

# ESTUDO DE ÍNDICES DE VEGETAÇÃO POR SENSORIAMENTO REMOTO PARA MONITORAMENTO DA CULTURA DE CANA-DE-AÇÚCAR

Gabriel Baptista Calil<sup>1</sup>

João Ricardo Favani<sup>2</sup>

Renata Bruna dos Santos Coscolin Favani<sup>2</sup>

## RESUMO

O cultivo da cana-de-açúcar caracteriza-se como uma das mais relevantes atividades na história do Brasil. O ciclo produtivo da cana foi responsável por moldar as relações sociais, trabalhistas e políticas que muito perduram até os dias atuais. Assim, estimar a produtividade da cana-de-açúcar de forma confiável e com antecedência em relação à sua colheita é de extrema importância para a tomada de decisões do produtor. Neste contexto, este trabalho teve como objetivo avaliar a utilização de índices de vegetação obtidos através de sensoriamento remoto, como ferramentas de identificação e monitoramento da produção e desenvolvimento da cana-de-açúcar. Foram utilizados índices de vegetação diversos, calculados por combinações de bandas multiespectrais capturadas por sensores do Sentinel-2, juntamente com a plataforma *Google Earth Engine*, demonstrando-se eficiente para o cálculo e a análise de 16 diferentes índices de vegetação, como NDVI, ARVI, SAVI, dentre outros. Em suma, o presente estudo reforçou a importância de se integrar o uso de métodos de sensoriamento remoto com técnicas de análises estatísticas avançadas como o PCA e o *K-means*, demonstrando-se satisfatório quanto a segmentação das áreas de interesse, possibilitando a observação de padrões homogêneos e a identificação de semelhanças entre as zonas de manejo. Os resultados indicaram que a combinação de ambos os métodos permitiu uma análise detalhada e precisa das condições do campo, possibilitando intervenções direcionadas e conseqüentemente, otimizando o cultivo e a produtividade da cana-de-açúcar.

Palavras-chaves: Análise de Componentes Principais; Bandas Multiespectrais; Sentinel-2; *Google Earth Engine*.

## 1 INTRODUÇÃO

Iniciado ainda no período colonial, o cultivo da cana-de-açúcar tem sido uma das mais relevantes atividades na história do Brasil. Além de ser um dos pilares da

---

<sup>1</sup> Discente no curso Big Data no Agronegócio na FATEC Pompéia, Pompéia-SP.

<sup>2</sup> Docentes do curso Big Data no Agronegócio na FATEC Pompeia, Pompéia-SP.

economia brasileira, o ciclo produtivo da cana foi responsável por moldar as relações sociais, trabalhistas e políticas que muito perduram até os dias atuais (Santana et al., 2022).

A associação de clima e solo adequados e contexto histórico-econômico propícios permitiu a ampla disseminação da cana-de-açúcar no Brasil, garantindo hoje ao país, de acordo com a Companhia Nacional de Abastecimento - CONAB, no levantamento realizado em agosto deste ano, o lugar de maior produtor mundial de cana-de-açúcar e segundo maior de etanol (CONAB, 2024).

A diversidade de produtos obtidos a partir da cana-de-açúcar alavanca a rentabilidade e sustentabilidade da cadeia suco energética, tendo o açúcar e o etanol como protagonistas, mas com elevada contribuição da bioeletricidade, dos fertilizantes orgânicos (vinhaça e torta de filtro), das leveduras, além de outros produtos que fazem parte da química verde (Ramos et al., 2024).

A produção de cana-de-açúcar do Brasil para a safra 2024/25, conforme o levantamento mais recente da CONAB, divulgado em novembro deste ano, está estimada em 678,67 milhões de toneladas, o que representa uma diminuição de 4,8% em relação à safra anterior. No entanto, a área destinada à colheita apresentou crescimento de 4,3%, alcançando 8,7 milhões de hectares, resultando em um aumento na extensão plantada. A produtividade média também sofreu queda, com previsão de 78.048 kg/ha, o que representa uma redução de 8,8% comparada à safra passada (CONAB, 2024).

Os principais fatores que contribuíram para essa diminuição incluem o impacto das condições climáticas adversas, como baixos índices pluviométricos e altas temperaturas, especialmente nas regiões Centro-Sul, que são responsáveis por 91% da produção nacional. Além disso, as queimadas nos canaviais, que afetaram diversas áreas no ciclo de produção, também prejudicaram o rendimento das lavouras.

A fim de se otimizar a produtividade de cana-de-açúcar, com o desenvolvimento de aplicações de sensoriamento remoto para o estudo da vegetação, diferentes índices de vegetação (IV's) têm sido propostos para avaliá-las, quantitativa e qualitativamente, por meio do uso de dados multiespectrais. Estes índices podem ser obtidos a partir de imagens de satélite e fornecem uma das melhores formas possíveis de se obter parâmetros biofísicos da vegetação, especialmente em grandes áreas (Moreira, 2019).

Dentre os vários índices desenvolvidos, o NDVI (Acrônimo do Inglês: Índice de

Vegetação por Diferença Normalizada) é um dos mais utilizados atualmente, uma vez que fornece estimativas de parâmetros biofísicos da vegetação, como Índice de Área Foliar (IAF), acúmulo de biomassa e produtividade. Outros índices também foram desenvolvidos com o objetivo de captar variações estruturais do dossel das culturas e também estimar o conteúdo de água foliar, como o EVI (Acrônimo do Inglês: Índice de Vegetação Melhorado) e o NDWI (Acrônimo do Inglês: Índice da Água por Diferença Normalizada), respectivamente (Maia, 2019). Para o cálculo desses índices, o uso de imagens de alta resolução é de extrema importância, e é nesse contexto que o Sentinel-2 se destaca.

O Sentinel-2 consiste numa constelação de três satélites que faz parte do programa *Copernicus* da Agência Espacial Europeia (ESA). O mesmo é composto por satélites idênticos (Sentinel-2A, Sentinel-2B e Sentinel-2C), tendo como intuito o monitoramento da superfície terrestre em alta resolução, fornecendo imagens multiespectrais através de 13 bandas espectrais que variam do visível ao infravermelho (ESA, 2024).

As imagens são capturadas a uma resolução espacial que varia de 10 a 60 metros, dependendo da banda, com um ciclo de revisita de 5 dias. Essas bandas possibilitam a realização de cálculos de diversos fatores relacionados ao solo e índices de vegetação como o NDVI, auxiliando na análise de uso do solo, recursos hídricos, cobertura vegetal e mudanças ambientais (ESA, 2024).

O *Google Earth Engine* (GEE) refere-se à uma plataforma de desenvolvimento em nuvem, desenvolvida pelo Google, cuja função é atuar como uma interface de processamento de dados geoespaciais, permitindo a execução de diversas funções, tais como a análise e visualização de quantidades robustas de dados. Possui uma grande quantidade de imagens de satélites e outras fontes de dados geoespaciais (GEE, 2024).

