

## MAPEAMENTO DA COBERTURA VEGETAL PARA ATUALIZAÇÃO CARTOGRÁFICA EM MARINGÁ/PR COM USO DE ABORDAGEM ESTATÍSTICA DO NDVI E ÁRVORE DE DECISÃO

**Américo José Marques**

Universidade Estadual de Maringá – UEM  
Departamento de Geografia, CCH-Gema, Maringá, PR, Brasil  
[aimarques74@hotmail.com](mailto:aimarques74@hotmail.com)

**Otávio Cristiano Montanher**

Universidade Estadual de Maringá – UEM  
Departamento de Geografia, CCH-Gema, Maringá, PR, Brasil  
[otaviocmontanher@yahoo.com.br](mailto:otaviocmontanher@yahoo.com.br)

### RESUMO

Este artigo tem como objetivo avaliar o uso de imagens orbitais do sensor MSI, satélite Sentinel 2A, para o mapeamento da vegetação na área correspondente à carta topográfica de Maringá (PR), folha SF-22-Y-D-II-3. O índice NDVI foi a variável espectral selecionada, e com uso do Google Earth Engine foram processadas nove imagens entre fev/2021 e jan/2022. As classificações se pautaram nas imagens-estatísticas de média e desvio-padrão, ao invés das imagens individuais. Foram testados dois métodos de classificação, um por fatiamento simples, e o outro método foi a árvore de decisão (algoritmo J48), em que foram usados 413 pontos de referência. Os resultados mostraram que o classificador por árvore de decisão apresentou resultados ligeiramente melhores do que o fatiamento simples, com índices Kappa iguais a 0,893 e 0,877, respectivamente. A árvore de decisão utilizou a média como principal variável, mas quando ela esteve entre 0,6678 e 0,7504 os pixels foram classificados também com o uso do desvio-padrão. O fatiamento simples classificou mais áreas como vegetação, enquanto a árvore de decisão classificou menos áreas. Enquanto o primeiro classificador seria mais indicado para o mapeamento de áreas de preservação, independentemente do porte da vegetação, o segundo seria mais indicado para o mapeamento de coberturas florestais.

**Palavras-chave:** GEE. Classificação. Escala. Uso do solo. Sentinel 2A.

### MAPPING OF VEGETABLE COVERAGE FOR CARTOGRAPHIC UPDATE IN MARINGÁ/PR USING THE NDVI STATISTICAL APPROACH AND DECISION TREE

### ABSTRACT

This article aims to evaluate the use of orbital images from the MSI sensor, Sentinel 2A satellite, to map the vegetation in the area corresponding to the Maringá sheet (PR), chart SF-22-Y-D-II-3. The NDVI index was the spectral variable selected, and nine images were processed using Google Earth Engine between February 2021 and January 2022. The classifications were based on the statistical images of mean and standard deviation, instead of individual images. Two classification methods were tested, one by simple slicing, and the other method was the decision tree (algorithm J48), in which 413 reference points were used. The results showed that the decision tree classifier presented slightly better results than the simple slicing, with Kappa indexes equal to 0.893 and 0.877, respectively. The decision tree used the mean as the main variable, but when it was between 0.6678 and 0.7504 the pixels were also classified using the standard deviation. Simple slicing classified more areas as vegetation, while the decision tree classified less areas. While the first classifier would be more suitable for mapping conservation areas, regardless of the size of the vegetation, the second would be more suitable for mapping forest cover.

**Keywords:** GEE, Classification. Scale. Use of the soil. Sentinel 2A.

### INTRODUÇÃO

O presente artigo faz parte de um projeto mais abrangente de atualização da carta topográfica de Maringá na escala de 1:50.000 (Folha SF-22-Y-D-II-3), editada primeiramente em 1972 através do convênio entre a Fundação IBGE e a Fundação Instituto de Terras e Cartografia do Estado do Paraná.

Tal projeto corresponde a pesquisa de cunho histórico-cartográfico que visa retornar praticamente meio século no tempo e apresentar tanto os elementos que outrora foram construídos no mundo físico real e materializados na carta, como também, apesar de representados nesse produto, não chegaram a ser efetivamente colocados em prática, ou ainda, que apesar de estarem presentes naquele momento, sofreram mudanças ao longo do tempo e, atualmente, não estão requalificados na carta.

Por isso, se pensar que os produtos cartográficos e, em específico, as cartas topográficas sejam meramente uma redução matemática e esquemática da superfície terrestre em um plano é, no mínimo, uma definição que leva em consideração somente aspectos relacionados à escala cartográfica e a simbologia dos mapas. Os produtos cartográficos são um espelho, uma representação gráfica de aspectos do mundo real mostrando uma realidade geográfica de um determinado período histórico que pode ter sido materializada ou não. Ímpar frisar também que refletem em um dado momento o avanço tecnológico da época de sua confecção (DUARTE, 2002).

Com a evolução das técnicas de fotografia analógica a partir do início do século XIX atrelada a intensificação do seu uso em aviões a partir do início do século XX para levantamentos aerofotogramétricos, a superfície terrestre começou a ser registrada numa perspectiva vertical e os mapas passaram a representar a superfície terrestre com maior precisão. No Brasil, na década de 1920, o trabalho sistemático de mapeamento na escala 1:1.000.000, conforme as especificações estabelecidas para a Carta do mundo ao Milionésimo (CIM) – e de onde se desdobra as demais cartas desse mapeamento sistemático, incluindo aqui as cartas na escala de 1:50.000 – resultou na edição da Carta do Brasil em um momento em que a fotogrametria dava seus primeiros passos em nosso país (ARCHELA, 2007).

Com advento do computador e, com ele, o desenvolvimento da interface gráfica em monitores de alta resolução, juntamente com a tomada das imagens por satélites a partir do início da década de 1970, permitiu-se que os mapas saíssem do mundo analógico e passassem para o mundo digital, estando hoje popularizados e presentes nas telas dos computadores, dos receptores GPS, dos celulares, dos tablets. Além disso, o armazenamento e processamento de dados na nuvem vem possibilitando o acesso e processamento de forma mais ágil aos produtos cartográficos nas mais variadas plataformas digitais. Um exemplo disso é a disponibilização das imagens de satélite e o seu processamento, gerando produtos praticamente prontos para a análise.

