

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

ANE CAROLINA SIMÕES

**MÉTODO DE GENERALIZAÇÃO PARA SISTEMA DE RECOMENDAÇÃO DE
ESPÉCIES NATIVAS PARA A PRÁTICA DE MELIPONICULTURA**

CURITIBA

2022

ANE CAROLINA SIMÕES

**MÉTODO DE GENERALIZAÇÃO PARA SISTEMA DE RECOMENDAÇÃO DE
ESPÉCIES NATIVAS PARA A PRÁTICA DE MELIPONICULTURA**

**Generalization method for recommendation system native species for the
practice of melliponiculture**

Trabalho de Conclusão de Curso de Graduação apresentado como requisito para obtenção do título de Bacharel em Sistemas de Informação do Curso de Bacharelado em Sistemas de Informação da Universidade Tecnológica Federal do Paraná.

Orientador: Prof^o Dr^o Luiz Celso Gomes Júnior,
Doutorado

Coorientador: Prof^a Dr^a Leyza Baldo Dorini,
Doutorado

CURITIBA

2022



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

ANE CAROLINA SIMÕES

**MÉTODO DE GENERALIZAÇÃO PARA SISTEMA DE RECOMENDAÇÃO DE
ESPÉCIES NATIVAS PARA A PRÁTICA DE MELIPONICULTURA**

Trabalho de Conclusão de Curso de Graduação
apresentado como requisito para obtenção do
título de Bacharel em Sistemas de Informação
do Curso de Bacharelado em Sistemas de
Informação da Universidade Tecnológica
Federal do Paraná.

Data de aprovação: 24 de junho de 2022

Ricardo Dutra da Silva
Doutorado
Universidade Tecnológica Federal do Paraná

Keiko Veronica Ono Fonseca
Doutorado
Universidade Tecnológica Federal do Paraná

Luiz Celso Gomes Júnior
Doutorado
Universidade Tecnológica Federal do Paraná

**CURITIBA
2022**

Dedico o presente trabalho aos meus pais e irmã, agradecendo toda a compreensão nos momentos mais difíceis, e ao meu marido, que acordou durante a madrugada para me fazer café e esteve comigo em todos os choros e alegrias do processo.

AGRADECIMENTOS

Certamente estes parágrafos não irão atender a todas as pessoas que fizeram parte dessa importante fase de minha vida. Portanto, desde já peço desculpas àquelas que não estão presentes entre essas palavras, mas elas podem estar certas que fazem parte do meu pensamento e de minha gratidão.

Agradeço aos meus orientadores Prof^o Dr^o Luiz Celso Gomes Júnior e Prof^a Dr^a Leyza Baldo Dorini, pela sabedoria e cuidado com que me guiaram, não somente durante o trabalho, mas por toda minha graduação.

Aos meus colegas que saíram do curso e, mesmo de longe, me incentivaram a continuar, João Paulo Stadler e Laura Madel Silva. Ao colega que me fez repensar os paradigmas que tinha com engenheiros, Brendon Leonam Pasquim. À colega que continuou me ajudando, mas a vida agitada nos impediu de manter mais contato, Carolina Lima Zanotto. Aos que não estiveram no curso, nem mesmo na área, mas durante o processo me disseram palavras de carinho e incentivo, Augusto Marques Fadanni e Larissa Cestari de Souza.

Gostaria de deixar registrado também, o meu reconhecimento aos meus tios, Vanessa Cruz, Jesus Rodrigues, e avó, Devair da Cruz, por todos os domingos de festa que tive de faltar ou sair mais cedo para poder me dedicar ao projeto e todos os momentos em que estive sentada escrevendo e programando, e cuidaram de mim.

Aos meus pais, Tais Adriana e Denio Simões, e irmã Amanda Simões por sempre me incentivarem a continuar na minha graduação, mesmo nos momentos mais difíceis, enfrentando milhares de dificuldades para eu conseguir me dedicar o máximo possível.

A todos aqueles com quem tive longas conversas de corredor na Universidade Tecnológica Federal do Paraná (UTFPR), sem vocês eu não teria força para continuar.

Ao meu marido, Gabriel Michalski, que cuidou de mim enquanto chorava por exaustão, e me motivava enquanto eu estava mais desanimada, me trazendo café e fazendo lanches durante a da madrugada.

*Sometimes I'll start a sentence and I don't even
know where it's going. I just hope I find it along
the way (SCOTT, 2008).*

RESUMO

A meliponicultura é a criação de abelhas sem ferrão para a produção de mel. Essa prática é importante para o meio ambiente, visto que as abelhas são parte significativa da polinização da flora, chegando a ser responsáveis pelo processo em 90% da Floresta Amazônica e, devido ao aumento das áreas urbanas, sua ocorrência tende a diminuir ao longo do tempo. Este estudo teve como objetivo principal desenvolver um sistema que, utilizando métodos de aprendizado de máquina e processamento de imagens, faça a generalização das ocorrências para a recomendação de espécies de abelhas nativas adequadas para a prática de meliponicultura em locais especificados pelo usuário. O sistema utiliza para treinamento do Mecanismo Global de Informação sobre Biodiversidade, do inglês, *Global Biodiversity Information Facility* (GBIF) sobre as ocorrências de abelhas e imagens de satélite para análise de áreas verdes, topografia e tipo do terreno ao redor da localização. Os dados alimentam modelos de árvore de regressão que estimam a viabilidade das espécies de abelhas na região. Os resultados das previsões realizadas pela árvore foram disponibilizados em forma de mapas, tendo em foco apenas o território brasileiro. Os mapas gerados são usados como base para uma análise qualitativa dos resultados.

Palavras-chave: abelhas nativas; meliponicultura; abelhas sem ferrão; processamento digital de imagens; aprendizado de máquina.

ABSTRACT

Meliponiculture is the breeding of stingless bees for the production of honey. This practice is important for the environment, since bees are a significant part of the pollination of flora, being responsible for 90% of the Amazon Forest (PEREIRA; SOUZA; LOPES, 2017) and, due to the increase in urban areas, their occurrence tends to decrease over time. The main objective of this study was to develop a system that, using machine learning and image processing methods, recommends native bee species suitable for meliponiculture in user-specified locations. The system uses to train GBIF on the occurrence of bees and satellite images for analysis of green areas, topography, and type of terrain around the location. The data feeds regression tree models that estimate the viability of bee species in the region. The results of the predictions made by the decision tree are available in the form of maps, focusing only on the Brazilian territory. The generated maps are used as a basis for a qualitative analysis of the results.

Keywords: native bees; meliponiculture; stingless bees; digital image processing; machine learning.

LISTA DE FIGURAS

Figura 1 – Gráfico de buscas sobre “meliponicultura” entre 2016 e 2021.	14
Figura 2 – Gráfico de buscas sobre “abelhas sem ferrão” entre 2016 e 2021.	14
Figura 3 – Imagem utilizada para obtenção de dados de NDVI	18
Figura 4 – Imagem utilizada para obtenção da classificação de terreno.	18
Figura 5 – Imagem utilizada para obtenção da altitude do local.	19
Figura 6 – Diagrama de arquitetura do sistema.	21
Figura 7 – Dados não filtrados de ocorrências recuperadas do GBIF.	22
Figura 8 – Dados filtrados de ocorrências recuperadas do GBIF.	23
Figura 9 – Ocorrências de abelhas sem ferrão no Brasil.	23
Figura 10 – Ocorrências de abelhas sem ferrão na América do Sul.	24
Figura 11 – Ocorrências de abelhas sem ferrão no Mundo.	24
Figura 12 – Mapa dos Biomas do Brasil.	25
Figura 13 – Mapa de calor referente às ocorrências da espécie <i>Tetragonisca Angustula</i> . .	26
Figura 14 – Densidade Estimada do Núcleo, do inglês, <i>Kernel Density Estimated</i> (KDE) criado a partir dos dados oriundos do GBIF.	27
Figura 15 – Imagem comparativa variando a quantidade de <i>features</i> e de pontos com e sem abelhas no treinamento.	29
Figura 16 – Imagem comparativa dos histogramas gerados com os dados demonstrados anteriormente.	30
Figura 17 – Modelo escolhido.	31

LISTA DE TABELAS

Tabela 1 – Espécies de abelhas com seu nome popular e científico.	17
Tabela 2 – Legenda de cores da Figura 4	19

LISTA DE ABREVIATURAS E SIGLAS

Siglas

EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
GBIF	Mecanismo Global de Informação sobre Biodiversidade, do inglês, <i>Global Biodiversity Information Facility</i>
KDE	Densidade Estimada do Núcleo, do inglês, <i>Kernel Density Estimated</i>
NDVI	Diferença Normalizada de Índice de Vegetação, do inglês, <i>Normalized Difference Vegetation Index</i>
NEO	Observação da Terra pela NASA, do inglês, <i>Nasa Earth Observation</i>
UTFPR	Universidade Tecnológica Federal do Paraná

SUMÁRIO

1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO	16
2.1	Melíponas	16
2.2	Árvores de decisão e regressão	16
2.3	Análise de imagens	17
3	TRABALHOS RELACIONADOS	20
4	ARQUITETURA E METODOLOGIA	21
4.1	Origem e Processamento de Dados	21
4.2	Análise Exploratória	22
4.3	Modelagem e Avaliação	24
5	RESULTADOS	28
6	CONCLUSÃO	32
	REFERÊNCIAS	33

1 INTRODUÇÃO

A meliponicultura é a prática de criar abelhas pertencentes à tribo *Meliponini*¹ e família *Hymenoptera*, as quais são produtoras de mel, cera e resina. Também conhecidas como abelhas sem ferrão, se diferem das abelhas-comuns (com ferrão, família *Apis*) por sua estrutura física, possuindo ferrões atrofiados que não conseguem ferir (PRESERVAÇÃO...). Além disso, diferem em estrutura de colmeias e ninhos, que podem ser encontrados em troncos de árvores, fendas de pedras ou no solo (SANTOS, 2010).

