



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DE RUSSAS
CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

EVANDRO RODRIGUES DE PAULA JUNIOR

RESTAURAÇÃO ECOLÓGICA INTELIGENTE ATRAVÉS DA ANÁLISE DE DADOS

RUSSAS

2024

EVANDRO RODRIGUES DE PAULA JUNIOR

RESTAURAÇÃO ECOLÓGICA INTELIGENTE ATRAVÉS DA ANÁLISE DE DADOS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação do Campus de Russas da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Ciência da Computação.

Orientadora: Profa. Dra. Tatiane Fernandes Figueiredo

RUSSAS

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

P346r Paula Junior, Evandro Rodrigues de.
Restauração Ecológica Inteligente Através da Análise de Dados / Evandro Rodrigues de Paula Junior. –
2024.
53 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Russas,
Curso de Ciência da Computação, Russas, 2024.

Orientação: Profa. Dra. Tatiane Fernandes Figueiredo.
Coorientação: Prof. Dr. Tiago Shizen Pacheco Toma..

1. Restauração ecológica. 2. Análise de dados.. 3. Computação aplicada. I. Título.

CDD 005

EVANDRO RODRIGUES DE PAULA JUNIOR

RESTAURAÇÃO ECOLÓGICA INTELIGENTE ATRAVÉS DA ANÁLISE DE DADOS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação do Campus de Russas da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Ciência da Computação.

Aprovada em: 04/10/2024

BANCA EXAMINADORA

Profa. Dra. Tatiane Fernandes
Figueiredo (Orientadora)
Universidade Federal do Ceará - UFC

Prof. Dr. Pablo Luiz Braga Soares
Universidade Federal do Ceará - UFC

Prof. Dr. Tiago Shizen Pacheco Toma
Universidade Federal de Minas Gerais - UFMG

AGRADECIMENTOS

O desenvolvimento deste trabalho de conclusão de curso contou com a ajuda de diversas pessoas, dentre as quais agradeço:

Aos meus pais, Maria e Evandro, meu eterno agradecimento pelo amor incondicional e pelo incentivo constante. Sem o suporte de vocês, esta conquista não seria possível.

À minha orientadora, Tatiane Fernandes, cujo apoio e orientação foram indispensáveis para a realização deste trabalho. Sua dedicação e expertise foram fundamentais para meu crescimento acadêmico e profissional.

Aos meus queridos amigos do Covil do Café — Jordão, Victor, Aurilene, Leonardo e Alan — agradeço pela amizade, pelas conversas inspiradoras e pelo companheirismo ao longo desta jornada. Vocês tornaram este caminho mais leve e significativo.

A realização desta pesquisa em Ciência da Computação foi essencial para meu desenvolvimento pessoal, acadêmico e profissional. Cada desafio superado contribuiu para meu aprimoramento e preparação para os futuros passos na minha carreira.

RESUMO

Este trabalho apresenta uma abordagem interdisciplinar entre computação e biologia, focada na Restauração Ecológica. A partir de uma base de dados composta por informações de espécies plantas, habitats, tipos de solo, cidades e localidades, foi realizada uma análise exploratória detalhada que permitiu extrair, como resultado, *insights* significativos e identificar padrões não evidentes em análises isoladas. Um dos principais resultados desta pesquisa foi o desenvolvimento de uma plataforma *online* que facilita o processo de restauração ecológica. Utilizando tecnologias modernas como React no *front-end* e Flask no *back-end*, a plataforma permite que usuários simulem listas de espécies para serem utilizadas na restauração ao fornecerem informações sobre o tipo de solo, habitat e localização geográfica. O algoritmo de simulação gera uma lista ranqueada de espécies adequadas para o ecossistema em questão, auxiliando na tomada de decisão para a restauração de áreas específicas. Os resultados obtidos contribuem para a área da restauração ecológica, fornecendo uma ferramenta prática que auxilia na seleção de espécies e promove a interação com dados relevantes para o sucesso de projetos de restauração.

Palavras-chave: restauração ecológica; análise de dados; plataforma *online*; seleção de espécies; computação aplicada.

ABSTRACT

This work presents an interdisciplinary approach between computing and biology, focused on Ecological Restoration. Based on a database composed of information on plant species, habitats, soil types, cities, and localities, a detailed exploratory analysis was conducted, which allowed extracting, as a result, significant *insights* and identifying patterns not evident in isolated analyses. One of the main results of this research was the development of an *online* platform that facilitates the process of ecological restoration. Utilizing modern technologies like React on the *front-end* and Flask on the *back-end*, the platform allows users to simulate lists of species to be used in restoration by providing information about soil type, habitat, and geographical location. The simulation algorithm generates a ranked list of suitable species for the ecosystem in question, aiding decision-making for the restoration of specific areas. The results obtained contribute to the field of ecological restoration by providing a practical tool that assists in species selection and promotes interaction with relevant data for the success of restoration projects.

Keywords: ecological restoration; data analysis; online platform; species selection; applied computing.

LISTA DE FIGURAS

Figura 1 – Exemplo de ecossistema de referência e três estágios de degradação. Fotos são das terras altas do sul do Brasil, Rio Grande do Sul (Fotos: J.-M. Hermann, Jan. 2012). Adaptado de Overbeck <i>et al.</i> (2013)	16
Figura 2 – Fluxograma da Análise de Dados (Simplificado). Fonte: elaborado pelo autor.	18
Figura 3 – Exemplo de matriz de correlação. Fonte: Miller (2019)	21
Figura 4 – Importância dos Ecossistemas de Referência. Adaptado de Toma <i>et al.</i> (2023)	23
Figura 5 – Fluxograma da Análise de Dados (Detalhado). Fonte: elaborado pelo autor.	27
Figura 6 – Exemplo dos dados na planilha de Lists. Fonte: elaborado pelo autor.	29
Figura 7 – Exemplo dos dados na planilha de AfterFlora. Fonte: elaborado pelo autor. .	30
Figura 8 – Resultado da junção dos <i>datasets</i> de Lists e Afterflora criados na seção 5.3.3. Fonte: elaborado pelo autor.	37
Figura 9 – Dados da cidade de Mariana, criados ao filtrar o <i>dataset</i> unificado. Fonte: elaborado pelo autor.	37
Figura 10 – Maiores IVs do solo do tipo afloramento ferruginoso. Fonte: elaborado pelo autor.	38
Figura 11 – Informações extras sobre afloramento ferruginoso. Fonte: elaborado pelo autor.	39
Figura 12 – Maiores IVs do solo do tipo afloramento quartzítico. Fonte: elaborado pelo autor.	40
Figura 13 – Informações extras sobre afloramento quartzítico. Fonte: elaborado pelo autor.	40
Figura 14 – Distribuição de habitats existentes na base de dados. Fonte: elaborado pelo autor.	41
Figura 15 – Exemplo de planilha gerado para <i>Forest Islands</i> . Fonte: elaborado pelo autor.	42
Figura 16 – Exemplo de planilha gerado para <i>Quartz Grasslands</i> . Fonte: elaborado pelo autor.	43
Figura 17 – Listas em Minas Gerais. Fonte: elaborado pelo autor.	43
Figura 18 – Listas na Bahia. Fonte: elaborado pelo autor.	43
Figura 19 – <i>Dataset</i> gerado ao final da primeira etapa do algoritmo. Fonte: elaborado pelo autor.	44
Figura 20 – <i>Dataset</i> gerado ao final da segunda etapa do algoritmo. Fonte: elaborado pelo autor.	45
Figura 21 – Tela de boas vindas da plataforma. Fonte: elaborado pelo autor.	45

Figura 22 – Tela de resumo sem filtros aplicados. Fonte: elaborado pelo autor.	47
Figura 23 – Tela de resumo com filtros aplicados. Fonte: elaborado pelo autor.	47
Figura 24 – Mapa contendo as espécies. Fonte: elaborado pelo autor.	48
Figura 25 – <i>Pop-up</i> contendo informações sobre a lista selecionada. Fonte: elaborado pelo autor.	48
Figura 26 – <i>Modal</i> contendo a listagem das espécies atreladas àquela lista. Fonte: elaborado pelo autor.	48
Figura 27 – Simulação de restauração. Fonte: elaborado pelo autor.	49

LISTA DE TABELAS

Tabela 1 – Análise quantitativa dos dados na base de dados unificada	37
Tabela 2 – Informações sobre afloramento ferruginoso na base de dados	38
Tabela 3 – Informações sobre afloramento quartzítico na base de dados	39

SUMÁRIO

1	INTRODUÇÃO	12
2	OBJETIVOS	14
2.1	Objetivo geral	14
2.2	Objetivo específicos	14
3	FUNDAMENTAÇÃO TEÓRICA	15
3.1	Restauração Ecológica	15
3.1.1	<i>Ecosistemas de Referência</i>	16
3.2	Análise de Dados	17
3.2.1	<i>Coleta de Dados</i>	18
3.2.2	<i>Pré-Processamento de Dados</i>	19
3.2.3	<i>Visualização de Dados</i>	20
4	TRABALHOS RELACIONADOS	22
4.1	<i>Utilização de ecossistemas de referência ótimos para restauração ecológica</i>	22
4.2	<i>Sistema de Recomendação de Plantio na Agricultura</i>	23
4.3	<i>Utilização de IoT e Aprendizado de Máquina para seleção de plantas na agricultura</i>	24
4.4	<i>Aplicando técnica de Ensembles para recomendação de plantio na Índia</i>	25
5	PROCEDIMENTOS METODOLÓGICOS	27
5.1	Entendimento do problema	27
5.2	Entendimento da base de dados	28
5.2.1	<i>Lists</i>	28
5.2.2	<i>AfterFlora</i>	30
5.3	Pré-processamento dos dados	31
5.3.1	<i>Seleção dos atributos mais importantes</i>	31
5.3.2	<i>Limpeza dos dados</i>	31
5.3.3	<i>Construção de novos atributos</i>	32
5.4	Análise exploratória de dados	33
5.5	Construção da simulação	33
5.5.1	<i>Primeira etapa</i>	33
5.5.2	<i>Segunda etapa</i>	34

5.6	Plataforma Online	35
6	RESULTADOS	36
6.1	Análise exploratória dos dados	36
6.1.1	<i>Resultados do pré-processamento dos dados</i>	36
6.1.2	<i>Análise quantitativa de atributos da base de dados</i>	37
6.1.3	<i>Análise quando solo é afloramento ferruginoso (ironstone)</i>	38
6.1.4	<i>Análise quando solo é afloramento quartzítico (quartzite)</i>	39
6.1.5	<i>Análise dos habitats</i>	41
6.1.6	<i>Análise das espécies x habitats</i>	42
6.1.7	<i>Análise das Cidades e Localidades</i>	43
6.1.8	<i>Análise sobre o funcionamento da simulação</i>	44
6.2	Plataforma web	45
6.2.1	<i>Funcionalidades</i>	45
6.2.1.1	<i>Resumo</i>	46
6.2.1.2	<i>Listas</i>	47
6.2.1.3	<i>Simulação</i>	49
7	CONCLUSÃO	50
	REFERÊNCIAS	51

1 INTRODUÇÃO

O ser humano, como espécie inteligente, tem como um de seus principais objetivos, se não o mais importante de todos, a sobrevivência, seja em relação a predadores ou em relação ao ambiente. No que remete ao ambiente, a humanidade atualmente está em luta para a mitigação das mudanças climáticas. Diversas abordagens para solucionar ou mitigar esse problemas já foram propostas e estão sendo executadas ao redor do mundo.

De acordo com Fawzy *et al.* (2020) as propostas podem ser categorizadas em três grandes abordagens. A primeira, usa das mitigações convencionais e utiliza-se de técnicas e tecnologias que buscam emitir menos carbono, como por exemplo energia renovável e energia nuclear. A segunda grande abordagem é a que foca na captura e sequestro de CO₂ da atmosfera, neste caso, a restauração ecológica desempenha um papel fundamental, enquanto a terceira abordagem gira em torno do princípio de alterar o equilíbrio de radiação da Terra por meio do gerenciamento da radiação solar e terrestre.

Tema central deste trabalho, a restauração ecológica desempenha um papel crucial no enfrentamento das mudanças climáticas. Como descrito por Harris *et al.* (2006), a restauração, incluindo a recuperação de ecossistemas degradados, é uma das estratégias mais eficazes para mitigar os impactos do aumento da concentração de dióxido de carbono na atmosfera. Ao restaurar áreas naturais, como florestas, savanas e zonas úmidas, o carbono é removido da atmosfera e armazenado em biomassa e solo, contribuindo significativamente para a redução dos gases de efeito estufa

Em paralelo aos avanços voltados para a restauração ecológica, há também um progresso significativo na área de dados e tecnologia, com a criação de grandes bases de dados globais que estão transformando a forma como a restauração é planejada e executada. De acordo com Ladouceur e Shackelford (2020), a síntese de dados em escalas globais tem o potencial de identificar padrões, reduzir incertezas e melhorar a previsibilidade dos resultados de restauração. Esses avanços oferecem ferramentas poderosas para integrar dados climáticos, topográficos e ecológicos, ajudando a otimizar as práticas de restauração ecológica e maximizando o impacto positivo dos esforços globais, como os da Década das Nações Unidas sobre a Restauração de Ecossistemas (2021-2030).

Este trabalho, por sua vez, tem como objetivo realizar uma análise exploratória sobre uma base de dados de espécies e ecossistemas com o objetivo de extrair novas informações bem como elaborar uma plataforma *online* para a disposição destes dados e, assim, realizar a

recomendação de espécies para restauração ecológica mediante entradas dadas pelo usuário. Tal abordagem leva em consideração diversos fatores como tipos de solo, diferentes tipos de habitats e a distribuição dos mesmos nas suas áreas de ocorrência.

A estrutura deste trabalho encontra-se da seguinte forma: na Seção 2 é apresentado o objetivo geral e os específicos da pesquisa; na Seção 3 são apresentados os assuntos essenciais que nortearão o restante do trabalho; na Seção 4 são elencados alguns trabalhos que se relacionam com este; e na Seção 5 é proposta a metodologia que será utilizada para abordar este problema; na Seção 6 serão abordados os resultados e, por fim, na Seção 7 será dada uma conclusão para os procedimentos e resultados trabalhados nessa pesquisa.

2 OBJETIVOS

2.1 Objetivo geral

Inspecionar, transformar e particionar uma base de dados dedicada a restauração ecológica para extração de novas informações que serão dispostas em um *dashboard* online, possibilitando a realização de simulações de seleção de espécies de plantas a serem utilizadas na restauração ecológica em determinadas regiões brasileiras.

2.2 Objetivo específicos

- Analisar o estado da arte para definir o problema central relacionado ao tema em estudo;
- Compreender o problema central e estudar o uso de tecnologias e técnicas de análise de dados para sua resolução;
- Limpar e normalizar a base de dados recebida;
- Analisar os dados recebidos de forma estatística e gráfica, buscando encontrar um maior entendimento das relações existentes;
- Gerar novas informações através do estudo das relações obtidas;
- Construir uma plataforma *online* onde o usuário pode visualizar e explorar as informações geradas e simular a seleção de espécies indicadas para a restauração ecológica por região.

3 FUNDAMENTAÇÃO TEÓRICA

Para um melhor entendimento do trabalho, nesta seção são abordados os principais conceitos necessários para o desenvolvimento desta pesquisa. Na seção 3.1, discute-se a definição de restauração ecológica. Na seção 3.2, é abordada a análise de dados. Nas suas subseções, exploram-se as subetapas da análise de dados: coleta de dados, processamento de dados e visualização de dados.

3.1 Restauração Ecológica

A Restauração Ecológica é um ramo da biologia que tem ganhado bastante destaque nos últimos anos. Para a Society for Ecological Restoration (2004), Restauração Ecológica pode ser definida como a aceleração e impulsionamento da recuperação de um ecossistema previamente deteriorado ou destruído. Ou seja, essa prática visa, através de várias metodologias, restaurar completamente ou parcialmente o estado original de um ecossistema.

Uma outra definição para o tema, vem do Centro Mundial de Monitoramento de Conservação das Nações Unidas UNEP-WCMC (2020) restauração ecológica é o processo de reverter a degradação dos ecossistemas para recuperar a sua funcionalidade ecológica, em benefício da natureza e das pessoas.