Uma das plataformas existentes atualmente é o Google Earth Engine (GEE) que oferece sua estrutura computacional permitindo aos usuários executar análises geoespaciais nos servidores da Google (GORELICK et al. 2017). A plataforma combina um catálogo de vários petabytes de imagens de satélite e conjuntos de dados geoespaciais com recursos de análise em escala planetária onde pesquisadores podem buscar os mais variados produtos para detectar mudanças, mapear tendências e quantificar diferenças na superfície da Terra (GEE, 2022).

O GEE tem sido utilizado em aplicações diversas, mas uma de suas grandes potencialidades é no mapeamento de culturas agrícolas e da cobertura vegetal. Em áreas com uso agrícola intenso, como a região de Maringá/PR, em que grande parte da vegetação original foi removida, é importante o monitoramento das áreas remanescentes para a fiscalização e preservação ambiental. Adicionalmente, como o GEE permite que o mapeamento seja feito rapidamente, o processo pode ser feito em intervalos de tempo reduzidos, pré-determinados, para observação de possíveis mudanças temporais, como, por exemplo o aumento ou a diminuição da cobertura vegetal.

De acordo com Tucker (1979), utilizando-se as diferentes propriedades e características das distintas coberturas vegetais existentes, podem-se distingui-las de acordo com os distintos comportamentos espectrais registrados nas imagens de satélite. Ponzoni (2001) relata que a resposta espectral da vegetação apresenta uma mistura complexa das características do dossel vegetativo, dos efeitos ambientais, da sombra, da umidade do solo, entre outras, que apresentam propriedades de reflexão própria, que permite a medição para a modulação das propriedades de reflexão da vegetação.

Outro aspecto importante é que a disponibilidade de várias bandas espectrais permite a exploração mais eficiente das características espectrais dos alvos, em conjunto com os modos de operação dos diversos sensores existentes (SILVA et al., 2009). Diversos tipos de agrupamento entre bandas vêm sendo desenvolvidos como as razões simples de normalizações e distâncias no espaço cartesiano e suas variantes, em geral realçam o componente espectral da vegetação e se correlacionam com parâmetros biofísicos da vegetação, como o Índice de Área Foliar - IAF, Índice de Vegetação Ajustado ao Solo – SAVI, Índice de Vegetação por Diferença Normalizada - NDVI (EIPHANIO et al., 1996).

Cabe destacar ainda que esses Índices de Vegetação (IVs) podem ser calculados automaticamente pelo GEE.

O objetivo desse artigo é avaliar a possibilidade do uso de imagens orbitais para a atualização cartográfica da vegetação presente na área correspondente à carta topográfica de Maringá, folha SF-22-Y-D-II-3. Para tanto, foram utilizados os recursos e produtos disponibilizados pela referida plataforma digital, imagens do sensor MSI, satélite Sentinel 2A, e diferentes abordagens de classificação, tendo como base o índice de vegetação NDVI. Em termos de metodologia e discussão dos resultados, os focos deste trabalho estão, respectivamente, sobre o uso de imagens-estatísticas, ao invés de imagens singulares, e sobre a interpretação das diferenças entre os resultados das duas classificações aplicadas. A avaliação dos resultados das classificações foi feita com base em pontos de controle aleatoriamente distribuídos ao longo da área de estudo, nos quais se observou a presença da classe vegetação.

## **METODOLOGIA**

### ***Área de estudo***

A área de estudo corresponde à carta topográfica de Maringá, escala 1:50.000, situada entre as latitudes de 23° 15' S e 23° 30' S e entre as longitudes de 51° 45' W e 52° 00' W (Figura 1). Até meados da década de 1970 essa área possuía uma diversidade de alvos como campos, áreas urbanas, corpos hídricos, vegetação e cultura permanente que, na sua maioria, era representada por plantação de café.

Devido à geada de 1975 (conhecida como geada negra) e a fatores econômicos mundiais traçados na época, essa cultura foi sendo substituída gradativamente por culturas agrícolas temporárias tais como milho, trigo, soja, entre outras. Na região, os ciclos predominantes são da plantação da soja na primeira safra (safra verão), com o plantio iniciando no mês de outubro e colheita a partir de janeiro e a segunda safra com a plantação de milho (também conhecida como “milho safrinha”), que tem início a partir de janeiro (FRANCO et al., 2013).

A vegetação original que recobria a área e hoje apresenta-se preservada em áreas de matas das mais variadas formas (matas ciliares, parques urbanos, reservas florestais rurais) faz parte da unidade biogeográfica pertencente à formação do conjunto da mata Atlântica, do domínio da floresta Estacional Semidecidual Submontana, abaixo de 500m e Montana, acima de 500m (IBGE, 1990), uma vez que a área ocupada pela carta apresenta altitudes que variam de 380m a aproximadamente 700m.

As três áreas urbanas principais da carta topográfica (Maringá, Sarandi e Marialva) se materializaram no divisor d'água entre duas bacias hidrográficas: a bacia hidrográfica do Pirapó e a bacia hidrográfica do Ivaí. A implantação nos divisores d'água das áreas urbanas, bem como da rodovia e ferrovia que interligam essas áreas, seguiu a sistemática definida pelo planejamento executado na área, a partir da década de 1940, pela Companhia Melhoramento Norte do Paraná.

Nesses divisores d'água o relevo é de planalto, apresentando patamares e mesetas com colinas suaves, características adequadas para a implantação de núcleos urbanos e que, neste caso, foi de primordial importância na determinação das características urbanísticas implantadas nestas cidades (MARQUES; BARROS, 2014).

