

## APLICAÇÃO DE MACHINE LEARNING NA PREVISÃO DA TAXA DIÁRIA DE QUEIMADAS NO CERRADO BRASILEIRO

Marcilene Soares do Nascimento <sup>1</sup>; Tamires Partélli Correia <sup>2</sup>; Jonathan da Rocha Miranda <sup>3</sup>

1 Marcilene Soares do Nascimento, Bolsista (CNPq), Bacharelado em Agronomia, IFMG Campus São João Evangelista, São João Evangelista - MG; [marcilenesoares98@outlook.com](mailto:marcilenesoares98@outlook.com)

2 Tamires Partélli Correia, Professora, IF Sudeste MG Campus de São João Del-Rei, São João Del-Rei – MG; [tamirespartelli@gmail.com](mailto:tamirespartelli@gmail.com)

3 Orientador: Jonathan da Rocha Miranda, IFMG Campus São João Evangelista; [Jonathan.rocha@ifmg.edu.br](mailto:Jonathan.rocha@ifmg.edu.br)

### RESUMO

Este trabalho propõem a aplicar a aprendizagem de máquina para a previsão de incêndios florestais, um problema ambiental premente, especialmente na Estação Ecológica de Uruçuí-Una, um ecossistema do Cerrado Brasileiro que tem experimentado um aumento na frequência de incêndios nas últimas décadas. O objetivo foi desenvolver um modelo de aprendizado de máquina que pudesse prever a ocorrência de incêndios florestais baseado em variáveis ambientais. Para atingir este objetivo, coletamos e analisamos dados de satélite do índice de vegetação (NDVI), temperatura, e déficit hídrico durante um período de vinte anos. Adotamos uma abordagem de aprendizado supervisionado, testando três algoritmos de aprendizado de máquina: Random Forest, SVM e Redes Neurais. O modelo Random Forest apresentou a melhor performance, identificando com precisão uma relação significativa entre a redução do NDVI, o aumento da temperatura e do déficit hídrico, e a incidência de incêndios na região. Concluímos que a aprendizagem de máquina tem um potencial significativo para prever incêndios florestais e deve ser considerada para implementação em estratégias de gestão ambiental e prevenção de incêndios. Entretanto, a generalização desses modelos para outros biomas necessita de mais pesquisas.

Palavras chave: Modelagem de Incêndios; Variáveis Climáticas; Sensoriamento Remoto

### INTRODUÇÃO:

O Cerrado, o segundo maior bioma brasileiro, cobre uma imensa região de 2.036.448 km<sup>2</sup>, sendo a savana com maior biodiversidade mundial. Porém, apenas 8,21% de sua área estão protegidos legalmente por Unidades de Conservação (UCs), com 2,85% de unidades de proteção integral e 5,36% de unidades de uso sustentável (LAURA *et al.*, 2021).

A Estação Ecológica Uruçuí-Una (ESECUNNA) enfrenta dificuldades para monitorar incêndios florestais devido à extensão territorial e à insuficiência de força de trabalho, mesmo com apoio do IBAMA (MEDEIROS; CUNHA, 2006). Além disso, a falta de dados confiáveis, infraestrutura insuficiente, escassez de patrulhamento, falta de apoio comunitário e atrasos preventivos intensificam o problema (PEREIRA *et al.*, 2004).

A detecção e o monitoramento eficientes de incêndios são cruciais para um controle efetivo e redução de custos (PRUDENTE; ROSA, 2010). O sensoriamento remoto, utilizando dados espaciais de satélites, torna-se uma ferramenta valiosa nesse contexto (WANG *et al.*, 2018). Entretanto, a atual metodologia de detecção de fontes de calor do INPE tem se mostrado inadequada para situações críticas (ABDO *et al.*, 2021).

Uma estratégia que eleva o alerta dos pontos de focos de calor poderia agilizar a resposta aos incêndios, direcionando equipes para as áreas de maior urgência. Além disso, a caracterização da vegetação suscetível à queima e a identificação das variáveis climáticas impactam diretamente na propagação dos incêndios (LAMCHIN *et al.*, 2020; NASCIMENTO *et al.*, 2012).

