

Avaliação de variáveis agronômicas e imagens do satélite Sentinel-2 para estimativa de produtividade da cana-de-açúcar a partir do algoritmo Random Forest

Rafaella Pironato Amaro^a, Ana Cláudia dos Santos Luciano^b

Departamento de Engenharia de Biosistemas - USP

^arafaella.amaro@usp.br

^banaluciano@usp.br

RESUMO: O objetivo deste trabalho, foi avaliar a importância de variáveis agronômicas e de imagens do satélite Sentinel-2 para estimativa da produtividade de cana-de-açúcar, utilizando o algoritmo Random Forest. Foram obtidos dados agronômicos referentes à variedade, estágio de corte, tipo de solo e relevo, além de dados provenientes das imagens de satélite referentes ao NDVI médio, máximo e o desvio padrão do NDVI de cada talhão. Foram criados dois modelos empíricos considerando: i) Variáveis agronômicas, ii) Variáveis agronômicas e imagens Sentinel-2. O modelo estimativo de produtividade apresentou R^2 igual a 0,64 e 0,83, RMSE de 10,17 e 7,0 ton/ha, para os modelos i e ii, respectivamente. A avaliação da importância das variáveis indicou que a variável estágio de corte foi a mais importante, seguida das variáveis variedade e NDVI médio do talhão. A combinação de variáveis agronômicas e de imagens de satélite trouxe melhorias na estimativa da produtividade de cana-de-açúcar.

PALAVRAS-CHAVE: Índice de vegetação, Modelo empírico, NDVI; Aprendizado de máquinas

ABSTRACT: The objective of this project was to evaluate the importance of agronomic variables and Sentinel-2 satellite images to estimate sugarcane yield using the Random Forest algorithm. We used agronomic data referring to the variety, cutting stage, soil type and relief, in addition to data from satellite images referring to the average, maximum NDVI and the standard deviation of the NDVI of each field. Two empirical models were created considering: i) Agronomic variables, ii) Agronomic variables and Sentinel-2 images. The model to estimate sugarcane yield showed R^2 equal to 0.64 and 0.83, RMSE of 10.17 and 7.0 ton/ha for models i and ii, respectively. The evaluation of the importance of the variables indicated that the variable cutting stage was the most important, followed by the variable variety and average NDVI of the field. The combination of agronomic variables and satellite images brought improvements to estimate sugarcane productivity.

KEYWORDS: Vegetation index, Empirical model, NDVI, Machine learning

1. Introdução

A cana-de-açúcar, que ocupa aproximadamente 8,7 milhões de hectares no Brasil [1], é uma cultura global com importância não apenas para a produção de açúcar, mas também é considerada uma das grandes alternativas para o setor de biocombustíveis devido ao grande potencial na produção de etanol e seus respectivos subprodutos [1, 2]. No Brasil, a cultura tem grande importância social e econômica para o país, pois gera empregos e exportação para o exterior. Nos últimos 20 anos, a produção e área plantada de cana-de-açúcar no Brasil duplicaram, apresentando um aumento de 1,8% em relação a 2019/2020

[1]. Apesar das oscilações climáticas, a produção de cana-de-açúcar atingiu cerca de 753 milhões de toneladas em 10,1 milhões de hectares na safra 2021 [3]. A região Centro-Sul representou o maior eixo produtivo do país, sendo responsável por 92% do total da cana-de-açúcar produzida [1]. No entanto, a cana-de-açúcar, por ser uma planta semi-perene, sofre com as influências do clima, que são oscilantes durante o ciclo de crescimento da cultura. Tais oscilações, ocorrem especialmente no aspecto das precipitações e na sua regularidade de distribuição, ao contrário do que acontece com as culturas anuais que são influenciadas pelo clima em períodos limitados [1]. Portanto, a cana-de-açúcar pode sofrer com eventuais déficits hídricos em algumas locali-

dades, impactando nos potenciais produtivos que podem variar em função da interação entre a época do ano em que ocorrem e, a fase do ciclo fenológico da cultura [3,4]. Neste sentido, o clima é um fator fundamental para o planejamento agrícola da produção de cana-de-açúcar. Os efeitos combinados da variabilidade natural do clima, as condições de população crescente, a perda de solo e as mudanças climáticas, exigem métodos que forneçam uma avaliação oportuna e precisa do crescimento e da produção das culturas e que contribuam para o aumento da sustentabilidade da produção de alimentos agrícolas [5, 6]. Neste contexto, a necessidade de um planejamento estratégico adequado, a previsão de safras de determinada cultura e o conhecimento de sua distribuição no espaço geográfico é de extrema importância para o planejamento do setor sucroenergético brasileiro. Além disso, o monitoramento da produção de cana-de-açúcar auxilia na criação de políticas públicas e segurança alimentar, o que impacta diretamente na melhoria da precisão e robustez dos sistemas de monitoramento de culturas [8].