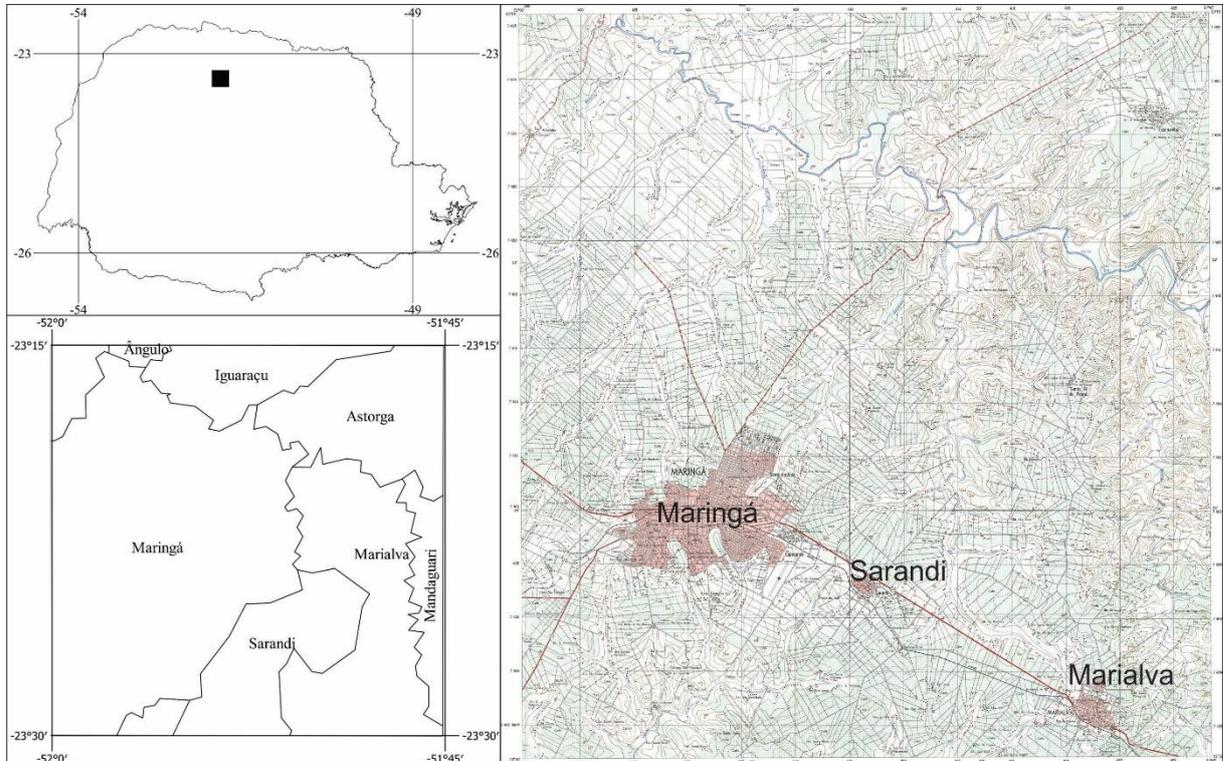
O limite entre as referidas bacias hidrográficas perfaz então padrões distintos com relação ao sistema de drenagem. A bacia do rio Pirapó, ao norte, é representada pelo curso d'água principal dessa bacia (rio Pirapó) e tem destaque na carta no que tange à hidrografia. Seus afluentes têm o sentido predominante sul-norte a partir do interflúvio onde se localizam os principais sítios urbanos e a rodovia e ferrovia que interligam essas localidades. Os cursos d'água pertencentes à bacia do rio Ivaí têm o sentido contrário, partindo do referido interflúvio.

### ***Escala cartográfica***

O Instrumento Multiespectral, ou MSI, presente no satélite Sentinel 2A possui 13 bandas espectrais, cobrindo as faixas espectrais do visível, infravermelho próximo e infravermelho de ondas curtas. Em termos de resolução espacial, o MSI possui quatro bandas com 10 m, seis com 20 m e três com resolução de 60 m. As bandas espectrais utilizadas nesta pesquisa correspondem às regiões do vermelho e do infravermelho próximo, ambas com 10 m de resolução espacial. A resolução radiométrica

do MSI é de 12 bits (quantização em 4096 níveis de brilho) e a resolução temporal é de cinco dias, considerando a frequência de revisita a partir da constelação dos dois satélites (ESA, 2022).

Figura 1 - Carta Topográfica de Maringá (Folha SF-22-Y-D-II-3): 1972.



Fonte - IBGE, 1972. Organização: os autores, 2022.

Para verificar se as imagens MSI/Sentinel 2A fornecem uma base de escala compatível com a carta topográfica em escala 1:50.000, foram realizadas duas avaliações. A primeira, de caráter bibliográfico, tendo como referência as diretrizes presentes em DSG (2016), que apresentam as resoluções espaciais (máxima, típica e mínima) e escalas cartográficas correspondentes para uma série de escalas entre 1:1.000 e 1:250.000. No caso da escala 1:50.000, as resoluções variam entre 2 (máxima) e 10,5 m (mínima), de modo que as bandas utilizadas neste estudo, de 10 m, condizem com o intervalo.

Outra forma de avaliação utilizou o método descrito em Boggione et al., (2009), que parte da dimensão relativa à resolução espacial efetiva, ou EIFOV, do sensor. Enquanto a resolução espacial nominal se refere à distância entre o centro de dois pixels vizinhos, por exemplo, a resolução espacial efetiva se refere à área realmente levantada, normalmente maior do que aquela circunscrita ao pixel. Tomando-se como base a banda 4 (região espectral do vermelho) do sensor MSI, seu EIFOV é de 22,06 m (RADOUX et al., 2016). Conhecido o EIFOV, a Equação 1 apresenta o denominador da escala, pressupondo erros de desenho (ou digitalização) de até 0,5 mm, seguindo um PEC (Padrão de Exatidão Cartográfica) classe A.

$$E = 1/2 \times res \times 1000 \quad (1)$$

E = escala (adimensional); res = resolução espacial efetiva (m).

Considerando o EIFOV mencionado, a maior escala cartográfica indicada seria de 1:11.030, de modo que atende a escala da carta topográfica, de 1:50.000.

