

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE GEOGRAFIA, GEOCIÊNCIAS E SAÚDE COLETIVA

MARIA CECÍLIA LEMES SANTOS

**MODELOS MULTIESPECTRAIS PARA A ESTIMATIVA
DE PRODUTIVIDADE DA CULTURA DA SOJA (*Glycine max L.*).**

Monte Carmelo

2025

MARIA CECÍLIA LEMES SANTOS

**MODELOS MULTIESPECTRAIS PARA A ESTIMATIVA
DE PRODUTIVIDADE DA CULTURA DA SOJA (*Glycine max L.*).**

Trabalho de Conclusão de Curso apresentado à
Universidade Federal de Uberlândia como
requisito parcial para obtenção do título de
bacharel em Engenharia de Agrimensura e
Cartográfica.

Orientador: Prof. Dr. George Deroco Martins

Monte Carmelo

2025

MARIA CECÍLIA LEMES SANTOS

**MODELOS MULTIESPECTRAIS PARA A ESTIMATIVA DE
PRODUTIVIDADE DA CULTURA DA SOJA (*Glycine max L.*)**

Trabalho de Conclusão de Curso
apresentado à Universidade Federal de
Uberlândia como requisito parcial para
obtenção do título de bacharel em
Engenharia de Agrimensura e Cartográfica.

Banca Examinadora:

Prof. Dr. George Deroco Martins
Universidade Federal de Uberlândia

Eng. Filipe Vieira da Silva
Universidade Federal de Uberlândia

Eng. Lucas Henrique Vicentini Viana de Carvalho
Universidade Federal de Uberlândia

AGRADECIMENTOS

Agradeço primeiramente a Deus por me conceder saúde, sabedoria e força para concluir esta etapa tão importante da minha vida. Sua presença constante me guiou e iluminou meus passos durante toda a jornada.

Aos meus pais e minha irmã, minha eterna gratidão pelo amor incondicional, apoio e incentivo em todos os momentos. Vocês são meu porto seguro e minha maior inspiração.

Ao meu companheiro, Lincon Rodrigues, pela paciência, compreensão e apoio constantes durante toda a realização deste trabalho e da minha formação. Compartilhar a vida com você tornou esse processo muito mais leve.

Aos meus amigos por estarem sempre ao meu lado, celebrando minhas conquistas e me oferecendo palavras de conforto nos momentos de dificuldade. A amizade de vocês é um presente inestimável. Em especial ao Filipe Vieira e a Rayssa Santos, que desenvolveram e participaram de cada detalhe desse trabalho, sem vocês minha trajetória não teria sido tão especial, amo vocês.

Ao meu orientador, Prof. Dr. George Deroco Martins, expresso minha mais profunda gratidão pela dedicação, paciência e conhecimento compartilhados. Sua orientação foi fundamental para o sucesso deste trabalho. Sua experiência e sabedoria foram essenciais para a construção deste conhecimento de minha formação.

À Universidade Federal de Uberlândia, agradeço pela oportunidade de realizar este trabalho e pelo ensino de excelência que recebi. A instituição me proporcionou um ambiente propício para o desenvolvimento intelectual e pessoal.

À empresa Lallemand Plant Care que forneceu os dados que possibilitaram a execução deste trabalho.

Agradeço também a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho. A cada um de vocês, minha sincera gratidão.

RESUMO

De acordo com a Companhia Nacional de Abastecimento (CONAB), atualmente o maior produtor de soja do mundo é o Brasil com uma produção de 151,4 milhões de toneladas. Para produções nessas proporções se faz necessário o uso do Sensoriamento Remoto para uma melhora no manejo através da localização de pragas, análise de deficiências do solo, como também para estimativa da produtividade. A estimativa da produtividade é uma ferramenta importante para o setor agroindustrial, pois auxilia em um manejo mais eficiente da cultura, na comercialização, no aumento da produtividade, além de otimizar a condução e simplificar o escoamento e armazenamento dos grãos. O atual processo para estimativa de produtividade demanda de muito tempo, equipes de campo e conhecimentos específicos. Essa técnica é executada em etapas no campo, começando com a contagem de 10 plantas, das vagens, dos grãos por vagem e por fim, é usada uma equação para estimativa de toda área. Esse trabalho avaliou e validou um novo método para estimativa de produtividade da soja, utilizando imagens multiespectrais, visto que, as imagens tornam possível o mapeamento em grande escala devido à dimensão de área que elas compreendem, possibilitando a estimativa de forma mais rápida, precisa e com menores custos. Além disso, foi verificado que o R4 (vagem completamente desenvolvida) é o melhor estágio fenológico para aquisição e predição da produtividade e também foi avaliado que uma série temporal não é capaz de melhorar o modelo de estimativa. Usando imagens multiespectrais do satélite Planet e pontos de produtividade georreferenciados, obtidos de forma pontual na área de estudo, localizada em Primavera do Leste, Mato Grosso, foi estimada a produtividade da soja e gerado um mapa de produtividade. Foram utilizadas as reflectâncias, os valores de produtividade, as 4 bandas da imagem (Red, Green, Blue e Near InfraRed) e um conjunto de 15 índices de vegetação para geração do modelo de estimativa, sendo eles: OSAVI, GOSAVI, ATSAVI, NDWI, NDVI, GNDVI, MTVI2, VARI, ARVI, RECI, GLI, GCI, SIPI, CVI, TGI. Os modelos foram desenvolvidos através dos algoritmos de aprendizado de máquina disponíveis no WEKA, onde o *Random Forest* apresentou um desempenho superior nas modelagens com um *Root Mean Squared Error* (RMSE) de 0,2186 e um coeficiente de correlação (R) de 0,8968. A metodologia proposta, baseada em imagens multiespectrais e algoritmos de aprendizado de máquina, oferece alta acurácia e eficiência, contribuindo para o planejamento e gestão otimizados da produção agrícola.

Palavras-chave: agricultura de precisão; soja; cultura temporária; modelos espectrais de predição da produtividade.

ABSTRACT

According to the National Supply Company (CONAB), Brazil is currently the largest soybean producer in the world, with a production of 151.4 million tons. To produce such large quantities, it is necessary to use Remote Sensing to improve management by locating pests, analyzing soil deficiencies, and also to estimate productivity. Estimating productivity is an important tool for the agroindustrial sector, as it helps in more efficient crop management, marketing, increasing productivity, in addition to optimizing the conduction and simplifying the flow and storage of grains. The current process for estimating productivity requires a lot of time, field teams, and specific knowledge. This technique is performed in stages in the field, starting with counting 10 plants, the pods, the grains per pod, and finally, an equation is used to estimate the entire area. This study evaluated and validated a new method for estimating soybean productivity using multispectral images, since the images allow for large-scale mapping due to the area dimension they encompass, enabling faster, more accurate and lower-cost estimation. In addition, it was found that R4 (fully developed pod) is the best phenological stage for acquiring and predicting productivity and it was also assessed that a time series is not capable of improving the estimation model. Using multispectral images from the Planet satellite and georeferenced productivity points, obtained punctually in the study area, located in Primavera do Leste, Mato Grosso, soybean productivity was estimated and generated a productivity map. The reflectances, productivity values, the 4 image bands (Red, Green, Blue and Near InfraRed) and a set of 15 vegetation indices were used to generate the estimation model, namely: OSAVI, GOSAVI, ATSAVI, NDWI, NDVI, GNDVI, MTVI2, VARI, ARVI, RECI, GLI, GCI, SIPI, CVI, TGI. The models were developed through the machine learning algorithms available in WEKA, where Random Forest presented a superior performance in the models with a Root Mean Squared Error (RMSE) of 0.2186 and a correlation coefficient (R) of 0.8968. The proposed methodology, based on multispectral images and machine learning algorithms, offers high accuracy and efficiency, contributing to the optimized planning and management of agricultural production.

Keywords: precision agriculture; soybean; temporary crops; spectral productivity prediction models.

LISTA DE ILUSTRAÇÕES

Figura 1 -	Estádios fenológicos da soja.....	16
Figura 2 -	Descrição dos estádios fenológicos e duração das fases.....	18
Figura 3 -	Esquema de um projeto completo de Aprendizado de Máquina.....	23
Figura 4 -	Mapa de localização da área de estudo.....	25
Figura 5 -	Limite da área de estudo e pontos de produtividade.....	27
Figura 6 -	Diagrama de fluxo.....	28
Figura 7 -	Relação pixel e ponto de produtividade.....	30
Figura 8 -	Gráficos de correlação dos IV's.....	31
Figura 9 -	Resultados dos processamentos.....	36
Figura 10 -	Avaliação da precisão do modelo: R e RMSE.....	37
Figura 11 -	Espacialização da produtividade.....	41

LISTA DE TABELAS

Tabela 1 -	Descrição sumária dos estádios vegetativos da soja.....	16
Tabela 2 -	Descrição sumária dos estádios reprodutivos da soja.....	17
Tabela 3 -	Índices para aplicação na agricultura.....	22
Tabela 4 -	Índices usados na criação dos modelos preditivos	32
Tabela 5 -	Bandas do Satélite Planet.....	33
Tabela 6 -	Resultado da série teporal.....	38
Tabela 7 -	Váriaveis que compõem o melhor modelo para estimativa.....	40

LISTA DE ABREVIATURAS E SIGLAS

ATSAVI -	Adjusted transformed soil-adjusted VI
AP –	Agricultura de Precisão
ARVI -	Atmospherically Resistant Vegetation Index
B -	Blue
CONAB –	Companhia Nacional de Abastecimento
CVI -	Chlorophyll Vegetation Index
EMBRAPA -	Empresa Brasileira de Pesquisa Agropecuária
EVI –	Enhanced Vegetation Index
G -	Green
GCI -	Green Chlorophyll Index
GLI	Green leaf index
GNDVI –	Green Normalized Difference Vegetation Index
GOSAVI -	Green Optimized Soil Adjusted Vegetation Index
IV -	Índice de vegetação
MTVI2 -	Modified Triangular Vegetation Index 2
NDVI –	Normalized Difference Vegetation Index
NIR -	Near InfraRed
NN –	Neural Net
NDWI -	Normalized Difference Water Index
R -	Red
ReCI -	Red Edge Chlorophyll Index
REM -	Radiação Eletromagnética
RF -	Random Forest
RMSE -	Root Mean Squared Error
RT -	Random Tree
SAVI -	Soil Adjusted Vegetation Index
SIPI -	Structure Insensitive Pigment Index
SVM –	Support Vector Machine
TGI -	Triangular Greenness Index
USDA -	United States Department of Agriculture
VARI -	Visible Atmospheric Resistant Index

SUMÁRIO

1 INTRODUÇÃO.....	12
2 OBJETIVOS.....	14
2.1 Objetivo geral.....	14
2.2 Objetivos específicos	14
3 REVISÃO BIBLIOGRÁFICA	15
3.1 Cultura da Soja	15
3.2 Sensoriamento remoto na cultura da soja	18
3.3 Índices utilizados na predição da produtividade	20
3.4 Aprendizado de máquinas na predição.....	23
4 MATERIAL E MÉTODOS	25
4.1 Área de estudo	25
4.2 Material	26
4.3 Métodos.....	27
4.3.1 Aquisição dos dados	29
4.3.2 Pré-processamento dos dados	29
4.3.3 Processamento dos dados	31
4.3.4 Validação e melhoramento do modelo	34
4.4 Representação Gráfica	35
5 RESULTADOS E DISCUSSÕES.....	36
5.1 Modelo com melhor desempenho.....	36
5.2 Estádio fenológico mais adequado para estimativa da soja.....	37
5.3 Série Temporal.....	38
5.4 Componentes do melhor modelo.....	39
5.5 Espacialização da produtividade da soja.....	41
6 CONCLUSÃO.....	43
REFERÊNCIAS	44

1 INTRODUÇÃO

De acordo com a United States Department of Agriculture (USDA, 2023), a produção mundial de soja na safra 2022/23 foi de 370,11 milhões de toneladas, com 136,54 milhões de hectares plantados. O Brasil é o maior produtor mundial segundo a Companhia Nacional de Abastecimento (CONAB, 2023), com uma produção 155.736,5 milhões de toneladas, em 44.031,7 milhões de hectares. Para a safra mundial de 2023/24, a USDA prevê um aumento de 32,68 milhões de toneladas do grão.

A soja (*Glycine max L*) é uma oleaginosa, pertencente à família *Fabaceae* e é uma cultura das mais importantes para a economia mundial, ocupando o 4º lugar dentre os cereais e grãos mais produzidos no mundo (ADAMA, 2021). Atualmente é o principal produto de exportação do Brasil com mais de 70% da produção exportada. Essa cadeia produtiva movimenta US\$ 100 bilhões/ano no país, a cada 100 dólares exportados, 14 são de soja (Aprosoja, 2016?). Considerando a importância da soja para a economia brasileira, torna-se fundamental obter informações que possam orientar as decisões relacionadas ao plantio, manejo e comercialização dessa cultura (Rizzi; Rudorff *et al.*, 2005 *apud* Andrade, 2014).

Dessa necessidade, surgiu a agricultura de precisão (AP), que se refere a aplicação de geotecnologias na agricultura que permitem um gerenciamento localizado dos cultivares (Lamparelli, 2022), a fim de melhorar a produção agrícola e consequentemente a produtividade dessa atividade. Dentre as técnicas utilizadas na AP, e mais especificamente na estimativa da produtividade, está o Sensoriamento Remoto (SR), que se refere à obtenção de imagens à distância, sobre a superfície terrestre que são adquiridas através de aparelhos denominados sensores remotos (Rudorff, 2011). Por meio da interação da radiação eletromagnética – planta – sensor, apresenta elevado potencial para identificação das condições dos cultivos agrícolas, uma vez que permite avaliação não destrutiva, de forma rápida e com custo relativamente baixo (Crusiol *et al.*, 2017). Através dessa avaliação realizada por SR, se torna possível a predição e estimativa da produtividade da lavoura, que é uma ferramenta necessária no manejo da colheita, no planejamento estratégico da comercialização, no direcionamento do armazenamento e no processamento da produção (Letras Ambientais, 2021).

Atualmente, a estimativa de produtividade da soja pode ser executada de forma eficiente e precisa. O desenvolvimento do processo acontece por meio de idas ao campo onde são primeiramente contabilizadas o número de plantas. Em seguida, deve-se contar as vagens e realizar a contagem da quantidade de grãos por vagem. Para finalizar é realizada a pesagem dos grãos. Pesquisas em Kentucky indicam que 2.500 sementes por quilo são uma estimativa do

peso médio, mas elas podem chegar a 3.400 sementes por quilo. Se o a estimativa de rendimento final parecer excepcionalmente alta ou baixa, pode haver ajustes a serem feitos quanto ao tamanho das sementes. Esse processo de estimativa deve ser repetido de cinco a dez vezes, caso as primeiras cinco estimativas de rendimento forem relativamente próximas, não há necessidade de se repetir o processo em mais locais. Caso sejam divergentes, é necessário a reprodução do processo em demais áreas (Lee, Herbek, 2005).

Equação 1 – Equação de estimativa da produtividade da soja

$$\frac{(Plants \text{ per acre}) * (Pods \text{ per plant}) * (Seeds \text{ per pod})}{\frac{Seeds \text{ per pound}}{Pound \text{ per bushel}}} = (Bushels \text{ per acre}) \quad (1)$$

Onde:

- Plants per acre = plantas por acre;
- Pods per plant = vagens por planta;
- Seeds per pod = sementes por vagem;
- Seeds per pound = sementes por libra;
- Pound per bushel = libra por alqueire;
- Bushels per acre = alqueires por acre.