Como apontado no artigo de Oliveira e Engel (2011), a restauração ecológica é uma área que vem se expandindo nas últimas décadas e como consequência disto, torna-se um aparato essencial para o gerenciamento, conservação e recuperação de diversos ecossistemas globalmente. Esse aparato, por sua vez, é embasado cientificamente na ecologia da restauração, que por sua vez é o agrupamento de práticas, aspectos políticos, tecnológicos, econômicos e socio-culturais.

De acordo com o artigo *International Principles & Standards for the Practice of Ecological Restoration* de Gann *et al.* (2019) a restauração ecológica possui 8 princípios fundamentais, são eles:

- Princípio 1: A restauração ecológica estimula o envolvimento de diversos atores;
- Princípio 2: A restauração ecológica se baseia em muitos tipos de conhecimento;
- Princípio 3: A prática da restauração ecológica é baseada em ecossistemas de referência nativos, considerando as mudanças ambientais;
- Princípio 4: A restauração ecológica apoia os processos de recuperação de ecossistemas;

- Princípio 5: A recuperação do ecossistema é avaliada por metas e objetivos claros, utilizando indicadores mensuráveis;
- Princípio 6: Restauração ecológica visa ao mais alto nível de recuperação alcançável.
- Princípio 7: A restauração ecológica ganha valor cumulativo quando aplicada em grandes escalas.
- Princípio 8: A restauração ecológica faz parte de um continuum de atividades restaurativas.

Este trabalho terá uma ênfase maior no terceiro princípio, que propõe o uso de ecossistemas de referência para o processo de restauração ecológica.

3.1.1 *Ecossistemas de Referência*

No artigo de Gann *et al.* (2019), é dito que um modelo de referência de ecossistema, como pode ser visto na primeira imagem da direita para a esquerda na figura 1, é um ecossistema natural que serve como meta ou guia para a restauração ecológica de um ecossistema degradado. A importância de um modelo de referência está na capacidade de orientar as ações de restauração, fornecendo informações sobre a estrutura, a composição, a diversidade e o funcionamento do ecossistema original.

Para Gann *et al.* (2019), os modelos de referência devem ser construídos com base em ecossistemas reais específicos que sejam objetos de esforços de conservação e restauração. Ademais, para a construção dos mesmos, é importante também a utilização de diversas fontes de informação e contar com a consulta de pessoas envolvidas, detalhes históricos e informações coletadas durante avaliações da área proposta como referência. Em suma, um modelo de referência ideal descreve a condição aproximada que o local teria se não tivesse sofrido degradação.

De acordo com Society for Ecological Restoration (2004), Ruiz-Jaen e Aide (2005) um modelo de referência também pode ajudar a avaliar o sucesso da restauração, comparando os indicadores ecológicos do ecossistema restaurado com os do ecossistema de referência.



Figura 1 – Exemplo de ecossistema de referência e três estágios de degradação. Fotos são das terras altas do sul do Brasil, Rio Grande do Sul (Fotos: J.-M. Hermann, Jan. 2012). Adaptado de Overbeck *et al.* (2013)

3.2 Análise de Dados

Análise de Dados, de acordo com Albright *et al.* (2011), foi um sinônimo para o termo estatística por muito tempo, porém, com a imensidão de dados que é gerado atualmente através de sensores, *clicks* na internet, transações, geração de conteúdo e outras formas, o termo ganhou uma significância maior. Ademais, Albright *et al.* (2011) menciona que essa área atualmente é e continuará a ser bastante relevante nos próximos anos.

Desde muitos anos atrás, a análise de dados já ganhava força, no artigo de Tukey (1962), é possível entender que na época já se havia uma ideia para onde área estava se expandindo. De acordo com o autor, a mesma pode ser definida como o conjunto de procedimentos para análise de dados, técnicas para interpretar os resultados desses procedimentos, formas de planejar a coleta de dados para tornar sua análise mais fácil, precisa ou acurada, e toda a estrutura e resultados da estatística que se aplicam à análise de dados.

No livro de Hand *et al.* (2001), Análise de Dados pode ser descrita como geração de hipóteses orientada à dados, onde ao examinar dados coletados, busca-se por estruturas que possam indicar relacionamentos profundos entre variáveis e problemas.

De acordo com Statista (2021), até 2020 o total de dados criados, capturados, copiados e consumidos globalmente chegou a 64.2 *zettabytes* e existe uma previsão que até 2025 esse total passe a ser maior que 180 *zettabytes*. Isso enfatiza, mais uma vez que, em decorrência da quantidade gigantesca de dados disposta de forma *online*, atualmente, a Análise de Dados, juntamente com Mineração de Dados e outras áreas voltadas à dados passaram a ser de extrema importância na hora de extrair novas informações e classificar todos esses dados.

No livro *Doing Data Science* de O'Neil e Schutt (2013) são elencados alguns passos que antecedem a Análise de Dados, tais como: definição do problema, coleta de dados, pré-processamento de dados e por fim a visualização desses dados, onde através destes, a etapa de análise pode ser realizada. Alguns destes passos estão listados na Figura 2 e serão melhor descritos nas subseções posteriores.

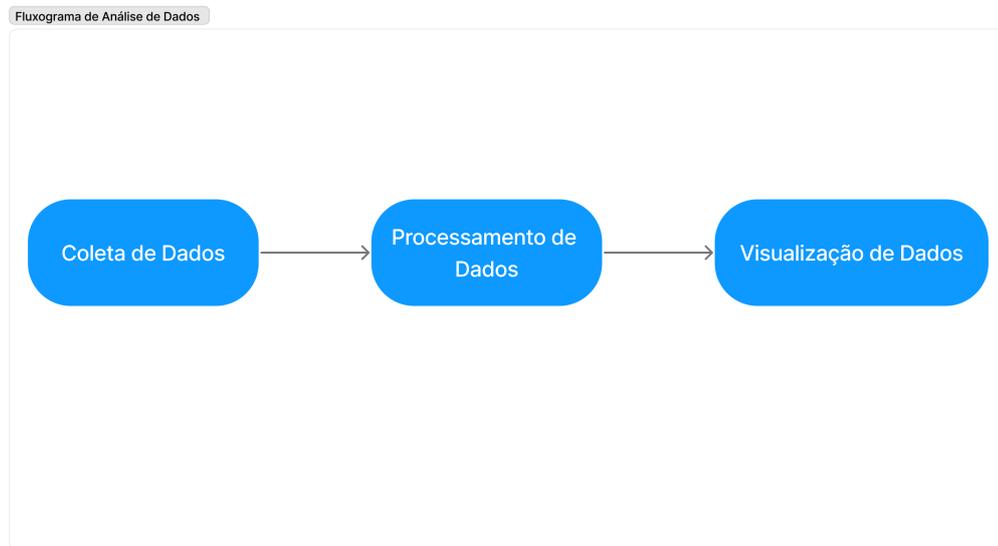


Figura 2 – Fluxograma da Análise de Dados (Simplificado). Fonte: elaborado pelo autor.

3.2.1 *Coleta de Dados*

De acordo com Shireen (2022) para a construção de sistemas inteligentes que operam sobre base de dados, tais como análises estatísticas e algoritmos de inteligência artificial requerem uma boa base de dados, portanto, esta é uma das partes mais importantes de um projeto.

A coleta de dados é o meio que nos permite registrar grandes quantidades de eventos e serve como um pontapé inicial para projetos de análise exploratória e aprendizado de máquina que consigam extrair padrões e relações não tão fáceis de serem identificadas a olho nu ou comprovadas cientificamente.

O produto desta etapa é a base de dados e ela pode ser montada de diversas formas. Segundo Shireen (2022), dados em texto, imagens, vídeos, áudio e outros podem ser utilizados para a construção da mesma. As etapas necessárias na coleta de dados, de acordo com o artigo de Bhandari (2020) podem ser listados da seguinte forma:

- Definir o foco da pesquisa;
- Escolher o método de coleta de dados;
- Planejar o procedimento de coleta de dados;
- Coletar os dados.

3.2.2 *Pré-Processamento de Dados*

Como citado na seção anterior, os dados de uma base de dados podem ser de diversos tipos, estruturados, não estruturados, numéricos, categóricos, textuais, entre outros. Porém, quando uma grande quantidade de dados é coletada, ocasionalmente podem existir problemas com a base de dados, como por exemplo, dados faltosos ou com erros. Portanto, é nesta etapa que os dados coletados são tratados afim de promover uma base de dados de qualidade, removendo diversas falhas que possam ter surgido durante a montagem da mesma.

Segundo Famili *et al.* (1997), o tratamento dos dados é uma etapa crucial para a análise de dados, pois é a através desta etapa que o refinamento da base de dados é feita para prevenir enganos durante a análise de dados. Muitas vezes são necessários diversos tratamentos nestas bases, principalmente ao trabalhar com dados do mundo real.

Para Famili *et al.* (1997) os motivos que levam ao pré-processamento de dados são:

- solucionar problemas de dados que possam impedir a realização de qualquer tipo de análise nos dados;
- compreender a natureza dos dados e realizar uma análise mais significativa;
- extrair conhecimentos mais significativos de um determinado conjunto de dados.

Alguns dos processos mais comuns dentro desta etapa de pré-processamento de dados são:

- **Limpeza de dados:** de acordo com Ridzuan e Wan Zainon (2019), pode ser definida como um conjunto de operações que podem ser realizadas afim de remover anomalias na base de dados e resolver inconsistências transformando os dados em um formato mais conciso com a realidade.
- **Transformação dos dados:** análises exploratórias, bem como algoritmos que utilizam-se de uma base de dados tendem a ser intrinsecamente relacionados a mesma. Dessa forma é necessário que os valores a serem estudados estejam escalonados de forma coerente. Segundo Raschka (2014), dados podem ser normalizados através de técnicas como *escore-Z*, *mínimo-máximo*, dados nominais e dados ordinais.

Em resumo, o pré-processamento de dados é uma etapa essencial que visa preparar os dados para análise, normalizando valores para uma mesma escala ou transformando dados de forma a criar uma base mais coerente e por consequência disso, melhorando sua qualidade, permitindo uma análise mais significativa e a extração de informações relevantes. É um processo fundamental para obter resultados confiáveis e tomar decisões mais precisas levando em

consideração uma boa base de dados.

3.2.3 *Visualização de Dados*

Visualizações efetivas permitirão que um cientista compreenda seus próprios dados e comunique suas percepções a outros. (TUKEY *et al.*, 1977) Outra área de grande importância da Análise de Dados é a Visualização de Dados, é nesta etapa, que acontece após o pré-processamento dos dados, que se encontra os resultados, gráficos e tabelas referentes as informações extraídas de uma base dados. É através da visualização que é possível explorar e comunicar informações complexas de forma mais intuitiva e acessível.

Para Heer e Shneiderman (2012) a visualização permite aos usuários explorar grandes volumes de dados, identificar tendências, padrões, anomalias e obter uma compreensão mais profunda dos dados. Além disso, segundo Segel e Heer (2011), a visualização ajuda a comunicar os resultados de forma clara e persuasiva, facilitando a comunicação entre especialistas e o público em geral.

De acordo com Sadiku *et al.* (2016), as técnicas mais comuns para amostragem dos dados em gráficos são:

- Gráfico de linha: mostra a relação entre itens. Pode ser usado para comparar mudanças ao longo de um período de tempo;
- Gráfico de barras: é usado para comparar quantidades de diferentes categorias;
- Gráfico de dispersão: é um gráfico bidimensional que mostra a variação de dois itens;
- Gráfico de pizza: isso é usado para comparar as partes de um todo.

Porém é importante mencionar que estes não são os únicas técnicas de visualização de dados. De acordo com Miller (2019) existem outros tipos como: histogramas, mapas de calor, mapas coropléticos, matrizes de correlação (a Figura 3 é um exemplo).



Figura 3 – Exemplo de matriz de correlação. Fonte: Miller (2019)

4 TRABALHOS RELACIONADOS

Este capítulo descreve trabalhos da literatura mais relevantes para a contextualização do problema proposto nesta pesquisa.

4.1 *Utilização de ecossistemas de referência ótimos para restauração ecológica*

O trabalho de Toma *et al.* (2023) introduz e aprofunda um conceito bem relevante para a restauração ecológica e este é o conceito de ecossistemas de referência ótimos. Que por sua vez são ecossistemas que continuaram naturais e ou tiveram pouquíssimo impacto no seu equilíbrio original.

Toma *et al.* (2023) ao definir ecossistemas de referência ótimos, estende alguns conceitos apresentados nos Padrões SER de Gann *et al.* (2019), onde um exemplo de ótimo se encaixa dentro três grandes características principais, são elas:

- As condições do ecossistema de referência candidato a categoria de ótima devem ser excelentes;
- No caso de candidatos que já tiveram algumas mudanças, caso estas sejam fáceis de reverter e/ou corrigir então a correção é desejável;
- E por fim, se ambas as afirmações acima forem dadas como verdade, o único ponto remanescente para um ecossistema ser classificado como ótimo, é a distância espacial entre este e o ecossistema a ser restaurado.

De forma resumida, ecossistemas de referência ótimos são nada mais que regiões intocadas ou quase não tocadas em relação a impactos antrópicos e que estes se encontrem espacialmente próximos dos locais alvos de restauração ecológica onde antes da degradação existia o mesmo ecossistema.

Na Figura 4, Toma *et al.* (2023) descreve três possíveis cenários na disponibilidade de ecossistemas de referência:

- **Cenário A:** referências ótimas estão amplamente disponíveis e há uma maior chance de sucesso da restauração;
- **Cenário B:** referências ótimas são escassas, entretanto, possuindo pelo menos uma, é possível utiliza-la como referência;
- **Cenário C:** não existem referências ótimas nas proximidades, ou seja, um modelo terá que ser construído e os riscos de falha na restauração são maiores.

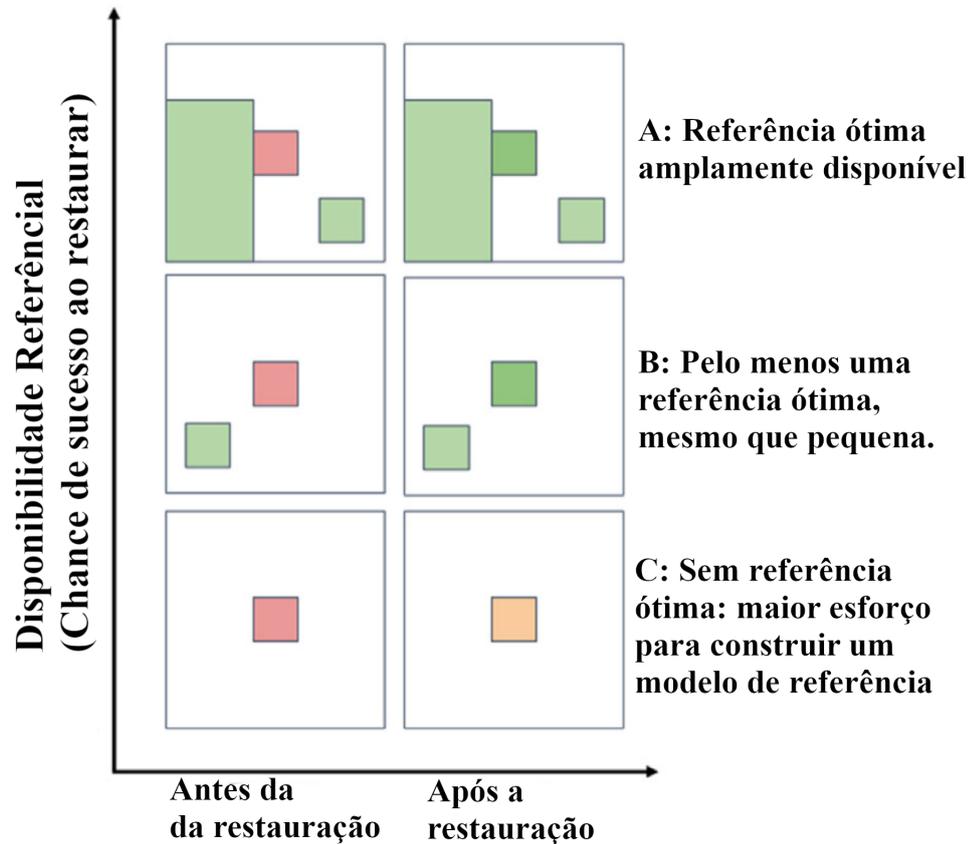


Figura 4 – Importância dos Ecossistemas de Referência. Adaptado de Toma *et al.* (2023)

O trabalho de Toma *et al.* (2023) serve de base fundamental para este trabalho, que busca integrar a análise de dados à construção de ecossistemas de referência, de forma que seja possível, através de algoritmos, detectar e indicar possíveis espécies de plantas que possam ser utilizadas para restauração, dada uma base de dados que será utilizada como referencial.

