



**NICOLAS PEREIRA DE SOUZA**

**USO DE AERONAVES REMOTAMENTE PILOTADAS NA  
DETECÇÃO DO BICHO MINEIRO NOS CAFEEIROS**

**LAVRAS-MG  
2022**

**NICOLAS PEREIRA DE SOUZA**

**USO DE AERONAVES REMOTAMENTE PILOTADAS NA DETECÇÃO DO BICHO  
MINEIRO NOS CAFEEIROS**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Engenharia Agrícola, área de concentração em Máquinas e mecanização para a obtenção do título de Mestre.

Prof. Dr. Fabio Moreira da Silva  
Orientador

**LAVRAS-MG  
2022**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da  
Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a)  
autor(a).**

Souza, Nicolas Pereira de.

Uso de aeronaves remotamente pilotadas na detecção do bicho mineiro nos cafeeiros / Nicolas Pereira de Souza. - 2022.

64 p. : il.

Orientador(a): Fabio Moreira da Silva.

Coorientador(a): Gabriel Araújo e Silva Ferraz.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2022.

Bibliografia.

1. Agricultura de precisão. 2. python. 3. Geoprocessamento. I. Silva, Fabio Moreira da. II. Silva Ferraz, Gabriel Araújo e. III. Título.

**NICOLAS PEREIRA DE SOUZA**

**USO DE AERONAVES REMOTAMENTE PILOTADAS NA DETECÇÃO DO BICHO  
MINEIRO NOS CAFEEIROS**

**USE OF REMOTELY PILOTTED AIRCRAFT IN THE DETECTION OF THE  
BICHO MINEIRO IN THE COFFEE TREES**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Engenharia Agrícola, área de concentração em Máquinas e mecanização para a obtenção do título de Mestre.

APROVADA em 27 de abril de 2022  
Dr. Fabio Moreira da Silva UFLA  
Dr. Marcelo de Carvalho Alves UFLA  
Dr. Kalill José Viana da Páscoa UFLA

Prof. Dr. Fabio Moreira da Silva  
Orientador

**LAVRAS-MG  
2022**

## AGRADECIMENTOS

À Universidade Federal de Lavras, por me ceder essa oportunidade, onde considero um privilégio ter conhecido Lavras e estudado aqui.

Ao Departamento de Engenharia Agrícola (DEA), e ao Programa de Pós-Graduação em Engenharia Agrícola da UFLA, pela oportunidade de realização do mestrado.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig) pela concessão da bolsa de estudos.

Ao professor Dr Fábio Moreira da Silva, por ter sugerido esse tema de elevada complexidade e por ter depositado tamanha confiança, além de compartilhar da sua vasta experiência de vida em forma de incríveis conselhos, além dos desafios superados quero levar comigo sua amizade como meu orientador e também um pouco de sua visão de mundo que sempre se apresentou de forma serena, focada e incrivelmente analítica.

Ao professor Dr. Gabriel Araújo e Silva Ferraz, por acreditar no potencial, e com sua paciência e otimismo me fazer acreditar e trilhar essa jornada.

Ao professor Dr. Alexandre dos Santos, por estar sempre disposto, e com sua capacidade de raciocínio extremamente acelerado me possibilitar ter abertura para falar de uma forma simples e clara das dificuldades.

Ao meu antigo professor Dr. na graduação Fausto Weimar Acerbi Júnior, “meu grande incentivador, que me ajudou a trilhar este caminho”.

À minha querida Esposa Marilaydy Martinez Hernandez, por estar sempre ao meu lado nos momentos bons e frustrantes e acima de tudo acreditando no meu potencial. É uma honra te ter ao meu lado.

A minha linda e encantadora filha, Luna Mayara de Souza Martinez, por estar sempre trazendo alegria e orgulho. É uma imensa gratidão acompanhar seu crescimento.

Aos amigos Hailton Carlos Pereira Souza, Vinicius Torres Ribeiro, Victor Augusto Torre Ribeiro o “Ranger Branco” e Jonatas Morais Puschmann meus agradecimentos pelo companheirismo e por todo real apoio desses anos.

Aos professores e funcionários do PPGEA/UFLA que contribuíram para este momento. Agradeço.

Enfim, a todos que não foram citados, mas que de alguma forma direta ou indiretamente, contribuíram para que eu chegasse até aqui.

## RESUMO

O sensoriamento remoto (SR) é uma técnica, cujo princípio fundamental consiste em obter dados de um objeto sem necessariamente ter o contato direto, é considerada uma excelente técnica para agricultura de precisão (AP), uma vez que pode detectar fenômenos no plantio, em um curto espaço de tempo e abranger grandes áreas se comparado aos métodos convencionais de campo. Portanto, detectar pragas como bicho-mineiro-do-cafeeiro (*Leucoptera coffeella*) se tornam um grande desafio, uma vez que é um problema fitossanitário do cafeeiro e que tem importância para o agronegócio brasileiro. Neste contexto, o objetivo é o monitoramento para sua detecção, utilizando uma aeronave remotamente pilotada (RPAs) com câmeras RGB e obtendo imagens de alta resolução combinadas com algoritmos de aprendizado de máquina (*machine learning*). O experimento foi conduzido na fazenda denominada Barro preto, localizada na cidade de Lavras, estado de Minas Gerais, Brasil. Os resultados alcançados foram a geração de 4 modelos derivado de *machine learning*, sendo o modelo ‘*ranger*’ derivado do *Random Forest* o de maior destaque, com acurácia de 90% em detectar a presença da infestação de bicho-mineiro por meio de imagem de RGB, com acerto de sensibilidade 88% de detectar o objeto em campo e especificidade de 91%, demonstrando que existe possibilidade real de fazer a gestão e manejo da lavoura cafeeira, através de imagens RGB, de baixo custo, com relação a infestação de bicho mineiro.

**Palavras-chave:** Agricultura de precisão. Auto ML Forester. Geoprocessamento. Opencv. Python.

## ABSTRACT

Remote sensing (SR) is a technique whose fundamental principle is to obtain data from an object without necessarily having direct contact. short time and cover large areas compared to conventional field methods. Therefore, detecting pests such as the coffee leaf miner (*Leucoptera coffeella*) becomes a great challenge, since it is a phytosanitary problem of coffee plants and is important for Brazilian agribusiness. In this context, the objective is monitoring for its detection, using a remotely piloted aircraft (RPAs) with RGB cameras and obtaining high resolution images combined with machine learning algorithms. The experiment was carried out on the farm called Barro preto, located in the city of Lavras, state of Minas Gerais, Brazil. The results achieved were the generation of 4 models derived from machine learning, being the 'ranger' model derived from Random Forest with 90% accuracy in detecting the presence of leaf miner infestation through RGB image, with sensitivity adjustment 88% of detecting the object in the field and specificity of 91% demonstrating that there is a real possibility of managing and managing the coffee crop, through low-cost RGB images, in relation to the miner infestation.

**Keywords:** Precision agriculture. Auto ML Forestry. Geoprocessing. Opencv. Python.

## LISTA DE FIGURAS

Figura 1 - Estrutura de árvore de decisão .....	21
Figura 2 - Estrutura de funcionamento do AUTOML.....	25
Figura 3 - Localização da área de estudo .....	27
Figura 4 - Fluxograma de atividades de cada etapa da metodologia.....	28
Figura 5 - Equipamentos de voo.....	29
Figura 6 - Imagem do <i>software</i> DroneDeploy (2021).....	30
Figura 7 - <i>Software</i> UAV forecast para verificar as condições do voo .....	30
Figura 8 - Fluxograma de atividades de cada etapa do processamento pix4d.....	32
Figura 9 - Visualização de etapa de calibração e criação de nuvem no pix4d .....	32
Figura 10 - Visualização do Ortomosaico e fim do processamento do pix4d.....	33
Figura 11 - Segmentação das imagens .....	33
Figura 12 – Exemplo de Vetorização das imagens .....	34
Figura 13 - Exemplo de banda padronizada .....	36
Figura 14 - Amostragem em campo das linhas no dia 06 de agosto 2021 .....	36
Figura 15 - Amostragem em campo das linhas no dia 08 de agosto 2021 .....	37
Figura 16 – Marcação dos pontos derivados das plantas com o <i>software</i> <i>avenza maps</i> .....	37
Figura 17 - Extração dos pixels .....	38
Figura 18 – Gráficos importância de recursos.....	45



## LISTA DE TABELAS

Tabela 1 - Parâmetros selecionados para o plano de voo .....	31
Tabela 2 - Voos realizados .....	31
Tabela 3 - Índices espectrais.....	34
Tabela 4 - Níveis de severidade e descrição de infestação do bicho mineiro .....	38
Tabela 5 - Dados organizados .....	39
Tabela 6 - Dados organizados com a retiradas das variáveis x e y .....	39
Tabela 7 – Métricas de teste do modelo de detecção na detecção de bicho mineiro em lavoura de café, a partir de imagens RGB adquiridas por sensor embarcado em aeronave remotamente pilotada (RPA) .....	42
Tabela 8 – Parâmetros de teste na detecção de bicho mineiro em lavoura de café, a partir de imagens RGB gerada por aeronave remotamente pilotada (RPA) .....	43
Tabela 9 – Resultados das métricas, a partir de imagens RGB gerada por aeronave remotamente pilotada (RPA) .....	44

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>10</b>
<b>1.1</b>	<b>Objetivos</b> .....	<b>11</b>
<b>1.1.1</b>	<b>Geral</b> .....	<b>11</b>
<b>1.1.2</b>	<b>Específicos</b> .....	<b>11</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> .....	<b>12</b>
<b>2.1</b>	<b>Características gerais do cafeeiro</b> .....	<b>12</b>
<b>2.2</b>	<b>Bicho-mineiro (<i>Leucoptera coffeella</i>)</b> .....	<b>13</b>
<b>2.3</b>	<b>Monitoramento e nível de controle para Bicho-mineiro (<i>Leucoptera coffeella</i>)</b> .....	<b>14</b>
<b>2.4</b>	<b>Agricultura de Precisão e sua contribuição na cafeicultura</b> .....	<b>15</b>
<b>2.5</b>	<b>Sensoriamento remoto</b> .....	<b>16</b>
<b>2.6</b>	<b>Aspecto gerais de aeronaves remotamente pilotadas e sua contribuição para agricultura</b> .....	<b>18</b>
<b>2.7</b>	<b>Técnicas de aprendizado de máquinas Regressões Lineares</b> .....	<b>19</b>
<b>2.8</b>	<b>Algoritmo de aprendizado Floresta Aleatória (<i>Randon Forest</i>)</b> .....	<b>20</b>
<b>2.9</b>	<b>Métricas de avaliação</b> .....	<b>22</b>
<b>2.10</b>	<b>Seleção de características</b> .....	<b>24</b>
<b>2.11</b>	<b>Software R</b> .....	<b>24</b>
<b>2.12</b>	<b>Auto ml forester</b> .....	<b>25</b>
<b>2.13</b>	<b>Pacote dalex do R</b> .....	<b>26</b>
<b>3</b>	<b>MATERIAL E MÉTODOS</b> .....	<b>27</b>
<b>3.1</b>	<b>Área de estudo</b> .....	<b>27</b>
<b>3.2</b>	<b>Obtenção de dados</b> .....	<b>27</b>
<b>3.3</b>	<b>Equipamentos para obtenção imagens</b> .....	<b>29</b>
<b>3.4</b>	<b>Planejamento de voo, parametros e datas de sobrevoos</b> .....	<b>29</b>
<b>3.5</b>	<b>Processamento de imagens</b> .....	<b>31</b>
<b>3.6</b>	<b>Segmentação dos indivíduos e vetorização</b> .....	<b>33</b>
<b>3.7</b>	<b>Aplicação de índices espectrais</b> .....	<b>34</b>
<b>3.8</b>	<b>Amostragens em campo</b> .....	<b>36</b>
<b>3.9</b>	<b>Entrada de dados software R</b> .....	<b>38</b>
<b>3.10</b>	<b>Construção dos modelos de dados</b> .....	<b>40</b>
<b>3.11</b>	<b>Parâmetros utilizadas para o modelo</b> .....	<b>40</b>
<b>3.12</b>	<b>Validando estatisticamente as amostras de treinando do algoritmo</b> .....	<b>40</b>
<b>3.13</b>	<b>Treinando as amostras</b> .....	<b>40</b>
<b>3.14</b>	<b>Testes de predição dos dados</b> .....	<b>40</b>
<b>3.15</b>	<b>Métricas utilizadas</b> .....	<b>41</b>
<b>3.16</b>	<b>Matriz de confusão</b> .....	<b>41</b>
<b>3.17</b>	<b>Contribuição de cada índice no modelo</b> .....	<b>41</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	<b>42</b>
<b>4.1</b>	<b>Resultado da construção dos modelos de dados</b> .....	<b>42</b>
<b>4.2</b>	<b>Resultados dos parâmetros utilizadas para o modelo</b> .....	<b>43</b>
<b>4.3</b>	<b>Resultados da matriz de confusão</b> .....	<b>44</b>
<b>4.4</b>	<b>Gráficos do comportamento do processo</b> .....	<b>45</b>
<b>5</b>	<b>CONCLUSÃO</b> .....	<b>47</b>

<b>REFERÊNCIAS .....</b>	<b>48</b>
<b>APÊNDICE A - Código fonte .....</b>	<b>58</b>

## 1 INTRODUÇÃO

O sensoriamento remoto (SR) é uma técnica, cujo princípio fundamental consiste em se obter dados de um objeto sem necessariamente ter o contato direto, é uma excelente técnica para agricultura de precisão (AP), uma vez que pode detectar fenômenos no plantio, em um curto espaço de tempo e abranger grandes áreas se comparado aos métodos convencionais de campo, auxiliando assim nas tomadas de decisões estratégicas do manejo da cultura. As imagens das áreas de interesse, podem ser detectadas utilizando diferentes sensores acoplados em plataformas, que se diferem de acordo com sua altitude, como satélites geoestacionários, balões dirigíveis, aviões, helicópteros, aeronaves de pequeno, médio ou grande porte pilotadas por controle remoto, entre outras, que podem ser utilizadas para coleta de dados.

A altitude é responsável diretamente pela resolução espacial, ou seja, quanto menor altura em relação ao objeto alvo melhor é o nível de detalhamento dessa imagem. Essas imagens por sua vez são imprescindíveis, pois além de detalhar a cultura alvo pode detectar anomalias em grandes áreas. Em comparação com o trabalho dos monitores de campo, profissionais responsáveis por identificar sinais de ataques de pragas ou doenças, essa atividade se torna mais árdua, pois dependem de percorrer toda a área de forma terrestre. Sendo assim, o tempo se torna chave para o sucesso da atividade, uma vez que quanto mais tempo for necessário para a atividade de identificar a infestação maior a probabilidade do crescimento da mesma.

A adoção de técnicas de SR utilizando mapeamento com imagens de alta resolução tem apresentados ganhos consideráveis de precisão, conforme a literatura científica vem demonstrando, isto ainda é potencializado quando técnicas são combinados com aprendizado de máquinas (*machine learning*), tais como algoritmos de *Support Vector Machine* (SVM), *Random Forest* (RF), Classificação Naïve Bayes (NB), entre outros. Isso tem possibilitado aos cientistas extrair padrões de dados para melhor entendimento das culturas, inclusive relacionadas aos ataques de pragas.

Nesse contexto, de monitoramento e detecção de ataques e sintomas de pragas na lavoura, o uso de Aeronaves Remotamente Pilotadas (RPA), conhecidas também como “*drone*” ou “*Vant*” (Veículo Aéreo Não Tripulado), podem ser usados em conjunto com aplicação de índices espectrais e treinamento de amostras, uma vez que podem captar dados e informações através de seus sensores embarcados, como câmeras e sensores multiespectrais. Portanto, esses equipamentos são mais acessíveis e possibilitam realizar missões programáveis, evitando problemas com nuvens e questões temporais. Estas aeronaves podem ser de diferentes modelos, asas fixa ou rotativa, o que implicará necessariamente na capacidade de imageamento e

autonomia do voo.

Portanto, o uso de RPA cada dia mais vem se consolidando nos processos de AP, graças a facilidade de obtenção de dados em campo. Pesquisas já realizadas para as culturas de milho, soja e cana de açúcar mostraram-se eficientes e ressaltaram as potencialidades desta tecnologia.

Entretanto, para a cafeicultura, as pesquisas ainda estão sendo realizadas, mas ainda existe uma grande demanda para compressão de fenômenos associados ataque de pragas, como a do bicho-mineiro-do-cafeeiro (*Leucoptera coffella*). O bicho-mineiro-do-cafeeiro é um problema fitossanitário do cafeeiro que é uma cultura de valor comercial agregado, importante para o agronegócio brasileiro.

Em vista, das peculiaridades desta praga, como época de reprodução, forma de amostragem e padrões da infestação, se faz necessário novas técnicas de monitoramento para sua detecção, e neste sentido o objetivo do trabalho foi detectar o ataque do bicho-mineiro-do-cafeeiro em imagens de alta resolução obtidas por RPA com o emprego de algoritmos de aprendizado de máquina (*machine learning*).

## **1.1 Objetivos**

### ***1.1.1 Geral***

Utilizar técnicas de sensoriamento remoto e geoprocessamento para detectar a presença de bicho-mineiro-do-cafeeiro.

### ***1.1.2 Específicos***

- a) Utilizar espectros RGB de câmera para detecção de bicho-mineiro-do-cafeeiro;
- b) Avaliar a efetividade do uso de *machine learning* para detectar a presença ou não do bicho-mineiro-do-cafeeiro.

## 2 REFERENCIAL TEÓRICO

### 2.1 Características gerais do cafeeiro

O cafeeiro (*Coffea* sp.) pertencente à família Rubiaceae, é uma planta de sub-bosque de porte baixo que apresentam sucessivas fases vegetativas e reprodutivas de aproximadamente 2 anos de ciclo fenológico. No primeiro ano, formam-se os ramos vegetativos, com gemas axilares nos nós, que por foto periodismo será transformada em gemas reprodutivas (GOUVEIA, 1984), e depois de alguns meses, essas gemas são induzidas a gemas foliares consequentemente essas gemas florais vão amadurecer e quando maduras entram em dormência e ficam prontas para a antese, quando ocorre um aumento substancial do potencial hídrico nas gemas dormentes (CAMARGO; CAMARGO, 2001). O segundo ano fenológico é marcado pela florada, formação dos “chumbinhos”, que precedem a expansão dos grãos, depois vem a fase de granação dos frutos e pôr fim a maturação do grão. Finalmente advém a senescência, morte dos ramos plagiotrópicos terminais e recomeça o ciclo.