Fonte: Lee; Herbeck (2005).

O método convencional de estimativa supracitado evidencia o quão moroso é realizá-lo, demandando tempo, recursos financeiros e de mão de obra, além de conhecimento específico da lavoura e da técnica. O método desenvolvido e avaliado neste trabalho (estimativa por meio de imagens multiespectrais) procura uma diminuição significativa no tempo gasto e nos custos. Podendo ser utilizado para estimar a produtividade em diversos estádios do desenvolvimento da planta, visto que o método convencional se restringe à fase reprodutiva próxima, otimizando assim o cálculo de produtividade da cultura da soja.

2 OBJETIVOS

2.1 Objetivo geral

O trabalho em questão tem como objetivo geral determinar o potencial de imagens multiespectrais adquiridas por satélites para a estimativa da produtividade da cultura da soja.

2.2 Objetivos específicos

Os objetivos específicos deste trabalho são:

- Determinar o mais adequado estágio fenológico para aquisição e predição da produtividade da soja;
- Determinar qual o modelo espectral apresenta menor erro para a estimativa;
- Avaliar se a análise de séries temporais é capaz de aumentar a acurácia da estimativa de produtividade;
- Gerar um mapa de produtividade.

3 REVISÃO BIBLIOGRÁFICA

3.1 Cultura da Soja

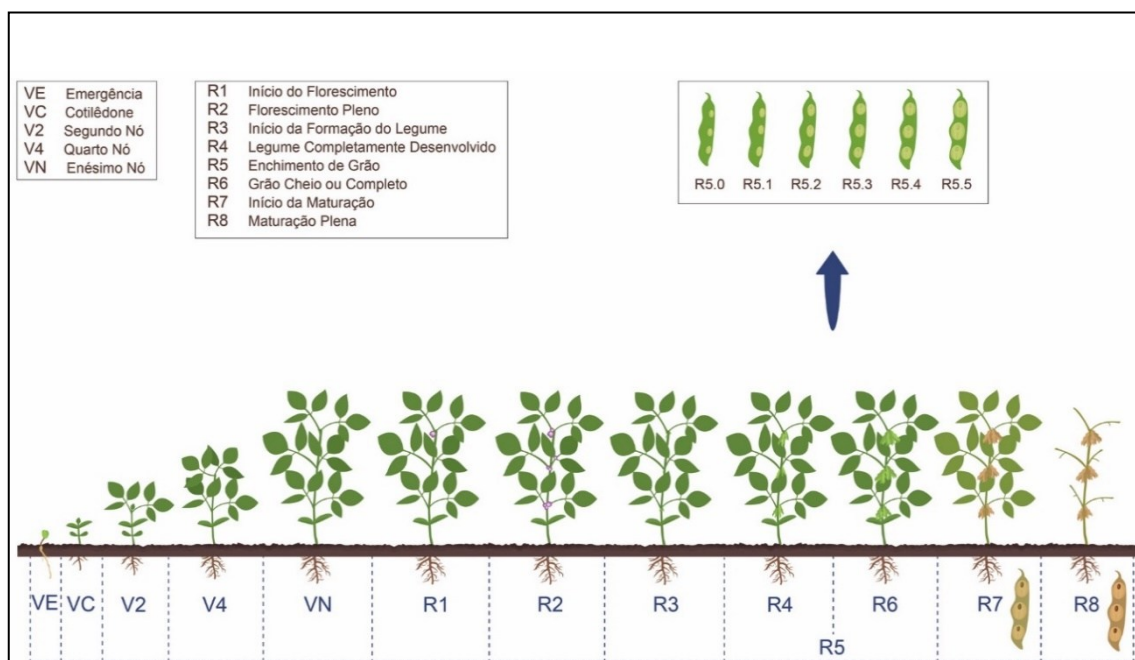
A soja é originária de uma região denominada Manchúria, localizada no nordeste da China. Entre o século XV e o começo do século XVII, os portugueses e espanhóis, seguidos por outros países europeus, exploraram o mundo de maneira intensiva em busca de novas rotas de comércio. Foi então que, em XVII, eles descobriram e levaram a planta da China para o continente europeu. A sua chegada à América se deu por volta dos anos de 1890. Nos Estados Unidos, a princípio, ela foi cultivada como forrageira, sendo utilizada na alimentação animal. No Brasil, a primeira notificação que se tem sobre a soja é de 1882, na Bahia, mas só em 1891 que novas cultivares foram introduzidas, visto que as primeiras não se adaptaram bem. Posteriormente, já nas décadas de 1940 e 1950, o cultivo se expandiu para o Paraguai, México e Argentina (ADAMA, 2021).

Seu nome científico é *Glycine max* L.; faz parte da família Fabaceae (leguminosas), assim como a ervilha, o feijão e a lentilha. Trata-se de um grão rico em proteínas, podendo ser consumido tanto por homens como por animais. No Brasil, é um grão muito plantado e tem formato arredondado com cor amarela. Contém de 18% a 20% de óleo e o farelo representa 79% (tem teor de proteína de 45%). Sua vagem pode ter até quatro sementes e sua estabilidade é limitada pelo ácido linolênico (Jardine, 2021).

A caracterização dos estádios de desenvolvimento da planta de soja é essencial para a descrição dos vários períodos que a lavoura atravessa durante o ciclo da cultura. O uso de uma linguagem unificada na descrição dos estádios de desenvolvimento agiliza o seu entendimento porque facilita a comunicação entre os diversos públicos envolvidos com a cultura de soja (Neumaier *et al*, 2000).

A metodologia de descrição dos estádios de desenvolvimento proposta por Fehr e Caviness (1977) é a mais utilizada no mundo inteiro, ela divide os estádios de desenvolvimento de soja em estádios vegetativos e estádios reprodutivos. Os estádios vegetativos são designados pela letra V e os reprodutivos pela letra R. Com exceção dos estádios VE (emergência) e VC (cotilédone) (Neumaier *et al*, 2000), como mostra a Figura 1.

Figura 1 - Estádios fenológicos da soja.



Fonte: Schwab; Alff (2013).

Os estádios vegetativos, apresentados na Tabela 1, descrevem o desenvolvimento vegetativo da soja. O estágio vegetativo denominado VE representa a emergência da plântula (pequena planta), VC Cotiledonar (cotilédone corresponde a primeira estrutura semelhante à folha (protófilo) ou o primeiro par de folhas embrionárias (Oliveira, [2015?]), V1 completo desenvolvimento das folhas e assim segue até Vn (Farias; Nepomuceno; Neumaier, 2007).

Tabela 1 - Descrição sumária dos estádios vegetativos da soja.

Estádio	Denominação	Descrição
VE	Emergencia	Cotilédones acima da superfície do solo
VC	Cotilédone	Cotilédones completamente abertos
V1	Primeiro nó	Folhas unifolioladas completamente desenvolvidas
V2	Segundo nó	Primeira folha trifoliolada completamente desenvolvida
V3	Terceiro nó	Segunda folha trifoliolada completamente desenvolvida
V4	Quarto nó	Terceira folha trifoliolada completamente desenvolvida
V5	Quinto nó	Quarta folha trifoliolada completamente desenvolvida
V6	Sexto nó	Quinta folha trifoliolada completamente desenvolvida
V...	...	...
Vn	Enésimo nó	Ante-enésima folha trifoliolada completamente desenvolvida

Fonte: Adaptado de Farias; Nepomuceno; Neumaier (2007).

Os estádios reprodutivos, dispostos na Tabela 2, descrevem detalhadamente o período florescimento-maturação. Os estádios reprodutivos abrangem quatro distintas fases do

desenvolvimento reprodutivo da planta, florescimento (R1 e R2), desenvolvimento da vagem (R3 e R4), desenvolvimento do grão (R5 e R6) e maturação da planta (R7 e R8). Na Tabela 2 são apresentados sumariamente os estádios reprodutivos da soja (Farias; Nepomuceno; Neumaier, 2007).

Tabela 2 - Descrição sumária dos estádios reprodutivos da soja.

Estádio	Denominação	Descrição
R1	Início do florescimento	Uma flor aberta em qualquer nó do caule (haste principal)
R2	Florescimento pleno	Uma flor aberta num dos 2 últimos nós do caule com folha completamente desenvolvida
R3	Início da formação da vagem	Vagem com 5mm de comprimento num dos 4 últimos nós do caule com folha completamente desenvolvida
R4	Vagem completamente desenvolvida	Vagem com 2mm de comprimento num dos 4 últimos nós do caule com folha completamente desenvolvida
R5	Início do enchimento do grão	Grão com 3mm de comprimento em vagem num dos 4 últimos nós do caule, com folha completamente desenvolvida
R6	Grão cheio ou completo	Vagem contendo grãos verdes preenchendo as cavidades da vagem de um dos 4 últimos nós do caule, com folha completamente desenvolvida
R7	Início da maturação	Uma vagem normal no caule com coloração de madura
R8	Maturação plena	95% das vagens com coloração de madura

Fonte: Adaptado de Farias, Nepomuceno, Neumaier (2007).

Na Figura 2 é possível acompanhar o desenvolvimento dos estádios fenológicos supracitados com uma média de dias de duração para cada fase, considerando uma safra de 120 dias.

superfície do solo e as plantações. A obtenção dessas informações só é possível devido ao desenvolvimento de sensores e equipamentos de alto desempenho. O sensoriamento remoto trabalha com sensores que utilizam como princípio básico a geração de dados por meio da detecção da REM refletida por uma superfície. Dessa maneira, podemos classificar dois tipos de sensores remotos: ativos e passivos (Totvs, 2020).

- Sensores ativos: não usam o Sol como fonte de radiação, trabalhando com uma fonte de REM própria, como um canhão de laser, capaz de projetar raios na superfície terrestre e calcular o tempo gasto para a reflexão,
- Sensores passivos: relacionam-se com informações externas preexistentes, coletando estímulos refletidos, principalmente, pela luz solar. As informações, então, são tratadas por algoritmos e sistemas integrados, que geram imagens e dados valiosos (Totvs, 2020).

Já a alguns anos é possível notar o aumento exponencial do uso do sensoriamento remoto para análise das lavouras de soja, não só para predição e estimativa da lavoura, mas também para diversos outros aspectos da cultura como estresse hídrico, biomassa, dentre outros. Observa-se que o uso de índices de vegetação (IV's) é muito voltado para análises diversas nos plantios como vigor da planta, doenças e vem sendo de importância absoluta para esse tipo de estudo (Totvs, 2020).

No trabalho de Andrade *et. al* (2022) foram utilizados modelos de regressão linear gerados através do índice NDVI (*Normalized Difference Vegetation Index* – Índice de Vegetação por Diferença Normalizada) para predição da produtividade de grãos de soja nos estádios de desenvolvimento V5-V6 (vegetativo) e R2 (reprodutivo) e obtiveram resultados promissores para predição da produtividade.

Já Fernandes, Veiverberg e Sebem (2009) propõem a estimativa de produtividade de soja por sensoriamento remoto em nível de talhão, avaliando a correlação entre o GNDVI (*Green Normalized Difference Vegetation Index* – Vegetação de Diferença Normalizada Verde) gerado por um sensor ótico orbital com a produtividade do cultivo de soja, assim gerando modelos matemáticos aplicados a estimativa de produção durante o ciclo da cultura. Demonstraram que é possível determinar a produtividade da cultura de soja antecipadamente por sensoriamento remoto em nível de talhão utilizando GNDVI.

No trabalho de Fernando Filho (2019), os dados de reflectância foram utilizados para monitoramento de mosca branca *Bemisia Tabaci* biótipo B (Hemiptera: Aleyrodidae) na soja cultivada sob diferentes níveis de reposição hídrica. O trabalho concluiu que é possível identificar plantas cultivadas sob diferentes níveis de reposição hídrica a partir da coleta dos

dados de reflectância dessas plantas, assim como é possível identificar plantas infestadas por ninfas de *Bemisia tabaci* biótipo B, com alto grau de precisão.

Enquanto Zampier, Gabardo e Guimarães (2018) avaliaram modelos de predição da ferrugem da soja gerados a partir de técnicas de aprendizado de máquina e de imagens de aeronaves remotamente pilotadas (RPA), utilizando as bandas Infravermelho próximo (NIR - do inglês, *Near Infrared*) do vermelho, verde e azul (RGB – do inglês, *Red, Green e Blue*).

Trindade, Alves e Noetzold (2017) partindo da hipótese que a utilização e análise de produtos do sensoriamento remoto, como os índices de vegetação (NDVI e EVI (*Enhanced Vegetation Index* – Índice de Vegetação Melhorado)), auxiliam na previsão da produtividade da lavoura e pode evitar gastos desnecessários do produtor em sua cultura no que diz respeito a correção dos nutrientes do solo, realizou um trabalho com objetivo de avaliar a correlação entre os índices NDVI e o índice EVI extraídos do sensor MODIS com o ciclo fenológico da soja a partir de um perfil espectro temporal, relacionando sua resposta espectral com a produtividade e os atributos físicos e químicos do solo em diferentes talhões comerciais de soja.

Dentre os índices de vegetação Trindade, Alves e Noetzold (2017) concluíram que o EVI foi mais eficaz para as correlações com os atributos físicos e químicos do solo. Com os valores de R^2 (coeficiente de determinação) observados entre a fase R5 e a produtividade final do talhão, pode-se pensar em uma possível relação entre o índice de vegetação em fase específica do ciclo fenológico (R5) da soja com a produtividade final.

3.3 Índices utilizados na predição da produtividade

As características espectrais de SR podem ser expressas de várias maneiras. Uma das formas mais simples, quando se trata de SR na faixa óptica do espectro eletromagnético, é a reflectância em uma única banda (como vermelho ou infravermelho). Outra forma muito comum é a combinação das informações de reflectância de cada banda individual de uma maneira específica, o que resulta em um índice de vegetação. (Wiegand *et al.*, 1991 *apud* Epiphany *et al.*, 1996).

O índice mais popular entre os profissionais do campo é o NDVI demonstrado pela Equação 2 abaixo. Ele é um índice amplamente utilizado para avaliar a saúde e a densidade da vegetação em uma determinada área. Os comprimentos de onda do infravermelho próximo e do vermelho são sensíveis à clorofila, um pigmento nas plantas que desempenha um papel importante na fotossíntese. A ideia por trás do NDVI é que as plantas saudáveis e densas absorvem a luz vermelha para a fotossíntese e refletem mais a luz infravermelha. Plantas

estressadas ou áreas sem vegetação tendem a ter valores de NDVI mais baixos (Planet Scope, 2023).

Equação 2 – Equação do Índice de NDVI

$$NDVI = \frac{(NIR-Red)}{(NIR+Red)} \quad (2)$$

Onde:

- NDVI = Índice de Vegetação por Diferença Normalizada;
- NIR = banda do infravermelho próximo;
- RED = banda do vermelho.

Fonte: Adaptado de Planet Scope (2023).

Para Jensen (1949) o NDVI é um índice importante para monitorar mudanças sazonais e interanuais na vegetação, pois reduz vários tipos de ruídos multiplicativos nas imagens de SR de múltiplas datas. No entanto, apresenta desvantagens, incluindo a sensibilidade a efeitos ruidosos aditivos, como radiação de trajetória atmosférica. Além disso, o NDVI pode perder sensibilidade em condições de alto IAF, causando saturação. Ele também é sensível a variações no substrato sob o dossel da vegetação, especialmente em áreas com substratos mais escuros (Jensen, 1949).

