

Análise e Clusterização da Flora Brasileira Ameaçada de Extinção

Juan P. F. M. Macedo¹, Matheus D. C. Sousa¹

¹Bacharelado Interdisciplinar em Ciência e Tecnologia

Universidade Federal Maranhão (UFMA) – São Luís, MA – Brazil

juan.mondego@discente.ufma.br, Matheus.donavan@discente.ufma.br

Abstract. *This study analyzed open data from the Ministry of Environment on endangered Brazilian flora up to 2021, aiming to compare threat categories between 2014 and 2021 and identify species clusters using the K-Means algorithm. Data acquisition, processing, exploratory analysis, and normalization were performed, including the conversion of categorical variables to numerical form. The results revealed significant variations in threat statuses and enabled segmentation into representative clusters, highlighting patterns that can support conservation management strategies and public policy formulation. The findings underscore the importance of computational methods for monitoring and preserving biodiversity.*

Resumo. *Este estudo analisou dados abertos do Ministério do Meio Ambiente sobre a flora brasileira ameaçada de extinção até 2021, com o objetivo de comparar as categorias de ameaça entre 2014 e 2021 e identificar grupos de espécies por meio do algoritmo K-Means. Foram realizadas a aquisição, tratamento, análise exploratória e normalização dos dados, incluindo a conversão de variáveis categóricas em numéricas. Os resultados revelaram variações significativas nos status de ameaça e possibilitaram a segmentação dos dados em clusters representativos, evidenciando padrões que podem subsidiar estratégias de manejo e políticas públicas de conservação. Os achados reforçam a importância da utilização de métodos computacionais para o monitoramento e a preservação da biodiversidade.*

Palavras-Chave: Biodiversidade; Conservação; Clusterização; Flora Ameaçada; K-Means.

1. Introdução

A conservação da biodiversidade é um dos maiores desafios contemporâneos, especialmente em um país extenso e ecologicamente diverso como o Brasil, que, apesar de abrigar uma das maiores riquezas biológicas do planeta, enfrenta sérias ameaças à integridade de seus ecossistemas. O acelerado declínio de espécies e a degradação ambiental impõem a necessidade de monitoramento constante e de políticas públicas efetivas para a preservação dos recursos naturais. Nesse contexto, a disponibilização de dados abertos pelo Ministério do Meio Ambiente (MMA) revela-se fundamental para a realização de análises que possam orientar ações de conservação.

A crescente aplicação de métodos computacionais na área ambiental tem possibilitado a análise de grandes volumes de dados, permitindo a identificação de padrões e tendências que, de outra forma, seriam difíceis de detectar. Em particular, técnicas de clusterização têm se destacado por sua capacidade de agrupar espécies com características semelhantes, contribuindo para a definição de áreas ou grupos prioritários para intervenção. Assim, este estudo propõe comparar as categorias de ameaça da flora brasileira entre 2014 e 2021 e identificar, por meio do algoritmo K-Means, grupos de espécies com padrões de variação similares, fornecendo subsídios quantitativos para futuras estratégias de conservação.

2. Objetivos

Este estudo tem como objetivos principais:

1. Analisar os dados referentes à flora brasileira ameaçada de extinção, com ênfase na comparação das categorias de ameaça registradas em 2014 e 2021;
2. Aplicar técnicas de clusterização, com destaque para o algoritmo K-Means, para identificar grupos de espécies que apresentem padrões de variação semelhantes em seu status de ameaça;
3. Integrar análises estatísticas e visualizações gráficas geradas por meio de códigos implementados no Google Colab, de forma a subsidiar futuras ações de conservação e a formulação de políticas públicas ambientais.

3. Fundamentação teórica

A efetividade das políticas de conservação depende diretamente da capacidade de monitorar e analisar dados ambientais de forma precisa. A disponibilização de bases de dados abertas pelo Ministério do Meio Ambiente (MMA) tem viabilizado a realização de análises quantitativas detalhadas, essenciais para identificar padrões e tendências de ameaça entre as espécies. Trabalhos correlatos, como o estudo “Geoprocessamento Aplicado à Conservação de Espécies Ameaçadas de Extinção: O Projeto Mico-Leão-da-Cara-Preta” (BRASIL, 1996), demonstraram que o uso de técnicas automatizadas pode facilitar a identificação de áreas prioritárias para intervenção, ainda que seu foco estivesse direcionado à fauna.