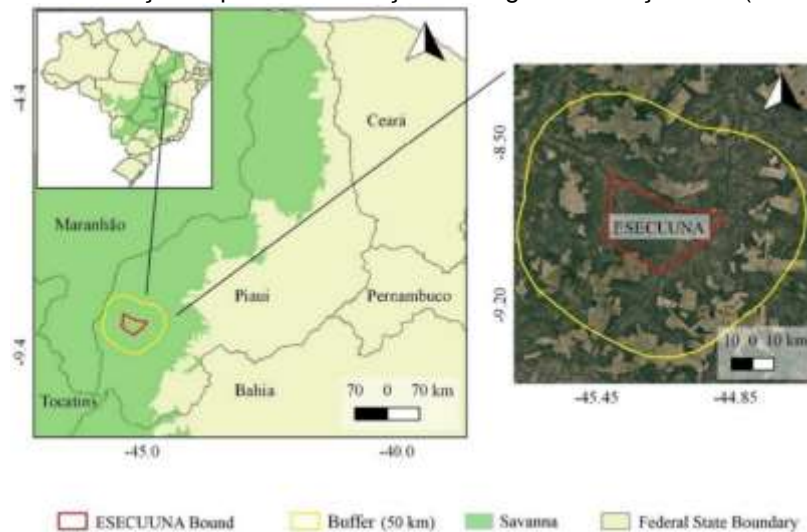
Contudo, o grande volume de variáveis pode dificultar a identificação de padrões relacionados à velocidade de queima e aos focos de incêndio. Nesse contexto, as ferramentas de aprendizado de máquina podem ser uma solução poderosa para a detecção de padrões em grandes conjuntos de dados (DE SOUSA *et al.*, 2022).

Este estudo propõe que, utilizando dados climáticos diários e índices de vegetação derivados de dados orbitais, é possível desenvolver um modelo que estime a velocidade de queima para um determinado foco de incêndio, permitindo a priorização eficaz das ações de combate. O objetivo é modelar uma base de dados históricos de queimadas e estimar a velocidade de propagação utilizando algoritmos de inteligência artificial.

## METODOLOGIA:

O estudo foi realizado na Estação Ecológica de Uruçuí-Una (ESECUUNA), uma vasta reserva ecológica situada no quadrante do Sudoeste Piauiense, no cerrado. As coordenadas específicas do local são: NE 44°57'49" W e 8°53'02" S, SE 45°11'37" W e 9°06'34" S, NO 45°23'02" W e 8°39'26" S, e SO 45°26'19" W e 8°54'24" S. Este recanto ecológico, que ocupa uma extensão de 135.000 hectares, está inserido na sub-região dos altos platôs piauienses, conforme ilustrado na Figura 1.

Figura. 1 . Localização espacial da Estação Ecológica de Uruçuí-Una (ESECUUNA).



Fonte: Miranda et al (2022)

Devido à vulnerabilidade da ESECUUNA a incêndios florestais, originados tanto internamente quanto em suas adjacências, foi estabelecida uma zona-tampão ou buffer de 50 km ao redor da unidade para o monitoramento intensivo. Esta estratégia amplia significativamente a área de estudo, elevando-a para uma totalidade de 17233,84 km<sup>2</sup>.

Foram empregados dados climáticos relativos à superfície terrestre abrangendo o período de 2001 a 2019. A seleção dos satélites utilizados foi orientada pelo critério de proporcionar as imagens de produtos processados com a maior resolução espacial e temporal disponível.

No âmbito das imagens que denotam áreas queimadas (MCD64A1), foi empregado um critério específico de filtragem, o qual foi orientado pela qualidade do pixel, indicada pela banda *ba\_qa*. Pixels apresentando um valor radiométrico igual a 1 foram considerados como portadores de informação confiável. A identificação das áreas queimadas foi efetuada por meio da seleção de pixels que apresentaram valor igual a 1 na imagem *ba\_qa* e superior a 1 na banda *Burn Area*. Este último valor corresponde à data do evento de queima, conforme relatado por (RODRIGUES *et al.*, 2019). A partir desta análise, a imagem foi categorizada em duas classes: áreas queimadas, denotadas por um valor de pixel igual a 1, e áreas não queimadas, denotadas por um valor de pixel igual a 0.