A prática de meliponicultura não é nova, visto que existem indícios de que o povo maia exercia essa prática para extração de mel e melhora da diversidade ambiental (RODRIGUES, 2005). Hoje em dia, as abelhas sem ferrão são comumente criadas nas Regiões Norte e Nordeste do Brasil, complementando a renda de diversas famílias que utilizam o mel e a cera para comercialização (SANTOS *et al.*, 2020).

As abelhas são necessárias para o meio ambiente, responsáveis pela polinização de 30% da flora presente na Caatinga e 90% da Floresta Amazônica (PEREIRA; SOUZA; LOPES, 2017). Segundo Kerr *et al.* (2001), o desmatamento e as queimadas representam os principais pontos de ameaça às espécies.

A diversidade de espécies de abelhas é importante em diversos aspectos da sociedade, como ecologia, economia e alimentação, visto que são responsáveis pela polinização de diversas categorias de plantas (SANTOS *et al.*, 2021).

Todavia, segundo Beringer (2018), há mais de uma década, pesquisadores e agricultores percebem uma queda na população de abelhas. As causas para essa diminuição são divergentes entre os estudiosos, mas a maioria converge para um ponto comum, a ação humana no meio ambiente. Dito isso, fica evidente a necessidade de ações para o cuidado e criação de diversas espécies de abelhas, preservando assim a flora local por meio do processo de polinização.

Recentemente, coincidindo com o início da pandemia de *COVID-19*, houve um aumento significativo no interesse sobre meliponicultura, ampliando as buscas na *web* sobre o tema. As Figuras 1 e 2, retiradas da plataforma *Google Trends*, mostram gráficos representativos do aumento nas pesquisas pelo termo “meliponicultura” entre maio e julho de 2020. Já o termo “abelhas sem ferrão” possui crescimento desde o início do ano de 2021.

No entanto, necessitam-se estudos para a criação adequada das espécies, considerando seu *habitat* comum, adequação de ambiente e redores à colônia, para não afetar os animais, além de não prejudicar a qualidade do mel produzido.

Este trabalho propõe um sistema que recebe a localização do usuário via uma aplicação *web* e utiliza a latitude e longitude para obter índices geográficos como topografia, categoria de vegetação e quantidade de vegetação na região. Esses dados são utilizados como *features* em árvores de regressão que retornam a probabilidade de cada espécie de abelha se adaptar ao

¹ nível de classificação taxonômica existente entre subfamília e gênero

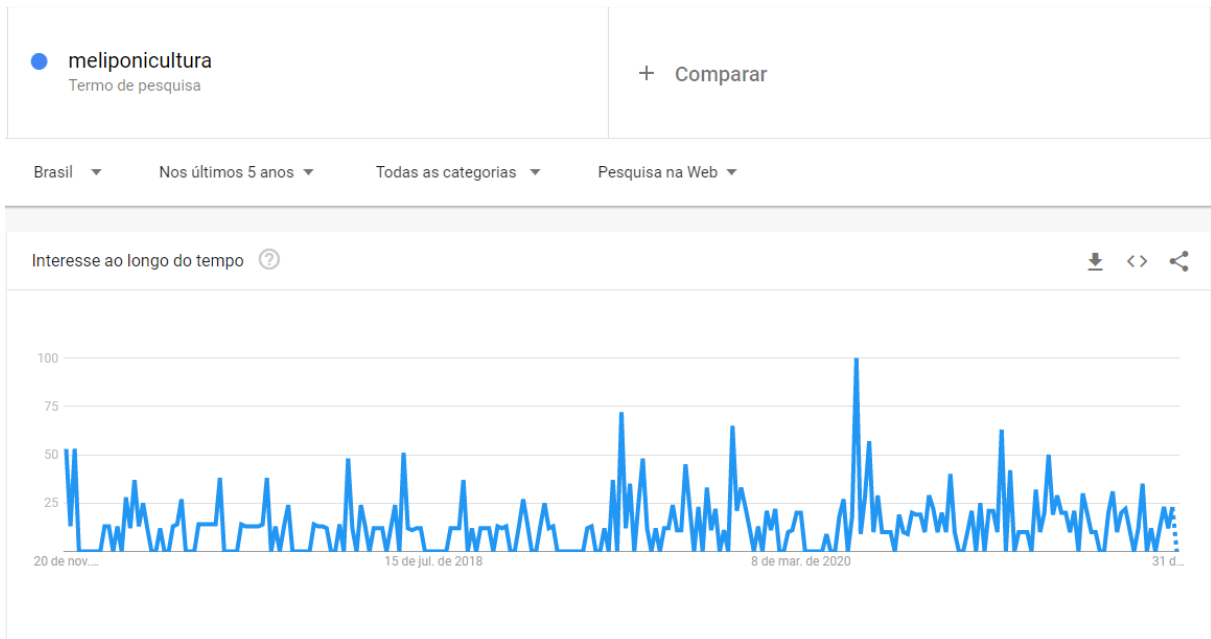


Figura 1 – Gráfico de buscas sobre “meliponicultura” entre 2016 e 2021.

Fonte: Google Trends.

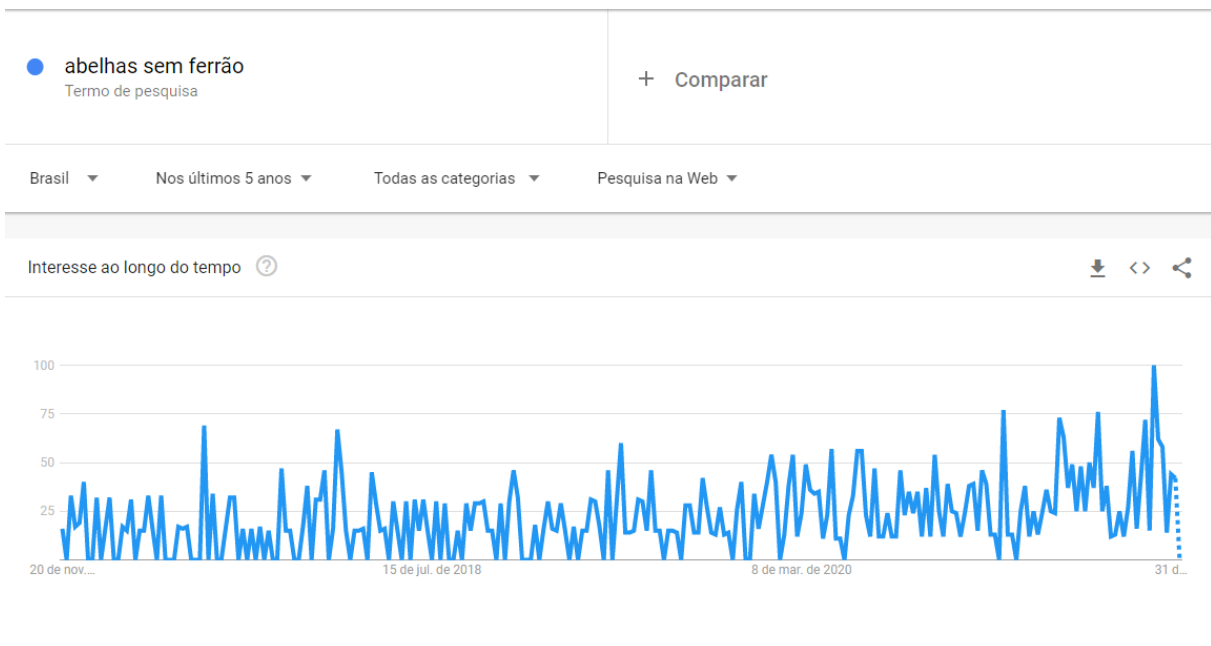


Figura 2 – Gráfico de buscas sobre “abelhas sem ferrão” entre 2016 e 2021.

Fonte: Google Trends.

local.

Para isso, foi necessário identificar índices de geografia e vegetação, utilizando imagens tipo *raster* retiradas do portal de imagens de satélite Observação da Terra pela NASA, do inglês, *Nasa Earth Observation* (NEO). Além disso, foi desenvolvida e treinada uma árvore de regressão capaz de estimar a adequação das espécies na região, generalizando os dados recebidos da plataforma.

Sendo assim, foi definida a estrutura do presente trabalho. O Capítulo 1 é composto pelo

contexto do trabalho, bem como seus objetivos e justificativa, seguido pelo Capítulo 2 descrevendo o referencial teórico utilizados como base para a escrita do texto. O Capítulo 3 descreve trabalhos e projetos que utilizam conceitos próximos e podem ser lidos para aprofundamento no tema. Já os Capítulos 4 e 5 são definidos pelos métodos utilizados e resultados obtidos durante o processo, finalizando o trabalho com as considerações finais.

2 REFERENCIAL TEÓRICO

O presente trabalho se baseia em três temas principais, melíponas, árvores de regressão e processamento de imagens, os combinando para obter um sistema preditor de adaptabilidade das abelhas. Todos os campos serão explorados neste capítulo.

2.1 Melíponas

Segundo Kerr *et al.* (2001), até o ano de 1838 as Melíponas eram as únicas espécies de abelhas presentes no território brasileiro. Nesse ano, o padre Manoel Severiano foi o responsável por introduzir as abelhas com ferrão, foi possível viabilizar a produção de ceras brancas para as velas.

Atualmente, a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) incentiva a prática de meliponicultura, destacando os benefícios gerados pela polinização e utilizando pesquisadores da área para desenvolver cartilhas, além de promover cursos gratuitos sobre o tema. Pereira, Souza e Lopes (2017) possuem alguns trabalhos publicados junto a EMBRAPA, nos quais citam abelhas sem ferrão e também explicam sobre estruturas de ninhos, morfologia dos insetos e sobre a alimentação dos meliponários.

Conforme dito por Abelhas... (), um ninho de jataí (*Tetragonisca angustula*), quando em um ambiente propício e possuindo flora adequada, pode produzir anualmente, um litro ou mais de mel, enquanto a abelha africanizada pode superar 20 litros em um ano. Além da demora para a produção, existem distinções entre os produtos de cada espécie, como consistência, cor, aroma e sabor. Por essas razões, o mel produzido por essas espécies é mais valorizado (CELLA; AMANDIO; FAITA, 2018). Segundo Senkovski (2017), o mel produzido pelas abelhas nativas pode custar R\$ 120 o quilo, representando uma fonte e renda que pode, em simultâneo, auxiliar na preservação ambiental.