De forma semelhante, o artigo “Espécies da fauna silvestre ameaçadas de extinção no Brasil: os planos de ação nacionais e suas contribuições para as metas globais de biodiversidade” (VERCILLO et al., 2022) destaca a importância da análise

quantitativa de dados como subsídio para a formulação de políticas públicas integradas de conservação.

Nesse cenário, a aplicação de métodos computacionais, com destaque para a clusterização, tem se consolidado como uma abordagem relevante. O algoritmo K-Means é amplamente utilizado devido à sua simplicidade e eficácia na segmentação de grandes conjuntos de dados em grupos homogêneos, facilitando a interpretação de padrões de ameaça (SILVA, 2003). A técnica envolve a conversão de variáveis categóricas em numéricas e a subsequente normalização dos dados, garantindo comparações consistentes e resultados confiáveis.

A integração dessas técnicas com bases de dados abertas permite a identificação não apenas de grupos de espécies com status de ameaça semelhante, mas também a geração de informações estratégicas que orientem políticas públicas mais eficazes para a conservação ambiental. Dessa forma, os estudos correlatos reforçam a relevância de abordagens quantitativas e computacionais no monitoramento da biodiversidade, servindo de base para a análise da flora ameaçada de extinção.

4. Metodologia

Os dados analisados foram extraídos da base de dados abertos do Ministério do Meio Ambiente (MMA), contendo informações sobre a flora brasileira ameaçada de extinção até o ano de 2021. As etapas metodológicas foram organizadas em cinco fases: aquisição e preparação dos dados; análise exploratória; seleção de variáveis; preparação para a clusterização; e aplicação do algoritmo K-Means.

4.1. Aquisição e Preparação dos Dados

Os dados foram extraídos de um arquivo CSV contendo informações sobre espécies da flora brasileira ameaçadas de extinção. A importação e manipulação dos dados foram realizadas com a biblioteca pandas no ambiente Google Colab. Foram efetuados os ajustes necessários no delimitador e na codificação, a fim de assegurar a correta interpretação das informações.

Posteriormente, realizou-se uma análise preliminar da estrutura do conjunto de dados, com a identificação e tratamento de valores nulos, etapa fundamental para garantir a integridade dos dados que embasariam as análises subsequentes.

Família (FB 2020)	0
Espécie (FB 2020)	0
Nome avaliação anterior	0
Categoria em 2014	1243
Sugestão de Categoria 2021	0
EW	3212
CR	2525
EN,	1364
VU	2535
Mesma Categoria de Ameaça 2014	1367
Aumentou a Categoria de Ameaça	3142
Reduziu a Categoria de Ameaça	3158
Entrou em Categoria de Ameaça	1969
Saíram da Categoria de Ameaça	3211
Saíram da Lista por serem sinonímias taxonômicas	3211

Figura 1. Plotagem de tabela simples no Google Colab

4.2. Análise Exploratória

A análise exploratória teve como objetivo familiarizar o pesquisador com as características e a distribuição dos dados, possibilitando a identificação de padrões e tendências. Para tanto, foram geradas as seguintes visualizações:

4.2.1. Gráfico de Barras Comparativo:

Este gráfico permite a comparação das categorias de ameaça registradas em 2014 e 2021, facilitando a identificação de variações e tendências ao longo do tempo.