4.2 Sistema de Recomendação de Plantio na Agricultura

Assim como em pesquisas anteriores, o trabalho de Athayde (2023) foi desenvolvido com o propósito de acompanhar o desenvolvimento de um sistema de recomendação de plantio voltado para a agricultura. Nele, o autor utiliza diversos algoritmos de aprendizado de máquina, assim como estatística e análise de dados, para fomentar a evolução de soluções inteligentes voltadas ao plantio.

Neste trabalho, o autor utiliza uma base de dados aberta e de domínio público construída na Índia, disponibilizada na plataforma *Kaggle*. Os atributos escolhidos pelo autor são: nitrogênio, potássio, fósforo, temperatura média, umidade média, pH do solo e índice pluviométrico.

Durante a etapa de pré-processamento de dados, o autor identificou diversos problemas comuns em bases de dados reais, como a presença de valores nulos e a distribuição desigual dos tipos de plantas. Foram realizadas análises estatísticas nos atributos, incluindo a criação de histogramas para verificação de simetria, o que pode revelar diferenças de escala entre os atributos e impactar negativamente os algoritmos aplicados posteriormente.

Para a criação do sistema de recomendação, Athayde (2023) desenvolveu modelos utilizando quatro algoritmos diferentes: *Random Forest*, *K-Nearest Neighbors*, *Support Vector Machine* e *Naive Bayes*. Diversas métricas de precisão foram utilizadas para determinar o melhor modelo a ser empregado.

A conclusão do trabalho foi que os modelos criados com *Random Forest* e *Naive Bayes* apresentaram os melhores resultados, atingindo valores acima de 99% em todas as métricas. No entanto, o modelo escolhido para continuidade foi o *Random Forest*, por lidar melhor com *outliers*, *overfitting* e *underfitting*.

Em relação a esta pesquisa, ambos os trabalhos são semelhantes na aplicação de técnicas de análise de dados para desenvolver sistemas de recomendação no contexto de plantio. Enquanto o trabalho de Athayde (2023) foca na agricultura, visando recomendar culturas agrícolas ideais com base em atributos do solo e clima, esta pesquisa concentra-se na restauração ecológica, buscando indicar espécies nativas adequadas para a recuperação de ecossistemas específicos.

Ambos os estudos destacam a importância de abordagens baseadas em dados para aprimorar processos decisórios, evidenciando como a integração entre computação e ciências ambientais pode gerar soluções eficazes para desafios distintos, mas relacionados ao uso sustentável do solo.

4.3 Utilização de IoT e Aprendizado de Máquina para seleção de plantas na agricultura

Com o avanço de tecnologias sensoriais, da ciência de dados e dos algoritmos de aprendizado de máquina, muito tem se avançado na forma como estudamos e exploramos dados atualmente.

O trabalho de Ruchirawya (2020) é voltado para o desenvolvimento de um sistema de recomendação de plantio que utiliza Internet das Coisas (IoT), aprendizado de máquina e processamento de linguagem natural para ajudar os agricultores a escolher as plantas mais adequadas para suas terras com base em fatores ambientais.

O sistema coleta dados de sensores acoplados a microcontroladores criados com arduino, como temperatura, luz solar, umidade do solo e pH, e os processa com os seguintes algoritmos de aprendizado: *Naive Bayes*, *Support Vector Machine* e *K-Means*, para prever o melhor tipo de planta para a área selecionada.

Os algoritmos são treinados com base em dados históricos e podem aprender com novos dados à medida que se tornam disponíveis. Isso permite que o sistema melhore sua precisão ao longo do tempo.

Os resultados desse trabalho mostram que o sistema é capaz de prever com precisão o melhor tipo de planta para uma determinada área com base nos fatores ambientais previamente mostrados. O sistema foi testado em várias localidades e apresentou resultados promissores em termos de melhoria da eficiência, produtividade e lucratividade da agricultura.

O trabalho de Ruchirawya (2020) assemelha-se a este trabalho pois, para o replantio de espécies é necessário utilizar-se de dados que compartilham da mesma estrutura dos abordados no trabalho do referido autor. A grande diferença do trabalho do autor para este e em sua aplicação. Nesta pesquisa os dados são coletados com o propósito de recomendação de espécies, porém, com limitantes estabelecidos ao se realizar restauração ecológica.

4.4 Aplicando técnica de Ensembles para recomendação de plantio na Índia

Agricultura de precisão é uma abordagem aplicada a agricultura que visa utilizar-se de técnicas modernas, algumas já citadas anteriormente, como: sensores, GPS, drones, entre outras. Isso com o propósito de atingir melhores resultados quanto ao plantio e também garantir uma supervisão maior do crescimento das plantas e observação do solo e clima.

Com isso, no trabalho de Pudumalar *et al.* (2017), o autor utiliza de uma técnica conhecida como *Ensemble*, onde o mesmo aplica diversos algoritmos de aprendizado de máquina em uma base de dados e toma uma decisão através de uma votação pela maioria. De forma resumida, os algoritmos classificam qual a melhor espécie que pode ser plantada naquele solo e a espécie mais votada entre os algoritmos é a escolhida.

A base de dados usada pelo autor foi coletada em Madurai, na Índia e é bem semelhante à base usada no trabalho anterior, de Ruchirawya (2020), porém com uma ênfase maior no solo, possuindo assim, algumas características a mais. Os atributos escolhidos foram: profundidade, textura, pH, cor do solo, permeabilidade, drenagem, retenção de água e erosão.

Quanto aos algoritmos de aprendizado de máquina que foram utilizados na técnica

de *Ensemble*, o autor optou pelos algoritmos *Random Tree*, *CHAID (Chi-squared Automatic Interaction Detection)*, *K-Nearest Neighbors* e *Naive Bayes*. A metodologia aplicada por Pudumalar *et al.* (2017) atingiu uma precisão de 88%.

O trabalho de Pudumalar *et al.* (2017) se relaciona muito bem com o proposto nesta pesquisa, semelhante ao anterior, esse trabalho também trata da criação de um sistema de recomendação de plantio baseado em dados.

5 PROCEDIMENTOS METODOLÓGICOS

Este capítulo apresenta a metodologia utilizada neste trabalho, que se resume na compressão da problemática em estudo, assim como a origem dos dados processados e seu entendimento. Além disso, será descrito também as etapas utilizadas para o pré-processamento de dados, procedimentos para a análise exploratória dos dados, a construção do algoritmo de ranqueamento das espécies para recomendação e a construção da plataforma *online*.

A Figura 5 representa, em alto nível, o fluxograma com todas as etapas da metodologia aplicada para geração dos resultados apresentados no capítulo seguinte.

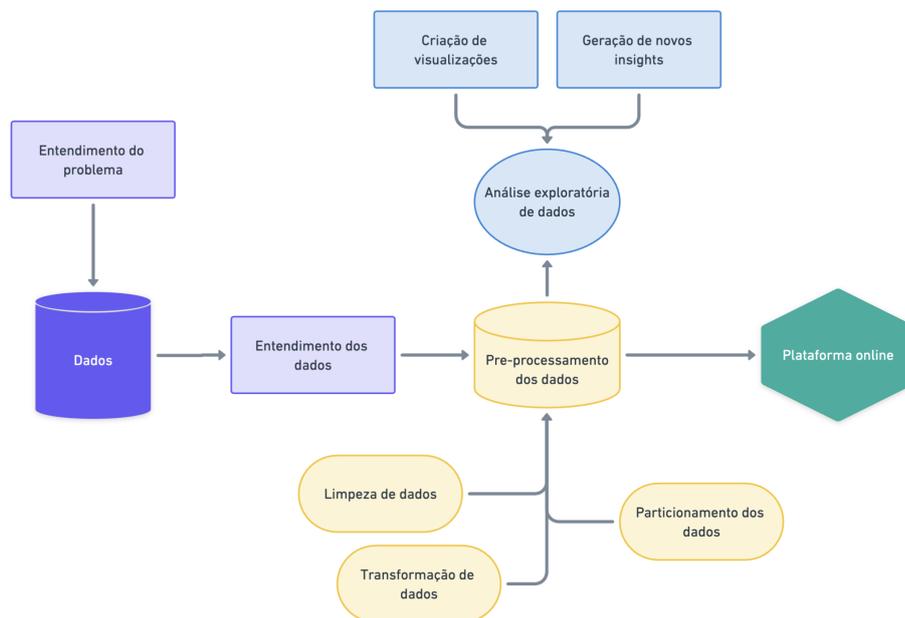


Figura 5 – Fluxograma da Análise de Dados (Detalhado). Fonte: elaborado pelo autor.

5.1 Entendimento do problema

Esta primeira etapa teve como objetivo central entender o problemática proposta e definir os objetivos a serem alcançados ao explorar e analisar a base de dados fornecida. Considerando o tema restauração ecológica, o objetivo central culminou na realização de uma análise exploratória sobre uma base de dados com listas de espécies brasileiras catalogadas por regiões e coletadas a partir de publicações científicas.

5.2 Entendimento da base de dados

Nesta etapa foi realizada uma exploração inicial dos dados disponibilizados pelos pesquisadores da Universidade Federal de Minas Gerais, incluindo sua organização e documentação. Durante essa fase, foi essencial compreender os dados, identificar informações relevantes e avaliar a qualidade dos mesmos. O objetivo central desta etapa foi descobrir padrões entre os dados, como estes se relacionavam entre si e entender como extrair novas informações a partir disso.

Durante a análise inicial da base de dados, foi observado que a base de dados foi construída em torno de listas de espécies coletadas em campo, por diversos pesquisadores. Estas listas foram catalogadas e divididas em várias planilhas.

A seguir são apresentadas os dados encontrados em cada uma das planilhas disponibilizadas:

- INFO: esta planilha contém algumas documentações básicas sobre a estrutura de toda a base de dados. Por exemplo: abreviações, e o nome e significado de cada coluna nas demais planilhas;
- Publications: esta planilha contém as diversas publicações que foram utilizadas para a construção das planilhas de Lists e AfterFlora. Cada publicação utilizada nesta base de dados gerou uma lista que contém dados específicos que serão discutidos posteriormente;
- Lists: esta planilha contém todas as 110 listas obtidas através destas publicações. Cada lista contém informações sobre: número da lista, identificador de publicação, referência da publicação, bioma, substrato, habitat, forma de vida, localidade, município, estado federal, latitude e longitude entre outras. Nem todas as colunas foram utilizadas nesta planilha;
- AfterFlora: esta planilha contém todas as entradas de cada lista, sendo a principal fonte de informação nas análises. A planilha está dividida nas colunas: código da lista, tipo da lista, substrato, habitat, iv100, nome científico e família.

5.2.1 Lists

Na planilha de Lists (figura 6), estão presentes todas as 110 listas obtidas através das publicações mencionadas anteriormente. Cada lista fornece informações específicas sobre bioma, solo, habitats, forma de vida, localidade, município, estado federal, latitude e longitude, entre outros.

listnumber	biome	substrate	habitat	growthform	locationname	municipality	federalstate	latitude	longitude	altitude
L0001	cerrado	ironstone	ferruginous_grasslands	all	Serra do Rola Moca	Brumadinho	MG	20°03'60"S	44°02'00"W	1450
L0002	cerrado	ironstone	ephemeral_pond	herbs	Chapada do Canga	Mariana	MG	20°9'35.29" S	43°25'44.78" W	910
L0003	cerrado	ironstone	ferruginous_grasslands	herbs	Chapada do Canga	Mariana	MG	20°9'48.62" S	43°26'8.43" W	910
L0004	cerrado	ironstone	ferruginous_grasslands	woody	Chapada do Canga	Mariana	MG	20°9'48.62" S	43°26'8.43" W	910
L0005	cerrado	ironstone	ferruginous_grasslands	all	Mina do Segredão	Sabará	MG	19°50'44" S	43°46'16" W	1240
L0006	cerrado	ironstone	canga_sensu_lato	all	Mina do Sapecado	Itabirito	MG	20°14'60.0"S	43°52'34.9"W	1300
L0007	cerrado	ironstone	ferruginous_grasslands	all	Fazenda Água Limpa	Conceição do Mato Dentro	MG	18°56'11.68" S	43°24'55.39" W	918
L0008	cerrado	quartzite	quartzitic_grasslands	all	Fazenda Boa Esperança	Conceição do Mato Dentro	MG	18°55'28.18" S	43°28'38.64" W	770
L0009	cerrado	ironstone	ferruginous_grasslands	woody	Serra da Calçada	Nova Lima	MG	20° 06'48.8" S	43°57'30.3"W	1389
L0010	cerrado	ironstone	ferruginous_grasslands	herbs	Serra da Calçada	Nova Lima	MG	20° 06'48.8" S	43°57'30.3"W	1389
L0011	cerrado	ironstone	ferruginous_grasslands	woody	Serra da Calçada	Brumadinho	MG	20°05'55.2" S	43°58'58.2" W	1389
L0012	cerrado	ironstone	ferruginous_grasslands	herbs	Serra da Calçada	Brumadinho	MG	20°05'55.2" S	43°58'58.2" W	1389
L0013	cerrado	quartzite	quartzitic_grasslands	woody	Serra da Calçada	Brumadinho	MG	20°06'59.5" S	43°59'27.8" W	1389
L0014	cerrado	quartzite	quartzitic_grasslands	herbs	Serra da Calçada	Brumadinho	MG	20°06'59.5" S	43°59'27.8" W	1389
L0015	cerrado	quartzite	quartzitic_outcrop	woody	Serra da Calçada	Brumadinho	MG	20°06'56.4" S	43°59'17.1" W	1389
L0016	cerrado	quartzite	quartzitic_outcrop	herbs	Serra da Calçada	Brumadinho	MG	20°06'56.4" S	43°59'17.1" W	1389
L0017	cerrado	quartzite	quartzitic_grasslands	all	Serra do Cipó (Cedro)	Santana do Riacho	MG	19°13'56.5" S	43°34'34.8" W	1100
L0018	cerrado	quartzite	quartzitic_grasslands	all	Serra do Cipó (Pedra do elefante)	Santana do Riacho	MG	19°17'43.0" S	43°33'17.4" W	1200
L0019	cerrado	quartzite	quartzitic_grasslands	all	Serra do Cipó (Q16)	Santana do Riacho	MG	19°17'49.6" S	43°35'28.2" W	1300
L0020	cerrado	quartzite	quartzitic_grasslands	all	Serra do Cipó (Alto Palácio)	Santana do Riacho	MG	19°16'59.3" S	43°32'08.9" W	1400
L0021	cerrado	quartzite	quartzitic_grasslands	all	Serra do Cipó, MG (sandy)	Santana do Riacho	MG	19°17'6.2" S	43°35'43.2" W	1156
L0022	cerrado	quartzite	quartzitic_grasslands	all	Serra do Cipó, MG (sandy)	Santana do Riacho	MG	19°17'20" S	43°34'58.6" W	1178
L0023	cerrado	quartzite	quartzitic_grasslands	all	Serra do Cipó, MG (sandy)	Santana do Riacho	MG	19°17'9.8" S	43°35'15.5" W	1188
L0024	cerrado	quartzite	quartzitic_grasslands	all	Serra do Cipó, MG (sandy)	Santana do Riacho	MG	19°17'47.3" S	43°35'24.1" W	1291
L0025	cerrado	quartzite	quartzitic_grasslands	all	Serra do Cipó, MG (sandy)	Santana do Riacho	MG	19°16'11.4" S	43°34'46.6" W	1091
L0026	cerrado	quartzite	quartzitic_grasslands	all	Serra do Cipó, MG (stony)	Santana do Riacho	MG	19°17'4.7" S	43°35'38.2" W	1162
L0027	cerrado	quartzite	quartzitic_grasslands	all	Serra do Cipó, MG (stony)	Santana do Riacho	MG	19°17'21.6" S	43°35'7.3" W	1273

Figura 6 – Exemplo dos dados na planilha de Lists. Fonte: elaborado pelo autor.