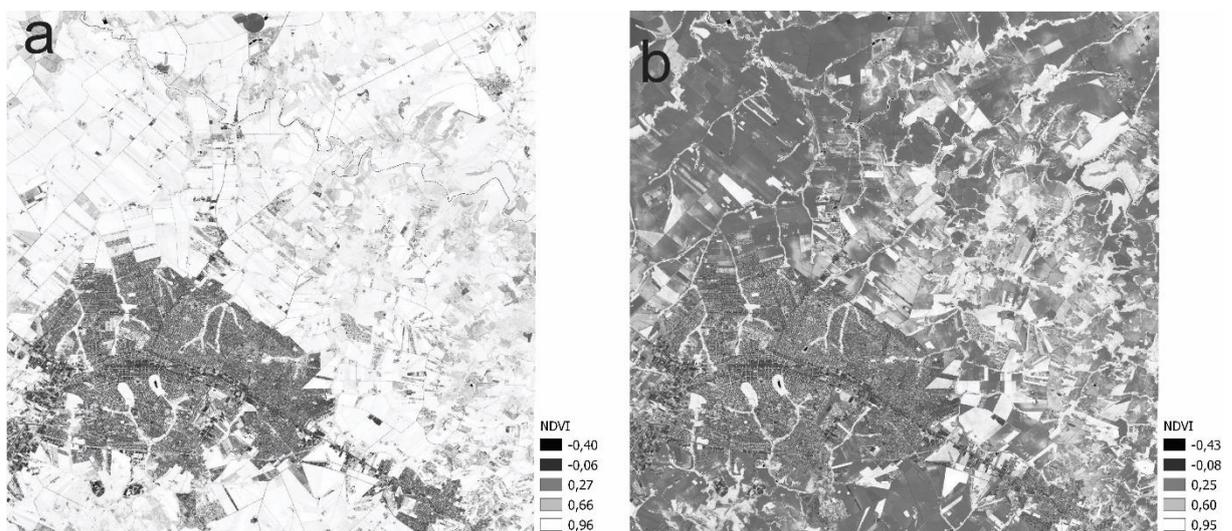
### Classificação

O processo de classificação constitui o foco deste trabalho, ao testar duas abordagens, uma por fatiamento (ou limiarização - ALVES et al., 1998) e outra com uso de árvore de decisão (FRIEDL e BRODLEY, 1997; RUIZ et al., 2014), em que o índice utilizado nos dois casos foi o NDVI (*Normalized Difference Vegetation Index*). O fatiamento é uma ferramenta simples, muitas vezes fornecendo resultados satisfatórios. No entanto, o fatiamento poderia gerar resultados inconsistentes se fosse usado o mesmo limiar em diferentes épocas do ano (devido ao menor vigor da vegetação durante o inverno) e há o problema de que os cultivos agrícolas podem apresentar valores de NDVI semelhantes ou até maiores do que as florestas, no auge do seu desenvolvimento. Estes dois casos foram constatados, por exemplo, nos valores de NDVI na imagem de 06 de fevereiro de 2021, quando a cultura da primeira safra estava no período da colheita (Figura 2a) e no mês de inverno, no período de entressafra (Figura 2b).

Para regiões com extensas áreas cultivadas, os erros de inclusão da classe “vegetação florestal” seriam grandes caso um limiar fosse aplicado em uma imagem obtida em momentos de pico de desenvolvimento dos cultivos agrícolas (soja: dezembro-janeiro, milho: abril-junho, considerando a área de estudo). Dessa forma, avaliou-se a opção de se trabalhar com as estatísticas básicas (média e desvio-padrão) dos valores de NDVI, considerando uma série de imagens.

Foram então utilizadas todas as imagens livres de cobertura de nuvens para a área de estudo, no período de fev/2021 a jan/2022, resultando em nove imagens, distribuídas ao longo do ano. Seleccionadas as imagens, calculou-se o NDVI e foram geradas imagens para a média e o desvio-padrão do NDVI. A hipótese que baseia o uso do NDVI médio é que, mesmo com os picos de desenvolvimento da agricultura, quando se toma uma série de imagens no decorrer do ano, os baixos valores de NDVI relativos ao solo exposto tendem a diminuir a média nas áreas de agricultura, quando comparadas com as áreas de vegetação natural. Por isso, valores acima de certo limiar associado ao NDVI médio são classificados como “vegetação florestal”, enquanto os valores abaixo do limiar são classificados como “outros”.

Figura 2 - Imagens MSI com aplicação do índice NDVI referentes à área abarcada pela carta topográfica de Maringá: a) 06/02/2021; b) 21/07/2021.



Organização - Os autores, 2022.

Já a hipótese que baseia o uso do desvio-padrão do NDVI é que as áreas de vegetação florestal tendem a possuir valores mais constantes no decorrer do ano em relação às áreas de agricultura. Por outro lado, em áreas urbanas o NDVI tenderia a ser mais constante do que a vegetação florestal, que devido ao seu ciclo fenológico, apresenta NDVI maior em algumas épocas e menor em outras, com a queda das folhas. Assim, testou-se o fatiamento simples a partir da imagem-média, considerando vegetação

as áreas com NDVI maior do que 0,66 (BACKES, 2010; LIRA et al, 2011), e as duas estatísticas foram utilizadas para a classificação via árvore de decisão.

A avaliação da acurácia das classificações seguiu práticas tradicionalmente utilizadas para esta finalidade, descritas em Lillesand et al. (2015). Inicialmente foram definidos pontos aleatórios sobre a extensão da área de estudo, observando em cada um deles a classe real (referência) e o resultado da classificação. Os resultados desse levantamento foram resumidos em uma matriz de confusão, e foram calculadas as acurácias do produtor e consumidor, erros de omissão e inclusão, acurácia global e o coeficiente Kappa. Para determinação da quantidade de pontos amostrais foi utilizada uma abordagem estatística clássica para proporções, visto que um dos principais resultados, a acurácia global, é uma proporção (equação 2):

$$n \geq \frac{N}{\frac{e^2(N-1)}{z_{\alpha}^2 \hat{p}(1-\hat{p})} + 1} \quad (2)$$

n = tamanho da amostra; N = tamanho da população; e = erro; z = valor associado com a distribuição normal padrão;  $\hat{p}$  = proporção previamente estimada.

Para aplicação da equação 2 primeiramente calculou-se o tamanho da população, em que cada pixel na imagem classificada foi considerado como um elemento de uma população finita. A área de estudo possui 25 x 25 km, ou seja, 625 km<sup>2</sup>, e cada pixel possui 10 x 10 m (100 m<sup>2</sup>). Convertendo-se ambas para a mesma unidade de área, observou-se que a região é coberta por 6.250.000 pixels. A equação 2 também necessita a definição de uma proporção dos erros previamente estimada, em que nos levantamentos exploratórios foi definida em 4%. Adicionalmente, foi adotado o nível de significância de 5% e uma margem de erro de  $\pm 1,89\%$ , obtendo como resultado uma quantidade amostral de 413.