Dentre os meios de monitoramento agrícola, encontra-se a estimativa de área e produtividade. A previsão de produtividade agrícola de acordo com os métodos tradicionais é conduzida por meio de pesquisas agrícolas ou por especialistas, com base em avaliações das condições das culturas, produção histórica da área e condições ambientais [9],[10]. Tais métodos são subjetivos, demorados e muitas vezes pouco representativos, devido ao pequeno tamanho das amostras, que não levam em consideração toda a variabilidade espacial dos talhões de produção [11]. Outra desvantagem desses métodos tradicionais é que demandam muito tempo e são onerosos, dado o grande número de pessoas envolvidas [12].

Para combater a subjetividade dos métodos tradicionais de predição da produtividade agrícola, e possibilitar a análise da variabilidade espacial e temporal, a estimativa baseada em modelos preditivos empíricos com imagens de satélite é uma alternativa promissora, que assiste os produtores

de cana-de-açúcar na tomada de decisão assertiva, auxiliando no manejo das áreas. As imagens de satélite têm sido amplamente utilizadas no monitoramento de culturas agrícolas para avaliação geral do estado da cultura da cana-de-açúcar [13], como por exemplo na estimativa da sua produtividade [14]. A estimativa de produtividade pode ser feita com base em dados agrônômicos, climáticos e a combinação com as imagens de satélite, utilizando técnicas estatísticas convencionais ou de aprendizado de máquina [15-17].

Algumas vantagens do uso de algoritmos de aprendizado de máquina, como floresta aleatória (RF - Random Forest), estão relacionadas à capacidade de usar uma grande quantidade e variedade de informações, como dados numéricos e categóricos, advindos da combinação de dados de sensoriamento remoto e dados agrônômicos [18]. O uso integrado de imagens de satélite com algoritmos de aprendizado de máquina, como o Random Forest, tem mostrado resultados promissores para prever a produção de culturas como trigo [19], soja [20] e da cana-de-açúcar [21] auxiliando na estimativa da produtividade de forma precisa, ao longo dos anos e, em diferentes condições ambientais.

Considerando a variabilidade das condições ambientais, a previsão do rendimento da cultura não é trivial, assim os modelos preditivos utilizando técnicas de mineração de dados e imagens de satélite podem acelerar o desenvolvimento e, melhoria da precisão e robustez desses sistemas de previsão de produtividade de forma regional e temporal. Ainda, o uso de modelos estimativos de produtividade com imagens de satélite assiste os produtores de cana-de-açúcar de forma espacial e temporal na tomada de decisão, auxiliando no manejo das áreas, reduzindo custos e melhorando a produtividade da lavoura.

O objetivo deste trabalho é avaliar o potencial de variáveis agrônômicas e de imagens do satélite Sentinel-2 para estimativa da produtividade de cana-de-açúcar, no estado de São Paulo, utilizando o algoritmo de aprendizado de máquina *Random Forest*.

2. Material e Métodos

2.1 Dados

A área de estudo está localizada na região de Catanduva, no centro do estado de São Paulo, no Brasil **figura 1**. O tipo de solo predominante é o argiloso, sendo as classes de solos com maior predominância os Latossolo Vermelho Amarelo e Latossolo Vermelho. Na área de estudo os ambientes de produção C e D são predominantes e, a produtividade média das últimas três safras foi de 72 ton/ha. Ao total há 3447 talhões com área média de 9,15 hectares. A classificação do clima da região é do tipo AW, caracterizado por ser tropical, com muito mais pluviosidade no verão que no inverno. Sendo a temperatura média de 23,3 °C e a média anual de pluviosidade de 1444 mm [22].