Devido a isso, foram gerados outros IV's com intuito de suprir as deficiências presentes no NDVI e de demonstrar outras características da vegetação como biomassa, estresse hídrico, cobertura do solo, maturidade da cultura, dentre diversas outras características. Como exemplo temos o índice SAVI (*Soil Adjusted Vegetation Index* – Índice de Vegetação Ajustado ao Solo), que melhora os resultados do índice NDVI.

O SAVI é o índice definido pela Equação 3 apresentada a seguir, que leva em consideração os efeitos do solo exposto nas imagens analisadas, para ajuste do NDVI quando a superfície não está completamente coberta pela vegetação. No qual: L é uma constante denominada de fator de ajuste do índice SAVI, podendo assumir valores de 0,25 a 1, dependendo da cobertura do solo (Gameiro *et al.*, 2016).

Equação 3 – Equação do IV SAVI

$$SAVI = \frac{(1+L)*(NIR-Red)}{(NIR+Red+L)} \quad \text{onde } L = 0,5 \quad (3)$$

Onde:

- SAVI = Índice de Vegetação Ajustado ao Solo;
- L = fator de ajuste;
- NIR = banda do infravermelho próximo;
- RED = banda do vermelho.

Fonte: Adaptado de Gameiro *et al.*, 2016

Observa-se que o índice SAVI é um derivado do índice NDVI, ambos se diferem apenas pelo acréscimo de uma constante, que através dela se faz possível a análise de outros aspectos no cálculo, dessa forma melhorando os resultados. Isso demonstra como é possível com pequenas alterações realizar melhoramentos nos resultados dos índices já existentes.

O Index database (Index Database, 2011) é um banco de dados para índices de SR. Ele possui mais de 500 índices armazenados em sua interface e conta com nome, abreviatura, aplicações, sensores e as referências, o que otimiza a busca do usuário já que a mesma pode ser feita não apenas pelo nome do IV.

Quando realizada uma pesquisa utilizando “Mostrar índices para o aplicativo selecionado” (do inglês - *Show Indices for selected Application*) e selecionando a opção “Agricultura” (do inglês - *Agriculture*), são apresentados os índices para aplicação na agricultura, demonstrando a organização e otimização das informações no banco de dados, como demonstrado na Tabela 3.

Tabela 3 - Índices para aplicação na agricultura.

Nr.	Name	Formula	Variables	Comment
1	Atmospherically Resistant Vegetation Index	$\frac{NIR - RED - g(RED - BLUE)}{NIR + RED - g(RED - BLUE)}$	NIR = [781:1399]	
2	Atmospherically Resistant Vegetation Index 2	$-0.18 + 1.17 \left(\frac{NIR - RED}{NIR + RED} \right)$		
3	Canopy Chlorophyll Content Index	$\frac{NIR - rededge}{NIR + rededge}$ $\frac{NIR - Red}{NIR + Red}$		
4	CASI NDVI	$\frac{((770:780) + (784:790)) - ((655:665) + (676:685))}{((770:780) + (784:790)) + ((655:665) + (676:685))}$		
5	CASI TM4/3	$\frac{((770:780) + (784:790))}{((655:665) + (676:685))}$		
6	Cellulose Absorption Index	$100 (0.5 (2030nm + 2210nm) - 2100nm)$		
7	Cellulose absorption index 2	$0,5 (2020nm + 2220nm) - 2100nm$		
8	Chlorophyll Absorption Ratio Index	$\left(\frac{700nm}{670nm} \right) \frac{\sqrt{(a \cdot 670 + 670nm + b)^2}}{(a^2 + 1)^{0.5}}$	b=(550nm-((700nm-550nm)/150*550)), a=(700nm-550nm)/150	
9	Chlorophyll Absorption Ratio Index 2	$\left(\frac{[(a \cdot 670) + (670) + b]}{(a^2 + 1)^{0.5}} \right) \left(\frac{700}{670} \right)$	a=((700)-[550])/150, b=[550]-(a*[550])	
10	Chlorophyll Green	$\left(\frac{700-800}{540-560} \right)^{(-1)}$		

Fonte: Index Database (2011).

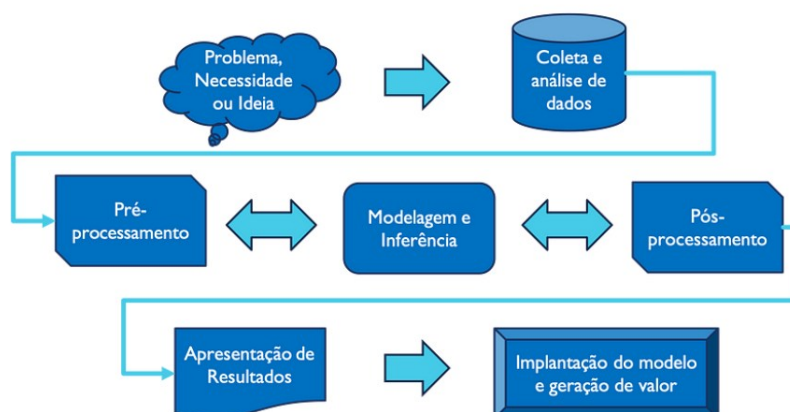
3.4 Aprendizado de máquinas na predição

O Aprendizado de Máquina é uma técnica de ciência de dados que permite que os computadores usem os dados existentes para prever futuros comportamentos, resultados e tendências. Usando o aprendizado de máquina, os computadores aprendem sem serem explicitamente programados (Microsoft, 2023). As ferramentas de aprendizado de máquina usam sistemas de Inteligência Artificial (IA) que podem identificar padrões e criar associações a partir da experiência com os dados (Microsoft, 2023). Basicamente, é uma representação matemática de dados e suas relações. Os dados podem ser dos mais variados (Parsons, 2021).

O aprendizado de máquina pode ser dividido em dois tipos: supervisionado e não-supervisionado. O aprendizado supervisionado, consiste no modelo (ou algoritmo) ser construído a partir dos dados de entrada, que são apresentados na forma de pares ordenados (entrada — saída desejada) (Escovedo, 2020). Já o aprendizado de máquina não-supervisionado não existe a informação dos rótulos históricos, ou seja, não temos as saídas desejadas a serem estimadas e, por este motivo, dizemos que nossos dados são não-rotulados. Assim, o algoritmo não recebe durante o treinamento os resultados esperados, devendo descobrir por si só, por meio da exploração dos dados, os possíveis relacionamentos entre eles (Escovedo, 2020).

Normalmente, se divide os dados de entrada em dois grupos: o conjunto de treinamento, que servirá para construir o modelo, e o conjunto de validação (ou teste), que servirá para verificar como o modelo se comporta com dados não vistos, de forma que nos permita ajustar se necessário, para a construção final do modelo a ser aplicado em novos dados que ainda não conhecemos a saída esperada (Escovedo, 2020). A seguir, na Figura 3 é apresentado um esquema de um projeto completo de Aprendizado de Máquina.

Figura 3 - Esquema de um projeto completo de Aprendizado de Máquina.



Fonte: Escovedo, Koshiyama (2020).

Os algoritmos podem ser classificados como métodos de regressão paramétricos, sendo os modelos de *Linear Regression* (LR) e os métodos de regressão não paramétricos sendo eles *Random Tree* (RT), *Random Forest* (RF), *Neural Net* (NN), *Support Vector Machine* (SVM) (Abreu Júnior, 2022).

Florestas aleatórias (do inglês - *Random Forest*) trata-se de uma combinação de preditores de árvores tais que cada árvore depende dos valores de um vetor aleatório amostrado de forma independente e com a mesma distribuição para todas as árvores da floresta (Breiman, 2001).

RT e RF possuem as mesmas características funcionais, porém a RF constrói múltiplas árvores de decisão sobre subconjuntos de dados inicializados e obtém o resultado através de votação majoritária para classificação ou média para regressão, enquanto a RT considera apenas uma árvore de decisão (Budu, 2023).

De forma geral, uma rede neural (do inglês – *Neural Net*) é uma máquina projetada para modelar a maneira como o cérebro humano realiza uma tarefa particular ou função de interesse; ela é normalmente implementada utilizando componentes eletrônicos ou é simulada por programação em um computador digital (Haykin, 2001).

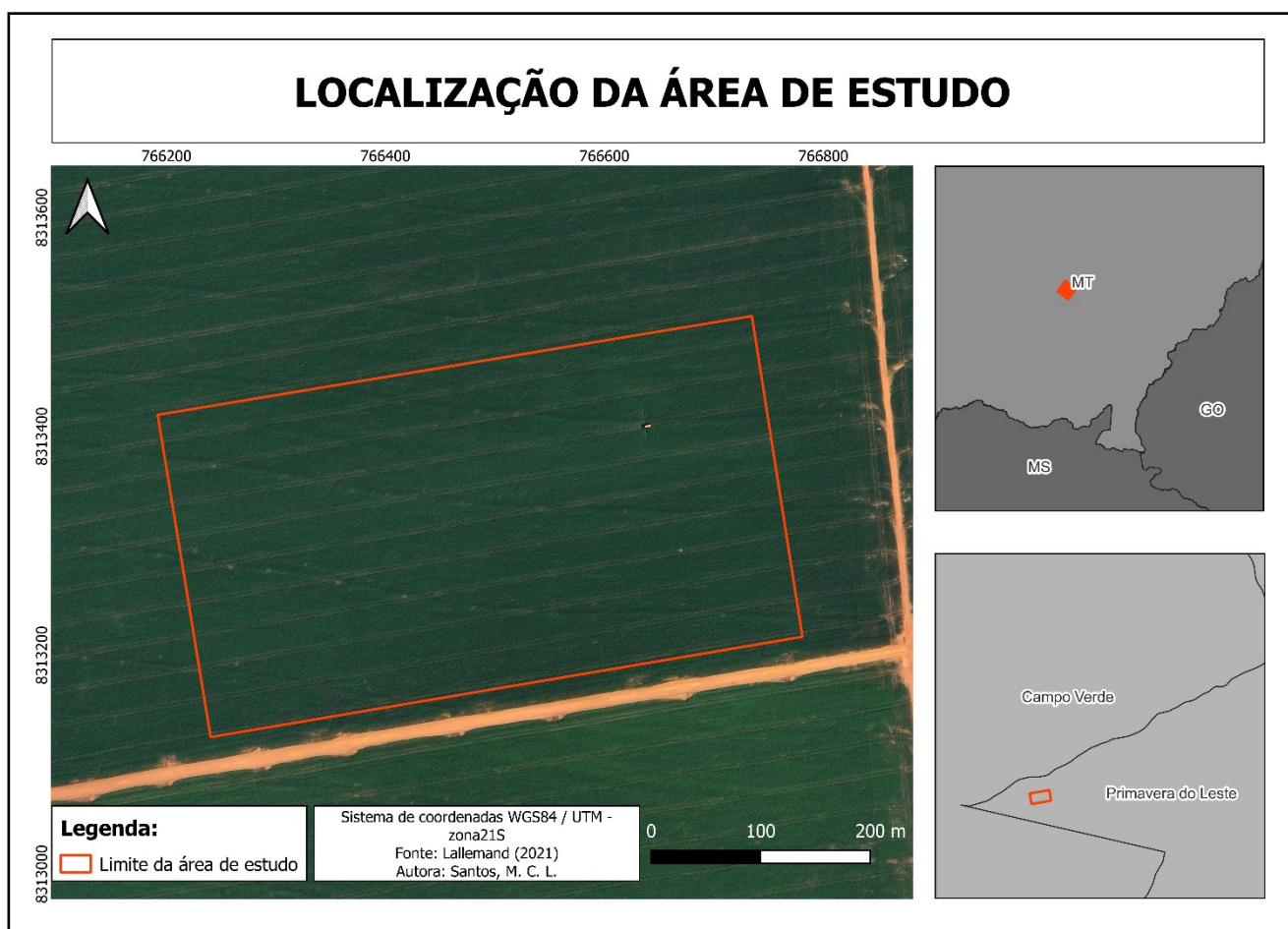
A Máquina de Vetores (do inglês – *Support Vector Machine*) é um algoritmo em que é plotado cada item de dado como um ponto no espaço n-dimensional (onde n é o número de recursos que se tem), com o valor de cada recurso sendo o valor de uma determinada coordenada. Então, é executada a classificação encontrando o hiperplano que melhor diferencia as duas classes. Os “Vetores de suporte” são simplesmente as coordenadas da observação individual. SVM é uma fronteira que melhor segrega as duas classes (hiperplano / linha) (Addan, 2019).

4 MATERIAL E MÉTODOS

4.1 Área de estudo

A área de estudo, contendo os pontos de produtividade, está localizada na cidade de Primavera do Leste, no estado do Mato Grosso. O município está inserido no bioma Cerrado, com clima predominante tropical sazonal (IBGE, 2019), temperaturas médias anuais em torno de 22-23°C e precipitação média anual entre 1200 e 1800 mm (USP, 2000). A lavoura em estudo, colhida em 02 de março de 2021, tem uma área aproximada de 16,386 ha e 9499 pontos geolocalizados com dados da colheita. Com uma produtividade média de 3,76 ton/ha, obtendo 2,24 ton/ha como mínima e uma máxima de 5,18 ton/ha. Abaixo a Figura 4 exibe um mapa demarcando o limite da área e a localização da região em estudo no presente trabalho.

Figura 4 - Mapa de localização da área de estudo.



Fonte: a Autora (2025).

4.2 Material

Para realização do trabalho foram utilizadas 9 imagens do satélite Planet em diferentes etapas do desenvolvimento da lavoura de soja, do plantio à colheita. A intenção foi estimar a produtividade em diversas datas, acompanhando os estádios fenológicos, para que fosse possível, ao final, comparar os resultados com os dados da colheita e determinar qual a melhor data de aquisição, aquela que mais se aproxima dos valores reais. Por isso a aquisição das imagens se fez tão importante para o trabalho, quanto maior a quantidade de imagens maior será a resolução temporal e o número de informações, possibilitando assim relacionar o estágio fenológico com as imagens.