Tendo como principal vantagem, a capacidade de processar, manipular e monitorar *petabytes* de dados sem a necessidade de equipamento especializado ou de uma infraestrutura computacional local, o GEE torna-se assim, a forma mais acessível para a realização de análises complexas, como por exemplo, o cálculo de índices de vegetação (GEE, 2024).

Devido à capacidade de redução da dimensionalidade dos dados e de identificação de padrões mais relevantes para a análise, a combinação das técnicas PCA (Acrônimo do Inglês: Análise de Componentes Principais) com o método *K-means* tem

sido amplamente utilizada tanto no monitoramento de culturas agrícolas quanto em diversas outras áreas, facilitando assim, a interpretação dos dados por este método. No contexto de análise de índices de vegetação, essa técnica possui grande utilidade, visto possibilitar uma análise mais detalhada dos agrupamentos relevantes para o desenvolvimento da cultura (Pereira et al., 2014).

Neste contexto, o objetivo deste trabalho foi avaliar a utilização de índices de vegetação como ferramentas auxiliares na identificação de ambientes de produção de cana-de-açúcar, apresentando metodologias para o desenvolvimento de estratégias que visam aprimorar o processo de definição dos ambientes de produção de cana-de-açúcar com base nas ferramentas de sensoriamento remoto, buscando alternativas e ferramentas para facilitar e agilizar a caracterização de novas áreas.

## **2 MATERIAIS E MÉTODOS**

Foram definidas como áreas de interesse para o presente trabalho, duas plantações de cana-de-açúcar encontradas em diferentes regiões do estado de São Paulo.

O talhão 1, encontra-se nas coordenadas geográficas médias ([-21.362483 de Latitude e -48.966211 de Longitude]), pela região norte do estado de São Paulo próximo ao município de Itajobi, possuindo uma área total de 7,31 hectares, com incidência média de temperatura anual de 23,75 °C.

O talhão de número 2, encontra-se nas coordenadas geográficas médias ([-22.534650 de Latitude e -50.684281 de Longitude]), próximo ao município de Macaraí com área total de 4.74 hectares e temperatura média anual de 23.09°C.z'

Ambas as áreas foram analisadas através de um algoritmo escrito dentro do GEE, utilizando a linguagem de programação *JavaScript*. De modo que, realizou-se uma função para mascarar a cobertura das nuvens presentes nas imagens posteriormente coletadas através do satélite de observação da terra, Sentinel-2.

Foi utilizado o sistema de bandas do Sentiel-2, que utiliza de um sensor multiespectral (MSI – Multispectral Instrument), para captar imagens em treze bandas espectrais diferentes, sendo responsáveis por fornecer as informações necessárias no cálculo dos índices de vegetação. Dentre essas informações, temos dados sobre a atmosfera, solo e água.

Igualmente, realizou-se o cálculo das médias dos índices para cada área de

interesse, sendo que, posteriormente, foi realizado um processamento nos dados coletados afim de se remover campos indesejados como valores não numéricos que afetariam no cálculo dos índices.

Por fim, os dados foram exportados em um formato CSV (*Comma Separated Value*), com o intuito de promoção da facilidade de sua utilização futura.

Através das análises realizadas em ambos os talhões, foram coletados dados referentes à dezesseis índices de vegetação, objetivando avaliar diferentes características relacionadas ao cultivo da cana-de-açúcar. Esses índices de vegetação estão representados na Tabela 1, que fornece uma visão geral de suas respectivas fórmulas e nomenclaturas:

**Tabela 1: Descrição e fórmulas de dezesseis índices de vegetação analisados no trabalho**

<b>Sigla</b>	<b>Nome</b>	<b>Tradução</b>	<b>Formula</b>
NDVI	<i>Normalized Difference Vegetation Index</i>	Índice de Vegetação por Diferença Normalizada	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$
RECI	<i>Red-Edge Chlorophyll Index</i>	Índice de Clorofila na Borda Vermelha	$(\text{NIR} / \text{RedEdge}) - 1$
NDRE	<i>Normalized Difference Red-Edge</i>	Índice de Diferença Normalizada na Borda Vermelha	$(\text{NIR} - \text{RedEdge}) / (\text{NIR} + \text{RedEdge})$
SAVI	<i>Soil-Adjusted Vegetation Index</i>	Índice de Vegetação Ajustado ao Solo	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red} + L) * (1 + L)$
MSAVI	<i>Modified Soil-Adjusted Vegetation Index</i>	Índice de Vegetação Ajustado ao Solo Modificado	$0.5 * (2 * \text{NIR} + 1 - \sqrt{(2 * \text{NIR} + 1)^2 - 8 * (\text{NIR} - \text{Red})})$
OSAVI	<i>Optimized Soil-Adjusted Vegetation Index</i>	Índice de Vegetação Ajustado ao Solo Otimizado	$(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red} + 0.16)$
GNDVI	<i>Green Normalized Difference Vegetation Index</i>	Índice de Diferença Normalizada do Verde	$(\text{NIR} - \text{Green}) / (\text{NIR} + \text{Green})$
NDWI	<i>Normalized Difference Water Index</i>	Índice de Diferença Normalizada da Água	$(\text{NIR} - \text{SWIR}) / (\text{NIR} + \text{SWIR})$
NDSI	<i>Normalized Difference Snow Index</i>	Índice de Diferença Normalizada da Neve	$(\text{Green} - \text{SWIR}) / (\text{Green} + \text{SWIR})$
ARVI	<i>Atmospherically Resistant Vegetation Index</i>	Índice de Vegetação Resistente à Atmosfera	$(\text{NIR} - (2 * \text{Red} - \text{Blue})) / (\text{NIR} + (2 * \text{Red} - \text{Blue}))$
EVI	<i>Enhanced Vegetation Index</i>	Índice de Vegetação Aprimorado	$2.5 * (\text{NIR} - \text{Red}) / (\text{NIR} + 6 * \text{Red} - 7.5 * \text{Blue} + 1)$
VARI	<i>Visible Atmospherically Resistant Index</i>	Índice Visível Resistente à Atmosfera	$(\text{Green} - \text{Red}) / (\text{Green} + \text{Red} - \text{Blue})$
LAI	<i>Leaf Area Index</i>	Índice de Área Foliar	$-\ln((\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})) * (\text{LAI}_{\text{max}} - 0) /$

NBR	<i>Normalized Burn Ratio</i>	Razão Normalizada da Queimada	LAI <sub>max</sub> (NIR - SWIR) / (NIR + SWIR)
SIPI	<i>Structure Insensitive Pigment Index</i>	Índice de Pigmento Insensível à Estrutura	(NIR - Blue) / (NIR - Red)
GCI	<i>Green Chlorophyll Index</i>	Índice de Clorofila Verde	NIR / Green - 1

Fonte: Elaborado pelo autor (2024)

Os dados no arquivo CSV são tratados por um algoritmo desenvolvido com a linguagem de programação *Python* (PYTHON, 2024) e a biblioteca *Pandas* focada em manipulação, processamento e análise de dados (PANDAS, 2024).

