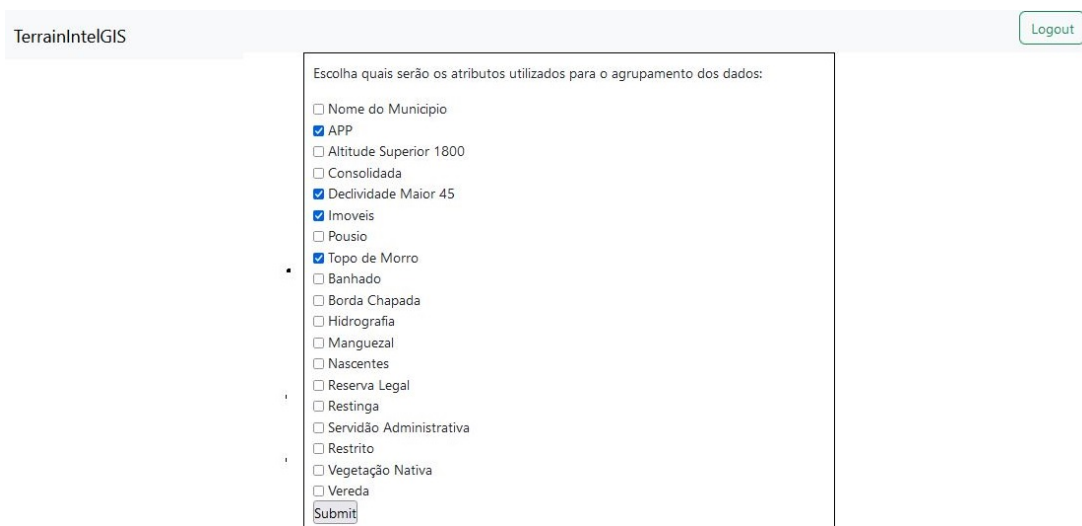


Figura 19 – Seleção dos atributos para serem considerados no Algoritmo *KMeans*.

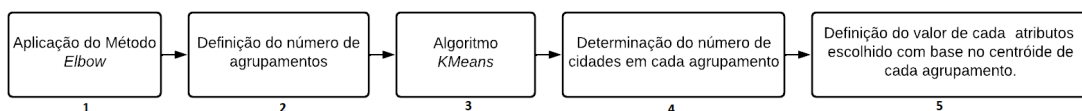
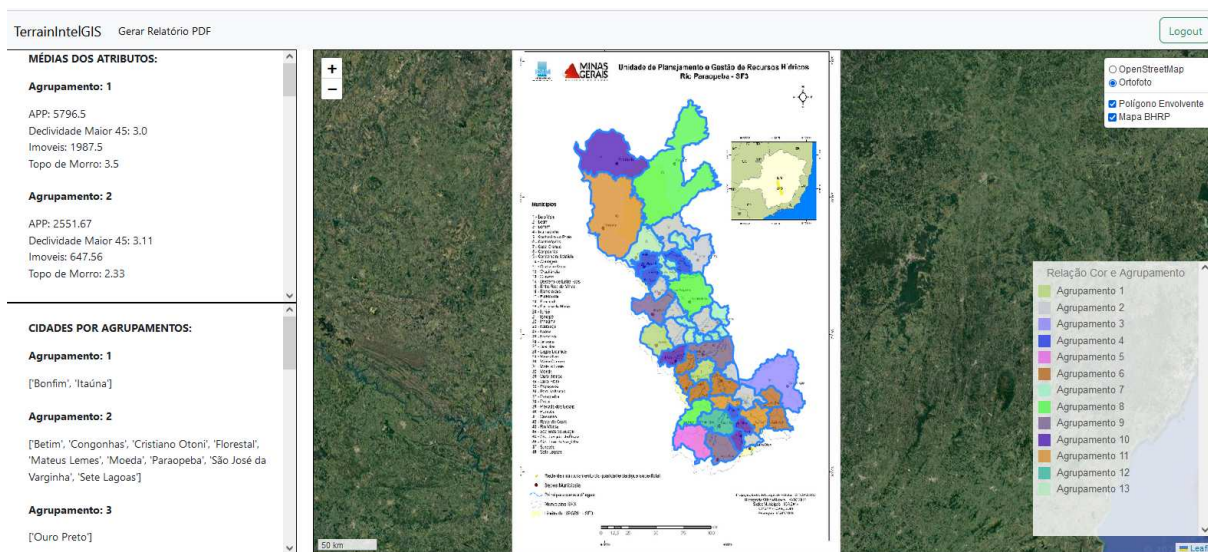


Figura 20 – Sequência de passos realizados pelo *backend* do sistema.

Além disso, outras duas funções são também executadas. A primeira relaciona cada uma das cidades com o agrupamento a qual pertence. Já a segunda, associa o valor de cada um dos atributos escolhidos pelo usuário com o seu respectivo agrupamento. Ambas funções são necessárias para a montagem da interface final do sistema, como é visto na Figura 21.

Por fim, após a execução de todas as funções implementadas no *backend* do sistema, usuário consegue visualizar a interface final. Conforme a Figura 21, todas as funcionalidades são exibidas e dispostas para o usuário visualizar e interagir, sendo a principal forma

de interação através do mapa. Além disso, conforme a Figura 30, pode ser visualizado também o mapa de calor, o que ajuda a visualizar o grau de importância de cada atributo em relação a cada um dos agrupamentos.



Mapa de Calor para os Agrupamentos Gerados

Figura 21 – Interface final do sistema.

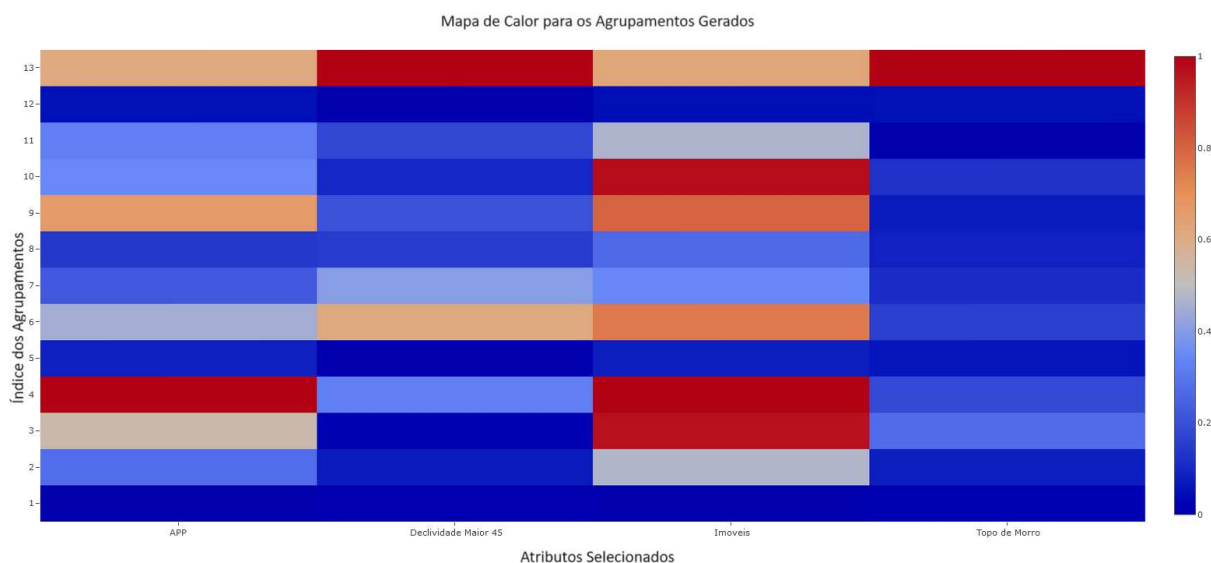


Figura 22 – Mapa de calor demonstrando a importância de cada atributo selecionado em cada um dos agrupamentos

### 5.1.2 Funcionalidades Implementadas

A Figura 23 destaca, na interface final do usuário, todas as ferramentas implementadas e disponíveis. Assim, as seguintes seções buscam elucidar cada uma delas.

#### 5.1.2.1 Gerar Relatório no formato *Portable Document Format - PDF*(1)

Permite exportar um resumo das informações apresentadas em tela (i.e. média dos atributos (1) e relação das cidades por agrupamento (2)). A geração de tal documento

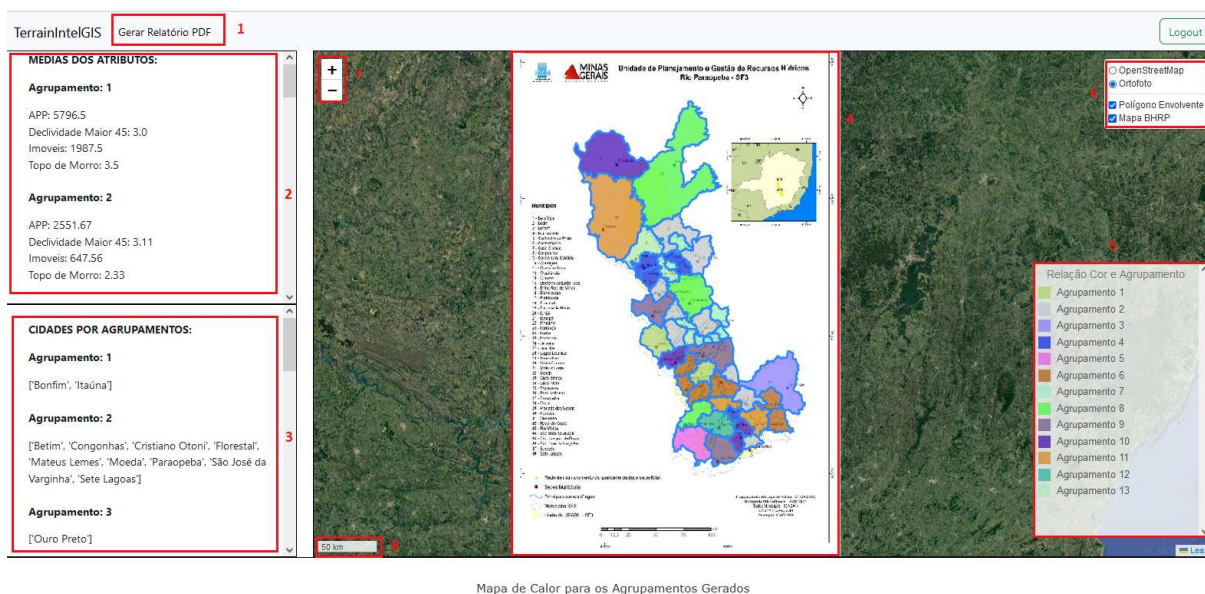


Figura 23 – Conjunto das funcionalidades desenvolvidas.

permite o compartilhamento de informações com outras pessoas que não possuem acesso ao sistema, além de funcionar como forma de *backup* e arquivamento da informação.

#### 5.1.2.2 Médias dos Atributos (2)

Para cada um dos atributos escolhidos pelo usuário, como pode ser verificado na Figura 19, é exibido, em relação a cada agrupamento formado, um valor de média. Esses valores correspondem ao centroide do respectivo agrupamento, que é um ponto representativo deles<sup>30</sup>. Como cada ponto do agrupamento representa um conjunto de atributos, logo, o centroide, que é o ponto que melhor representa o agrupamento<sup>31</sup>, representa os melhores valores dos atributos daquele grupo.

#### 5.1.2.3 Cidades por Agrupamentos (3)

Da mesma forma, para cada um dos agrupamentos formados, é exibida a relação das cidades que fazem parte dele. Dessa forma, o usuário tem a transparência de quais são os membros daquele agrupamento e que possuem como média os valores indicados em (1).

#### 5.1.2.4 Delimitação da Região de Interesse (4)

A delimitação da região de interesse foi feita através da utilização de um arquivo geoJson contendo a geometria da BHRP, que apresenta todos os 48 municípios da região. Além disso, utilizou-se também um mapa auxiliar do Comitê da Bacia Hidrográfica do Rio Paraopeba<sup>32</sup>. Para ser disponível, tal mapa seguiu o fluxo determinado pelo módulo 01 da arquitetura, igual é visto na Figura 14. Vale ressaltar que para ser visualizado nessa interface, o sistema consumiu um serviço WMS (*Web Feature Service*) do servidor de mapas.

<sup>30</sup> <<https://learn.microsoft.com/pt-br/azure/machine-learning/component-reference/k-means-clustering?view=azureml-api-2>>

<sup>31</sup> <<https://www.inf.ufpr.br/dagoncalves/IA06.pdf>>

<sup>32</sup> <<https://cbhsaofrancisco.org/comites-de-afluentes/cbh-do-rio-paraopeba-sf3-minas-gerais/>>

### 5.1.2.5 Relação Cor e Agrupamento (5)

Visando facilitar a compreensão final do usuário, cada cidade é preenchida com uma cor distinta. Tal cor representa a qual agrupamento a cidade pertence. Assim, no canto inferior direito, foi disponibilizado a relação da cor com o agrupamento. Tais cores são geradas de forma automática pelo sistema conforme a página é recarregada.

### 5.1.2.6 Funcionalidades Internas da Biblioteca de Mapas Interativos

Para desenvolvimento do mapa principal, utilizou a biblioteca *Leaflet*<sup>33</sup>. Optou-se por tal biblioteca por ser leve, personalizável, ter suporte a vários provedores de mapas, existir uma quantidade enorme de *plugins* disponíveis e ter uma comunidade bem ativa, além de ser *opensource*. Através dessa biblioteca, as seguintes funcionalidades podem ser facilmente disponibilizadas:

- Controle de Camadas (6): Possibilita que camadas-base e camadas de sobreposição sejam ativadas e desativadas por meio da seleção no painel;
- Zoom (7): Permite aproximar ou afastar em pontos do mapa;
- Barra de Escala (8): Fornece ao usuário dimensões reais do terreno com base nos níveis de zoom definido.

Na próxima seção, é apresentado um estudo de caso para mostrar a configuração e o funcionamento do sistema aqui desenvolvido em um cenário de desastre real. Assim, busca-se contextualizar o desastre e detalhar o passo-a-passo do sistema.