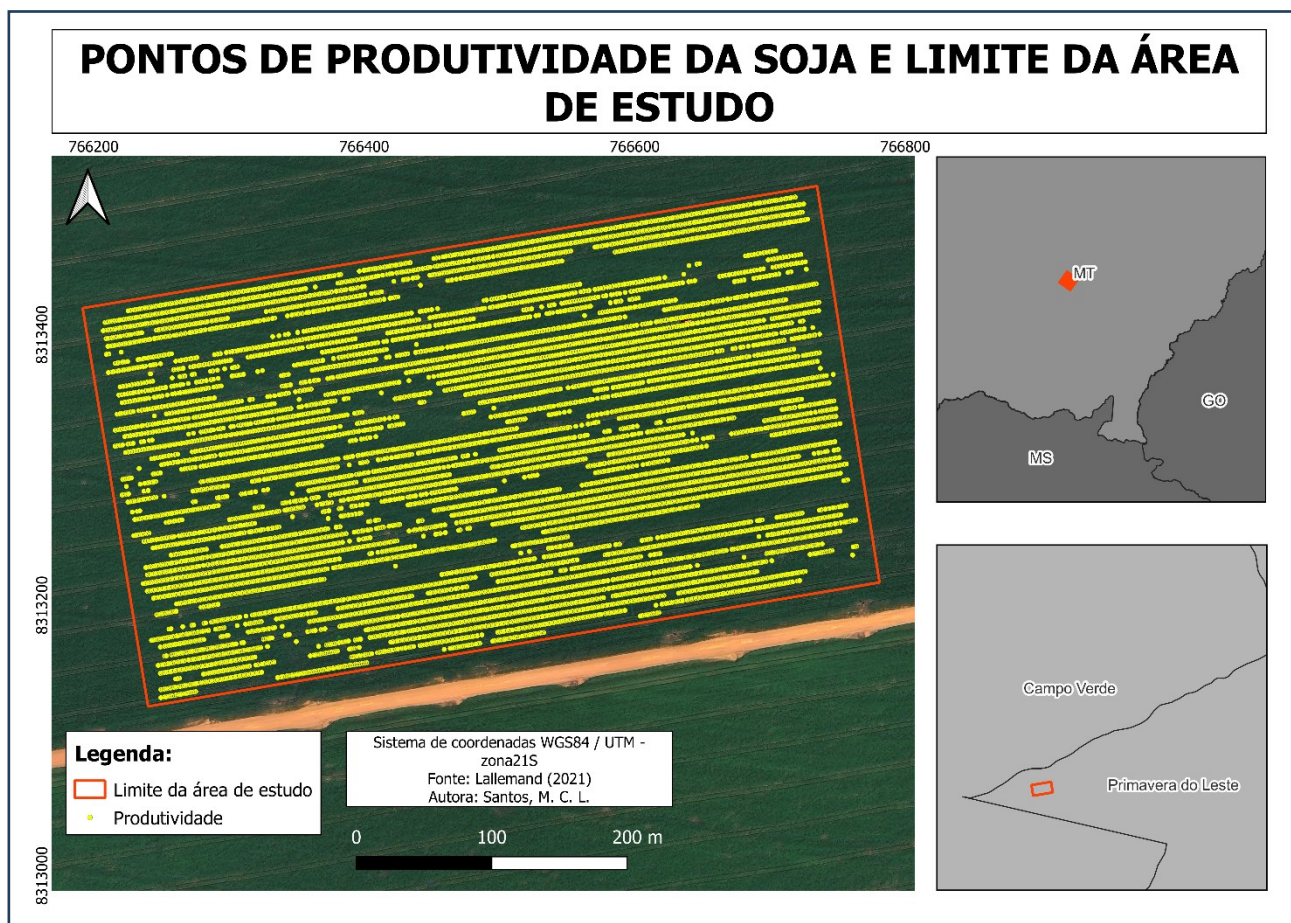
As imagens têm resolução espacial de 3.0m a 3.9m e inclui 4 bandas espectrais com variação de 455 a 860 nanômetros (Planet Labs, 2024). As imagens Planet se demonstraram adequadas para a finalidade do trabalho devido ao número de bandas, que tornou possível a realização do cálculo de diversos índices, monitoramento de ampla área de regiões agrícolas em todo o mundo, alta resolução espacial, com frequência temporal diária ou semanal (Planet Labs, 2024). Isso possibilitou a avaliação ao longo do desenvolvimento da lavoura de soja.

Foram necessários também os dados de produtividade georreferenciados, obtidos por meio de uma colhedora com um sistema georreferenciador, e o limite da lavoura disponibilizados pela Lallemand Plant Care em arquivos no formato *shapefile* (.shp) e *Keyhole Markup Language* (.kml).

Na realização do processamento e espacialização dos dados vetoriais, foi empregado um ambiente de Sistema de Informação Geográfica (SIG), o software QGIS na versão Desktop 3.16.16. que possui código aberto e é um projeto oficial da *Open Source Geospatial Foundation* (OSGeo); uma ferramenta que possibilita a criação e a manipulação de dados vetoriais (ex.: pontos, linhas, polígonos) e matriciais (ex.: raster) (QGIS, 201-?).

Foi usado também o software WEKA (*Waikato Environment for Knowledge Analysis*) em sua versão 3.8.6, que se trata de uma coleção de algoritmos de aprendizado de máquina voltado para de mineração de dados. Ele tem ferramentas que permitem a preparação de dados, classificação, regressão, clustering, mineração de regras de associação e visualização. Ele também se trata de um software de código aberto e foi emitido pela GNU – *General Public License*. (Waikato, 201?). Por fim, foi utilizado o software Excel para o cálculo dos Índices de Vegetação. A seguir, na Figura 5, o mapa com o limite da área de estudo e a localização dos pontos de produtividade.

Figura 5 - Limite da área de estudo e pontos de produtividade.

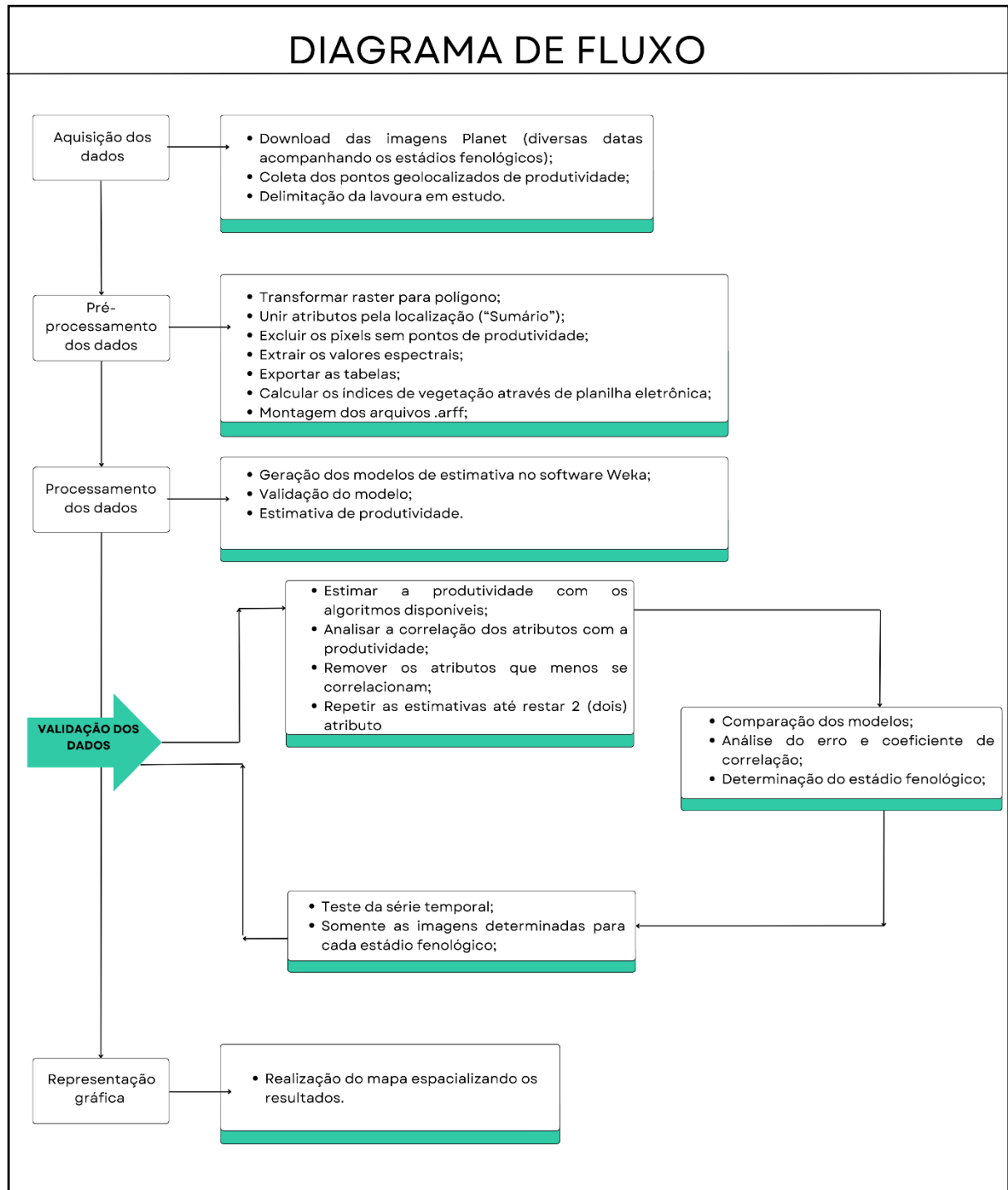


Fonte: a Autora (2025).

4.3 Métodos

A abordagem adotada para execução desse trabalho foi segmentada em 4 etapas: aquisição dos dados, pré-processamento dos dados, processamento dos dados e representação gráfica. Abaixo é exposto um diagrama de fluxo que ilustra o desdobramento da metodologia utilizada (Figura 6).

Figura 6 - Diagrama de fluxo.



Fonte: a Autora (2025).

4.3.1 Aquisição dos dados

A aquisição dos dados se deu início com a obtenção das imagens do satélite da Planet Labs pode apresentar desafios devido ao caráter comercial da empresa, que disponibiliza suas imagens de alta resolução mediante assinatura. No entanto, é possível acessar esses dados de forma gratuita por meio de programas específicos, dependendo da finalidade do estudo. Uma alternativa viável é o *Education and Research Program* (Programa de Educação e Pesquisa) ou *Disaster Response Program* ou (Programa de Resposta a Desastres), que concede acesso a imagens para determinadas categorias de usuários. Entre os beneficiários estão: pesquisadores acadêmicos vinculados a universidades, organizações sem fins lucrativos e ONGs, Instituições governamentais e iniciativas de resposta a desastres, que podem obter imagens para análises e monitoramento.

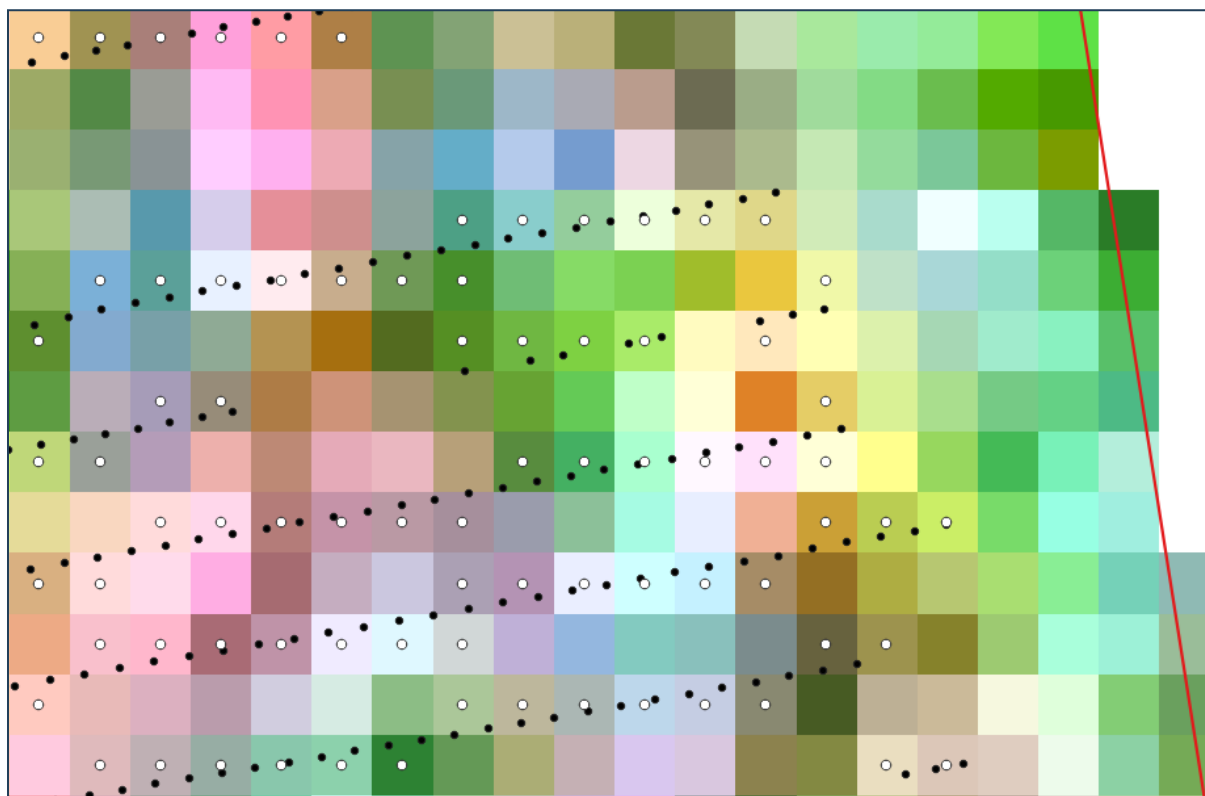
O processo de solicitação ocorre por meio do site oficial da Planet Labs, para isso, é necessário enviar uma solicitação detalhada, justificando o uso das imagens e o propósito do projeto (Planet Labs, 2024).

Com isso, foi utilizado o polígono discriminando o limite da área de estudo, para determinar quais imagens abrangem a região, e filtros de seleção, como data e de cobertura de nuvens. Foram baixadas ao menos uma imagem para cada estágio fenológico, porém ao utilizar os filtros, as imagens que não demonstraram potencial para o estudo devido a presença de nuvens ou defeitos em seu enquadramento foram descartadas. Assim, foram utilizadas 9 imagens abrangendo os estágios fenológicos: VE, VC, V1, V2, R1, R2, R3, R4, R5 e R6.

4.3.2 Pré-processamento dos dados

Primeiramente, foi necessário vincular o pixel da imagem com os pontos de produtividade. Dado que o pixel tem dimensão de 3m x 3m, há ocorrência de vários pontos de produtividade dentro de um mesmo pixel ou de pixels sem nenhum ponto de produtividade. Para resolver isso, no software QGIS, foi necessário separar os pixels que continham pontos de produtividade dos que não tinham, transformá-los de raster para polígono e através da ferramenta “Unir atributos pela localização (Sumário)” adicioná-los a média da produtividade como um atributo à tabela da camada vetorial gerada (Figura 7).

Figura 7 – Relação pixel e ponto de produtividade.



Fonte: a Autora (2025).

Em seguida, também no software QGIS, foi apropriado que ocorresse a geração dos centroides de cada pixel seguindo com a extração da reflectância, utilizando a ferramenta “Add raster values to points”. Então a tabela de atributo foi exportada em um formato aceito pelo Excel, para que pudesse ser feito o cálculo dos índices.

Foi necessário que houvesse a separação dos pontos para validação e treinamento do modelo, de acordo com a metodologia proposta nesse trabalho. Essa técnica é conhecida como “*Holdout*”, indicada para casos que se tenha grande quantidade de dados e em caso de pouco tempo para realização do trabalho e/ou para o início da criação de modelos (Allibhai, 2018; Shah, 2017). Os dados então foram separados no em um grupo com 80% dos pontos para a criação do modelo e outro com 20% dos pontos excedentes para a validação do modelo que será gerado, sendo a relação 20/80 muito usada pelos usuários da técnica (Allibhai, 2018). A geração desses arquivos será realizada no software QGIS utilizando a ferramenta de seleção aleatória dos pontos.