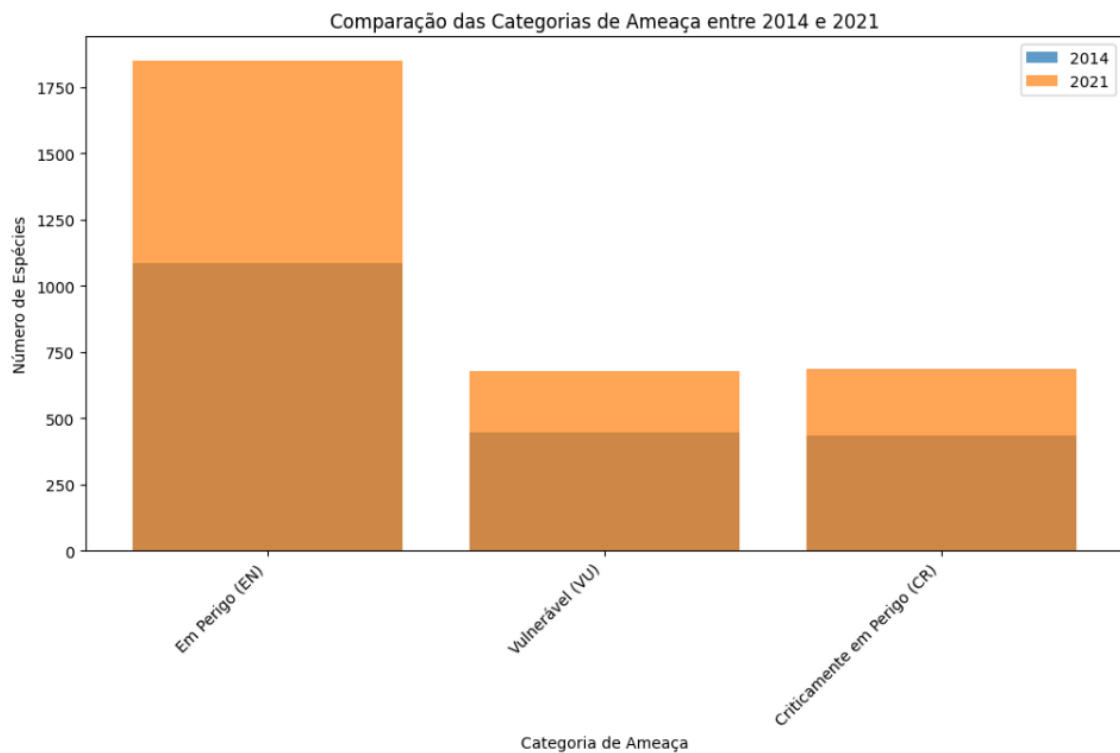


Figura 2. Comparação entre as categorias de ameaça entre 2014 e 2021 (produzido pelo autor com o google colab).

4.2.2. Gráfico de Colunas das Famílias Botânicas:

Com o intuito de evidenciar quais famílias abrigam maior número de espécies ameaçadas, foi elaborado um gráfico de colunas que apresenta o ranking das principais famílias.

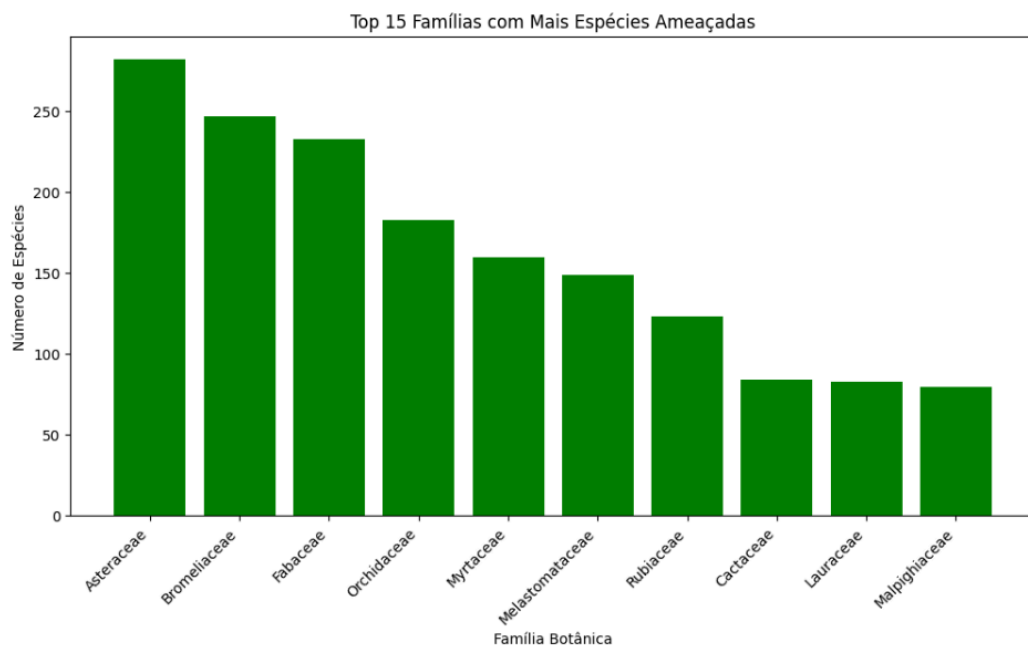


Figura 3. Top 15 famílias com mais espécies ameaçadas (produzido pelo autor com o google colab).