Até o momento da escrita do trabalho, pode-se afirmar que existem três tribos de abelhas sem ferrão catalogadas, *Meliponini*, *Trigonini* e *Lestrimelitini*. Todavia, como a primeira possui maior quantidade de documentações sobre o tema, será utilizada apenas essa tribo para o estudo. Apesar de abelhas nativas serem conhecidas por seus jargões populares (muitas vezes repetidos em diferentes espécies), possuem seus nomes científicos identificados conforme descritos na Tabela 1 (SILVA *et al.*, 2019).

2.2 Árvores de decisão e regressão

O aprendizado de máquina visa o desenvolvimento de ferramentas e técnicas que possam adquirir conhecimentos e tomar decisões baseadas em contexto e conhecimento prévio (MONARD; BARANAUSKAS, 2003). Sua análise é baseada em estatísticas e informações

Nome Popular	Nome Científico
Abelha-Limão	Lestrimelitta limão
Boca-de-Sapo	Partamona helleri
Borá	Tetragona clavipes
Guarupu	Melipona bicolor
Irapuã	Trigona spinipes
Jataí	Tetragonisca angustula
Jataí-da-Terra	Paratrigona subnuda
Lambe-Olhos	Leurotrigona muelleri
Mandaçaia	Melipona mandacaia
Manduri	Melipona marginata
Marmelada Amarela	Frieseomelitta varia
Mombucão	Cephalotrigona capitata
Tubuna	Scaptotrigona bipunctata
Uruçu	Melipona scutellaris

Tabela 1 – Espécies de abelhas com seu nome popular e científico.

Fonte: Pereira, Souza e Lopes (2017).

aprendidas com bases de treinamento, que podem estar já rotuladas (no caso de aprendizado supervisionado) ou não rotuladas (no caso de aprendizado não supervisionado).

A árvore de decisão é um dos muitos métodos de classificação existentes, esta é definida por uma estrutura onde cada um dos nós internos representa uma “questão” e os níveis abaixo representam as respostas. O nó mais ao topo da árvore representa a raiz e os últimos são chamados folhas, esses exibem os resultados das decisões realizadas pela árvore (ROKACH; MAIMON, 2005).

No presente trabalho, foi utilizada uma variação das árvores de decisão, conhecidas como árvores de regressão, que se diferem por utilizarem dados não *booleanos* e sim predizendo os valores dentro de um intervalo identificado nos dados de treinamento.

2.3 Análise de imagens

O processamento de imagens foi utilizado no trabalho para análise de arquivos *raster*, imagens onde os dados são representados em forma de matriz (GBIF, 2022).

Diferença Normalizada de Índice de Vegetação, do inglês, *Normalized Difference Vegetation Index* (NDVI) sinaliza a quantidade e a saúde da vegetação no local. O valor pode variar no intervalo [-1,1] sendo que -1 representa a inexistência de vegetação e 1 representa uma flora saudável (JUNIOR, 2020). A Figura 3 mostra os valores de NDVI no globo, sendo que os pontos onde o verde é mais escuro estão mais próximos de 1 e mais claros perto de -1. Podemos observar a grande presença da cor verde-escura no Brasil, principalmente na região Norte onde está presente a Amazônia.

O índice de classificação de terreno utiliza cores para diferenciar as categorias, con-

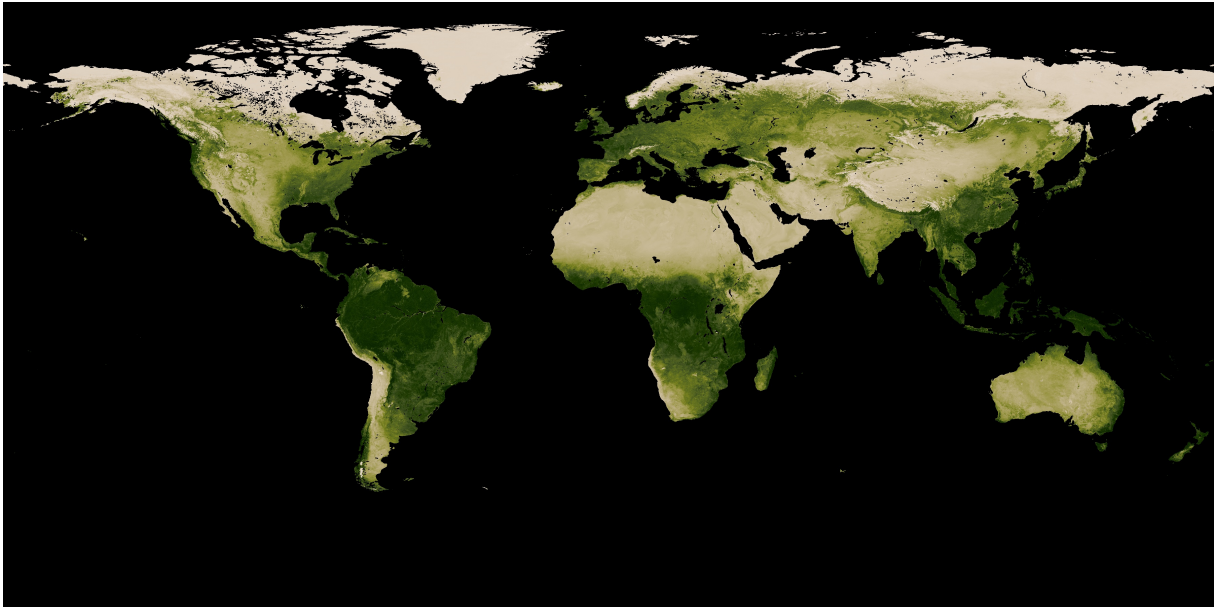


Figura 3 – Imagem utilizada para obtenção de dados de NDVI
Fonte: NASA (2022).

forme demonstrado na Figura 4. Sendo que cada uma possui um valor designado, conforme lido na Tabela 2.

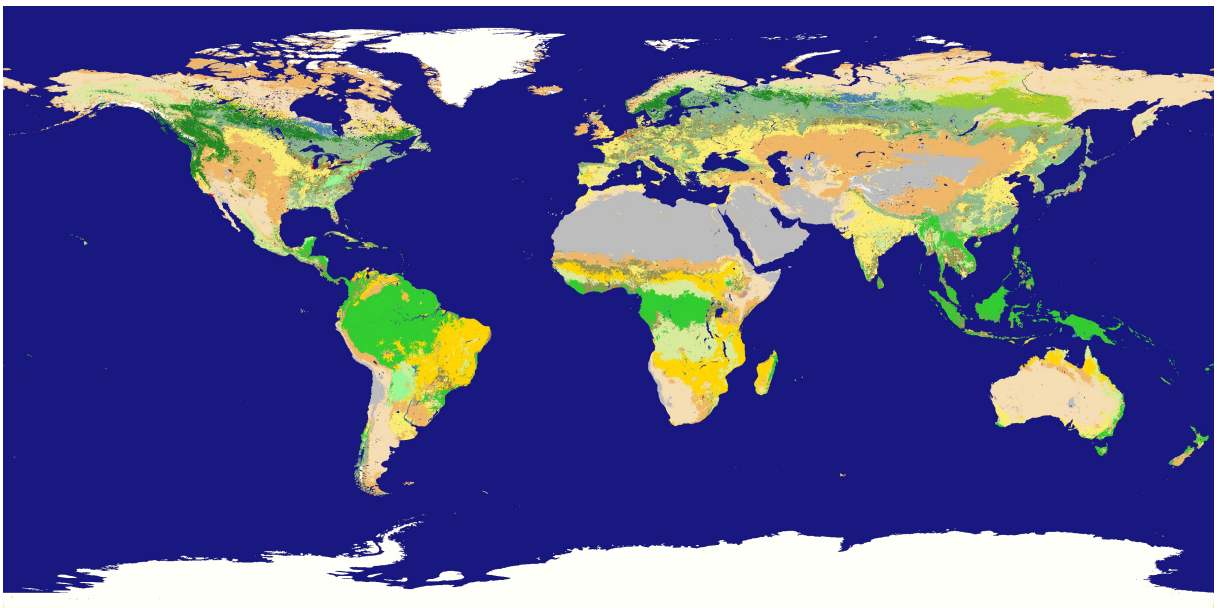


Figura 4 – Imagem utilizada para obtenção da classificação de terreno.
Fonte: NASA (2022).

A topografia utiliza de um intervalo entre 0 e 6400, sendo essa a altitude em metros de cada ponto. Na Figura 5 é possível observar que o verde refere-se a pontos mais próximos de zero, portanto próximo ao nível do mar, e locais mais próximos ao branco demonstram maior altitude.

Valor	Título	Cor	Valor	Título	Cor
0	Água		9	Savanas	
1	Floresta perene de folhas pequenas		10	Campo	
2	Floresta perene de folhas largas		11	Pantanos	
3	Floresta caducifolia de folhas pequenas		12	Campo de Cultivo	
4	Floresta caducifolia de folhas largas		13	Área Urbana	
5	Florestas Mistas		14	Vegetação Intocada	
6	Matagais fechados		15	Área permanentemente congelada	
7	Matagais abertos		16	Vegetação Esparsa	
8	Savanas Arborizadas				

Tabela 2 – Legenda de cores da Figura 4

Fonte: Sulla-Menashe e Friedl (2018).