Na planilha de dados em Lists, cada coluna que foi utilizada nesta pesquisa significa:

- listnumber: número da lista conforme foi adicionada à base de dados;
- biome: bioma onde a lista foi coletada. Para esta pesquisa, somente o bioma Cerrado foi utilizado;
- substrate: substrato, remete ao tipo de solo de onde foi coletado os dados da lista. Dentro da base de dados 2 tipos, sendo eles, afloramento ferruginoso (*ironstone*) e afloramento quartzítico (*quartzite*);
- habitat: indica o tipo de habitat em que a lista foi coletada. Dentro da base existem 7 tipos, são eles: *Ferruginous Grassland*, *Ephemeral Pond*, *Candéal*, *Quartz Grassland*, *Quartz Outcrop*, *Forest Island* e *Peat Bog*;
- growthform: forma de vida das espécies contidas na lista. Na base existem 3 tipos, lenhosas (*woody*), herbáceas (*herb*) e todas (*all*);
- locationname: nome da localidade;
- municipality: município;
- federalstate: estado federal. Na base só foi analisado os estados de Minas Gerais (MG) e Bahia (BH);
- latitude: latitude em graus, minutos e segundos;

- longitude: longitude em graus, minutos e segundos.

5.2.2 AfterFlora

Na planilha de AfterFlora (figura 7), são apresentadas as entradas de cada lista, sendo a fonte principal para as análises realizadas. As colunas incluem informações que ligam com a planilha Lists, tais como número da lista, publicação, entre outras.

listnumber	list_type	substrate	ecotype_GW	publicationref	iv100	scientific.name	family
L0001	Herb	ironstone	Ferruginous_grassland	jacobi_et_al_2008	12.26	Andropogon ingratus Hack.	Poaceae
L0001	Herb	ironstone	Ferruginous_grassland	jacobi_et_al_2008	9.66	Bulbostylis fimbriata (Nees) C.B.Clarke	Cyperaceae
L0001	Herb	ironstone	Ferruginous_grassland	jacobi_et_al_2008	9.22	Cattleya caulescens (Lindl.) Van den Berg	Orchidaceae
L0001	Herb	ironstone	Ferruginous_grassland	jacobi_et_al_2008	5.13	Acianthera teres (Lindl.) Borba	Orchidaceae
L0001	Herb	ironstone	Ferruginous_grassland	jacobi_et_al_2008	4.16	Paspalum scalare Trin.	Poaceae
L0001	Herb	ironstone	Ferruginous_grassland	jacobi_et_al_2008	4.05	Vellozia graminea Pohl	Velloziaceae
L0001	Herb	ironstone	Ferruginous_grassland	jacobi_et_al_2008	3.35	Struthanthus flexicaulis (Mart.) Mart.	Loranthaceae
L0001	Herb	ironstone	Ferruginous_grassland	jacobi_et_al_2008	1.36	Lagenocarpus rigidus Nees	Cyperaceae
L0001	Herb	ironstone	Ferruginous_grassland	jacobi_et_al_2008	1.15	Portulaca hirsutissima Cambess.	Portulacaceae
L0001	Herb	ironstone	Ferruginous_grassland	jacobi_et_al_2008	0.69	Epidendrum secundum Jacq.	Orchidaceae
L0001	Herb	ironstone	Ferruginous_grassland	jacobi_et_al_2008	0.28	Axonopus siccus (Nees) Kuhlms.	Poaceae
L0002	Herb	ironstone	Ephemeral_pond	pereira_2010_diss_uvf	65.37	Vellozia graminea Pohl	Velloziaceae
L0002	Herb	ironstone	Ephemeral_pond	pereira_2010_diss_uvf	9.69	Axonopus siccus (Nees) Kuhlms.	Poaceae
L0002	Herb	ironstone	Ephemeral_pond	pereira_2010_diss_uvf	7.5	Rhynchospora consanguinea (Kunth) Boeckeler	Cyperaceae
L0002	Herb	ironstone	Ephemeral_pond	pereira_2010_diss_uvf	5.13	Bulbostylis eleocharioides Kral & M.T.Strong	Cyperaceae
L0002	Herb	ironstone	Ephemeral_pond	pereira_2010_diss_uvf	3.18	Rhynchospora setigera (Kunth) Griseb.	Cyperaceae
L0002	Herb	ironstone	Ephemeral_pond	pereira_2010_diss_uvf	2.55	Axonopus conduplicatus G.A. Black	Poaceae
L0002	Herb	ironstone	Ephemeral_pond	pereira_2010_diss_uvf	1.89	Vellozia tragacantha (Mart. ex Schult. & Schult.f.) Mart. ex Seub.	Velloziaceae
L0002	Herb	ironstone	Ephemeral_pond	pereira_2010_diss_uvf	1.77	Trichantheum distichophyllum (Spreng.) Zuloaga & Morrone	Poaceae
L0002	Herb	ironstone	Ephemeral_pond	pereira_2010_diss_uvf	1.5	Paspalum pilosum Lam.	Poaceae
L0002	Herb	ironstone	Ephemeral_pond	pereira_2010_diss_uvf	1.47	Cambessedesia hilariana (Kunth) DC.	Melastomataceae
L0003	Herb	ironstone	Ferruginous_grassland	pereira_2010_diss_uvf	38.63	Axonopus siccus (Nees) Kuhlms.	Poaceae
L0003	Herb	ironstone	Ferruginous_grassland	pereira_2010_diss_uvf	12.87	Hoplocryptanthus schwackeanus (Mez) Leme, S.Heller & Zizka	Bromeliaceae
L0003	Herb	ironstone	Ferruginous_grassland	pereira_2010_diss_uvf	12.85	Dyckia rariflora Schult. & Schult.f.	Bromeliaceae
L0003	Herb	ironstone	Ferruginous_grassland	pereira_2010_diss_uvf	10.22	Trichantheum wettsteinii (Hack.) Zuloaga & Morrone	Poaceae
L0003	Herb	ironstone	Ferruginous_grassland	pereira_2010_diss_uvf	8.29	Lagenocarpus rigidus Nees	Cyperaceae
L0003	Herb	ironstone	Ferruginous_grassland	pereira_2010_diss_uvf	5.94	Vellozia graminea Pohl	Velloziaceae
L0003	Herb	ironstone	Ferruginous_grassland	pereira_2010_diss_uvf	4.01	Gomesa gracilis (Lindl.) M.W.Chase & N.H.Williams	Orchidaceae
L0003	Herb	ironstone	Ferruginous_grassland	pereira_2010_diss_uvf	2.9	Acianthera teres (Lindl.) Borba	Orchidaceae

Figura 7 – Exemplo dos dados na planilha de AfterFlora. Fonte: elaborado pelo autor.

Na planilha de dados em AfterFlora, cada coluna que foi utilizada nesta pesquisa significa:

- listnumber: código representando de qual lista vem essa espécie;
- list_type: indica em qual tipo da lista a espécie foi atrelada;
- substrate: substrato atrelado à espécie;
- ecotype-2 (habitat): habitat atrelado à espécie;
- iv100: índice de importância da espécie (média de densidade relativa, frequência relativa e dominância relativa; soma de 100%.)
- scientific.name: nome científico da espécie;
- family: família da espécie;

5.3 Pré-processamento dos dados

Nesta etapa da pesquisa, ocorreu a preparação da base de dados para ser utilizada nas etapas subsequentes. Esta etapa foi considerada crucial, pois visava aprimorar a qualidade da base de dados. Durante esse processo, foram removidos dados incorretos, preenchidos valores ausentes, descartadas colunas desnecessárias e aplicadas normalizações, entre outros procedimentos de tratamento. Nas próximas subseções será mostrado o que foi realizado em cada etapa do pré-processamento.

5.3.1 Seleção dos atributos mais importantes

Na etapa de pré-processamento foram selecionados os atributos mais relevantes para o escopo da pesquisa, descartando atributos considerados dispensáveis e eventuais metadados que não teriam utilidade dentro do escopo desta pesquisa. Assim, apenas as planilhas *Lists* e *Afterflora* foram consideradas importantes para a construção das análises e da plataforma. Como citado na Seção 5.2, os dados contidos e que utilizados na análise realizada foram:

- Planilha Lists: *list*, *biome*, *substrate*, *habitat*, *location*, *city*, *state*, *latitude*, *longitude* e *altitude*;
- Planilha Afterflora: *list*, *type*, *substrate*, *habitat*, *iv100*, *scientificname* e *family*.

5.3.2 Limpeza dos dados

Nesta etapa ocorreu a remoção *outliers*, ou seja, valores que se diferenciavam significativamente dos demais presentes na base de dados, sendo também. Neste processo, foram eliminados dados faltantes e valores discrepantes em relação ao esperado para determinada coluna, assim como outras falhas identificadas na base de dados.

Inicialmente, algumas listas dentro da planilha Lists tiveram de ser inteiramente removidas, incluindo suas ocorrências na planilha AfterFlora. O motivo para se deu pela falta de dados e/ou incoerência dos dados coletados, tornando inviável a utilização das listas. Após a remoção das listas defeituosas restaram um total de 101 listas.

As colunas latitude e longitude na planilha Lists precisaram ser ajustadas. Para estas, os ajustes consistiram em basicamente garantir que todas as linhas possuíssem o formato correto de coordenada em graus, minutos e segundos (por exemplo: 20°05'55.2" S) para depois serem convertidas para graus decimais (por exemplo: -43.582944). Esta conversão nos valores

de latitude e longitude foram necessárias pois a biblioteca de visualização de mapas tanto nas visualizações quanto na plataforma, depende que a entrada, as coordenadas neste caso, obedecam este formato.

5.3.3 *Construção de novos atributos*

Nesta última etapa do pré-processamento de dados ocorreu a criação de novos dados a partir de atributos já existentes nas planilhas assim como a criação de *datasets* menores e mais específicos.

O passo inicial que houve nesta etapa foi a junção das planilhas de Lists e AfterFlora. Estas que por sua vez já haviam passado por um pré-processamento para remover colunas desnecessárias para as análises que serão discutidas mais adiante. A criação desse *dataset* unificado foi de grande importância, pois as análises exploratórias, o algoritmo de simulação e a plataforma *online* dependem desses dados unificados.

Também é válido mencionar que as colunas que foram migradas para esse novo *dataset* são: *list*, *type*, *family*, *scientificname*, *iv100*, *substrate*, *habitat*, *state*, *city*, *location*, *latitude* e *longitude*, essa informação é valiosa, pois mais a frente algumas novas planilhas, buscas e algoritmos se utilizarão dessas colunas.

Após a criação desse *dataset* unificado foi realizado uma reordenação de todas as linhas levando em consideração a coluna de *iv100* em ordem decrescente. Em seguida, diversos *datasets* menores e mais específicos foram criados, como por exemplo:

- *datasets* para cada tipo de *substrate*, trazendo uma visualização mais específica sobre as entradas dependendo do tipo de solo;
- *datasets* para cada tipo de *habitat*, especificando ainda mais os dados de cada *habitat*;
- *datasets* para cada cidade e suas localidades, tornando possível saber quais espécies existem em cada localidade e cidade;
- *datasets* para cada tipo de *habitat* focando nas espécies destes *habitats*, trazendo dados de frequência, índices de importância e o *ranking* que aquela espécie está em importância para aquele *habitat*;

O objetivo da criação de *datasets* mais específicos, a partir da união de duas planilhas de dados com uma grande densidade de dados, foi obter um maior nível de especificidade e proporcionar que consultas sejam feitas de forma mais otimizada em uma base reduzida com apenas informações relevantes.

5.4 Análise exploratória de dados

Nesta etapa, foram criadas diversas visualizações a partir dos diferentes *datasets* gerados na Seção 5.3.3. Como os dados foram organizados para atingir um nível maior de especificidade, foi possível gerar visualizações cobrindo vários cenários, tais como:

- informações sobre tipo da lista, famílias, espécies e habitats mais populares em cada tipo de solo;
- espécies com maiores *iv100* de cada solo;
- informações sobre tipo da lista, famílias, espécies populares e quais destas tinham um maior *iv100* para cada *habitat*;
- mesma visualização do item anterior, porém aplicado à cada cidade;

Através de cada uma das visualizações geradas, foi possível obter novas informações que, anteriormente não era tão claras ao olhar as planilhas de forma isolada. Essas informações serão discutidas na próxima seção de resultados.

5.5 Construção da simulação

Nesta etapa, foi implementada a simulação do processo de restauração, seguindo o fluxo de interação que o usuário deve realizar dentro da plataforma. A simulação é realizada a partir das informações fornecidas pelo usuário, tais como tipo de solo, habitat e coordenadas de latitude e longitude do local a ser restaurado. A partir das escolhas do usuário o algoritmo cria uma nova planilha seguindo duas etapas.

5.5.1 Primeira etapa

Nesta etapa são realizadas filtragens baseadas em algumas das entradas do usuário, além do cálculo da distância de cada espécie para o ponto de restauração definido e o uso de penalizações aplicadas ao *iv100* e a frequência destas espécies.

Dado o *dataset* unificado como entrada, considerando o tipo de solo, *habitat*, latitude e longitude do ponto de restauração, os passos a seguir são realizados:

1. Filtragem da base de dados considerando o tipo de solo e o *habitat*;
2. Cálculo da distância entre o ponto de restauração e as espécies selecionadas no passo anterior, utilizando a fórmula de Haversine para cálculo de distância entre duas coordenadas na Terra, resultando em uma nova coluna na planilha chamada de *distance_to_target*;

3. Penalização do valor *iv100* de cada espécie em relação à distância até o local de restauração, criando a coluna *penalized_iv*. Penalização calculada com: $\frac{iv100}{distance_to_target}$ para cada linha do *dataset*;
4. Penalização da frequência de cada espécie em relação à distância até o local de restauração, criando a coluna *penalized_frequency*. Penalização calculada com: $\frac{1}{distance_to_target}$ para cada linha do *dataset*

É importante mencionar que a primeira etapa gera sempre um novo *dataset* contendo os seguintes dados: *list*, *type*, *scientificname*, *iv100*, *substrate*, *habitat*, *latitude*, *longitude*, *distance_to_target*, *penalized_iv*, *penalized_frequency* também é importante mencionar que há espécies que podem aparecer mais de uma vez neste *dataset* em decorrência de existirem em diversas listas.

5.5.2 Segunda etapa

Na segunda etapa, a partir do *dataset* criado na primeira etapa, mais algumas transformações são realizadas, além de uma normalização dos dados. O resultado desta etapa é o algoritmo final para realizar simulações de restauração ecológica.

Os passos realizados nessa etapa são:

1. Criação da coluna *metaFrequency*, contendo o somatório das frequências penalizadas;
2. Criação da coluna *ivSum*, contendo o somatório dos valores *iv100* penalizados;
3. Balizamento das colunas *metaFrequency* e *ivSum*, gerando as colunas *scoreFrequency* e *scoreIV*;
4. Criação da coluna *scoreAvg* que guarda o cálculo da média obtida por: $\frac{scoreFrequency + scoreIV}{2}$;
5. Criação da coluna de *ranking* em ordem crescente, onde o maior *scoreAvg* corresponde à espécie mais recomendada.

É importante mencionar que esta etapa também gera um novo *dataset*, agora específico para o *habitat* que o usuário inseriu como entrada e totalmente parametrizado pelos dados na base de dados e a latitude e longitude também fornecidas pelo usuário.

Portanto, ao executar as etapas acima para qualquer entrada válida do usuário, obtém-se uma lista de espécies ranqueadas, das mais às menos recomendadas, que poderão ser utilizadas pelos pesquisadores durante o processo de restauração ecológica.