4.2.3. Gráfico de Pizza:

Para ilustrar as alterações no status de ameaça (por exemplo, entrada na lista, manutenção, aumento ou redução da categoria), desenvolveu-se um gráfico de pizza.

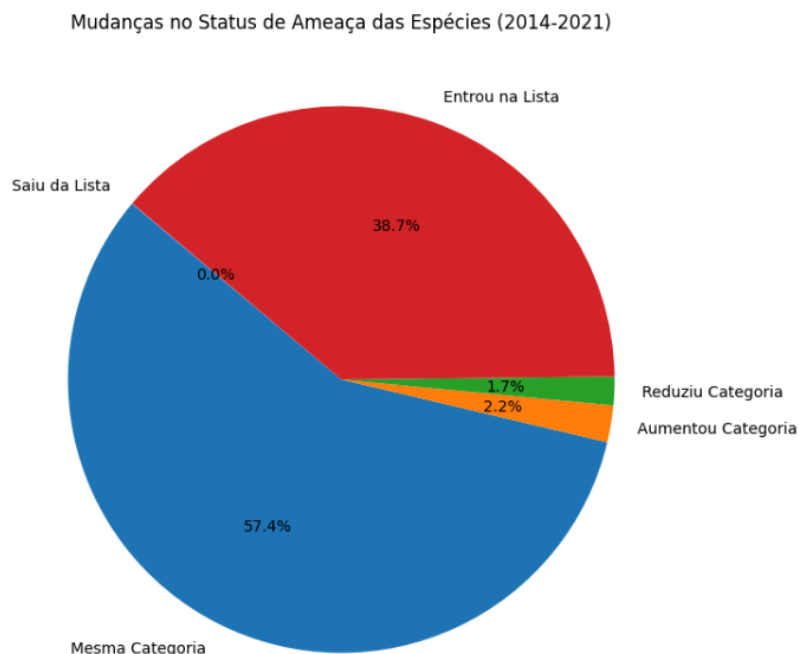


Figura 4. Mudanças no status de ameaça das espécies entre 2014 e 2021 (produzido pelo autor com o google colab).

Essas visualizações forneceram uma visão preliminar dos dados, permitindo identificar tendências que nortearam as etapas subsequentes do estudo.

4.3. Seleção das Variáveis para a Análise

Para a análise de clusterização, optou-se por utilizar três variáveis consideradas fundamentais para compreender a evolução do status de ameaça das espécies ao longo do tempo:

4.3.1. Categoria de Ameaça (2014):

Essa variável reflete o nível de risco atribuído a cada espécie na avaliação anterior, apresentando valores como "Vulnerável (VU)", "Em Perigo (EN)" ou "Criticamente em Perigo (CR)". Ela é crucial para estabelecer uma linha de base que possibilite a comparação com dados posteriores.

4.3.2. Sugestão de Categoria (2021):

Representa a nova classificação de cada espécie na avaliação mais recente, com possíveis valores como "Em Perigo (EN)", "Extinto na Natureza (EW)" ou "Menos Preocupante (LC)". Comparar essa variável com a de 2014 permite inferir se houve agravamento ou melhoria na situação de conservação.

4.3.3. Entrou ou Mudou de Categoria:

Essa variável abrange informações sobre espécies que passaram a ser consideradas ameaçadas ("Entrou em Categoria de Ameaça") e aquelas que tiveram seu status modificado, seja com aumento ou redução do nível de risco ("Aumentou a Categoria de Ameaça" e "Reduziu a Categoria de Ameaça"). Essa abordagem possibilita a identificação de padrões, como espécies que migraram de "Vulnerável" para "Críticamente em Perigo", as que permaneceram estáveis ou aquelas que, eventualmente, melhoraram e saíram da lista de ameaça.

A escolha dessas variáveis permite uma avaliação aprofundada da evolução do risco para cada espécie, embasando a aplicação do método de clusterização para identificar grupos com comportamentos similares.

4.4. Preparação dos Dados para Clusterização

Para realizar a clusterização, selecionaram-se as variáveis consideradas mais relevantes para a análise dos status de ameaça:

- Categoria em 2014
- Sugestão de Categoria 2021
- Entrada em Categoria de Ameaça

Como os métodos de clusterização requerem dados numéricos, as variáveis categóricas foram convertidas para um formato numérico, utilizando o LabelEncoder da biblioteca scikit-learn. Adicionalmente, os valores nulos presentes na variável “Entrada em Categoria de Ameaça” foram tratados e transformados em indicadores binários, a fim de evitar interferências na análise.