O café possui elevada importância para o agronegócio brasileiro, segundo Cunha *et al.* (2017) o setor é responsável por movimentar mais de 35 bilhões de dólares anualmente à nível mundial, além de que o Brasil possui condições geográficas, climáticas e de dimensão territorial adequadas para a cultura, contando com incentivos à pesquisa, que geram resultados significativos em produtividade e qualidade (CONSÓRCIO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO DO CAFÉ - CBP&D/CAFÉ, 2021). Os principais estados produtores no país são Minas Gerais, Espírito Santo, São Paulo, Paraná, Bahia e Rondônia (CUNHA *et al.*, 2017). Atualmente, o Brasil é o maior produtor mundial de café é o segundo maior mercado consumidor, atrás somente dos Estados Unidos (ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ – ABIC, 2021).

Por ter valor agregado e ser um dos principais produtos agrícolas, e ainda com potencial de crescimento, também existem fatores que podem limitar seu desenvolvimento, como ataques de pragas, doenças, bienalidade, clima, entre outros. Dentro dos fatores limitantes, pragas são sem dúvida um dos maiores agentes envolvidos em especial no Brasil, destaque-se o ataque do bicho mineiro do cafeeiro (*Leucoptera coffeella*). Logo compreender sua dinâmica em campo auxilia a tomada de decisões mais eficientes para o seu combate.

## 2.2 Bicho-mineiro (*Leucoptera coffeella*)

O cafeeiro pode ter um número elevado de artrópodes pragas, que atacam partes diferentes da planta, sendo responsáveis por danos diretos aos frutos, ou mesmo indiretos as plantas, as quais podem acarretar perdas na produção e na qualidade do produto (MARQUES, 2017). No Brasil, as mais importantes, pragas do cafeeiro são: o bicho mineiro do cafeeiro (GUÉRIN-MÉNEVILLE, 1842) (Lepidoptera: *Lyonetiidae*), e a broca do café *Hypothenemus hampei* (Coleóptera: *Scolytidae*) (MATIELLO *et al.*, 2010). De acordo com Reis, Souza e Venzon (2002) os ácaros vermelhos, *Oligonychus ilicis* (Tetranychidae) e plano, *Brevipalpus phoenicis* (Tenuipalpidae) são também considerados pragas primárias.

Sua origem é provinda do continente africano e suas primeiras aparições no Brasil data de 1851 (MESQUITA *et al.*, 2016). O inseto adulto é um microlepidóptero, mariposa possui 6,5 mm de envergadura e 2,2 mm de comprimento de coloração branco prateado, com manchas circulares pretas e halo amarelo nas extremidades das asas anteriores, tendo maior presença em regiões cafeeiras, com elevadas temperaturas e de maior déficit hídrico do Brasil (REIS; SOUZA, 1986; REIS; SOUZA; VENZON, 2002; SOUZA; REIS; RIGITANO, 1998; MACHADO *et al.*, 2014; SILVA *et al.*, 2014).

Possui hábito crepuscular noturno (SOUZA; REIS; RIGITANO, 1998), com preferência para voar entre 16 e 20 horas (CONCEIÇÃO, 2005). Durante o dia, as mariposas ocultam-se sob as folhas dos cafeeiros na metade inferior das plantas, procurando ambientes de microclima mais úmido. No crepúsculo, as fêmeas realizam a postura endofítica na parte adaxial das folhas.

Uma fêmea, durante a sua vida, é capaz de colocar mais de 50 ovos, em um período embrionário variando de 5 a 21 dias, dependendo das condições de temperatura e umidade (SOUZA; REIS; RIGITANO, 1998). Logo após a eclosão, as lagartas perfuram a cutícula e a epiderme superior da folha que se encontra em contato com o ovo, penetrando diretamente no parênquima paliçádico onde permanecem se alimentando de células desse tecido até o abandono para transformação em crisálida. Em uma única lesão, podem ser encontradas uma ou mais lagartas e a presença de mais de uma lagarta deve-se à coalescência das lesões (REIS; SOUZA; MELLES, 1984). As regiões destruídas vão secando e a área atacada vai aumentando com o próprio desenvolvimento da lagarta. É no período larval, portanto, que o bicho-mineiro causa injúrias às plantas (SOUZA; REIS; RIGITANO, 1998). As minas têm formato arredondado e coloração castanho-clara, com o centro das lesões mais escuro, resultado do acúmulo de excreções; a epiderme do limbo superior, no local da lesão, destaca-se com facilidade (REIS; SOUZA, 1998).

Este inseto é favorecido, por sistemas agrícolas em condições climáticas de baixa umidade relativa e temperatura elevada. As injúrias são provocadas pelas lagartas e estas afetam a produtividade devido à redução da área foliar, e em altas infestações provocam desfolha nas plantas. Na região Sul de Minas Gerais foi constatado uma redução de 50% na produção devido a 67% de desfolha ocorrida em outubro, época de floração do cafeeiro (REIS; SOUZA, 2002). O bicho-mineiro se alimenta de área foliar da planta, o que causa pequenas minas nas folhas gerando o nome empregado pelos cafeicultores, essas minas nas folhas são geradas pela forma de lagartas, transformando-se em áreas necrosadas e conseqüentemente mortas, com a área fotossintetizadora menor existe a queda e grande desfolha, nos períodos mais prolongados de seca do ano.

Alguns autores como Caixeta *et al.* (2004), citam em seu trabalho que a intensidade dos ataques da *L. coffella* ao cafeeiro podem estar associadas, em função do estado fisiológico que implica na nutrição da planta, além de outros fatores como: clima, densidade de plantio, tratos culturais, presença de predadores e parasitoides.

Sendo assim, esse tipo de infestação é controlável desde que não ultrapasse 30% de ataque nas folhas, aonde é detectados minas com lagarta vivas, quando o ataque chega a proporções de 40% é uma fase de predação, o que indica que nesse caso não pode ser mais controlada.

Diante de sua importância no cultivo de café, o manejo populacional de bicho-mineiro é imprescindível para a redução dos danos causados, o seu controle é geralmente realizado por meio de inseticidas, entretanto, tem-se buscado o manejo integrado de pragas (MIP), que além de reduzir o uso de químicos em cultivos, promove o monitoramento eficiente de populações de pragas, a fim de determinar o momento correto das aplicações (BACCA *et al.*, 2006).

A partir do monitoramento eficiente, é possível conhecer a variabilidade espacial das populações de pragas, e com essa informação pode ser elaborado e executados programas de amostragem para o manejo correto, pois permite a identificação da dinâmica populacional, possibilitando tomada de decisão fundamentada em maior número de informações. Todavia, ainda são poucos os estudos sobre monitoramento e distribuição espacial do bicho mineiro em campo, os quais são peças chaves para o desenvolvimento de um monitoramento e controle mais eficiente (RIFFEL *et al.*, 2012).

### **2.3 Monitoramento e nível de controle para Bicho-mineiro (*Leucoptera coffeella*)**

O monitoramento da população da praga em campo é necessário para auxiliar as



tomadas de decisão em relação ao controle, sendo assim, a amostragem de insetos se torna uma forte ferramenta nessa etapa, pois a partir dela será feito, aplicação ou outro tipo de controle (SANTOS; GARCIA; ALVES, 1995).

Sabe-se que o bicho mineiro, é influenciado por fatores climáticos, presença ou não de plantas daninhas, ou parasitoides e predadores, pela condução das lavouras (lavouras mais arejadas tendem a favorecer o ataque da praga), aplicação de fungicidas cúpricos, ciclo bienal do cafeeiro e problemas nutricionais, entre outros. Essa combinação de fatores agindo isolados ou em conjunto, poderão determinar maiores ou menores infestações da praga em determinadas áreas da plantação e em certos períodos do ano (REIS; SOUZA, 1998).

Deste modo, o conhecimento da dinâmica do padrão espacial e temporal do bicho-mineiro, em cultivos de cafeeiro pode ser importante, no contexto do controle biológico dessa praga, para um manejo mais preciso, tendo em vista a produção mais sustentável.

Face ao exposto, ainda existem dúvidas com relação a melhor maneira de amostragem em campo do bicho-mineiro, pois as pesquisas são insuficientes para avaliar a infestação em relação a sua área real e em um tempo hábil.

Uma das possíveis soluções para essa lacuna é o emprego de ferramentas de sensoriamento remoto em diferentes níveis orbitais (orbital e suborbital), técnicas que atualmente podem ser chamadas de agricultura de precisão (AP). Os valores de reflectância aliados com os índices espectrais, podem indicar tendências, comparar comportamentos de forma temporal e espacial, como vigor ou estresse, identificar falhas e inferir se existe o ataque do bicho-mineiro, mitigando assim as limitações do campo.

#### **2.4 Agricultura de Precisão e sua contribuição na cafeicultura**

Agricultura de precisão pode ser definido como uma tecnologia de informações empregada em práticas agrícolas para auxiliar o tratamento de variabilidade espacial da lavoura (BERNARDI *et al.*, 2015), na qual compreende as etapas de coleta de dados, análise e interpretação, geração de recomendações, aplicação no campo e avaliação dos resultados (GEBBERS; ADAMCHUK, 2010).

O uso de técnicas comumente vista na ciência de sensoriamento remoto, estão cada vez mais se popularizando dentro de AP, em especial as pesquisas onde o foco está na cultura do café. Por exemplo, (Moreira *et al.*, 2019) que aborda a caracterização da cultura por meio de uso de sensoriamento remoto, detecção de anomalias utilizando imagens do satélite Sentinel 2, inovações para o manejo sustentável de pragas e doenças (VENZON *et al.*, 2021), entre outros.

Contudo, ainda assim, existem poucos trabalhos que tratem de ataques de pragas e ainda menos quando o tema é combinado com bicho mineiro no cafeeiro e RPAs. Segundo em um trabalho de revisão de literatura sobre as pesquisas desenvolvidas sobre café, ainda o uso de RPAs para o monitoramento de cafezais, são muito poucos utilizados, sendo que esse tipo de tecnologia teve início nas pesquisas a partir do ano 2006. Porém, Silva e Alves (2013) com um livro onde aborda “Cafeicultura de Precisão”, demonstram que o cenário pode está mudando, pois com a capacidade de embarcar sensores RGB e multiespectrais é possível obter o mapeamento, com uma enorme redução de tempo e poder gerar produtos de altíssimo nível. Além de outras aplicações pontuais e precisas, como identificação de doenças, pragas, desfolha, que podem ser estudadas para o avanço da cafeicultura.

## 2.5 Sensoriamento remoto

O sensoriamento remoto (SR) é definido por Fitz (2008) como técnica que utiliza sensores para a captação e registro à distância, sem o contato direto, da energia refletida ou absorvida pela superfície terrestre. Além do sensoriamento remoto existem também geotecnologias utilizadas na AP que auxiliam a detecção de variabilidade espacial de campo sendo os Sistemas de Informação Geográfica (SIGs); Sistema Global de Navegação por Satélite (GNSS); sensores de: solo, planta, produtividade; dispositivos de aplicação à taxa variável; entre outros que podem auxiliar na obtenção de dados e dar suporte localizado no manejo agrícola (LUANA *et al.*, 2020).

Agricultura sem dúvida possui uma dinâmica ímpar que exige um constante monitoramento com objetivo de identificar e antecipar possíveis problemas. O SR amplia consideravelmente as possibilidades desse monitoramento, pois com o uso de satélites a nível orbitais aquele cujo os sensores utilizados para captar informação encontram-se a bordo de satélites que podem prover imagens de média, ou alta resolução espacial. As imagens de média resolução são majoritariamente utilizados na agricultura de precisão, sendo provindas das imagens Landsat 8, que por sua vez, são gratuitas e contribuem em programas de monitoramento agrícola, para previsões de safra, detecção de ataque de ferrugem (*Hemileia vastatrix*) e alterações na refletância espectral para detecção de pragas.

Apesar do SR possuir muitas vantagens, existem algumas situações que sua aplicação é prejudicada como a obtenção de dados contínuos para previsão de safra, análise de déficit nutricional, acompanhamento sistemático de crescimento da cultura, detecção de caso emergências de acompanhamento de culturas, entre outros. Existe também imagens derivadas

de satélites pagos, cujo sua resolução espectral tem ganhos em cm quando comparados com as imagens de satélite gratuitas de resolução em metros, o que implica limitações custos elevados, falta de flexibilidade operacional e baixa resolução espacial e temporal (WHITEHEAD; HUGENHOLTZ, 2014). Outro fator que interfere na aquisição de imagens em nível orbital, são as condições climáticas, em dias nublados, pois há o impedimento da passagem da energia solar e consequentemente a perda de informação de dados da superfície (HONKAVAARA *et al.*, 2013).

Com ocorrências dessas situações relatadas, algumas alternativas são testadas no meio da ciência de SR, como emprego de plataformas dirigíveis que auxiliem na obtenção de dados remotos, como por exemplos dirigíveis cheios de hélio existem algumas iniciativas de criar projetos nacionais como AIRSHIP do Brasil no intuito de aplicar essa tecnológica em diferentes áreas, como meio ambiente, levantamento de dados, logística.

Entretanto ainda existem muitos empecilhos para real aplicação desse tipo de tecnologia, como áreas industriais para construção do equipamento, leis para regulamentação, real demanda do setor por esse tipo de modal e com isso operacionalmente para aplicação na agricultura acaba sendo inviável. Outra aplicação são uso de balões que atualmente estão mais direcionados para obtenção de dados meteorológicos, sendo que já foram testados 1859 segundo em projetos militares para tomadas de posições inimigas por meio de várias câmaras convenientemente fixadas ao balão.

Portanto, com aderência cada dia maior das RPA (aeronaves remotamente pilotadas), no cotidiano, existe forte potencial para uso em AP, pois promovem as seguintes vantagens: baixo custo em relação às plataformas orbitais; capacidade de realizar missões; capacidade de operar em condições adversas; segurança ao piloto (WHITEHEAD; HUGENHOLTZ, 2014).

Sendo assim, as RPA estão sendo estudadas e utilizadas para obtenção de imagens de alta resolução temporal (adquiridas várias vezes por dia), resolução espacial (em centímetros e até em milímetros) além de baixos custos operacionais comparados com aeronaves tripuladas e principalmente os satélites de alta resolução espacial (HARDIN; HARDIN, 2010; HONKAVAARA *et al.*, 2013; HUNT *et al.*, 2005; LALIBERTE; RANGO, 2011; LELONG; PINET; POILVE´, 1998; NEBIKER *et al.*, 2008; RANGO *et al.*, 2009; TORRES-SÁNCHEZ *et al.*, 2014; XIANG; TIAN, 2011b). Podendo ser aplicadas em áreas menores e em locais pontuais com facilidade de obtenção de dados em menor tempo, acompanhando o crescimento de diversas culturas.

## 2.6 Aspecto gerais de aeronaves remotamente pilotadas e sua contribuição para agricultura

Nos últimos anos principalmente a partir de 2013 as aeronaves remotamente pilotadas, destacaram-se por possuir inúmeras vantagens, assim fornecendo um novo meio de obtenção de dados na agricultura e na área ambiental. Os sistemas de aeronaves não tripuladas, assumem duas nomenclaturas normalmente, aeronave remotamente pilotada (RPA) ou de veículo aéreo não tripulado (Vant), e graças a sua grande flexibilidade de manuseio, operação, alta resolução espacial, espectral, temporal e baixo custo de aquisição, e manutenção e também rápida aquisição de dados (ZHANG; KOVACS, 2012; MULLA, 2013; ZHANG; WALTERS; KOVACS, 2014; DENG *et al.*, 2018).

As RPA estão se tornando uma alternativa de aquisição de informação, de dados e de baixo custo quando comparados com satélites de alta resolução ou aeronaves tripuladas. Existem na literatura ainda outras formas de designar as RPA sendo Unmanned Aerial Vehicles (UAV) (CANDIAGO *et al.* 2015; GRENZDÖRFFER; ENGEL; TEICHERT, 2008; LÓPEZ-GRANADOS *et al.*, 2016; PEÑA *et al.*, 2013; TOKEKAR *et al.*, 2016), Remotely Piloted Aircraft Systems (RPAS) (CRISTEA *et al.*, 2016), Remotely Piloted Aircraft (RPA) (GILES, 2016; ZAJKOWSKI *et al.*, 2016; (HARDIN; HARDIN, 2010; SIEBERT; TEIZER, 2014), Unmanned Aircraft Systems (UAS) (WHITEHEAD; HUGENHOLTZ, 2014; Remotely Operated Aircraft (ROA) (UYSAL; TOPRAK; POLAT, 2015), Vant (Veículos Aéreos Não Tripulados) (MEDEIROS *et al.*, 2008).

No contexto da AP, estudos com a utilização de imagens de RPA estão demonstrando grandes avanços, e contribuindo significativamente, para os entendimentos da variabilidade espacial das culturas, pois graças as técnicas aplicadas de tecnologia da informação, combinadas com múltiplas fontes de dados, tem proporcionado uma estratégia mais assertiva nos fatores que influenciam as decisões associada à produção agrícola (CANDIAGO *et al.*, 2015). As múltiplas fontes de dados são derivadas de técnicas de sensoriamento remoto que envolvem a mensuração sem contato direto com o objeto, por meio da radiação refletida ou emitida de campos agrícolas, cuja intensidade depende de padrões climáticos e das propriedades das plantas que estão sendo investigadas (MOZGERIS *et al.*, 2018).

Os parâmetros mais aplicados atualmente nos estudos são, a transmitância de luz através de folhas, a refletância de luz na superfície das folhas e a fluorescência de clorofila e polifenóis (DEMOTES-MAINARD *et al.*, 2008). As principais áreas que atuam o sensoriamento remoto na AP são: mapeamento (TORRES-SÁNCHEZ *et al.*, 2015; SANTESTEBAN *et al.*, 2017;

ALLRED *et al.*, 2018), monitoramento; AGÜERA VEGA *et al.*, 2015; JANNOURA *et al.*, 2015; LONG; MCCARTHY; JENSEN, 2016; NORIEGA; ANDERSON, 2016), sensoriamento remoto (XIANG; TIAN, 2011a; TORRES-SÁNCHEZ *et al.*, 2013; ZARCO-TEJADA *et al.*, 2013; TOKEKAR *et al.*, 2016; ALSALAM *et al.*, 2017; ARROYO *et al.*, 2017; SANTESTEBAN *et al.*, 2017; ALLRED *et al.*, 2018). Contudo, ainda existem áreas, que não se aplicam o uso de RPA como semeadura, colheita. Entretanto, áreas de irrigação e detecção e controle de pragas estão aumentando (ALBORNOZ; GIRALDO, 2017; FAIÇAL *et al.*, 2017; ROMERO *et al.*, 2018).