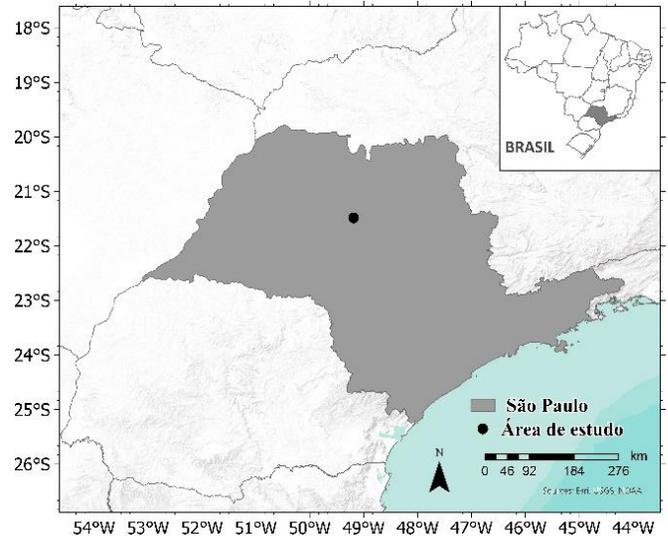
Foram obtidos dados agronômicos de uma empresa parceira para os talhões da área de estudo. Para cada talhão foram obtidos os dados referentes à variedade de cana-de-açúcar, estágio de corte, TCH (tonelada de cana-de-açúcar por hectare), tipos de solo e relevo. No total há 53 variedades de cana-de-açúcar na área de estudo, o estágio de corte no local está entre o 1º até o 8º corte. O relevo varia entre suave ondulado a ondulado. Os dados são referentes à safra de cana-de-açúcar de 2018-2019, que compreende 01 de abril de 2018 até 31 de março de 2019.

Todos os dados foram organizados para retirada de inconsistência e falhas (*outliers*) e, portanto, foi feita uma padronização das informações. Essa análise de consistência foi realizada no software estatístico R [23]. Após o tratamento dos dados, foram obtidos um total de 2691 talhões.

As imagens de satélite foram obtidas por meio do sensor orbital multiespectral MSI (*Multispectral Imager*) a bordo do satélite Sentinel-2A. Essas imagens possuem resolução temporal de 5 dias e 13 bandas espectrais variando de 443 a 2190 nm, com resolução espacial de 10 m para as bandas do vermelho (B4) e infravermelho próximo (B8). Além disso, as imagens possuem correção atmosférica. Foram selecionadas as imagens de refletância de superfície das bandas 4 e 8, as quais foram processadas para o dia 09/07/2018. Em seguida, foi calculado o índice espectral *Normalized Difference Vegetation Index* - NDVI [24] em cada talhão. Por fim, para cada talhão foi calculado os valores de NDVI médio, NDVI máximo e

desvio padrão do NDVI.

Fig. 1 – Localização da área de estudo.



2.2 Modelagem da produtividade da cana-de-açúcar

Os dados agronômicos e os dados das imagens de satélite foram integrados por meio da criação de modelos empíricos, utilizando o algoritmo de regressão Random Forest (RF), implementado no software estatístico R [23]. A técnica RF se baseia na aplicação de bagging em árvores de decisão, com uma importante extensão - além de mostrar os registros, o algoritmo também mostra as árvores. Nas árvores de decisão tradicionais, para determinar como criar uma sub-repartição de uma repartição, o algoritmo escolhe uma variável e um ponto de divisão por meio da minimização de um critério a ser escolhido. Entretanto, no caso do RF, em cada estágio do algoritmo, a escolha de uma variável é limitada a um subconjunto aleatório de variáveis. Dessa forma, quando comparado com o algoritmo de árvore básico, o algoritmo de RF adiciona mais dois passos: o *bagging* e a amostragem bootstrap das variáveis em cada divisão [25]. O RF tem sido usado para previsão de produtividade, devido à sua capacidade de lidar com alta dimensionalidade de dados, detecção de outliers, robustez contra overfitting e a possibilidade de estudar a importância da variável de entrada em um modelo calibrado [26]. Para a construção dos modelos foram utilizados 80% dos dados para calibração e 20% para validação. Foram criados

dois modelos empíricos: i) O Modelo 1 utilizou apenas os dados agronômicos, como o estágio de corte, variedade de cana-de-açúcar, tipo de solo e relevo; ii) o Modelo 2 considerou os dados agronômicos do Modelo 1 (estágio de corte, variedade de cana-de-açúcar, tipo de solo e relevo) e, foi adicionado dados de imagem de satélite, como os valores médios do índice de vegetação NDVI, desvio padrão do NDVI e valores máximos do índice de vegetação NDVI. As métricas de avaliação utilizadas foram o coeficiente de determinação (R^2) e Raiz quadrada média dos erros (RMSE - *Root Mean Square Error*). Para a determinação das métricas de avaliação o TCH estimado pelo modelo, utilizando os 20% dos dados do conjunto de validação, foi comparado ao TCH medido pela usina. Por fim, foi avaliada a importância das variáveis utilizando o algoritmo *Random Forest* [27].