De início, realizou-se um pré-processamento, com o intuito de que os dados indesejados e prejudiciais para a análise, como dados não numéricos ou não relacionados aos índices de vegetação fossem removidos. Também foram utilizadas as bibliotecas *Scikit-learn* (SCIKIT-LEARN, 2024), focada em algoritmos e técnicas de aprendizado de máquina, e *Numpy* (NUMPY, 2024), cujo intuito é facilitar grandes operações matemáticas, como por exemplo, o trabalho com matrizes numéricas.

Uma função do *scikit-learn* chamada '*StandardScaler*', foi utilizada inicialmente, realizando o importante papel de normalizar os valores numéricos, garantindo assim, que todas as variáveis possuíssem a mesma escala (Pedregosa et al., 2011). Tal procedimento faz-se necessário para a realização da próxima etapa, referente ao cálculo do PCA, pois este possui uma grande sensibilidade em relação a variações na escala dos valores (Shlens, 2014).

Posteriormente, utilizou-se o método *K-means* para a classificação das áreas de interesse, dividindo as mesmas em grupos ou *clusters*, com base na similaridade das características observadas pelo PCA. (Wang & Xu, 2018). Do mesmo modo, desenvolveu-se o chamado método *Elbow*, para identificação da quantidade de *clusters* necessários para cada amostra de dados, esse método é fundamental para a determinação do número ideal de *clusters* a serem utilizados, sendo uma das abordagens mais simples e eficientes para tal contexto (Syakur et al., 2018).

Por fim, foi efetuado primeiramente, uma normalização dos dados pela função '*MinMaxScaler*' do *scikit-learn*, para posteriormente ser realizado o método *K-means* em si, sendo o mesmo também proveniente a partir do *scikit-learn*.

Ao serem executadas todas as fases do sistema em ambas as plantações de cana-de-açúcar, foi realizada a coleta de imagem em dois períodos distintos para cada uma delas, sendo os mesmos compreendidos entre as datas 01 a 31 de novembro de

2022 (período mensal), e 01 de novembro de 2021 a 30 de novembro de 2022 (período anual).

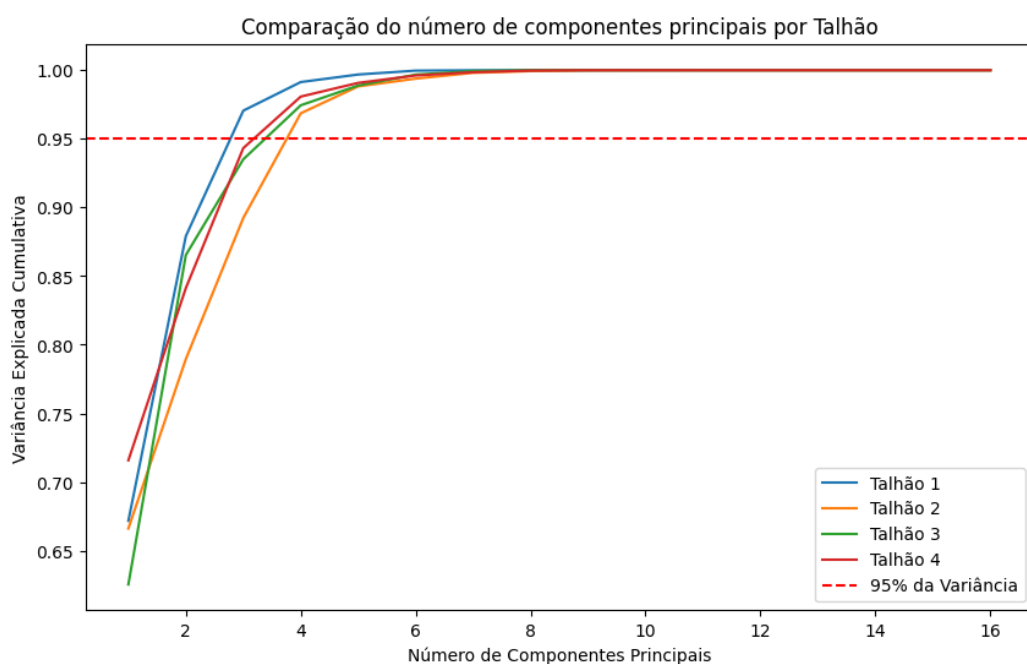
### 3 RESULTADOS E DISCUSSÃO

Com o intuito de comparação entre os padrões de relevância nos resultados, foram coletadas o total de 6 imagens no período de um mês e 68 imagens no período de um ano para o primeiro talhão. Do mesmo modo, coletou-se 2 imagens no período de um mês e 36 imagens no período de um ano para o segundo talhão.

A quantidade de imagens foi afetada em virtude da função de máscara de cobertura de nuvens. Assim, ressalta-se que, para uma análise mais satisfatória, foram coletadas apenas imagens que possuíam até vinte por cento de cobertura de nuvens.

Subdividiram-se os índices de vegetação coletados de ambos os talhões em componentes principais com base na sua relevância, entendida como a capacidade de explicação de cada índice em relação a variabilidade observada pela análise dos dados. Da mesma forma, foi calculada uma média dos valores dos autos vetores de cada base de dados, retornados pelo sistema em formato de dados textuais e gráficos, conforme representado no Gráfico 1.

