#### UNIVERSIDADE FEDERAL DE VIÇOSA

#### CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS

#### DEPARTAMENTO DE INFORMÁTICA

CARLOS HENRIQUE TAVARES BRUMATTI

##### MINERAÇÃO DE DADOS GEOESPACIAIS PÚBLICOS: ANÁLISE DO CADASTRO AMBIENTAL RURAL COMO FERRAMENTA DE APOIO À TOMADA DE DECISÃO EM PROBLEMAS AMBIENTAIS LIGADOS À BACIA HIDROGRÁFICA DO RIO PARAOPEBA

VIÇOSA

2021

CARLOS HENRIQUE TAVARES BRUMATTI

##### MINERAÇÃO DE DADOS GEOESPACIAIS PÚBLICOS: ANÁLISE DO CADASTRO AMBIENTAL RURAL COMO FERRAMENTA DE APOIO À TOMADA DE DECISÃO EM PROBLEMAS AMBIENTAIS LIGADOS À BACIA HIDROGRÁFICA DO RIO PARAOPEBA

Projeto de Pesquisa para o curso de Mestrado *stricto sensu* em Ciência da Computação da Universidade Federal de Viçosa

Orientador: Julio Cesar Soares dos Reis

Co-orientador: Jugurta Lisboa Filho

viÇOSA

2021

Sumário

1. **INTRODUÇÃO**

O desastre ambiental 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. Segundo Godoy e Dias (2021), o impacto causado pelo desastre ambiental provocou sérias consequências não somente na região de ocorrência, mas em todo o percurso dos rejeitos até a sua chegada ao mar. Estas consequências não se restringem somente às questões ambientais, mas sim socioeconômicas, políticas e humanas. 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 há uma proposta da criação de um fundo de US$20 bilhões, ao longo de 10 anos, pelas empresas envolvidas.

Uma das áreas atingidas consequentemente com o desastre foi a Bacia Hidrográfica do Rio Paraopeba (BHRP) – Minas Gerais, Brasil. Abrangendo uma área de 12.054,25 km² e contando com um total de 1.318.885 milhão de habitantes (IBGE/CENSO 2010) distribuídos ao longo de 48 municípios, a BHRP é um dos principais afluentes da Bacia Hidrográfica Federal do Rio São Francisco, fazendo parte da chamada Circunscrição Hidrográfica SF3, conforme a Deliberação Normativa CERH-MG nº 66, de 17 de novembro de 2020. As principais atividades econômicas desses municípios giram em torno desse recurso hídrico (IGAM, 2013), destacando-se as atividades extrativistas, industriais e agropecuárias.

Atualmente, um conjunto considerável de dados ambientais é produzido através do Cadastro Ambiental Rural (CAR). Os dados referentes às propriedades são informados de maneira eletrônica por todos aqueles que possuem uma propriedade rural ou que são responsáveis legais pelas mesmas. Sendo assim, estes dados possuem a capacidade de descrever quantitativamente e qualitativamente uma dada região rural. Utilizando técnicas adequadas, é possível aplicar o Processo de Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Database – KDD*) a fim de extrair informações, verificar hipóteses e descobrir novos padrões nos dados, como proposto por Fayyad, Piatetsky-Shapiro, Smyth (1996).

Miller e Han (2009) definem o processo de KDD como sendo constituído por quatro etapas não lineares: Seleção, Pré-Processamento, Redução e Projeção, e Interpretação e Comunicação dos Dados. A primeira etapa é responsável por determinar os conjuntos de dados que serão utilizados e definir quais as varáveis de interesse para verificação. A segunda etapa é onde ocorre a “limpeza” dos dados. Nela, os ruídos e os dados duplicados são eliminados e os dados faltantes são determinados, além de serem definidas também outras bases de dados para obter dados complementares de interesse. Na terceira etapa, o volume total de dados é reduzido, buscando assim trabalhar somente com um conjunto representativo dos dados gerais, o que facilita a manipulação posterior deles pelos algoritmos de mineração de dados em si. Na última etapa, é obtido o resultado das manipulações e esses valores são interpretados em relação ao contexto geral do problema.

Hand (1999), por outro lado, definiu a mineração de dados como uma etapa independente no processo de KDD, consistindo em um conjunto de ideias, ferramentas e métodos de várias outras áreas, como a Estatística e o Aprendizado de Máquina, por exemplo, para a aplicação em um conjunto de dados. Dessa forma, a mineração de dados é um conceito multidisciplinar, voltado para o entendimento do problema e o uso das melhores técnicas disponíveis.

Com base nos conceitos apresentados até agora, esse trabalho busca aplicar as técnicas de mineração de dados sobre os dados geoespaciais disponíveis no CAR em relação à BHRP, buscando assim verificar como ocorre a distribuição das principais coberturas vegetais e hídricas ao longo da bacia e a relação existente entre elas e os aspectos socioeconômicas da região, evidenciando assim a importância da bacia para as cidades locais.