Em seguida, os dados foram normalizados por meio do StandardScaler, garantindo a padronização das escalas e prevenindo que diferenças de magnitude comprometessem o desempenho do algoritmo de clusterização.

4.5. Aplicação do Algoritmo de Clusterização

Para a segmentação dos dados, optou-se pela aplicação do algoritmo K-Means, devido à sua simplicidade e eficácia na identificação de grupos homogêneos em conjuntos numéricos. Foi estipulado o uso de 3 clusters, conforme definido experimentalmente, e os resultados da clusterização foram adicionados ao DataFrame original por meio de uma nova coluna intitulada “Cluster”, que facilita a análise comparativa dos grupos formados.

	Família (FB 2020)	Espécie (FB 2020)	Cluster
0	Acanthaceae	Aphelandra espirito-santensis Profice & Wassh.	0
1	Acanthaceae	Aphelandra margaritae E.Morren	2
2	Acanthaceae	Aphelandra maximiliana (Nees) Benth.	0
3	Acanthaceae	Aphelandra rigida Glaz. ex Mildbr.	1
4	Acanthaceae	Aphelandra stephanophysa Nees	2
5	Acanthaceae	Dyschoriste lavandulacea (Nees) Kuntze	0
6	Acanthaceae	Dyschoriste smithii Leonard	0
7	Acanthaceae	Justicia clauseniana (Nees) Profice	0
8	Acanthaceae	Justicia genuflexa Nees & Mart.	2
9	Acanthaceae	Justicia meyeniana (Nees) Lindau	1

Figura 5. Amostra dos dados com a coluna “Cluster” atribuída (produzido pelo autor com o google colab).

Posteriormente, calcularam-se as médias das variáveis para cada cluster, permitindo a caracterização e interpretação dos grupos formados. Para complementar, elaborou-se um gráfico de dispersão que ilustra a distribuição dos clusters no espaço das variáveis normalizadas, facilitando a visualização dos padrões emergentes .

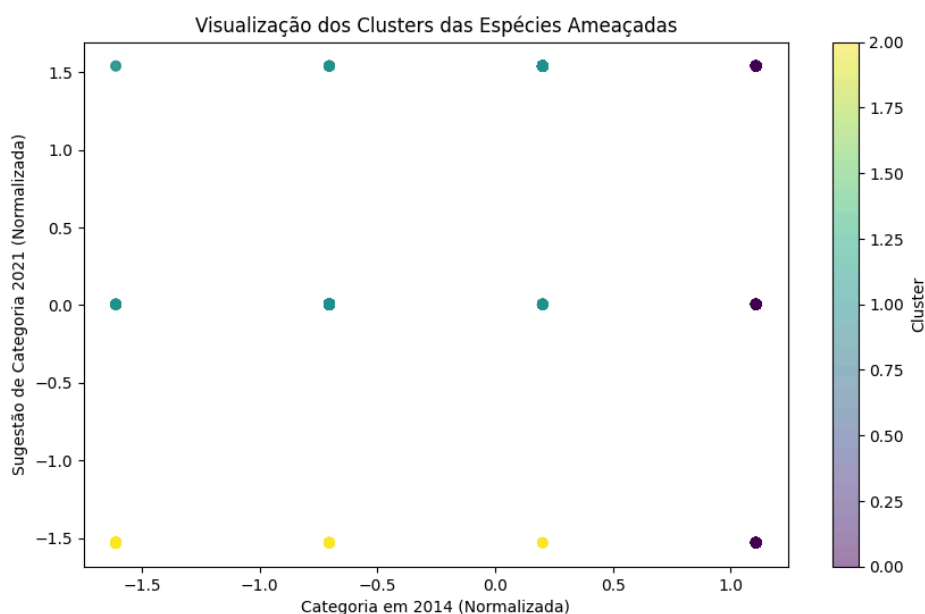


Figura 6. Gráfico de dispersão dos Clusters (produzido pelo autor com o google colab).

A integração dos procedimentos descritos, da aquisição e tratamento dos dados à aplicação da clusterização, possibilitou uma análise razoável dos padrões de ameaça na flora brasileira, fornecendo embasamento quantitativo para futuras ações de conservação e para a formulação de políticas públicas ambientais.