**Gráfico 1: Quantidade de componentes principais que explicam até 95% da variância de cada amostra de dados**

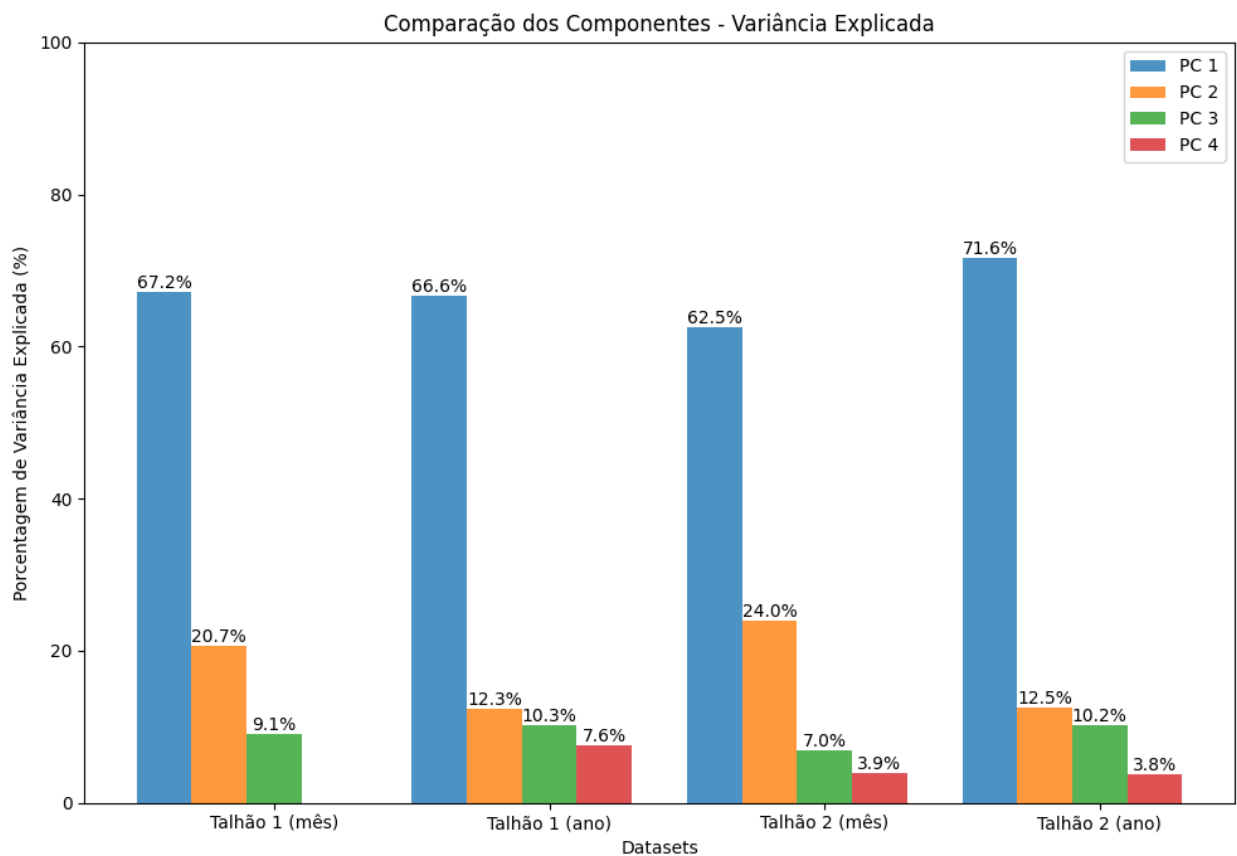


Fonte: Elaborado pelo autor (2024)

Através da análise do Gráfico 1, é possível observar que ambos os casos se situavam na faixa de dois a quatro componentes principais para explicar a porcentagem previamente definida de 95%.

Por este motivo, definiu-se uma pequena margem de erro de porcentagem acima daquela definida anteriormente, para a realização do próximo processo de comparação de variância de cada componente, conforme pode ser visualizado no Gráfico 2.

**Gráfico 2: Comparação dos componentes principais explicando a porcentagem de variância em cada amostra**



Fonte: Elaborado pelo autor (2024)

Realizados os ajustes necessários, a variabilidade foi explicada em três componentes principais (PC) na primeira amostra e quatro componentes nas demais, compondo assim, aproximadamente, 95% da variância.

No entanto, em todos os casos, é possível observar uma diferença na relação do primeiro componente principal (PC1) com os demais apresentados. O PC1 se trata do componente responsável por capturar a maior parte da variação total dos dados,

representando assim as características mais dominantes e relevantes para a análise. Ressalta-se o mesmo ser responsável por explicar mais de 60% da variância em todos os quatro casos, tendo uma contribuição ainda maior no último caso, onde sua porcentagem ultrapassa a margem de 71% (Chang, 2023).

Tais resultados sugerem que grande parte da variação dos índices de vegetação coletados pode ser atribuída a algum fator dominante como por exemplo, a saúde geral da vegetação, que pode ser influenciada por combinações de nutrientes, disponibilidade hídrica ou intensidade luminosa, formando assim, uma forte correlação entre os índices. Isso pode ser explicado pelo comportamento consistente da vegetação em resposta as variáveis ambientais ao longo do tempo, como por exemplo, disponibilidade da água, nutrientes ou condições relacionadas ao clima e fatores meteorológicos.

Em se tratando de PCA, geralmente, o primeiro componente tende a explicar 90% ou mais da variância ou variabilidade (Jolliffe, 2016).

Assim, observados os valores obtidos pelo PC, conclui-se que os mesmos não são totalmente válidos para uma análise geral, pois todos apresentaram uma média de 65% da variância ou variabilidade.

Diante do exposto, realizou-se um estudo aprofundado nos índices de vegetação de cada amostra de dados com o intuito de se identificar quais possuem maior relevância para a sua composição. Desta forma, foram selecionados os cinco índices de maior relevância para o primeiro componente de cada *dataset* (Tabela 2).

**Tabela 2 – Contribuição dos 5 índices de vegetação mais relevantes para o PC1 de cada amostra**

<b>Amostra de Dados</b>	<b>Índice</b>	<b>Autovetores</b>
Talhão 01 (mês)	<b>MSAVI</b>	<b>0,304298</b>
Talhão 01 (mês)	<b>NDVI</b>	<b>0,304096</b>
Talhão 01 (mês)	<b>OSAVI</b>	<b>0,304096</b>
Talhão 01 (mês)	<b>SAVI</b>	<b>0,304096</b>
Talhão 01 (mês)	<b>ARVI</b>	<b>0,301215</b>
Talhão 01 (ano)	SAVI	0,305582
Talhão 01 (ano)	OSAVI	0,305581
Talhão 01 (ano)	NDVI	0,305581
Talhão 01 (ano)	ARVI	0,300066
Talhão 01 (ano)	NDRE	0,298786
Talhão 02 (mês)	<b>NDVI</b>	<b>0,311646</b>

Talhão 02 (mês)	<b>OSAVI</b>	<b>0,311646</b>
Talhão 02 (mês)	<b>SAVI</b>	<b>0,311646</b>
Talhão 02 (mês)	<b>MSAVI</b>	<b>0,310557</b>
Talhão 02 (mês)	<b>ARVI</b>	<b>0,308597</b>
Talhão 02 (ano)	SAVI	0,294393
Talhão 02 (ano)	OSAVI	0,294392
Talhão 02 (ano)	NDVI	0,294392
Talhão 02 (ano)	ARVI	0,293823
Talhão 02 (ano)	NDRE	0,287775

Fonte: Elaborado pelo autor (2024)

Através da análise dos dados representados na Tabela 2, é possível identificar alguns padrões importantes. Em primeiro lugar observa-se o fato de que entre os 16 índices de vegetação utilizados para análise, somente 6 aparecem como relevantes dentro dos PC1 de cada amostra.

Desse modo, pode-se deduzir que em ambos os casos, independente de local ou época, existe um padrão de relevância dos índices para a cultura da cana-de-açúcar. Tal fato é ainda mais reforçado ao se analisar o valor de contribuição de cada índice, extremamente próximos em todos os casos.

Esse fato pode ser tratado como uma vantagem para o cultivo da cultura, pois os fatores mais influentes para seu crescimento se mantêm próximos em ambos casos (mensais e anuais), o que de certa forma facilita o planejamento do plantio.

Identificada tal consistência, os usos da análise de índices de vegetação podem ser amplamente aplicáveis em diferentes talhões e regiões. Já que índices de vegetação, como NDVI e SAVI, são ferramentas eficazes no monitoramento de biomassa e carbono em áreas de plantio de cana-de-açúcar, podendo fornecer estimativas valiosas das condições da vegetação e das alterações ambientais ao longo do tempo, corroborando assim, com a consistência observada dos padrões de relevância dos primeiros componentes principais (Polonio, 2015).