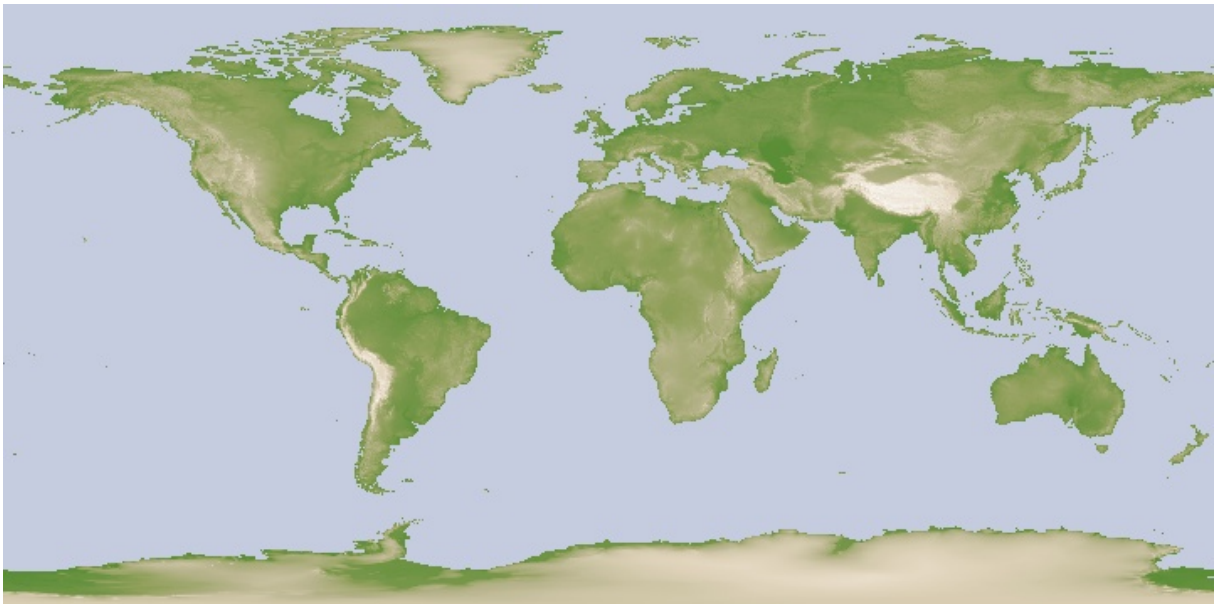


Figura 5 – Imagem utilizada para obtenção da altitude do local.

Fonte: NASA (2022).

3 TRABALHOS RELACIONADOS

Tratando-se da questão do início da atividade de meliponicultura, a Associação Brasileira de Estudos de Abelhas (A.B.E.L.H.A) desenvolveu um aplicativo ¹ para auxiliar no entendimento sobre o tema com artigos sobre mel, polinização e conservação das abelhas. Para os usuários já produtores, o aplicativo *Beeh App* ², desenvolvido pela empresa homônima, propõe um sistema de controle de enxames para os meliponicultores, organizando datas de alimentação e para colheita.

Esses projetos diferem do presente trabalho devido ao foco de desenvolvimento, visto que o aplicativo desenvolvido pela *Beeh App* visa auxiliar pessoas que já praticam a meliponicultura e o sistema da A.B.E.L.H.A informa pessoas interessadas em iniciar a prática com artigos e imagens. O produto visado como resultado deste trabalho facilitará a escolha das abelhas, não somente com textos, mas também utilizando de tecnologias para entender quais as espécies adequadas para a região.

Em Curitiba, a prefeitura criou o projeto “Jardins de Mel”, que teve seu início no ano de 2018, visando conscientizar a população sobre a importância das abelhas sem ferrão. Para isso, foram espalhados 53 meliponários, com diferentes espécies, em parques da cidade. O trabalho de ensinar sobre os benefícios das melíponas também foi aplicado em escolas, o projeto “Abelhas” propôs uma extensão do “Jardins de Mel”, apresentado em mais de 30 unidades de ensino como Centros Municipais de Educação Infantil (CMEI) e Escolas Municipais da cidade, promovendo o incentivo para as crianças pesquisarem sobre o assunto, além de serem instalados meliponários ao redor das unidades de ensino para que os alunos cuidassem e tivessem maior proximidade com os insetos (CURITIBA, 2019).

No âmbito de processamento de imagens de satélite, utiliza-se como inspiração o projeto *Sunroof* ³ (traduzido como Teto solar), guiado pela empresa *Google*, a qual visa incentivo do uso de painéis solares onde, utilizando imagens de satélites, verifica-se a incidência solar diária na localização escolhida pelo usuário, retornando a quantidade de horas de sol por ano e assim fazendo uma estimativa de quantos dólares seriam salvos caso opte por instalar os painéis e o tamanho do painel necessário para instalação. A similaridade entre o *Sunroof* e o projeto proposto no artigo baseia-se apenas na utilização de imagens de satélite em conjunto com análises de dados feita, porém diferem no escopo.

¹ Disponível em <https://abelha.org.br/abelha-app- agora-disponivel-tambem-na-app-store/>

² Disponível em <https://beeh.app/>

³ Disponível em: <https://sunroof.withgoogle.com/>

4 ARQUITETURA E METODOLOGIA

A estrutura do sistema dá-se por quatro componentes principais: uma aplicação *web* desenvolvida em *JavaScript*, sistema em *Python* para o *backend*, banco de ocorrência de espécies obtido via GBIF e arquivos do tipo *raster*, conforme representado pela Figura 6.

O fluxo na imagem consiste na interação entre o usuário e a aplicação *web* que envia os dados de localização do usuário ao sistema em *Python*. Este possui registros do banco de dados de espécies e os arquivos *raster*, para serem analisados pela árvore de regressão. Após o processamento, o sistema retorna ao usuário quais as abelhas com melhor adaptabilidade à região.

Apesar da aplicação *web* não ter sido implementada, os demais itens estão presentes no sistema, sendo assim, a parte em *javascript* poderá ser adicionada futuramente, em possíveis complementos do trabalho.

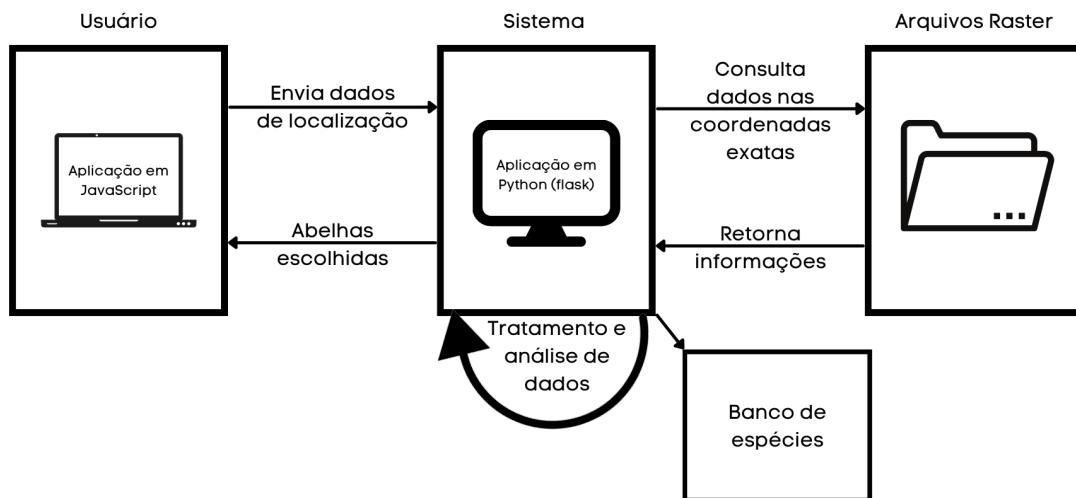


Figura 6 – Diagrama de arquitetura do sistema.

Fonte: Produção própria.

4.1 Origem e Processamento de Dados

Foram utilizados os dados presentes no GBIF, onde os pesquisadores notificam quando identificada a ocorrência de espécies, fornecendo informações sobre sua localização para o mapeamento. Foi necessário filtrar os dados recuperados na base, pois alguns apresentavam informações sobre a pessoa que fez o registro da ocorrência, o que é irrelevante para a pesquisa. Também foram descartados registros sem localização ou de espécies que não são

recomendadas para a prática de meliponicultura (PEREIRA; SOUZA; LOPES, 2017).

Os dados foram tratados e plotados utilizando *Python* e as bibliotecas *Matplotlib*, *Pandas* e *GeoPandas* que auxiliaram e facilitaram a visualização dos dados. As informações contidas na tabela foram separadas em cinco colunas: gênero (*genus*), espécie (*species*), o código do país (*countryCode*), latitude (*decimalLatitude*) e longitude (*decimalLongitude*), sendo que cada linha da tabela representa uma ocorrência de uma espécie de abelha sem ferrão registrada por uma pessoa envolvida na pesquisa. Além disso, foi criada uma coluna adicional chamada *geometry*, que se refere à combinação dos dados de latitude e longitude em um objeto do tipo *point* pertencente à biblioteca *GeoPandas*.

Quando feita a impressão dos dados, obtém-se o total de 207.491 linhas, sem utilizar os filtros de espécie, conforme pode ser observado na Figura 7.

	genus	species	countryCode	decimalLatitude	decimalLongitude	geometry
0	Scaptotrigona	Scaptotrigona hellwegeri	MX	23.638811	-106.360156	POINT (-106.36016 23.63881)
1	Tetragonula	Tetragonula collina	VN	11.433537	107.422995	POINT (107.42300 11.43354)
2	Trigona	Trigona fulviventris	MX	20.795870	-89.035887	POINT (-89.03589 20.79587)
3	Tetragona	Tetragona clavipes	BR	-15.741608	-47.756392	POINT (-47.75639 -15.74161)
4	Tetragonula	Tetragonula iridipennis	LK	7.303551	80.650977	POINT (80.65098 7.30355)
...
321012	Tetragonisca	Tetragonisca angustula	CO	11.223000	-74.186000	POINT (-74.18600 11.22300)
321013	Tetragonisca	Tetragonisca angustula	CO	11.223000	-74.186000	POINT (-74.18600 11.22300)
321014	Tetragonisca	Tetragonisca angustula	CO	11.223000	-74.186000	POINT (-74.18600 11.22300)
321015	Tetragonisca	Tetragonisca angustula	CO	11.223000	-74.186000	POINT (-74.18600 11.22300)
321016	Tetragonisca	Tetragonisca angustula	CO	11.223000	-74.186000	POINT (-74.18600 11.22300)

207491 rows x 6 columns

Figura 7 – Dados não filtrados de ocorrências recuperadas do GBIF.