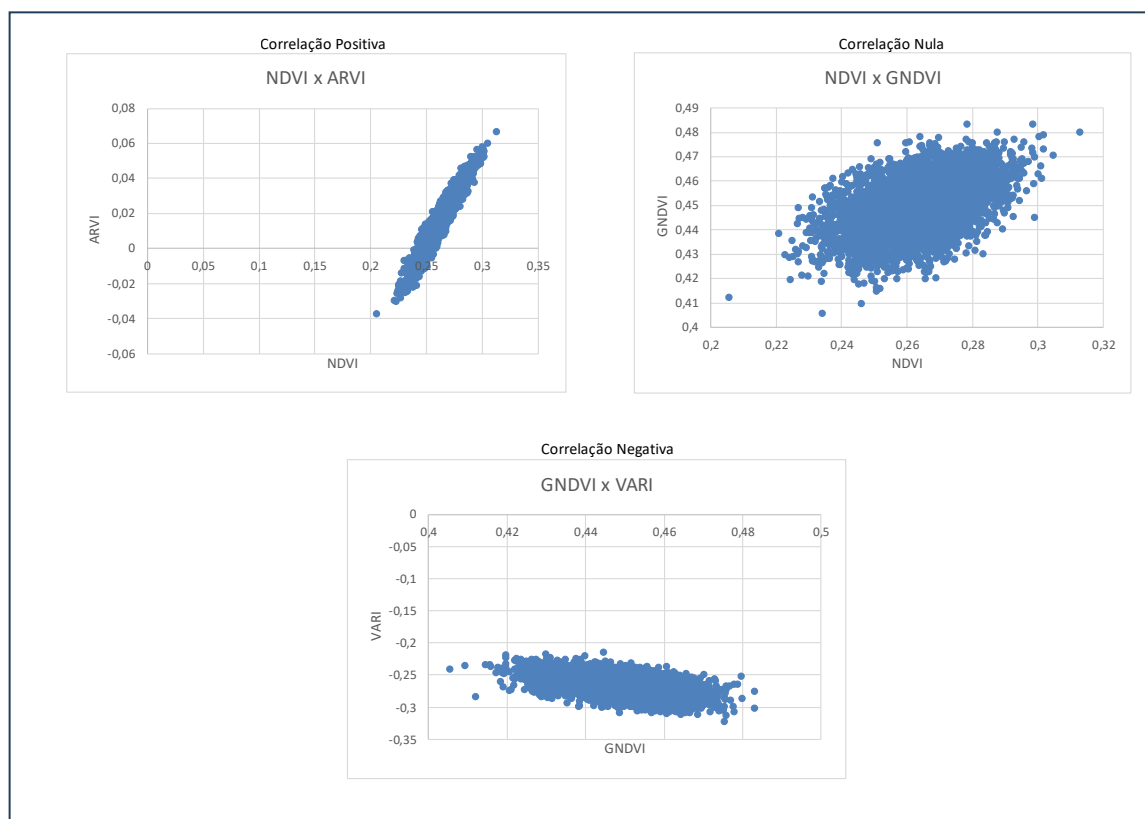
Para finalizar o pré-processamento realizou-se a preparação do arquivo de entrada do WEKA que permitiu o teste dos modelos. Neste arquivo deve conter as medidas de

produtividade, os valores espectrais de cada banda relacionados aos pontos e os IV's que foram definidos.

4.3.3 Processamento dos dados

Para o processamento dos dados, os IV's foram computados no Excel e em seguida, foram calculados os coeficientes de variação e foram plotados os gráficos de dispersão daqueles que evidenciam as mesmas características da planta (Figura 8). Os que demonstraram graficamente uma correlação positiva, evidenciam forte correlação entre os dados e em vista disso elegeram-se somente um dos IV's, usando como critério de escolha aquele que apresentou o maior coeficiente de variação, a fim de evitar a redundância e garantir a variabilidade dos dados. Em contrapartida, para os IV's que apresentaram uma correlação nula, que não existe correlação, ou correlação negativa entre os IV's foi admitido o uso de ambos. Isso revela que mesmo ressaltando a mesma característica da planta eles estão agregando informações distintas ao modelo.

Figura 8 - Gráficos de correlação dos IV's.



Fonte: a Autora (2025).

Inicialmente, foram calculados 21 índices, dos quais, com base na metodologia proposta, foram selecionados os 14 mais relevantes, além do NDVI, cuja importância é amplamente reconhecida em diversas pesquisas. Na Tabela 4, estão dispostos os 15 IV's utilizados.

Tabela 4 – Índices usados na criação dos modelos preditivos.

Índices	Equação	Referência
OSAVI – Optimized Soil Adjusted Vegetation Index	$\frac{(NIR - R)}{(NIR + R + 0,16)}$	Sergieieva (2024).
ATSAVI - Adjusted transformed soil-adjusted VI	$a \frac{NIR - a * R - b}{a * NIR + R - a * b + X(1 + a^2)}$	Baret, Guyot (1991).
GOSAVI - Green Optimized Soil Adjusted Vegetation Index	$\frac{NIR - G}{NIR + G + Y}$	Rondeaux, Steven, Baret, (1996).
ARVI – Atmospherically Resistant Vegetation Index	$\frac{(NIR - (2 * R) + B)}{(NIR + (2 * R) + B)}$	Kaufman, Tanré (1992).
GNDVI – Green Normalized Difference Vegetation Index	$\frac{(NIR - G)}{(NIR + G)}$	Gitelson et al., (1996).
MTVI2 - Modified Triangular Vegetation Index 2	$\frac{1,5[2,5(NIR - B) - 2,5(R - G)]}{\sqrt{(2NIR + 1)^2 - 6NIR - 5\sqrt{(NIR - 0,5)}}$	Haboudane et al. (2003).
GLI - Green leaf index	$\frac{2 * G - R - B}{2 * G + R + B}$	Hunt Jr. et al., (2013).
VARI – Visible Atmospheric Resistant Index	$\frac{(G - R)}{(G + R - B)}$	Sergieieva (2024).
NDWI – Normalized Difference 860/1240 Normalized Difference Water Index	$\frac{(G - NIR)}{(G + NIR)}$	Gao (1996).
ReCI – Red Edge Chlorophyll Index	$\left(\frac{NIR}{R}\right) - 1$	Sergieieva (2024).
GCI – Green Chlorophyll Index	$\frac{NIR}{G} - 1$	Ataíde (2024).
SIPI – Structure Insensitive Pigment Index	$\frac{NIR - B}{NIR - R}$	Penuelas et al. (1995).

CVI – Chlorophyll Vegetation Index	$\text{NIR} * \frac{R}{G^2}$	Vincini; Frazzi; D'Alessio (2008).
TGI – Triangular Greenness Index	$-0,5*((665-490)*(R-G)-(665-560)*(R+B))$	McKinnon & Hoff, (2017).

Fonte: a Autora (2025).

Onde:

Tabela 5 – Bandas do Satélite Planet.

Bandas Planet	Comprimento de onda (µm)
Blue - B	455–515 nm
Green - G	500–590 nm
Red - R	590–670 nm
Near Infrared - NIR	780–860 nm

Fonte: a Autora (2025).

Em seguida, foram gerados os arquivos de entrada padrão WEKA, do tipo .arff (*Attribute – Relation File Format*) composto pelos dados já separados e os IV's calculados. O arquivo continha um cabeçalho, registrando os tipos de atributos e seus valores, e uma seção com os dados separados por vírgula.

O software WEKA dispõem de algoritmos de aprendizado de máquina para uso na estimativa de uma variável de interesse, no presente trabalho, a produtividade. A RL, NN, RF, RT e o SVM são os algoritmos implementados que foram utilizados no seu formato padrão para geração do modelo. Após o processamento, como resultado são disponibilizados a estimativa, o erro, o coeficiente de correlação (R), o *Root Mean Squared Error* (RMSE) e outros parâmetros relevantes. Ademais, na ferramenta “*Select Attributes*” é possível ranquear os atributos de acordo com a correlação de cada um com a produtividade, da maior para menor, possibilitando a exclusão dos mesmos, evitando assim uma possível confusão no algoritmo que influencie negativamente a estimativa.

Então, na página inicial do WEKA foi adicionado o arquivo contendo 80% dos dados e na aba “Classify” foi selecionado a opção “Supplied test set” e adicionado então o arquivo com os 20% dos dados restantes. Em seguida, foram testados os algoritmos de aprendizado de máquina e anotado em uma planilha os valores de RMSE e R de cada estimativa. As fórmulas do R e RMSE são calculados segundo as equações 4 e 5 respectivamente.

Equação 4 – Equação do Índice de R

$$R = \frac{\sum_i^n (xi - \bar{x})(yi - \bar{y})}{\sqrt{\sum_i^n (xi - \bar{x})^2 \sum_i^n (yi - \bar{y})^2}} \quad (4)$$

Equação 5 – Equação do RMSE.

$$RMSE = \sqrt{\frac{1}{n} \sum_i^n (yi - \bar{y})^2} \quad (5)$$

Onde:

- xi e yi = valores medidos;
- $\bar{x}$ e $\bar{y}$ = valores estimados.

Fonte: Adaptado de IDB (2025).

Realizou-se então a remoção gradual das variáveis, uma a uma, com base na classificação de R dos atributos em relação à produtividade, realizada na aba “Select Attributes” através da ferramenta “Attribute Evaluator”, “CorrelationAttributeEval”. Para cada exclusão, novas estimativas foram geradas utilizando todos os algoritmos previamente testados. O processo de remoção foi interrompido quando o conjunto de variáveis se reduziu a apenas duas, possibilitando a identificação das bandas e IVs mais relevantes para a acurácia da estimativa.

4.3.4 Validação e melhoramento do modelo

A validação dos modelos foi realizada através da análise do R e do RMSE, onde o modelo que apresentasse o menor erro e maior coeficiente de correlação seria definido como o que obteve a melhor performance para os dados e melhor acurácia. Dessa forma, o estádio fenológico que melhor estima a produtividade da soja foi determinado com base no modelo escolhido com os parâmetros citados anteriormente.

Com o estádio fenológico determinado, foi iniciado o processo de melhoramento dos modelos gerados, que ocorreu através do teste de uma série histórica, para averiguar se o uso dela resulta em melhora do modelo. Essa análise foi realizada unicamente com o algoritmo que obteve o melhor desempenho na fase de determinação do estádio fenológico.

Diferentemente da abordagem anterior, que se limitava a uma data específica do desenvolvimento, neste teste, o desempenho da reflectância espectral e dos IV's foi analisado em toda a extensão do período de plantio. Essa análise abrangente permitiu a construção de um modelo mais completo e representativo.

Com as análises realizadas, foi necessário exportar os resultados encontrados para que seja possível realizar a representação gráfica.

4.4 Representação Gráfica

Com a conclusão das análises, foi criada uma camada vetorial contendo todos os pixels, incluindo os utilizados para treinamento, validação e aqueles com produtividade nula, previamente removidos para a geração dos modelos. Foi extraído a reflectância de todos eles, calculados os IV'S no Excel e então foi gerado então um arquivo Weka (.arff) com os IV's calculados e um "?" ao final, onde ficara a produtividade anteriormente.

Já no Weka, utilizando o modelo com melhor desempenho, o arquivo de 80% submetido na aba principal foi o utilizado na melhor modelagem e na aba "Classify", "Supplied test set" foi submetido o arquivo com todos os pontos. Esse processo permitiu a espacialização da produtividade em toda a lavoura.

Por fim, o resultado obtido no processamento, foi formatado em uma tabela (.csv) e foi aberto no software QGIS, onde foram plotados e um layout de impressão foi elaborado para a geração do mapa de produtividade da soja utilizando a simbologia "Graduado", no modo "Intervalos iguais".

5 RESULTADOS E DISCUSSÕES

5.1 Modelo com melhor desempenho

A avaliação dos modelos multiespectrais para a estimativa da produtividade da cultura da soja revelou que o Random Forest obteve o melhor desempenho em todas as imagens processadas. Entre os diferentes períodos analisados, a imagem capturada em 08 de janeiro apresentou os melhores coeficientes de correlação (R), chegando a 0,8968 e erros quadráticos médios (RMSE) apresentando 0,2186 (Figura 9).

Figura 9 – Resultados dos processamentos.

Data estimada: 08/01/2021												Ranked attributes:	
Algoritmo	R	RMSE	Atributo Removido	Algoritmo	R	RMSE	Atributo Removido	Algoritmo	R	RMSE	Atributo Removido		
LR	0,3568	0,4438	NENHUM	LR	0,3601	0,4429	GOSAVI	LR	0,2685	0,4564	VARI	0,2914	14 NDWI
MP	0,4859	0,6123	NENHUM	MP	0,5617	0,5795	GOSAVI	MP	0,2614	0,7141	VARI	0,2612	17 SIPI
SMO	0,4296	0,4292	NENHUM	SMO	0,3891	0,4386	GOSAVI	SMO	0,2706	0,4666	VARI	0,0897	11 MTVI2
RF	0,8968	0,2186	NENHUM	RF	0,888	0,2339	GOSAVI	RF	0,8612	0,2562	VARI	0,0462	18 CVI
RT	0,8725	0,2399	NENHUM	RT	0,8398	0,2703	GOSAVI	RT	0,8151	0,2895	VARI	0,0121	12 GU
LR	0,3568	0,4438	NIR	LR	0,3601	0,4429	GNDVI	LR	0,2682	0,4564	GLI	-0,1742	13 VARI
MP	0,5618	0,6017	NIR	MP	0,5608	0,5792	GNDVI	MP	0,2493	0,7234	GLI	-0,2309	1 B
SMO	0,3934	0,4386	NIR	SMO	0,3894	0,4386	GNDVI	SMO	0,2702	0,4667	GLI	-0,2427	8 NDVI
RF	0,8954	0,2201	NIR	RF	0,8875	0,234	GNDVI	RF	0,862	0,2561	GLI	-0,2428	5 OSAVI
RT	0,8539	0,2557	NIR	RT	0,8448	0,2657	GNDVI	RT	0,814	0,2893	GLI	-0,2443	15 RECI
LR	0,3568	0,4438	G	LR	0,3601	0,4429	ARVI	LR	0,2682	0,4564	CVI	-0,2472	9 ARVI
MP	0,5635	0,6016	G	MP	0,4209	0,6625	ARVI	MP	0,2446	0,7316	CVI	-0,2314	10 GNDVI
SMO	0,3938	0,4385	G	SMO	0,3893	0,4386	ARVI	MP	0,2704	0,4666	CVI	-0,2914	7 GOSAVI
RF	0,8934	0,222	G	RF	0,8894	0,232	ARVI	RF	0,8627	0,2561	CVI	-0,2946	16 GCI
RT	0,8632	0,2504	G	RT	0,8313	0,2775	ARVI	RT	0,8161	0,2892	CVI	-0,3071	19 TGI
LR	0,3568	0,4438	ATSAVI	LR	0,3601	0,4429	RECI	LR	0,2505	0,4587	MTVI2	-0,323	3 R
MP	0,5549	0,6033	ATSAVI	MP	0,3449	0,6802	RECI	MP	0,2187	0,7184	MTVI2	-0,3474	6 ATSAVI
SMO	0,3936	0,4386	ATSAVI	SMO	0,3709	0,4427	RECI	SMO	0,2517	0,4673	MTVI2	-0,3573	2 G
RF	0,8981	0,2191	ATSAVI	RF	0,8892	0,2311	RECI	RF	0,8437	0,2669	MTVI2	-0,4142	4 NIR
RT	0,8512	0,2603	ATSAVI	RT	0,8571	0,2542	RECI	RT	0,8318	0,2752	MTVI2		
LR	0,3574	0,4435	R	LR	0,3601	0,4429	OSAVI	LR	0,2508	0,4585	SIP1		
MP	0,4473	0,641	R	MP	0,344	0,6822	OSAVI	MP	0,2198	0,7205	SIP1		
SMO	0,3936	0,4386	R	SMO	0,3705	0,4427	OSAVI	SMO	0,2508	0,4679	SIP1		
RF	0,8992	0,2214	R	RF	0,8899	0,2293	OSAVI	RF	0,7504	0,3128	SIP1		
RT	0,8655	0,2463	R	RT	0,8539	0,2575	OSAVI	RT	0,7924	0,3008	SIP1		
LR	0,3601	0,4429	TGI	LR	0,3601	0,4429	NDVI						
MP	0,4367	0,6385	TGI	MP	0,3417	0,6804	NDVI						
SMO	0,3974	0,4368	TGI	SMO	0,3698	0,4428	NDVI						
RF	0,8846	0,2372	TGI	RF	0,8905	0,2309	NDVI						
RT	0,8474	0,2638	TGI	RT	0,8584	0,254	NDVI						
LR	0,3601	0,4429	GCI	LR	0,2681	0,4565	B						
MP	0,5527	0,5906	GCI	MP	0,2639	0,7179	B						
SMO	0,389	0,4387	GCI	SMO	0,2698	0,4668	B						
RF	0,8869	0,2345	GCI	RF	0,8614	0,2567	B						
RT	0,8619	0,2492	GCI	RT	0,8259	0,2821	B						

Fonte: a Autora (2025).