Na sequência da identificação e demarcação das áreas queimadas, procedeu-se à segmentação desta categoria. Este procedimento permitiu a vetorização em polígonos representativos das cicatrizes causadas pelas queimadas. Tal metodologia foi aplicada a cada imagem que representava um evento de queima ocorrido entre os anos de 2001 a 2019. Mediante este vasto conjunto de vetores, foi realizada a combinação de todos os polígonos, objetivando a reorganização dos dados de uma perspectiva mensal para uma perspectiva anual.

A seleção dos focos de queima validados demandou uma transformação dos pixels correspondentes para pontos representativos dos focos de queima validados. Em cada um desses pontos, foi estabelecida uma área de interesse (buffer) com um raio de 3 km para fins de análise.

Esse raio se baseou na estimativa dos raios médios calculados a partir da área média das cicatrizes de queima. A referência para essa determinação veio do trabalho de Miranda *et al.* (2022), que efetuaram um levantamento das áreas queimadas na Estação Ecológica de Uruçuí-Una (ESECUUNA) durante o mesmo período abrangido por este estudo.

Assim, a aplicação do buffer de 3 km ao redor de cada ponto representando um foco de queima validado possibilita a captação da variação espacial das variáveis ambientais e climáticas que possivelmente estão influenciando a ocorrência e o comportamento do fogo naquele local específico.

Deste modo, para o foco de queima detectado, foi inserido a informação para modelagem da TDQ por meio de técnicas de machine learning. A variável preditiva foi a TDQ e as preditoras serão as variáveis NDVI, precipitação, evapotranspiração, déficit hídrico, velocidade do vento e temperatura.

Foi utilizado a biblioteca de aprendizado de máquina Scikit-learn (PEDREGOSA *et al.*, 2011), codificada na linguagem Python, para implementar os algoritmos de classificação SVM, Random Forest e Multilayer Perceptron.

Os algoritmos implementados serão submetidos à validação cruzada estratificada em 10 dobras, um procedimento que envolve múltiplas iterações dos algoritmos, em que os conjuntos de dados de treinamento e teste são reorganizados a cada rodada, garantindo uma representação equitativa de cada classe amostral. Os resultados finais da classificação foram determinados pela média dos resultados de todas as iterações, conforme recomendado por (HALL *et al.*, 2009).

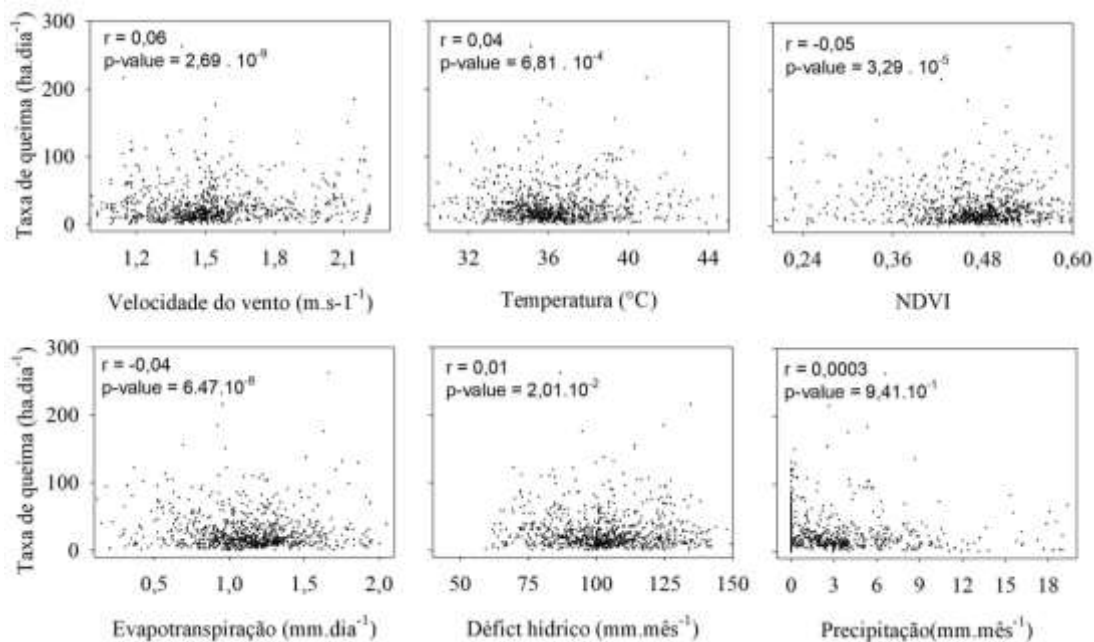
Para avaliar os resultados obtidos pelos modelos ajustados foram adotadas as medidas de Erro Absoluto Médio (MAE) e Erro Quadrático Médio (MSE) que permitem verificar a diferença entre os valores observados e estimados, possibilitando quantificar a acurácia obtida em cada ajuste.