5.6 Plataforma *Online*

Nesta etapa, foi desenvolvida a plataforma *online*, que é o foco principal desta pesquisa, pois é através dela que o usuário realizará o processo de restauração ecológica. A plataforma foi construída utilizando o *framework* React como ferramenta principal no *front-end* e Flask como principal no *back-end*, além de diversas outras bibliotecas que auxiliaram no desenvolvimento. A comunicação entre o *front-end* e o *back-end* ocorre de forma tradicional, com o site se comunicando diretamente com o servidor por meio de chamadas HTTP.

O procedimento descrito na Seção 5.5 foi implementado na funcionalidade de simulação da plataforma. No entanto, ele foi levemente adaptado para integrar-se adequadamente ao *back-end* e ao *front-end* da aplicação, garantindo uma experiência de qualidade ao usuário. Essas adaptações permitiram que o algoritmo de simulação funcionasse de maneira eficiente dentro da arquitetura da plataforma, proporcionando resultados precisos e rápidos durante a interação do usuário com a ferramenta.

Além disso, as outras funcionalidades como *Overview* e *Lists* foram construídas com base também na base de dados unificada gerada anteriormente. Essas funcionalidades foram implementadas de forma a auxiliar os pesquisadores a analisarem de forma mais dinâmica os dados coletados até então.

6 RESULTADOS

Este capítulo apresenta os resultados obtidos nesta pesquisa, abrangendo desde a análise exploratória dos dados até o desenvolvimento da plataforma *online*. Também é apresentado de forma mais detalhada os resultados referentes à construção da plataforma, destacando como os dados foram organizados e integrados para permitir que os usuários simulem cenários de restauração ecológica e consultem informações relevantes.

6.1 Análise exploratória dos dados

Nesta seção será discutido os resultados principais que foram construídos durante o processo metodológico no que diz respeito aos dados e como estes se relacionam, apresentado ao final quais foram os artefatos gerados através do pre-processamento de dados.

6.1.1 Resultados do pré-processamento dos dados

Após os procedimentos realizados na seção 5.3 novos *datasets* mais específicos foram criados facilitando a realização das análises apresentadas. Os procedimentos realizados na seção 5.3.3 geraram um total de 77 novos *datasets* a partir da base de dados unificada. Esses *datasets* estão organizados da seguinte forma:

- por tipo de solo: afloramento ferruginoso e afloramento quartzítico;
- por *habitat*: Candeal, Ephemeral Pond, Ferruginous Grasslands, Forest Islands, Peat Bog, Quartz Grasslands e Quartz Outcrop;
- por espécie x *habitat*;
- por cidade e localidade;

A Figura 8 é apresentada um exemplo de *dataset* unificado (da junção das planilhas de Lists e AfterFlora). Os demais *datasets* geradas segue a mesma estrutura apresentada no exemplo exposto, apenas com modificações de algumas filtragens. Por exemplo, no *dataset* que corresponde a cidade de Mariana, haverão somente entradas do *dataset* unificado onde *city* é Mariana, como pode ser visto na Figura 9.

	list	type	family	scientificname	iv100	substrate	habitat	state	city	location	latitude	longitude
1	L0039	Woody	Velloziaceae	Vellozia ramosissima L.B.Sm.	76.1	ironstone	Ferruginous Grassland	MG	Conceição do Mato Dentro	Fazenda Fala Melo	-18.976257	-43.401292
2	L0073	Woody	Ericaulaceae	Paepalanthus planifolius (Bong.) Körn.	66.74	quartzite	Peat Bog	MG	Conceição do Mato Dentro	PE Serra do Intendente	-19.18	-43.581083
3	L0002	Herb	Velloziaceae	Vellozia graminea Pohl	65.37	ironstone	Ephemeral Pond	MG	Mariana	Chapada do Canga	-20.159803	-43.429106
4	L0049	Woody	Asteraceae	Eremanthus incanus (Less.) Less.	62.96	ironstone	Candéal	MG	Conceição do Mato Dentro	Fazenda Água Santa	-19.086444	-43.372028
5	L0047	Woody	Asteraceae	Eremanthus incanus (Less.) Less.	54.16	ironstone	Candéal	MG	Conceição do Mato Dentro	Fazenda Água Santa	-19.089611	-43.365833
6	L0059	Woody	Velloziaceae	Vellozia scabrosa L.B.Sm. & Ayensu	54.12	ironstone	Ferruginous Grassland	MG	Conceição do Mato Dentro	Fazenda Tameiros	-19.012083	-43.397111
7	L0067	Woody	Velloziaceae	Vellozia variabilis Mart. ex Schult. & Schult.f.	52.71	quartzite	Quartz Grassland	MG	Santana do Riacho	Serra do Cipó	-19.281389	-43.594639
8	L0036	Woody	Velloziaceae	Vellozia ramosissima L.B.Sm.	52.33333333	quartzite	Quartz Grassland	MG	Conceição do Mato Dentro	Fazenda Boa Esperança	-18.93029	-43.48343
9	L0045	Woody	Asteraceae	Eremanthus incanus (Less.) Less.	48.95	ironstone	Candéal	MG	Conceição do Mato Dentro	Fazenda Tameiros	-19.004583	-43.389417
10	L0062	Herb	Velloziaceae	Vellozia graminea Pohl	47.68	ironstone	Ephemeral Pond	MG	Catas Altas	Fazenda Matadouro	-20.134917	-43.390444

Figura 8 – Resultado da junção dos *datasets* de Lists e Afterflora criados na seção 5.3.3. Fonte: elaborado pelo autor.

	list	type	family	scientificname	iv100	substrate	habitat	state	city	location	latitude	longitude
1	L0040	Herb	Asteraceae	Achyrocline satureioides (Lam.) DC.	0.11	ironstone	Ferruginous Grassland	MG	Mariana	Alegria 7	-20.175389	-43.513528
2	L0040	Herb	Orchidaceae	Acianthera teres (Lindl.) Borba	2.06	ironstone	Ferruginous Grassland	MG	Mariana	Alegria 7	-20.175389	-43.513528
3	L0003	Herb	Orchidaceae	Acianthera teres (Lindl.) Borba	2.9	ironstone	Ferruginous Grassland	MG	Mariana	Chapada do Canga	-20.163506	-43.435675
4	L0041	Herb	Orchidaceae	Acianthera teres (Lindl.) Borba	2.76	ironstone	Ferruginous Grassland	MG	Mariana	Alegria 7	-20.175889	-43.516083
5	L0040	Herb	Ericaulaceae	Actinocephalus bongardii (A.St.-Hil.) Sano	0.11	ironstone	Ferruginous Grassland	MG	Mariana	Alegria 7	-20.175389	-43.513528
6	L0041	Woody	Ericaceae	Agarista pulchella var. cordifolia (Meisn.) Judd	0.82	ironstone	Ferruginous Grassland	MG	Mariana	Alegria 7	-20.175889	-43.516083
7	L0040	Woody	Ericaceae	Agarista pulchella var. cordifolia (Meisn.) Judd	0.59	ironstone	Ferruginous Grassland	MG	Mariana	Alegria 7	-20.175389	-43.513528
8	L0041	Herb	Araceae	Anthurium minarum Sakur. & Mayo	1.31	ironstone	Ferruginous Grassland	MG	Mariana	Alegria 7	-20.175889	-43.516083
9	L0040	Herb	Araceae	Anthurium minarum Sakur. & Mayo	0.34	ironstone	Ferruginous Grassland	MG	Mariana	Alegria 7	-20.175389	-43.513528
10	L0040	Herb	Poaceae	Apochloa polioophylla (Renvoize & Zuloaga) Zuloaga & Morrone	3.74	ironstone	Ferruginous Grassland	MG	Mariana	Alegria 7	-20.175389	-43.513528

Figura 9 – Dados da cidade de Mariana, criados ao filtrar o *dataset* unificado. Fonte: elaborado pelo autor.

6.1.2 Análise quantitativa de atributos da base de dados

A seguir é apresentada a análise inicial da base de dados unificada, com foco nas características quantitativas dos dados obtidos. A Tabela 1 resume as principais métricas observadas na base de dados, apresentando uma visão geral dos elementos contidos e sua distribuição.

Tabela 1 – Análise quantitativa dos dados na base de dados unificada

Característica	Quantidade
Qntd. de listas utilizadas	101
Qntd. de dados na base unificada	3969
Qntd. de espécies	1127
Qntd. de substratos	2
Qntd. de habitats	7
Qntd. de entradas quando substrato é afloramento ferruginoso	689
Qntd. de entradas quando substrato é afloramento quartzítico	3280
Qntd. de localidades	34
Qntd. de cidades	15
Qntd. de estados	2

Fonte: Elaborado pelo autor.

A Tabela 1 destaca diversas métricas-chave, incluindo o número total de espécies documentadas, a distribuição dessas espécies em diferentes solos e *habitats*. Além disso, apresenta-se o número de localidades e municípios registrados na base de dados.

6.1.3 Análise quando solo é afloramento ferruginoso (ironstone)

Dentro da base de dados um dos dois tipos de solo presentes é o afloramento ferruginoso. Essas áreas contêm uma alta exposição de rochas e solos ricos em ferro na superfície terrestre e são facilmente caracterizados por sua coloração avermelhada ou acastanhada.

Uma breve análise sobre os dados envolvendo esse tipo de solo estão dispostos na Tabela 2:

Tabela 2 – Informações sobre afloramento ferruginoso na base de dados

Característica	Quantidade
Quantidade de habitats	3
Quantidade de espécies	287
Quantidade de cidades	7
Quantidade de localidades	12

Fonte: Elaborado pelo autor.

Além disso, é possível visualizar através da Figura 10 algumas das espécies com os maiores IVs e através da Figura 11 mais algumas informações sobre o que compõe o afloramento ferruginoso.

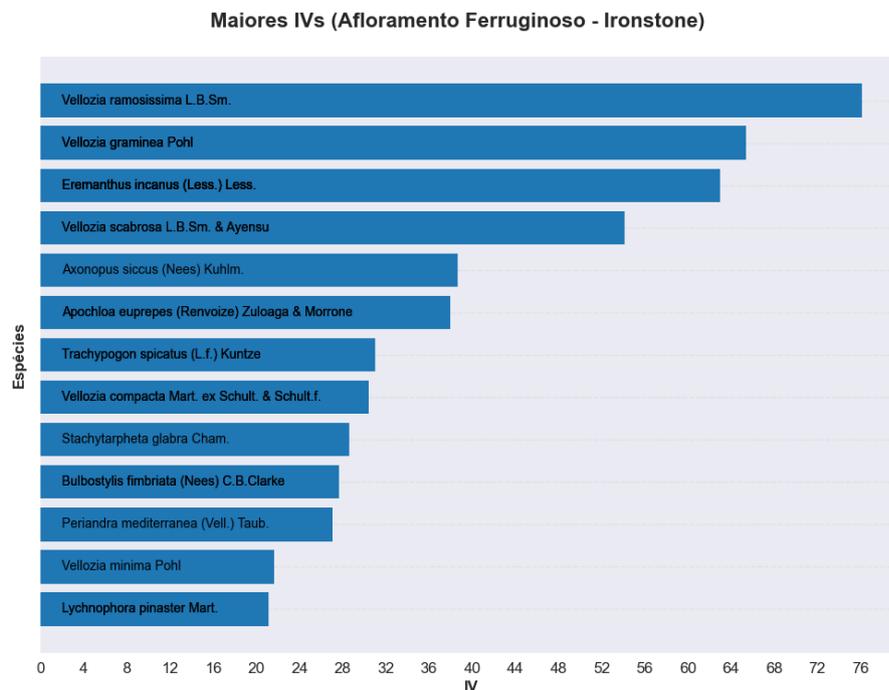


Figura 10 – Maiores IVs do solo do tipo afloramento ferruginoso. Fonte: elaborado pelo autor.

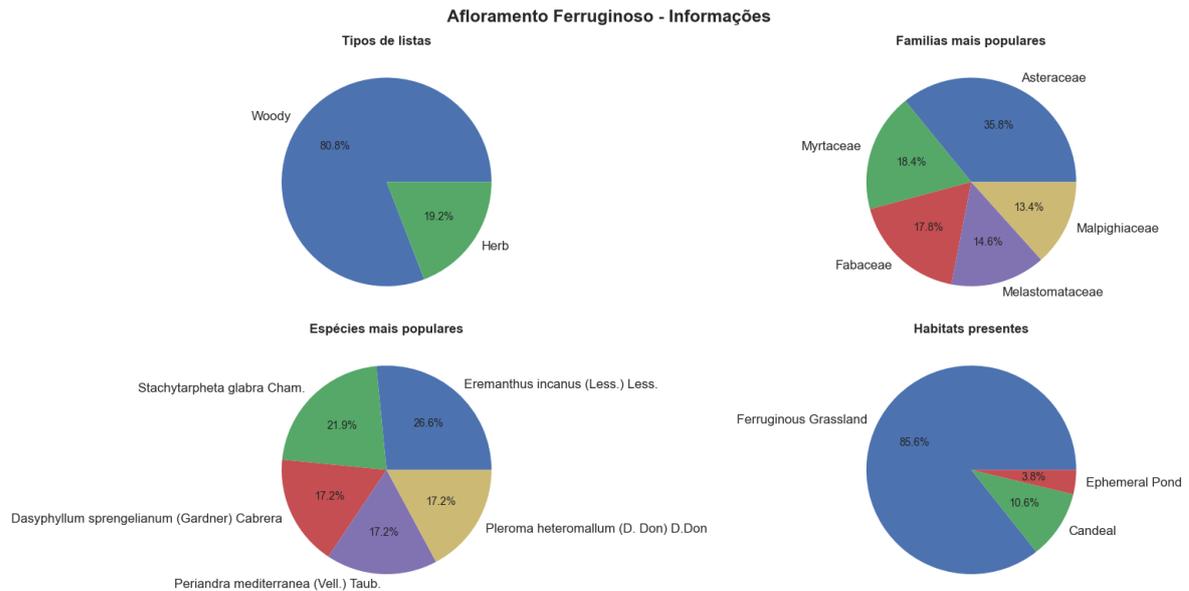


Figura 11 – Informações extras sobre afloramento ferruginoso. Fonte: elaborado pelo autor.

Ao analisar a Figura 11 percebe-se que o tipo predominante de espécie no *dataset* é do tipo lenhosa (*woody*), a família mais popular é a *Asteraceae*, a espécie mais popular é a *Eremanthus incanus* e por fim, os habitats *Ferruginous Grassland*, *Ephemeral Pond* e *Candéal* são os únicos presentes.

6.1.4 Análise quando solo é afloramento quartzítico (*quartzite*)

Um outro tipo de solo também presente na base de dados é o afloramento quartzítico, este refere-se a áreas onde há exposição de rochas quartzíticas na superfície terrestre. As rochas quartzíticas são formadas predominantemente por quartzo e resultam da metamorfose de arenitos ricos em quartzo.

Uma breve análise sobre os dados envolvendo esse tipo de solo estão dispostos na Tabela 3:

Tabela 3 – Informações sobre afloramento quartzítico na base de dados

Característica	Quantidade
Quantidade de habitats	4
Quantidade de espécies	986
Quantidade de cidades	10
Quantidade de localidades	23

Fonte: Elaborado pelo autor.

Também foi possível visualizar através da Figura 12 algumas das espécies com os

maiores IVs e através da Figura 13 mais algumas informações sobre o que compõe o afloramento quartzítico.