5. Resultados e Discussão

A análise dos dados revelou resultados que contribuem significativamente para a compreensão dos padrões de ameaça que afetam a flora brasileira. A seguir, apresentam-se os principais achados e suas respectivas interpretações.

5.1. Resultados

Os gráficos gerados a partir dos códigos implementados evidenciam diversas tendências relevantes:

5.1.1. Categorias de Ameaça (Figura 2):

O gráfico comparativo entre os anos de 2014 e 2021 demonstrou variações significativas na distribuição das categorias de ameaça. Observou-se que um número considerável de espécies apresentou aumento no risco, enquanto outras mantiveram ou reduziram sua categoria de ameaça. Tal variação sugere que modificações nos critérios de classificação e intervenções de manejo podem estar influenciando o status de conservação das espécies.

5.1.2. Famílias Botânicas com Maior Número de Espécies Ameaçadas (Figura 3):

A ordenação das famílias botânicas permitiu identificar os grupos com maior concentração de espécies ameaçadas, indicando que determinadas famílias exigem atenção prioritária. Esses resultados reforçam a necessidade de direcionar esforços de conservação para os grupos taxonômicos que abrigam um elevado número de espécies em risco.

5.1.3. Dinâmica do Status de Ameaça (Figura 4):

O gráfico de pizza evidenciou as proporções relativas das mudanças no status das espécies, como ingresso na lista, manutenção ou reclassificação do risco. Essa visualização ilustra a dinâmica contínua e, por vezes, complexa do status de ameaça, ressaltando a importância do monitoramento periódico.

5.1.4. Clusterização dos Dados:

A aplicação do algoritmo K-Means, com a definição de 3 clusters, permitiu segmentar os dados em grupos homogêneos. O cálculo das médias das variáveis por cluster e a visualização por meio de um gráfico de dispersão (Figura 6) possibilitaram caracterizar os grupos formados. De maneira geral, os clusters podem ser interpretados como:

Variável	Cluster 0	Cluster 1	Cluster 2
Categoria em 2014	3.000000	1.264078	0.066038
Entrou em Categoria de Ameaça	1.000000	0.000000	0.000000
Sugestão de Categoria 2021	0.996782	1.270550	0.000000

Figura 7. Tabela das médias de variáveis dentro de cada Cluster (produzida pelo autor).

- Cluster 0: Espécies cujo status de ameaça se manteve estável.
- Cluster 1: Espécies que apresentaram aumento no risco, sugerindo uma deterioração nas condições de conservação.
- Cluster 2: Espécies que evidenciaram uma mudança para categorias de risco menos graves ou uma melhora relativa no status.

5.1.5. Análise Detalhada dos Clusters e possíveis soluções:

A avaliação dos clusters fornece uma perspectiva diferenciada sobre o cenário de ameaça das espécies e orienta prioridades para intervenções conservacionistas. No caso do Cluster 0, verifica-se que as espécies se mantêm em situação crítica, com índices elevados de ameaça persistentes desde avaliações anteriores. Esse comportamento, possivelmente associado a condições ambientais degradadas em regiões como a Mata Atlântica e o Cerrado, ressalta a necessidade de intervenções imediatas e mais intensificadas.

O Cluster 1 agrupa espécies que, apesar de apresentarem níveis moderados de risco, não demonstraram variações significativas em seus status. Essa estabilidade, embora possa ser interpretada como ausência de agravamento, não reflete uma recuperação efetiva, evidenciando a importância de manter e, se possível, reforçar os esforços de manejo para evitar futuras deteriorações.

Em contrapartida, o Cluster 2 caracteriza-se por espécies com índices de risco praticamente inexistentes, que permanecem em condição segura. Ainda que este grupo demonstre resultados positivos, o monitoramento contínuo é indispensável para assegurar que tais condições se mantenham e para que suas estratégias possam servir de referência para outras áreas.

De modo geral, a diferenciação dos clusters orienta as ações futuras pois torna claras pelo menos três ações fundamentais: concentrar esforços no aprimoramento das condições das espécies do Cluster 0, investigar os fatores que perpetuam a estabilidade do Cluster 1 e garantir o acompanhamento das espécies do Cluster 2. Esses direcionamentos são essenciais para o desenvolvimento de políticas públicas e estratégias de manejo ambiental mais eficazes.