Esses pontos aleatórios foram utilizados de duas formas: primeiramente, para a classificação por fatiamento da média se observou o resultado da classificação e a classe real. Essa base foi utilizada para o cálculo da acurácia global, bem como outras métricas de acurácia para a classificação por fatiamento. Posteriormente, para cada ponto aleatório onde já se conhecia a classe real, se obteve o valor médio e o desvio-padrão do NDVI. Essa base de dados (já convertida em dados tabelados) foi utilizada no software WEKA 3.8 (FRANK et al., 2016), em que o algoritmo J48 foi aplicado para geração de uma árvore de decisão. O J48 é a implementação em linguagem JAVA de um algoritmo de indução de árvores de decisão amplamente difundido, o C4.5, criado na década de 1990. Esse algoritmo foi muito difundido por suas propriedades de pouca restrição quanto às características dos atributos (trabalha com dados qualitativos, contínuos e discretos) e não exige alguma distribuição de probabilidade para os dados (VIEIRA et al., 2018). Uma profundidade máxima da árvore foi definida em três, para evitar o super ajustamento aos dados.

## RESULTADOS E DISCUSSÃO

Os resultados da avaliação da classificação por meio do fatiamento simples estão resumidos nas Tabelas 1 e 2. Obteve-se uma acurácia global de 95,88% (17 pontos incorretamente classificados em 413 observados), desta forma, o erro global é de  $4,12 \pm 1,89\%$ . Com relação ao índice Kappa, pode-se afirmar que o resultado da classificação aplicada, de 0,877, está bem acima do esperado para uma classificação aleatória, estando no intervalo de uma classificação quase perfeita (LANDIS; KOCH, 1977).

Avaliando-se as acurácias do produtor e consumidor, bem como os respectivos erros de omissão e inclusão para as duas classes, foram observadas acurácias altas, bem como baixos erros. No entanto, os erros associados com uma classificação chamam atenção em relação aos demais, de 92 pontos classificados como "Vegetação", na realidade 12 pertenciam à classe "Outros", resultando em um erro de inclusão de 13,04%. Dessa forma, o erro mais frequente deste mapeamento refere-se às áreas que possuem outros tipos de coberturas, mas que foram classificadas como vegetação arbórea. Os outros erros (comissão e inclusão) estão abaixo de 6%, sendo considerados baixos para a proposta deste trabalho. A árvore de decisão ajustada por meio do algoritmo J48 está apresentada na Figura 3.

Tabela 1 - Matriz de confusão: classificação por fatiamento simples.

Classificação	Referência		
	Vegetação	Outros	Total
Vegetação	80	12	92
Outros	5	316	321
Total	85	328	413

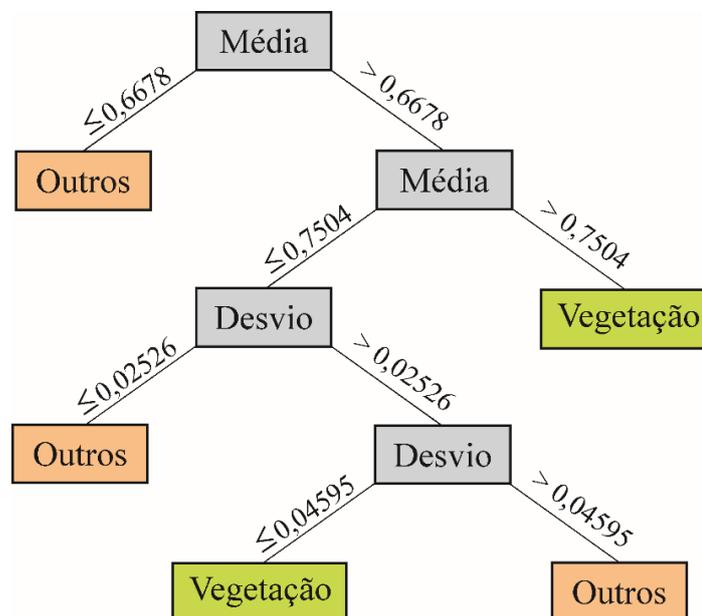
Organização - Os autores, 2022.

Tabela 2 - Estatísticas relativas à acurácia da classificação: fatiamento simples.

Estatística	Vegetação	Outros
Acurácia do produtor	94,12	96,34
Acurácia do consumidor	86,96	98,44
Erro de omissão	5,88	3,66
Erro de inclusão	13,04	1,56
Índice Kappa	0,877	

Elaboração - Os autores, 2022.

Figura 3 - Árvore de decisão para classificação da vegetação florestal. As variáveis “Média” e “Desvio” se referem à média e ao desvio-padrão do NDVI para uma série de nove imagens adquiridas entre fev/2021 e jan/2022.



Organização - Os autores, 2022.

O primeiro aspecto que chama atenção na árvore de decisão ajustada aos 413 pontos observados (Figura 3) é que o principal fator que distingue as duas classes é o NDVI médio, de modo que, abaixo de 0,6678 o pixel é classificado como não pertencendo à classe vegetação. Esse valor é muito próximo do apresentado pela bibliografia (BACKES, 2010; LIRA et al, 2011), que forneceu base para a classificação por fatiamento, de 0,66. Esse fato faz com que as duas classificações tenham resultados semelhantes, em um primeiro momento, com diferenças específicas ao intervalo de NDVI entre 0,6678 e 0,7504. Isso porque valores de NDVI médio acima de 0,7504 foram classificados diretamente como

“Vegetação”. A árvore de decisão utiliza o desvio-padrão como uma variável auxiliar para classificação de pixels com NDVI médio entre 0,6678 e 0,7504. Para esses pixels, os que possuem menor desvio-padrão (abaixo de 0,02526), foram classificados como “Outros”, assim como os que possuem desvio-padrão acima de 0,04595. No intervalo entre esses dois limiares, os pixels foram classificados como “Vegetação”.

Entende-se que o desvio-padrão baixo está associado com alvos espectralmente invariantes, ou que variam muito pouco ao longo do tempo. Alguns alvos localizados em áreas urbanas normalmente possuem essa característica, como o asfalto, concreto, telhados, entre outras coberturas. A vegetação florestal, por outro lado, tende a possuir variações sazonais de NDVI, entre as épocas de maior e menor vigor. Por outro lado, um desvio-padrão muito alto está relacionado com alvos que se modificam bastante ao longo do tempo, tendo como principal exemplo na área de estudo a agricultura de ciclo curto, como a soja e o milho. Dessa maneira, em termos de comparação entre todos os alvos da cena, a classe “Vegetação” possui valores de desvio-padrão intermediários.

Os resultados da avaliação da classificação por meio de árvore de decisão estão resumidos nas Tabelas 3 e 4. Obteve-se uma acurácia global de 96,61% (14 pontos incorretamente classificados em 413 observados), desta forma, o erro global é de  $3,39 \pm 1,89\%$ . Com relação ao índice Kappa, o resultado de 0,893 foi ligeiramente superior ao calculado para a classificação por fatiamento.