Com isso, verifica-se o potencial da aplicação de RPA na AP, mas ainda existem lacunas na detecção de pragas e ataques e quando verificado estudos em cafeicultura essa lacuna se torna ainda maior. Uma possibilidade que vem se desenvolvendo para sanar esses problemas é o emprego de técnicas de aprendizado de máquinas (*Machine learning*) que utilizam de várias ferramentas estatísticas como regressões lineares, algoritmos como Floresta aleatória (Random Forest).

## 2.7 Técnicas de aprendizado de máquinas Regressões Lineares

O modelo de regressão linear é utilizado para realizar a medição ou previsão entre uma ou mais variáveis independente/explicativas e a variável dependente, geralmente são denotadas por  $x$  e  $y$  respectivamente. Podendo ser classificadas como simples ou múltiplas (Yue *et al.*, 2007).

Segundo alguns autores (MONTGOMERY; PECK; VINING, 2012) a regressão linear simples é composta por três elementos sendo a primeira variável dependente  $\beta_0$  conhecido com o intercepto, a segunda variável independente  $x$  representado por  $\beta_1$ , o pôr fim coeficiente angular representado por  $\varepsilon$  que é o erro aleatório Equação 2.1.

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (2.1)$$

Para encontrar os coeficientes  $\beta_0$  e  $\beta_1$ , utiliza-se o método dos mínimos quadrados. A Equação 2.2 apresenta o cálculo para encontrar o coeficiente  $\beta_1$ , em que  $n$  representa o número de amostras.

$$\beta_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad (2.2)$$

O coeficiente  $\beta_0$  pode ser calculado conforme Equação 2.3, em que  $\bar{y}$  refere-se à média dos valores de  $y$ , e  $\bar{x}$  a média dos valores de  $x$ . O cálculo dos coeficientes tem como objetivo minimizar a soma dos quadrados dos desvios (LAPPONI, 2005).

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad (2.3)$$

A regressão linear múltipla é aplicada quando há mais de uma variável explicativa ( $x$ ) no modelo, e pode ser descrita conforme Equação 2.4. Em que  $\beta_0$ ,  $\beta_1$  e  $\beta_n$  são os coeficientes de regressão e  $x_1$ ,  $x_2$  e  $x_n$ , são as variáveis explicativas e  $\varepsilon$  o erro aleatório (MONTGOMERY; PECK; VINING, 2012).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_n x_n + \varepsilon \quad (2.4)$$

O modelo de regressão linear múltipla pode ser reescrito em forma matricial, conforme Equação 2.5 (MORAIS, 2010). Em que  $Y$  é o vetor das observações,  $X$  é uma matriz das variáveis independentes,  $\beta$  um vetor dos coeficientes e  $\varepsilon$  o vetor dos erros.

$$Y = X\beta + \varepsilon \quad (2.5)$$

Assim como a regressão linear simples, a regressão múltipla também utiliza o método dos mínimos quadrados para estimar os coeficientes de regressão.

$$\beta = (X^T X)^{-1} X^T \quad (2.6)$$

## 2.8 Algoritmo de aprendizado Floresta Aleatória (*Random Forest*)

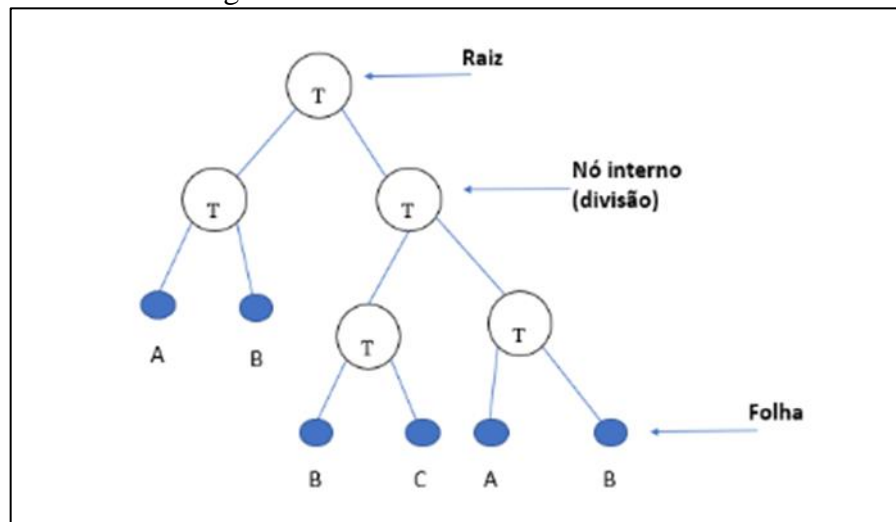
O algoritmo de Floresta aleatória (Random Forest-RF) é uma forma de aprendizado de máquina supervisionado, utilizado para resolver problemas de classificação e regressão, pois, possuem alto desempenho com apenas alguns parâmetros para ajustes (GENUER *et al.*, 2017).

O algoritmo funciona agrupando os resultados ou previsões, de várias árvores de decisões, treinadas individualmente, no intuito de criar o melhor modelo preditivo para resolver o mesmo problema, diminuindo a variância e o viés.

A floresta aleatória usa decisões finais, ou folhas, de cada nó para chegar a uma conclusão própria desta forma, cada folha, nós e árvores representam e respectivamente todos os atributos, possíveis valores para os atributos e as distintas classes. Portanto, o conjunto de

dados é classificado subdividindo-o sequencialmente de acordo com as estruturas de decisão definidas pela árvore (FIGURA 1) (FRIEDL; BRODLEY, 1997).

Figura 1 - Estrutura de árvore de decisão



Fonte: Adaptado de Friedl e Brodley (1997).

Com esse processo envolvido a precisão do modelo é potencializada, tendo em vista que está observando os resultados de muitas árvores de decisão diferentes e encontrando uma média. Sua aplicação pode ser feita em diferentes campos do conhecimento como sensoriamento remoto, diagnóstico médico por imagem, predição rápida de movimentos, etc. (BELGIU; DRAGUT, 2016; CRIMINISI *et al.*, 2013; GENUER *et al.*, 2017).

Ao fim as árvores de decisão podem ser denominadas como modelos estatísticos que através de um treinamento supervisionado são capazes de prever dados e classificá-los.

Para que um algoritmo funcione de maneira satisfatória é necessário o entendimento da sua lógica de programação, no caso da floresta aleatória é utilizado dois coeficientes o de Gini e o de Entropia.

O primeiro determina as ramificações que são mais prováveis de ocorrer, baseado no grau de heterogeneidade dos dados. Logo, pode ser utilizado para medir a impureza de um nó seu. Este coeficiente é dado por:

$$Gini(S) = 1 - \sum_{i=1}^c (p_i)^2 \quad (2.7)$$

Onde  $p_i$  representa a probabilidade dos pontos de classe que está sendo observada no conjunto de dados e  $c$  representa o número de classe.

O segundo coeficiente serve para medir o quão impuro ou não homogêneo é o conjunto

de dados. Por exemplo, a entropia igual a zero significa que o conjunto de dados são completamente homogêneo, ou um quando for igualmente dividida.

Dado um conjunto de dados (S) que pode ter c classes distintas, a entropia de S será dada por:

$$Entropia (S) = 1 - \sum_{i=1}^c p_i \log_2 p_i \quad (2.8)$$

Outra característica relacionada à entropia é o ganho de informação. O ganho de informação é baseado na redução da entropia depois que um conjunto de dados é dividido em um atributo. Construir uma árvore de decisão envolve encontrar o atributo que retorna o maior ganho de informação, ou seja, os ramos mais homogêneos.

Para arquitetar uma árvore de decisão, os treinamentos são feitos através de entradas e saídas. O princípio das árvores de decisão é ‘dividir para conquistar’, ou seja, um problema complexo é separado em problemas mais simples (CREPALDI *et al.*, 2011).

Desta maneira, os pontos positivos que influenciam na escolha do algoritmo de Florestas Aleatórias são a capacidade de suportar dados de elevada dimensionalidade de maneira satisfatória, além de ser possível trabalhar com dados ausentes e ainda ter precisão.

## 2.9 Métricas de avaliação

As métricas de avaliação são elementos importantes no aprendizado de máquinas, pois sem elas não é possível estimar o quanto modelo pode estar errado; do modelo, resumindo. As métricas, erro médio quadrático (MSE), erro médio absoluto (MAE), raiz quadrada do erro médio quadrático (RMSE) e coeficiente de determinação ( $R^2$ ) são os mais comumente utilizadas em modelos de previsão (HUYNH *et al.*, 2020). A raiz do erro logarítmico quadrático médio (RMSLE) também é um bom indicador para previsão, pois ele serve para avaliar o desempenho das previsões, de acordo com Salaken *et al.* (2015).

O erro médio quadrático (MSE) calcula a média dos erros ao quadrado. Desta forma, julga os diferentes graus entre os valores previstos e reais. Quanto menor o MSE, melhor o previsor (GUO *et al.*, 2003). O MSE pode ser calculado conforme Equação 2.9, onde  $y_i$  representa o valor real,  $\tilde{y}_i$  o valor previsto no número de elementos.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2 \quad (2.9)$$



O RMSE é uma métrica que possui apenas valores positivos, pois os erros são elevados ao quadrado antes de realizar a média. Desta forma, maiores erros terão uma maior penalização (GUO *et al.*, 2003; SALAKEN *et al.*, 2015). O RMSE pode ser expresso conforme Equação 2.10. Em que  $y_i$  representa o valor observado,  $\tilde{y}_i$  o valor previsto e  $n$  o número de amostras.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2} \quad (2.10)$$

A raiz do erro logarítmico quadrático médio (RMSLE) é uma métrica que ao contrário do RMSE, atribui uma penalização maior aos erros nas quais as previsões deveriam ser mais próximas de zero (SALAKEN *et al.*, 2015). O RMSLE é descrito na Equação 2.11.

$$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\text{LOG}(y_i + 1) - \text{LOG}(\tilde{y}_i + 1))^2} \quad (2.11)$$

O erro médio absoluto (MAE) é uma métrica que dá o mesmo peso para todas as diferenças entre o valor real e o valor previsto. Pode ser expresso conforme Equação 2.12.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \tilde{y}_i| \quad (2.12)$$

Além das métricas descritas anteriormente, pode ser aplicado também o coeficiente de determinação ( $R^2$ ), que se refere a proporção de variação da variável dependente (explicada), pela variável independente. O resultado do  $R^2$  deve ser avaliado com cautela, sendo que valores grandes podem indicar que o modelo se ajustou bem aos dados, por outro lado, valores pequenos pode indicar que o modelo não se adequou aos dados (MONTGOMERY; RUNGER, 2013).

O  $R^2$  é calculado conforme Equação 2.13, em que o somatório dos quadrados dos resíduos (Res) pode ser calculado de acordo com a Equação 2.14, e a soma total dos quadrados  $SS(Total)$  calculado conforme a Equação 2.15. Desta forma, temos que  $y_i$  representa o valor observado,  $\hat{y}_i$  o valor previsto,  $\bar{y}$  a média e  $n$  o número de amostras.

$$SS(Res) = \sum_{j=1}^n (y_j - \hat{y}_j)^2 \quad (2.13)$$

$$SS(Total) = \sum_{j=1}^n (y_j - \bar{y})^2 \quad (2.14)$$

$$R^2 = 1 - \frac{SS(Res)}{SS(Total)} \quad (2.15)$$

## 2.10 Seleção de características

Algumas bases de dados, podem possuir muitas características (atributos, variáveis), mas, uma parte pode ser irrelevante ou redundante para o conjunto de dados (FACELI *et al.*, 2011). O objetivo da seleção de características ou atributos, é selecionar um subconjunto a partir de um conjunto de dados, com a finalidade de reduzir características irrelevantes, facilitar a visualização e compreensão dos dados, e assim obter um melhor desempenho na execução dos algoritmos de aprendizado e conseqüentemente fornecer bons resultados de previsão (CHANDRASHEKAR; SAHIN, 2014; GUYON; ELISSEEFF, 2003).

Existem várias técnicas que se aplicam a seleção de características, e visam encontrar o subconjunto ótimo de características. Na avaliação do desempenho do subconjunto podem-se utilizar três abordagens: a Embutida (Embedded), o Filtro e o Wrapper (FACELI *et al.*, 2011).

Na abordagem embutida, a seleção de características é parte interna e comum do algoritmo de aprendizado. As árvores de decisão são um exemplo de algoritmo de aprendizado que utiliza esse método para seleção de atributos (FACELI *et al.*, 2011). Na abordagem baseada em filtro, a finalidade como o próprio nome sugere, é filtrar ou fazer uma heurística na busca de atributos irrelevantes do conjunto, através de algum critério, antes de aplicar o algoritmo de aprendizagem de máquina (JOHH; KOHAVI; PLEGER, 1994). Sendo assim, o algoritmo de aprendizagem receberá, como entrada os atributos relevantes fornecidos pela saída do filtro. A vantagem de se utilizar técnicas de filtragem, de acordo com Saeys, Inza e Larrañaga (2007), é que elas são computacionalmente rápidas, simples e independentes dos algoritmos de aprendizagem.

A abordagem baseada em wrapper utiliza o algoritmo de aprendizado de máquina, para selecionar e avaliar diferentes subconjuntos de características do conjunto de treinamento. Desta forma, o subconjunto que obtiver um melhor desempenho no processo de aprendizagem é definido como o subconjunto de características (STÁNCZYK; JAIN, 2015).

## 2.11 Software R

O R é um ambiente de desenvolvimento, e uma linguagem usada para descrição de problemas estatísticos de dados de forma gráfica. Foi criado na Universidade de Auckland, Nova Zelândia e está disponível sob a licença GPL e as versões binárias estão disponíveis para

as principais plataformas, como Windows, Java e Linux.

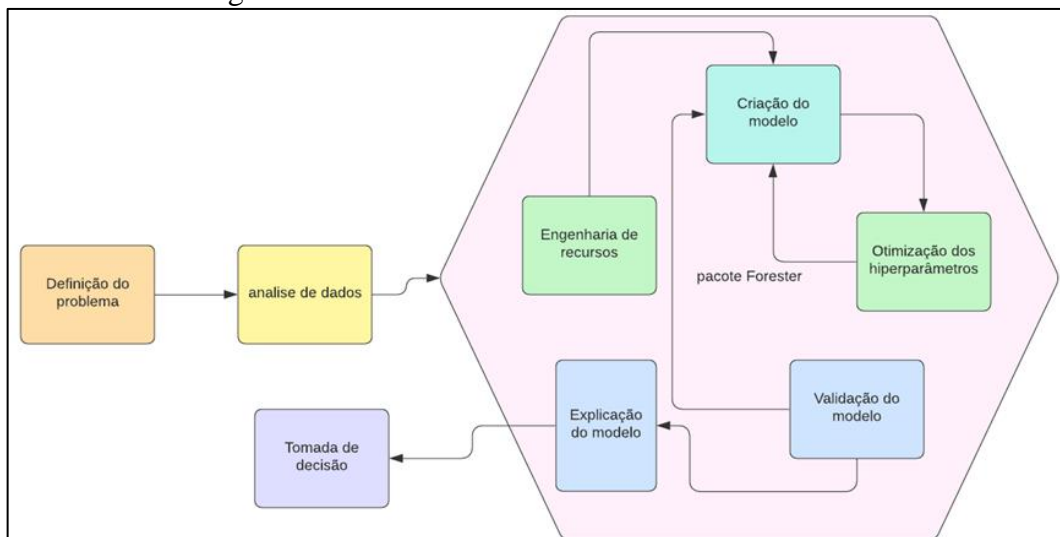
Uma das principais características do R é sua alta capacidade de expansão, baseada no uso dos pacotes, bibliotecas que implementam funções específicas para diversas áreas de estudo. Um conjunto básico de pacotes é incluído inicialmente com a instalação de R, mas muitos outros estão disponíveis na rede de distribuição do R. R é largamente utilizado entre estatísticos e analistas de dados para desenvolver aplicativos de diversas áreas da estatística e análise de dados (MONTANÕ, 2016).

## 2.12 Auto ml forester

O pacote forester resume automaticamente, etapas de complexidade elevada como engenharia de recursos, quais hiper parâmetros escolher, validação e pôr fim a explicação dos resultados. Com isso não há necessidade de criar um objeto específico para cada modelo, O pacote pode lidar com a estrutura de dados comuns, como quadro de dados, matrizes, tabelas de dados. Ele executa a engenharia de recursos para que possa integrara os dados (TIEN, 2021).

Outro ponto interessante da utilização é a interface do usuário que se torna simples uma vez que existe a penas 3 parâmetros obrigatórios para criar o modelo e sua otimização automática dos hiper parâmetros sem precisar ficar fazendo tentativas e erros. De acordo com a métrica do resultado final o modelo é escolhido de forma automática dentro dos modelos criados para melhor explicar seu comportamento, fornecendo explicações que desempenham um papel crucial na eliminação da relutância e no aumento da confiança dos tomadores de decisão podendo gerar explicações de modelos a nível local ou global (TIEN, 2021).

Figura 2 - Estrutura de funcionamento do AUTOML



Fonte: Do Autor (2022)

### 2.13 Pacote dalex do R

O pacote DALEX é um pacote R com um conjunto de ferramentas que ajudam a fornecer explicações descritivas de aprendizado de máquina que variam de métodos de interpretabilidade globais a locais (DALEX, 2022).

As funções de makes são indiscutivelmente o núcleo do pacote. Seu objetivo é simplificar o processo de criação de modelos. Usando essas funções, você pode criar modelos baseados em árvore de decisão em um curto período de tempo.

### 3 MATERIAL E MÉTODOS

#### 3.1 Área de estudo

O experimento foi realizado, na fazenda Barro Preto, localizada no município de Lavras, situado entre as coordenadas  $-21.199910^{\circ}\text{S}$  e  $-45.002519^{\circ}\text{O}$ . Em relação a classificação de Köppen-Geiser, a região apresenta o clima Cwa, (subtropical úmido, com verões quentes e úmidos e invernos frios e secos) com temperaturas médias anuais do ar de  $19,4^{\circ}\text{C}$  e precipitação média anual total de 1530 mm (DAMATTA *et al.*, 2007).

O experimento foi realizado em cultura implantada em meados de agosto de 2017 (aproximadamente 4 anos?) com a espécie *Coffea arábica* L., com espaçamento de  $3,30 \times 0,60\text{m}$ , contudo o experimento foi delimitado em uma área de  $100 \times 100\text{m}$  totalizando 25 linhas de plantio e média de 167 plantas por linhas (FIGURA 3).