5.2. Discussão

Os resultados obtidos reforçam a utilidade de integrar métodos computacionais à análise de dados ambientais. A variação observada nas categorias de ameaça entre 2014 e 2021 indica que o cenário de conservação é dinâmico e que fatores como intervenções governamentais e alterações nos critérios de avaliação podem influenciar significativamente o status das espécies. A identificação das famílias botânicas mais afetadas ressalta a importância de direcionar esforços de manejo para grupos taxonômicos específicos, o que é fundamental para a priorização das ações de conservação.

A segmentação dos dados por meio do algoritmo K-Means, conforme destacado por SILVA (2003), demonstrou ser uma ferramenta eficaz para a identificação de padrões ocultos nos dados. A interpretação dos clusters permite não apenas reconhecer quais espécies estão em maior risco, mas também identificar aquelas que podem estar se

beneficiando de ações de preservação, servindo de base para uma avaliação crítica das estratégias de conservação adotadas.

Além disso, os resultados do gráfico de pizza (Figura 4) evidenciam a necessidade de um monitoramento constante, já que a dinâmica dos status de ameaça reflete a influência de múltiplos fatores, desde variações ambientais até intervenções de políticas públicas. A análise conjunta dos resultados gráficos e estatísticos fornece uma visão abrangente dos desafios e avanços na proteção da biodiversidade brasileira.

Em síntese, os achados deste estudo demonstram que, mesmo diante das limitações inerentes à base de dados, é possível extrair informações valiosas que podem orientar a formulação de políticas públicas de conservação. A combinação de análises exploratórias e técnicas de clusterização proporciona um entendimento robusto dos padrões de ameaça, contribuindo para a discussão e implementação de estratégias de manejo ambiental mais eficazes.

6. Conclusão

O presente estudo demonstrou que a integração de dados abertos do Ministério do Meio Ambiente com técnicas computacionais, como a clusterização via algoritmo K-Means, é uma abordagem eficaz para identificar padrões de ameaça na flora brasileira. A comparação entre os registros de 2014 e 2021 evidenciou variações importantes nos status de ameaça, permitindo a identificação de grupos de espécies com comportamentos distintos – desde aquelas que mantiveram sua categoria de risco até as que apresentaram aumento ou redução na ameaça.

A aplicação do K-Means possibilitou a segmentação dos dados em três clusters, cuja análise revelou informações relevantes para a priorização de esforços de conservação. Tais resultados não apenas contribuem para uma compreensão mais profunda dos desafios enfrentados pela biodiversidade, mas também fornecem subsídios quantitativos para a formulação de políticas públicas ambientais mais direcionadas e eficazes.

Entretanto, reconhece-se que o estudo possui limitações, como a falta de informações detalhadas quanto à origem geográfica dos dados, o que pode restringir a abrangência das conclusões. Assim, futuras pesquisas poderão integrar variáveis ambientais e socioeconômicas adicionais, além de considerar análises em escalas espaciais mais refinadas, a fim de aprimorar a interpretação dos padrões de ameaça.

Em síntese, a metodologia aplicada reforça a importância do uso de técnicas de análise de dados para o monitoramento da biodiversidade e destaca o potencial dos métodos de clusterização como ferramenta de apoio à conservação. Esses achados contribuem para o avanço do conhecimento na área e sugerem caminhos promissores para a melhoria contínua das estratégias de manejo e proteção da flora brasileira.

7. Referências

BRASIL. *Geoprocessamento aplicado à conservação de espécies ameaçadas de extinção: o projeto mico-leão-da-cara-preta*. (1996). Disponível em: <http://www.carto.eng.uerj.br/fgeorj/segeo1996/114/index.htm>. Acesso em: 20 fev. 2025.

SOUZA, J. F.; SILVA, M. A. F. da. *A técnica de cluster como ferramenta para a gestão ambiental*. SciELO – Sociedade e Estado, v. 18, n. 1-2, dez. (2003). Disponível em: <https://www.scielo.br/j/se/a/pQqJXVN6SqXMSTFgG75XxnC/?lang=pt>. Acesso em: 20 fev. 2025.

VERCILLO, U. E. et al. *Espécies da fauna silvestre ameaçadas de extinção no Brasil: os planos de ação nacionais e suas contribuições para as metas globais de biodiversidade*. Revista de Meio Ambiente e Desenvolvimento, v. 59, p. 461-488, jan./jun. (2022). Disponível em: <https://revistas.ufpr.br/made/article/download/77521/46564>. Acesso em: 20 fev. 2025.