Fonte: Produção própria.

Feitos os filtros nas espécies, obteve-se uma diminuição para 30.158 registros, sendo 14% dos dados originais, conforme podemos observar na Figura 8.

4.2 Análise Exploratória

Após o processamento dos dados, foi representada graficamente a distribuição dos dados, utilizando os mapas do Brasil, América do Sul e do mundo, conforme as Figuras 9, 10 e 11. Cada ponto na imagem representa um registro na tabela de ocorrências e cada uma das cores demonstra uma espécie, descrita na legenda por seu nome científico (Tabela 1).

A partir da distribuição, se entende a preferência das espécies pelo bioma da Mata Atlântica além de, parte do Cerrado e Mata Amazônica representados na Figura 12. Os pontos

	genus	species	countryCode	decimallatitude	decimallongitude	geometry
3	Tetragona	Tetragona clavipes	BR	-15.741608	-47.756392	POINT (-47.75639 -15.74161)
7	Tetragonisca	Tetragonisca angustula	BZ	17.539444	-89.111019	POINT (-89.11102 17.53944)
9	Tetragonisca	Tetragonisca angustula	BR	-15.763198	-47.880518	POINT (-47.88052 -15.76320)
12	Tetragonisca	Tetragonisca angustula	BR	-20.082891	-40.786885	POINT (-40.78688 -20.08289)
19	Tetragonisca	Tetragonisca angustula	HN	14.841200	-89.134223	POINT (-89.13422 14.84120)
...
321012	Tetragonisca	Tetragonisca angustula	CO	11.223000	-74.186000	POINT (-74.18600 11.22300)
321013	Tetragonisca	Tetragonisca angustula	CO	11.223000	-74.186000	POINT (-74.18600 11.22300)
321014	Tetragonisca	Tetragonisca angustula	CO	11.223000	-74.186000	POINT (-74.18600 11.22300)
321015	Tetragonisca	Tetragonisca angustula	CO	11.223000	-74.186000	POINT (-74.18600 11.22300)
321016	Tetragonisca	Tetragonisca angustula	CO	11.223000	-74.186000	POINT (-74.18600 11.22300)

30158 rows x 6 columns

Figura 8 – Dados filtrados de ocorrências recuperadas do GBIF.

Fonte: Produção própria.

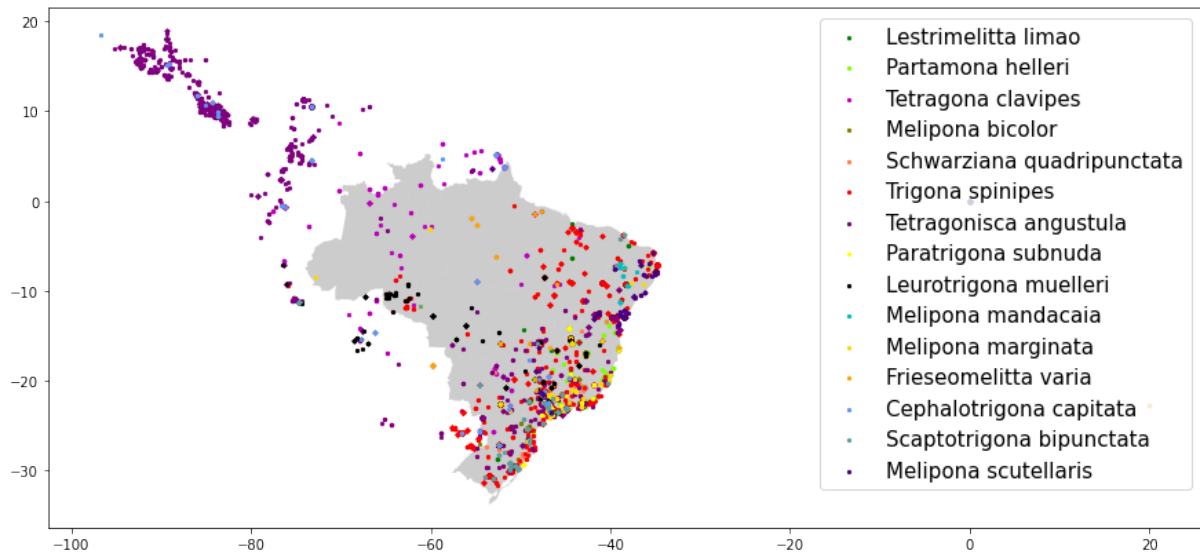


Figura 9 – Ocorrências de abelhas sem ferrão no Brasil.

Fonte: Produção própria.

demonstrados concentram-se nas regiões Sul, Sudeste e Nordeste do país, onde também estão pontuados os grandes centros urbanos, como as cidades de São Paulo e Rio de Janeiro, podendo implicar em tendências nas distribuições, visto que esses lugares representam a maior quantidade de pesquisadores (CNPQ, 2017).

Após isso, se iniciou o processo de análise e plotagem dos dados, onde foram criados mapas de calor separados por espécies para melhor observação da distribuição conforme observado na Figura 13 que se utiliza a espécie *Tetragonisca Angustula* como exemplo.

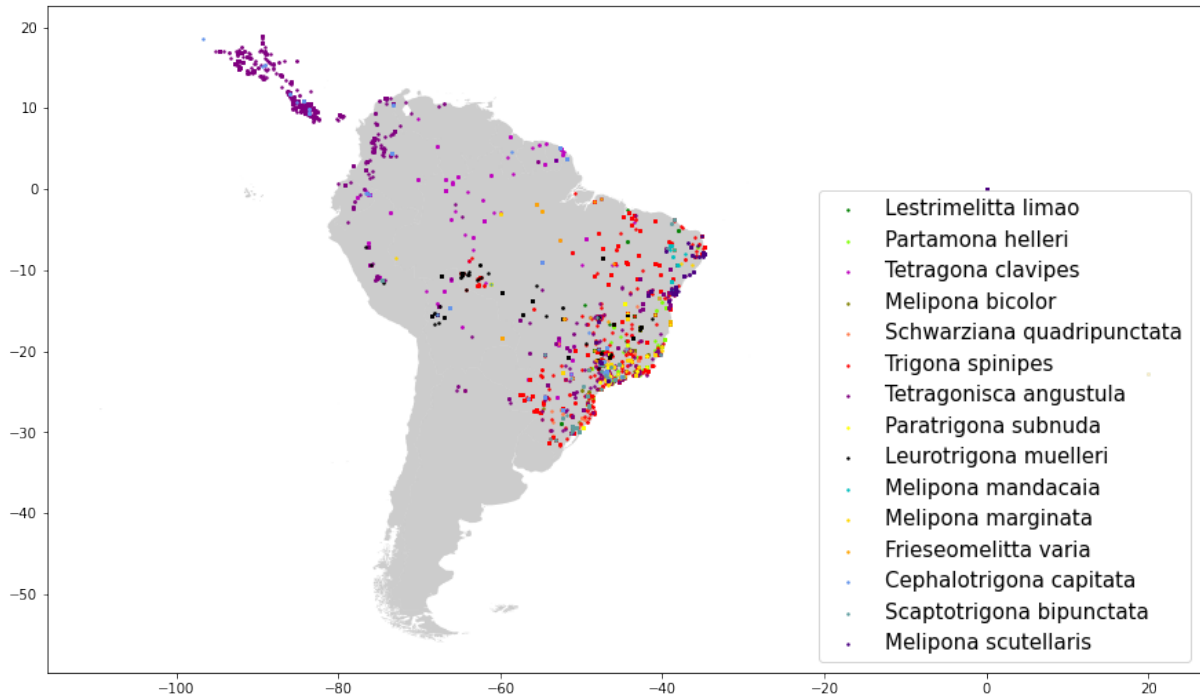


Figura 10 – Ocorrências de abelhas sem ferrão na América do Sul.

Fonte: Produção própria.

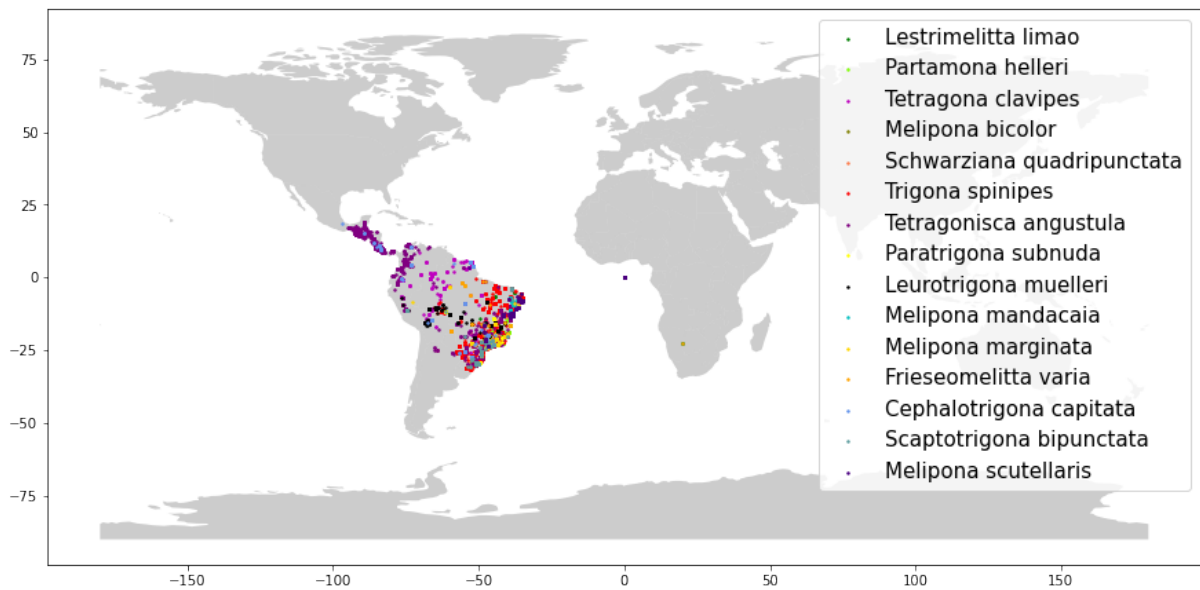


Figura 11 – Ocorrências de abelhas sem ferrão no Mundo.