Figura 3 - Localização da área de estudo

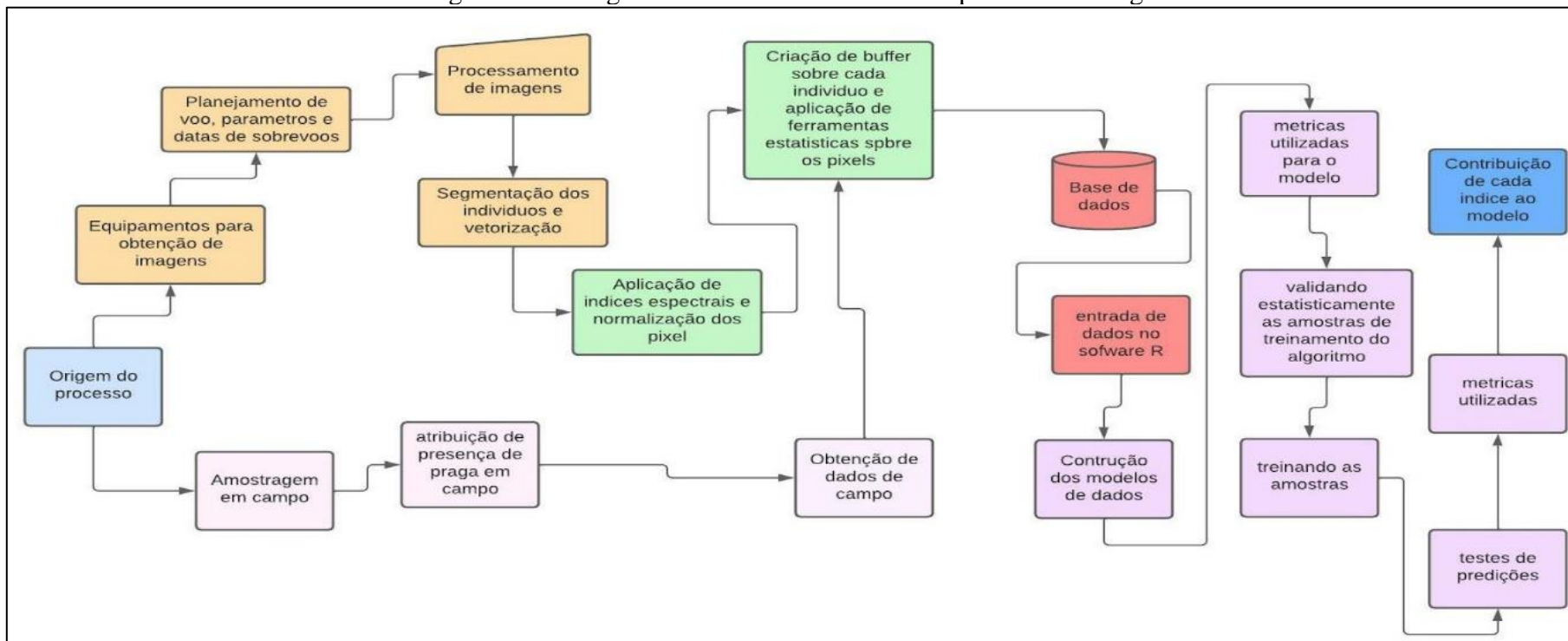


Fonte: Do Autor (2022)

#### 3.2 Obtenção de dados

Para obtenção da base de dados para treinamento do algoritmo, se realizaram uma série de processos simultâneos tanto na parte da área de voo como na parte de campo (FIGURA 4).

Figura 4 - Fluxograma de atividades de cada etapa da metodologia

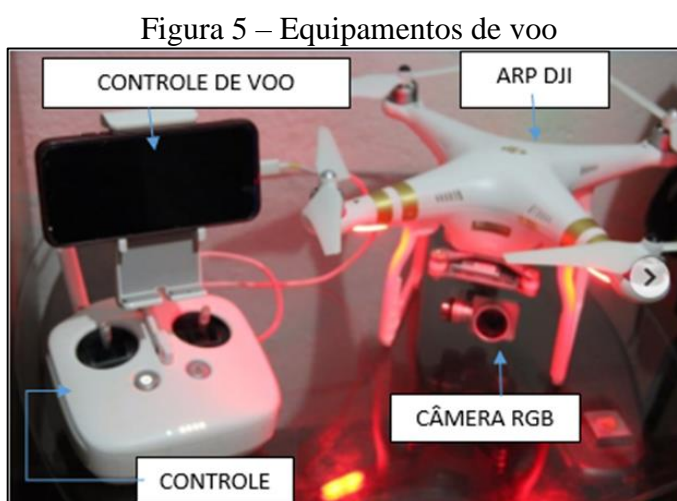


Fonte: Do Autor (2022)

### 3.3 Equipamentos para obtenção imagens

A obtenção de imagens da lavoura cafeeira, teve início em agosto e foi finalizado no começo de dezembro do ano de 2021, a partir de uma aeronave remotamente pilotada (RPA) com câmera RGB convencional, cujo o controle é realizado por meio de um sistema de posicionamento global por satélite (GNSS) que está integrado ao controle remoto, permitindo assim voos autônomos. No mesmo dia de obtenção das imagens, foram coletados dados de campo para as etapas subsequentes.

O modelo de RPA utilizado foi DJI phantom 3 pro (FIGURA 5), classificado como um quadricoptero, com uma bateria de autonomia de 23 minutos de voo e presença de Gimbal para estabilização da câmera, que auxilia a correção das fotos durante o voo, essas obtidas perpendicularmente ao solo.



Fonte: Do Autor (2022)

A câmera possui resolução de 12.4 megapixels (4000 x 3000) definidos no espectro visível RGB (Red-R, Green-G, Blue- B) e resolução radiométrica de 8bits, com lentes 20 mm com abertura ótica de f2/8, modelo 1/2.3" CMOS Pixels, sendo toda essa estrutura acoplada ao gimbal que por sua vez está ligado ao corpo do equipamento. As imagens capturadas foram armazenadas em um micro sd de 32 gb de classe 10.

### 3.4 Planejamento de voo, parâmetros e datas de sobrevoos

Neste estudo, a etapa de planejamento de voo foi feita no escritório, levado para área do experimento e sincronizado com a RPA, em algumas ocasiões foram feitos pequenos ajustes



nos parâmetros em campo, para determinar a rota correta, sobreposição da imagem, altura e velocidade de sobrevoo. Utilizou-se do *software* DroneDeploy (2021) Figura 6, uma vez que para tipo de RPA, seja ele de asa fixa ou rotativa, haverá um *software* ou aplicativo compatível para planejar, e executar missões como por exemplo: DJI Ground Station Pro Skydrones, Precision Flight, Pix4D Capture, dentre outros.

Figura 6 - Imagem do *software* DroneDeploy (2021)



Fonte: Do Autor (2022)

Outro *software* utilizado, foi o UAV forecast (FIGURA 7) que permite verificar as condições climáticas de voo, e também se existe algumas áreas restritivas, derivadas de rotas comerciais de aviões, assim sendo possível uma execução de voo responsável e seguro.

Figura 7 - *Software* UAV forecast para verificar as condições do voo



Fonte: Do Autor (2022)



Dentro do plano de voo, a sobreposição lateral e frontal foi adotada de 85%, uma vez que os primeiros testes demonstraram problemas de distorção, quando adotados valores menores que esse. A sobreposição é um fator que interfere na precisão, e qualidade do produto final, porém tudo depende do objetivo e do produto final, que também implica em aumento do tempo de captura de imagens, maior quantidade de fotos, para processamento da nuvem de pontos e por fim um ortomosaico, com maior quantidade de *pixel* o que resulta em uma difícil visualização em *softwares* de Sistemas de Informações Geográficas (SIG). Nestas condições para melhor compreensão de todos os parâmetros adotados foi construído uma tabela para exemplificar (TABELA 1).

Tabela 1 - Parâmetros selecionados para o plano de voo

<b>Parâmetros</b>	<b>Plano de voo</b>
Sobreposição lateral	85%
Sobreposição horizontal	85%
Altitude	30 m
Velocidade	3 m/s <sup>-1</sup>
Tempo estimado de voo	20 min 42 s
Área imageada	2.40 ha
Quantidade imagens por voo	350
Bateria	1
Resolução espacial (gsd)	1.20 cm/pixel

Fonte: Do Autor (2022)

Após a definição dos parâmetros do plano de voo foi feita a execução dos mesmos, os voos foram executados nos meses de agosto a dezembro, totalizando 40 voos distribuídos 3 vezes a semana (TABELA 2). A realização do horário de voos foram entre 11:00 e 13:00 horas para que não houvesse influência de sombreamento, uma vez que, esse horário o sol está posicionado perpendicularmente a superfície da terra, segundo Bater (2011) as imagens das câmeras sofrem influência de mudanças de iluminação e épocas sazonais.

Tabela 2 - Voos realizados

<b>Mês</b>	<b>Data dos voos</b>
Agosto	4,11,09,13,18,20,24,25,26,27
Setembro	3,6,8,10,15,17,22,24,28,30
Novembro	4,6,8,11,13,15,20,22,27,29
Dezembro	3,5,8,10,12,15,17,19,22,25

Fonte: Do Autor (2022)

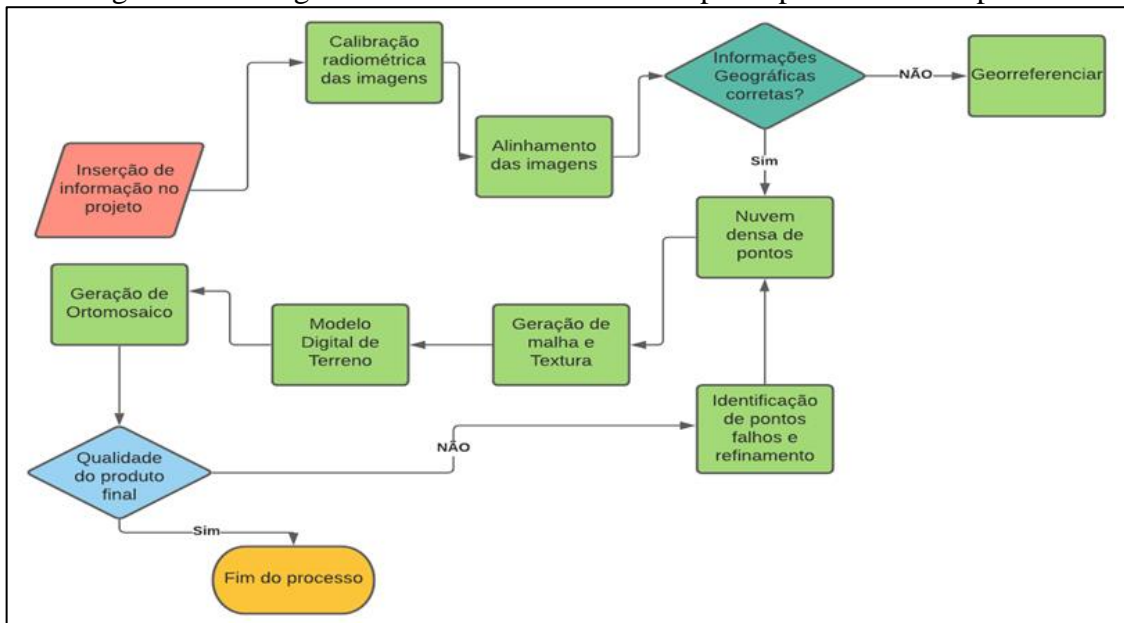
### 3.5 Processamento de imagens

Para etapa de processamento das imagens utilizou-se um computador com processador

i5 de 7 geração, memória RAM de 32 GB e sistema operacional Windows 10, arquitetura 64 bits em conjunto com uma placa dedicada de 2 GB da marca NVIDIA GeForce linha 940 MX.

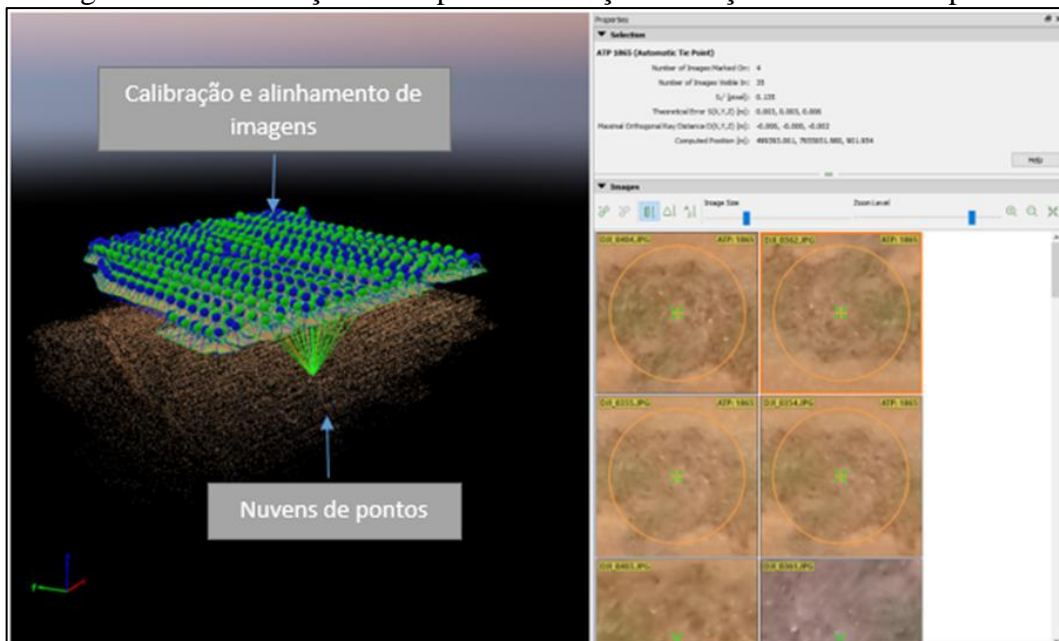
O processamento de imagens da RPA (FIGURA 8), foi executado usando o *software* Pix4D versão 4.5.6 na qual seu processo acontece da seguinte maneira: Calibração da câmera, alinhamento das imagens, geração de pontos de nuvem para modelos digitais de elevação (MDT) e de superfície (MSD) (FIGURA 9) e geração do ortomosaico (FIGURA 10).

Figura 8 - Fluxograma de atividades de cada etapa do processamento pix4d



Fonte: Do Autor (2022)

Figura 9 - Visualização de etapa de calibração e criação de nuvem no pix4d



Fonte: Do Autor (2022)

Figura 10 - Visualização do Ortomosaico e fim do processamento do pix4d



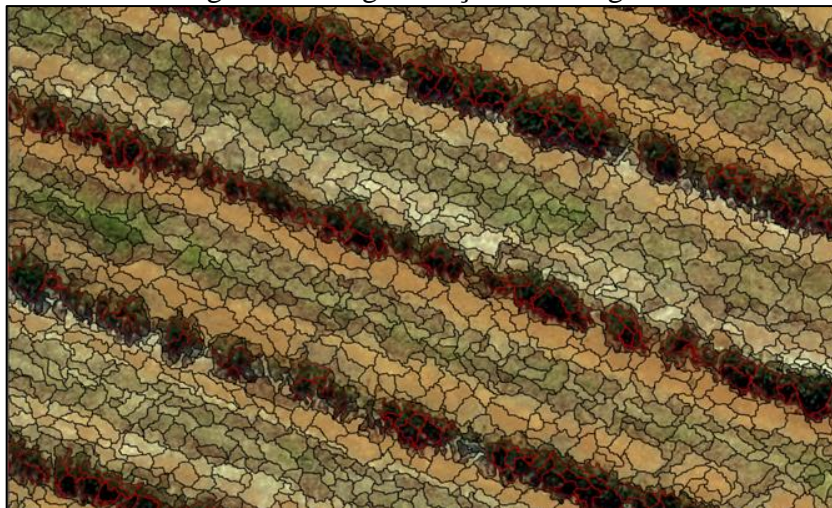
Fonte: Do Autor (2022)

### 3.6 Segmentação dos indivíduos e vetorização

Com os produtos gerados, utilizou-se a biblioteca OpenCV, combinado com a linguagem Python e executada no ambiente Google Colab, que é uma plataforma com processamento em nuvem, que pode ser acessado de qualquer dispositivo com conexão a internet, para geração de segmentação dos indivíduos e assim conseguir extrair os *pixels* que realmente importam.

A segmentação das imagens, utilizou do algoritmo Detector de bordas Canny, com limiares maiores de (70,150) e utilizando uma função Gaussiana para suavizar as bordas e diminuir os ruídos (FIGURA 11).

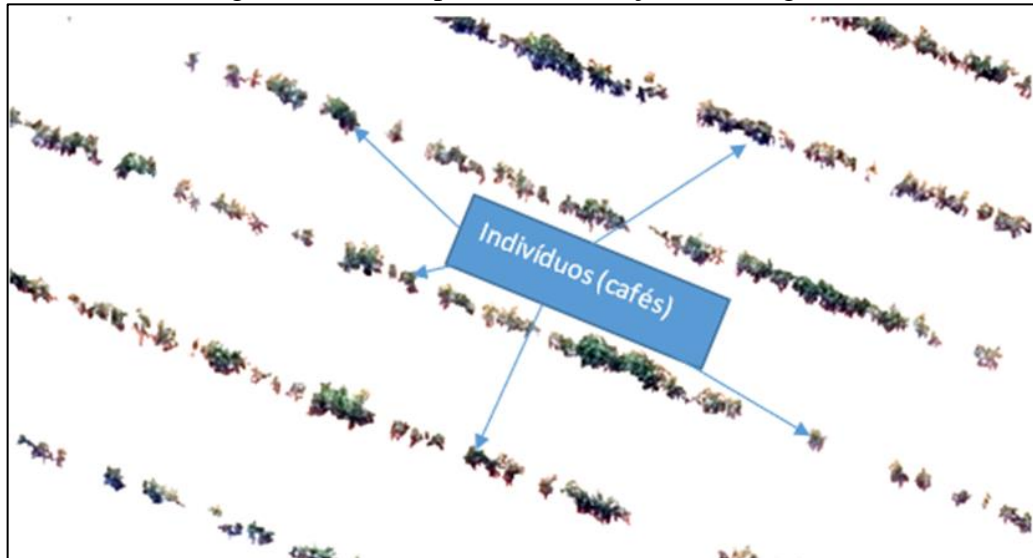
Figura 11 - Segmentação das imagens



Fonte: Do Autor (2022)

Com o *software* de SIG Qgis 3.22.1 Biatowieza foram feitas as etapas de vetorização (FIGURA 12) e inseridos as planilhas de dados para futuras análises estatísticas.

Figura 12 – Exemplo de Vetorização das imagens



Fonte: Do Autor (2022)

### 3.7 Aplicação de índices espectrais

Foram utilizados índices espectrais de vegetação na faixa do visível e as bandas separadas, para identificar a melhor resposta para detecção do bicho mineiro (TABELA 3).