## RESULTADOS E DISCUSSÕES:

Ao investigar a correlação direta das variáveis analisadas com a taxa diária de queima, a velocidade do vento e a temperatura mostraram uma relação proporcional à taxa diária de queima, com coeficientes de correlação de 0,06 e 0,04, respectivamente (Figura 2). Isso indica que quanto maiores os valores de temperatura e velocidade do vento, maior a taxa de queima.

Por outro lado, o NDVI e a evapotranspiração apresentaram uma correlação inversamente proporcional à taxa de queima, com coeficientes de  $-0,05$  e  $-0,04$ , respectivamente. Isso sugere que a taxa de queima tende a ser menor quando esses índices são maiores. As variáveis déficit hídrico e precipitação mostraram uma correlação significativa com a taxa diária de queima, apesar da tendência observada anteriormente.

Figura 2. Relação da taxa de queima e as variáveis climática



Estudos conduzidos por Santos e Ranieri (2013) observaram uma correlação relevante entre os fatores climáticos e a ocorrência de incêndios na ESECUUNA, sendo a umidade relativa do ar e a precipitação os elementos de maior influência. Seguindo a mesma linha de análise, Durigan e Ratter (2016) constataram que temperaturas superiores a 30°C, aliadas a ventos fortes e umidade do ar inferior a 30%, compõem o perfil climático das áreas que sofreram incêndios de grande escala na região de savana. Lamchin *et al.*

(2020) realizaram uma análise usando a tendência de Mann-Kendall para a vegetação entre os anos de 1982 e 2014 e observaram uma correlação inversa entre a temperatura e o vigor vegetativo. Por outro lado, a precipitação demonstrou uma correlação positiva e direta. Estes dados reafirmam a importância dos elementos climáticos na prevenção e controle de incêndios florestais. A correlação percebida entre a redução do NDVI e o aumento do déficit hídrico sugere que as condições de seca podem ter sido um fator contribuinte para a maior vulnerabilidade da vegetação ao fogo. Este cenário foi corroborado por Miranda *et al.* (2022) que avaliaram a tendência negativa de Mann-Kendall e constataram que a repetição de incêndios no mesmo local tem provocado uma degradação progressiva do ambiente.

O Random Forest obteve o melhor desempenho em relação a Rede neural e o SVM, este dado sugere que o modelo Random Forest foi capaz de prever a taxa diária de queima com uma precisão relativamente alta e com um erro reduzido (Figura 3, Tabela 1). O SVM, por outro lado, teve o pior desempenho neste sentido o modelo teve dificuldade em prever a taxa diária de queima com precisão. Do mesmo modo a Rede Neural também apresentou um desempenho inferior, embora tenha se saído melhor que o SVM, ainda ficou muito atrás do Random Forest em termos de precisão e confiabilidade.

Figura 3. Desempenho dos algoritmos de machine learning na estimativa da taxa de queima em função das variáveis climáticas