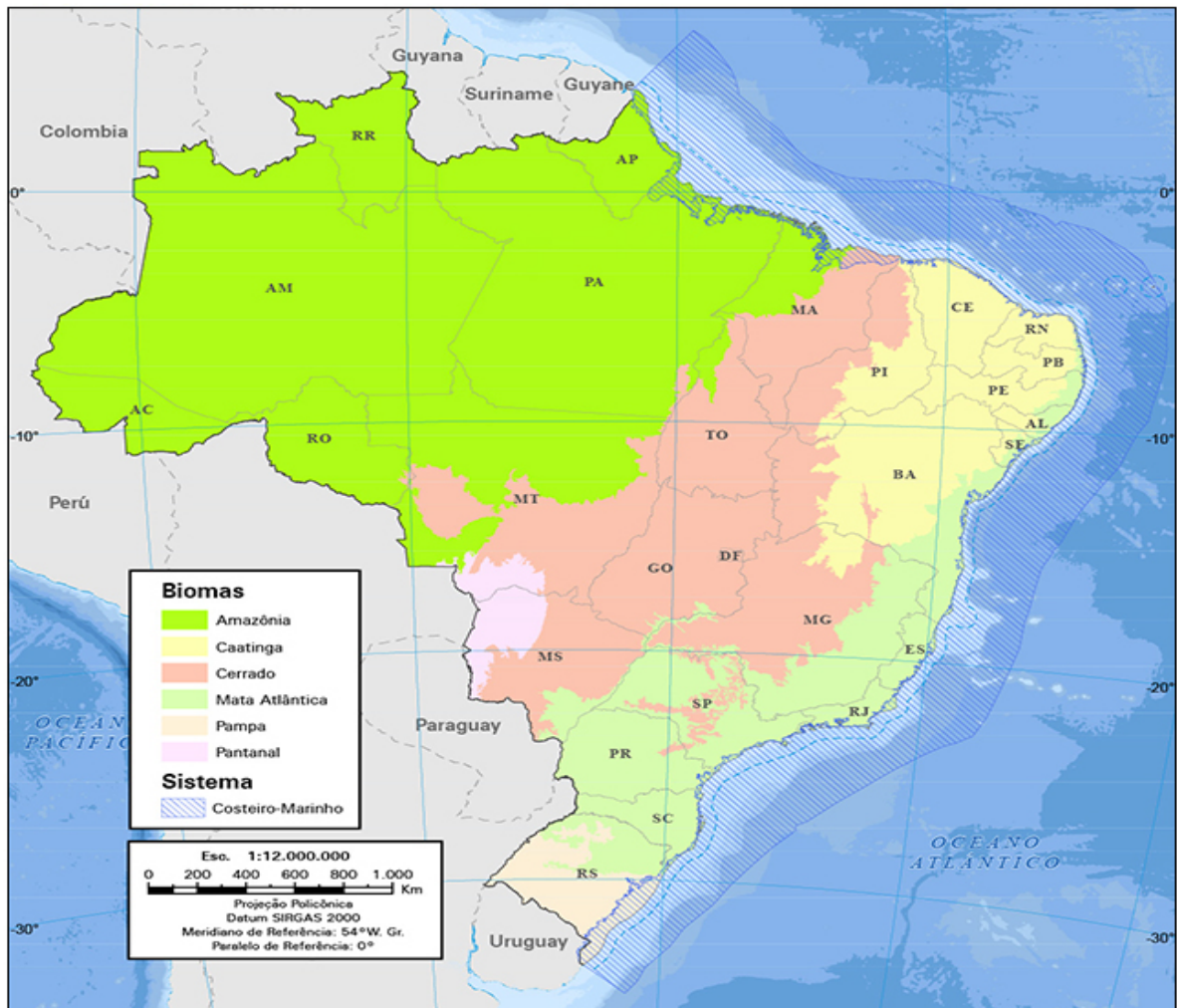
Fonte: Produção própria.

4.3 Modelagem e Avaliação

Os principais desafios na modelagem do problema são: a inexistência de dados adequados para o treinamento, visto a sub-notificação das ocorrências, a inexistência de critérios objetivos para a seleção das *features* e a geração de exemplos em locais sem abelhas.

Os dados de ocorrência das espécies representam parcialmente a distribuição e contém os vieses descritos na Seção 4.2. Portanto, o primeiro desafio foi determinar uma generalização

Biomias brasileiros



Fonte: IBGE

Figura 12 – Mapa dos Biomas do Brasil.

Fonte: IBGE (LAURA, 2021).

aproximada da distribuição das espécies. Para isto foi utilizado o KDE (Figura 14), obtendo assim a probabilidade de existirem abelhas nas regiões determinadas. As informações retornadas foram usadas posteriormente nas funções de predição, essenciais para o desenvolvimento das árvores de regressão.

Outro desafio relacionado com a falta de dados representativos é a inexistência de um critério objetivo para a seleção das *features*. Deste modo, foram criados diversos modelos variando o número de *features* e alternando entre mais ou menos pontos com ocorrências na base de treinamento.

Por fim, para produzir exemplos de regiões sem abelhas para treinar o algoritmo, selecionamos pontos onde o KDE obteve resultado menor que 0,0001. O número foi definido após testes com a base, em que se observou valores menores que 0,0001 em pontos onde não haveria possibilidade de haver abelhas como, por exemplo, o oceano Atlântico. Todavia, quando utilizados valores menores que o citado, o resultado dos modelos permanecia o mesmo, con-



**Figura 13 – Mapa de calor referente às ocorrências da espécie *Tetragonisca Angustula*.
Fonte: Produção Própria.**

vencionando assim o uso deste.

Como não existe um critério definitivo para a avaliação do sistema, optamos por uma abordagem qualitativa a partir da análise visual de mapas com os resultados das variações do modelo. O objetivo é identificar indícios de que os modelos estão generalizando adequadamente a extensão das regiões de ocorrência das espécies.

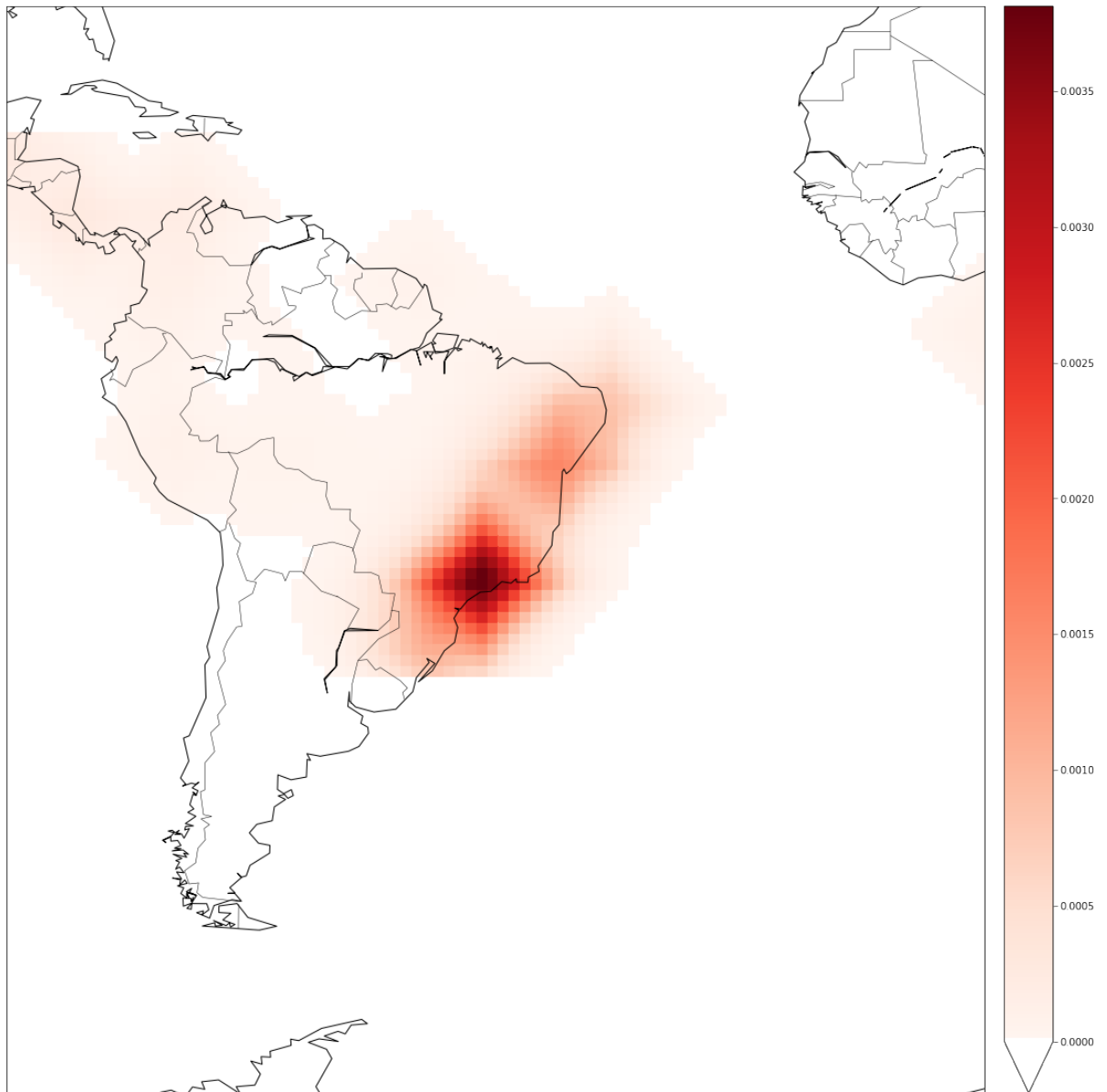


Figura 14 – KDE criado a partir dos dados oriundos do GBIF.