Outro ponto fundamental a ser observado nos dados analisados refere-se ao NDVI, que, apesar de ser o índice de vegetação comumente mais utilizado em grande parte das pesquisas e trabalhos em geral, devido a sua simplicidade e eficácia, possui um valor de relevância que, na maioria dos casos, pode se equiparar com alguns outros índices distintos. No contexto do presente estudo, outros índices se mostraram mais relevantes para a análise da cana-de-açúcar que o próprio NDVI, como foi o caso do SAVI e MSAVI.

Neste contexto, é de grande importância a utilização de índices de vegetação

distintos, obtidos de imagens suborbitais, em diferentes fases do ciclo de desenvolvimento da cana-de-açúcar, como variáveis preditoras para a estimativa de produtividade, visto o conhecimento de que a estimativa da produtividade desta cultura, de forma confiável e com antecedência em relação à colheita faz-se importante para a tomada de decisão do produtor (Speranza et al., 2023).

Da mesma forma, observou-se que a utilização de índices como o SAVI (Acrônimo do Inglês: Índice De Vegetação Ajustado Ao Solo), o MSAVI (Acrônimo do Inglês: Índice De Vegetação Ajustado Ao Solo Modificado) e o OSAVI (Acrônimo do Inglês: Índice De Vegetação Ajustado Ao Solo Otimizado) foram tão importantes quanto o NDVI na composição dos dados coletados. Os mesmos foram responsáveis por valores próximos, e até mesmo maiores (em alguns casos), que o NDVI, no quesito de relevância.

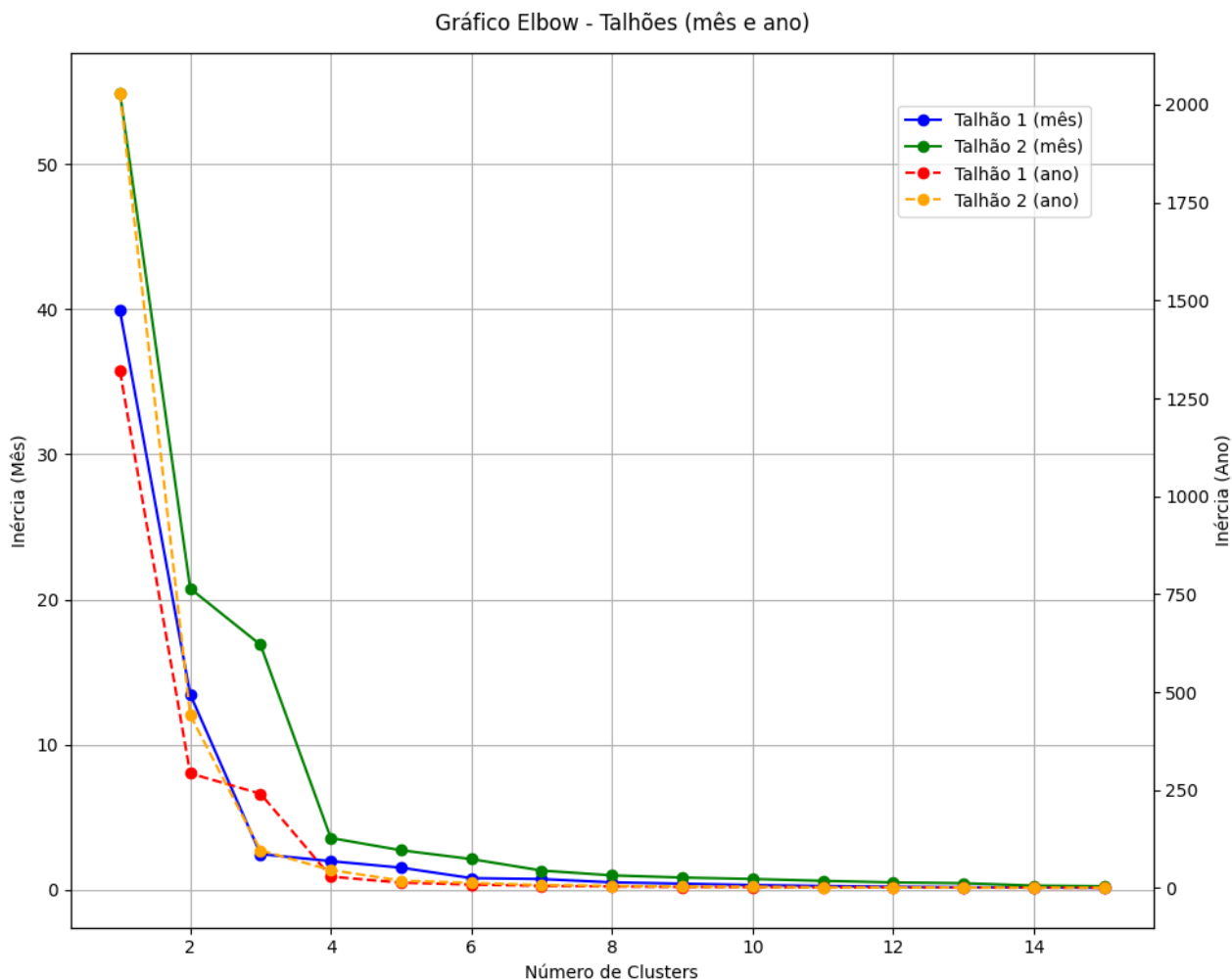
Igualmente, verificou-se a importância da utilização de índices como o ARVI (Acrônimo do Inglês: Índice De Vegetação Resistente À Atmosfera) e o NDRE (Acrônimo do Inglês: Índice De *RedEdge* Por Diferença Normalizada) no monitoramento de culturas como a cana-de-açúcar ou semelhantes.

O ARVI acrescenta melhorias de precisão das análises de vigor da vegetação através de correções relacionadas a influências atmosféricas, como poluição e poeira. Da mesma forma, o NDRE tem a função de proporcionar análises de dados de nutrição vegetal e estado de estresse hídrico, auxiliando assim, no aprimoramento das técnicas de manejo do campo (Ataíde, 2024).

Esses índices são utilizados para estimar a saúde das plantas em diferentes estágios de desenvolvimento. Ao combiná-los com outros índices como o próprio NDVI, é possível se obter uma visão mais completa do estado fisiológico das culturas, permitindo intervenções mais precisas para otimizar a produtividade (Dias, 2022).