## 5.2 Caso de Estudo: Configuração da Plataforma *TerrainIntelGIS* para a Região da BHRP, área afetada pelo desastre de Brumadinho

### 5.2.1 Desastre de Brumadinho

No dia 25 de janeiro de 2019, ocorreu o rompimento da Barragem I, da mina Córrego do Feijão, em Brumadinho (Minas Gerais). Tal barragem era de responsabilidade da mineradora Vale S.A. e possuía cerca de 12 milhões de metros cúbicos de lama e rejeitos da mineração de ferro (OLIVEIRA; ROHLFS; GARCIA, 2019). Como consequência desse rompimento, estima-se que cerca de 270 pessoas morreram e 03 continuam até hoje desaparecidas<sup>34</sup>. Além dos óbitos e desaparecimentos, os resíduos do rompimento se estenderam por cerca de quase 270 hectares<sup>35</sup> (2,7 km<sup>2</sup>), sendo que desse total, 51% eram vegetação nativa, 19% eram áreas ocupadas por atividades com alta circulação de pessoas e 13% por atividades agropecuárias (PEIXOTO; ASMUS, 2020), resultando dessa forma em impactos ambientais, sociais e econômicos nas áreas atingidas.

<sup>33</sup> <<https://leafletjs.com/>>

<sup>34</sup> <<https://g1.globo.com/mg/minas-gerais/noticia/2023/01/25/quatro-anos-da-tragedia-em-brumadinho-270-mortes-tres-desaparecidos-e-nenhuma-punicao.ghtml>>

<sup>35</sup> <<https://www.canalrural.com.br/noticias/rompimento-destruiu-270-hectares/>>

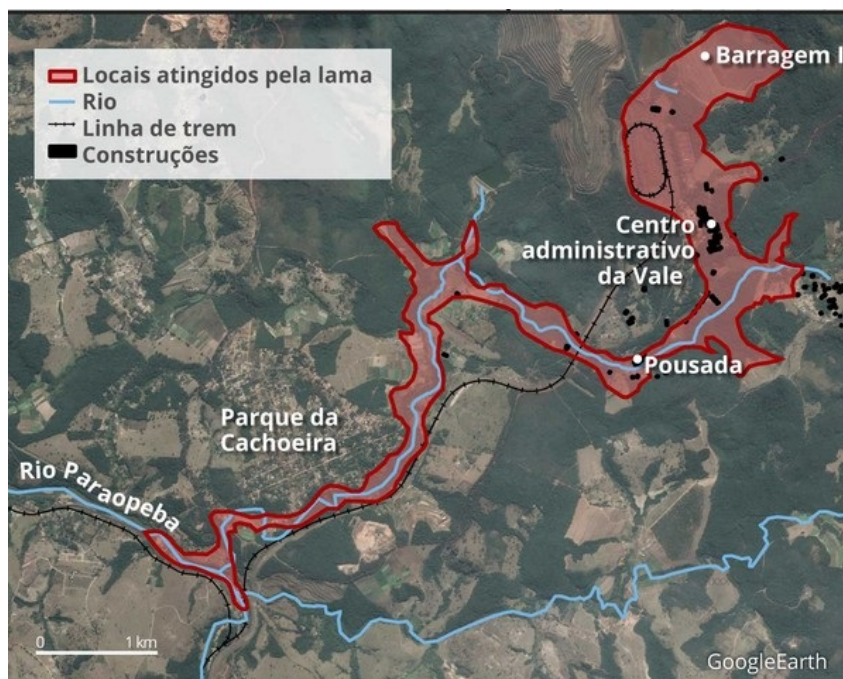


Figura 24 – Área do desastre de Brumadinho.

### 5.2.2 Área de Estudo

A área de estudo corresponde a BHRP, que é uma região localizada na parte sudeste do estado de Minas Gerais. Com uma área de 12.054,25 km<sup>2</sup>, abrange cerca de 48 municípios do estado. Com o rompimento da barragem, cerca de 1,3 milhão de habitantes foram impactados de forma direta ou indireta pelos rejeitos (SOARES, 2021).

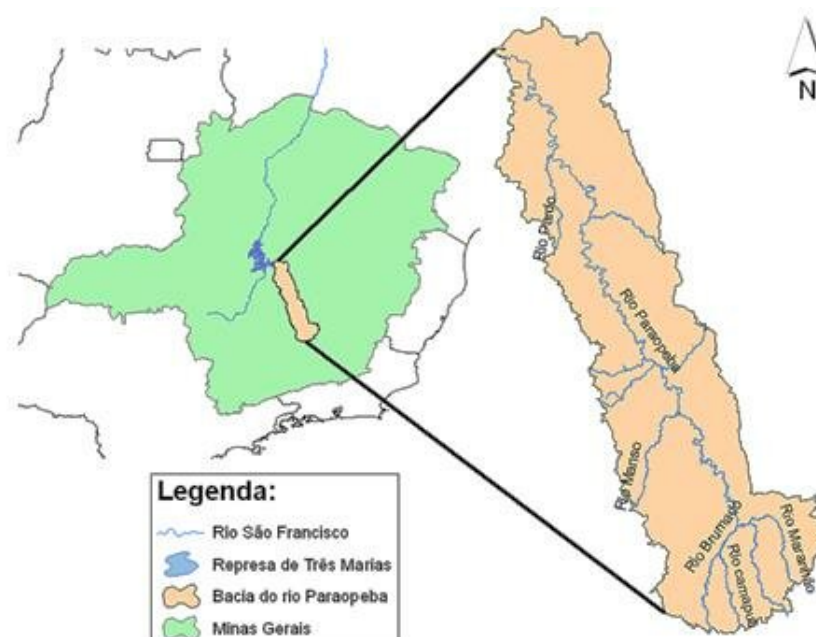


Figura 25 – Bacia Hidrográfica do Rio Paraopeba.

### 5.2.3 Fonte dos Dados

Para esse estudo, foram coletados dados públicos provenientes do Cadastro Ambiental Rural (CAR<sup>36</sup>) e do Instituto Brasileiro de Geografia e Estatística (IBGE<sup>37</sup>) para uso na plataforma. Dessa forma, busca-se caracterizar e refletir os aspectos demográficos, ambientais, sociais e econômicos da área. A relação dos dados, assim como sua classificação e fonte de origem, podem ser vistas conforme as Tabelas 2 e 3, da Seção 4.

### 5.2.4 Configurando a plataforma *TerrainIntelGIS*

#### 5.2.4.1 Cenário 01: A plataforma definindo o número de agrupamentos de forma automatizada

Nesse exemplo, foi realizado o uso do sistema no caso em que o usuário pretende deixar que a escolha do número de agrupamentos seja feita pela própria aplicação, através do Método *Elbow* implementado, conforme explicado na Subseção 5.1.1 desse capítulo. Assim, o primeiro passo que foi realizado foi selecionar a opção "Não", para o campo "Informar o número de agrupamentos?". Após isso, foi feito o *upload* do arquivo *csv* de nome "ibge+car.csv" e clicado em "Fazer a consulta".



Upload do arquivo .csv

Ícones

Informar o número de agrupamentos? Não ▾

Procurar... ibge+car.csv

Fazer a consulta Redefinir

Figura 26 – Configuração inicial do sistema.

Nesse momento, a aplicação disponibiliza em tela todos os atributos contidos no arquivo *csv* que foi realizado o *upload*. Cada coluna do arquivo é então convertida em uma opção com o *checkbox* associado. A Figura 27 exhibe a relação dos atributos do arquivo *csv* convertidos para campos do tipo *checkbox* dentro da aplicação.

Das opções disponíveis, a Figura 28 destaca quais foram as selecionadas para o exemplo.

Após clicado em "Submit", o sistema executou todo o passo-a-passo, conforme também explicado na Subseção 5.1.1. Em seguida, foi exibido para o usuário a tela final do sistema (Figura 29). Foi exibido também o mapa de calor (Figura 30), auxiliando o usuário na visualização e interpretação do resultado.

<sup>36</sup> <<https://www.car.gov.br/publico/imoveis/index>>

<sup>37</sup> <<https://cidades.ibge.gov.br/>>

TerrainIntelGIS Logout

Escolha quais serão os atributos utilizados para o agrupamento dos dados:

- Nome do Município
- APP
- Altitude Superior 1800
- Consolidada
- Declividade Maior 45
- Imóveis
- Pousio
- Topo de Morro
- Banhado
- Borda Chapada
- Hidrografia
- Manguezal
- Nascentes
- Reserva Legal
- Restinga
- Servidão Administrativa
- Restrito
- Vegetação Nativa
- Vereda
- População estimada(2021)
- População último censo(2010)
- Densidade demografica(2010)
- Salario medio mensal(2019)
- Pessoal ocupado(2019)
- Populacao ocupada(2019)
- Percentual da populacao com renda de ate meio salario(2010)
- Taxa escolarizacao 6 a 14 anos(2010)
- IDEB anos iniciais ensino fundamental(2019)
- IDEB anos finais ensino fundamental(2019)
- Matriculas ensino fundamental(2020)
- Matriculas ensino medio(2020)

Figura 27 – Tela com todos os atributos do arquivo *csv* após *upload*.

Tabela 8 – Relação dos atributos selecionados para o exemplo.

Nome do atributo	Orgão de origem
APP	CAR
Imóveis	CAR
Hidrografia	CAR
Reserva Legal	CAR
População último censo (2010)	IBGE
Densidade demográfica (2010)	IBGE
Pessoal ocupado (2019)	IBGE
População ocupada (2019)	IBGE
IDHM (2010)	IBGE
Área territorial (2021)	IBGE

#### 5.2.4.1.1 Conclusões do Exemplo 01

Nesse exemplo, o valor ideal determinado para o número de agrupamentos foi  $K=3$ . Um resumo do resultado obtido pode ser visto na Tabela 9. Nela, para cada um dos agrupamentos, foram destacados as cidades que o compõem, assim como os valores que

TerrainIntelGIS Logout

Escolha quais serão os atributos utilizados para o agrupamento dos dados:

- Nome do Município
- APP
- Altitude Superior 1800
- Consolidada
- Declividade Maior 45
- Imoveis
- Pousio
- Topo de Morro
- Banhado
- Borda Chapada
- Hidrografia
- Manguezal
- Nascentes
- Reserva Legal
- Restinga
- Servidão Administrativa
- Restrito
- Vegetação Nativa
- Vereda
- População estimada(2021)
- População último censo(2010)
- Densidade demografica(2010)
- Salario medio mensal(2019)
- Pessoal ocupado(2019)
- Populacao ocupada(2019)
- Percentual da populacao com renda de ate meio salario(2010)
- Taxa escolarizacao 6 a 14 anos(2010)
- IDEB anos iniciais ensino fundamental(2019)
- IDEB anos finais ensino fundamental(2019)
- Matriculas ensino fundamental(2020)
- Matriculas ensino medio(2020)
- Docentes ensino fundamental(2020)
- Docentes ensino medio(2020)
- Numeros estabelecimentos ensino fundamental(2010)
- Numeros estabelecimentos ensino medio(2020)
- PIB per capita(2019)
- Percentual receitas fontes externas(2015)
- IDHM(2010)
- Total receitas realizadas(2017)
- Total receitas empenhadas(2017)
- Mortalidade infantil(2019)
- Internacoes por diarreias(2016)
- Estabelecimentos SUS(2009)
- Area territorial(2021)
- Esgotamento adequado(2010)
- Arborizacao vias publicas(2010)
- Urbanizacao vias publicas(2010)
- Bioma

Submit

Figura 28 – Seleção dos atributos de interesse pelo usuário.

melhor os representam, conforme as propriedades selecionadas.

Analisando o mapa de calor obtido (Figura 30), pode-se afirmar que:

- O agrupamento 1 é formado principalmente por extensos municípios (valor grande de Área Territorial), predominantemente rurais (altos valores para APP e Imoveis). Aliado a isso, o baixo número da população faz com que a densidade demográfica também seja baixa. Por apresentar muitos imóveis rurais, supõem-se que em sua maioria sejam famílias que praticam uma agricultura de subsistência. Dessa forma, percebe-se que são cidades que necessitam diretamente dos recursos hídricos disponíveis para o desenvolvimento da economia local.
- O agrupamento 2 já é caracterizado por cidades menores e com a população predominantemente urbana (baixo valor de Área Territorial e baixos valores de APP e Imoveis, respectivamente). Como a área territorial é menor, consequentemente esses municípios apresentam uma alta densidade demográfica. Além disso, como existe uma área rural menor, os empregos estão provavelmente concentrados em torno dos setores secundário e terciário da economia.

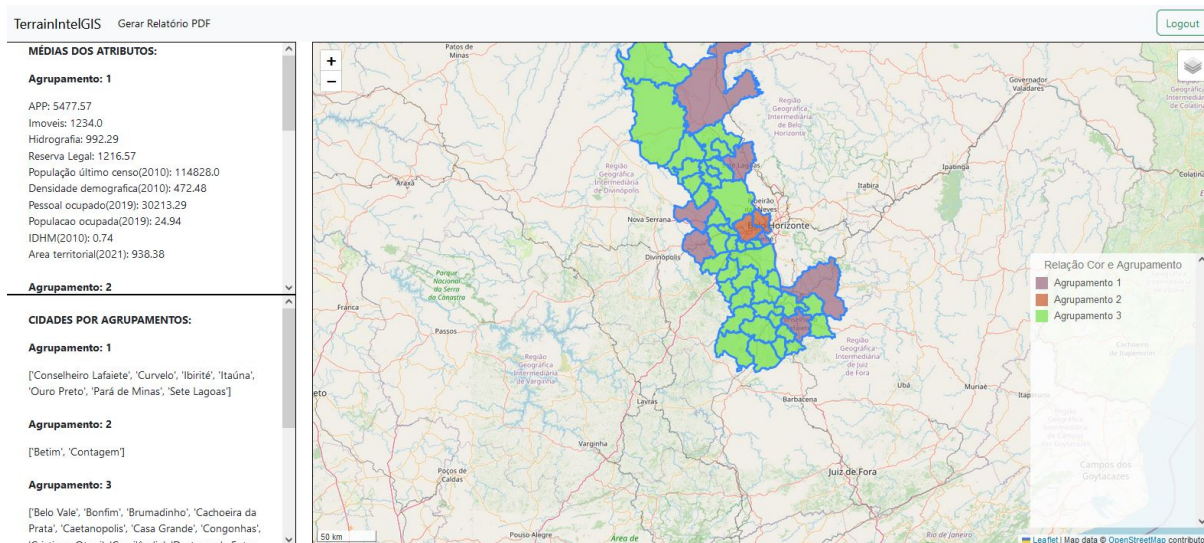


Figura 29 – Disponibilização dos resultados após aplicação das técnicas pelo *backend* do sistema.