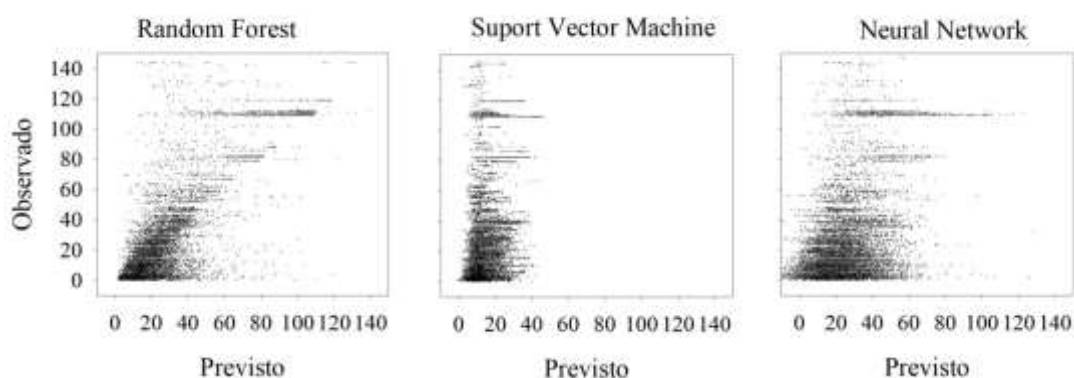


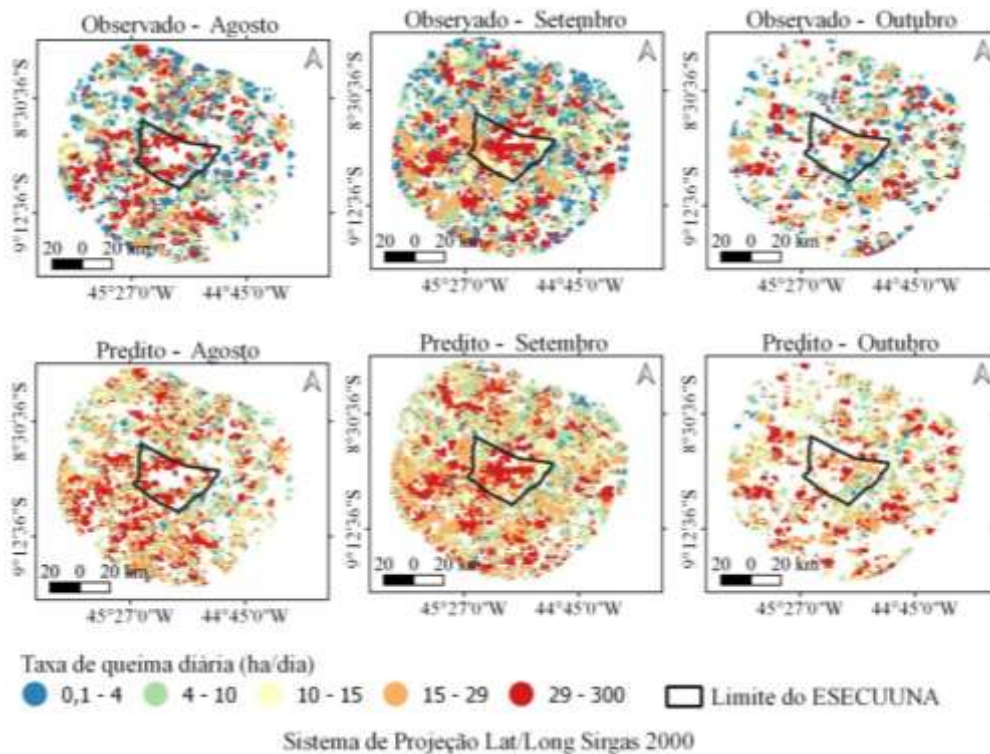
Tabela 1. Métricas de avaliação dos algoritmos de machine learning na estimativa da taxa de queima em função das variáveis climáticas

Regressores	MAE	MSAE	R <sup>2</sup>
Random Forest	5,533	306,182	0,758
SVM	8,194	1319,182	0,041
Rede Neural	10,884	1097,075	0,134

No interior da ESECUUNA, constata-se que grande parte dos focos de queima se enquadra na classe de maior taxa de queima, superior a 29 hectares por dia (Figura 4). Tal fenômeno pode ser atribuído à dificuldade de ação rápida no combate aos focos de incêndio, tendo em vista que a região é predominantemente atendida por brigadistas. Esta situação permite que o fogo se alastre, atingindo assim uma taxa de queima expressivamente alta.

Por outro lado, na região externa à ESECUUNA, observa-se a predominância de classes de menor taxa de queima. Isto provavelmente decorre do fato de que as queimadas em propriedades rurais podem ser rapidamente controladas pelos próprios proprietários. Essa prontidão na resposta e ação localizada favorece a manutenção da taxa de queima em níveis mais baixos.