O primeiro passo para o estudo do *K-means* foi a utilização do método *Elbow* para cada uma das amostras de dados, com o intuito de identificar o número de *clusters* necessários para realização do *K-means*, conforme demonstrado no Gráfico 3:

### Gráfico 3: Número de *clusters* necessários para cada *K-means* através do método *Elbow*



Fonte: Elaborado pelo autor (2024)

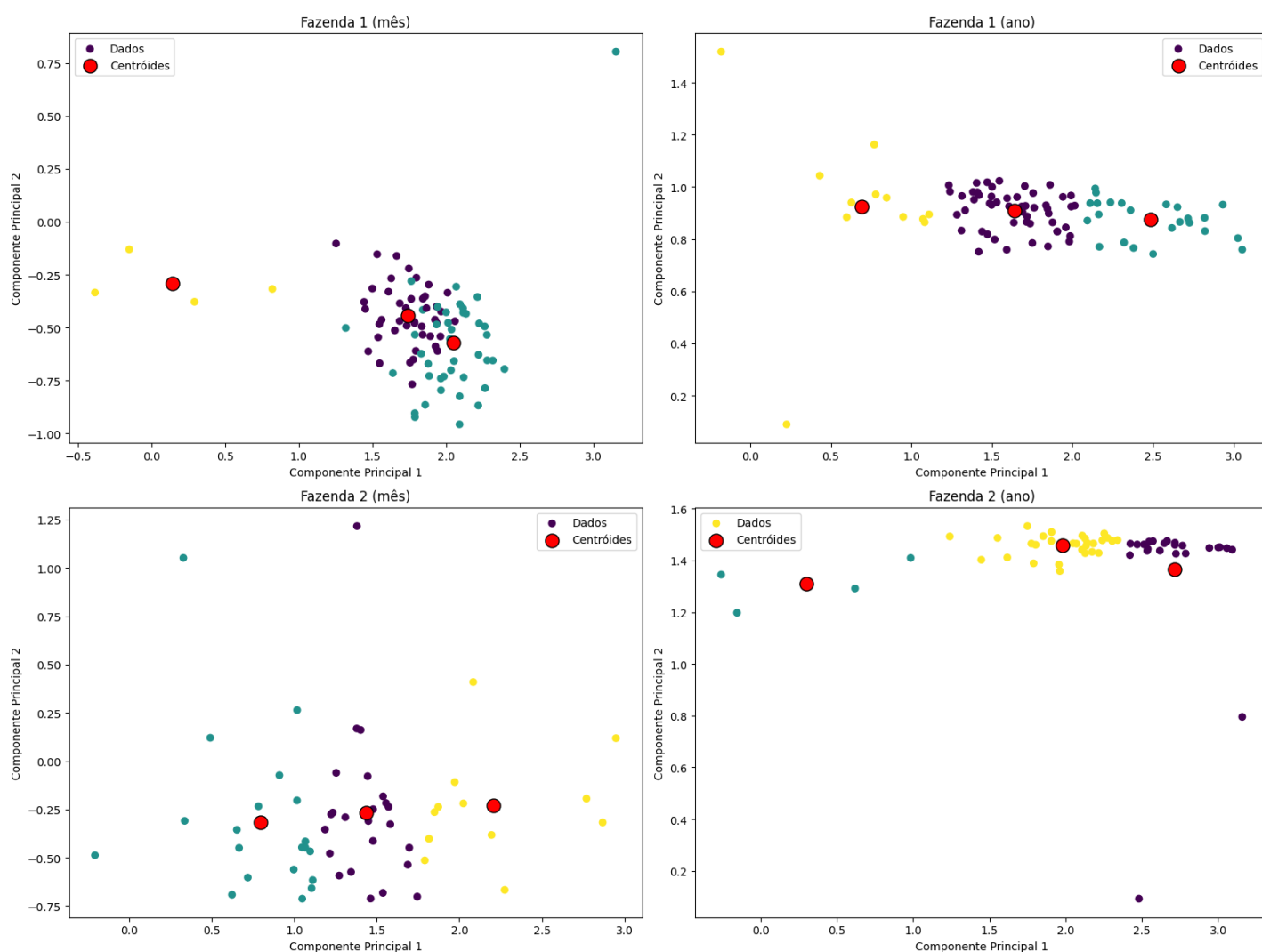
Conforme descrito no Gráfico 3, a utilização do método *Elbow* demonstrou, através da comparação da inércia (soma dos erros quadráticos), o número de *clusters* recomendados para cada *dataset*.

Observou-se uma queda acentuada de variação entre 2 e 4 *clusters*, com maior destaque em ambos os casos, aproximadamente no número 3. Assim, de acordo com os resultados exibidos no gráfico, o número ideal de *clusters* (K) para a aplicação do *K-means* em todos os casos seria de 3.

Após tal ponto, a adição de mais *clusters* resultaria em uma diminuição mínima da inércia, indicando que a segmentação adicional dos *clusters* não traria ganhos significativos em termo de ajustes para a análise. Portanto, o cotovelo (*elbow*) por volta de 3 *clusters* representa a escolha mais eficiente para balancear a precisão da análise.

A partir do resultado obtido no método *Elbow*, foi realizado o *K-means* de duas amostras de cada talhão, com  $K = 3$ . (Figura 1).

**Figura 1: Comparação dos *clusters* do *K-means* ( $K = 3$ ) realizado entre as 4 amostras de dados.**



Fonte: Elaborado pelo autor (2024)

No exemplo demonstrado na figura 1, os pontos coloridos representam os grupos de cada *cluster*, sendo cada cor responsável por um dos 3 *clusters* utilizados. Já o círculo vermelho representa o centroide (ponto central) de cada *cluster*.

Prontamente faz-se perceptível uma certa diferença entre os padrões de agrupamento *dos clusters* nas amostras mensais e anuais de cada talhão, visto que, nas amostras mensais, os *clusters* são exibidos de forma mais dispersa, sugerindo uma variação maior nos índices de vegetação em curtos períodos. Este fato pode estar relacionado a diversos fatores, como mudanças climáticas em curtos períodos ou condições meteorológicas específicas, por exemplo (Chang, 2023).

Em contraste, as amostras anuais possuem um agrupamento mais homogêneo, indicando uma menor variação em um período mais longo, demonstrando que a cultura da cana-de-açúcar apresenta padrões de desenvolvimento com maior nível de estabilidade.

Tais variações nos períodos mensais e anuais refletem a diferença do desenvolvimento da vegetação em curto e longo prazo, já que as culturas tendem a absorver as flutuações sazonais a curto prazo. São principais influências nessas variações, os fatores ambientais imediatos, como disponibilidade de água e estresse hídrico, refletindo diretamente no estado geral da saúde da planta.

Enquanto perante as variações anuais, isso é diferente, pois a longo prazo, a vegetação tende a se estabilizar mais facilmente (Araújo, 2023).

Outro ponto importante a ser destacado, foi a eficiência do uso de 3 *clusters* na separação e agrupamento das áreas de interesse, utilizando-se o PCA para o processamento inicial dos dados e características espectrais dos índices de vegetação analisados. O papel do PCA foi destacar as variações mais relevantes entre as características espectrais, simplificando os dados afim de facilitar a identificação de padrões.