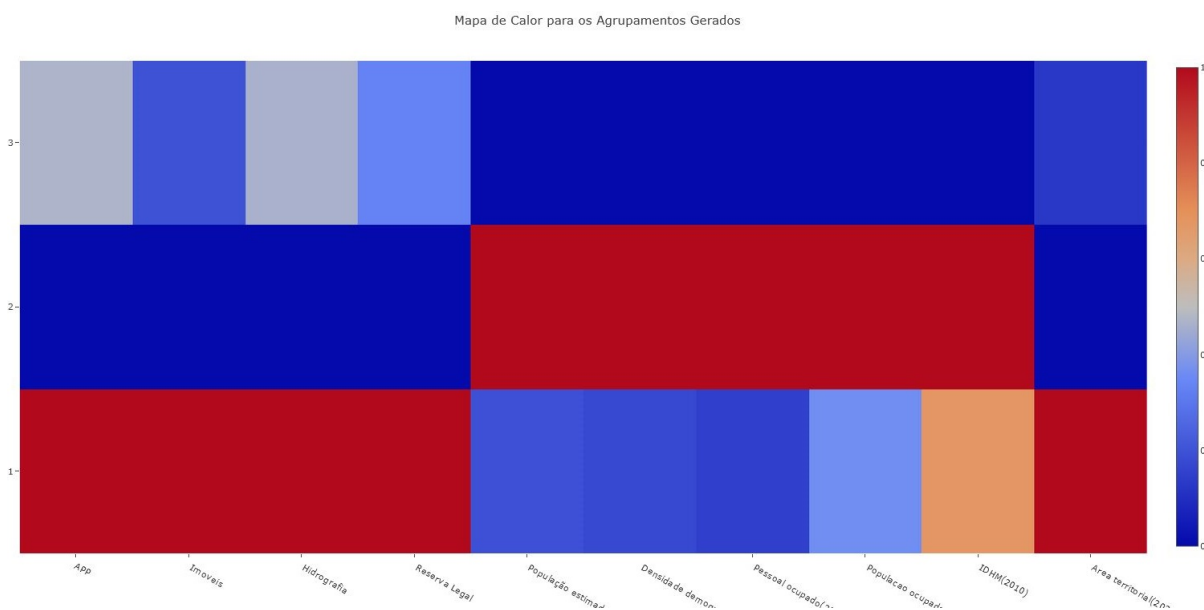


Figura 30 – Mapa de calor gerado.

- Por fim, o agrupamento 3 se caracteriza por possuir cidades de tamanho intermediário, com moderada concentração de APP, rios e áreas com mata nativa. Por apresentar poucos imóveis rurais e uma população empregada, em sua maioria, de forma informal, supõem-se que a maioria das propriedades seja de porte grande e que empreguem muitas dessas pessoas.

#### 5.2.4.2 Cenário 02: O usuário informando o número de agrupamentos

Nesse outro exemplo, foi realizado a escolha do número de agrupamento de forma manual, considerando  $K=2$ . Para isso, selecionou-se o valor "Sim" no campo "Informar o número de agrupamentos?". Em seguida, digitou-se o número de agrupamentos pretendido, e em seguida foi realizado o *upload* do arquivo *csv*.

Após clicar em "Fazer consulta", foram selecionados os atributos de interesse para

Tabela 9 – Resultados obtidos após a realização do exemplo 01.

Número do Agrupamento	Cidades do Agrupamento	Valores médios obtidos
1	Conselheiro Lafaiete, Curvelo, Ibirité, Itaúna, Ouro Preto, Pará de Minas, Sete Lagos	APP: 5477,57
		Imoveis: 1234,0
		Hidrografia: 992,29
		Reserva Legal: 1216,57
		População último censo (2010): 11428,0
		Densidade demográfica (2010): 472,48
		Pessoal ocupado (2019): 30213,29
		População ocupada (2019): 24,94%
		IDHM (2010): 0,74
		Área territorial (2021): 938,38
2	Betim, Contagem	APP: 1644,5
		Imoveis: 616,0
		Hidrografia: 297,5
		Reserva Legal: 474,5
		População último censo (2010): 490765,5
		Densidade demográfica (2010): 2096,56
		Pessoal ocupado (2019): 165768,5
		População ocupada (2019): 29,55%
		IDHM (2010): 0,75
		Área territorial (2021): 269,32
3	Belo Vale, Bonfim, Brumadinho, Cachoeira da Prata, Caetanópolis, Casa Grande, Congonhas, Cristiano Ottoni, Crucilândia, Desterro de Entre Rios, Entre Rios de Minas, Esmeraldas, Felixlândia, Florestal, Fortuna de Minas, Igarapé, Inhaúma, Itatiaiuçu, Itaverava, Jeceaba, Juatuba, Lagoa Dourada, Maravilhas, Mario Campos, Mateus Lemes, Moeda, Ouro Branco, Papagaios, Paraopeba, Pequi, Piedade dos Gerais, Pompéu, Queluzito, Resende Costa, São Brás do Suaçuí, São Joaquim de Bicas, São José da Varginha, Sarzedo	APP: 3736,51
		Imoveis: 785,1
		Hidrografia: 670,05
		Reserva Legal: 769,67
		População último censo (2010): 13669,95
		Densidade demográfica (2010): 72,58
		Pessoal ocupado (2019): 3197,54
		População ocupada (2019): 19,14%
		IDHM (2010): 0,68
		Área territorial (2021): 368,63

**Upload do arquivo .csv**

Ícones

Informar o número de agrupamentos?  ▾

Digite o valor:

ibge+car.csv

Figura 31 – Definição manual do número de agrupamentos.

o exemplo. Nesse caso, os atributos selecionados foram os mesmos disponíveis na Tabela 8 anterior. Por fim, foi exibida a tela final do sistema, contendo as médias dos valores de cada agrupamento e o mapa de calor gerado, conforme as Figuras 32 e 32, respectivamente.

#### 5.2.4.2.1 Conclusões do Exemplo 02

Nesse segundo exemplo, o valor foi definido  $K = 2$ , sendo escolhido de forma arbitrária. Um resumo do resultado obtido pode ser visto na Tabela 10. Assim como para

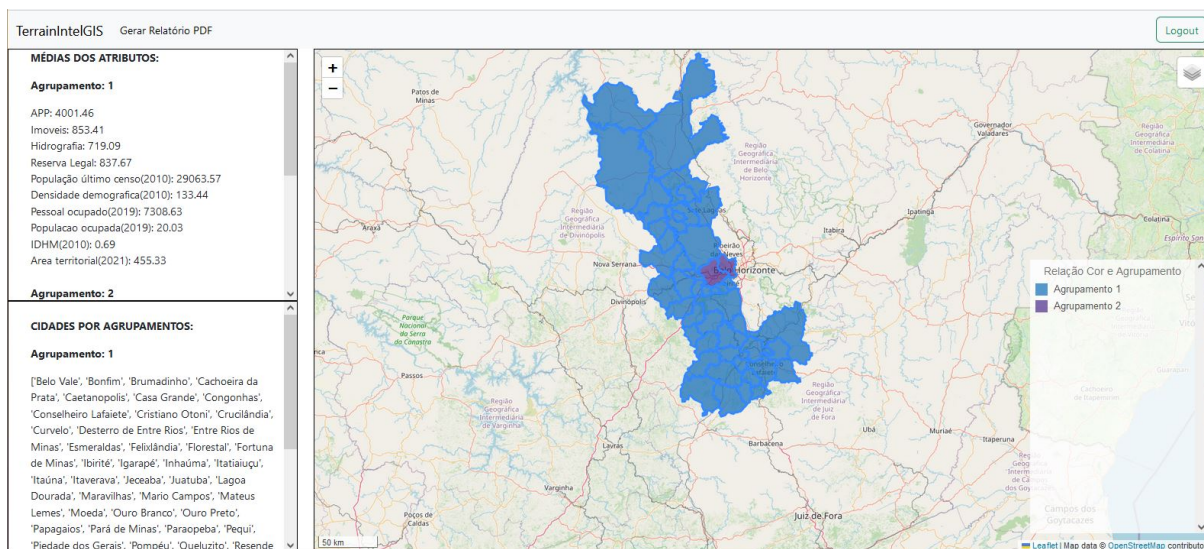


Figura 32 – Tela principal do sistema.

Mapa de Calor para os Agrupamentos Gerados

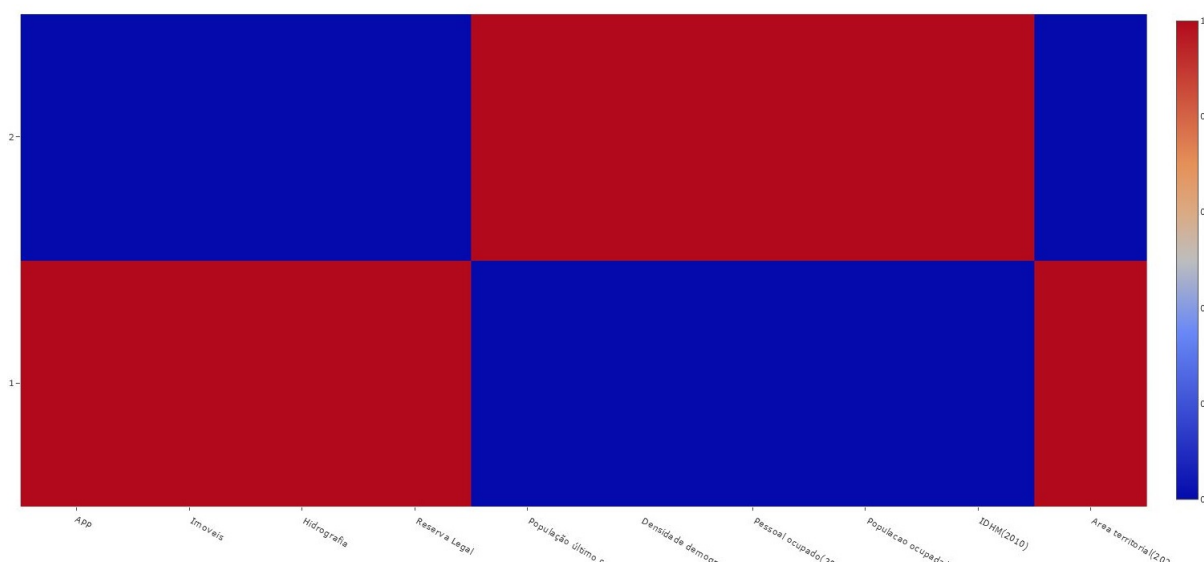


Figura 33 – Mapa de calor gerado para o exemplo definido.

o cenário anterior, para cada um dos dois agrupamentos, foram destacados as cidades que os compõem, assim como os valores que melhor o representam, conforme as propriedades selecionadas.

Com base no mapa de calor representado na Figura 33, pode-se analisar em relação a cada um dos agrupamentos formados:

- O agrupamento 1 apresenta características rurais, uma vez que conta com altos valores para atributos como o número de APP, Imóveis e Hidrografia. Além disso, as cidades apresentam uma área territorial elevada e baixa densidade demográfica, o que são características típicas de municípios localizados mais no interior dos estados.
- Já o agrupamento 2, ao contrário do anterior, caracteriza-se por possuir um valor mais elevado para atributos como população e IDH. Supõe-se que os municípios

Tabela 10 – Resultados obtidos após a realização do exemplo 02.

Número do Agrupamento	Cidades do Agrupamento	Valores Médios Obtidos
1	Belo Vale, Bonfim, Brumadinho, Cachoeira da Prata, Caetanópolis, Casa Grande, Congonhas, Conselheiro Lafaiete, Cristiano Ottoni, Crucilândia, Curvelo, Desterro de Entre Rios de Minas, Esmeraldas, Felixlândia, Florestal, Fortuna de Minas, Ibirité, Igarapé, Inhaúma, Itatiaiuçu, Itaúna, Itaverava, Jeceaba, Juatuba, Lagoa Dourada, Maravilhas, Mário Campos, Mateus Lemes, Moeda, Ouro Branco, Ouro Preto, Papagaios, Pará de Minas, Paraopeba, Pequi, Piedade dos Gerais, Pompéu, Queluzito, Resende Costa, Rio Manso, São Brás do Suaçuí, São Joaquim de Bicas, São José da Varginha, Sarzedo e Sete Lagoas	APP: 4001,96
		Imóveis: 853,41
		Hidrografia: 719,09
		Reserva Legal: 837,67
		População Último Censo (2010): 29063,57
		Densidade Demográfica (2010): 133,44
		Pessoal Ocupado (2019): 7308,63
		População Ocupada (2019): 20,03%
		IDHM (2010): 0,69
		Área Territorial (2021): 455,33
		APP: 1644,5
		Imóveis: 616,0
2	Betim, Contagem	Hidrografia: 297,5
		Reserva Legal: 474,5
		População Último Censo (2010): 490765,5
		Densidade Demográfica (2010): 2096,56
		Pessoal Ocupado (2019): 165768,5
		População Ocupada (2019): 29,55%
		IDHM (2010): 0,75
		Área Territorial (2021): 269,32

sejam localizados mais próximos a capitais, uma vez que apresentam um alto valor para a densidade demográfica.

### 5.3 Avaliação Preliminar do Sistema

Nesta seção, busca-se apresentar os resultados de uma avaliação quantitativa realizada juntamente a um laboratório de pesquisa do Departamento de Solos (DPS) da UFV. Tal laboratório foi escolhido tendo em vista que uma das suas áreas de pesquisa é justamente em relação à recuperação de áreas degradadas por desastres ambientais.

#### 5.3.1 Metodologia

Toda a avaliação aconteceu por meio digital, utilizando para isso o *Google Forms*, uma versão do sistema implementada e um arquivo *csv* previamente preenchido. Tal avaliação foi conduzida em três etapas. No primeiro momento, cada um dos participantes realizou a leitura das Orientações Gerais. Ao concordarem com os termos, eles foram submetidos a um questionário Pré-Teste. Através desse questionário, buscou-se conhecer a experiência deles com esse tipo de ferramenta como a qual estava sendo proposta. Após esse primeiro questionário, os participantes seguiram para a próxima etapa, que consistiu em um conjunto de tarefas, duas tarefas propostas, sendo que tais tarefas compreendiam todas as funcionalidades implementadas no sistema.

Por fim, os participantes foram orientados a preencherem um questionário Pós-Teste. Nesse momento, tal questionário continha um conjunto de 26 perguntas, sendo 23 de múltiplas escolhas e 3 perguntas abertas. Em relação às perguntas de múltiplas escolhas, 10 delas possuíam escala de 1 a 7, conforme definido pelo *System Usability Scale* (SUS) (JOHN, 1996) e 13 delas estavam com escala definida conforme o *System Acceptance Scale* (SAS) (LAAN; HEINO; WAARD, 1997). Por fim, as perguntas abertas eram voltadas para identificação de alguma propriedade ou valor presente em tela, buscando assim verificar a compreensão de tudo o que era exibido para o usuário. Todo o material está disponível também nos apêndices A, B, C e D.