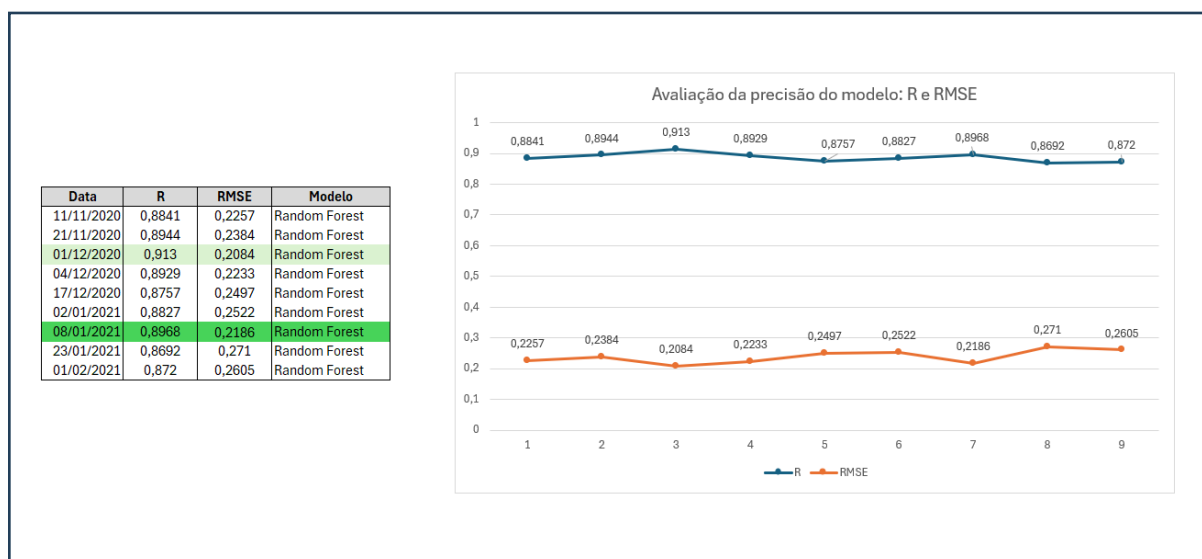
O alto valor de R indica uma forte correlação entre os valores preditos pelo modelo e os valores observados em campo, demonstrando a robustez do Random Forest na modelagem da produtividade da soja. Além disso, o baixo RMSE evidencia a precisão das estimativas, reduzindo a margem de erro e tornando o modelo confiável para aplicações práticas.

A eficiência do RF está alinhada à literatura, pois Breiman (2001) destaca que esse algoritmo consegue minimizar o erro de previsão através de um conjunto de árvores de decisão, aumentando sua robustez. Além disso, Fernandes, Veiverberg e Sebem (2009) demonstraram que IVs derivados de sensores orbitais, como o GNDVI, podem ser correlacionados com a produtividade de soja em diferentes fases fenológicas, reforçando a adequação dos modelos baseados em aprendizado de máquina.

5.2 Estádio fenológico mais adequado para estimativa da soja

Ao analisar os resultados dos melhores modelos, evidenciados na Figura 10, ficou explícito que o período de 08 de janeiro foi o momento mais adequado para a aquisição de imagens para a estimativa da produtividade da soja. Nessa data, a lavoura encontrava-se, possivelmente, no estágio R4 (vagem completamente desenvolvida), fase que ocorre entre 65 e 75 dias após o plantio e que é considerada crítica para o desenvolvimento e definição do rendimento da cultura.

Figura 10 – Avaliação da precisão do modelo: R e RMSE.



Fonte: a Autora (2025).

Nesta fase, fisiologicamente, a cultura já atingiu o pico da área foliar, otimizando a interceptação da radiação solar, um fator essencial para a fotossíntese. O conteúdo hídrico elevado também favorece a distinção espectral entre áreas com diferentes níveis de produtividade, tornando as imagens capturadas mais representativas da realidade da lavoura. (Neumaier *et al*, 2000). Além disso, há um acúmulo significativo de biomassa, refletindo diretamente no rendimento final, uma vez que o crescimento e a translocação de fotoassimilados para as vagens estão em plena atividade (Bueno, 2022).

Em relação aos fatores bióticos, observa-se uma menor incidência de doenças foliares severas, uma vez que a planta ainda não atingiu a maturação podendo comprometer a saúde da lavoura e alterar a resposta espectral (Godoy, 2017). Além disso, há um menor impacto de pragas tardias, como percevejos, que afetam mais intensamente as fases finais do enchimento de grãos, o que poderia interferir nas estimativas caso as imagens fossem adquiridas tardiamente

(Oliveira, 2024). Além disso, há uma menor interferência do efeito de senescência, uma vez que em estádios mais avançados, como R5 em diante, as folhas começam a amarelar (Oliveira, 2024), alterando a refletância do dossel vegetal e tornando as estimativas menos precisas.

Entre os fatores abióticos, destaca-se a estabilidade climática observada em janeiro, quando, na maioria das regiões produtoras de soja, há umidade suficiente no solo e temperaturas adequadas para o desenvolvimento da cultura (Lulu; Zolin, 2021). Esse cenário favorece a aquisição de imagens sem grandes interferências de estresses hídricos severos ou temperaturas extremas, que poderiam comprometer a precisão dos modelos.

Dessa forma, a escolha do período de 08 de janeiro para a aquisição das imagens se mostrou estrategicamente favorável, garantindo maior correlação entre os dados espectrais e a produtividade final da soja. Essa informação é essencial para o aprimoramento de modelos de estimativa agrícola, permitindo um monitoramento mais eficiente da cultura e possibilitando tomadas de decisão mais assertivas para o manejo da lavoura.

Ressalta-se que as estimativas realizadas próximo ao plantio, não possuem a mesma precisão que as estimativas feitas em outros momentos do ciclo devido ao fato da planta estar no início da fase vegetativa, onde ainda não possui muitas folhas e nós (Farias; Nepomuceno; Neumaier, 2007). Nesses casos o solo se encontra mais exposto, o que acaba influenciando nas respostas espectrais das imagens e tende a saturar alguns índices, como NDVI (Jensen, 1949).

5.3 Série Temporal

A análise da série temporal foi realizada para verificar se a incorporação de imagens de diferentes datas poderia aprimorar a estimativa da produtividade, permitindo capturar dinâmicas fenológicas da cultura e minimizar variações ocasionais. No entanto, conforme indicado na Tabela 6, os resultados mostraram que não houve uma melhora significativa nos coeficientes de correlação e no RMSE.

Tabela 6 – Resultado da série temporal.

R	RMSE
0,8819	0,2412

Fonte: a Autora (2025).

Isso pode ser explicado por diversos fatores. Primeiramente, a alta correlação da produtividade com o estágio R4, fase fenológica em que a biomassa e o teor de água da planta

atingem níveis ideais para diferenciação espectral. Como essa fase já apresenta forte correlação com a produtividade, a inclusão de imagens de outros períodos pode ter inserido apenas informações redundantes no modelo.

Além disso, o efeito de saturação espectral (Jensen, 1949) nos estágios mais avançados do ciclo da soja, especialmente a partir do enchimento de grãos (R5 e R6), reduz a sensibilidade do modelo a variações sutis na produtividade, limitando o impacto positivo da adição de imagens posteriores à data mais relevante.

Outro fator relevante é a eficiência do algoritmo Random Forest, que já demonstrou alto desempenho na modelagem com uma única data, sugerindo que a informação disponível na imagem específica de 08 de janeiro foi suficiente para capturar os padrões de produtividade da cultura. Como esse método trabalha com múltiplas árvores de decisão (Budu, 2023), a introdução de mais variáveis temporais pode não ter agregado novas informações relevantes para o modelo.

Dessa forma, os resultados reforçam a importância da seleção criteriosa da data de aquisição das imagens para a estimativa da produtividade da soja. Em vez de utilizar múltiplas imagens ao longo do ciclo da cultura, a escolha do momento mais adequado para a obtenção dos dados espectrais se mostrou mais eficiente para aprimorar a modelagem preditiva. Isso tem implicações diretas para futuros estudos e aplicações práticas, sugerindo que a otimização da data de sensoriamento pode ser uma estratégia mais eficaz do que o simples aumento da quantidade de imagens na série temporal.

5.4 Componentes do melhor modelo

A seleção criteriosa das variáveis espectrais utilizadas na modelagem foi essencial para alcançar a melhor estimativa de produtividade da soja. O modelo com melhor desempenho foi obtido utilizando o conjunto total de bandas espectrais e IV's calculados neste trabalho, apresentados na tabela 7. Tendo destaque o NDWI como a variável com maior correlação com a produtividade da cultura.

Tabela 7 – Variáveis que compõem o melhor modelo para estimativa.

VÁRIAVEIS PRESENTES NO MODELO COM MELHOR DESEMPENHO.				
R	ATSAVI	NDVI	ARVI	SIPI
G	OSAVI	GNDVI	RECI	CVI
B	GOSAVI	MTVI2	GLI	TGI
NIR	NDWI	VARI	GCI	

Fonte: a autora (2025).

A identificação do NDWI (Normalized Difference Water Index) como o índice de maior correlação com a produtividade reforça a importância do conteúdo hídrico da vegetação na estimativa do rendimento da cultura. Durante o estágio R4, a soja apresenta alta demanda por água, sendo este um fator determinante para o enchimento das vagens e a manutenção do metabolismo vegetal. Dessa forma, o NDWI, que é sensível à umidade da vegetação, demonstrou ser um indicador altamente eficaz para a predição da produtividade. Isso foi demonstrado por Trindade (2018) em seu trabalho, onde correlacionou diversos IV'S com a produtividade de grãos de soja em uma fazenda em Querência (MT), e constatou que as melhores relações foram observadas com os índices associados ao conteúdo de água das folhas, sendo eles o NDWI e WBI.

Além do NDWI, os demais índices selecionados contribuíram significativamente para a modelagem, fornecendo informações complementares sobre diferentes aspectos fisiológicos da lavoura:

- **NDVI, GNDVI e RECI** – Relacionados à quantidade de clorofila e à saúde da vegetação, indicam o vigor e o potencial fotossintético das plantas (Andrade *et al.*, 2022; Gitelson *et al.*, 1996; Sergieieva, 2024).
- **OSAVI, GOSAVI e ATSAVI** – Ajustam a resposta espectral para reduzir os efeitos do solo exposto, aumentando a precisão em áreas com menor cobertura vegetal (Sergieieva, 2024; Rondeaux, Steven, Baret, 1996; Baret, Guyot, 1991).
- **MTVI2 e VARI** – Sensíveis à estrutura da vegetação e à variabilidade da biomassa ao longo do ciclo da cultura (Haboudane *et al.*, 2003; Sergieieva, 2024).
- **ARVI e SIPI** – Corrigem distorções causadas por aerossóis e estresses ambientais, aprimorando a detecção da produtividade real (Esri, 2020; Auster Tecnologia, 2020).
- **GLI, GCI, CVI e TGI** – Relacionados ao teor de pigmentos, como a clorofila e os carotenoides, permitindo a diferenciação entre plantas saudáveis e estressadas (Hunt Jr. *et al.*, 2013; Ataíde, 2024; Vincini, Frazzi, D'Alessio 2008; Starý, 2020).

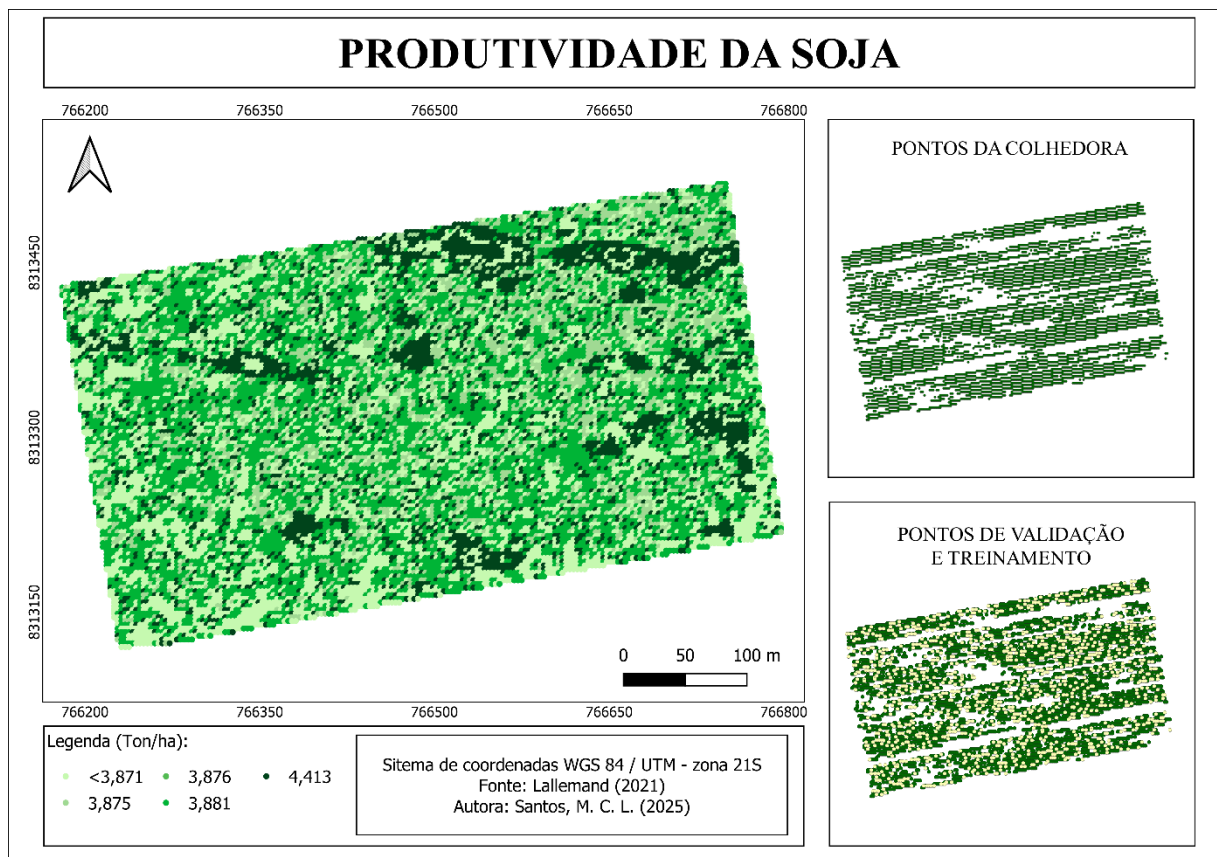
O uso combinado dessas variáveis espectrais possibilitou uma modelagem robusta, maximizando a precisão na estimativa da produtividade da soja.