A definição dos centroides, como observado na Figura 1, pode ser interpretada como uma representação das diferentes zonas de manejo presentes em cada talhão analisado nas amostras. Nesta situação, cada um dos *clusters* obtidos pelo *K-means* representariam áreas do talhão que possuem características parecidas, podendo assim, ser tratadas de forma homogênea em termos de práticas agrícolas como adubação ou irrigação.

A eficiência dessa técnica tem sido observada em diversos estudos no contexto da agricultura de precisão de várias culturas, principalmente para a identificação de

zonas de manejo homogêneas. Como exemplo, o estudo de Rodrigues (2011), concluiu que o uso de clusters foi fundamental para destacar zonas de uma plantação de café que possuíam maior necessidade de aplicação de insumos agrícolas, como fertilizantes e irrigação, reduzindo custos e otimizando a produção.

Observando o resultado do *K-means*, nota-se a presença de alguns “*outliers*”, ou pontos isolados presentes nos *clusters*, os quais se referem a observações que não seguem o padrão predominante dos agrupamentos gerados dentro do conjunto de dados.

No contexto da análise, os *outliers* podem representar áreas da plantação que possuem algumas características atípicas em relação ao restante do talhão. Isso se dá, geralmente, devido a variações severas nos índices de vegetação, que resultam de eventos localizados como propagação de pragas, falhas na irrigação, estresse hídrico severo ou condições ambientais únicas.

Os *outliers* possuem devida importância para a análise, especialmente no contexto do agronegócio, pois sua capacidade de identificação das áreas da plantação em condições anormais, ajudam os manejadores do campo a tomarem as devidas providências no local.

A identificação e o tratamento dessas áreas discrepantes da análise são fundamentais para garantir a eficiência da estratégia de manejo adequada a ser escolhida para a plantação.

O algoritmo *K-means* tem demonstrado grande potencial para detectar a ocorrência dessas anomalias de forma eficiente, devido à sua capacidade de segmentação de grandes conjuntos de dados.

Assim, a realização de uma análise adequada dos resultados obtidos pelos *clusters* permite a avaliação multivariada dos dados coletados, como por exemplo, a detecção de valores anômalos e discrepantes de forma mais robusta que métodos estatísticos tradicionais. O que a torna assim, uma ferramenta de grande valor para a análise de *datasets* em diferentes contextos (Silva et al., 2019).

## 4 CONSIDERAÇÕES FINAIS

No presente trabalho avaliou-se a utilização de índices de vegetação obtidos através de sensoriamento remoto, como ferramentas de identificação e monitoramento da produção da cana-de-açúcar. A combinação de imagens obtidas a partir do Sentinel-2 com a plataforma *Google Earth Engine*, se mostrou eficiente para o cálculo e a análise de 16 diferentes índices de vegetação, como NDVI, ARVI, SAVI, dentre outros. Esses índices são utilizados para estimar a saúde das plantas em diferentes estágios de desenvolvimento. O uso das técnicas PCA e *K-means* foi satisfatório quanto a segmentação das áreas de interesse, possibilitando a observação de padrões homogêneos e a identificação de semelhanças entre as zonas de manejo. Portanto, a partir da análise dos índices, conclui-se a demonstração da sua utilidade no auxílio da produção desta cultura, devido a caracterização das áreas de plantio, proporcionando uma abordagem tecnológica e eficiente para a realização do monitoramento contínuo e gestão estratégica de plantações da cultura de cana-de-açúcar.

Como proposta de aprimoramento deste estudo, sugere-se a integração dos dados obtidos dos índices de vegetação com modelos de aprendizado de máquina, juntamente com técnicas mais avançadas de predição da produtividade, poderá aprimorar ainda mais a capacidade de previsão e auxílio no manejo da plantação em si, permitindo abordagens ainda mais otimizadas no controle e manutenção da plantação.

## REFERÊNCIAS

ARAÚJO, A. S. A. Estudo da dinâmica da vegetação na América do Sul: mudanças de longo prazo, efeitos antrópicos e controles climáticos. 2023. Tese (Doutorado em Ciências) – Instituto de Física, Universidade de São Paulo, São Paulo, 2023. Disponível em: <https://www.teses.usp.br/teses/disponiveis/43/43134/tde-29092023-230453/>. Acesso em: 19 out. 2024.

ATAÍDE, M. V. R. Uso de drone para identificar a dominância de gramíneas invasoras e vegetação nativa em áreas de restauração do Cerrado. 2024. 87 p. Dissertação (Mestrado em Ciências Florestais) - Universidade de Brasília, Brasília, 2024. Disponível em:

[http://repositorio2.unb.br/jspui/bitstream/10482/48415/1/MarcosViniciusRezendeDeAtaide\\_DISSERT.pdf](http://repositorio2.unb.br/jspui/bitstream/10482/48415/1/MarcosViniciusRezendeDeAtaide_DISSERT.pdf). Acesso em: 21 out, 2024.

CHANG, C.-T.; CHIANG, J.-M.; DAI, J. Remote sensing of climate-vegetation dynamics and their effects on ecosystems. *Remote Sensing*, v. 15, n. 5097, p. 1-7, 2023. Disponível em: <https://doi.org/10.3390/rs15215097>. Acesso em: 21 out, 2024.

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. Acompanhamento da safra brasileira de cana-de-açúcar. Brasília, DF, v. 12, nº 2, agosto 2024. Disponível em: <https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar>. Acesso em: 22 out, 2024.

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. Acompanhamento da safra brasileira de cana-de-açúcar | v.12 – safra 2024/25, n.3 – Terceiro levantamento | novembro 2024. Disponível em: <https://www.conab.gov.br/info-agro/safras/cana/boletim-da-safra-de-cana-de-acucar>. Acesso em: 03 dez, 2024.

DIAS, E. A. Classificação da produtividade de soja e avaliação de grupos de maturação por meio de imagens multiespectrais. 2022. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa, 2022. Disponível em: <https://locus.ufv.br/bitstreams/c4674e2e-3024-4f09-8ee3-f605484ce684/download>. Acesso em: 15 out, 2024.

ENGESAT. Sentinel-2: Informações sobre a constelação de satélites Sentinel-2. Disponível em: <https://www.engesat.com.br/sentinel-2/>. Acesso em: 24 set. 2024.

ESA – European Space Agency. Copernicus Sentinel-2. Disponível em: [https://www.esa.int/Applications/Observing\\_the\\_Earth/Copernicus/Sentinel-2](https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2). Acesso em: 24 set. 2024.

GEE - Google Earth Engine. Disponível em: <https://earthengine.google.com/>. Acesso em: 16 out. 2024.

JOLLIFFE, I. T. *Principal Component Analysis: A Review and Recent Developments*.

*Philosophical Transactions of the Royal Society. A*, v. 374, n. 2065, 2016. Disponível em: <https://royalsocietypublishing.org/doi/epdf/10.1098/rsta.2015.0202>. Acesso em: 09 out, 2024.