## 5.3.2 Resultados Obtidos

### 5.3.2.1 Questionário Pré-Teste

No primeiro momento, buscou-se compreender a familiaridade dos possíveis usuários com esse tipo de ferramenta. Dessa forma, conclui-se que o sistema juntamente com a metodologia aqui desenvolvida representa, até certo ponto, uma novidade para a área, uma vez que 100% dos entrevistados nunca utilizaram nenhum tipo de plataforma semelhante, como pode ser visto na Figura 34.



Figura 34 – Gráfico gerado com as respostas do Pré-Teste

### 5.3.2.2 Questionário Pós-Teste

Os questionários SAS e SUS foram aplicados no pós-teste. Sendo assim, os seguintes resultados foram obtidos:

#### 5.3.2.2.1 Questionário SAS

A aplicação desse questionário buscou validar aspectos relacionados a facilidade (Figuras 35, 36, 37 e 38), a utilidade (Figuras 39, 40, 41 e 42) e a intenção de uso da aplicação (Figuras 43 e 44). Dessa forma, foram obtidas as seguintes respostas.

Conforme as Figuras, é possível concluir que a utilidade desse sistema é verdadeiramente notável e que dessa forma busca facilitar a vida das pessoas, otimizando processos. Os benefícios práticos desse sistema são evidentes, mas sua verdadeira grandeza reside na capacidade de inspirar confiança nos usuários, uma vez que apresenta bons valores de intenção de uso pelos possíveis usuários.

#### 5.3.2.2.2 Questionário SUS

Já a aplicação desse tipo de questionário buscou metrificar aspectos de efetividade, eficiência e satisfação<sup>38</sup>.

<sup>38</sup> <<https://brasil.uxdesign.cc/o-que-%C3%A9-o-sus-system-usability-scale-e-como-us%C3%A1-lo-em-seu-site-6d63224481c8>>

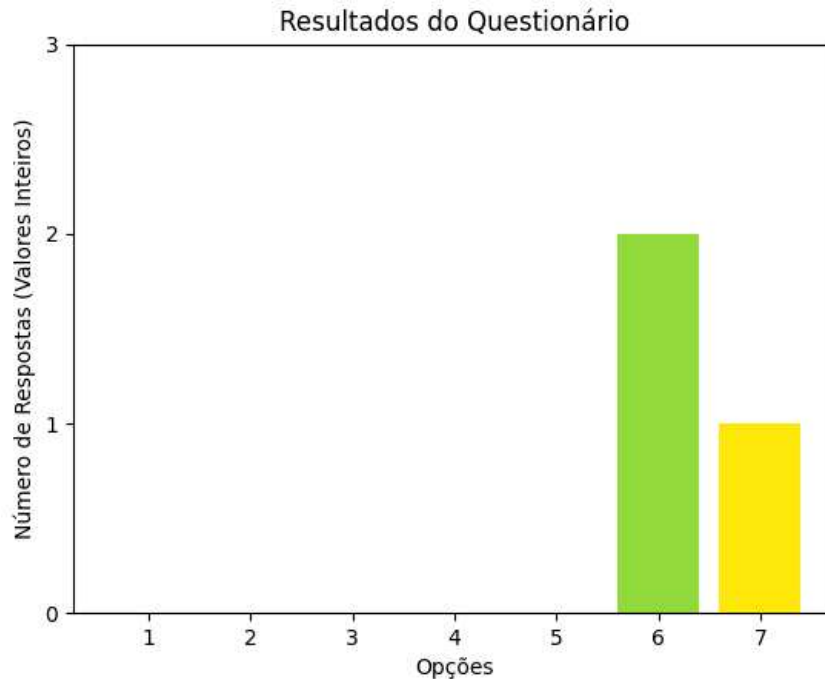


Figura 35 – Minha interação com a ferramenta *TerrainIntelGIS* foi clara e compreensível.

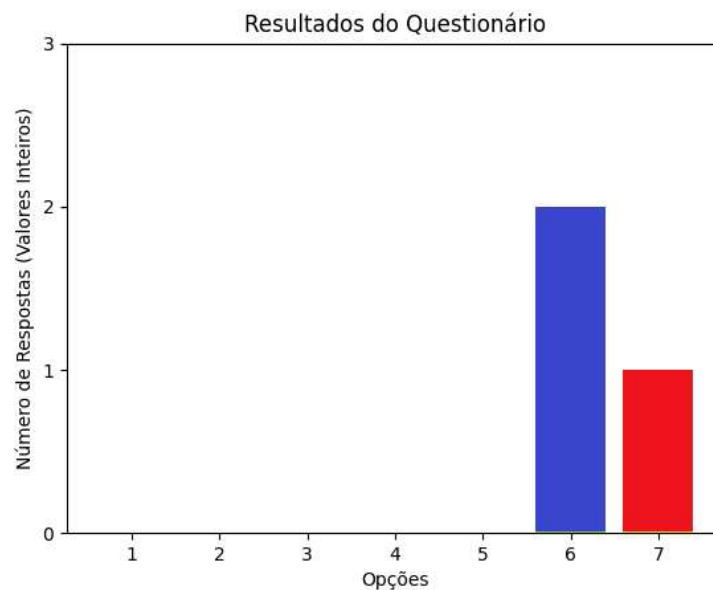


Figura 36 – Interagir com a ferramenta *TerrainIntelGIS* não exige muito do meu esforço mental.

Assim, avaliando as respostas sob a ótica da heurística de Nielsen (NIELSEN, 1993), pode-se concluir que:

- Considerando a Figura 46, o sistema poderia ser mais simples, contando com um número menor de informações disponíveis em uma mesma tela, o que facilitaria a memorização do conjunto de para o usuário;
- Pela Figura 50, haviam ainda *bugs* na ferramenta que precisam ser identificados e corrigidos, eliminando assim essas inconsistências;

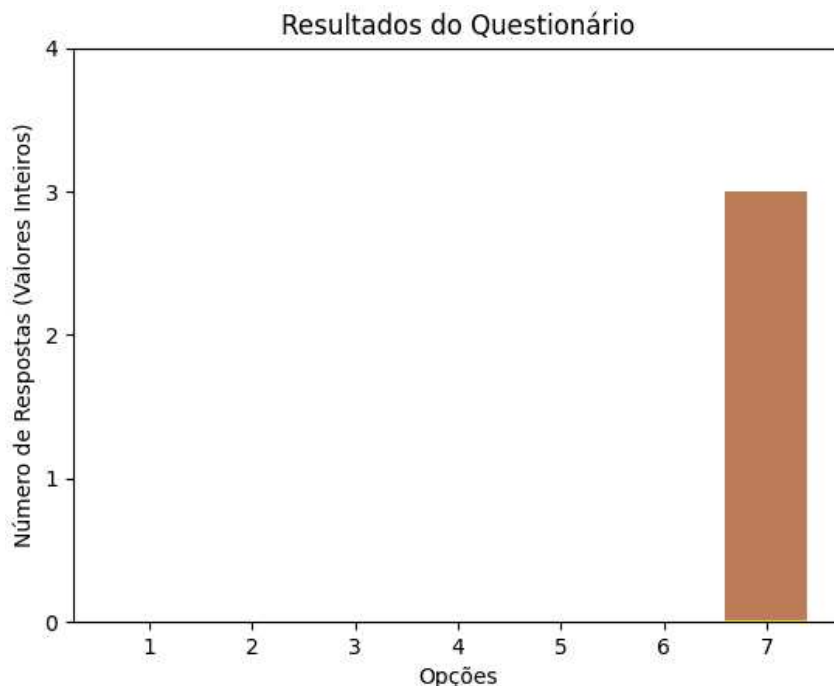


Figura 37 – Considero a ferramenta *TerrainIntelGIS* fácil de usar.

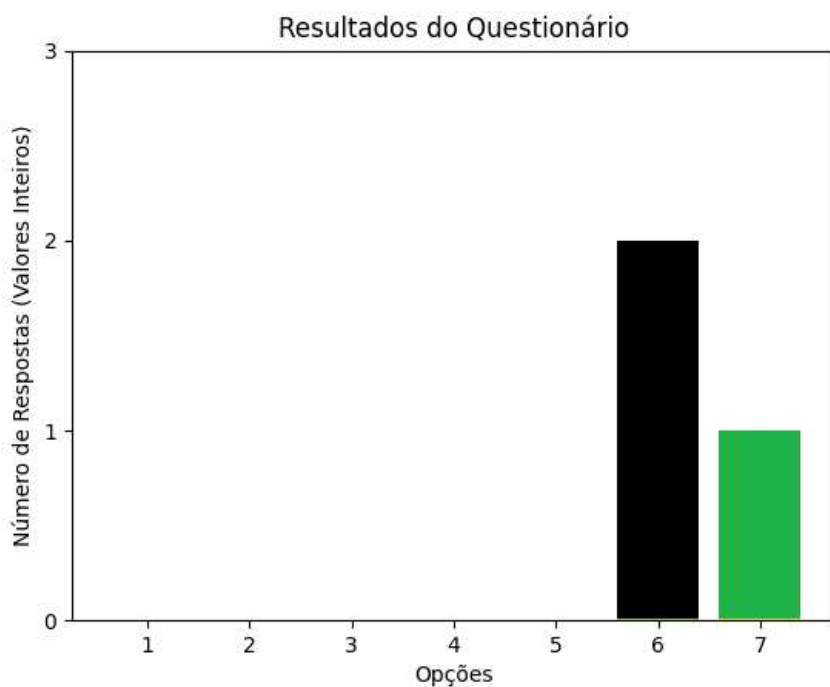


Figura 38 – Eu acho fácil conseguir que a ferramenta *TerrainIntelGIS* faça o que eu quero que ela faça.

- Tendo em vista as Figuras 49, 50 e 52, pode-se concluir que as funcionalidades desenvolvidas são interessantes sim e cumprem o que se propõe a fazer;
- Tendo em vista as Figuras 47, 48, 51 e 54, o sistema é bem claro, não necessitando de que os usuários tenham que aprender novos conceitos relacionados para fazerem o uso da ferramenta;
- Por fim, levando em conta as Figuras 45, 48 e 53, o sistema pode sim ser utilizado

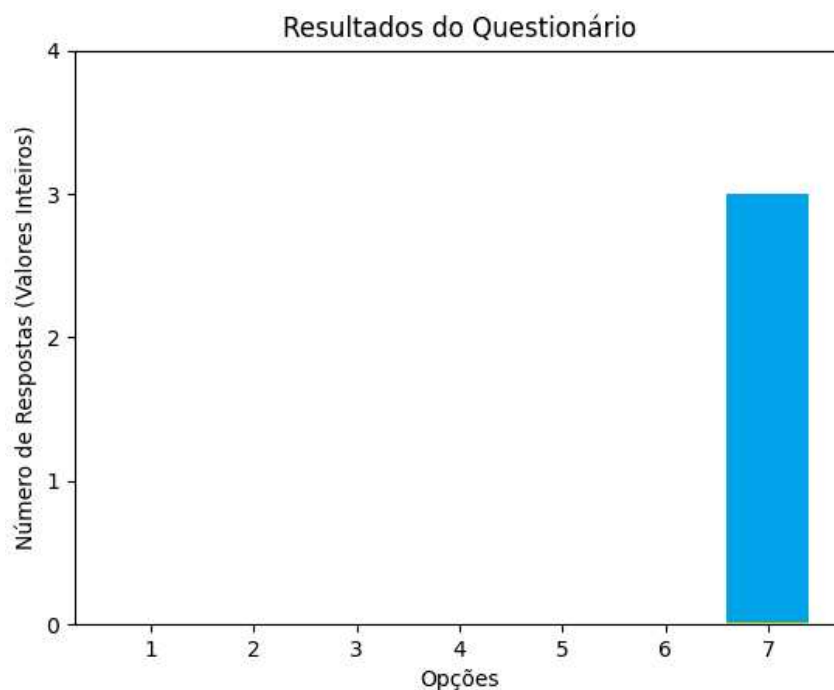


Figura 39 – Utilizar a ferramenta *TerrainIntelGIS* pode melhorar o meu desempenho na atividade de agrupar cidades da região de interesse.

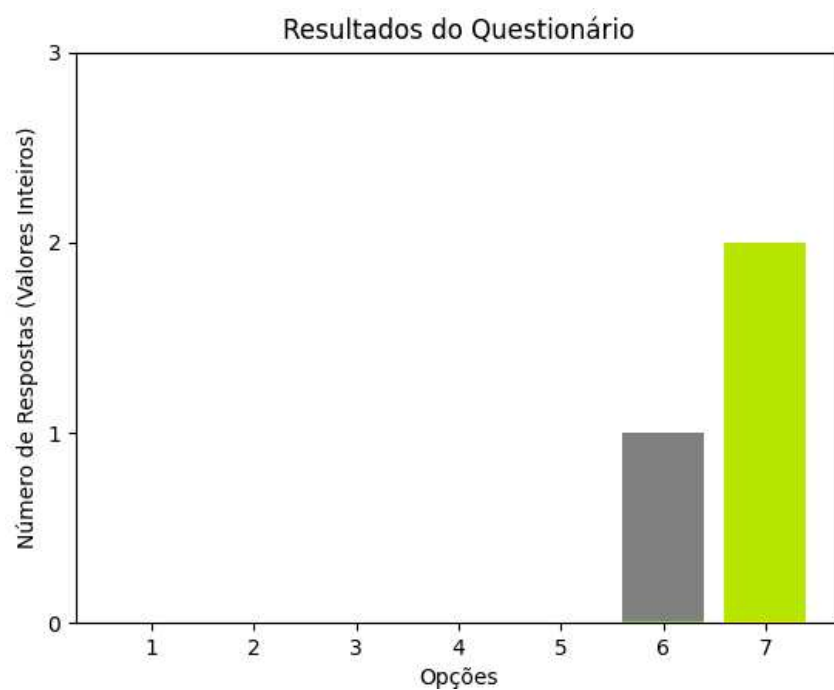


Figura 40 – Usar a ferramenta *TerrainIntelGIS* no meu trabalho pode melhorar a minha produtividade na atividade de agrupar cidades da região de interesse.

pelos usuários, podendo até mesmo fazer parte do fluxo de trabalho do grupo.

De maneira geral, o sistema pode ser sim considerado útil para os propósitos que se dispões a fazer, uma vez que ele é intuitivo, possui funcionalidades consideradas interessantes e apresenta intenção de uso por parte dos possíveis usuários do grupo.