1. **REFERENCIAL TEÓRICO**
   1. **CADASTRO AMBIENTAL RURAL (CAR)**
   2. **SISTEMA NACIONAL DE CADASTRO AMBIENTAL RURAL (SICAR)**
   3. **MINERAÇÃO DE DADOS**
   4. **BACIA HIDROGRÁFICA DO RIO PARAOPEBA (MG)**
   5. **DADOS GEOESPACIAIS**
2. **Hipóteses de trabalho + avaliação dos resultados**
3. **METODOLOGIA**
4. **CRONOGRAMA**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Atividade** | **Jan** | **Fev** | **Mar** | **Abr** | **Mai** | **Jun** | **Jul** | **Ago** | **Set** | **Out** | **Nov** | **Dez** |
| Revisão de literatura | x | x | x | x | x | x | x | x | x | x |  |  |
| Seleção | x | x | x |  |  |  |  |  |  |  |  |  |
| Pré-processamento |  | x | x | x |  |  |  |  |  |  |  |  |
| Transformação |  |  |  |  | x |  |  |  |  |  |  |  |
| Mineração de Dados |  |  |  |  | x | x | x | x |  |  |  |  |
| Interpretação |  |  |  |  |  |  |  | x | x |  |  |  |
| Interface de Visualização |  |  |  |  |  |  | x | x |  |  |  |  |
| Escrita de Artigos e Submissões |  |  | x | x | x | x | x | x | x | x | x |  |
| Defesa de Dissertação |  |  |  |  |  |  |  |  |  |  |  | x |

(Tabela 0x – Cronograma do Projeto)

1. **ORÇAMENTO**

|  |  |  |
| --- | --- | --- |
| **Especificação das Despesas** | **Valor** | **Fonte de Recursos** |
| PESSOAL |  |  |
| Bolsa de Estudos | 18 x 1.500,00 = 27.000,00 | Fonte Própria |
| PUBLICAÇÃO |  |  |
| Traduções | 1.200,00 | Fonte Própria |
| Taxas de Publicações | 1.000,00 | Fonte Própria |
| Viagens para congressos | 1.200,00 x 3 = 3.600,00 | Fonte Própria |
| MATERIAIS |  |  |
| Material de Escritório | 100,00 | Fonte Própria |
| TOTAL | 32.900,00 | Fonte Própria |

(Tabela 0x: Orçamento previsto)

1. **REFERÊNCIAS**

CHRISTEN, Martin. **Geospatial Analysis for the Smart City**. In: BIG DATA BBQ, 2021, Mannheim. Disponível em: < https://github.com/martinchristen/bigdatabbq2021 >. Acesso em: 23 out. 2021.

WESTRA, Erik. **Python geospatial development**. Birmingham: Packt Publishing, 2010.

LISBOA F., J. Porto Alegre: CPGCC da UFRGS, 1995. 1995. 69p, (TI-491)

COUTO, Rodrigo de Souza. Estruturação, modelagem e implementação de banco de dados geoespaciais para o cadastro ambiental rural. 2017.

ARAUJO, Marcos Gabriel Pereira. Aplicação de técnicas de mineração de dados para análise climática no estado do Tocantins. 2021.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37-37, 1996.

PIB do setor agropecuário registrou alta de 5,7% no primeiro trimester. Disponível em:< https://www.gov.br/agricultura/pt-br/assuntos/noticias/pib-do-setor-agropecuario-registrou-alta-de-5-7-no-primeiro-trimestre>. Acesso em: 23 out. 2021

BORGES, Sérgio. O desastre da barragem de rejeitos em Mariana, Minas Gerais: aspectos socioambientais e de gestão na exploração de recursos minerais. **Cuadernos de Geografía: Revista Colombiana de Geografía**, v. 27, n. 2, p. 301-312, 2018.

ESPINDOLA, Haruf Salmen; NODARI, Eunice Sueli; SANTOS, Mauro Augusto dos. Rio Doce: riscos e incertezas a partir do desastre de Mariana (MG). **Revista Brasileira de História**, v. 39, p. 141-162, 2019

**CAR**

EMBRAPA TERRITORIAL. **Agricultura e preservação ambiental**: uma análise do cadastro ambiental rural. Campinas, 2020. Disponível em: < www.embrapa.br/car >. Acesso em:  28 out. 2021

**What is geospatial data ?** Disponível em: https://www.ibm.com/topics/geospatial-data. Acesso em: 05 novem. 2021.

BARBOSA, Francisco Antônio Rodrigues 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

GODOY, Sandro Marcos; DIAS, Mateus Benites. 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

**Bacia do Rio Paraopeba.**Disponível em: http://www.feam.br/component/content/article/901-bacia-do-rio-paraopeba. Acesso em: 10 novem. 2021.

**QUALIDADE DAS ÁGUAS SUPERFICIAIS DE MINAS GERAIS EM 2013.** Disponível em: http://www.igam.mg.gov.br/images/stories/qualidade\_aguas/2014/resumo-executivo-2013.pdf. Acesso em: 11 novem. 2021.

GOODCHILD, Michael F. Citizens as sensors: the world of volunteered geography. **GeoJournal**, v. 69, n. 4, p. 211-221, 2007

**DELIBERAÇÃO NORMATIVA CERH-MG Nº 66, DE 17 DE NOVEMBRO DE 2020**. Disponível em: http://www.siam.mg.gov.br/sla/download.pdf?idNorma=52900. Acesso em: 13 novem. 2021

HAND, David J. Statistics and data mining: intersecting disciplines. **Acm Sigkdd Explorations Newsletter**, v. 1, n. 1, p. 16-19, 1999.

MILLER, Harvey J.; HAN, Jiawei. **Geographic data mining and knowledge discovery**. CRC press, 2009.