Tabela 3 - Matriz de confusão: classificação por árvore de decisão.

Classificação	Referência		
	Vegetação	Outros	Total
Vegetação	75	4	79
Outros	10	324	334
Total	85	328	413

Elaboração - Os autores, 2022.

Tabela 4 - Estatísticas relativas à acurácia da classificação: árvore de decisão.

Estatística	Vegetação	Outros
Acurácia do produtor	94,94	97,01
Acurácia do consumidor	88,24	98,78
Erro de omissão	5,06	2,99
Erro de inclusão	11,76	1,22
Índice Kappa	0,893	

Elaboração - Os autores, 2022.

Tomando-se como base as estatísticas apresentadas nas Tabelas 1 a 4, o classificador por árvore de decisão retornou resultados ligeiramente superiores, mas muito semelhantes. Ao se analisar os resultados contínuos das classificações, realmente se observou similaridade entre os resultados. No entanto, o classificador por fatiamento gerou mais áreas de vegetação, enquanto a árvore de decisão foi mais restritiva. Isso pode ser observado nas Figuras 4 e 5.

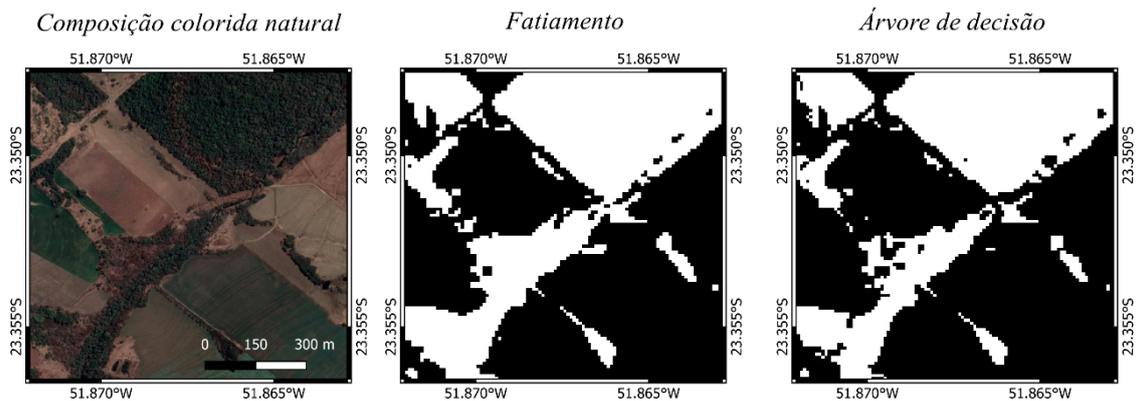
Em parte, a menor seleção de áreas para a classe “Vegetação” pela Árvore de decisão é devido ao primeiro limiar de corte desse classificador ser um pouco superior ao valor de 0,66 do fatiamento. No entanto, avaliando-se os valores de média e desvio-padrão do NDVI em algumas áreas amostrais, observou-se que o principal fator que determinou a classe “Outros” em áreas onde na classificação por fatiamento era “Vegetação”, foi o desvio-padrão. Em alguns casos, onde há vegetação mista entre gramíneas e arbustivas, encontrou-se um desvio-padrão maior do que em áreas com vegetação florestal. Isso fez com que essas áreas fossem classificadas como “Outros” pela árvore de decisão. Essa distinção se deve possivelmente pela diferença de comportamento desses tipos de vegetação no período mais seco e frio do ano, em que principalmente as gramíneas perdem bastante vigor.

Outros casos em que a árvore de decisão não classificou como “Vegetação” onde essa classe foi determinada na classificação por fatiamento, se encontram nas bordas das áreas de floresta. As bordas foram mais incluídas como “Vegetação” no fatiamento, e determinadas como “Outros” na árvore por decisão. Isso possivelmente ocorreu devido ao fato de que entre os imageamentos os pixels que constituem bordas não possuem sempre a mesma proporção entre vegetação e outras classes, mesmo

com o georreferenciamento adequado entre as imagens (ressalta-se que o georreferenciamento não foi aplicado, pois as imagens já se encontram devidamente georreferenciadas no banco de dados do GEE). Dessa forma, nessas áreas de borda o desvio-padrão foi maior do que nas áreas internas dos polígonos de áreas florestais. Exemplos desses casos podem ser observados nas Figuras 4 e 5.

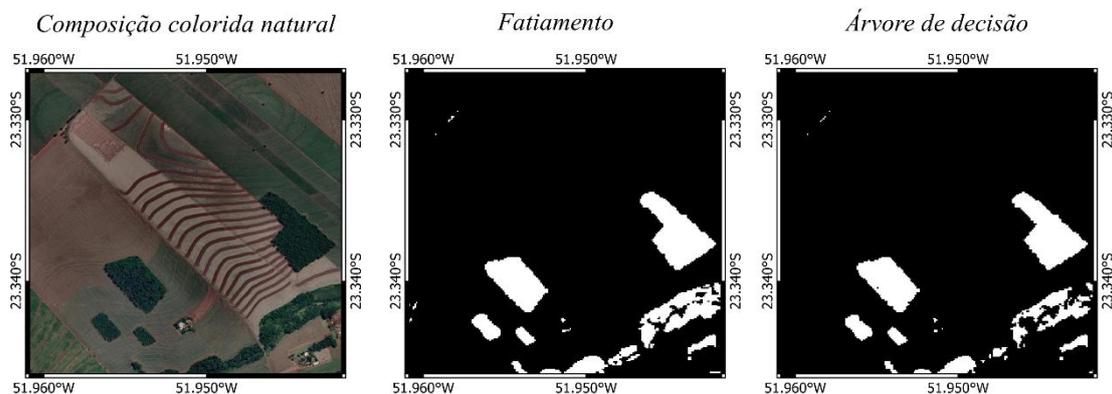
Um motivo adicional para que as bordas das áreas de vegetação tenham sido menos incluídas na classe “Vegetação” pela árvore por decisão pode ser o sombreamento. Dada a resolução espacial das imagens utilizadas (10 m), a sombra das árvores sobre os alvos adjacentes seria passível de captação. Visto que o sombreamento muda de tamanho e de direção durante o ano, devido às variações dos ângulos zenital e azimutal de incidência da radiação solar, isso faria com que nessas áreas as reflectâncias variassem mais, sem que se alterassem as propriedades dos alvos. Dessa forma, espera-se que os valores de NDVI nessas áreas variem mais do que nas “áreas-core” de vegetação, ou dos alvos adjacentes. Uma variação maior do NDVI implica em um desvio-padrão maior, de modo que por meio da árvore de decisão esses locais sejam classificados como “Outros”. Ressalta-se que o sombreamento entre alvos é reconhecidamente um fenômeno que diminui a precisão de classificadores, principalmente em imagens de alta resolução espacial. Exemplos de pesquisas que lidam com os problemas derivados do sombreamento em classificações de vegetação podem ser consultados em Song et al. (2015) e Lopatin et al. (2019).