Para validar a metodologia desenvolvida e o sistema implementado, realizou-se

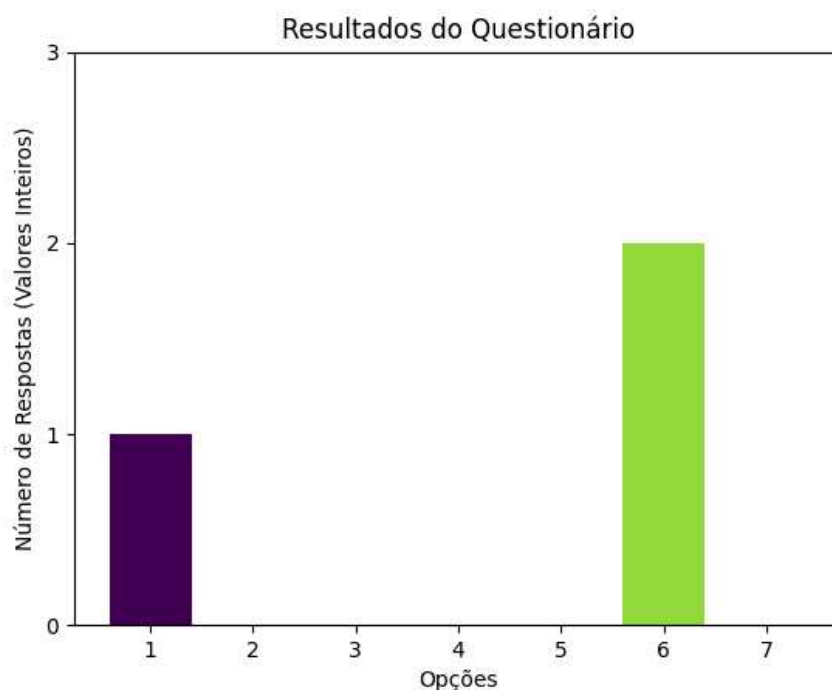


Figura 41 – Usar a ferramenta *TerrainIntelGIS* pode aumentar a minha eficácia no meu trabalho.

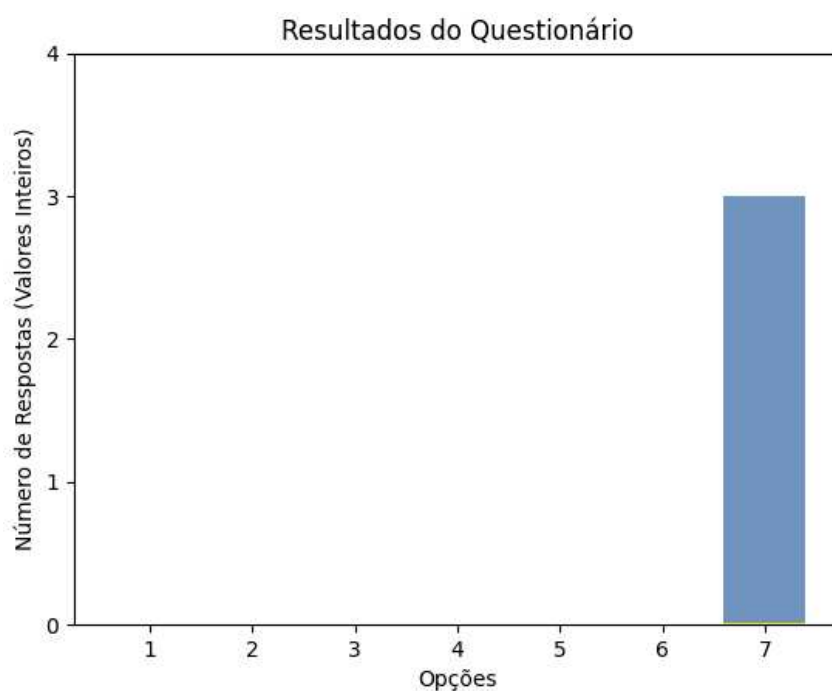


Figura 42 – Eu considero a ferramenta *TerrainIntelGIS* será útil para agrupar cidades de uma mesma região de interesse.

uma pesquisa quantitativa com um grupo de pesquisa do Departamento de Solos (DPS) da UFV. Tal grupo foi escolhido tendo em vista que uma das suas áreas de pesquisa é justamente em relação à recuperação de áreas degradadas por desastres ambientais.

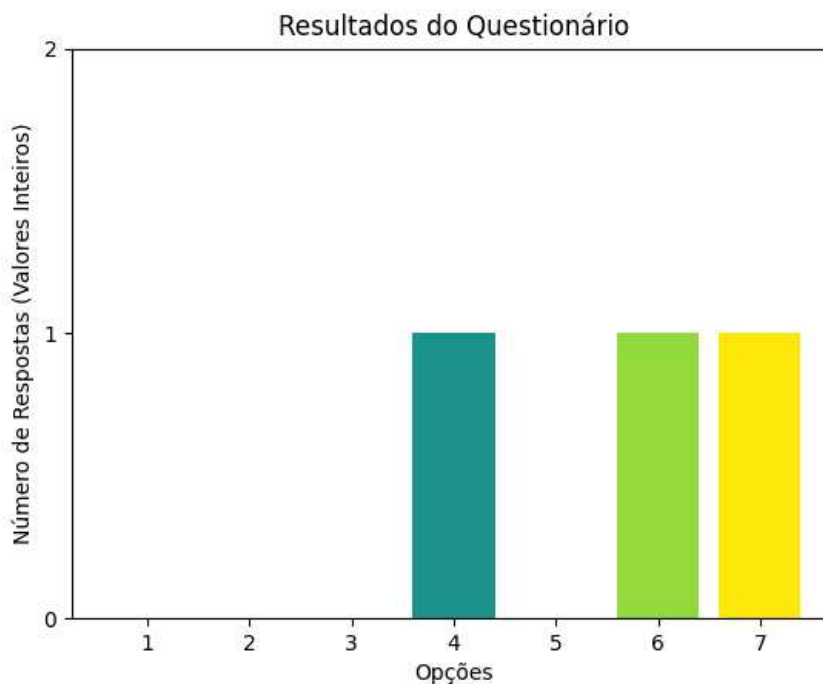


Figura 43 – Se eu tiver acesso à ferramenta *TerrainIntelGIS*, eu pretendo utilizá-la.

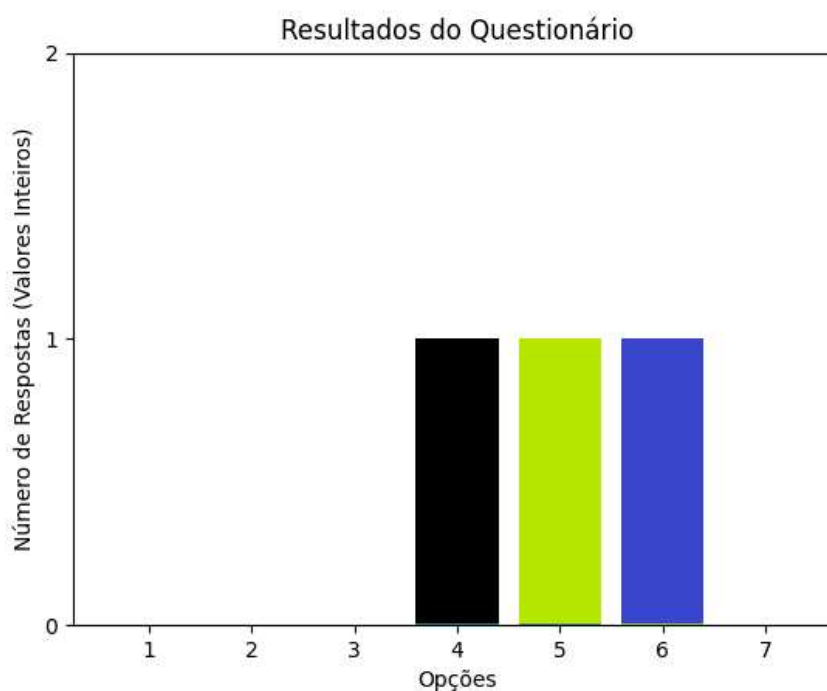


Figura 44 – Levando em conta que poderei escolher usar ou não a ferramenta *TerrainIntelGIS*, eu prevejo que irei utilizá-la.

### 5.3.3 Metodologia

Sendo assim, a avaliação aconteceu por meio digital, utilizando para isso o *Google Forms*, uma versão do sistema implementada e um arquivo *csv* previamente preenchido. Tal avaliação foi conduzida em três etapas. No primeiro momento, cada um dos participantes realizou a leitura das Orientações Gerais. Ao concordarem com os termos, eles foram submetidos a um questionário Pré-Teste. Através desse questionário, buscou-se conhecer a experiência deles com esse tipo de ferramenta como a qual estava sendo proposta. Após

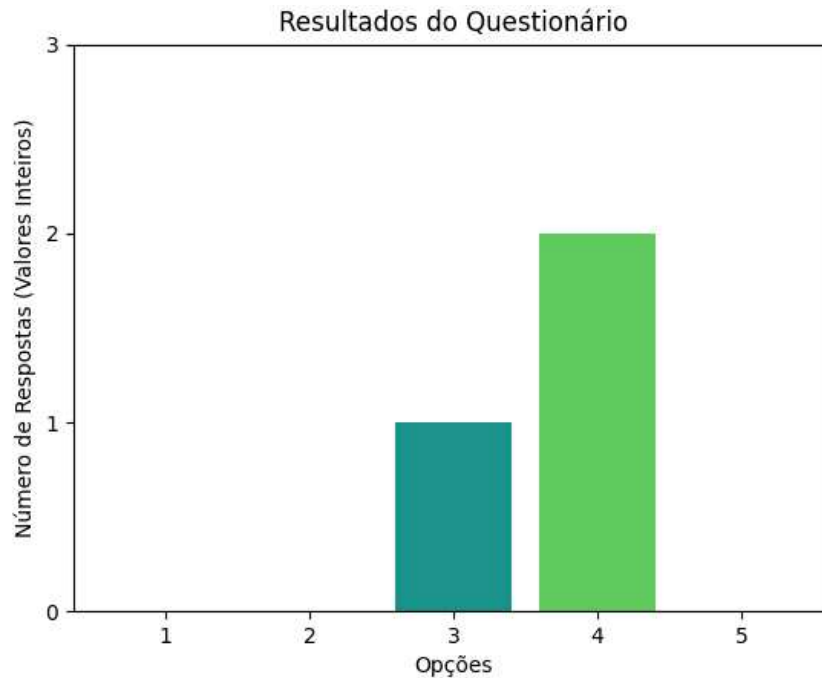


Figura 45 – Eu penso que gostaria de usar esse sistema com frequência.

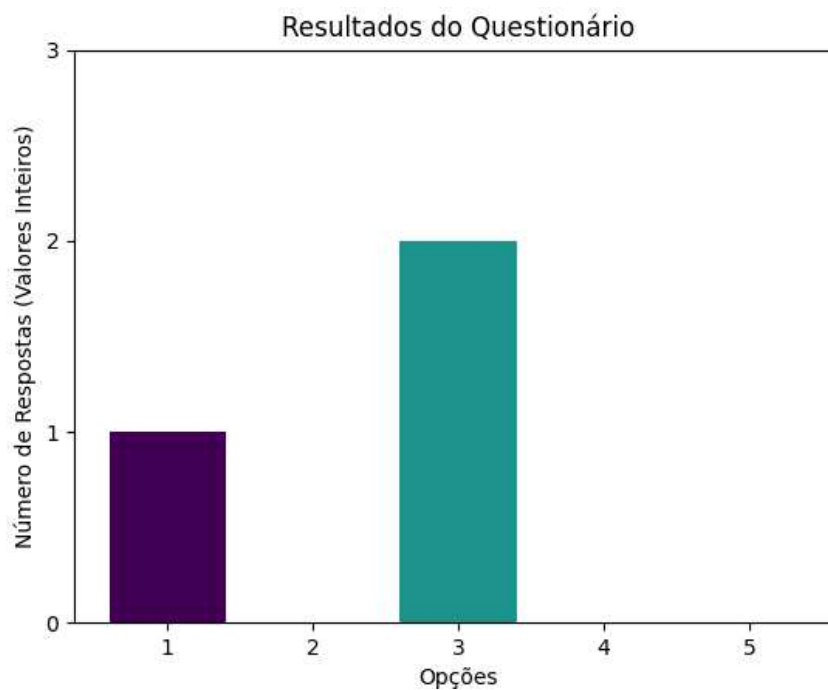


Figura 46 – Eu achei o sistema desnecessariamente complexo.

esse primeiro questionário, os participantes seguiram para a próxima etapa, que consistiu em um conjunto de tarefas, duas tarefas propostas, sendo que tais tarefas compreendiam todas as funcionalidades implementadas no sistema.

Por fim, os participantes foram orientados a preencherem um questionário Pós-Teste. Nesse momento, tal questionário continha um conjunto de 26 perguntas, sendo 23 de múltiplas escolhas e 3 perguntas abertas. Em relação às perguntas de múltiplas escolhas, 10 delas possuíam escala de 1 a 7, conforme definido pelo *System Usability Scale* (SUS)(JOHN, 1996) e 13 delas estavam com escala definida conforme o *System Accep-*

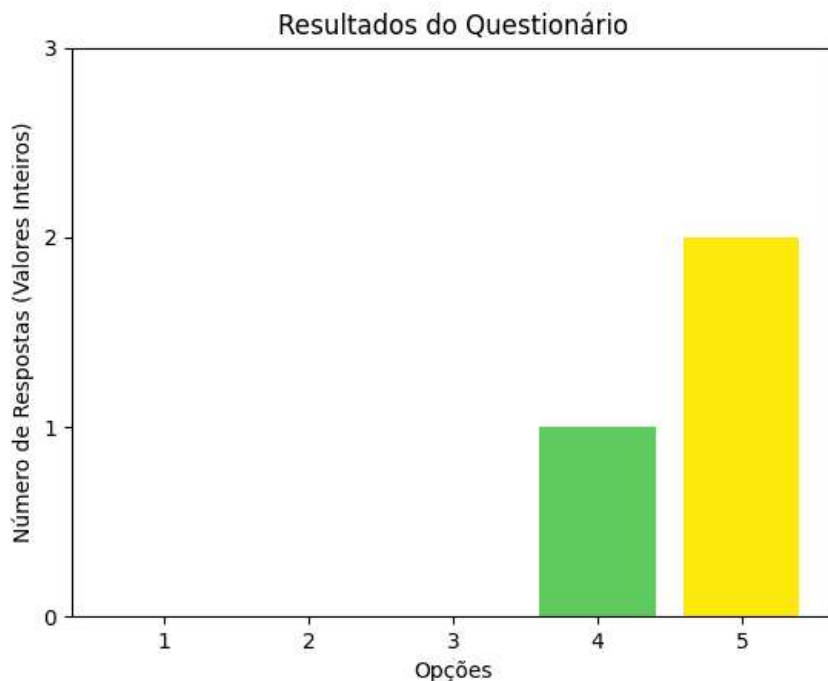


Figura 47 – Eu achei o sistema fácil de usar.