MAIA, F. C. O. Utilização de índices de vegetação para identificação de ambientes de produção de cana-de-açúcar. Brasília: Faculdade de Agronomia e Medicina Veterinária, Universidade de Brasília, 2019, 86 p. Dissertação de Mestrado. Disponível em:

[http://www.realp.unb.br/jspu/bitstream/10482/35383/1/2019\\_FelipeCardosodeOliveiraMaia.pdf](http://www.realp.unb.br/jspu/bitstream/10482/35383/1/2019_FelipeCardosodeOliveiraMaia.pdf). Acesso em: 15 out, 2024.

MOREIRA, D. S. S. Monitoramento de Áreas de Cana-de-Açúcar por Sensoriamento Remoto. Orientador: PhD. Giancarlo Santilli. Trabalho de Conclusão de Curso – Universidade de Brasília – UnB. Faculdade UnB Gama – FGA, 2019. Disponível em: [https://bdm.unb.br/bitstream/10483/24043/1/2019\\_DanielSampaioSantosMoreira\\_tcc.pdf](https://bdm.unb.br/bitstream/10483/24043/1/2019_DanielSampaioSantosMoreira_tcc.pdf). Acesso em: 15 out, 2024

NUMPY, *Numpy: The fundamental package for scientific computing with Python*. 2024. Disponível em: <https://numpy.org/>. Acesso em: 26 set. 2024.

PANDAS, *Pandas Documentation*. 2024. Disponível em: <https://pandas.pydata.org/docs/index.html>. Acesso em: 26 set. 2024.

PEDREGOSA, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. Disponível em: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>. Acesso em: 21 out, 2024.

PEREIRA, R. L. et al. Classificação de modos de operação em sistemas de bombeio mecânico baseada em PCA e *K-Means*. In: CONGRESSO BRASILEIRO DE AUTOMAÇÃO INTELIGENTE, 2014, Natal: UFRN, 2014. Disponível em: [https://www.researchgate.net/publication/275042965\\_CLASSIFICA\\_CAO\\_DE\\_MODOS\\_DE\\_OPERACAO\\_EM\\_SISTEMAS\\_DE\\_BOMBEIO\\_MECANICO\\_BASEADA\\_EM](https://www.researchgate.net/publication/275042965_CLASSIFICA_CAO_DE_MODOS_DE_OPERACAO_EM_SISTEMAS_DE_BOMBEIO_MECANICO_BASEADA_EM)

\_PCA\_E\_K-MEANS. Acesso em: 20 out, 2024.

POLONIO, V. D. Índices de vegetação na mensuração do estoque de carbono em áreas com cana-de-açúcar. 2015. Dissertação (Mestrado em Agronomia: Energia na Agricultura) – Universidade Estadual Paulista, Faculdade de Ciências Agrônômicas, Botucatu, 2015. Disponível em: <https://repositorio.unesp.br/server/api/core/bitstreams/a8b70a49-57d2-4c06-ba70-35ca224fd752/content>. Acesso em: 16 out, 2024

PYTHON. The Python tutorial. 2024. Disponível em: <https://docs.python.org/3/tutorial/index.html>. Acesso em: 26 set. 2024.

RAMOS, N. P.; et al. Regionalização dos perfis “típicos” da produção de cana-de-açúcar para uso no RenovaBio. EMBRAPA – Meio Ambiente. Jaguariúna, SP / Maio 2024. Disponível em: <https://www.gov.br/anp/pt-br/assuntos/renovabio/arq/arquivos-estudos-relatorio-e-seminarios/relatoriofinalcanadeacucar.pdf>. Acesso em: 03 dez. 2024.

RODRIGUES, G. S. S. C., 1963. A trajetória da cana-de-açúcar no Brasil [recurso eletrônico]: perspectivas geográfica, histórica e ambiental / Gelze Serrat de Souza Campos Rodrigues, Jurandyr Luciano Sanches Ross. – Uberlândia: EDUFU, 2020. Disponível em: [https://edufu.ufu.br/sites/edufu.ufu.br/files/edufu\\_a\\_trajetoria\\_da\\_cana-de-acucar\\_no\\_brasil\\_2020\\_ficha\\_corrigida.pdf](https://edufu.ufu.br/sites/edufu.ufu.br/files/edufu_a_trajetoria_da_cana-de-acucar_no_brasil_2020_ficha_corrigida.pdf). Acesso em: 22 out. 2024.

RODRIGUES JUNIOR, F. A.; VIEIRA, L. B.; QUEIROZ, D. M.; SANTOS, N. T. Geração de zonas de manejo para cafeicultura empregando-se sensor SPAD e análise foliar. Revista Brasileira de Engenharia Agrícola e Ambiental, v. 15, n. 8, p. 778-787, 2011. Disponível em: <http://www.agriambi.com.br>. Acesso em: 20 out. 2024.

SANTANA, T.; et al. História e produtos da cana-de-açúcar: um breve relato. Ciências Agrárias: o avanço da ciência no Brasil - ISBN 978-65-5360-092-8 - Editora Científica Digital-[www.editoracientifica.org](http://www.editoracientifica.org)-Vol.3-Ano 2022. Disponível em: <https://downloads.editoracientifica.com.br/articles/220207897.pdf>. Acesso em: 22 out.

2024.

SCIKIT-LEARN, *Scikit-learn: Machine Learning in Python*. 2024. Disponível em: <https://scikit-learn.org/stable/index.html>. Acesso em: 26 set. 2024.

SHLENS, J. (2014). A tutorial on Principal Component Analysis. arXiv preprint arXiv:1404.1100. Disponível em: <https://arxiv.org/pdf/1404.1100>. Acesso em 21 out, 2024

SILVA, P. B.; FERREIRA, S.; KOKKINOGENIS, Z.; ANDRADE, M. Identificação de *outliers* em dados de acidentes de trânsito no Brasil: análise de cluster versus métodos estatísticos. *Revista de Engenharia e Tecnologia*, v. 11, n. 1, p. 103-115, 2019. Disponível em: <https://revistas.uepg.br/index.php/ret/article/view/12880/209209210918>. Acesso em: 20 out, 2024.

SPERANZA E.; et al. Importância de Índices de Vegetação para Modelos de Estimativa de Produtividade em Cana-de-açúcar. In: ANAIS DO XX SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 2023, Florianópolis. Anais eletrônicos. São José dos Campos, INPE, 2023. Disponível em: <https://proceedings.science/sbsr-2023/trabalhos/importancia-de-indices-de-vegetacao-para-modelos-de-estimativa-de-produtividade?lang=pt-br> Acesso em: 21 out. 2024.

SYAKUR, M. A. et al. *Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster*. *IOP Conf. Series: Materials Science and Engineering*, v. 336, p. 012017, 2018. Disponível em: <https://iopscience.iop.org/article/10.1088/1757-899X/336/1/012017/pdf>. Acesso em: 20 out, 2024.

WANG, C., & XU, D. (2018). An improved K-means clustering algorithm based on density. *IEEE Access*, 6, 12029-12035. Disponível em: <https://www.mdpi.com/1099-4300/23/11/1550>. Acesso em: 21 out, 2024.