4. Resultados e Discussão

O Modelo 1, com variáveis agronômicas, apresentou R^2 igual a 0,64 e RMSE igual a 10,2 ton/ha. A dispersão dos dados pode ser vista na **figura 2a**. Resultados similares, R^2 de 0,73, foram encontrados para estudo utilizando variáveis agronômicas como o tipo de solo, largura do sulco, rendimento da parcela em relação ao último ano, variedade de cana-de-açúcar, irrigação, controle epidêmico, fertilização, e volume de chuva [28]. Outros estudos de produtividade da cana-de-açúcar, obtiveram variação do erro médio absoluto obtido (MAE) entre 4,6 e 7,5 ton/ha, ou seja, valores de RMSE entre 2,1 e 2,7 ton/ha, próximos ao do presente estudo. Os autores consideraram variáveis de produção e manejo, além do clima ocorrido no período de análise a fim de avaliar modelos como Redes Neurais Artificiais, Máquinas de Vetor de Suporte, Árvores de Regressão Impulsionada e *Random Forest* [29]. Ambos os autores, associaram a produtividade com variáveis climatológicas e atingiram melhores resultados demonstrados pelas métricas de avaliação.

Dentre as variáveis avaliadas, pelo modelo 1, o estágio de corte foi a variável mais importante, seguido pela variedade de cana-de-açúcar, tipo de solo e, por fim relevo (**tabela 1**). Resultado similar foi encontrado por outros autores, que realizando um estudo com árvore de decisão para uma única unidade de usina de açúcar no

oeste do estado de São Paulo, constatou que o número de cortes e teor de argila na camada superior do solo (até 25 cm) são os principais fatores que afetam a produtividade da cana-de-açúcar [30]. Ainda, a avaliação de mais de uma técnica de mineração de dados, utilizando variáveis agronômicas e clima, para identificar e ordenar as principais variáveis que condicionam a produtividade da cana-de-açúcar, mostrou que o número de cortes foi o fator mais importante por todas as técnicas de mineração de dados [16].

Quanto à variedade de cana-de-açúcar sabe-se que a mesma é um indicador direto da produtividade, uma vez que as diferenças entre as variedades contribuem de forma significativa na variabilidade da produtividade [31]. Isto porque cada variedade de cana-de-açúcar tem uma característica específica. Além disso, ressalta-se que há uma interação significativa entre variedades e cortes sucessivos na produtividade da cana-de-açúcar [32], o que reflete diretamente na produtividade.

O relevo foi a variável com menor relevância em relação às outras variáveis. Tal fato pode ser explicado devido à baixa variabilidade das feições do terreno, que permaneceram entre suave ondulado e ondulado e acabam não influenciando, nesse estudo, de forma significativa na variabilidade da produtividade.

Ao adicionarmos outras variáveis ao modelo empírico de produtividade, como os valores médios, desvio padrão e valores máximos do índice de vegetação NDVI, o modelo apresentou melhores ajustes (R^2 igual a 0,83), ver **figura 2b**. O R^2 obtido aumentou para 0,83 e o RMSE diminuiu para 7,0 ton/ha em comparação ao modelo 1. Tal fato, demonstra que as variáveis provenientes de imagens de satélite trouxeram um ganho de informação na criação de um modelo empírico de estimativa da produtividade de cana-de-açúcar. Estudos similares, encontraram valores de R^2 igual a 0,94 para cana-de-açúcar utilizando imagens do satélite Landsat-8, e os índices de vegetação que combinam a banda do vermelho e infravermelho como o NDVI, Índice de Vegetação Aprimorado (EVI) e Índice de Vegetação Ajustado ao Solo (SAVI), além do Índice de Vegetação Normalizada do Verde (GNDVI) [13]. Apesar das diferenças dos resultados dos modelos, vale ressaltar que estes estudos consideraram mais de um índice de vegetação ao longo de séries temporais, ao in-

vés de uma única data, que foi constatado pelos autores estimar de forma mais precisa em comparação a um único índice de vegetação e uma data específica de coleta da imagem de satélite.