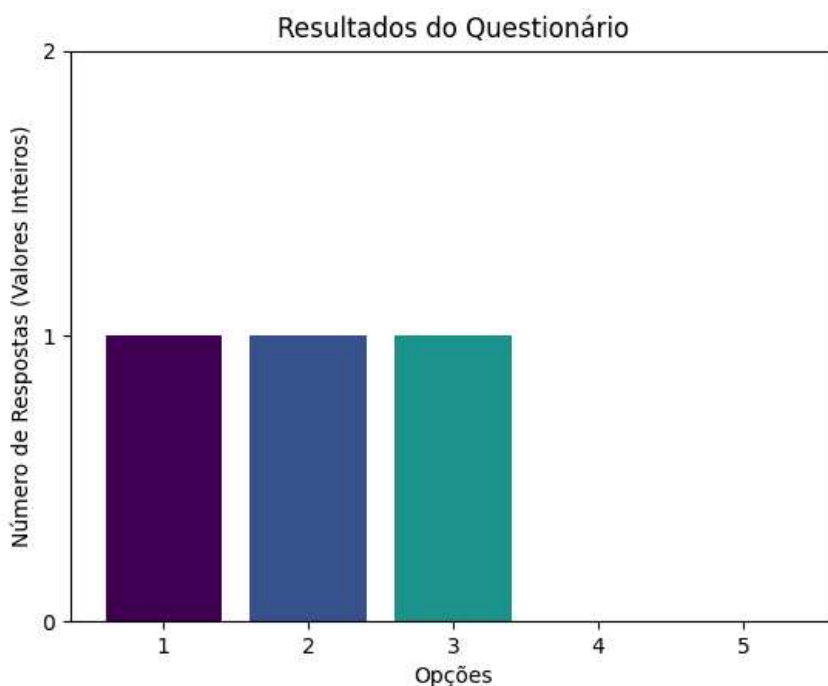


Figura 48 – Eu acho que precisaria de ajuda de uma pessoa com conhecimentos técnicos para usar o sistema.

*tance Scale* (SAS)([LAAN; HEINO; WAARD, 1997](#)). Por fim, as perguntas abertas eram voltadas para identificação de alguma propriedade ou valor presente em tela, buscando assim verificar a compreensão de tudo o que era exibido para o usuário.

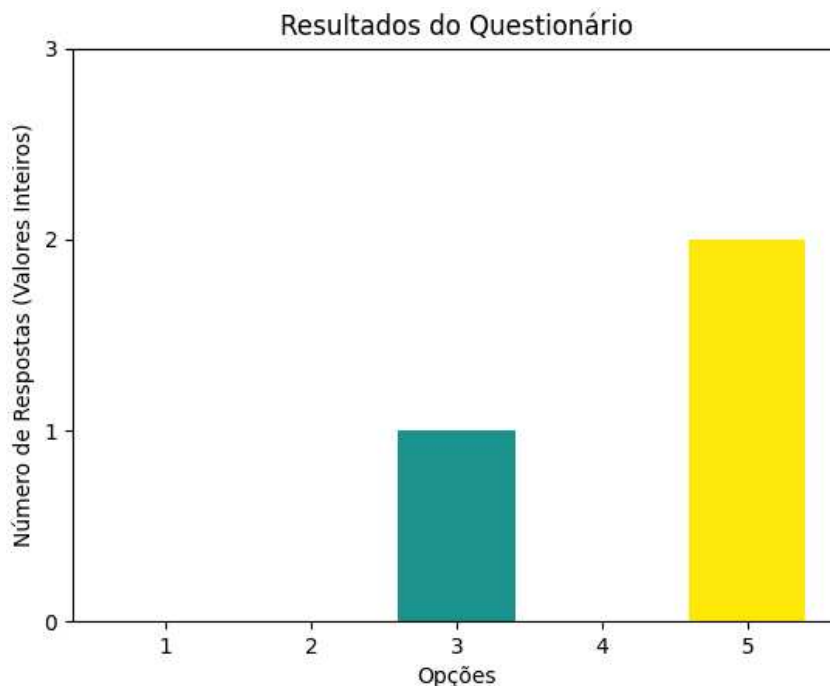


Figura 49 – Eu achei que as várias funções do sistema estão bem integradas.

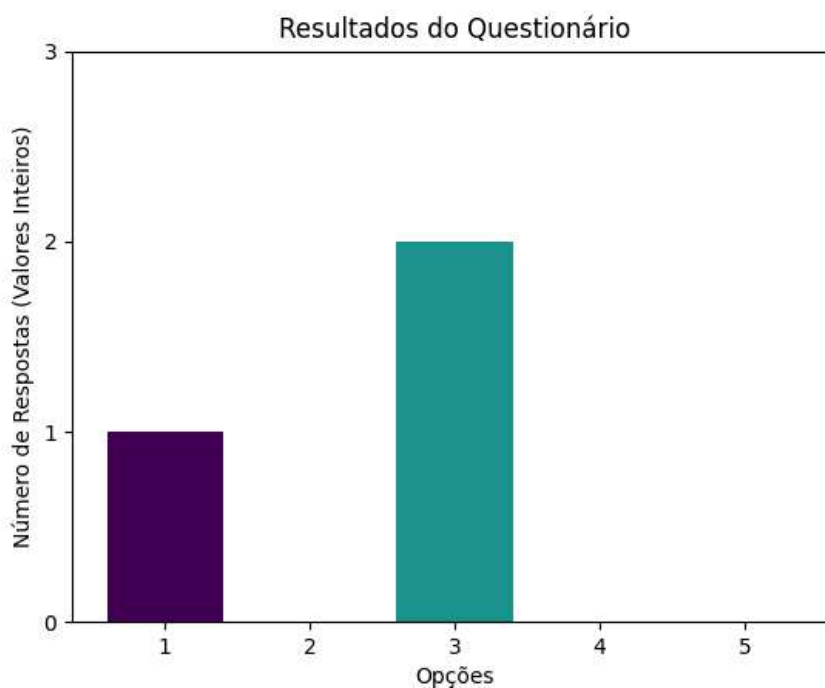


Figura 50 – Eu achei que existem inconsistências no sistema.

### 5.3.4 Resultados Obtidos

#### 5.3.4.1 Questionário Pré-Teste

A Tabela 11 ilustra o questionário pré-teste realizado:

#### 5.3.4.2 Questionário Pós-Teste

Os questionários SAS e SUS foram aplicados no pós-teste. Sendo assim, os seguintes resultados foram obtidos:

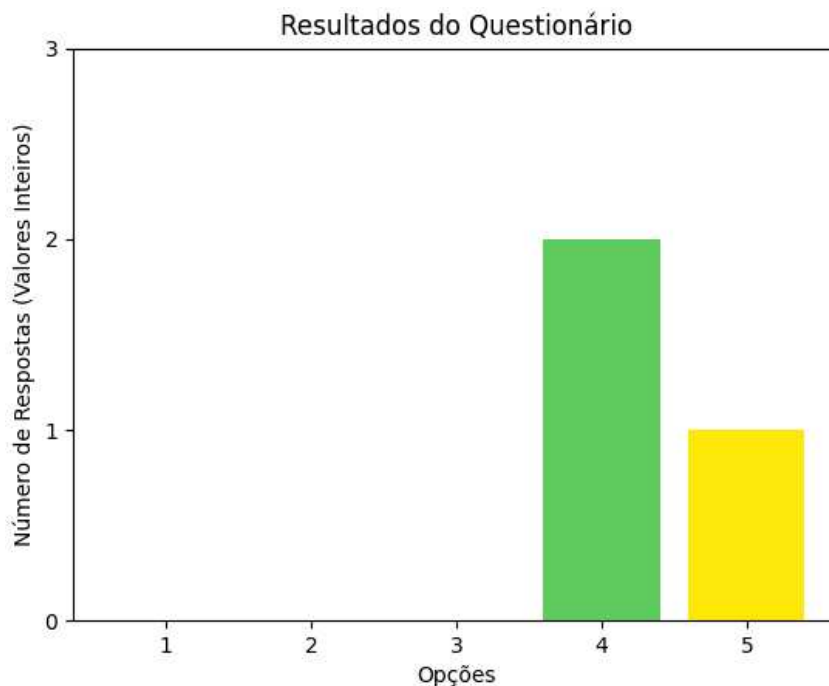


Figura 51 – Eu imagino que a maioria das pessoas aprenderiam a usar o sistema rapidamente.

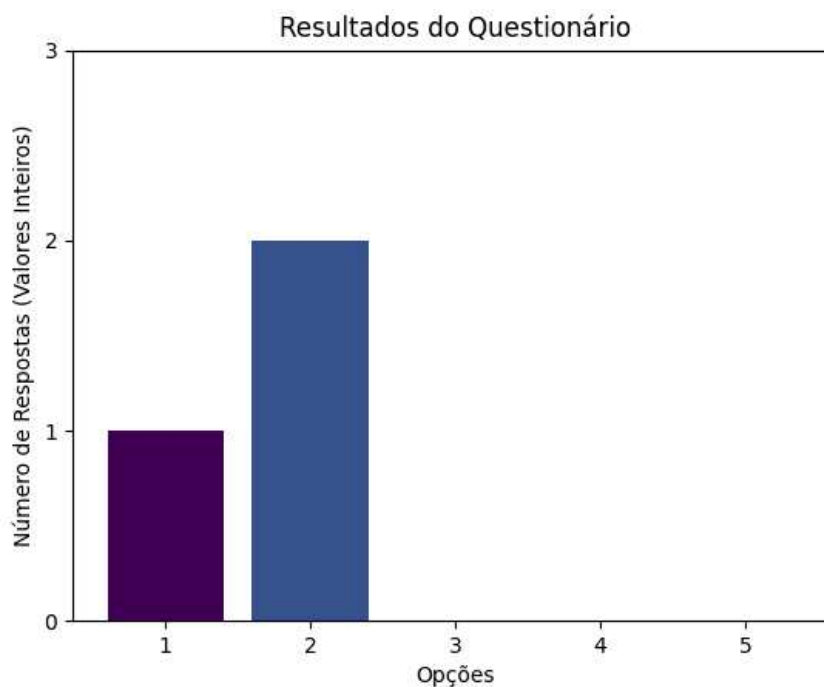


Figura 52 – Eu achei o sistema desajeitado de usar.

Tabela 11 – Exemplo do questionário pré-teste aplicado.

Pergunta	Opções	Resposta (%)
Você já utilizou alguma outra plataforma voltada para auxiliar o usuário no processo de alocação de recursos financeiros para a recuperação de cidades com base no agrupamento delas?	Sim	0
	Não	100
Caso sua resposta anterior seja "Sim", qual é a sua experiência com a utilização desse tipo de plataforma?		0

### 5.3.4.2.1 Questionário SAS

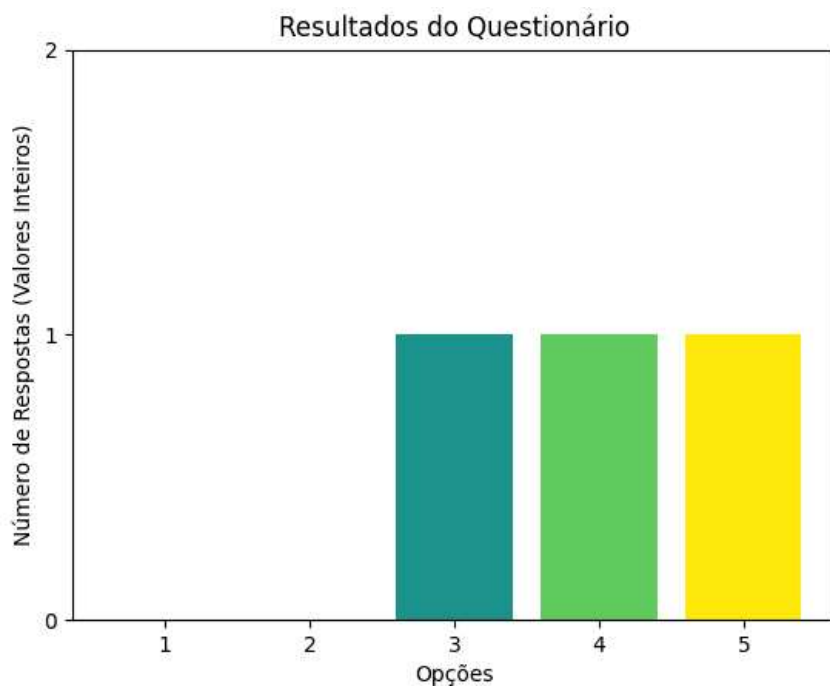


Figura 53 – Eu me senti bastante confiante ao usar o sistema.

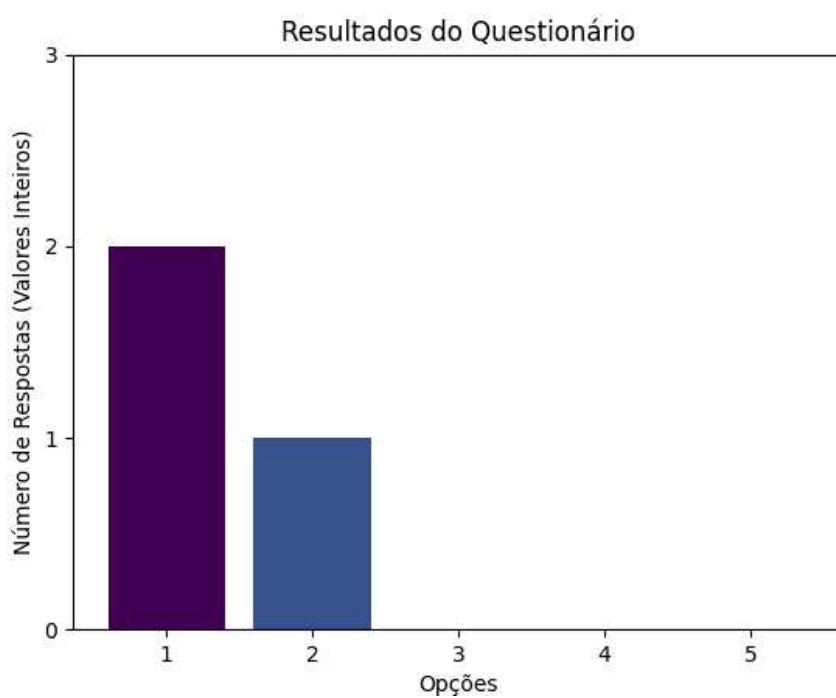


Figura 54 – Eu precisei aprender várias coisas novas antes de conseguir usar o sistema.