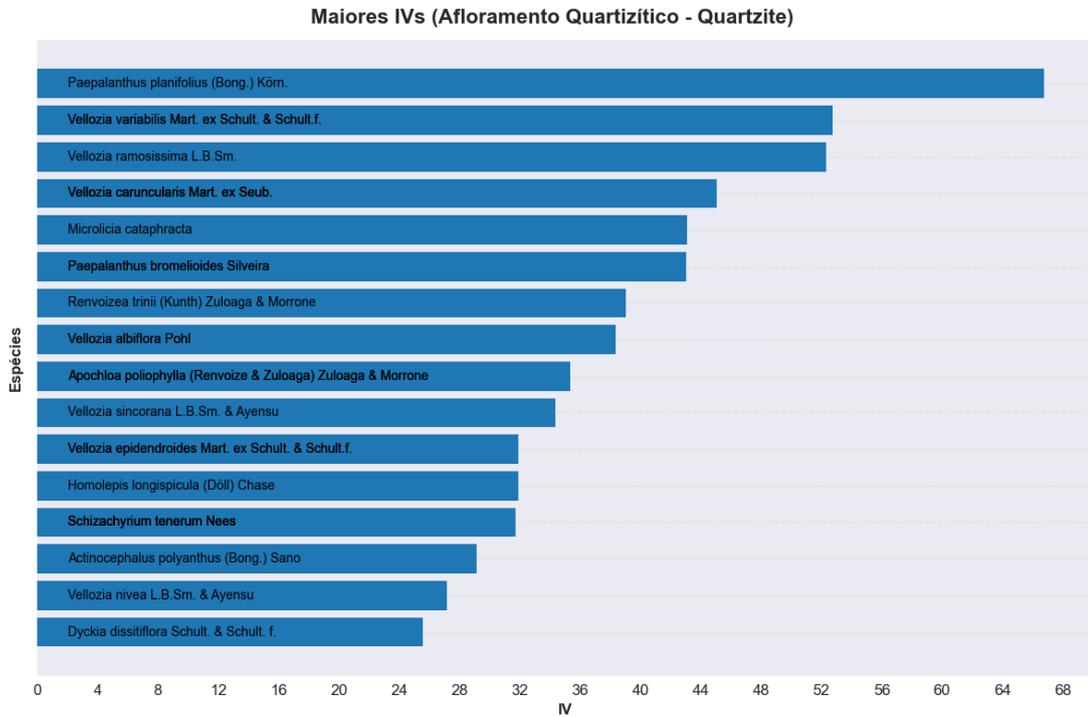


Figura 12 – Maiores IVs do solo do tipo afloramento quartzítico. Fonte: elaborado pelo autor.

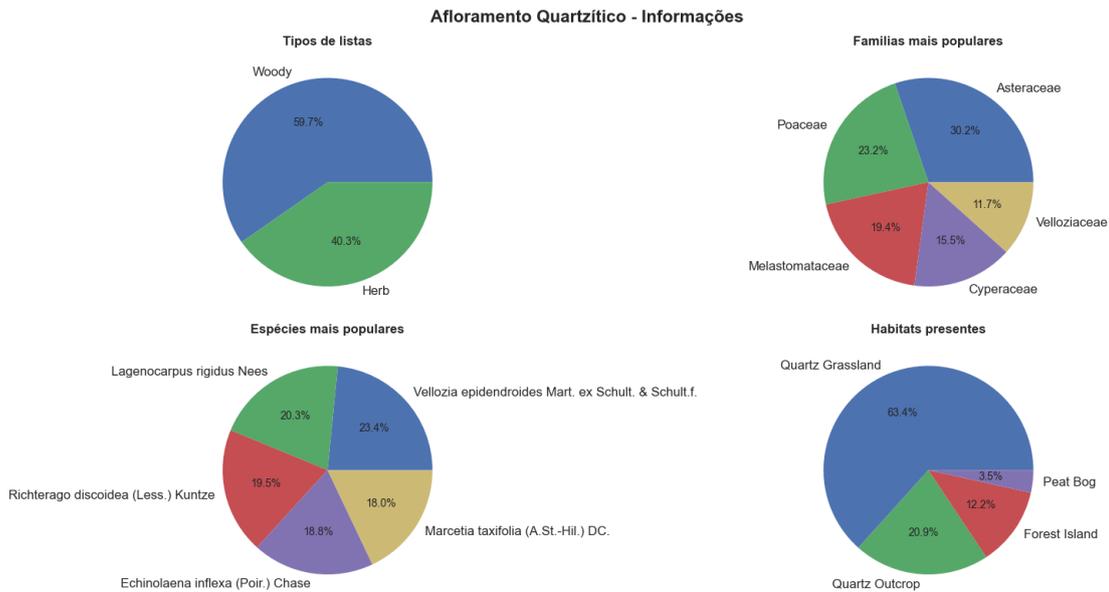


Figura 13 – Informações extras sobre afloramento quartzítico. Fonte: elaborado pelo autor.

Ao analisar a Figura 13 percebe-se que o tipo predominante de espécie no *dataset* também é do tipo lenhosa (*woody*), a família mais popular também é a *Astaraceae*, a espécie

mais popular é a *Vellozia epidendroides* e por fim, os habitats *Quartz Grassland*, *Quartz Outcrop*, *Forest Island* e *Peat Bog* são os únicos presentes.

6.1.5 Análise dos habitats

A base em questão apresenta uma diversidade notável de habitats, fornecendo uma visão abrangente da riqueza ambiental encontrada em diferentes áreas estudadas. Esses ecossistemas abarcam uma ampla gama de ambientes naturais e substratos geológicos que influenciam diretamente a flora e a fauna presentes.

De acordo com a Figura 14, os habitats mais populares na base de dados são predominantemente os associados a afloramentos quartzíticos, demonstrando uma frequência considerável de ocorrência. Dentre esses, o *Quartz Grassland* desponta como o ecossistema mais frequente, presente em 2078 registros, seguido por *Quartz Outcrop* com 687 registros.

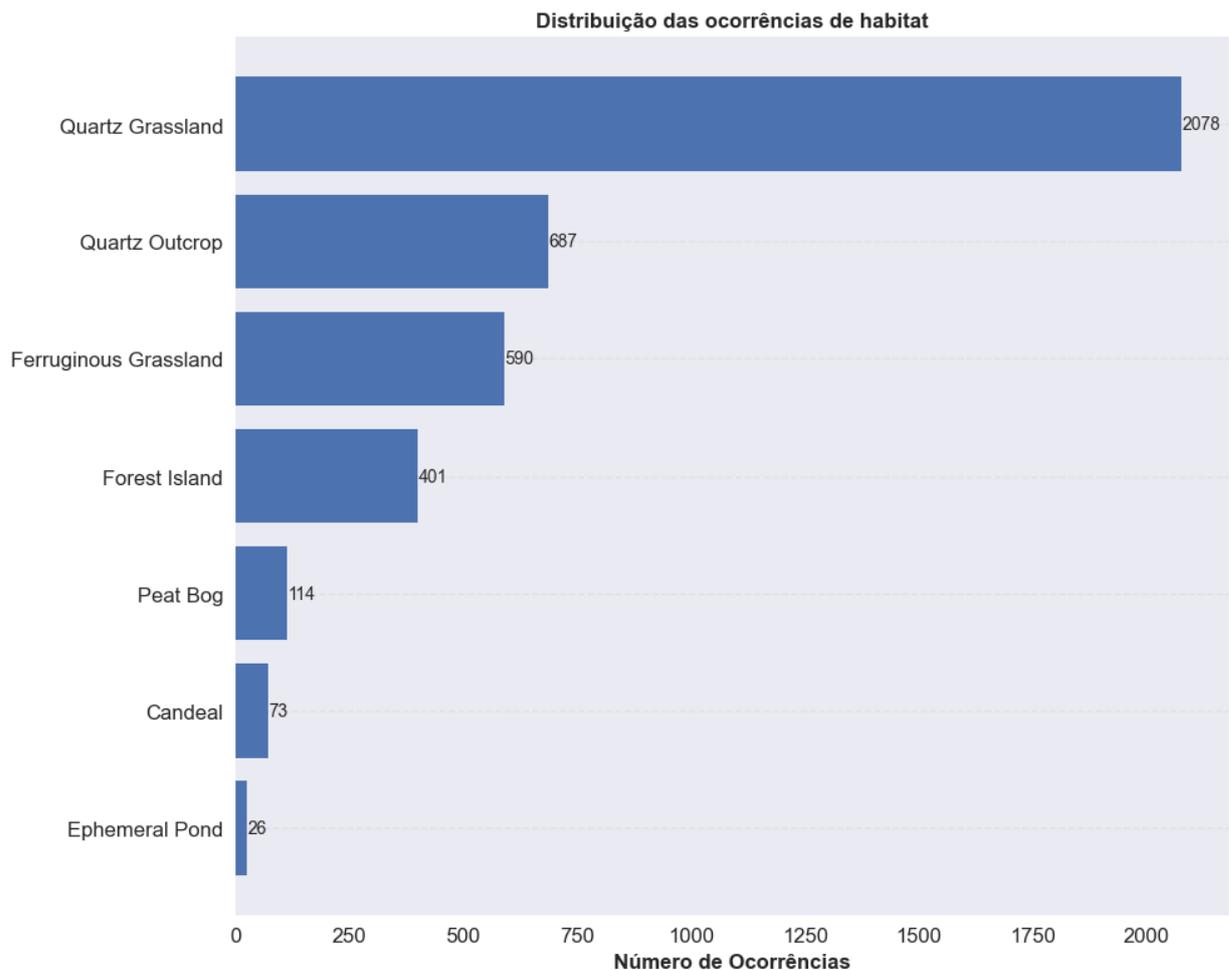


Figura 14 – Distribuição de habitats existentes na base de dados. Fonte: elaborado pelo autor.

Esses números sugerem uma prevalência notável de áreas com características de

quartzo, indicando uma representatividade significativa desses ambientes nos dados coletados, porém é importante salientar que os habitats do afloramento ferruginoso, como *Ferruginous Grassland*, *Candeal* e *Ephemeral Pond* ainda representam 17,36% das ocorrências na base de dados.

6.1.6 Análise das espécies x habitats

O objetivo desta análise é ter uma visão mais específica de como as espécies estão distribuídas entre os *habitats* na base de dados. Por tanto um algoritmo semelhante ao desenvolvido na seção 5.5 foi desenvolvido, a única diferença sendo que não haveria penalização por distância, uma vez que o propósito aqui era entender a relação das espécies com os habitats.

Como exemplo, foi aplicado o algoritmo sem a penalização por distância para o *habitat Forest Islands* e os resultados obtidos são descritos através das Figura 15.

	scientificname	metaFrequency	ivSum	scoreFrequency	scoreIV	scoreAvg	rank
1	<i>Protium spruceanum</i> (Benth.) Engl.	7	34.51	10.0	10.0	10.0	1
2	<i>Casearia decandra</i> Jacq.	7	24.92	10.0	7.221095334685598	8.6105476673428	2
3	<i>Cabralea canjerana</i> (Vell.) Mart.	7	22.31	10.0	6.464792813677195	8.232396406838598	3
4	<i>Guatteria australis</i> A.St.-Hil.	7	20.509999999999998	10.0	5.943204868154158	7.971602434077079	4
5	<i>Endlicheria paniculata</i> (Spreng.) J.F.Macbr.	7	19.02	10.0	5.5114459576934225	7.755722978846711	5
6	<i>Psidium cattleianum</i> Sabine	7	17.15	10.0	4.9695740365111565	7.484787018255578	6
7	<i>Monteverdia gonoclada</i> (Mart.) Biral	7	16.39	10.0	4.7493480150680965	7.374674007534049	7
8	<i>Myrcia amazonica</i> DC.	7	13.61	10.0	3.9437844103158506	6.971892205157925	8
9	<i>Nectandra oppositifolia</i> Nees	6	14.98	8.571428571428571	4.340770791075051	6.4560996812518106	9
10	<i>Siphoneugena densiflora</i> O.Berg	7	9.85	10.0	2.854245146334396	6.427122573167198	10

Figura 15 – Exemplo de planilha gerado para *Forest Islands*. Fonte: elaborado pelo autor.

Através desta planilha é possível verificar que, por exemplo, a espécie *Protium spruceanum* ocorreu 7 vezes dentro da planilha do *habitat Forest Islands*, totalizando o somatório de *ivSum* em 34,51. Para a normalização da coluna *scoreIV*, como esta espécie teve o maior *ivSum* sua nota começa com 10.0 e todas as demais espécies são normalizadas apartir desta espécie. Como a espécie também obteve *scoreAvg* 10.0 logo ela recebe o *ranking* de 1.

Um outro exemplo de aplicação desse algoritmo alternativo pode ser visto na Figura 16.

	scientificname	metaFrequency	ivSum	scoreFrequency	scoreIV	scoreAvg	rank
1	Vellozia epidendroides Mart. ex Schult. & Schult.f.	28	130.30120588399998	10.0	8.490542300081657	9.245271150040828	1
2	Vellozia caruncularis Mart. ex Seub.	16	153.466293764	5.7142857142857135	10.0	7.857142857142857	2
3	Echinolaena inflexa (Poir.) Chase	19	126.797188752	6.7857142857142865	8.262217431730535	7.523965858722411	3
4	Homolepis longispicula (Döll) Chase	12	146.018122515	4.285714285714286	9.514670546455381	6.900192416084833	4
5	Vellozia sincorana L.B.Sm. & Ayensu	9	153.05016999999998	3.2142857142857144	9.972885005964898	6.5935853601253065	5
6	Paspalum erianthum Nees ex Trin.	16	106.559634079	5.7142857142857135	6.943520395616453	6.328903054951083	6
7	Vellozia albiflora Pohl	16	100.809464624	5.7142857142857135	6.568834247018729	6.141559980652222	7
8	Schizachyrium tenerum Nees	14	107.638823729	5.0	7.013841351673395	6.006920675836698	8
9	Lagenocarpus rigidus Nees	18	84.634263	6.428571428571429	5.514843743483525	5.9717075860274775	9
10	Tatianyx arnatices (Trin.) Zuloaga & Soderstr.	10	119.426276434	3.5714285714285716	7.781922238746012	5.676675405087292	10

Figura 16 – Exemplo de planilha gerado para *Quartz Grasslands*. Fonte: elaborado pelo autor.

Podemos concluir que as planilhas geradas geram valor aos usuários, pois ajudam a compreender a importância de cada espécie para o ambiente em que a mesma aparece, esta etapa é fundamental para processo de recomendação, uma vez que o uso de uma espécie no *habitat* correto aumenta a chance de sucesso do processo de restauração.

6.1.7 Análise das Cidades e Localidades

Como mencionado anteriormente na Seção 6.1.2, a base de dados é composta por informações de 2 estados, 15 cidades e 34 localidades. Todas as 101 listas utilizadas para extrair as entradas da base de dados estão bem distribuídas pelos territórios de Minas Gerais e da Bahia, como pode ser visto nas Figuras 17 e 18.

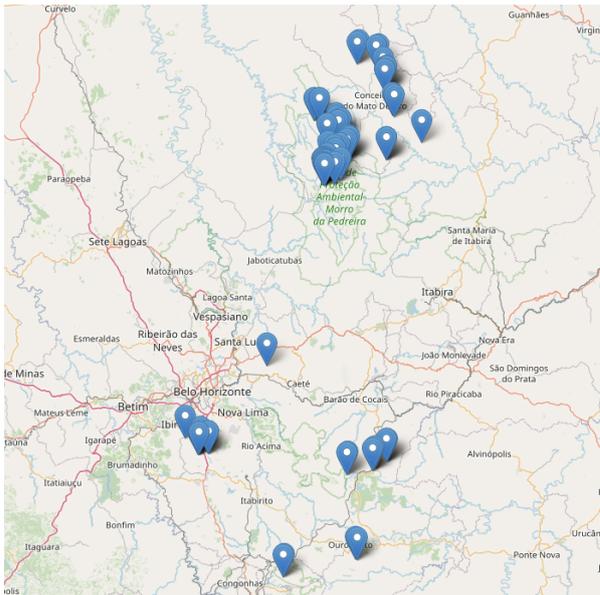


Figura 17 – Listas em Minas Gerais. Fonte: elaborado pelo autor.

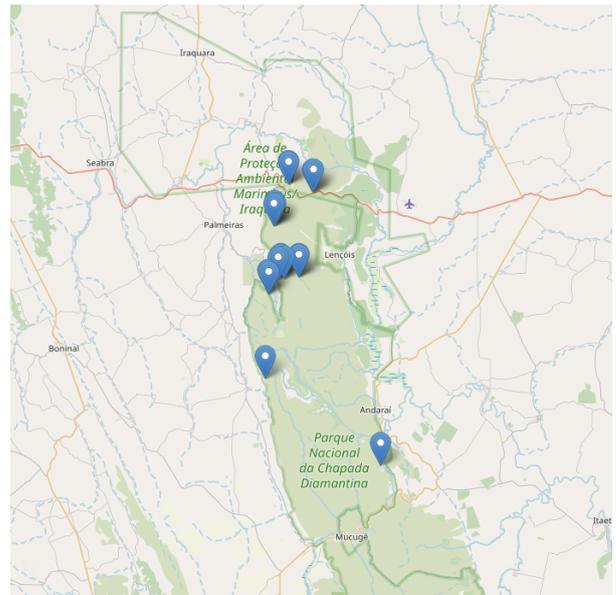


Figura 18 – Listas na Bahia. Fonte: elaborado pelo autor.