Tabela 3 - Índices espectrais

Índice de Vegetação	Equação Matemática	Referência
Normalized Green Red Difference Index	$NGRDI = \frac{Green-Red}{Green+Red}$	Tucker (1979)
Redness Index	$RI = \frac{Red-Green}{Red+Green}$	Escadafal e Huete (1991)
Excess Green Index	$ExG = 2\ Green - Red - Blue$	Woebbeeke <i>et al.</i> (1995)
Ground Level Image Analysis	$GLI = \frac{(Green-Red)+(Green-Blue)}{(Green+Red)+(Green+Blue)}$	Louhaichi, Borman e Johnson (2001)
Visible Atmospherically Resistant Index	$VARI_{GREEN} = \frac{Green-Red}{Green+Red-Blue}$	Gitelson <i>et al.</i> (2002)
Excess Red Vegetative Index	$ExR = 1.4\ Red - Green$	Meyer, Hindman e Laksmi (1999)
Vegetative	$(2)_{VEG} = \frac{Green}{RED^\alpha BLUE^{(1-\alpha)}}$	Marchant e Onyango (2002)
Color Index of Vegetation Extraction	$CIVE = 0.441 * Red - 0.811 * Green + 0.385 * Blue + 18,78745$	Kataoka <i>et al.</i> (2003)
Excess Green Minus Excess Red	$ExGR = ExG - ExR$	Camargo Neto (2004)

Triangular Greenness Index	$(1)_{TGI} = -0,5[(\lambda Red - \lambda Blue)(Red - Green) - (\lambda Red - \lambda Green)(Red - Blue)]$	Hunt <i>et al.</i> (2005)
Combination	$COM = 0.36ExG + 0.47CIVE + 0.17VEG$	Guerrero <i>et al.</i> (2012)
Visible-band Difference Vegetation Index	$VDVI = \left( \frac{2 * Green - Red - Blue}{2 * Green + Red + Blue} \right)$	Wang <i>et al.</i> (2015)
IAF	$IAF = -25.838(\sqrt{red} + Blue)^2 - \sqrt{Green} + 2.354$	Miranda (2017)
Red	RED	-
Green	Green	-
Blue	Blue	-

Legenda: (1)TGI= Os valores de foram: Azul (B): 450nm, Verde(G):560 nm, vermelho (R): 650 nm.  
(2)= 0,667  
Fonte: Do Autor (2022)

Com a obtenção das imagens já vetorizadas, foi aplicado uma correção de refletância nos *pixels* para correção, em função das variações das imagens ocasionadas pelas diferentes datas de voo (XU; PU; GUO, 2020). A equação utilizada para cada imagem foi:

$$Dn_{std} = (DN - \overline{DN}) / std_{DN} \quad (3.1)$$

Onde e  $Dn_{std}$  é o valor de pixel padronizado,  $DN$  é o valor de pixel original,  $\overline{DN}$  é o valor médio de todos os valores de pixel, em uma única banda e  $std(x)$  é o desvio padrão para todos os valores de *pixel* em uma única banda.

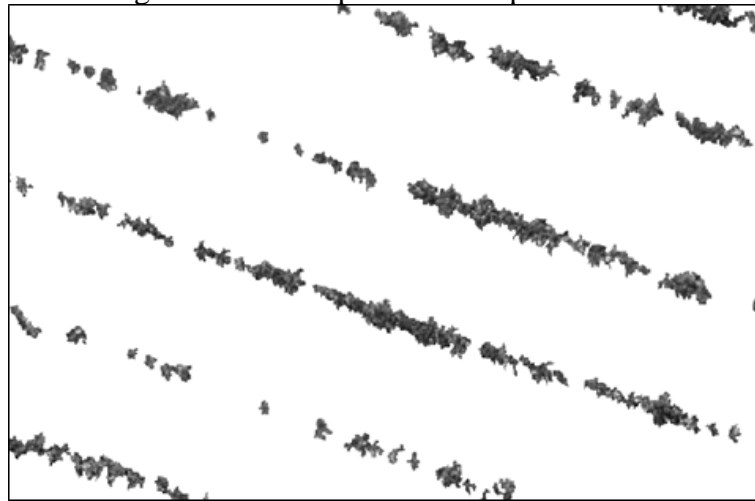
Depois de cada banda ser padronizada, todos os valores de pixel se enquadram na faixa de -1 a 1. as imagens padronizadas, são depois normalizadas de acordo com a equação 3.2 como imagens com valores de pixel variam de 0 a 255.

$$Dn_{nor} = 255 * (Dn_{std} - \min Dn_{std}) / (\max(Dn_{std}) - \min(xDn_{std})) \quad (3.2)$$

Onde  $Dn_{nor}$  é o valor de pixel normalizado,  $Dn_{std}$  é o valor de pixel padronizado,  $\min(Dn_{std})$  é o valor mínimo de todos os valores de pixel em cada banda padronizada e  $\max(Dn_{std})$  é o máximo de todos os valores de pixel em cada banda padronizada (FIGURA 13).



Figura 13 – Exemplo de banda padronizada

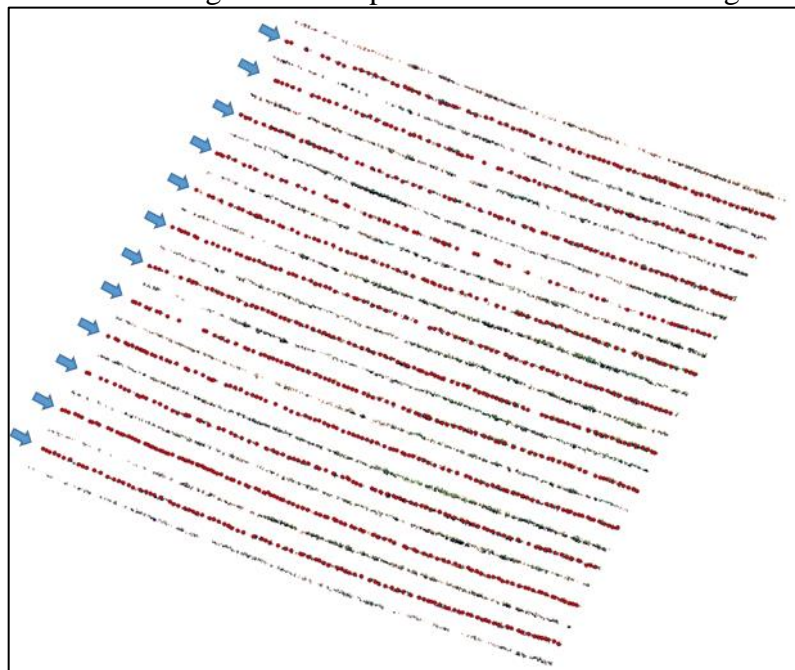


Fonte: Do Autor (2022)

### 3.8 Amostragens em campo

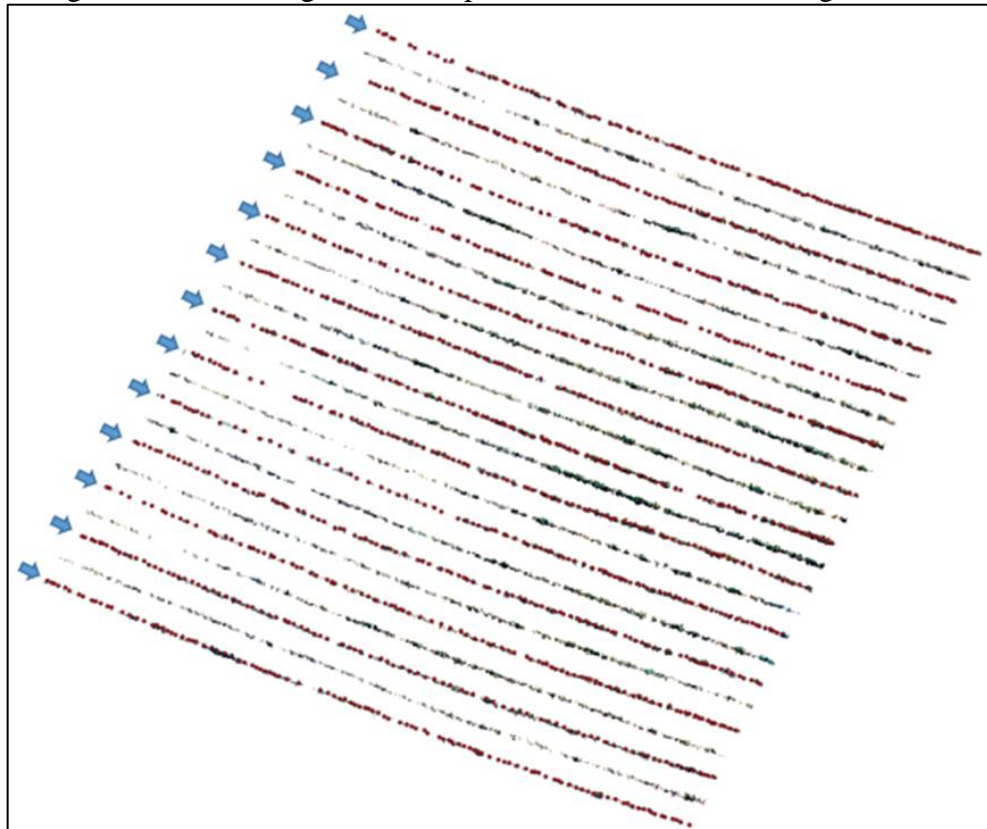
As amostragens em campo da infestação de bicho do mineiro (*Leucoptera coffella*) foram feitas de forma sistemática, mas dentro da linha optou-se por selecionar indivíduos randomicamente e alternando o início da linha de plantio (Figura 12 e 13) tudo isso em conjunto com as datas de voo, no período de agosto a dezembro de 2021, a fim de acompanhar o desenvolvimento da infestação na lavoura. Foram coletados uma média de 1200 de plantas de café para cada voo (FIGURA 14).

Figura 14 - Amostragem em campo das linhas no dia 06 de agosto 2021



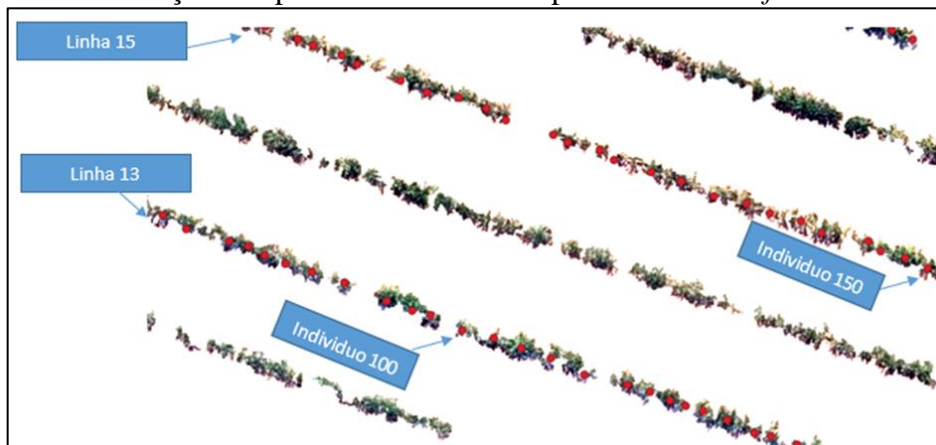
Fonte: Do Autor (2022)

Figura 15 - Amostragem em campo das linhas no dia 08 de agosto 2021



Fonte: Do Autor (2022)

As plantas amostradas foram marcadas com pontos derivados *software* Avenza maps, contudo sua precisão depende exclusivamente do gps do *smartphone*, o que implica em um alto nível de deslocamento no plano real, para minimizar isso e ter certeza da planta amostrada, foi atribuído para cada linha do plantio, um número correspondente a sua sequência, de 1 a 25, e foi anotado o número correspondente do início do indivíduo na linha (FIGURA 16) e depois cruzado com a imagem e anotações para assegurar amostragem correta.

Figura 16 – Marcação dos pontos derivados das plantas com o *software* avenza maps

Fonte: Do Autor (2022)

A forma de analisar os sintomas em campo, da infestação da praga do bicho mineiro (*Leucoptera coffella*) foram adaptados para escala, aonde cada indivíduo é avaliado de forma visual e atribuído um valor para a severidade do ataque, sendo classificados nesse trabalho em dois níveis de severidade (TABELA 4).

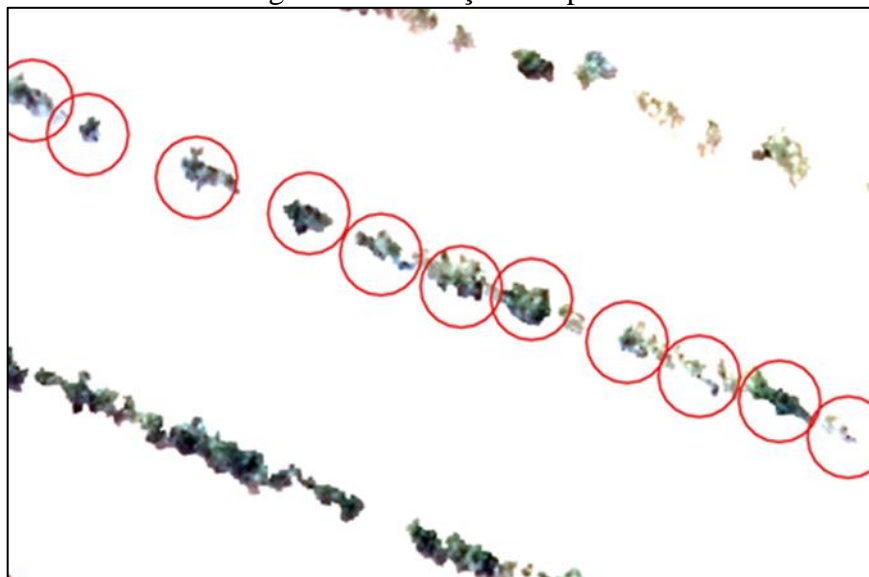
Tabela 4 - Níveis de severidade e descrição de infestação do bicho mineiro

Nível de Severidade	Descrição
0	Sem ocorrência de <i>Leucoptera coffella</i>
1	Com ocorrência de <i>Leucoptera coffella</i>

Fonte: Do Autor (2022)

O intuito de gerar esse banco foi capacitar a criação de uma área de zoneamento com 0,6m de raio executada pela ferramenta denominada *buffer* do Qgis 3.22.1 (FIGURA 17) para correta extração do pixel.

Figura 17 - Extração dos pixels



Fonte: Do Autor (2022)

### 3.9 Entrada de dados *software R*

Com a etapa anterior finalizada, utilizou-se a ferramenta *zonal statistics* do mesmo programa, para aplicar a estatística descritiva (média, mediana, máximo e mínimo), aonde os dados da média foram suficientes para criação do modelo como será abordado em resultados.

Ao fim do processamento foram ao todo 17 (dezessete) variáveis devidamente normalizadas e 11986 (onze mil novecentos e oitenta seis) indivíduos coletados, com a planilha organizada (TABELA 5).



Tabela 5 - Dados organizados

<b>Classe</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>...</b>
x	"499.458.032"	"4994570145"	"4994559378"	"4994546634"	...
y	"7655633888"	"7655634235"	"7655634551"	"7655635025"	...
cive	0.565	0	0.532	0.498	...
exg	0.45	0	0.486	0.505	...
exgr	0.383	0	0.4	0.5	...
exr	0.631	0	0.636	0.472	...
gli	0.359	0	0.353	0.45	...
iaf	0.808	0	0.836	0.851	...
ngrdi	0.69	0	0.687	0.724	...
ri	0.31	0	0.313	0.276	...
tgi	0.755	0	0.719	0.883	...
varig	0.743	0	0.739	0.773	...
vdvi	0.751	0	0.761	0.764	...
r	0.471	0	0.431	0.440	...
g	0.508	0	0.502	0.424	...
b	0.472	0	0.405	0.441	...

Fonte: Do Autor (2022)

Utilizou-se do *software* R 4.1.2 com as seguintes bibliotecas: Dplyr; Dalex; Forester; Ranger e o Caret inseridos, para o treinamento do algoritmo de aprendizado de máquina. Após a leitura dos dados pela função Read.csv foi executado, uma função para retirada de duas variáveis que foram as coordenadas x e y uma vez que poderiam influenciar sobre análise dos dados (TABELA 6).

Tabela 6 - Dados organizados com a retiradas das variáveis x e y

<b>Classe</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>...</b>
cive	0.565	0	0.532	0.498	...
exg	0.45	0	0.486	0.505	...
exgr	0.383	0	0.4	0.5	...
exr	0.631	0	0.636	0.472	...
gli	0.359	0	0.353	0.45	...
iaf	0.808	0	0.836	0.851	...
ngrdi	0.69	0	0.687	0.724	...
ri	0.31	0	0.313	0.276	...
tgi	0.755	0	0.719	0.883	...
varig	0.743	0	0.739	0.773	...
vdvi	0.751	0	0.761	0.764	...
r	0.471	0	0.431	0.440	...
g	0.508	0	0.502	0.424	...
b	0.472	0	0.405	0.441	...

Fonte: Do Autor (2022)

### 3.10 Construção dos modelos de dados

Com o uso do comando no R o pacote auto *machine learning* forester (AutoML forester), analisou os dados das estatísticas descritivas, tendo a média como melhor parâmetro para a criação dos modelos de determinação de presença ou não do bicho mineiro. O pacote usa alguns modelos prévios para suas análises sendo eles Lightgbm, Catboost, Ranger, Xgboost, Mlr3, Caret H2O, Rpart.

### 3.11 Parâmetros utilizadas para o modelo

Com as variáveis devidamente processadas, foi feito um comando no R para que pudesse usar o melhor modelo na próxima etapa, pois o pacote AutoML forester trabalha de forma randômica, portanto, se os dados fossem rodados novamente sem essa etapa é possível que pudessem existir pequenas diferenças nos valores dos modelos.

### 3.12 Validando estatisticamente as amostras de treinando do algoritmo

Para evitar equívocos sobre os resultados, foram abertos os dados do processo anterior e executada uma repetição de 99 vezes conhecido como Bootstrap, a qual é do tipo com reposição, possibilitando assim melhor análise dos dados, após isso dividiu-se a base de dados para testes e validação.

### 3.13 Treinando as amostras

Para o treinamento das amostras foi destinado 90% da base para treinar o algoritmo e 10% para sua validação por meio dos comandos em R.

### 3.14 Testes de predição dos dados

Foram feitos os testes de predição, com o objetivo de verificar o erro estimado gerado sobre o modelo ajustado em relação a base de dados não ajustadas, e assim ver o quanto o modelo está tendo assertividade. As operações foram realizadas no ambiente R e foram utilizadas as métricas Accuracy (%), Sensitivity ou positivo verdadeiro (%), Specificity ou falso verdadeiro (%) obtidas por *confusion matrix*.