#### 5.3.4.2.2 Questionário SUS

Perguntas	Possíveis Respostas (%)						
	1	2	3	4	5	6	7
Minha intenção com a ferramenta TerrainIntelGIS foi clara e compreensível.						66,66	33,33
Interagir com a TerrainIntelGIS não exige muito do meu esforço mental.						66,66	33,33
Considero que a ferramenta TerrainIntelGIS fácil de usar.							100
Eu acho fácil conseguir que a ferramenta TerrainIntelGIS faça o que eu quero que ela faça.						66,66	33,33
Utilizar a ferramenta TerrainIntelGIS pode melhorar o meu desempenho na atividade de agrupar cidades da região de interesse.							100
Usar a ferramenta TerrainIntelGIS no seu trabalho pode melhorar a minha produtividade na atividade de agrupar cidades da região de interesse.						33,33	66,66
Usar a ferramenta TerrainIntelGIS pode aumentar a minha eficácia no meu trabalho.	33,33					66,66	
Eu considero a ferramenta TerrainIntelGIS será útil para agrupar cidades de uma mesma região de interesse.							100
Se eu tiver acesso à ferramenta TerrainIntelGIS, eu pretendo utilizá-la.				33,33		33,33	33,33
Levando em conta que poderei escolher usar ou não a ferramenta TerrainIntelGIS, eu prevejo que irei utilizá-la.				33,33	33,33	33,33	

Perguntas	Possíveis Respostas (%)				
	1	2	3	4	5
Eu penso que gostaria de usar esse sistema com frequência.			33,33	66,66	
Eu achei o sistema desnecessariamente complexo.	33,33		66,66		
Eu achei o sistema fácil de usar.				33,33	66,66
Eu acho que precisaria de ajuda de uma pessoa com conhecimentos técnicos para usar o sistema.	33,33	33,33	33,33		
Eu achei que as várias funções do sistema estão bem integradas.			33,33		66,66
Eu achei que existem inconsistência no sistema.	33,33		66,66		
Eu imagino que a maioria das pessoas aprenderiam usar o sistema rapidamente.				66,66	33,33
Eu achei o sistema desajeitado de usar.	33,33	66,66			
Eu me senti bastante confiante ao usar o sistema.			33,33	33,33	33,33
Eu precisarei aprender várias coisas novas antes de conseguir usar o sistema.	66,66	33,33			

## 6 Conclusão

O presente trabalho buscou propor e validar uma metodologia que auxiliasse no processo de tomada de decisão em relação à alocação dos recursos financeiros para a recuperação de áreas atingidas por desastres ambientais. Durante o seu desenvolvimento, percebeu-se que a ocorrência de tais desastres é algo que vêm se intensificando nos últimos anos, devido principalmente às mudanças climáticas e todo o desequilíbrio causado por ela, sendo assim importante o desenvolvimento de tal metodologia. Complementar a isso, existe ainda a complexidade de tais cenários, uma vez que as possibilidades de consequências são variadas, como econômicas, sociais e ambientais, além de que as áreas atingidas não são locais isolados e sim um conjunto interligado de ecossistemas e comunidades.

Para facilitar o contato dos possíveis usuários com a metodologia, foi elaborado e desenvolvido um *software* que facilitasse isso. Tal ferramenta consiste em um SIG denominado *TerrainIntelGIS*. Através dele, o usuário pode entrar em contato com ferramentas de MD sem conhecimento prévio, já que se buscou implementar algo que fosse simples de usar e que abstraísse ao máximo tais conceitos. O código-fonte do sistema pode ser acessado no seguinte link: <<https://github.com/TavaresCarlos/TerrainIntelGIS>>. O *link* para a versão inicial disponibilizada para testes está disponível em: <<http://www.webgisbrasil.com.br/cases-sucesso.html>> e um vídeo de demonstração se encontra em <<https://youtu.be/7seAYuxYfqk>>. Para validação do sistema e da metodologia, foi elaborado um caso de estudo em um cenário real, considerando para isso a Bacia Hidrográfica do Rio Paraopeba, região atingida pelo desastre de Brumadinho, em 2019.

Ao longo da implementação de tal ferramenta, foi realizado também a validação dela pelos potenciais usuários. Ao todo, foram realizados duas avaliações preliminares com um grupo de estudos do Departamento de Solos (DPS) da própria UFV. A primeira avaliação, consistiu em verificar a viabilidade de tal projeto, coletando sugestões e ideias. Já a última, foi a validação da ferramenta implementada, utilizando para isso metodologias como o SAS e o SUS. O número de respostas obtido foi relevante e essencial para entender a percepção do usuário em relação a ferramenta.

Em resumo, os resultados mostram que dados públicos são uma importante fonte de consulta e que podem ser utilizados para caracterizar de várias formas as áreas que sofreram algum tipo de desastre natural, seja ou em concomitância com dados privados de uma organização ou de forma única. Além disso, a metodologia elaborada se mostrou efetiva em alcançar os resultados pretendidos, permitindo caracterizar e agrupar cidades semelhantes de alguma forma.

### 6.1 Contribuições Científicas

Os resultados obtidos durante o desenvolvimento deste trabalho geraram as seguintes publicações científicas:

- Uma metodologia para apoio à tomada de decisão em cenários de desastres ambientais utilizando dados públicos, apresentado no *Brazilian Symposium on Geoinformatics* (GEOINFO), em 2022;

- *TerrainIntelGis: A Geographic Intelligence System to Support Decision-Making in Environmental and Natural Disaster Scenarios.*, apresentado no Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia), em 2023.

Além das publicações, dois resumos oriundos dos resultados obtidos durante o estudo foram selecionados e apresentados no Simpósio de Integração Acadêmica (SIA), edições de 2022 e 2023.

## 6.2 Potenciais Impactos Sociais

O *TerrainIntelGIS* é um SIG proposto com o uso exclusivo de tecnologias *open-source*. Dessa forma, busca-se evitar parte de gastos futuros envolvendo licenças ou planos. Além disso, buscou-se abstrair os conceitos de MD por meio do sistema, por entender que o usuário alvo pode não conhecer profundamente tais técnicas.

De forma geral, a aplicação aqui descrita é interessante por estar alinhada com os Objetivos de Desenvolvimento Sustentável (ODSs)<sup>39</sup> estabelecidos em 2015 pela Organização das Nações Unidas (ONU), em relação aos seguintes objetivos:

- Água Potável e Saneamento (6)
- Trabalho Decente e Crescimento Econômico (8)
- Indústria, Inovação e Infraestrutura (9)
- Redução das Desigualdades (10)
- Cidades e Comunidades Sustentáveis (11)
- Ação Contra a Mudança Global do Clima (13)
- Vida na Água (14)
- Vida Terrestre (15)
- Parcerias e Meios de Implementação (17)

Vale destacar ainda que, dos ODSs citados, o de número 13 (i.e., “Ação Contra a Mudança Global do Clima”)<sup>40</sup> merece destaque aqui, pois, entende-se que as mudanças climáticas que vêm ocorrendo levam ao desequilíbrio do meio ambiente e, conseqüentemente, ao aumento do número de desastres ambientais. Além disso, sabe-se que uma das causas das mudanças climáticas é a exploração excessiva dos recursos naturais, o que contribui ainda mais para a ocorrência desse tipo de desastre.

<sup>39</sup> <<https://brasil.un.org/pt-br/sdgs>>

<sup>40</sup> <<https://brasil.un.org/pt-br/sdgs/13>>

## 6.3 Trabalhos Futuros

É importante destacar que sistemas semelhantes são essenciais para a compreensão total da dimensionalidade do evento, característica essa possível através do uso de mapas interativos. Além disso, a tomada de decisão através do uso de dados e de técnicas de MD permite a caracterização real da região, possibilitando uma melhor tomada de decisão pelos responsáveis.

Como trabalhos futuros, planeja-se investigar alternativas que possam facilitar ainda mais a implementação do sistema por usuários finais interessados, utilizando-se a estratégia de containerização também no Módulo 01 do sistema, assim como permitir a leitura e MD em formatos distintos como imagens e arquivos geoespaciais, como *geoJSON* e *shapefiles*. Além disso, é interessante ainda explorar o uso de outras técnicas de Agrupamento, permitindo até mesmo que a própria ferramenta indique qual é a técnica mais adequada para um determinado conjunto de dados, e ainda explorar o uso de outras técnicas de MD neste mesmo cenário, como Predição, Associação e Séries Temporais. Em relação a escalabilidade do modelo, é importante a realização de investigações no que se refere ao tempo de execução e resposta para o usuário dado um grande volume de dados e atributos que possam vir a ser utilizados pelo mesmo. Pretende-se ainda evoluir as funcionalidades implementadas, permitindo assim a exportação de relatórios mais robustos, contando com informações referentes às técnicas de MD utilizadas. Por fim, a medida que novas funcionalidades vão sendo implementadas, é necessário também a realização de testes com usuários, buscando validar as novas funcionalidades.

# Referências

- ANNAS, S.; RAIS, Z. k-means and gis for mapping natural disaster prone areas in indonesia. In: EUROPEAN ALLIANCE FOR INNOVATION. *MSCEIS 2019: Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019, 12 October 2019, Bandung, West Java, Indonesia*. [S.l.], 2020. p. 340.
- BARBOSA, F. A. R.; MAIA-BARBOSA, P. M.; NASCIMENTO, A. M. A.; RIETZLER, A. C.; FRANCO, M. W.; PAES, T. A.; REIS, M.; MOURA, K. A. F.; DIAS, M. F.; ÁVILA, M. de P. et al. O desastre de mariana e suas consequências sociais, econômicas, políticas e ambientais: porque evoluir da abordagem de gestão dos recursos naturais para governança dos recursos naturais? *Arquivos do Museu de História Natural e Jardim Botânico da UFMG*, v. 24, n. 1-2, 2015.
- BRO, R.; SMILDE, A. K. Principal component analysis. *Analytical methods*, Royal Society of Chemistry, v. 6, n. 9, p. 2812–2831, 2014.
- BRUMATTI, C. H. T.; FIGUEIREDO, E. L.; DORIGUETO, L. F.; LISBOA-FILHO, J. Um framework para desenvolvimento de sistemas de gerenciamento de informações de desastres naturais. In: SBC. *Anais Estendidos do XVII Simpósio Brasileiro de Sistemas de Informação*. [S.l.], 2021. p. 13–16.
- BRUMATTI, C. H. T.; SCHAEFER, M. A. R.; FILHO, J. L.; REIS, J. C. Uma metodologia para apoio à tomada de decisão em cenários de desastres ambientais utilizando dados públicos. In: *GEOINFO*. [S.l.: s.n.], 2022. p. 329–334.
- CAMILO, C. O.; SILVA, J. C. d. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. *Universidade Federal de Goiás (UFG)*, v. 1, n. 1, p. 1–29, 2009.
- CARVALHO, H. M. Aprendizado de máquina voltado para mineração de dados : árvores de decisão. *Monografia (graduação) Universidade de Brasília, Faculdade UnB Gama, Curso de Engenharia de Software*, 2014.
- CHEN, Y.; OUYANG, X.; AGAM, G. Changenet: Learning to detect changes in satellite images. In: *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. [S.l.: s.n.], 2019. p. 24–31.
- DALLU, A. M. Risk heat map. KASNEB, 2020.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37–37, 1996.
- FILHO, J. L.; IOCHPE, C. Um estudo sobre modelos conceituais de dados para projeto de bancos de dados geográficos. *Revista IP-Informática Pública*, v. 1, n. 2, p. 37–90, 1999.
- GODOY, S. M.; DIAS, M. B. O desastre ambiental de mariana e o papel da fundação renova na reparação dos danos. *Direito e Desenvolvimento*, v. 12, n. 1, p. 37–48, 2021.
- HAN, J.; PEI, J.; TONG, H. *Data mining: concepts and techniques*. [S.l.]: Morgan kaufmann, 2022.

- HUANG, X.; LI, Z.; WANG, C.; NING, H. Identifying disaster related social media for rapid response: a visual-textual fused cnn architecture. *International Journal of Digital Earth*, Taylor & Francis, 2019.
- JOHN, B. Sus: a "quick and dirty" usability scale. *Usability evaluation in industry*, Taylor and Francis, p. 189–194, 1996.
- K-MEANS. 2019. <<https://www.inf.ufpr.br/dagoncalves/IA06.pdf>>. Acessado em: 02/12/2023.
- LAAN, J. D. V. D.; HEINO, A.; WAARD, D. D. A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, Elsevier, v. 5, n. 1, p. 1–10, 1997.
- LAI, J. Z.; LIAW, Y.-C. Improvement of the k-means clustering filtering algorithm. *Pattern Recognition*, Elsevier, v. 41, n. 12, p. 3677–3681, 2008.
- LAI, Y.; LIU, J. Optimization study on initial center of k-means algorithm. *Computer Engineering and Applications*, v. 44, n. 10, p. 147–149, 2008.
- LIRA, K. C.; OLIVEIRA, M. A. d.; MAGALHAES, R. P.; GONÇALVES, E. J. T. Utilizando mineração de dados e sistemas multiagentes na análise da evasão em educação a distância por meio do perfil dos alunos. In: *BRAZILIAN CONFERENCE ON INTELLIGENT SYSTEM (BRACIS)*. [S.l.: s.n.], 2016. v. 5, p. 1–12.
- MACQUEEN, J. et al. Some methods for classification and analysis of multivariate observations. In: OAKLAND, CA, USA. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. [S.l.], 1967. v. 1, n. 14, p. 281–297.
- MARRAKESH. *Natural Disasters Force 26 Million People into Poverty and Cost \$ 520bn in Losses Every Year, New World Bank Analysis Finds*. 2016. Disponível em: <https://www.worldbank.org/en/news/press-release/2016/11/14/natural-disasters-force-26-million-people-into-poverty-and-cost-520bn-in-losses-every-year-new-world-bank-analysis-finds>. Acessado: 22 jan. 2023.
- MEIRA, C. Processo especializado de descoberta de conhecimento em bases de dados para a modelagem de doenças de plantas-versão 1.0. Campinas: Embrapa Informática Agropecuária, 2012., 2012.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, v. 1, n. 1, p. 32, 2003.
- NIELSEN, J. *Engineering Usability*. [S.l.]: Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.
- NONATO, R. T. *Aplicação de mineração de dados na identificação de áreas cultivadas com cana-de-açúcar em imagens de sensoriamento remoto no Estado de São Paulo*. Tese (Doutorado) — [sn], 2010.
- OLIVEIRA, W. K. d.; ROHLFS, D. B.; GARCIA, L. P. *O desastre de Brumadinho e a atuação da Vigilância em Saúde*. [S.l.]: SciELO Brasil, 2019. e20190425 p.
- PEIXOTO, S. V.; ASMUS, C. I. R. F. O desastre de brumadinho e os possíveis impactos na saúde. *Ciência e Cultura*, Sociedade Brasileira para o Progresso da Ciência, v. 72, n. 2, p. 43–46, 2020.