Figura 4. Espacialização da taxa de queima para os focos de calor estimado pelo Random Forest e comparação com os dados observado



O algoritmo Random Forest foi particularmente eficaz na previsão da taxa diária de queima, levando em consideração as variáveis abordadas. O destaque desse modelo pode ser atribuído à sua notável capacidade de capturar interações complexas entre as variáveis e à sua robustez frente ao ruído e aos outliers nos dados, conforme destacado por Rodriguez Galiano *et al.* (2012) em suas pesquisas em classificação de uso do solo.

Contrastando com o Random Forest, os algoritmos SVM e Redes Neurais fundamentam-se em princípios de otimização de funções matemáticas. Embora Karimi *et al.* (2008) tenham relatado resultados promissores com o SVM na estimativa de parâmetros biofísicos das culturas, esses algoritmos podem encontrar desafios quando os dados não se adaptam bem a funções matemáticas regressivas ou apresentam interações complexas entre variáveis.

Os resultados obtidos neste estudo também enfatizam a importância de monitorar constantemente as condições climáticas e superficiais, uma vez que esses dados alimentam o modelo e permitem uma previsão precisa da taxa de queima (PAUSAS, 2004). Tais informações são fundamentais para o planejamento de ações de prevenção, bem como para a rápida resposta a incêndios emergentes.

Portanto, estes resultados contribuem não só para a ciência do comportamento do fogo, mas também para a gestão prática e as estratégias de combate a incêndios, facilitando a proteção eficaz dos ecossistemas e das comunidades humanas dos efeitos devastadores dos incêndios florestais (MCCAFFREY, 2004).

## CONCLUSÃO:

O potencial do uso de técnicas de machine learning na previsão de incêndios florestais, e sugere sua implementação como uma ferramenta estratégica na gestão de recursos naturais e no combate a incêndios. Tal ferramenta pode facilitar a previsão e detecção de incêndios, permitindo ações de prevenção e combate mais eficientes e bem direcionadas, o que contribuirá para a proteção mais eficaz dos ecossistemas e comunidades humanas.

Também ressalta-se a importância da consideração dos dados climáticos e da vegetação na modelagem e previsão de incêndios, indicando que ações de manejo e prevenção de incêndios devem considerar esses fatores.

Contudo, é importante frisar que, apesar dos resultados promissores, mais estudos são necessários para refinar esses modelos de previsão e expandi-los para outras regiões e biomas. Além disso, o machine learning é apenas uma das várias ferramentas que podem ser usadas para enfrentar o problema dos

incêndios florestais, que é complexo e multidimensional, requerendo uma abordagem multifacetada e interdisciplinar para a sua solução efetiva.

#### REFERÊNCIAS BIBLIOGRÁFICAS:

ABDO, R.; FERRO-FAMIL, L.; BOUTET, F.; ALLAIN-BAILHACHE, S. Analysis of the double-bounce interaction between a random volume and an underlying ground, using a controlled high-resolution poltomosar experiment. **Remote Sensing**, [s. l.], v. 13, n. 4, p. 1–20, 2021.

DE SOUSA, M. B.; FARIAS, R. H. L.; VILELA, O. C.; DE SOUZA BARBOSA, E. M.; PEREIRA, A. C.; JATOBA, E. B.; NETO, A. C.; DE MELO FILHO, J. B. ESTIMATIVA DE IRRADIÂNCIA MÉDIA HORÁRIA A PARTIR DE IMAGENS DO SATÉLITE GOES-16 BASEADA NO MODELO DE TARPLEY DE 1979. In: CONGRESSO BRASILEIRO DE ENERGIA SOLAR-CBENS, 2022, [s. l.], . **Anais [...]**. [S. l.: s. n.], 2022. p. 1–10.

DOS SANTOS, M. R. R.; RANIERI, V. E. L. Criteria for analyzing environmental zoning as an instrument in land use and spatial planning. **Ambiente e Sociedade**, [s. l.], v. 16, n. 4, p. 43–60, 2013.

DURIGAN, G.; RATTER, J. A. The need for a consistent fire policy for Cerrado conservation. **Journal of Applied Ecology**, [s. l.], v. 53, n. 1, p. 11–15, 2016.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The WEKA data mining software. **ACM SIGKDD Explorations Newsletter**, New York, NY, USA. v. 11, n. 1, p. 10–18, 2009. Disponível em: <http://doi.acm.org/10.1145/1656274.1656278>.