Dessa forma, os resultados evidenciam que a umidade da vegetação, a biomassa e a quantidade de clorofila são fatores-chave para a predição da produtividade da soja de forma pontual, e que a utilização de índices adequados, aliados a um modelo eficiente como o Random Forest, pode fornecer estimativas altamente confiáveis para o monitoramento da cultura.

5.5 Espacialização da produtividade da soja

A espacialização dos dados de produtividade foi realizada com o objetivo de visualizar a distribuição espacial das estimativas ao longo da lavoura. No entanto, a tentativa de interpolação dos valores preditos não gerou um resultado visualmente satisfatório, o que levou à opção de representar os dados de forma pontual (Figura 11). Essa abordagem permitiu manter a fidelidade das estimativas obtidas pelo modelo, sem distorções artificiais que poderiam comprometer a interpretação dos resultados.

Figura 11 – Mapa de produtividade da soja.



Fonte: a Autora (2025).

A principal dificuldade na interpolação pode estar relacionada à distribuição espacial dos pontos amostrados e à variabilidade dos dados. Como a produtividade da soja pode apresentar padrões complexos, influenciados por fatores como variabilidade do solo, umidade, declividade e manejo agrícola, a interpolação pode gerar suavizações que não refletem com precisão as variações reais dentro da lavoura. Além disso, a alta densidade de pontos pode ter contribuído para a geração de uma superfície interpolada com ruídos, dificultando a identificação clara das tendências espaciais.

Ao optar pela representação pontual, foi possível observar com maior nitidez as regiões de maior e menor produtividade. Os maiores valores foram concentrados na região Nordeste, possivelmente devido à menor influência de fatores externos, como o efeito de borda, que impactou negativamente a produtividade na região Sul, onde há a presença de uma estrada. Essa diferença espacial pode estar relacionada tanto a condições edafoclimáticas quanto a fatores operacionais da lavoura.

A espacialização dos dados na forma pontual também permite comparações mais diretas com outros fatores ambientais e agronômicos da lavoura, como índices de vegetação e variáveis edafoclimáticas, facilitando a análise e a tomada de decisão para o manejo da cultura. Assim, os resultados reforçam a importância da escolha criteriosa da técnica de espacialização, considerando as características da área estudada e os padrões de distribuição dos dados.

6 CONCLUSÃO

A estimativa da produtividade agrícola por meio de imagens multiespectrais e aprendizado de máquina representa um avanço significativo em relação aos métodos tradicionais, proporcionando maior precisão e reduzindo a necessidade de trabalho braçal. Este estudo demonstrou que, para a cultura da soja, a combinação de índices espectrais específicos e o uso do algoritmo Random Forest resultaram em estimativas altamente acuradas, reforçando o potencial do sensoriamento remoto na agricultura de precisão.

Os resultados obtidos podem ser expandidos para outras culturas agrícolas, desde que sejam realizados ajustes nos índices espectrais e na escolha do período fenológico mais relevante. A abordagem metodológica utilizada pode ser adaptada para culturas como milho, trigo e cana-de-açúcar, explorando a relação entre variáveis espectrais e produtividade em diferentes cenários e condições ambientais.

Além disso, há espaço para aprimoramentos na modelagem, como a inclusão de dados climáticos e edáficos para enriquecer o treinamento do modelo, melhorando sua capacidade preditiva. Testes com algoritmos híbridos ou técnicas de deep learning também podem ser explorados para comparar a eficiência de diferentes abordagens.

Outro ponto relevante para estudos futuros é a otimização da interpolação espacial. Como a espacialização dos dados apresentou limitações visuais neste trabalho, métodos mais robustos de interpolação, como Krigagem e Redes Neurais, podem ser investigados para melhorar a representatividade espacial das estimativas.

Por fim, os achados deste estudo reforçam a importância da escolha criteriosa da data de aquisição das imagens, mostrando que, no caso da soja, a análise de um único momento fenológico se mostrou mais eficaz do que a utilização de séries temporais. No entanto, pesquisas futuras podem testar a viabilidade dessa abordagem em outras culturas e cenários agrícolas.

Dessa forma, este trabalho não apenas valida a eficácia do uso de aprendizado de máquina na estimativa de produtividade da soja, mas também abre caminho para novas pesquisas e aplicações, contribuindo para a evolução do sensoriamento remoto na gestão agrícola e no planejamento estratégico da produção.

REFERÊNCIAS

- ABREU JÚNIOR, M. A. C. **Estimativa de produtividade da cultura do milho a partir de imagens multiespectrais**. 2022. Dissertação (Pós-Graduação em Agricultura e Informações Geoespaciais) – Universidade Federal de Uberlândia – Campus Monte Carmelo, Monte Carmelo, 2022. Disponível em: <https://repositorio.ufu.br/handle/123456789/36220>. Acesso em: 25 de out. 2023.
- ADAMA. **Guia completo sobre a cultura da soja**. [S. l.], 22 nov. 2021. Portal: ADAMA. Disponível em: <https://portaladama.com/guia-cultura-da-soja/#:~:text=A%20cultura%20desse%20gr%C3%A3o%20%C3%A9,al%C3%A9m%20de%20diversos%20outros%20alimentos>. Acesso em: 16 de out. de 2023.
- ADDAN, Diego. **Support Vector Machine**. Unibrasil, 2019. [PowerPoint]. Disponível em: <https://www.inf.ufpr.br/dagoncalves/IA07.pdf>. Acesso em: 06 de nov. de 2023.
- ALLIBHAI, J. **Hold-out vs. cross-validation in machine learning**. [S. l.], 2018. Portal: Medium. Disponível em: <https://medium.com/@eijaz/holdout-vs-cross-validation-in-machine-learning-7637112d3f8f>. Acesso em: 30 maio 2023.
- ANDRADE, R. G. *et. al.* **Uso do sensoriamento remoto na obtenção de parâmetros biofísicos em plantios de soja**. In: BERNARDI, A. C. de C.; NAIME, J. de M.; RESENDE, A. V. de; BASSOI, L. H.; INAMASU, R. Y. (ed.). *Agricultura de precisão: resultados de um novo olhar*. Brasília, DF: Embrapa, 2014. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1002012/uso-do-sensoriamento-remoto-na-obtencao-de-parametros-biofisicos-em-plantios-de-soja>. Acesso em: 21 de set. 2023.
- ANDRADE, T. G. *et. al.* **Rev. Caatinga. Soybean yield prediction using remote sensing in southwestern Piauí State, Brazil**. Mossoró, v. 35, n. 1, p. 105 –116, jan./mar., 2022. Disponível em: <https://periodicos.ufersa.edu.br/caatinga/article/view/10255/10822>. Acesso em: 22 de out. 2023.
- APROSOJA BRASIL. **A soja: economia**. Brasília, [2016?]. Portal: Aprosoja Brasil. Disponível: <https://aprosojabrasil.com.br/a-soja/economia/#:~:text=A%20soja%20%C3%A9%20o%20principal,maior%20exportador%20mundial%20do%20gr%C3%A3o>. Acesso em: 21 de set. 2023.
- ATAÍDE, M.V.R. **Uso de drone para identificar a dominância de gramíneas invasoras e vegetação nativa em áreas de restauração do cerrado**. 2024, Dissertação (Mestrado) – Departamento de Engenharia Florestal, Universidade de Brasília, Brasília. Disponível em: http://www.rlbea.unb.br/jspui/bitstream/10482/48415/1/MarcosViniciusRezendeDeAtaide_DISSERT.pdf. Acesso em: 05 de jan. de 2025.
- AUSTER TECNOLOGIA. **Além do NDVI: um panorama sobre os índices de vegetação na agricultura**. [S. l.], 2020. Disponível em: <https://www.austertecnologia.com/single-post/indices-de-vegeta%C3%A7ao-na-agricultura#:~:text=%C3%8Dndice%20de%20Pigmento%20Insensitivo%20%C3%A0,%2C8%20a%201%2C8>. Acesso em: 05 de jan. de 2025.

BARET, F.; GUYOT, G.; MAJOR, D. J. **TSAVI**: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation. *In*: CANADIAN SYMPOSIUM ON REMOTE SENSING, 12.; 10-14 Julho 1989. Vancouver, BC, Canada, **Anais[...]**. Vancouver: IEEE, 2002. p. 1355-1358. DOI: 10.1109/IGARSS.1989.576128. Disponível em: https://www.researchgate.net/publication/3679422_TSAVI_A_vegetation_index_which_minimizes_soil_brightness_effects_on_LAI_and_APAR_estimation#fullTextFileContent. Acesso em: 05 de jan. de 2025.

BREIMAN, L. Kluwer Academic Publishers. Manufactured in The Netherlands. **Random Forests** - Berkeley. v. 45, p. 5–32, 2001. Disponível em: <https://link.springer.com/article/10.1023/a:1010933404324>. Acesso em: 25 de out. de 2023.

BUDU, E. **Random Forest vs. Extremely Randomized Trees**. [S. l.], 2023. Portal: Baeldung. Disponível em: <https://www.baeldung.com/cs/random-forest-vs-extremely-randomized-trees>. Acesso em 18 maio 2023.

BUENO, D. **Enchimento de grãos eficaz**. 2022, [S.l.]. Portal: Agrotécnico. Disponível em: <https://agrotecnico.com.br/enchimento-de-graos-eficaz/>. Acesso em: 07 de fev. de 2025.

CONAB. Soja. **Acompanhamento da safra brasileira de grãos**, Brasília, v. 10, n. 9, p. 78-85, jun. 2023. Disponível em: <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos?limitstart=0>. Acesso em: 19 de set. 2023.

CRUSIOL, L.G.T. *et. al.* **Sensoriamento remoto aplicado ao monitoramento do déficit hídrico na cultura da soja**. *In*: JORNADA ACADEMICA DA EMBRAPA SOJA, 12., 2017, Londrina. Resumos expandidos[...]. Londrina: Embrapa Soja, 2017. p. 119-127.

EPIPHANIO, J. C. N. *et. al.* Índices de vegetação no sensoriamento remoto da cultura do feijão. **Revista Pesquisa Agropecuaria Brasileira**. Brasília, v. 31, n. 6, p. 445-454, jun. 1996. Disponível em: <https://seer.sct.embrapa.br/index.php/pab/article/view/4503/1789#>. Acesso em: 23 de out. de 2023.

ESCOVEDO, T.; KOSHIYAMA, A. S. **Introdução a Data Science**: algoritmos de machine learning e métodos de análise. São Paulo: Ed. Casa do Código, 2020.

ESCOVEDO, T. **Machine Learning**: conceitos e modelos — parte II: aprendizado não-supervisionado. [S. l.], 2020. Portal: Medium. Disponível em: <https://tatianaesc.medium.com/machine-learning-conceitos-e-modelos-parte-ii-aprendizado-n%C3%A3o-supervisionado-fb6d83e4a520>. Acesso em: 16 de fev. de 2024.

ESCOVEDO, T. **Machine Learning**: conceitos e modelos — parte I: aprendizado supervisionado. [S. l.], 2020. Portal: Medium. Disponível em: [https://tatianaesc.medium.com/machine-learning-conceitos-e-modelos-f0373bf4f445#:~:text=Podemos%20dividir%20o%20aprendizado%20em,\(entrada%20%E2%80%94%20sa%C3%ADda%20desejada\)](https://tatianaesc.medium.com/machine-learning-conceitos-e-modelos-f0373bf4f445#:~:text=Podemos%20dividir%20o%20aprendizado%20em,(entrada%20%E2%80%94%20sa%C3%ADda%20desejada)). Acesso em: 23 de out. de 2023.

EOS Data Analytics. **Bandas espectrais no sensoriamento remoto**. [S. l.], [202-?]. Portal: EOS Data Analytics. Disponível em: <https://eos.com/pt/make-an-analysis/#:~:text=A%20banda%20NIR%20%C3%A9%20ideal,na%20maioria%20das%20outras%20bandas>. Acesso em: 05 de jan. de 2025.

ESRI. **Atmospherically Resistant Vegetation Index (ARVI)**. [S. l.], 2020. Portal: ESRI. Disponível em: <https://agriculture.africageoportal.com/content/56a6db4cdd1c46988a1411d0365fd5f7/about>. Acesso em: 05 de jan. de 2025.

FARIAS, J. R. B.; NEPOMUCENO, A. L.; NEUMAIER, N. **Ecofisiologia da soja**. Circular Técnica. Londrina, 2007, p. 8, set. 2007. Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/470308/ecofisiologia-da-soja>. Acesso em: 19 de out. de 2023.

FERNANDES, P; VEIVERBERG, K. V; SEBEM, E. **Determinação da Produtividade de Soja por Sensoriamento Remoto em Nível de Talhão**. In: Simpósio Brasileiro de Sensoriamento Remoto, 14, Natal, p. 2009 – 2015, abr. 2009. Disponível em: <http://mar.tecnico.unicamp.br/col/dpi.inpe.br/sbsr@80/2008/11.16.19.52/doc/2009-2015.pdf>. Acesso em: 22 de out. de 2023.

FEHR, W.R., CAVINESS, C.E. **Stages of soybean development**. Ames: Iowa State University, 1977. (Special Report, 80). Disponível em: <https://dr.lib.iastate.edu/bitstreams/13bd0d8f-66ff-4d0e-a0e3-a70c2c47f6f3/download>. Acesso em: 06 de fev. de 2025.

FILHO, Fernando Henrique Lost. **Sensoriamento remoto para monitoramento de mosca branca, Bemisia tabaci biótipo B (Hemiptera: Aleyrodidae) em soja**. 2019. Dissertação (Mestrado em Entomologia) Universidade de São Paulo - Escola Superior de Agricultura “Luis de Queiroz”, Piracicaba, 2019. Disponível em: https://www.teses.usp.br/teses/disponiveis/11/11146/tde-14052019-164149/publico/Fernando_Henrique_Lost_Filho_versao_revisada.pdf. Acesso em: 22 de out. de 2023.

GAMEIRO, Samuel *et al.* Avaliação da cobertura vegetal por meio de índices de vegetação (NDVI, SAVI e IAF) na Sub-Bacia Hidrográfica do Baixo Jaguaribe, CE. **Terræ**, [S. l.], v. 13, p. 15-22, 2016. Disponível em: <https://www.ige.unicamp.br/terrae/V13/PDFv13/TE067-2.pdf>. Acesso em: 23 de out. de 2023.

GAO, B.C. NDWI – A normalized difference water index for remote sensing of vegetation liquid water from space. **Remote Sensing of Environment**, [S. l.], v. 58, n. 3, p. 257–266, 1996. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0034425796000673>. Acesso em: 10 de jan de 2025.

GITELSON, A.A., KAUFMAN, Y., MERZLYAK, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote Sensing of Environment**, [S. l.], v. 58, n.3, p. 289-298, 1996. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0034425796000727>. Acesso em: 05 de jan. de 2025.