As cidades com maior representatividade na base de dados são Santana do Riacho,

responsável por 1.759 entradas, correspondendo a 44,31% dos dados, seguida por Palmeiras, Conceição do Mato Dentro, Mariana, Brumadinho, entre outras.

Em relação às localidades, as que possuem mais ocorrências são a Serra do Cipó, com 715 entradas na base de dados, seguida pela Serra do Intendente, Serra da Calçada, Serra do Ouro Branco, entre outras.

6.1.8 Análise sobre o funcionamento da simulação

Durante a seção 5.5 foi construído um algoritmo para recomendar espécies que podem ser utilizadas durante o processo de restauração ecológica. Esse algoritmo é essencial para esta pesquisa e fundamental para a funcionalidade de simulação da plataforma *online*.

Esse algoritmo, como dito anteriormente, leva em consideração a entrada do usuário na plataforma, pois, para recomendar espécies necessita-se saber algumas informações, como tipo de solo, *habitat*, latitude e longitude do local à ser restaurado.

Na Figura 19 pode ser visualizado o *dataset* que é gerado ao final da primeira etapa do algoritmo, onde são incluídas as colunas relacionadas à distância ao alvo de restauração bem como as colunas que trazem o valor de importância e frequência penalizadas pela distância.

	list	type	scientificname	iv100	substrate	habitat	latitude	longitude	distance_to_target	penalized_iv	penalized_frequency
1	L0067	Woody	Vellozia variabilis Mart. ex Schult. & Schult.f.	52.71	quartzite	Quartz Grassland	-19.281389	-43.594639	185.9789214252587	0.2834192154468598	0.005376953432875351
2	L0036	Woody	Vellozia ramosissima L.B.Sm.	52.33333333	quartzite	Quartz Grassland	-18.93029	-43.48343	157.38283306218955	0.3325225014174232	0.006353933148508337
3	L0083	Woody	Vellozia caruncularis Mart. ex Seub.	45.08	quartzite	Quartz Grassland	-19.148056	-43.550556	174.63774664983492	0.25813434303174804	0.0057261389314939674
4	L0075	Woody	Vellozia caruncularis Mart. ex Seub.	44.45	quartzite	Quartz Grassland	-19.156361	-43.550667	175.45219625785933	0.2533453610046154	0.005699558177831617
5	L0079	Woody	Paepalanthus bromelioides Silveira	43.0	quartzite	Quartz Grassland	-19.163611	-43.535	176.9274281154748	0.24303750107041228	0.005652034908614239
6	L0094	Herb	Renvoizea trinii (Kunth) Zuloaga & Morrone	39.01077	quartzite	Quartz Grassland	-12.451686	-41.472456	663.559384275725	0.0587901714969795	0.001507024124286178
7	L0020	Herb	Vellozia albiflora Pohl	38.36	quartzite	Quartz Grassland	-19.283139	-43.535806	188.75986366158858	0.2032211681863278	0.005297736396932424
8	L0020	Woody	Vellozia albiflora Pohl	38.36	quartzite	Quartz Grassland	-19.283139	-43.535806	188.75986366158858	0.2032211681863278	0.005297736396932424
9	L0064	Herb	Apochioa polioophylla (Renvoize & Zuloaga) Zuloaga & Morrone	35.35	quartzite	Quartz Grassland	-19.281611	-43.591611	186.13139767443377	0.18991959680994502	0.005372548707494909
10	L0103	Woody	Vellozia sincorana L.B.Sm. & Ayensu	34.36326	quartzite	Quartz Grassland	-12.625689	-41.503322	644.8234427351031	0.05329095954428042	0.001550812104098401

Figura 19 – *Dataset* gerado ao final da primeira etapa do algoritmo. Fonte: elaborado pelo autor.

Já, na Figura 20 pode ser visualizado o *dataset* que é gerado ao final da segunda etapa do algoritmo. Com uma estrutura diferente do que havia sido criada no *dataset* unificado. Aqui, somente as colunas *metaFrequency*, *ivSum*, *scoreFrequency*, *scoreIV*, *scoreAvg* e *rank* estão presentes.

	scientificname	metaFrequency	ivSum	scoreFrequency	scoreIV	scoreAvg	rank
1	Vellozia epidendroides Mart. ex Schult. & Schult.f.	0.15081395064742747	0.6979473794272821	10.0	8.586781651217212	9.293390825608606	1
2	Vellozia caruncularis Mart. ex Seub.	0.08511909883002211	0.81281603256832	5.643980445085837	10.0	7.821990222542919	2
3	Homolepis longispicula (Döll) Chase	0.06430932328415086	0.7832624621442817	4.264149503946956	9.636405173620217	6.950277338783586	3
4	Echinolaena inflexa (Poir.) Chase	0.08700586873145733	0.608734845142176	5.769086238902358	7.4892081449071295	6.629147191904744	4
5	Paspalum erianthum Nees ex Trin.	0.0854371086531408	0.5778401763107166	5.66506667893579	7.109113909636706	6.3870902942862475	5
6	Schizachyrium tenerum Nees	0.07629021157085557	0.605714604831001	5.058564625046303	7.452050409452137	6.25530751724922	6
7	Vellozia albiflora Pohl	0.08538825272018297	0.5342938010840369	5.661827195270776	6.573366908078645	6.117597051674711	7
8	Tatianyx arnatices (Trin.) Zuloaga & Soderstr.	0.0535598211437806	0.639534764585917	3.55138373564609	7.8681366872787635	5.709760211462426	8
9	Vellozia variabilis Mart. ex Schult. & Schult.f.	0.06014844079944498	0.508497743177278	3.98825443808311	6.256000408487723	5.122127423285416	9
10	Lagenocarpus tenuifolius (Boeckeler) C.B.Clarke	0.06913577194031191	0.44173392095172953	4.58417617491749	5.43461131734751	5.0093937461325	10

Figura 20 – Dataset gerado ao final da segunda etapa do algoritmo. Fonte: elaborado pelo autor.

6.2 Plataforma web

Conforme detalhado na Seção 5.6 da metodologia, a plataforma é essencial para a comunicação do usuário com o algoritmo de recomendação das espécies, permitindo que o usuário realize simulações de restauração ecológica com base em entradas como tipo de solo, habitat e coordenadas geográficas. Um exemplo da tela de boas vindas pode ser visto na Figura 21.

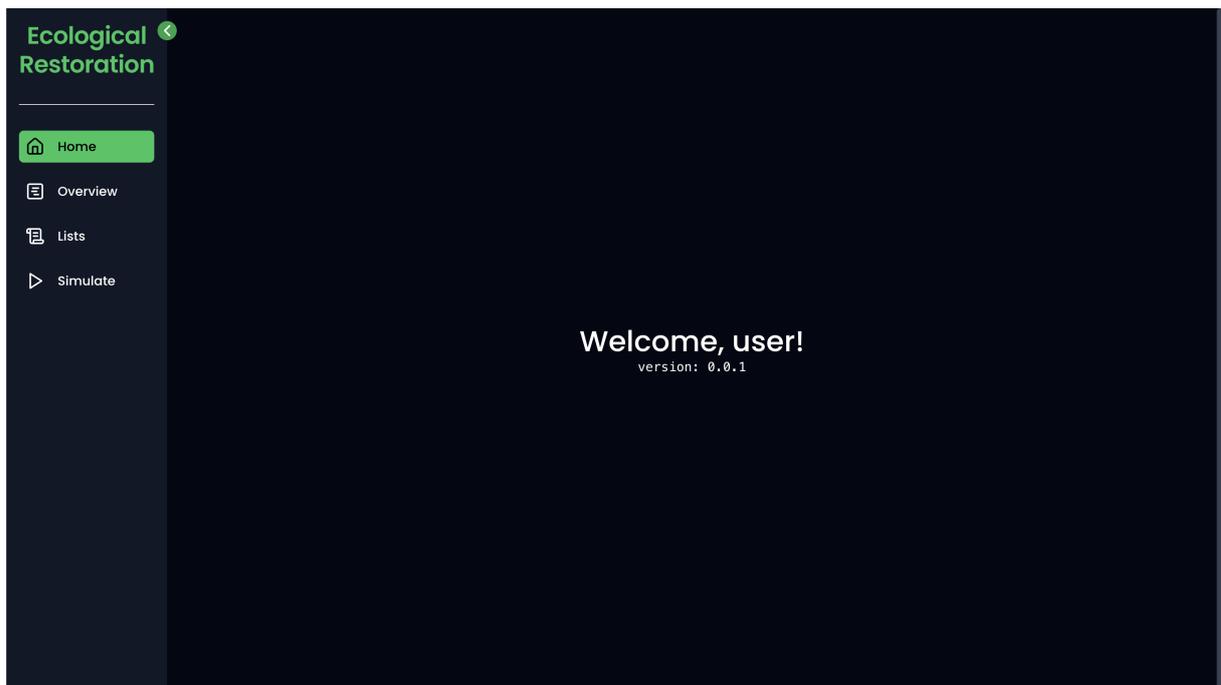


Figura 21 – Tela de boas vindas da plataforma. Fonte: elaborado pelo autor.

6.2.1 Funcionalidades

A plataforma oferece diversas funcionalidades que permitem aos usuários explorar e interagir com os dados de restauração ecológica. Entre as principais, destaca-se a possibilidade de consultar informações detalhadas sobre listas de espécies, localidades, substratos e habitats,

além de realizar filtragens avançadas para uma busca direcionada. Essas funcionalidades estão organizadas em telas intuitivas que apresentam os dados em tabelas e mapas interativos, permitindo uma compreensão visual e detalhada dos dados.

Além disso, a plataforma inclui uma funcionalidade de simulação, onde o usuário pode definir um local de restauração, escolhendo substrato, habitat e coordenadas geográficas. A partir dessas informações, o sistema gera uma lista ranqueada das espécies mais adequadas para a restauração da área, contribuindo para a tomada de decisões baseadas em dados científicos.

6.2.1.1 *Resumo*

Essa primeira tela, permite que o usuário consulte e aplique filtros na base de dados unificada gerada nas etapas anteriores, incluindo informações sobre listas de espécies, substratos, *habitats*, cidades e localidades.

Os dados são organizados em uma tabela, na qual o usuário pode visualizar campos como código da lista, estado, cidade, local, latitude, longitude, substrato, habitat, nome científico da espécie e o valor de importância (IV).

Há também nesta tela, na parte superior indicadores quantitativos disponíveis da base de dados. Estes indicadores são fixos e representam algumas informações principais referentes a base de dados.

A Figura 22 demonstra como a tela fica quando nenhum filtro é aplicado e a figura 23 demonstra como a listagem fica quando os filtros listas, substratos e *habitats* são aplicados. Cada filtro também possui a possibilidade de múltipla seleção e a tabela contém paginação para quando mais de uma página é necessária, tendo em vista que cada página mostra 10 entradas da base de dados.

List	State	City	Location	Latitude	Longitude	Substrate	Habitat	Scientific Name	IV
L0039	MG	Conceição do Mato Dentro	Fazenda Faia Melo	-18.976257	-43.401292	ironstone	Ferruginous Grassland	Vellozia ramosissima L.B.Sm.	76.1
L0073	MG	Conceição do Mato Dentro	PE Serra do Intendente	-19.18	-43.581083	quartzite	Peat Bog	Paepalanthus planifolius (Bong.) Körn.	66.74
L0002	MG	Mariana	Chapada do Canga	-20.159803	-43.429106	ironstone	Ephemeral Pond	Vellozia graminea Pohl	65.37
L0049	MG	Conceição do Mato Dentro	Fazenda Água Santa	-19.086444	-43.372028	ironstone	Candead	Eremanthus incanus (Less.) Less.	62.96
L0047	MG	Conceição do Mato Dentro	Fazenda Água Santa	-19.089611	-43.365833	ironstone	Candead	Eremanthus incanus (Less.) Less.	54.16
L0059	MG	Conceição do Mato Dentro	Fazenda Tamelos	-19.012083	-43.397111	ironstone	Ferruginous Grassland	Vellozia scabrosa L.B.Sm. & Ayensu	54.12
L0067	MG	Santana do Riacho	Serra do Cipó	-19.281389	-43.594639	quartzite	Quartz Grassland	Vellozia variabilis Mart. ex Schult. & Schult.f.	52.71
L0036	MG	Conceição do Mato Dentro	Fazenda Boa Esperança	-18.93029	-43.48343	quartzite	Quartz Grassland	Vellozia ramosissima L.B.Sm.	52.33333333
L0045	MG	Conceição do Mato Dentro	Fazenda Tamelos	-19.004583	-43.389417	ironstone	Candead	Eremanthus incanus (Less.) Less.	48.95

Figura 22 – Tela de resumo sem filtros aplicados. Fonte: elaborado pelo autor.

List	State	City	Location	Latitude	Longitude	Substrate	Habitat	Scientific Name	IV
L0039	MG	Conceição do Mato Dentro	Fazenda Faia Melo	-18.976257	-43.401292	ironstone	Ferruginous Grassland	Vellozia ramosissima L.B.Sm.	76.1
L0039	MG	Conceição do Mato Dentro	Fazenda Faia Melo	-18.976257	-43.401292	ironstone	Ferruginous Grassland	Eremanthus incanus (Less.) Less.	16.26666667
L0039	MG	Conceição do Mato Dentro	Fazenda Faia Melo	-18.976257	-43.401292	ironstone	Ferruginous Grassland	Kielmeyera lathrophyton Saddi	2.916666667
L0039	MG	Conceição do Mato Dentro	Fazenda Faia Melo	-18.976257	-43.401292	ironstone	Ferruginous Grassland	Micropholis gnaphalocladus (Mart.) Pierre	2.38
L0039	MG	Conceição do Mato Dentro	Fazenda Faia Melo	-18.976257	-43.401292	ironstone	Ferruginous Grassland	Humiria balsamifera (Aubl.) A.St.-Hil.	1.296666667
L0039	MG	Conceição do Mato Dentro	Fazenda Faia Melo	-18.976257	-43.401292	ironstone	Ferruginous Grassland	Ternstroemia brasiliensis Cambess.	1

Figura 23 – Tela de resumo com filtros aplicados. Fonte: elaborado pelo autor.

6.2.1.2 Listas

Nesta tela, o usuário visualiza um mapa que apresenta a localização geográfica de todas as listas de espécies registradas na base de dados (101 listas no total), conforme pode ser visto na figura 24. Cada marcador no mapa corresponde a uma lista específica, permitindo que o

usuário interaja com essas informações de forma visual.

Ao clicar em um marcador, uma janela *pop-up* (figura 25) exibe informações detalhadas sobre a lista selecionada, incluindo a cidade, local, coordenadas, altitude, substrato, *habitat* e a opção de visualizar as espécies associadas àquela lista.

A partir da opção “*See species*”, o usuário pode consultar uma lista completa das espécies presentes na lista correspondente (figura 26). Essa visualização permite uma compreensão espacial dos dados de restauração ecológica, facilitando a exploração das informações.