- PENG, B.; LIU, X.; MENG, Z.; HUANG, Q. Urban flood mapping with residual patch similarity learning. In: *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. [S.l.: s.n.], 2019. p. 40–47.
- PHILLIPS, J. L.; COLVIN, M. E.; NEWSAM, S. Validating clustering of molecular dynamics simulations using polymer models. *BMC bioinformatics*, BioMed Central, v. 12, n. 1, p. 1–23, 2011.
- SANTANA, G. V.; SOUZA, A. N. de; GASTALDELLO, D. S.; JUNIOR, P. da C.; GIFALLI, A.; IKESHOJI, M. A. Análise de perfis de consumo residencial para implementação da tarifa branca. *Simpósio Brasileiro de Sistemas Elétricos-SBSE*, v. 1, n. 1, 2020.
- Serviço Geológico do Brasil. *Dashboard - Risco Geológico*. 2023. Disponível em: <https://geoportal.cprm.gov.br/portal/apps/opsdashboard/index.html>. Acessado: 22 jan. 2023.
- SOARES, A. L. C. *BACIA Hidrográfica DO RIO PARAOPEBA: ANÁLISE INTEGRADA DOS DIFERENTES IMPACTOS ANTRÓPICOS*. [S.l.]: Universidade Federal de Minas Gerais, Brasil, 2021.
- Superintendência de Comunicação Integrada do Ministério Público de Minas Gerais. *Rompimento da barragem de Fundão, em Mariana: resultados e desafios cinco anos após o desastre*. 2024. Acessado: 19 mar. 2024. Disponível em: <https://www.mpmg.mp.br/portal/menu/comunicacao/noticias/rompimento-da-barragem-d-e-fundao-em-mariana-resultados-e-desafios-cinco-anos-apos-o-desastre.shtml>.
- SUPRIYADI, B.; WINDARTO, A. P.; SOEMARTONO, T. et al. Classification of natural disaster prone areas in indonesia using k-means. *International Journal of Grid and Distributed Computing*, NADIA, v. 11, n. 8, p. 87–98, 2018.
- THANT, A. A.; AYE, S. M.; MANDALAY, M. Euclidean, manhattan and minkowski distance methods for clustering algorithms. *Int. J. Sci. Res. Sci. Eng. Technol*, v. 7, 2020.
- THORNDIKE, R. Who belongs in the family? *Psychometrika*, Springer, v. 18, n. 4, p. 267–276, 1953.
- VIRMANI, D.; TANEJA, S.; MALHOTRA, G. Normalization based k means clustering algorithm. *arXiv preprint arXiv:1503.00900*, 2015.
- WITTEN, I. H.; FRANK, E. Data mining: practical machine learning tools and techniques with java implementations. *Acm Sigmod Record*, ACM New York, NY, USA, v. 31, n. 1, p. 76–77, 2002.
- YADAV, J.; SHARMA, M. A review of k-mean algorithm. *Int. J. Eng. Trends Technol*, Citeseer, v. 4, n. 7, p. 2972–2976, 2013.
- ZHANG, H.; YU, H.; LI, Y.; HU, B. Improved k-means algorithm based on the clustering reliability analysis. In: ATLANTIS PRESS. *2015 International Symposium on Computers & Informatics*. [S.l.], 2015. p. 2516–2523.

---

ZHANG, X.; ZHANG, G.; LIU, P. Improved k-means algorithm based on clustering criterion function. *Jisuanji Gongcheng yu Yingyong(Computer Engineering and Applications)*, North China Computing Technology Institute,| a No. 26, P. O. Box 619| c , v. 47, n. 11, p. 123–127, 2011.

# Apêndices

# APÊNDICE A – Termo de Consentimento

Você está sendo convidado a participar, como voluntário(a), da pesquisa "Avaliação da Qualidade de Uso do *TerrainIntelGIS*", conduzida por Carlos Henrique Tavares Brumatti, estudante de mestrado em Ciência da Computação na UFV - Universidade Federal de Viçosa, sob a supervisão dos professores Júlio Reis e Jugurta Lisboa, e com auxílio da professora Maria Lucia Villela. A informação neste termo visa ajudá-lo a decidir se irá ou não participar da pesquisa.

O propósito deste estudo é desenvolver uma metodologia de apoio ao usuário especialista para o processo de tomada de decisão para a alocação de recursos financeiros destinados à recuperação de áreas afetadas por desastres ambientais. Além dessa metodologia, esse trabalho objetiva também desenvolver um Sistema de Inteligência Geográfica que reflita toda a metodologia adotada e facilite a interação do usuário com a mesma.

Esta pesquisa está sendo realizada com pesquisadores especialistas no contexto de solos, do Departamento de Solos da Universidade Federal de Viçosa (UFV).

O motivo que nos leva a desenvolver esta pesquisa é a busca por feedback dos possíveis usuários, visando identificar e realizar possíveis melhorias na usabilidade do sistema, corrigindo bugs identificados e otimizando ainda mais o código.

Caso você concorde em participar desta pesquisa, você terá acesso a um questionário, cujo tempo de resposta estimado é de aproximadamente 15 minutos. As questões vão coletar suas percepções relacionadas à facilidade de uso, utilidade e intenção de uso, bem como à usabilidade da plataforma *TerrainIntelGIS*.

O preenchimento deste questionário não representará qualquer risco de ordem física. Os riscos relacionados com sua participação podem envolver a possibilidade de constrangimento ou desconforto ao responder às questões do questionário. Como benefícios que poderão ser trazidos pela presente pesquisa, há a possibilidade de desenvolvimento do conhecimento e as respectivas influências tecnológicas na vida dos possíveis usuários do sistema.

Para participar deste estudo você não terá nenhum custo, nem receberá qualquer vantagem financeira. Apesar disso, diante de eventuais dados, identificados e comprovados, decorrentes da pesquisa, você tem assegurado o direito à indenização. Esta pesquisa é voluntária, e cabe a você decidir se irá ou não participar desse estudo. Se você escolher não participar deste estudo, não haverá consequências. Você poderá também interromper sua participação a qualquer momento, sem sofrer nenhum tipo de penalidade. Os resultados da pesquisa estarão à sua disposição quando finalizada. Você não será identificado(a) em nenhuma publicação que possa resultar. Seu nome ou material que indique sua participação não serão liberados sem a sua permissão.

Os dados e instrumentos desta pesquisa ficarão arquivados com o pesquisador responsável por um período de 5 (cinco) anos após o término da pesquisa, e depois desse tempo serão destruídos. Os pesquisadores tratarão a identidade do participante com padrões profissionais de sigilo e confidencialidade, atendendo à legislação brasileira, em especial, à resolução 466/2012 do conselho nacional de saúde, e utilizarão as informações somente para fins acadêmicos e científicos.

Você deverá acessar o seu e-mail institucional para responder esta pesquisa. Assim, conforme o que está sendo apresentado a você pelo sistema, e de acordo com suas experiências na plataforma, solicitamos que você dedique atenção às ações que devem ser realizadas e responda o mais sincero(a) possível as perguntas apresentadas.

Você pode contatar Carlos Henrique Tavares Brumatti pelo e-mail carlos.h.tavares@ufv.br, responsável por esta pesquisa, para falar sobre preocupações, tirar dúvidas ou dar sugestões.

As orientações gerais que regem essa avaliação podem ser consultadas no seguinte link: <[https://drive.google.com/file/d/1J2qfqt\\_m20NZXlxZQv3JCTTeAZOxAciu/view?usp=drive\\_link](https://drive.google.com/file/d/1J2qfqt_m20NZXlxZQv3JCTTeAZOxAciu/view?usp=drive_link)>

## APÊNDICE B – Tarefas solicitadas aos usuários ao longo do teste.

Nesta seção, você será orientado a executar algumas pequenas tarefas no sistema *TerrainIntelGIS*. A partir dessas tarefas, é que serão coletadas informações relacionadas à sua experiência de uso do sistema.

É importante ressaltar que suas opiniões, tanto em relação ao sistema quanto em relação ao teste do qual você está participando, são muito bem-vindas, e o que queremos avaliar aqui é a interface do sistema, e não o seu desempenho como usuário. Portanto, por favor, não tenha receio de nos dizer o que você realmente pensa.

### **Cadastro e Login no Sistema**

Para poder executar as tarefas solicitadas no teste, é necessário primeiramente que você faça o seu cadastro no sistema e forneça as informações solicitadas inicialmente, através das seguintes interações:

1. Para acessar o sistema, clique sobre TerrainIntelGIS (primeiro link no site de apoio) ou copie e cole no seu navegador o seguinte link: [<http://ip172-18-0-18-ckohrqcsnmng00f20cj0-3000.direct.labs.play-with-docker.com/>](http://ip172-18-0-18-ckohrqcsnmng00f20cj0-3000.direct.labs.play-with-docker.com/)
2. Clique no botão para realizar o cadastro no sistema.
3. Na nova tela, informe os dados solicitados.
4. Após informar os dados, finalize o cadastro.
5. Na tela principal, informe o login e senha, conforme solicitados.
6. Após informar os dados, clique para entrar

### ***TerrainIntelGIS***

A seguir, é descrito um conjunto de tarefas necessárias para a realização desse Teste de Usuário. Recomenda-se que a leitura das tarefas e a execução das mesmas ocorra de forma paralela, alternando-se entre as janelas.

### **TAREFA 01: AGRUPAMENTO DOS DADOS, INFORMANDO O NÚMERO DE GRUPOS**

Uma vez dentro da aplicação, agora é necessário informar alguns parâmetros obrigatórios. Para esse exemplo, é necessário informar o número de agrupamentos desejados. Assim:

1. Assim, você deve escolher  $K = 3$  como número de agrupamentos e faça o upload do arquivo csv de teste, disponível no site (link: [<http://www.webgisbrasil.com.br/cases-sucesso.html>](http://www.webgisbrasil.com.br/cases-sucesso.html))
2. Após o sistema processar a consulta, escolha os seguintes atributos: APP, Imóveis, Hidrografia, Vegetação Nativa, População Estimada (2021) e Área Territorial (2021).

3. Observe também o mapa principal e a relação das cores com cada um dos agrupamentos.
4. Ainda no mapa, selecione a opção "Ortofoto" e marque a opção "Mapa BHRP" e observe.
5. Por fim, avalie mentalmente o mapa de calor gerado.
6. Faça o download do relatório.
7. Saia do sistema.

### **TAREFA 02: AGRUPAMENTO DOS DADOS, NÃO INFORMANDO O NÚMERO DE GRUPOS**

Para esse exemplo agora, não será informado o número de agrupamentos desejados. Assim:

1. Agora, defina como não o desejo de informar o número de agrupamentos e faça o upload do arquivo csv de teste, disponível no site (link: <http://www.webgisbrasil.com.br/cases-sucesso.html>)
2. Após o sistema processar a consulta, escolha os seguintes atributos: APP, Imóveis, Hidrografia, Vegetação Nativa, População Estimada (2021) e Área Territorial (2021).
3. Observe também o mapa principal e a relação das cores com cada um dos agrupamentos.
4. Ainda no mapa, selecione a opção "Ortofoto" e marque a opção "Mapa BHRP" e observe.
5. Por fim, avalie mentalmente o mapa de calor gerado.
6. Faça o download do relatório.
7. Saia do sistema.

## APÊNDICE C – Avaliação *System Acceptance Scale (SAS)*

Para cada uma das afirmativas abaixo, selecione a coluna referente ao número que melhor expressa o seu grau de concordância, conforme a escala a seguir: (1) Discordo Totalmente, (2) Discordo Fortemente, (3) Discordo Parcialmente, (4) Nem Concordo nem Discordo, (5) Concordo Parcialmente, (6) Concordo Fortemente, (7) Concordo Totalmente.

### **Facilidade de Uso Percebida:**

1. Minha interação com a ferramenta *TerrainIntelGIS* foi clara e compreensível
2. Interagir com a ferramenta *TerrainIntelGIS* não exige muito do meu esforço mental
3. Considero a ferramenta *TerrainIntelGIS* fácil de usar
4. Eu acho fácil conseguir que a ferramenta *TerrainIntelGIS* faça o que eu quero que ela faça

### **Utilidade Percebida:**

1. Utilizar a ferramenta *TerrainIntelGIS* pode melhorar o meu desempenho na atividade de agrupar cidades da região de interesse.
2. Usar a ferramenta *TerrainIntelGIS* no meu trabalho pode melhorar a minha produtividade na atividade de agrupar cidades da região de interesse.
3. Usar a ferramenta *TerrainIntelGIS* pode aumentar a minha eficácia no meu trabalho.
4. Eu considero a ferramenta *TerrainIntelGIS* será útil para agrupar cidades de uma mesma região de interesse.

### **Intenção de Uso:**

1. Se eu tiver acesso à ferramenta *TerrainIntelGIS*, eu pretendo utilizá-la.
2. Levando em conta que poderei escolher usar ou não a ferramenta *TerrainIntelGIS*, eu prevejo que irei utilizá-la.

## APÊNDICE D – Avaliação *System Usability Scale* (SUS)

Por favor, a seguir, marque a opção que reflete a sua resposta imediata a cada afirmação. Não pense muito tempo sobre cada afirmação. Se você não souber como responder, marque a opção 3.

1. Eu penso que gostaria de usar esse sistema com frequência.
2. Eu achei o sistema desnecessariamente complexo.
3. Eu achei o sistema fácil de usar.
4. Eu acho que precisaria de ajuda de uma pessoa com conhecimentos técnicos para usar o sistema.
5. Eu achei que as várias funções do sistema estão bem integradas.
6. Eu achei que existem inconsistências no sistema.
7. Eu imagino que a maioria das pessoas aprenderiam usar o sistema rapidamente.
8. Eu achei o sistema desajeitado de usar.
9. Eu me senti bastante confiante ao usar o sistema.
10. Eu precisei aprender várias coisas novas antes de conseguir usar o sistema.