Figura 4 - Comparação entre as classificações (branco equivale à classe “Vegetação” e o preto se refere à classe “Outros”. Região norte do município de Maringá/PR.



Organização - Os autores, 2022.

Figura 5 - Comparação entre as classificações (branco equivale à classe “Vegetação” e o preto se refere à classe “Outros”. Região central do município de Maringá/PR.



Organização - Os autores, 2022.

Os dois casos em que a árvore de decisão incluiu menos pixels como “Vegetação”, em relação ao fatiamento, fazem com que esse classificador retorne como resultado uma classificação mais voltada às áreas com cobertura florestal, excluindo vegetações de menor porte e as bordas, onde há mistura

espectral com outros alvos. Por outro lado, o classificador por fatiamento abrange esses casos, fornecendo áreas maiores e mais contínuas de vegetação. Para o mapeamento de limites de áreas de preservação e áreas onde não há cultivo agrícola, mesmo que a vegetação seja de menor porte, o classificador por fatiamento seria mais adequado. Por outro lado, se o objetivo é identificar áreas com cobertura florestal, o classificador por árvore de decisão apresentado seria o mais indicado.

É importante mencionar que as duas técnicas de classificação avaliadas não diferenciam remanescentes florestais de áreas com silvicultura. Na área de estudo as áreas que possuem esse uso do solo possuem pequenas extensões, nas quais predomina o plantio de *Eucalyptus*. Considerando essas áreas de plantio, caso as árvores ainda estejam pouco desenvolvidas, com boa porção de solo exposto, provavelmente ambas as técnicas retornarão a classe “Outros” para a silvicultura, devido ao NDVI médio provavelmente estar abaixo dos limiares de classificação (0,66 e 0,6678). Por outro lado, com o desenvolvimento das árvores e a consequente cobertura do solo pela biomassa, essas áreas seriam classificadas em conjunto com a classe “Vegetação”.

## CONCLUSÕES

Este trabalho permite obter as seguintes conclusões:

- i) O uso de imagens-estatísticas de média e de desvio-padrão de NDVI trouxe vantagens no processo de classificação da vegetação em relação ao uso de imagens individuais, principalmente devido à dinâmica temporal dos alvos agrícolas. O uso do GEE como uma base georreferenciada de dados orbitais já pré-processados (em unidade de reflectância de superfície), tornou simples e ágil a obtenção das imagens-estatísticas;
- ii) O limiar consultado na bibliografia de NDVI para classificação da vegetação em outras áreas (0,66), foi muito próximo ao principal limiar de diferenciação empiricamente ajustado por meio do algoritmo J48 na área de estudo deste trabalho, o que apoia os bons resultados com o uso deste limiar;
- iii) Considerando o classificador por árvore de decisão, que apresentou resultados ligeiramente melhores do que o fatiamento simples, a estatística média foi a principal variável. Considerando o NDVI médio, valores abaixo de 0,6678 foram classificados diretamente como “Outros”, e acima de 0,7504 foram classificados diretamente como “Vegetação”. Valores médios entre esses dois limiares foram classificados também com o uso do desvio-padrão do NDVI;
- iv) O fatiamento simples classificou mais áreas como vegetação, enquanto a árvore de decisão classificou menos áreas. Em parte, essa diferença se baseia nos diferentes limiares de corte (0,66 e 0,6678), mas também no desvio-padrão, que sendo mais baixo (pixels com NDVI pouco variável), ou mais alto (pixels com NDVI bastante variáveis), o local não é classificado como vegetação.
- v) Observando-se a espacialização dos resultados das duas classificações utilizadas, nota-se que o fatiamento simples fornece áreas mais contínuas de vegetação, enquanto a árvore de decisão fornece áreas mais fragmentadas. Enquanto o primeiro classificador seria mais indicado para o mapeamento de áreas de preservação, independentemente do porte da vegetação, o segundo seria mais indicado para o mapeamento de coberturas florestais.

Considerando os resultados dessa pesquisa, o próximo passo é seu uso para a atualização cartográfica da área abrangida pela carta topográfica de Maringá, considerando o tema vegetação. Outras áreas com contextos semelhantes de uso e cobertura do solo podem utilizar as abordagens descritas neste artigo para uma rápida classificação da cobertura florestal.

## REFERÊNCIAS

- ALVES, D. S.; PEREIRA, J. L. G.; SOUSA, C. L.; SOARES, J. V.; MOREIRA, J. C.; ORTIZ, J. O.; SHIMABUKURO, Y. E.; YAMAGUCHI, F. Análise comparativa de técnicas de classificação de imagens do sensor Landsat/TM para caracterização de áreas desflorestadas. *In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO*, 9., 1998, Santos. **Anais** [...] Santos: INPE, 1998. p. 1409-1419.
- ARCHELA, R. S. **Evolução histórica da cartografia no Brasil: Instituições, formação profissional e técnicas cartográficas**. Revista Brasileira de Cartografia No 59/03, Dezembro, 2007, pp. 213-223. <https://doi.org/10.14393/rbcv59n3-44889>

BACKES, K. S. Variações do índice de vegetação por diferença normalizada (NDVI) do sensor modis associadas a variáveis climáticas para o estado do Rio Grande do Sul. Mestrado em Engenharia Florestal, UFSM, 2010.

BOGGIONE, G. A.; SILVA, M. V. A.; CARVALHO JUNIOR, N. R.; TELES, T. L.; NAZARENO, N. R. X. **Definição da escala em imagens de sensoriamento remoto: uma abordagem alternativa**. XIV Simpósio Brasileiro de Sensoriamento Remoto: Instituto Nacional de Pesquisas Espaciais, 2009, p. 1739-1746. Disponível em: <http://mar.te.sid.inpe.br/col/dpi.inpe.br/sbsr%4080/2008/11.18.03.28/doc/1739-1746.pdf>. Acesso em 15 mar. 2022.