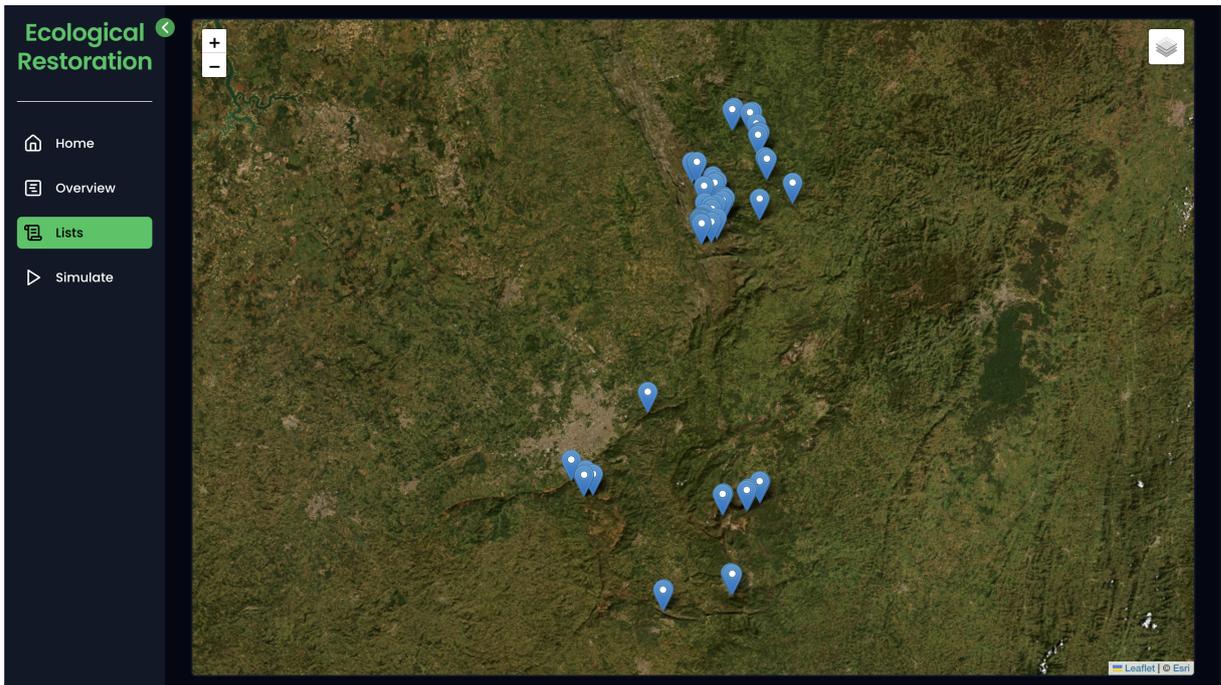


Figura 24 – Mapa contendo as espécies. Fonte: elaborado pelo autor.

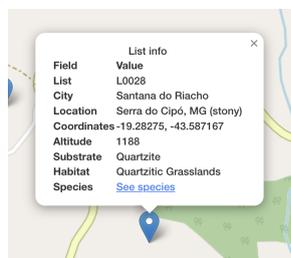


Figura 25 – *Pop-up* contendo informações sobre a lista selecionada. Fonte: elaborado pelo autor.

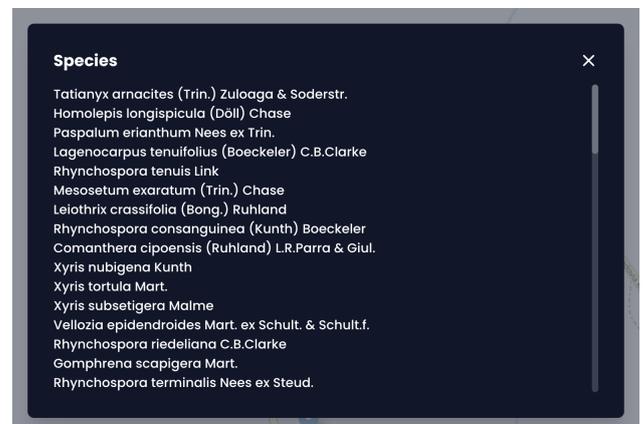
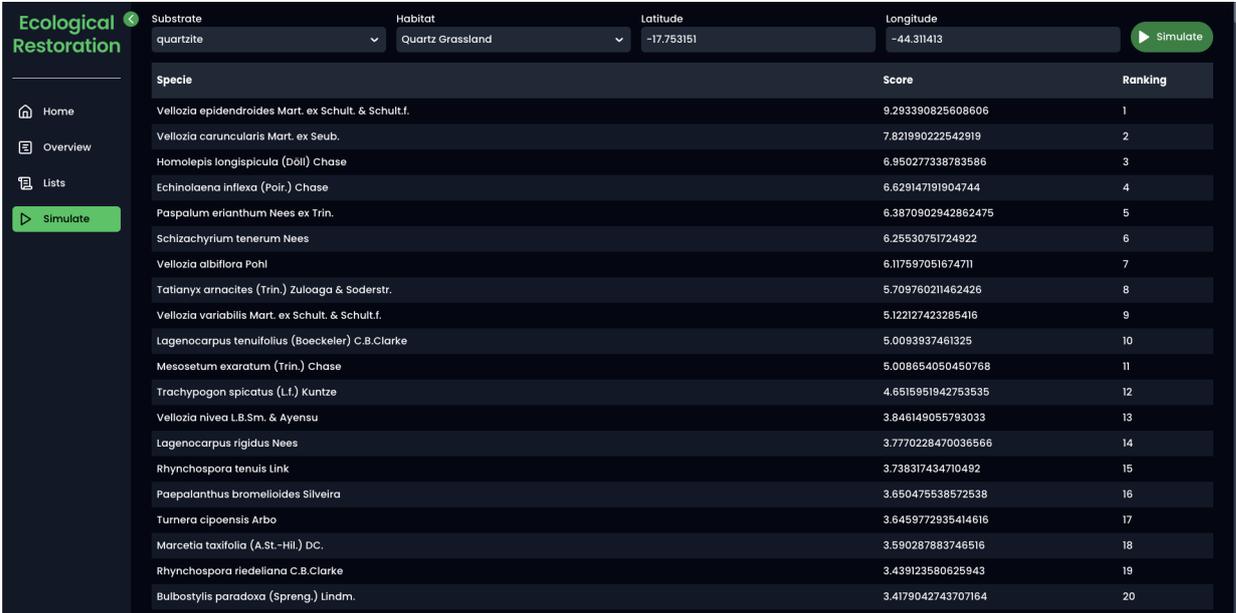


Figura 26 – *Modal* contendo a listagem das espécies atreladas àquela lista. Fonte: elaborado pelo autor.

6.2.1.3 Simulação

Nesta tela, o usuário pode realizar uma simulação para determinar as espécies mais adequadas para a restauração ecológica de uma área específica. A simulação é realizada com base na seleção de um substrato, um tipo de habitat, e na inserção da latitude e longitude do local a ser restaurado.

Após a configuração, o usuário pode realizar a simulação, e o servidor retorna uma lista ranqueada de espécies que podem ser plantadas no local, como pode ser visto na figura 27. A lista inclui o nome científico das espécies, uma pontuação (*score*) atribuída a cada uma, e sua posição no *ranking* de adequação.



Specie	Score	Ranking
Vellozia epidendroides Mart. ex Schult. & Schult.f.	9.293390825608606	1
Vellozia caruncularis Mart. ex Seub.	7.821990222542919	2
Homolepis longispicula (Doll) Chase	6.950277338783586	3
Echinolaena inflexa (Poir.) Chase	6.629147191904744	4
Paspalum erianthum Nees ex Trin.	6.3870902942862475	5
Schizachyrium tenerum Nees	6.25530751724922	6
Vellozia albiflora Pohl	6.117597051674711	7
Tatianyx arnaces (Trin.) Zuloaga & Soderstr.	5.709760211462426	8
Vellozia variabilis Mart. ex Schult. & Schult.f.	5.122127423285416	9
Lagenocarpus tenuifolius (Boeckeler) C.B. Clarke	5.0093937461325	10
Mesosetum exaratum (Trin.) Chase	5.008654050450768	11
Trachypogon spicatus (Lf.) Kuntze	4.6515951942753535	12
Vellozia nivea L.B.Sm. & Ayensu	3.846149055793033	13
Lagenocarpus rigidus Nees	3.7770228470036566	14
Rhynchospora tenuis Link	3.738317434710492	15
Paepalanthus bromelioides Silveira	3.650475538572538	16
Turnera cipoensis Arbo	3.6459772935414616	17
Marcetia taxifolia (A.St.-Hill.) DC.	3.590287883746516	18
Rhynchospora riedeliana C.B. Clarke	3.439123580625943	19
Bulbostylis paradoxa (Spreng.) Lindm.	3.4179042743707164	20

Figura 27 – Simulação de restauração. Fonte: elaborado pelo autor.

Essa funcionalidade é o produto principal decorrente desta pesquisa e é uma ferramenta valiosa para auxiliar na tomada de decisões em projetos de restauração ecológica, facilitando a escolha das espécies com maior potencial de sucesso para a recuperação do ecossistema.

7 CONCLUSÃO

Este trabalho apresentou uma abordagem interdisciplinar entre computação e biologia, focada na restauração ecológica. A partir de uma extensa base de dados composta por informações de espécies, habitats, tipos de solo, cidades e localidades, foi realizada uma análise exploratória detalhada que permitiu evidenciar novas informações significativas e identificar padrões não evidentes em análises isoladas.

Além da análise exploratória foi também desenvolvida uma plataforma *online* da qual os usuários podem interagir, de forma dinâmica com os dados cadastrados na base de dados. Também é possível realizar simulações de seleção de espécies de plantas para a restauração dentro da plataforma, o que garante que os biólogos responsáveis pela execução da restauração tenham consigo uma ferramenta inteligente e poderosa capaz de auxiliar na tomada de decisões.

Em suma, este trabalho contribuiu para a área de restauração ecológica ao fornecer uma ferramenta prática que auxilia na tomada de decisões, permitindo que profissionais selecionem espécies de forma mais informada e eficiente. A plataforma desenvolvida não apenas facilita o acesso a dados relevantes, mas também promove a interação e a visualização de informações cruciais para o sucesso de projetos de restauração.

Como perspectivas futuras, sugere-se a expansão da base de dados para incluir novos estados, cidades e diferentes tipos de *habitats*, bem como a incorporação de fatores adicionais no algoritmo de simulação, como condições climáticas e interações entre espécies. A melhoria contínua da plataforma, incluindo a adição de visualizações mais interativas e funcionalidades avançadas, poderá aumentar ainda mais seu impacto e utilidade na comunidade científica e profissional.

REFERÊNCIAS

- ALBRIGHT, S. C.; WINSTON, W. L.; ZAPPE, C. J.; BROADIE, M. N. **Data analysis and decision making**. [S.l.]: Citeseer, 2011. v. 577.
- ATHAYDE, M. P. Sistema de recomendação de plantio utilizando aprendizado de máquina. Universidade Federal de Uberlândia, 2023.
- BHANDARI, P. **Data Collection | Definition, Methods and Examples**. 2020. Disponível em: <https://www.scribbr.com/methodology/data-collection>. Acesso em: 13 out. 2023.
- FAMILI, A.; SHEN, W.-M.; WEBER, R.; SIMOUDIS, E. Data preprocessing and intelligent data analysis. **Intelligent Data Analysis**, IOS Press, v. 1, n. 1, p. 3–23, 1997. ISSN 1571-4128.
- FAWZY, S.; OSMAN, A. I.; DORAN, J.; ROONEY, D. W. **Strategies for mitigation of climate change: a review**. 2020. 2069–2094 p. Disponível em: <https://doi.org/10.1007/s10311-020-01059-w>. Acesso em: 24 set. 2023.
- GANN, G. D.; MCDONALD, T.; WALDER, B.; ARONSON, J.; NELSON, C. R.; JONSON, J.; HALLETT, J. G.; EISENBERG, C.; GUARIGUATA, M. R.; LIU, J.; HUA, F.; ECHEVERRIA, C.; GONZALES, E. K.; SHAW, N.; DECLEER, K.; DIXON, K. W. **International Principles and Standards for the Practice of Ecological Restoration**. 2nd. ed. [S.l.: s.n.], 2019. S1–S46 p. Restoration Ecology.
- HAND, D. J.; MANNILA, H.; SMYTH, P. **Principles of data mining (adaptive computation and machine learning)**. [S.l.]: MIT Press, 2001.
- HARRIS, J. A.; HOBBS, R. J.; HIGGS, E.; ARONSON, J. Ecological restoration and global climate change. **Restoration Ecology**, v. 14, n. 2, p. 170–176, 2006.
- HEER, J.; SHNEIDERMAN, B. Interactive dynamics for visual analysis. **Communications of the ACM**, v. 55, p. 45–54, 04 2012.
- LADOUCEUR, E.; SHACKELFORD, N. The power of data synthesis to shape the future of the restoration community and capacity. **Restoration Ecology**, v. 29, n. 1, p. e13251, 2020.
- MILLER, K. **17 Data Visualization Techniques All Professionals Should Know**. 2019. Disponível em: <https://online.hbs.edu/blog/post/data-visualization-techniques>. Acesso em: 24 set. 2023.
- OLIVEIRA, R. E. d.; ENGEL, V. L. A restauração ecológica em destaque: um retrato dos últimos vinte e oito anos de publicações na área. **Oecologia Australis**, v. 15, n. 2, p. 303–315, 2011.
- O'NEIL, C.; SCHUTT, R. **Doing data science: Straight talk from the frontline**. [S.l.]: "O'Reilly Media, Inc.", 2013.
- OVERBECK, G.; HERMANN, J.-M.; ANDRADE, B.; BOLDRINI, I.; KIEHL, K.; KIRMER, A.; KOCH, C.; KOLLMANN, J.; MEYER, S.; MÜLLER, S.; NABINGER, C.; PILGER, G.; TRINDADE, J. P.; VÉLEZ-MARTIN, E.; WALKER, E.; ZIMMERMANN, D. G.; PILLAR, V. Restoration ecology in brazil - time to step out of the forest. **Natureza & conservação revista brasileira de conservação da natureza**, v. 11, p. 92–95, 07 2013.

PUDUMALAR, S.; RAMANUJAM, E.; RAJASHREE, R. H.; KAVYA, C.; KIRUTHIKA, T.; NISHA, J. Crop recommendation system for precision agriculture. In: **2016 Eighth International Conference on Advanced Computing (ICoAC)**. [S.l.: s.n.], 2017. p. 32–36.

RASCHKA, S. **About Feature Scaling and Normalization – and the effect of standardization for machine learning algorithms**. 2014. Disponível em: https://sebastianraschka.com/Articles/2014_about_feature_scaling.html. Acesso em: 15 ago. de 2023.

RIDZUAN, F.; Wan Zainon, W. M. N. **A Review on Data Cleansing Methods for Big Data**. 2019. 731-738 p. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050919318885>. Acesso em: 16 ago. de 2023.

RUCHIRAWYA, T. Crop recommendation system. In: . [S.l.: s.n.], 2020.

RUIZ-JAEN, M. C.; AIDE, T. M. Restoration success: How is it being measured? **Restoration Ecology**, v. 13, n. 3, p. 569–577, 2005.

SADIKU, M.; SHADARE, A. E.; MUSA, S. M.; AKUJUOBI, C. M.; PERRY, R. Data visualization. **International Journal of Engineering Research And Advanced Technology (IJERAT)**, v. 2, n. 12, p. 11–16, 2016.

SEGEL, E.; HEER, J. Narrative visualization: Telling stories with data. **IEEE transactions on visualization and computer graphics**, v. 16, p. 1139–48, 01 2011.

SHIREEN, M. **Data Collection: The First Step to Building Intelligent Applications**. 2022. Disponível em: <https://indiaai.gov.in/article/data-collection-for-machine-learning-and-ai/>. Acesso em: 10 set. de 2023.

Society for Ecological Restoration. **The SER International Primer on Ecological Restoration**. [S.l.]: SER International, 2004. Version 2.

STATISTA. **Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025**. 2021. Disponível em: <https://www.statista.com/statistics/871513/worldwide-data-created>. Acesso em: 28 maio de 2023.

TOMA, T. S. P.; OVERBECK, G. E.; MENDONÇA, M. de S.; FERNANDES, G. **Optimal references for ecological restoration: the need to protect references in the tropics**. 2023. 25-32 p. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2530064423000032>. Acesso em: 19 out. de 2023.

TUKEY, J. W. The future of data analysis. **The annals of mathematical statistics**, JSTOR, v. 33, n. 1, p. 1–67, 1962.

TUKEY, J. W. *et al.* **Exploratory data analysis**. [S.l.]: Reading, MA, 1977. v. 2.

UNEP-WCMC. **10 years to boost ecosystem restoration for people and planet**. 2020. Disponível em: <https://www.unep-wcmc.org/en/news/10-years-to-boost-ecosystem-restoration-for-people-and-planet>. Acesso em: 13 out. 2023.