### 3.15 Métricas utilizadas

As métricas utilizadas são demonstradas pelas equações 1,2,3 sendo Accuracy (%), Sensitivity ou positivo verdadeiro (%), Specificity ou falso verdadeiro (%) obtidas pela matriz de confusão.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (3.3)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3.4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3.5)$$

Para TP (*True Positive*) representando uma correta predição positiva, FP (*False Positive*) uma incorreta predição positiva, TN (*True Negative*) uma correta predição negativa e, FN (*False Negative*) uma incorreta predição negativa.

### 3.16 Matriz de confusão

Com os testes de predição, foi executado uma matriz de confusão para compressão dos resultados finais e os resultados foram salvos por uma linha de comando.

### 3.17 Contribuição de cada índice no modelo

O pacote *forester* permite em conjunto com DALEX gerar gráficos para interpretação dos resultados e assim verificar de forma mais precisa sua contribuição para o processo, com isso em mente foi criado no R uma linha de comando para verificar os possíveis gráficos.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Resultado da construção dos modelos de dados

Segundo a metodologia proposta, no Item 3.9 como se verifica na Tabela 7 os resultados de precisão (*precision*) tiveram melhor desempenho no modelo ranger (99,98%), sendo que os demais modelos tiveram decréscimo variando de 20% (Xgboost) a 29,1% (Lightgbm) quando comparado com o primeiro modelo.

Tabela 7 – Métricas de teste do modelo de detecção na detecção de bicho mineiro em lavoura de café, a partir de imagens RGB adquiridas por sensor embarcado em aeronave remotamente pilotada (RPA)

Modelos	<i>Precision</i> (%)	<i>Recall</i> (%)	F1 (%)	<i>Accuracy</i> (%)	<i>Auc</i>
Ranger	99,98	98,17	99,07	99,17	99,08
Xgboost	79,63	41,50	54,56	68,75	66,37
Catboost	70,54	23,97	35,78	61,09	57,85
Lightgbm	70,07	39,10	50,19	64,91	62,65

Fonte: Do autor (2022)

A modelo *random forest* ‘ranger’ seria o mais preciso em uma medição, já que nesse caso seria menor a variabilidade entre os valores obtidos, apresentando dessa maneira uma pequena dispersão.

A variável *recall* é compreendida, como uma métrica que indica a relação entre as previsões positivas, realizadas corretamente e todas as previsões que realmente são positivas, portanto, nessa situação o modelo *random forest* ‘ranger’ apresentou o melhor resultado de 98,17% o que indica que a segmentação, na escolha dos pontos corretos da nuvem, ou seja, os pontos que deveriam ser verdadeiros e positivos, foram rotulados como tal. Já o modelo *Catboost* apresentou o pior resultado sendo de 23,97% e os demais possuíram variações entre 39,1% *Lightgbm* e 41,50% *Xgboost* o que significa que a escolha dos pontos corretos em nuvens obteve baixo acerto na classificação do alvo.

A métrica F1 é a menos intuitiva de compreensão, contudo, ela é uma métrica que combina *Precision* e *recall*, e é derivada de uma média harmônica, quando os dois valores do cálculo são iguais. Essa média, gera resultados muito próximos da média “comum”. No entanto, sempre que os valores são diferentes, essa média se aproxima mais dos valores menores. Portanto, ela expressa bem *dataset* com classes desproporcionais, e quanto maior o valor melhor o resultado do modelo, sendo assim o modelo *random forest* ‘ranger’ apresentou o melhor valor 99,07%, e o *Catboost* apresentou o menor valor 35,78% os modelos *Xgboost* e *Lightgbm* se

saíram um pouco melhor que o Catboost, mas não ultrapassaram os 50% o que demonstra que também não alcançaram valores relevante para modelagem.

Outra variável foi a *accuracy*, de forma geral quanto mais acurado o processo de medição, mais próximo está o resultado da medida do valor tido como verdadeiro, de referência, de modo que, a acurácia está ligada à presença de tendenciosidade e quanto maior a precisão menor a variabilidade entre as medidas. O modelo *random forest* ‘ranger’ obteve a maior acuracidade 99,08% e os demais não ultrapassaram o valor de 67%.

Por fim, a variável auc que obteve melhor resposta, foi novamente o modelo *random forest* ‘ranger’ 99,08% o que indica o quão bem o modelo separa as classes da variável resposta, baseado na sua distribuição acumulada, portanto quanto maior o valor melhor será o modelo, contudo, o segundo melhor modelo foi o Xgboost 66,37%, Lightgbm 62,65% seguido pelo pior resultado que foi o Catboost.

Portanto, de acordo com os modelos de aprendizagem, gerados para detectar a presença do bicho mineiro, o que se destacou com resultados acima de 99% em todas variáveis avaliadas (Precision, recall, f1, Accuracy e Auc) foi modelo *random forest* ‘ranger’.

#### 4.2 Resultados dos parâmetros utilizadas para o modelo

Segundo a metodologia proposta, no Item 3.10 como se verifica na Tabela 8 o número de árvore de processos (NUMBER OF TREES) foi de 500, esse parâmetro é importante, pois, vários estudos afirmam que, para obtenção de resultados satisfatórios o padrão adotado pode variar de 300 a 400 (DURO; FRANKLIN; DUBÉ, 2012). Porém, segundo Liaw e Wiener (2002), valores acima desses mencionados podem proporcionar, um resultado estável. Contudo, números mais elevados que esses não prejudicam o modelo, apenas aumenta o tempo de processamento.

Tabela 8 – Parâmetros de teste na detecção de bicho mineiro em lavoura de café, a partir de imagens RGB gerada por aeronave remotamente pilotada (RPA)

Type:	Classification
NUMBER OF TREES:	500
SAMPLE SIZE:	9588
NUMBER OF INDEPENDENT VARIABLES:	14
MTRY nó:	3
TARGET NODE SIZE:	1

Fonte: Do autor (2022)

O sample size é o parâmetro das amostras que foram treinadas, nesse caso para o

treinamento foi utilizado 90% da base para treinamento o que resulta no valor de 9588 indivíduos com ou sem presença de bicho mineiro.

Em relação ao parâmetro *mtry*, o valor foi de 3, contudo alguns trabalhos como utilizam uma outra forma de calcular seguindo a orientação dos autores Duro, Franklin e Dubé (2012) para detecção de classificação de incêndios, a comparação pode ser válida, pois a metodologia utiliza-se da ideia de uma classificação binária, ou seja, está ou não queimado da mesma forma que este trabalho se utiliza da presença ou não do bicho mineiro.

O parâmetro que NUMBER OF INDEPENDENT VARIABLES foram avaliados 11 índices espectrais em RGB e mais 3 imagens separadas RGB totalizando 14.

Em relação ao TARGET NODE SIZE esse parâmetro a implementações de floresta aleatória permite, controlar a profundidade e a complexidade das árvores individuais. Isso geralmente inclui hiper parâmetros, como tamanho do nó, profundidade máxima, número máximo de nós terminais ou o tamanho do nó necessário para permitir divisões adicionais. O tamanho do nó é provavelmente o hiper parâmetro mais comum para controlar a complexidade da árvore, e a maioria das implementações usa os valores padrão no caso 1 de um para classificação e cinco para regressão, pois esses valores tendem a produzir bons resultados (DÍAZ-URIARTE; ALVAREZ DE ANDRÉS, 2006).

#### 4.3 Resultados da matriz de confusão

Segundo a metodologia proposta, no Item 3.16 como se verifica na Tabela 9, a Acurácia foi de 90%, tratando-se de excelente resultado de detecção. Safonova *et al.* 2019 encontram resultados satisfatórios de acurácia 94,12; 91,95; 96,39; e 98,77% na detecção de árvores de *Abies sibirica Ledeb (Abeto siberiano)* atacadas pelo besouro de casca (*Polygraphus proximus* Blandford), contudo Qin *et al.* (2021) para identificação de surtos do nematoide *Bursaphelenchus xylophilus*, causador da murcha do pinheiro, obteve uma acurácia média de 79,33%.

Tabela 9 – Resultados das métricas, a partir de imagens RGB gerada por aeronave remotamente pilotada (RPA)

<i>Accuracy</i>	<i>sensitivity</i>	<i>Specificity</i>
90%	88%	91%

Fonte: Do autor (2022)

A Sensibilidade obtida foi de 88% o que indica um valor relativamente alto. Segundo Saito e Rehmsmeier (2015) esse indicador demonstra o nível de taxa de acerto na detecção do

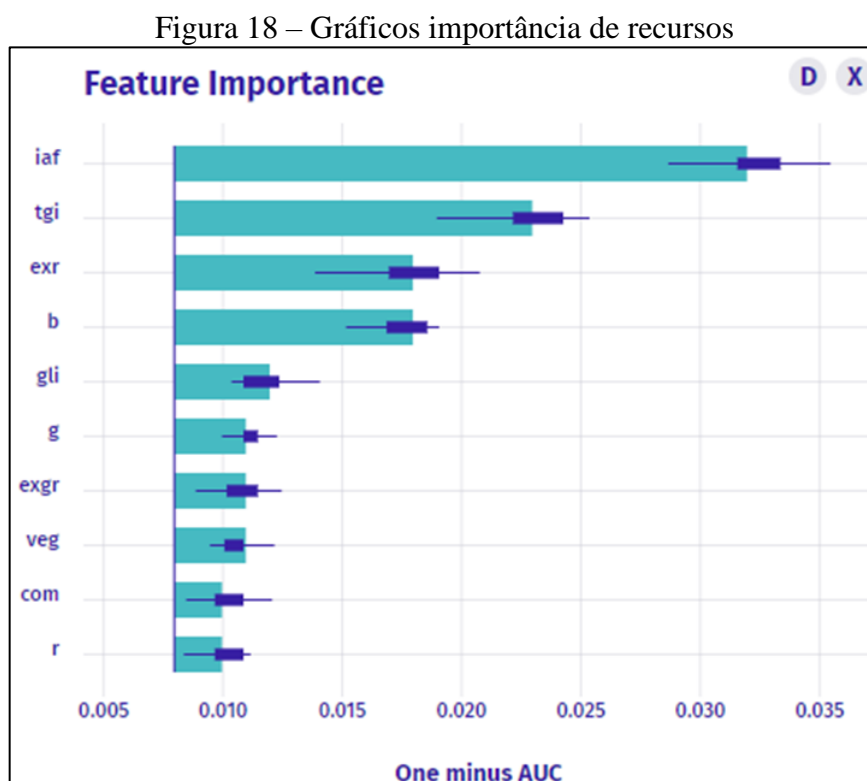
objeto para a classe de interesse, no caso se houve ou não a presença do bicho mineiro.

O indicador especificidade, ou conhecido como falso verdadeiro, apresentou valor de 91% considerado alto, pois um modelo de aprendizado de máquinas ser alto os valores da métrica citadas devem estar acima de 85%.

Para detecção de bicho mineiro usando técnicas de *machine learning*, os resultados foram satisfatórios o que corrobora com outras pesquisas de detecção de pragas como apresentada por que utilizou *machine learning* rede neural YOLOv4 para detecção de cochonilhas (Coccidae e Diaspididae) em tempo real para diversas culturas agrícolas. com a implementação Faster-RCNN 89% em seus resultados de detecção, de árvores mortas pela murcha do pinheiro também apresentaram proposta, de métodos de detecção de surtos de pragas em florestas baseadas em CNNs.

#### 4.4 Gráficos do comportamento do processo

Segundo a metodologia proposta, no Item 3.19 foi feito um gráfico (FIGURA 18), da contribuição de cada índices para o modelo.



Fonte: Do autor (2022)

O gráfico importância de recursos (*Feature Importance*) é uma representação global.

Isso significa que ele ao analisar todas as observações, nesse trabalho 14 índices, retornou os mais relevantes, para a geração do modelo, onde quatro índices se destacaram, que foram: *random forest* ‘ranger’, tgi, exr e a banda do azul (b).

O iaf foi o que obteve maior contribuição, sendo acima de 0.030. esse índice pode ser um indicador biofísico e que possibilita verificar se a planta sofre algum tipo de estresse em sua área foliar, o que colabora com ataque do bicho-mineiro, além que o iaf pode, ser utilizado para detecção de doenças em plantas (ALMEIDA *et al.*, 2015), detecção de estresse em plantas (SILVA *et al.*, 2015), estimativa de biomassa de culturas agrícolas (SOUZA *et al.*, 2019), mapeamento de culturas agrícolas (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE, 2016). O segundo índice que contribuiu com o modelo é o tgi (*Triangular Greenness Index*) na faixa de 0.020 a 0.025, trabalhos feitos com o mesmo índice, foram executados para serem usados para explicar diferenças sutis na saúde das árvores causadas por várias doenças (GARZA *et al.*, 2020), essas modificações sutis na área foliar, também contribuem para o entendimento do ataque do bicho mineiro, uma vez que a folha ao receber o ataque necrosa partes e modifica a coloração na folha.

E por fim a banda espectral azul (b) e o índice exr contribuíram na mesma faixa para o modelo sendo de 0.015 a 0.018. Já os demais ficaram abaixo da faixa de 0.015 em especial a banda do vermelho (r) que teve o menor alcance 0.010.



## 5 CONCLUSÃO

Com relação à detecção de infestação do bicho mineiro na lavoura cafeeira, utilizando imagens RGB, gerada por aeronave remotamente pilotada, o Auto ML Forester gerou 4 melhores modelos de aprendizagem, com Acurácia que variou de 70 a 90%.

O modelo *random forest* ‘ranger’ com 500 árvores (NTREE), 3 nós (MTRY) e com uma quantidade de 9588 amostras processadas (SAMPLE SIZE) Ee 14 variáveis analisadas foi o que alcançou a maior Acurácia, de 90%, Especificidade de 88% e Sensibilidade de 91%, tratando-se de valores elevados de aprendizagem de máquina, que representam confiabilidade.

Os resultados alcançados demonstram ser possível, fazer a gestão e manejo da lavoura cafeeira, através de imagens RGB, de baixo custo, com relação a infestação de bicho mineiro.

## REFERÊNCIAS

- AGÜERA VEGA, F. *et al.* Multi-temporal imaging using an unmanned aerial vehicle for monitoring a sunflower crop. **Biosystems Engineering**, v. 132, p. 19-27, 2015. Disponível em: <https://doi.org/10.1016/j.biosystemseng.2015.01.008>. Acesso em: 19 mai. 2022.
- ALBORNOZ, C.; GIRALDO, L. F. Trajectory design for efficient for crop irrigation with a UAV. In: **IEEE COLOMBIAN CONFERENCE ON AUTOMATIC CONTROL**, 3., Cartagena, 2017. Proceedings. Cartagena: IEEE, 2017. p. 1-6. Editors Diego Patino and Eugenio Yime. 3rd CCAC. Disponível em: [https:// DOI: 10.1109/CCAC.2017.8276401](https://doi.org/10.1109/CCAC.2017.8276401). Acesso em: 19 mai. 2022.
- ALLRED, B. *et al.* Effective and efficient agricultural drainage pipe mapping with UAS thermal infrared imagery: a case study. **Agricultural Water Management**, v. 197, p.132-137, 2018. Disponível em: <https://doi.org/10.1016/j.agwat.2017.11.011>. Acesso em: 19 mai. 2022.
- ALMEIDA, A. Q. de *et al.* Índice de Área Foliar de Eucalyptus Estimado por Índices de Vegetação Utilizando Imagens TM - Landsat 5. **Floresta e Ambiente**, v. 22, n. 3, p. 368–376, 2015. Disponível em: <https://doi.org/10.1590/2179-8087.103414>. Acesso em: 19 maio 2022.
- ALSALAM, B. H. Y. *et al.* Autonomous UAV with vision based on-board decision making for remote sensing and precision agriculture. In: **IEEE AEROSPACE CONFERENCE**, 2017, Big Sky. Proceedings. Big Sky: IEEE, 2017. p. 1-12. Disponível em: <https://doi.org/10.1109/AERO.2017.7943593>. Acesso em: 19 mai. 2022.
- ARROYO, J. A. *et al.* Assessing nitrogen nutrition in corn crops with airborne multispectral sensors. In: BENFERHAT, S.; TABIA, K.; ALI, M. (Ed.). **Advances in Artificial Intelligence: from theory to practice**. Cham: Springer, 2017. p. 259-267. Disponível em: [https://doi.org/10.1007/978-3-319-60045-1\\_28](https://doi.org/10.1007/978-3-319-60045-1_28). Acesso em: 19 mai. 2022.
- ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ – ABIC. **O café brasileiro na atualidade**. ABIC, 2021. Disponível em: <https://www.abic.com.br/tudo-de-cafe/o-cafe-brasileiro-na-atualidade/#:~:text=O%20Brasil%20%C3%A9%20o%20maior,h%C3%A1%20mais%20de%2015%20anos>. Acesso em: 19 mai. 2022.
- BATER, C. W. *et al.* Using digital time-lapse cameras to monitor species-specific understorey and overstorey phenology in support of wildlife habitat assessment. **Environmental Monitoring and Assessment**, [s.l.], v. 180, n. 1, p. 1-13, set. 2011. Disponível em: <https://link.springer.com/article/10.1007/s10661-010-1768-x>. Acesso em: 19 mai. 2022.
- BELGIU, M.; DRAGUT, L. Random Forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, Amsterdam, v. 114, p. 24–31, 2016. Disponível em: <https://doi.org/10.1016/j.isprsjprs.2016.01.011>. Acesso em: 19 mai. 2022.
- BERNARDI, A. C. D. C. *et al.* Ferramentas de agricultura de precisão como auxílio ao manejo da fertilidade do solo. **Cadernos de Ciência e Tecnologia**, Brasília, v. 32, n. 1/2, p. 211-227, ago. 2015. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/>

/publicacao/1037525/ferramentas-de-agricultura-de-precisao-como-auxilio-ao-manejo-da-fertilidade-do-solo. Acesso em: 19 mai. 2022.

CAIXETA, S. L. *et al.* Nutrição e vigor de mudas de cafeeiro e infestação por bicho mineiro. **Ciência Rural**, v. 34, n. 5, p. 1429-1435, 2004. Disponível em: <https://doi.org/10.1590/S0103-84782004000500016>. Acesso em: 19 mai. 2022.

CAMARGO NETO, J. **A combined statistical** - soft computing approach for classification and mapping weed species in minimum-tillage systems. Dissertation (Doctor) - University of Nebraska, Lincoln, 2004. Disponível em: <https://digitalcommons.unl.edu/dissertations/AAI3147135>. Acesso em: 02 fev. 2022.