DIRETORIA DE SERVIÇO GEOGRÁFICO – DSG. **Norma para especificação técnica para produtos de conjunto de dados Geoespaciais (et-pcdg)**. MINISTÉRIO DA DEFESA. EXÉRCITO BRASILEIRO. DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA. 2. ed. 2016.

DUARTE, P. A. **Fundamentos de Cartografia**. 2ª. Ed. Florianópolis: Ed. Da UFSC, 2002.

EPIPHANIO, J.C.N.; GLERIAN, J.; FORMAGIO, A.R.; RUDORFF, B.F.T. Índices de vegetação no sensoriamento remoto da cultura de feijão. Pesquisa Agropecuária brasileira, v.31, n.6, p.445-454, 1996.

EUROPEAN SPACE AGENCY - ESA. **Sentinel-2 MSI UserGuide**. Disponível em: <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi>. Acesso em 20 set. 2022.

FRANCO, A. A. N.; MARQUES, O. J.; VIDIGAL FILHO, P. S. **Sistemas de produção de milho safrinha no Paraná**. XII Seminário Nacional Milho Safrinha. Embrapa: Dourados-MS, 2013.

FRANK, E; HALL, M. A.; WITTEN, I. H. The WEKA Workbench. Online Appendix for "**Data Mining: Practical Machine Learning Tools and Techniques**", Morgan Kaufmann, Fourth Edition, 2016.

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

GEE – Google Earth Engine. Disponível em: <https://earthengine.google.com/>. Acesso em 08 mar. 2022.

GORELICK, N.; HANCHER, M.; DIXON, M.; ILYUSHCHENKO, S.; THAU, D.; MOORE, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, v. 202, p. 18-27, 2017. DOI. <https://doi.org/10.1016/j.rse.2017.06.031>.

IBGE. Carta Topográfica de Maringá. 1:50.000. 1972. Disponível em: <https://biblioteca.ibge.gov.br/>. Acesso em: 12 fev. 2022.

\_\_\_\_\_. **Geografia do Brasil: Região Sul**. Rio de Janeiro: Fundação Instituto Brasileira de Geografia e Estatística, Dir. Geociências. 1990.

LANDIS, J.; KOCH, G. **The measurement of observer agreement for categorical data, Washington, USA**. Biometrics, v. 33, n. 1, p. 159-174, 1977. <https://doi.org/10.2307/2529310>

LILLESAND, T. M.; KIEFER, R. W.; CHIPMAN, J. W. **Remote sensing and image interpretation**. 7 ed. Nova York: John Wiley & Sons, 2015.

LIRA, M. V. et al. Variação sazonal da cobertura vegetal em áreas do submédio São Francisco a partir de dados NDVI e imagens TM-LandSat 5. **Enciclopédia Biosfera**, v.7, n. 12, 2011.

LOPATIN, J.; DOLOS, K.; KATTENBORN, T.; FASSNACHT, F. E. How canopy shadow affects invasive plant species classification in high spatial resolution remote sensing. **Remote Sensing in Ecology and Conservation**, v. 5, n. 4, p. 302 – 317, 2019. <https://doi.org/10.1002/rse2.109>

MARQUES, A. J.; BARROS, M. V. F. **Quantificação das áreas verdes urbanas do município de Maringá, PR**. Anais do VII Congresso Brasileiro de Geógrafos. Vitória/ES, 2014.

MENESES, P. R. **Princípios de Sensoriamento Remoto**. In: MENESES, P. R.; ALMEIDA, T. Introdução ao processamento de imagens de sensoriamento remoto. Brasília, 2012.

NIGRO, G. T.; FERREIRA, M. E. M. C. **Avaliação da recomposição florestal das áreas de fundos de vale na bacia do Córrego Cleópatra/Moscados, Maringá - PR**. Boletim de geografia, v. 39, p. 144-163, e58252, 10 nov. 2021. DOI. <https://doi.org/10.4025/bolgeogr.v39.a2021.e58252>

PONZONI, F. J. **Comportamento espectral da vegetação. In. Sensoriamento Remoto: Reflectância dos alvos naturais.** Brasília: UnB, EMBRAPA, p.157-199, 2001.

RADOUX, J.; CHOMÉ, G.; JACQUES, D.C.; WALDNER, F.; BELLEMANS, N.; MATTON, N.; LAMARCHE, C.; D'ANDRIMONT, R.; DEFOURNY, P. Sentinel-2's Potential for Sub-Pixel Landscape Feature Detection. *Remote Sensing*, v. 8, n. 6, p. 488-510. 2016. DOI.<https://doi.org/10.3390/rs8060488>

RUIZ, L. F. C.; CATEN, A. T.; DALMOLIN, R.S. D. **Árvore de decisão e a densidade mínima de amostras no mapeamento da cobertura da terra.** *Ciência Rural*, v.44, n.6, p.1001-1007, 2014. <https://doi.org/10.1590/S0103-84782014000600008>

SILVA, C. A. V. da. **Uso do Sensoriamento Remoto através de Índices de Vegetação NDVI, SAVI e IAF na microrregião de Itamaracá – PE.** In: Anasi XIV Simpósio brasileiro de Sensoriamento Remoto, Natal, Brasil, 25 – 30 abril 2009, INPE, p. 3079-3085.

SONG, W.; MU, X.; YAN, G.; HUANG, S. Extracting the Green Fractional Vegetation Cover from Digital Images Using a Shadow-Resistant Algorithm (SHAR-LABFVC). *Remote Sensing*, v. 7, p. 10425-10443, 2015. <https://doi.org/10.3390/rs70810425>

TUCKER, C. J. **Red and photographic infrared linear combinations for monitoring vegetation.** *Remote Sens. Environ.* **8**, 127–150, 1979. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)

VIEIRA, E. M. A.; NEVES, N. T. A. T.; OLIVEIRA, A. C. C.; MORAES, R. M.; NASCIMENTO, J. A. Avaliação da performance do algoritmo J48 para construção de modelos baseados em árvores de decisão. *Revista Brasileira de Computação Aplicada*. v. 10, n.2, p. 80–90, 2018. <https://doi.org/10.5335/rbca.v10i2.8078>

---

Recebido em: 27/04/2022

Aceito para publicação em: 14/10/2022