KARIMI, Y.; PRASHER, S. O.; MADANI, A.; KIM, S. Application of support vector machine technology for the estimation of crop biophysical parameters using aerial hyperspectral observations. **Canadian Biosystems Engineering / Le Genie des biosystems au Canada**, [s. l.], v. 50, n. 7, p. 13–20, 2008.

LAMCHIN, M.; WANG, S. W.; LIM, C. H.; OCHIR, A.; PAVEL, U.; GEBRU, B. M.; CHOI, Y.; JEON, S. W.; LEE, W. K. Understanding global spatio-temporal trends and the relationship between vegetation greenness and climate factors by land cover during 1982–2014. **Global Ecology and Conservation**, [s. l.], v. 24, p. e01299, 2020.

LAURA, V. A. de; NETO, M. M. G.; ALMEIDA, R. G. De; ALVES, F. V. O eucalipto em sistemas de integração lavoura-pecuária-floresta (ILPF) no Bioma Pampa. **Embrapa**, [s. l.], p. 1110–1132, 2021.

MCCAFFREY, S. Thinking of wildfire as a natural hazard. **Society and Natural Resources**, [s. l.], v. 17, n. 6, p. 509–516, 2004.

MEDEIROS, F. C.; CUNHA, A. M. C. Plano Operativo de Prevenção e Combate aos Incêndios Florestais da Estação Ecológica de Uruçui-Una/PI. **Ministério do Meio Ambiente-MMA**. 15p, [s. l.], 2006.

MIRANDA, J. da R.; SILVA, R. G. da; JUVANHOL, R. S. Forest fire action on vegetation from the perspective of trend analysis in future climate change scenarios for a Brazilian savanna region. **Ecological Engineering**, [s. l.], v. 175, p. 106488, 2022.

NASCIMENTO, C. R.; MIRANDA, J. da R.; SALLES, G. de B. Gis Applied To the Study of Temporal Recovery of Burned Areas in the Municipality of Pai Pedro Located in Region North of the State of Minas Gerais. **Journal of Hyperspectral Remote Sensing**, [s. l.], v. 2, n. 5, p. 98–109, 2012.

PAUSAS, J. G. Changes in fire and climate in the eastern Iberian Peninsula (Mediterranean Basin). **Climatic Change**, [s. l.], v. 63, n. 3, p. 337–350, 2004.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, É. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, [s. l.], v. 12, n. Oct, p. 2825–2830, 2011.

PEREIRA, C. A.; FIEDLER, N. C.; DE MEDEIROS, M. B. Análise De Ações De Prevenção E Combate Aos Incêndios Florestais Em Unidades De Conservação Do Cerrado. **Floresta**, [s. l.], v. 34, n. 2, 2004.

PRUDENTE, T. D.; ROSA, R. Detecção De Incêndios Florestais No Parque Nacional Da Chapada Dos Veadeiros E Área De Entorno. **Caminhos de Geografia**, [s. l.], v. 11, n. 35, p. 209–221, 2010.



ISSN 2558-6052

RODRIGUES, J. A.; LIBONATI, R.; PEREIRA, A. A.; NOGUEIRA, J. M. P.; SANTOS, F. L. M.; PERES, L. F.; SANTA ROSA, A.; SCHROEDER, W.; PEREIRA, J. M. C.; GIGLIO, L.; TRIGO, I. F.; SETZER, A. W. How well do global burned area products represent fire patterns in the Brazilian Savannas biome? An accuracy assessment of the MCD64 collections. **International Journal of Applied Earth Observation and Geoinformation**, [s. l.], v. 78, p. 318–331, 2019.

RODRIGUEZ-GALIANO, V. F.; GHIMIRE, B.; ROGAN, J.; CHICA-OLMO, M.; RIGOL-SANCHEZ, J. P. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal of Photogrammetry and Remote Sensing**, [s. l.], v. 67, n. 1, p. 93–104, 2012.

WANG, L.; YU, D.; LIU, Z.; YANG, Y.; ZHANG, J.; HAN, J.; MAO, Z. Study on NDVI changes in Weihe Watershed based on CA–Markov model. **Geological Journal**, [s. l.], v. 53, p. 435–441, 2018.