CAMARGO, Â. P. de; CAMARGO, M. B. P. de. Definição e esquematização das fases fenológicas do cafeeiro arábica nas condições tropicais do Brasil. **Bragantia**, v. 60, n. 1, p. 65-68, 2001. Disponível em: <https://doi.org/10.1590/S0006-87052001000100008>. Acesso em: 19 mai. 2022.

CANDIAGO, S. *et al.* Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. **Remote Sensing**, Basel, v. 7, n. 4, p. 4026-4047, abr. 2015. Disponível em: <https://doi.org/10.3390/rs70404026>. Acesso em: 19 mai. 2022.

CHANDRASHEKAR, G.; SAHIN, F. A survey on feature selection methods. **Computers and Electrical Engineering**, v. 40, n. 1, p. 16-28, 2014. Disponível em: <https://doi.org/10.1016/j.compeleceng.2013.11.024>. Acesso em: 19 mai. 2022.

CONCEIÇÃO, C. H. C. **Biologia, dano e controle do bicho-mineiro em cultivares de café arábica**. 2005. 86 p. Dissertação (Mestrado em Agricultura Tropical e Subtropical) - Instituto Agrônômico, Campinas, 2005. Disponível em: <http://www.iac.sp.gov.br/areadoinstituto/posgraduacao/repositorio/buscar.php>. Acesso em: 19 mai. 2022.

CONSÓRCIO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO DO CAFÉ - CBP&D/CAFÉ. **Região Sudeste produz 87,5% dos Cafés do Brasil em 2020**. Consórcio Pesquisa Café, 2021. Disponível em: <http://www.consorciopequisacafe.com.br/index.php/imprensa/noticias/1037-2021-01-20-17-12-34>. Acesso em: 19 mai. 2022.

CREPALDI, P. G. *et al.* Um estudo sobre a árvore de decisão e sua importância na habilidade de aprendizado. **Revista Eletrônica do Instituto de Ensino Superior de Londrina**, v. 14, n. 1, p. 1-14, 2011. Disponível em: <https://www.inesul.edu.br/revista/index.php?vol=15>. Acesso em: 19 mai. 2022.

CRIMINISI, A. *et al.* Regression forests for efficient anatomy detection and localization in computed tomography scans. **Medical Image Analysis**, Amsterdam, v. 17, n. 8, p. 1293-1303, 2013. Disponível em: <https://doi.org/10.1016/j.media.2013.01.001>. Acesso em: 19 mai. 2022.

CRISTEA, L. *et al.* Multiple criteria analysis of remotely piloted aircraft systems for monitoring the crops vegetation status. **In IOP Conference Series: Materials Science and Engineering**, [s.l.], v. 147, n. 1, jan. 2016. Disponível em: <https://doi.org/10.1088/1757->

899X/147/1/012059. Acesso em: 19 mai. 2022.

CUNHA, A. C. *et al.* Sistema computacional web para controle de gestão da produção de café. **Coffee Science**, Lavras, v. 12, n. 3, p. 344 - 354, jul. /set. 2017. Disponível em: <http://www.sbicafe.ufv.br/handle/123456789/9125>. Acesso em: 19 mai. 2022.

DAMATTA, F. M. *et al.* Ecophysiology of coffee growth and production. **Brazilian Journal of Plant Physiology**, Campos dos Goytacazes, v. 19, p. 485- 510, 2007. Disponível em: <https://doi.org/10.1590/S1677-04202007000400014>. Acesso em: 19 mai. 2022.

DEMOTES-MAINARD, S. *et al.* Indicators of nitrogen status for ornamental woody plants based on optical measurements of leaf epidermal polyphenol and chlorophyll contents. **Scientia Horticulturae**, v. 115, p. 377-385, 2008. Disponível em: <https://doi.org/10.1016/j.scienta.2007.10.006>. Acesso em: 19 mai. 2022.

DENG, L. *et al.* UAV-based multispectral remote sensing or precision agriculture: a comparison between different cameras. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 146, p.1 24-136, 2018. Disponível em: <https://doi.org/10.1016/j.isprsjprs.2018.09.008>. Acesso em: 19 mai. 2022.

DÍAZ-URIARTE, R.; ALVAREZ DE ANDRÉS, S. Gene selection and classification of microarray data using random forest. **BMC Bioinformatics**, v. 7, n. 3, p. 1-13, 2006. Disponível em: <https://doi.org/10.1186/1471-2105-7-3>. Acesso em: 19 maio 2022.

**DRONEDEPLOY**. Site institucional da DroneDeploy. 2021. Disponível em: <https://www.dronedeploy.com/>. Acesso em: 19 mai. 2022.

DURO, D. C.; FRANKLIN, S. E.; DUBÉ, M. G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. **Remote Sensing of Environment**, v. 118, n. 1, p. 259-272, 2012. Disponível em: <https://doi.org/10.1016/j.rse.2011.11.020>. Acesso em: 19 maio 2022.

ESCADAFAL, R.; HUETE, A. Etude des propriétés spectrales des sols arides appliquée à l'amélioration des indices de végétation obtenus par télédétection. **Comptes Rendus de l'Académie des Sciences**, 312, p. 1385-1391, 1991. Disponível em: <https://www.documentation.ird.fr/hor/fdi:34390>. Acesso em: 19 mai. 2022.

FACELI, K. *et al.* **Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina**. 1. ed. Rio de Janeiro: Livros Técnicos e Científicos Editora Ltda, 2011.

FAIÇAL, B. S. *et al.* An adaptive approach for UAV-based pesticide spraying in dynamic environments. **Computers and Electronics in Agriculture**, v. 138, p. 210-223, 2017. Disponível em: <https://doi.org/10.1016/j.compag.2017.04.011>. Acesso em: 19 mai. 2022.

FITZ, P. R. **Geoprocessamento sem complicação**. São Paulo: Oficina de Textos, 2008.

FRIEDL, M. A.; BRODLEY, C. E. Decision tree classification of land cover from remotely sensed data. **Remote Sensing of Environment**, New York, v. 61, n. 3, p. 399-409, 1997. Disponível em: [https://doi.org/10.1016/S0034-4257\(97\)00049-7](https://doi.org/10.1016/S0034-4257(97)00049-7). Acesso em: 19 mai. 2022.

GARZA, B. N. *et al.* Quantifying Citrus Tree Health Using True Color UAV Images. **Remote Sensing**, v. 12, n. 1, p. 170, 2020. Disponível em: <https://doi.org/10.3390/rs12010170>. Acesso em: 19 mai. 2022.

GEBBERS, R.; ADAMCHUK, V. I. Precision agriculture and food security. **Science**, New York, v. 327, n. 5967, p. 828-831, fev. 2010. Disponível em: <https://doi.org/10.1126/science.1183899>. Acesso em: 19 mai. 2022.

GENUER, R. *et al.* RandomForests for Big Data. **Big Data Research**, Amsterdam, v. 9, p. 28-46, 2017. Disponível em: <https://doi.org/10.1016/j.bdr.2017.07.003>. Acesso em: 19 mai. 2022.

GILES, D. K. Use of Remotely Piloted Aircraft for Pesticide Applications: Issues and Outlook. **Outlooks on Pest Management, Saffron Walden**, v. 27, n. 5, p. 213-216, out. 2016. Disponível em: [https://doi.org/10.1564/v27\\_oct\\_05](https://doi.org/10.1564/v27_oct_05). Acesso em: 19 mai. 2022.

GITELSON, A. A. *et al.* Novel algorithms for re-mote estimation of vegetation fraction. **Remote Sensing of Environment**, v. 80, n. 1, p. 76-87, 2002. Disponível em: [https://doi.org/10.1016/S0034-4257\(01\)00289-9](https://doi.org/10.1016/S0034-4257(01)00289-9). Acesso em: 19 mai. 2022.

GOUVEIA, N. M. **Estudo da diferenciação e crescimento de gemas florais de /Coffea arábica/ L.; observações sobre antese e maturação dos frutos**. Campinas: Universidade Estadual de Campinas. Instituto de Biologia, 1984. Disponível em: <http://repositorio.ufla.br/>. Acesso em: 19 maio 2022.

GRENZDÖRFFER, G. J.; ENGEL, A.; TEICHERT, B. The photogrammetric potential of low-cost UAVs in forestry and agriculture. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, **Proceedings...** Beijing, v. 37, p. 1207-1214, jan. 2008. Disponível em: [https://www.isprs.org/proceedings/XXXVII/congress/1\\_pdf/206.pdf](https://www.isprs.org/proceedings/XXXVII/congress/1_pdf/206.pdf). Acesso em: 19 maio 2022.

GUERRERO, J. M. *et al.* Support Vector Machines for crop/weeds identification in maize fields. **Expert Systems with Applications**, v. 39, n. 12, p. 11149–11155, 2012. Disponível em: <https://doi.org/10.1016/j.eswa.2012.03.040>. Acesso em: 19 maio 2022.

GUO, G. *et al.* KNN Model-Based Approach in Classification. In: MEERSMAN, R.; TARI, Z.; SCHMIDT, D. C. (eds) **On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003. Lecture Notes in Computer Science**, vol 2888. Springer, Berlin, Heidelberg, 2003, p. 986–996. Disponível em: [https://doi.org/10.1007/978-3-540-39964-3\\_62](https://doi.org/10.1007/978-3-540-39964-3_62). Acesso em: 19 maio 2022.

GUYON, I.; ELISSEEFF, A. An Introduction of Variable and Feature Selection. **Journal of Machine Learning Research**, Brookline, v. 3, p. 1157-1182, 2003. Disponível em: <https://doi.org/10.5555/944919.944968>. Acesso em: 19 maio 2022.

HARDIN, P. J.; HARDIN, T. J. Small-scale remotely piloted vehicles in environmental research. **Geography Compass**, [s.l.], v. 4, n. 9, p. 1297–1311, set. 2010. Disponível em: <https://doi.org/10.1111/j.1749-8198.2010.00381.x>. Acesso em: 19 maio 2022.

HONKAVAARA, E. *et al.* Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight UAV spectral camera for precision agriculture. **Remote Sensing**, Basel, v. 5, n. 10, p. 5006-5039, out. 2013. Disponível em: <https://doi.org/10.3390/rs5105006>. Acesso em: 19 maio 2022.

HUNT, E. R. *et al.* Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. **Precision Agriculture**, v. 6, p.3 59-378, 2005. Disponível em: <https://doi.org/10.1007/s11119-005-2324-5>. Acesso em: 19 maio 2022.

HUYNH, A. N. L. *et al.* Near Real-Time Global Solar Radiation Forecasting at Multiple Time-Step Horizons Using Long Short-Term Memory Network. **Energies**, Basel, v. 13, n. 14, 2020. Disponível em <https://doi.org/10.3390/en13143517>. Acesso em: 19 mai. 2022.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Levantamento Sistemático da Produção Agrícola**. Rio de Janeiro v.29 n.12 p. 1-82, 2016. Disponível em: <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=76>. Acesso em: 19 maio 2022.

JANNOURA, R. *et al.* Monitoring of crop biomass using true colour aerial photographs taken from a remote controlled hexacopter. **Biosystems Engineering**, v. 129, p. 341-351, 2015. Disponível em: <https://doi.org/10.1016/j.biosystemseng.2014.11.007>. Acesso em: 19 maio 2022.

JOHH, G. H.; KOHAVI, R.; PLEGER, L. IrrelevantFeatures and the Subset Selection Problem. In: **Proceedings of the Eleventh International Conference on Machine Learning**, New Brunswick, 1994, p. 121-129. Disponível em: <https://doi.org/10.1016/B978-1-55860-335-6.50023-4>. Acesso em: 19 mai. 2022.

KATAOKA, *et al.* Crop growth estimation system using machine vision. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003). **Proceedings...** Kobe, Japan, 2003, p. b1079-b1083, v. 2. Disponível em: <https://doi.org/10.1109/AIM.2003.1225492>. Acesso em: 19 mai. 2022.

LALIBERTE, A. S.; RANGO, A. Image processing and classification procedures for analysis of sub-decimeter imagery acquired with an unmanned aircraft over arid rangelands. **GIScience e Remote Sensing**, [s.l.], v. 48, n. 1, p. 4-23, mai. 2011. Disponível em: <https://doi.org/10.2747/1548-1603.48.1.4>. Acesso em: 19 mai. 2022.

LAPPONI, J. C. **Estatística usando Excel**. Rio de Janeiro: Elsevier, 2005.

LELONG, C. C. D.; PINET, P. C.; POILVE', H. Hyperspectral imaging and stress mapping in agriculture: A case study on wheat in Beauce (France). **Remote Sensing of Environment**, [s.l.], v. 66, p. 179–191, nov. 1998. Disponível em: [https://doi.org/10.1016/S0034-4257\(98\)00049-2](https://doi.org/10.1016/S0034-4257(98)00049-2). Acesso em: 19 mai. 2022.

LIAW, A.; WIENER, M. Classification and Regression by randomForest. **R News**, v. 2/3, p. 18-22. Disponível em: <https://cran.r-project.org/doc/Rnews/>. Acesso em: 19 maio 2022.

LONG, D.; MCCARTHY, C.; JENSEN, T. Row and water front detection from UAV

thermal-infrared imagery for furrow irrigation monitoring. In: **IEEE International Conference on Advanced Intelligent Mechatronics (AIM)**, 2016, Banff. Proceedings. Banff: IEEE, 2016. p. 300-305. Disponível em: <https://doi.org/10.1109/AIM.2016.7576783>. Acesso em: 19 mai. 2022.

LÓPEZ-GRANADOS, F. *et al.* Object-based early monitoring of a grass weed in a grass crop using high resolution UAV imagery. **Agronomy for Sustainable Development**, [s.l.], v. 36, n. 4, p. 67, dez. 2016. Disponível em <https://link.springer.com/article/10.1007/s13593-016-0405-7>. Acesso em: 19 mai. 2022.

LOUHAICHI, M.; BORMAN, M. M.; JOHNSON, D. E. Spatially Located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. **Geocarto International**, v. 16, n. 1, p. 65-70, 2001. Disponível em: <https://doi.org/10.1080/10106040108542184>. Acesso em: 19 mai. 2022.

MACHADO, J. L. *et al.* Pragas do cafeeiro: bioecologia e manejo integrado. **Informe Agropecuário**, Belo Horizonte, v. 35, n. 280, p. 7-13. 2014.

MARCHANT, J. A.; ONYANGO, C. M. Shadow-invariant classification for scenes illuminated by daylight. **Journal of the Optical Society of America A**, v. 17, n. 11, p. 1952-1961, 2000. Disponível em: <https://doi.org/10.1364/JOSAA.17.001952>. Acesso em: 19 mai. 2022.

MARQUES, K. B. S. C. **Infestação e parasitismo de *Leucoptera coffeella* (Guérin-Mèneville & Perrottet, 1842) em cafeeiros em transição agroecológica**. 2017. 59 p. Dissertação (Mestrado em Agronomia/Entomologia) - Universidade Federal de Lavras, Lavras, 2017. Disponível em: <http://repositorio.ufla.br/>. Acesso em: 19 mai. 2022.

MATIELLO, J. B. *et al.* **Cultura de café no Brasil**: manual de recomendações. Rio de Janeiro; Varginha: SARC/PROCAFÉ, 2010.

MEDEIROS, F. A. *et al.* Utilização de um veículo aéreo não-tripulado em atividades de imageamento georeferenciado. **Ciência rural**, Santa Maria, v. 38, p. 2375- 2378, nov. 2008. Disponível em: <https://doi.org/10.1590/S0103-84782008000800046>. Acesso em: 19 mai. 2022.

MESQUITA, C. M. *et al.* **Manual do café**: distúrbios fisiológicos, pragas e doenças do cafeeiro (*Coffea arabica* L.). Belo horizonte: EMATER-MG, 2016. Disponível em: <https://revistacafeicultura.com.br/?mat=62999>. Acesso em: 19 mai. 2022.

MEYER, G. E.; HINDMAN, T. H.; LAKSMI, K. Machine vision detection parameters for plant species identification. In: Precision Agriculture and Biological Quality. **Proceedings...** Boston, MA, United States, 1999, p. 327-335; v. 3543. Disponível em: <https://doi.org/10.1117/12.336896>. Acesso em: 19 mai. 2022.

MIRANDA, R. de Q. **Avaliação integrada da variação espacial e temporal do balanço hídrico na Caatinga**. Tese (Doutorado em Desenvolvimento e Meio Ambiente) - Universidade Federal de Pernambuco, Recife, 2017. Disponível em: <https://repositorio.ufpe.br/handle/123456789/25188>. Acesso em: 02 fev. 2022.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to Linear Regression Analysis**. 5. ed. [S.l.]: Wiley, 2012.

MONTGOMERY, D. C.; RUNGER, G. C. **Applied Statistics and Probability for Engineers**. 6. ed. [S.l.]: John Wiley, 2013.

MORAIS, N. F. **Análise de regressão linear com estudo de caso em acidentes de trânsito**. 2010. Monografia (Bacharel em Estatística) - Universidade Estadual da Paraíba, Campina Grande, Paraíba, 2010. Disponível em: <http://dspace.bc.uepb.edu.br/jspui/handle/123456789/3787>. Acesso em: 19 mai. 2022.

MOZGERIS, G. *et al.* Imaging from manned ultra-light and unmanned aerial vehicles for estimating properties of spring wheat. **Precision Agriculture**, v. 19, p. 876-894, 2018. Disponível em: <https://link.springer.com/article/10.1007/s11119-018-9562-9>. Acesso em: 19 mai. 2022.

MULLA, D. J. Twenty-Five Years of Remote Sensing in Precision Agriculture: Key Advances and Remaining Knowledge Gaps. **Biosystems Engineering**, v.1 14, p. 358-371, 2013. Disponível em: <https://doi.org/10.1016/j.biosystemseng.2012.08.009>. Acesso em: 19 mai. 2022.

NEBIKER, S. *et al.* A light-weight multispectral sensor for micro UAV: Opportunities for very high resolution airborne remote sensing. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, [s.l.], v.37, p.1193–1200, jan. 2008. Disponível em: <https://www.isprs.org/publications/archives.aspx>. Acesso em: 19 mai. 2022.

NORIEGA, A.; ANDERSON, R. P. Linear-optimization-based path planning algorithm for an agricultural UAV. In: **AIAA INFOTECH@AEROSPACE CONFERENCE**, 2016, San Diego. Proceedings. [San Diego]: AIAA, 2016. v. 1, p. 1003. Disponível em: <https://doi.org/10.2514/6.2016-1003>. Acesso em: 19 mai. 2022.

PEÑA, J. M. *et al.* Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images. **PloS one**, [s.l.], v. 8, n. 10, p. 77151, out. 2013. Disponível em: <https://doi.org/10.1371/journal.pone.0077151>. Acesso em: 19 mai. 2022.