GODOY, C. V. **Manejo de doenças na cultura da soja**. In: WORKSHOP CTC AGRICULTURA, 16., 2017, Rio Verde. Agricultura - Resultados 2017. Rio Verde: Centro Tecnológico Comigo, 2017. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1074366/1/ID380530001.pdf>. Acesso em: 07 de fev. de 2025.

HABOUDANE, D. et al. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. **Remote sensing of environment**, [S. l.], v. 90, n. 3, p. 337-352, 2004. DOI: <https://doi.org/10.1016/j.rse.2003.12.013>. Disponível em: https://www.researchgate.net/publication/222046420_Hyperspectral_vegetation_indices_and_Novel_Algorithms_for_Predicting_Green_LAI_of_crop_canopies_Modeling_and_Validation_in_the_Context_of_Precision_Agriculture#fullTextFileContent. Acesso em: 05 jan. 2025.

HAYKIN, S. **Redes neurais: princípios e práticas**. 2. ed. São Paulo: Bookman, 2001. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=bhMwDwAAQBAJ&oi=fnd&pg=PP1&dq=Rede++neural&ots=09ttJKQRJq&sig=cX_InZNbr2ZV8tXBq7v0-dn9T_0#v=onepage&q=Rede%20%20neural&f=false. Acesso em: 06 de nov. de 2023.

HUETE, A. R. A soil vegetation adjusted index (SAVI). **Remote sensing of environment**, [S. l.], v. 25, n. 3, p. 295-309, 1988. Disponível em: <https://www.sciencedirect.com/science/article/pii/003442578890106X>. Acesso em: 05 de jan. de 2025.

HUNT JR., E. R.; DORAISWAMY, P. C.; MCMURTREY, J. E.; DAUGHTRY, C. S. T.; PERRY, E. M. A visible band index for remote sensing leaf chlorophyll content at the canopy scale. **International Journal of Applied Earth Observation and Geoinformation**, [S. l.], v. 21, p. 103-112, 2013. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0303243412001791>. Acesso em: 05 de jan. de 2025.

IBGE. **Cidades**. Disponível em: <https://cidades.ibge.gov.br/brasil/mt/primavera-do-leste/panorama>. Acesso em: 23 de out. de 2023.

INDEX DATABASE. **A database for remote sensing indices**. By The IDB Project, [S. l.], 2011. Disponível em: <https://www.indexdatabase.de/>. Acesso em: 27 de out. de 2023.

JARDINE, J. G.; BARROS, T. D. **Agroenergia – Soja**. Brasília, 2021. Portal: Embrapa. Disponível em: <https://www.embrapa.br/agencia-de-informacao-tecnologica/tematicas/agroenergia/biodiesel/materias-primas/soja#:~:text=A%20soja%20%C3%A9%20uma%20planta,por%20homens%20como%20por%20animais>. Acesso em: 13 de out. de 2023.

JENSEN, J.R. **Sensoriamento Remoto do Ambiente: uma perspectiva em recursos terrestres**. 2. ed., São José dos Campos: Parenteses, 2009. Disponível em: <https://pdfcoffee.com/sensoriamento-remoto-do-ambiente-john-r-jensen-pdf-free.html>. Acesso em: 23 de out. de 2023.

KAUFMAN, Y.J.; TANRÉ, D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIs. **IEEE Transactions on Geoscience and Remote Sensing**, [S. l.], v. 30, n. 2, p. 261-270, 1992. DOI: 10.1109/36.134076. Disponível em: <https://ieeexplore.ieee.org/document/134076>. Acesso em: 05 de jan. de 2025.

LAMPARELLI, R. A. C. **Agricultura de precisão**. Brasília, 2022. Portal: Embrapa. Disponível em: <https://www.embrapa.br/agencia-de-informacao->

tecnologica/cultivos/cana/producao/avanco-tecnologico/agricultura-de-precisao. Acesso em: 13 de out. de 2023.

LEE, C.; HERBEK J. **Estimating Soybean Yield**. In: University of Kentucky College of Agriculture, Food and Environment, Frankfort, 2005. Disponível em: <https://www2.ca.uky.edu/agcomm/pubs/agr/agr188/agr188.pdf> Acesso em: 25 de out. de 2023.

LETRAS AMBIENTAIS. **Os 9 passos para se estimar a produtividade agrícola por satélites**. Ponta Verde, 2021. Portal: Letras Ambientais. Disponível em: <https://www.letrasambientais.org.br/posts/os-9-passos-para-se-estimar-a-produtividade-agricola-por-satelites#>. Acesso em: 13 de out. de 2023.

LULU, J. , ZOLIN, C. A. **Boletim Agrometeorológico**. Sinop, 2021. Portal: Embrapa. Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1132615/1/2021-cpamt-jl-boletim-agrometeorologico-15.pdf>. Acesso em: 07 de fev. de 2025.

MANITOBA PULSE & SOYBEAN GROWERS. **Soybean Plant Development Guide**. Carman, 2018. Portal: Manitoba Pulse & Soybean Growers. Disponível em: <https://manitobapulse.ca/2016/08/soybean-plant-development/>. Acesso em: 17 de fev. de 2025.

MCKINNON, T.; HOFF, P. Comparing RGB-Based Vegetation Indices With NDVI For Drone Based Agricultural Sensing. **AgribotixLLC**, [S.l.], 2017. Disponível em: <https://www.semanticscholar.org/paper/Comparing-RGB-Based-Vegetation-Indices-With-NDVI-McKinnon/5ec3ec2a92e61e49dbc6b33f9f0d170b313128f5>. Acesso em: 05 de jan. 2025.

MICROSOFT. **O que é o aprendizado de máquina?** [S. l.], 2023. Portal: Microsoft. Disponível em: <https://learn.microsoft.com/pt-br/azure/cloud-adoption-framework/innovate/best-practices/machine-learning>. Acesso em: 23 de out. 2023.

MULLER, E.; DÉCAMPS, H. Modeling soil moisture-reflectance. **Remote sensing of environment**, [S. l.], v. 76, n. 2, p. 173 – 180, 2001. Disponível em: <https://www.sciencedirect.com/science/article/pii/S003442570000198X?via%3Dihub>. Acesso em: 05 de jan. 2025.

NEUMAIER, N.; NEPOMUCENO, A. L.; FARIAS, J. R. B, OYA, T. 1 Estádios de desenvolvimento da cultura de soja. In: BONATO, E. R. (ed.) **Estresses em soja**. Passo Fundo: Embrapa, p. 19-44, 2000. Disponível em: <https://www.alice.cnptia.embrapa.br/handle/doc/456809>. Acesso em: 19 de out. de 2023.

OLIVEIRA, C. **Conheça cada estágio fenológico da soja em dias e seus manejos**. Porto Alegre, 2024. Portal: Aegro. Disponível em: <https://blog.aegro.com.br/estadio-fenologico-da-soja/#:~:text=definir%20os%20est%C3%A1dios,-VE,a%20forma%C3%A7%C3%A3o%20das%20folhas%20verdadeiras>. Acesso em: 07 de fev. de 2025.

OLIVEIRA, L. E. M. **Temas em fisiologia vegetal: morfologia de sementes**. Lavras, [2015?]. Portal: Universidade Federal de Lavras. Disponível em:

[http://www.ledson.ufla.br/metabolismo-da-germinacao/morfologia-de-sementes/#:~:text=A\)%20Cotil%C3%A9dones,embrion%C3%A1rias%2C%20das%20Angiospermas%20e%20Gimnospermas](http://www.ledson.ufla.br/metabolismo-da-germinacao/morfologia-de-sementes/#:~:text=A)%20Cotil%C3%A9dones,embrion%C3%A1rias%2C%20das%20Angiospermas%20e%20Gimnospermas). Acesso em: 27 de out. 2023.

PARSONS, C. **O que é um modelo de machine learning**. [S. l.], 2021. Portal: NVIDIA. Disponível em: <https://blog.nvidia.com.br/2021/09/28/o-que-e-um-modelo-de-machine-learning/#:~:text=O%20Que%20%C3%A9%20um%20Modelo%20de%20ML%3F,os%20mecanismos%20matem%C3%A1ticos%20da%20AI>. Acesso em: 23 de out. de 2023.

PENUELAS, J.; BARET, F.; FILELLA, I. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. **Photosynthetica**, [S. l.], v. 31, p. 221-230, 1995. Disponível em: https://www.researchgate.net/publication/229084513_Semi-Empirical_Indices_to_Assess_CarotenoidsChlorophyll-a_Ratio_from_Leaf_Spectral_Reflectance. Acesso em: 05 de jan. 2025.

PRESTES *et. al.* Avaliação de metodologia de estimativa de produtividade de soja por meio de aeronave não tripulada e técnica de aprendizado de máquina baseada em regressão. **Mundi Engenharia, Tecnologia e Gestão**. Paranaguá, 2020. v. 5, n. 3, p. 245-01, 245-12, 2020. DOI:10.21575/25254782rmetg2020vol5n31253. Disponível em: <https://revistas.ifpr.edu.br/index.php/mundietg/article/view/1012/877>. Acesso em: 10 de jan. de 2025.

PLANET LABS. **Data Driven Precision Agriculture with Planet**. 2024. Portal: Planet Labs. Disponível em: <https://www.planet.com/industries/agriculture/>. Acesso em: 16 dez. de 2024.

PLANET SCOPE. **O que é o NDVI?** São Paulo, 2023. Portal: SCCON GEOSPATIAL. Disponível em: <https://www.scon.com.br/o-que-e-ndvi/>. Acesso em: 23 de out. de 2023.

PONZONI, F. J.; SHIMABUKURO, Y. E.; KUPLICH, T. M.. **Sensoriamento remoto da vegetação**. 2. ed. atual. e ampl. São Paulo: Oficina de Textos, 2012.

QGIS. **QGIS: a liderança do SIG de código aberto**. [S. l.], [201-?]. Portal: QGIS. Disponível em: https://www.qgis.org/pt_BR/site/about/index.html. Acesso em: 28 out de 2023.

RONDEAUX, G., STEVEN, M., BARET, F. Optimization of Soil-Adjusted Vegetation Indices. **Remote sensing of environment**, [S. l.], v. 55, n. 2, p. 95-107, 1996, Pages 95-107. DOI: 10.1016/0034-4257(95)00186-7. Disponível em: https://www.researchgate.net/publication/222741732_Optimization_of_Soil-Adjusted_Vegetation_Indices. Acesso: 05 de jan de 2025.

RUDORFF, B. F. T. **Produtos de Sensoriamento Remoto**. Inpe, 2011. Disponível em: https://rbrebello.files.wordpress.com/2011/02/rudorff_-_produtos_de_sensoriamento_remoto.pdf. Acesso em: 13 de out. de 2023.

SCHWAB, N. T.; ALFF, R. **Escala fenológica da soja**. [S. l.], 2013. Portal: Elevagro. Disponível em: <https://elevagro.com/conteudos/fotos/escala-fenologica-da-soja>. Acesso em: 19 de out. de 2023.

- SERGIEIEVA, K. **Índices de vegetação para impulsionar soluções digitais**. [S.l.], 2024. Portal: EOS Data Analytics. Disponível em: <https://eos.com/pt/blog/indices-de-vegetacao/>. Acesso em: 05 de jan. de 2025.
- SHAH, T. **About train, validation and test sets in machine learning**. [S. l.], 2017. Portal: Medium. Disponível em: <https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7>. Acesso em: 30 maio 2023.
- STARÝ, K. *et al.* Comparing RGB - based vegetation indices from UAV imageries to estimate hops canopy área. **Agronomy Research**, [S. l.], v. 18, n. 4, p. 2592-2601, 2020. DOI: <https://doi.org/10.15159/AR.20.169>. Disponível em: <https://dspace.emu.ee/server/api/core/bitstreams/47a5099f-5d81-4261-b9c9-068216fb6b18/content>. Acesso em: 05 de jan. de 2025.
- TOTVS. **Gestão agrícola: Como funciona o sensoriamento remoto na agricultura?** [S. l.], 2020. Portal: TOTVS. Disponível em: <https://www.totvs.com/blog/gestao-agricola/sensoriamento-remoto-na-agricultura/#:~:text=O%20sensoriamento%20remoto%20na%20agricultura%20utiliza%20instrumentos%20de%20alta%20performance,de%20produ%C3%A7%C3%A3o%20na%20regi%C3%A3o%2C%20etc>. Acesso em: 22 de out. de 2023.
- Trindade, F. S.; Alves, M. C.; Noetzold, R. Análise de perfis espectro temporais de NDVI e EVI no ciclo fenológico da cultura de soja e sua aplicação em sojicultura de precisão. *In*: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO -SBSR, 18., Santos, mai. 2017. **Anais [...]**. Santos: SBSR, 2017, p. 3878 - 3885. Disponível em: <http://marte2.sid.inpe.br/col/sid.inpe.br/marte2/2017/10.27.13.20.29/doc/59427.pdf?metadatarpository=sid.inpe.br/marte2/2017/10.27.13.20.30&mirror=dpi.inpe.br/marte2/2013/05.17.15.03.08>. Acesso em: 22 de out. de 2023.
- TRINDADE, F. S. **Monitoramento da cultura da soja em agrossistema com espectrorradiômetro orbital de resolução moderada**. 2018. 86 f. Dissertação (Mestrado em Engenharia Agrícola)-Universidade Federal de Lavras, Lavras, 2018. Disponível em: http://repositorio.ufla.br/jspui/bitstream/1/29693/2/DISSERTA%C3%87%C3%83O_Monitoramento%20da%20cultura%20da%20soja%20em%20agrossistema%20com%20espectrorradi%C3%B4metro%20orbital%20de%20resolu%C3%A7%C3%A3o%20moderada.pdf. Acesso em: 10 de jan. de 2025.
- USDA. **World agricultural production**. Foreign Agricultural Service, [S. l.], 2023, p. 1-41, Oct. 2023. Disponível em: <https://apps.fas.usda.gov/psdonline/circulars/production.pdf>. Acesso em: 19 de set. 2023.
- USP. **Aspectos do Cerrado - clima**. São Paulo, 2000. Portal: Universidade de São Paulo. Disponível em: http://ecologia.ib.usp.br/cerrado/aspectos_clima.htm. Acesso em: 26 out. de 2023.
- VINCINI, M.; FRAZZI, E.; D'ALESSIO, P. A broad-band leaf chlorophyll vegetation index at the canopy scale. **Precision Agriculture**, [S. l.], v. 9, p. 303-319, 2008. DOI: <https://doi.org/10.1007/s11119-008-9075-z>. Disponível em: https://www.researchgate.net/publication/225681715_A_broad-band_leaf_chlorophyll_vegetation_index_at_the_canopy_scale. Acesso em: 05 jan. 2025.

WAIKATO. **Weka 3**: Software de aprendizado de máquina em Java. [S. l.], 201?. Portal: Waikato University. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/index.html>. Acesso em: 28 out de 2023.

ZAMPIER, D. P.; GABARDO, G., GUIMARÃES, A. M. Modelo de predição da ferrugem da soja baseado em técnicas de aprendizado de máquina e imagens de rpa. *In*: ENCONTRO ANUAL DE INICIAÇÃO CIENTÍFICA, 27., 2018, Ponta Grossa. **Anais [...]**. Ponta Grossa: UEPG, 2018. Disponível em: https://siseve.apps.uepg.br/storage/eaic2018/9_David_Patrick_Zampier-153646140165686.pdf. Acesso em: 22 de out. 2023.