Fonte: Produção Própria.

5 RESULTADOS

Durante a geração dos modelos, foram desenvolvidas diversas bases de treinamento, que foram comparados entre diversas combinações, onde se buscavam modelos que fossem coesos à base recebida e ao KDE desenvolvido. Além disso, viu-se a necessidade de a comparação final ser entre modelos com muitas e com poucas abelhas no treinamento, para analisar o comportamento da árvore utilizando essas bases. Por fim, foram obtidos os quatro modelos:

- 80% dos pontos sem abelhas e 20% com abelhas, utilizando todas as *features* (NDVI, média NDVI, desvio padrão NDVI, topografia e categoria do terreno).
- 5% dos pontos sem abelhas e 95% com abelhas, utilizando todas as *features*
- 80% dos pontos sem abelhas e 20% com abelhas, utilizando apenas a *feature* de média NDVI
- 5% dos pontos sem abelhas e 95% com abelhas, utilizando apenas a *feature* de média NDVI

Os resultados são observados na Figura 15.

Apesar da semelhança entre as quatro imagens, pode-se observar uma distribuição mais esparsa nas plotagens que utilizam apenas a *feature* de média. Também vê-se os modelos em que não há filtro das *features* tem grande semelhança ao KDE (Capítulo 4 - Seção 4.3) com maior destaque para as regiões Sul e Sudeste.

Além disso, obtiveram-se os histogramas de cada uma das imagens desenvolvidas, demonstrando com mais clareza a distribuição dos valores, visto na Figura 16. Neste, podemos observar uma grande quantidade de pontos resultando em zero. Isso se atribui a toda a parte de oceano presente no mapa, além de outros países, onde não haviam registros de abelhas e sua vegetação não é semelhante à brasileira.

Ao final das análises, foi definido o uso do modelo com todas as *features* treinado com 5% dos dados com pontos que não possuem ocorrências de abelhas e 95% com registro de presença de melíponas (Figura 17). Este mostrou melhor coerência nos dados, quando comparado aos dados retornados obtidos pelo GBIF e ao KDE.

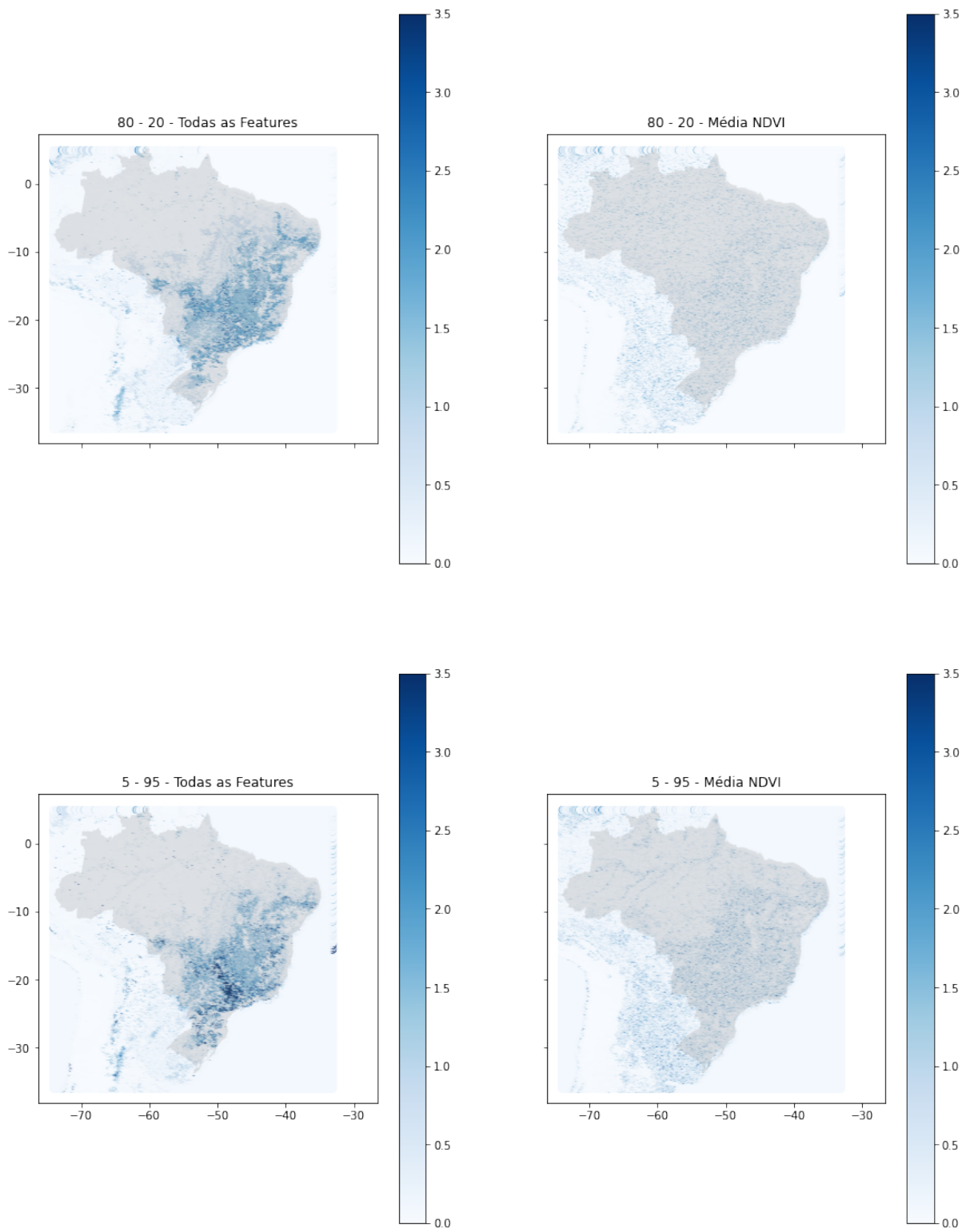


Figura 15 – Imagem comparativa variando a quantidade de *features* e de pontos com e sem abelhas no treinamento.

Fonte: Produção Própria.

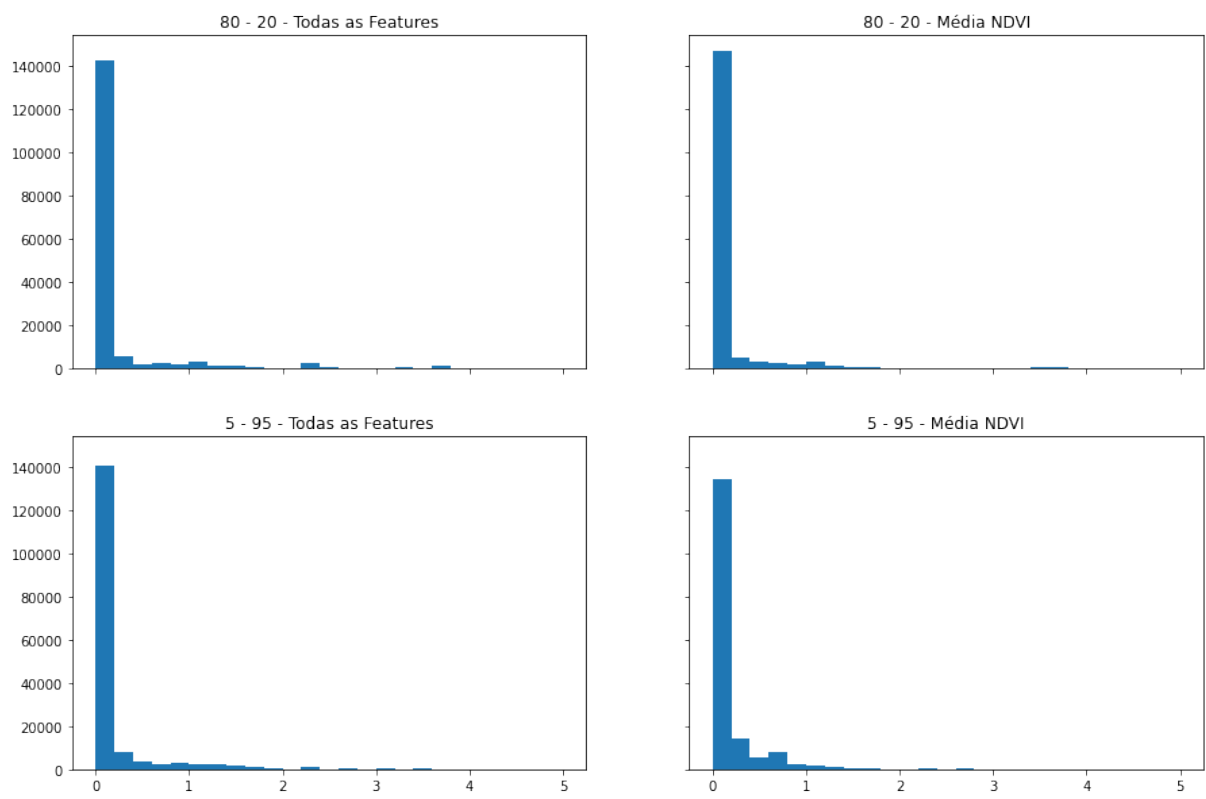


Figura 16 – Imagem comparativa dos histogramas gerados com os dados demonstrados anteriormente.

Fonte: Produção Própria.