QIN, J. *et al.* A chlorinated copolymer donor demonstrates a 18.13% power conversion efficiency. **Journal of Semiconductors**, v. 42, n. 1, p. 010501, 2021. Disponível em: <https://doi.org/10.1088/1674-4926/42/1/010501>. Acesso em: 19 maio 2022.

RANGO, A. *et al.* Unmanned aerial vehicle-based remote sensing for rangeland assessment, monitoring, and management. **Journal of Applied Remote Sensing**, [s.l.], v. 3, n. 1, p. 033542, ago. 2009. Disponível em <https://doi.org/10.1117/1.3216822>. Acesso em: 19 mai. 2022.

REIS, P. R.; SOUZA, J. C. de. Manejo integrado das pragas do cafeeiro em Minas Gerais. **Informe Agropecuário**, Belo Horizonte, v. 19, n. 193, p. 17-25, 1998.

REIS, P. R.; SOUZA, J. C. de. Pragas do cafeeiro. In: RENA, A. B. *et al.* (Ed.). **Cultura do cafeeiro**: fatores que afetam a produtividade. Piracicaba: POTAFOS, 1986. p. 323-378.



REIS, P. R.; SOUZA, J. C. de; MELLES, C. C. A. Pragas do cafeeiro. **Informe Agropecuário**, Belo Horizonte, v. 10, n. 109, p. 3-57, 1984.

REIS, P. R.; SOUZA, J. C. de; VENZON, M. Manejo ecológico das principais pragas do cafeeiro. **Informe Agropecuário**, v. 23, n. 214/215, p. 83-99, 2002.

REIS, P. R.; SOUZA, J. C. Insetos na folha. **Cultivar**, Pelotas, v. 4, n. 38, p. 30-33, 2002.

RIFFEL, C. T. *et al.* Densidade amostral aplicada ao monitoramento georreferenciado de lagartas desfolhadoras na cultura da soja. **Ciência Rural**, v. 42, n. 12, p. 2112-2119, 2012. Disponível em: <https://doi.org/10.1590/S0103-84782012005000116>. Acesso em: 19 mai. 2022.

ROMERO, M. *et al.* Vineyard water status estimation using multispectral imagery from an UAV platform and machine learning algorithms for irrigation scheduling management. **Computers and Electronics in Agriculture**, v. 147, p. 109-117, 2018. Disponível em: <https://doi.org/10.1016/j.compag.2018.02.013>. Acesso em: 19 mai. 2022.

SAEYS, Y.; INZA, I.; LARRAÑAGA, P. A review of feature selection techniques in bioinformatics. **Bioinformatics**, Oxford, v. 23, n. 19, p. 2507-2517, 2007. Disponível em: <https://doi.org/10.1093/bioinformatics/btm344>. Acesso em: 19 mai. 2022.

SAFONOVA, A. *et al.* Detection of Fir Trees (*Abies sibirica*) Damaged by the Bark Beetle in Unmanned Aerial Vehicle Images with Deep Learning. **Remote Sens**, v. 11, n. 6, p. 643, 2019. Disponível em: <https://doi.org/10.3390/rs11060643>. Acesso em: 19 maio 2022.

SAITO, T.; REHMSMEIER, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. **PLoS ONE**, v. 10, n. 3, p. e0118432, 2015. Disponível em: <https://doi.org/10.1371/journal.pone.0118432>. Acesso em: 19 maio 2022.

SALAKEN, S. M. *et al.* Forecasting Bike Sharing Demand Using Fuzzy Inference Mechanism. **Neural Information Processing**, p. 567-574, 2015. Disponível em: [https://link.springer.com/chapter/10.1007/978-3-319-26555-1\\_64](https://link.springer.com/chapter/10.1007/978-3-319-26555-1_64). Acesso em: 19 mai. 2022.

SANTESTEBAN, L.G. *et al.* High-resolution UAV-based thermal imaging to estimate the instantaneous and seasonal variability of plant water status within a vineyard. **Agricultural Water Management**, v. 183, p. 49-59, 2017. Disponível em: <https://doi.org/10.1016/j.agwat.2016.08.026>. Acesso em: 19 mai. 2022.

SIEBERT, S.; TEIZER, J. Mobile 3D mapping for surveying earthwork projects using an Unmanned Aerial Vehicle (UAV) system. **Automation in Construction**, v. 41, p. 1-14, maio 2014. Disponível em: <https://doi.org/10.1016/j.autcon.2014.01.004>. Acesso em: 19 mai. 2022.

SILVA, F. M. da.; ALVES, M. de C. **Cafecultura de precisão**. 1. ed. Lavras: Ed, UFLA 2013. n. 227 p. 1-25.

SILVA, L.F.M. *et al.* Área foliar e índice de área foliar de plantas de tomateiro em função de níveis de reposição de evapotranspiração da cultura. **In: CONIRD – Congresso Nacional de**

- Irrigação e Drenagem**, São Cristovão, 2015, p.1676-1681. Disponível em: <http://www.abid.org.br/cd-xxv-conird/xxvconird.html>. Acesso em: 19 mai. 2022.
- SILVA, R. A. *et al.* Pragas do cafeeiro: bioecologia e manejo integrado. **Informe Agropecuário**, Belo Horizonte, v. 35, n. 280, p. 7-13. 2014.
- SOUZA, A. L. P. *et al.* Avaliação de Três Métodos de Obtenção do Índice de Área Foliar para Cultura da Soja. **Nativa**, v. 7, n. 3, p. 284, 2019. Disponível em: <https://doi.org/10.31413/nativa.v7i3.7545>. Acesso em: 19 mai. 2022.
- SOUZA, J. C.; REIS, P. R. RIGITANO, R. L. de O. **Bicho mineiro do cafeeiro**: biologia, danos e manejo integrado. 2. ed. Belo Horizonte: EPAMIG,1998. (EPAMIG, Boletim Técnico, 54).
- STÁNCZYK, U.; JAIN, L. C. **Feature selection for data and Pattern Recognition**. New York: Springer, 2015.
- TIEN, L. **forester**: An AutoML R package for Tree-based Models. R-bloggers, 2021. Disponível em: <https://www.r-bloggers.com/2021/09/forester-an-automl-r-package-for-tree-based-models/>. Acesso em: 19 mai. 2022.
- TOKEKAR, P. *et al.* Sensor planning for a symbiotic UAV and UGV system for precision agriculture. **IEEE Transactions on Robotics**, [s.l.], v. 32, n. 6, p. 1498-1511, 2016. Disponível em: <https://doi.org/10.1109/IROS.2013.6697126>. Acesso em: 19 mai. 2022.
- TORRES-SÁNCHEZ, J. *et al.* Configuration and specifications of an unmanned aerial vehicle (UAV) for early site specific weed management. **PLoS ONE**, v.8, e58210, 2013. Disponível em: <https://doi.org/10.1371/journal.pone.0058210>. Acesso em: 19 mai. 2022.
- TORRES-SÁNCHEZ, J. *et al.* High-throughput 3-d monitoring of agricultural-tree plantations with unmanned aerial vehicle (UAV) technology. **PLoS ONE**, v.10, e0130479, 2015. Disponível em <https://doi.org/10.1371/journal.pone.0130479>. Acesso em: 19 mai. 2022.
- TORRES-SÁNCHEZ, J. *et al.* Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. **Computers and Electronics in Agriculture**, [s.l.], v. 103, p. 104- 113, abr. 2014. Disponível em: <https://doi.org/10.1016/j.compag.2014.02.009>. Acesso em: 19 mai. 2022.
- TUCKER, C. J. Red and photographic infrared linear combinations for monitoring vegetation. **Remote Sensing of Environment**, v. 8, n. 2, p. 127-150, 1979. Disponível em: [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0). Acesso em: 19 maio 2022.
- UYSAL, M.; TOPRAK, A. S.; POLAT, N. DEM generation with UAV Photogrammetry and accuracy analysis in Sahitler hill. **Measurement**, [s.l.], v. 73, p. 539-543, 2015. Disponível em: <https://doi.org/10.1016/j.measurement.2015.06.010>. Acesso em: 19 mai. 2022.
- VENZON, M. *et al.* **Controle alternativo de pragas e doenças**: opção ou necessidade. Belo Horizonte: EPAMIG, 2021.
- WANG, X. *et al.* Extraction of vegetation information from visible unmanned aerial vehicle

images. **Transactions of the Chinese Society of Agricultural Engineering**, v. 31, n. 5, p. 152–159, 2015. Disponível em: <https://doi.org/10.3969/j.issn.1002-6819.2015.05.022>. Acesso em: 19 maio 2022.

WHITEHEAD, K.; HUGENHOLTZ, C. H. Remote sensing of the environment with small unmanned aircraft systems (UASs), part 1: A review of progress and challenges. **Journal of Unmanned Vehicle Systems**, [s.l.], v. 2, n. 3, p. 69-85, 2014. Disponível em: <https://doi.org/10.1139/juvs-2014-0006>. Acesso em: 19 mai. 2022.

WOEBBECKE, D. M. *et al.* Color Indices for Weed Identification Under Various Soil, Residue, and Lighting Conditions. **Transactions of the ASAE**, v. 38, n. 1, p. 259-269, 1995. Disponível em: <https://doi.org/10.13031/2013.27838>. Acesso em: 19 maio 2022.

XIANG, H.; TIAN, L. Development of a low-cost agricultural remote sensing system based on an autonomous unmanned aerial vehicle (UAV). **Biosystems Engineering**, v. 108, n. 2, p. 174-190, 2011a. Disponível em: <https://doi.org/10.1016/j.biosystemseng.2010.11.010>. Acesso em: 19 mai. 2022.

XIANG, H.; TIAN, L. Method for automatic georeferencing aerial remote sensing (RS) images from an unmanned aerial vehicle (UAV) platform. **Biosystems Engineering**, [s.l.], v. 108, n. 2, p. 104-113, 2011b. Disponível em: <https://doi.org/10.1016/j.biosystemseng.2010.11.003>. Acesso em: 19 mai. 2022.

XU, D.; PU, Y.; GUO, X. A Semi-Automated Method to Extract Green and Non-Photosynthetic Vegetation Cover from RGB Images in Mixed Grasslands. **Sensors**, 20(3), 6870, 2020. Disponível em: <https://doi.org/10.3390/s20236870>. Acesso em: 19 mai. 2022.

YUE, L. *et al.* Demand Forecasting by Using Support Vector Machine. **Third International Conference on Natural Computation (ICNC 2007)**, 2007, p. 272-276. Disponível em: <https://doi.org/10.1109/ICNC.2007.324>. Acesso em: 19 mai. 2022.

ZAJKOWSKI, T.J. *et al.* Evaluation and use of remotely piloted aircraft systems for operations and research-RxCADRE 2012. **International Journal of Wildland Fire**, v. 25, n. 1, p. 114-128, 2016. Disponível em: <https://doi.org/10.1071/WF14176>. Acesso em: 19 mai. 2022.

ZARCO-TEJADA, P. *et al.* Estimating leaf carotenoid content in vineyards using high resolution hyperspectral imagery acquired from an unmanned aerial vehicle (UAV). **Agricultural and Forest Meteorology**, v. 171-172, p. 281-294, 2013. Disponível em: <https://doi.org/10.1016/j.agrformet.2012.12.013>. Acesso em: 19 mai. 2022.

ZHANG, C.; KOVACS, J. M. The application of small unmanned aerial systems for precision agriculture: a review. **Precision agriculture**, [s.l.], v. 13, n. 6, p. 693-712, 2012. Disponível em: <https://link.springer.com/article/10.1007/s11119-012-9274-5>. Acesso em: 19 mai. 2022.

ZHANG, C.; WALTERS, D.; KOVACS, J.M. Applications of low altitude remote sensing in agriculture upon farmers' requests— a case study in Northeastern Ontario, Canada. **PLoS ONE**, v.9, e112894, 2014. Disponível em: <https://doi.org/10.1371/journal.pone.0112894>. Acesso em: 19 mai. 2022.

## APÊNDICE A - Código fonte

```

#-----
# Coffee leaf-miner v0.1
#-----

#Packages -----
library(dplyr); library(DALEX);library(forester);library(ranger);library(caret)

# Dataset -----
ds<-read.csv("https://raw.githubusercontent.com/Leprechault/trash/main/bicho_F.csv")
str(ds)
# 'data.frame':    11986 obs. of  17 variables:
# $ classe: int  2 2 2 1 1 2 1 2 1 2 ...
# $ x     : chr "499.458.032" "4.994.570.145" "4.994.559.378" "4.994.546.634" ...
# $ y     : chr "7.655.633.888" "7.655.634.235" "7.655.634.551" "7.655.635.025" ...
# $ cive  : num 0.565 0 0.532 0.498 0.543 ...
# $ exg   : num 0.45 0 0.486 0.505 0.47 ...
# $ exgr  : num 0.383 0 0.4 0.5 0.418 ...
# $ exr   : num 0.631 0 0.636 0.472 0.586 ...
# $ gli   : num 0.359 0 0.353 0.45 0.382 ...
# $ iaf   : num 0.808 0 0.836 0.851 0.818 ...
# $ ngrdi : num 0.69 0 0.687 0.724 0.691 ...
# $ ri    : num 0.31 0 0.313 0.276 0.309 ...
# $ tgi   : num 0.755 0 0.719 0.883 0.79 ...
# $ varig : num 0.743 0 0.739 0.773 0.743 ...
# $ vdvi  : num 0.751 0 0.761 0.764 0.752 ...
# $ r     : num 0.471 0 0.431 0.403 0.441 ...
# $ g     : num 0.508 0 0.502 0.424 0.478 ...
# $ b     : num 0.471 0 0.431 0.403 0.441 ...

# Remove x and y, convert classe to factor

```

```

ds<-ds[,-c(2:3)]
ds$classe<-as.factor(ds$classe)
str(ds)
# 'data.frame':      11986 obs. of  15 variables:
# $ classe: Factor w/ 2 levels "1","2": 2 2 2 1 1 2 1 2 1 2 ...
# $ cive  : num  0.565 0 0.532 0.498 0.543 ...
# $ exg   : num  0.45 0 0.486 0.505 0.47 ...
# $ exgr  : num  0.383 0 0.4 0.5 0.418 ...
# $ exr   : num  0.631 0 0.636 0.472 0.586 ...
# $ gli   : num  0.359 0 0.353 0.45 0.382 ...
# $ iaf   : num  0.808 0 0.836 0.851 0.818 ...
# $ ngrdi : num  0.69 0 0.687 0.724 0.691 ...
# $ ri    : num  0.31 0 0.313 0.276 0.309 ...
# $ tgi   : num  0.755 0 0.719 0.883 0.79 ...
# $ varig : num  0.743 0 0.739 0.773 0.743 ...
# $ vdvi  : num  0.751 0 0.761 0.764 0.752 ...
# $ r     : num  0.471 0 0.431 0.403 0.441 ...
# $ g     : num  0.508 0 0.502 0.424 0.478 ...
# $ b     : num  0.471 0 0.431 0.403 0.441 ...

# Build the model pipeline
bicho_model_general <- forester::forester(data = ds,
target = "classe", type = "classification",
metric = "precision", tune = FALSE)

# model precision recall f1 accuracy auc
# -----
# Ranger 0.9997651 0.9817762 0.9906890 0.9916562 0.9907929
# XGBoost 0.7963701 0.4149942 0.5456476 0.6875261 0.6637126
# CatBoost 0.7053632 0.2396770 0.3577824 0.6109720 0.5785288
# LightGBM 0.7007028 0.3910035 0.5019248 0.6491448 0.6265887

# The best model based on rmse metric is Ranger.

```

```

# Save the better model
saveRDS(bicho_model_general$model,"bicho_ranger_model_final.rds")

# Open the model
model_F <- readRDS("bicho_ranger_model_final.rds")
model_F

# Ranger result

# Call:
# ranger::ranger(form, data = data, classification = is_classif)

# Type: Classification
# Number of trees: 500
# Sample size: 9588
# Number of independent variables: 14
# Mtry: 3
# Target node size: 1
# Variable importance mode: none
# Splitrule: gini
# OOB prediction error: 49.10 %

#-----
# Bootstrap for accuracy, sensibility and specificity mean and SD

res.svm<-NULL

for(s in 1:99){

# Splitting the dataset into the Training set and Test set
RES<-ds
RES$id <- 1:nrow(RES)

```

```

# Training using 90% for cal/train and 10% for val
train<-RES %>% sample_n(length(RES[,1])*0.9)
train<-as.data.frame(train)

#Using 10% for val = 181
rest.RES <- dplyr::anti_join(RES, train, by = 'id')
test<-rest.RES %>% sample_n(length(RES[,1])*0.1)
test<-as.data.frame(test)
train<-train[,-c(ncol(train))]
test<-test[,-c(ncol(test))]

model_F<- ranger(
  formula      = classe ~.,
  data         = train,
  num.trees    = 500,
  mtry         = 3,
  min.node.size = 1,
  sample.fraction = 1.0
)

test_predictions <- predict(model_F, data = test[,-c(1)])
# FORESTER MECHANISM: 1 -> 1 and 2 -> 0.
test_predictions$predictions <-ifelse(test_predictions$predictions == 1, "1","2")

# Making the Confusion Matrix
cm_val = table(test[,1], test_predictions$predictions)
CM<-confusionMatrix(cm_val, positive = "2") #Sensitivity and specificity
res.svm<-rbind(res.svm,c(s,CM$overall[1],CM$byClass[1],CM$byClass[2],CM$byClass[3],
CM$byClass[4],CM$byClass[5],CM$byClass[6],CM$byClass[7],CM$byClass[8],CM$byClass[9],
CM$byClass[10],CM$byClass[11]))
res.svm<-data.frame(res.svm)
colnames(res.svm)<-c("simulation","accuracy", "sensitivity",
"specificity","pospredvalue","negpredvalue",

```

```

"precision", "recall", "F1", "prevalence", "detection rate", "detection prevalence",
"balanced accuracy")
  # Save svm results
write.csv(res.svm, "res ranger simulation s99 v2.csv", row.names = FALSE)
  }
  #
  # Final results
result<-read.csv("res ranger simulation s99.csv")
  # Accuracy mean and SD
  mean(result$accuracy)
  #[1] 0.8967893
  smd <- function(x, ...) sd(x, ...)/sqrt(length(x)-1)
  smd(result$accuracy)
  #[1] 0.0007765801
  # Sensitivity mean and SD
  mean(result$sensitivity)
  #[1] 0.888928
  smd(result$sensitivity)
  #[1] 0.001146505
  # Specificity mean and SD
  mean(result$specificity)
  #[1] 0.9071427
  smd(result$specificity)
  #[1] 0.001222945
  #-----

```