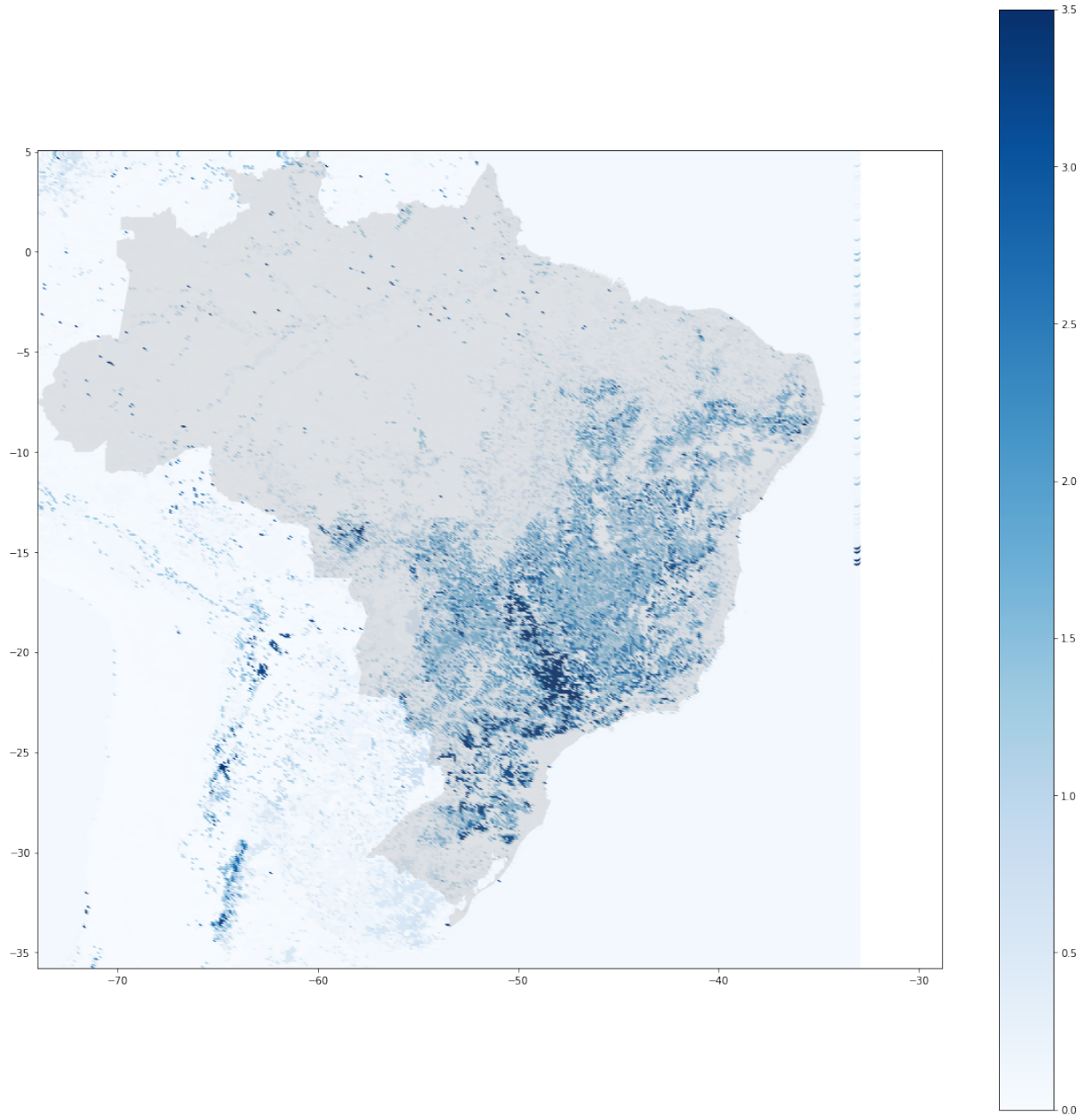


Figura 17 – Modelo escolhido.
Fonte: Produção Própria.

6 CONCLUSÃO

A prática de meliponicultura, além de benéfica ao meio ambiente, devido polinização realizada pelas abelhas, pode ser utilizada como fonte de renda ou, até mesmo, como uma atividade manual para diminuir o estresse. O sistema proposto contribui para a difusão da prática, facilitando o processo para meliponicultores iniciantes e, também para criadores experientes, que desejam aumentar sua colônia.

Dados os objetivos propostos no Capítulo 1, se desenvolveu um sistema capaz de prever, com base nos resultados de uma árvore de decisão, a melhor espécie de abelha sem ferrão para a prática da meliponicultura, com base nas características de terreno e vegetação do local em que se deseja desenvolver a colônia.

Com base nas imagens apresentadas no Capítulo 5, podemos afirmar a preferência das espécies de abelhas das regiões de grandes centros, contudo, a predição foi afetada pela distribuição dos pontos recebidos pelo GBIF, onde os grandes centros urbanos possuem maior número de ocorrências do que em regiões com menos pesquisadores. Apesar disso, pode-se considerar sucesso do modelo, visto que as predições estão corretas com base nos dados recebidos.

Como continuação, seria necessária revisão dos dados recebidos e tratados intuindo a melhora na qualidade dos mapas desenvolvidos. Além do auxílio de profissionais da área da Biologia, podendo assim delinear com mais clareza as necessidades das abelhas sem ferrão.

REFERÊNCIAS

- ABELHAS Indígenas sem Ferrão. [S.l.: s.n.].
- BERINGER, J. e. a. **Revista Eletrônica Científica da UERGS**. 2018. Disponível em: <http://revista.uergs.edu.br/index.php/revuergs/article/view/1686/411>.
- CELLA, I.; AMANDIO, D. T. T.; FAITA, M. R. Meliponicultura. **Boletim Didático**, v. 1, p. 56, out. 2018. Disponível em: <https://publicacoes.epagri.sc.gov.br/BD/article/view/408>.
- CNPQ, C. N. d. D. C. e. T. **Por região - Plataforma Lattes - CNPq**. 2017. Disponível em: <http://lattes.cnpq.br/web/dgp/por-regiao1>.
- CURITIBA, P. M. d. **AbeLinhas do Conhecimento - Secretaria da Educação**. 2019. Disponível em: <https://educacao.curitiba.pr.gov.br/conteudo/abelinhas-do-conhecimento/10496>.
- GBIF. 2022. Disponível em: https://docs.qgis.org/2.14/pt_BR/docs/gentle_gis_introduction/raster_data.html.
- JUNIOR, A. C. Tudo o que você precisa saber sobre o ndvi: Perguntas respostas para a agricultura 4.0 – parte 1 - geografia das coisas. In: . [s.n.], 2020. Disponível em: <https://geografiadascoisas.com.br/artigos/tudo-o-que-voce-precisa-saber-sobre-o-ndvi-parte-1/>.
- KERR, W. E. *et al.* Aspectos pouco mencionados da biodiversidade amazônica. In: . [S.l.: s.n.], 2001. v. 6, n. 12, p. 20–41.
- LAURA, A. Ibge educa jovens. In: . [s.n.], 2021. Disponível em: <https://educa.ibge.gov.br/jovens/conheca-o-brasil/territorio/18307-biomas-brasileiros.html>.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. In: . [S.l.]: Manole Ltda, 2003. v. 1, n. 1, p. 32.
- NASA. NASA Earth Observations (NEO), 2022. Disponível em: <https://neo.gsfc.nasa.gov/>.
- PEREIRA, F. d. M.; SOUZA, B. d. A.; LOPES, M. d. R. Criação de abelhas-sem-ferrão. In: . [S.l.]: Teresina: Embrapa Meio-Norte,[2017]., 2017.
- PRESERVAÇÃO de Abelha Sem Ferrão no Semiárido Através da Criação Racional. In: . [s.n.]. Disponível em: http://www.prac.ufpb.br/anais/lcbeu_anais/anais/meioambiente/abelhas.pdf.
- RODRIGUES, A. d. S. Etnoconhecimento sobre abelhas sem ferrão: saberes e práticas dos índios guarani mbyá na mata-atlântica. In: . [s.n.], 2005. Disponível em: <https://teses.usp.br/teses/disponiveis/91/91131/tde-15072005-150814/pt-br.php>.
- ROKACH, L.; MAIMON, O. Decision trees. In: _____. **Data Mining and Knowledge Discovery Handbook**. Boston, MA: Springer US, 2005. p. 165–192. ISBN 978-0-387-25465-4. Disponível em: https://doi.org/10.1007/0-387-25465-X_9.
- SANTOS, A. B. Abelhas nativas: polinizadores em declínio. In: . [S.l.: s.n.], 2010. v. 8, n. 3, p. 103–106.
- SANTOS, C. F. dos *et al.* Diversidade de abelhas sem ferrão e seu uso como recurso natural no Brasil: permissões e restrições legais consorciadas a políticas públicas. **Revista Brasileira de Meio Ambiente**, v. 9, n. 2, 2021.

SANTOS, G. G. *et al.* A meliponicultura atua na conscientização ambiental e incrementa a renda de comunidades carentes no maranhão / meliponicultura works to raise environmental awareness and increase the income of poor communities in maranhão. In: . [s.n.], 2020. v. 6, n. 8, p. 63578–63585. Disponível em: <https://www.brazilianjournals.com/index.php/BRJD/article/view/15888/13034>.

SCOTT, M. 2008. Disponível em: https://www.primevideo.com/detail/0S7YJ4PTPWHP2C3ZSIU2BI4F9D/ref=atv_dp_season_select_s5?language=pt_BR.

SENKOVSKI, A. C. **Mel de abelhas sem ferrão ganha mercado e vale “preço de ouro” - A.B.E.L.H.A.** 2017. Disponível em: <https://abelha.org.br/mel-de-abelhas-sem-ferrao-ganha-mercado-e-vale-preco-de-ouro/>.

SILVA, G. R. d. *et al.* Pesquisas com abelhas-sem-ferrão (hymenoptera: Meliponini) e aplicabilidade dos marcadores moleculares: Uma revisão sistemática da literatura. **Pubvet**, v. 13, n. 1, p. 1–19, Jan 2019. Disponível em: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/207236/1/PesquisasAbelhaSemFerraomarcadoresMolecularesFabiaMello.pdf>.

SULLA-MENASHE, D.; FRIEDL, M. **User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product.** [s.n.], 2018. Disponível em: https://modis-land.gsfc.nasa.gov/pdf/MCD12Q1_C6_Userguide04042018.pdf.