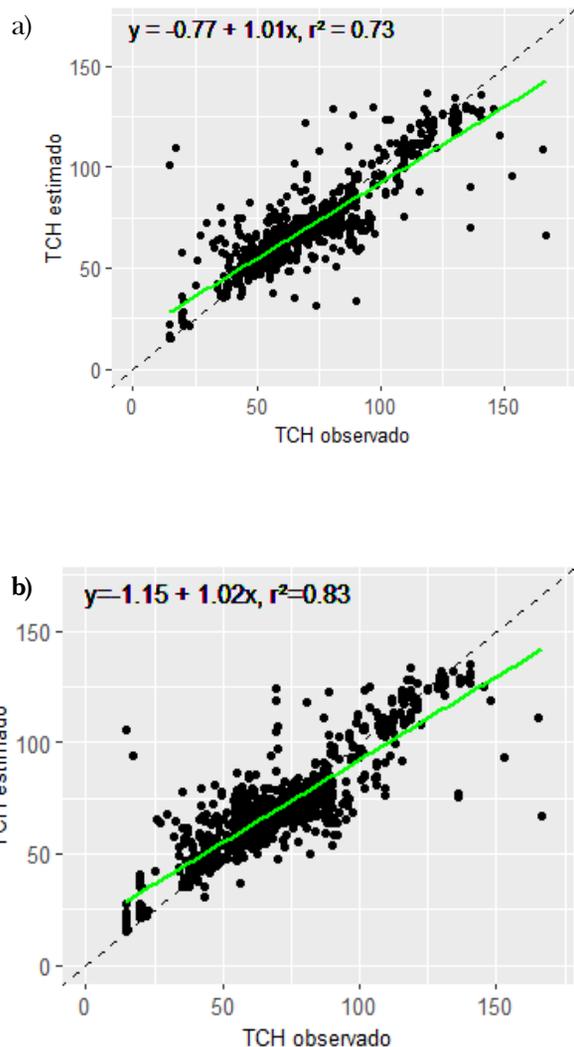
Quanto às importâncias de variáveis, para o Modelo 2, as variáveis que tiveram maior importância na estimativa de produtividade foram similares ao Modelo 1, ou seja, o estágio de corte e a variedade de cana-de-açúcar. No entanto, as variáveis relacionadas ao NDVI tiveram maior importância que o tipo de solo e relevo **tabela 1**. Os índices de vegetação, como o NDVI, estão relacionados principalmente à abundância de cobertura vegetal verde e biomassa e são sensíveis a variações nas respostas de estresse do dossel da planta, cultivares e práticas de gestão [33, 34]. Além disso, os dados de índices de vegetação obtidos no período do seu máximo crescimento, têm relação direta com a produtividade da cultura da cana-de-açúcar [35]. Outros estudos mostraram a importância de índices como o NDVI para a previsão da produtividade da cana-de-açúcar. Mulianga et all. [36] obtiveram RMSE inferior a 5 ton/ha para estimativa da produtividade utilizando o NDVI do satélite MODIS (Moderate Resolution Imaging Spectroradiometer) e dados agrônômicos, por meio de um modelo de regressão linear. Já Fernandes et all. [37] também mostraram a importância do NDVI ao utilizá-lo para prever a produtividade da cana-de-açúcar no Estado de São Paulo, para isso foram utilizadas imagens do sensor MODIS e um modelo de conjunto de redes neurais artificiais. O R² obtido nesse estudo foi de 0,61.

Tab. 1: Importância das variáveis.

Ordem de importância	Modelo 1	Modelo 2
1	Estágio de corte	Estágio de corte
2	Variedade de cana-de-açúcar	Variedade de cana-de-açúcar
3	Tipo de solo	NDVI médio
4	Relevo	Desvio padrão NDVI
5	-	NDVI máximo
6	-	Tipo de solo
7	-	Relevo

Fonte: Autor.

Fig. 2 – Gráfico de dispersão para o modelo 1 considerando apenas variáveis agrônômicas (a) e considerando variáveis agrônômicas e de satélite -modelo 2 (b).



5. Conclusões

Foram criados modelos empíricos de produtividade utilizando informações agrônômicas e imagens do satélite Sentinel-2, a partir do algoritmo *Random Forest*. Os resultados demonstraram que a combinação de variáveis agrônômicas e de imagens de satélite trouxe melhorias na estimativa da produtividade de cana-de-açúcar. As variáveis de maior importância para os modelos empíricos foram as variáveis estágio de corte, variedade de ca-

na-de-açúcar e NDVI médio do talhão. Estudos futuros devem ser realizados para a integração de informações climáticas e outros índices de vegetação. Por fim, a utilização de séries temporais de imagens para estimativa e acompanhamento da produtividade pode trazer ganhos ao monitoramento da cana-de-açúcar.

Referências

- [1] CONAB. Acompanhamento da Safra Brasileira de Cana-de-Açúcar – Quarto Levantamento da safra 2020/21. Companhia Nacional de Abastecimento, v.7, 2021, p. 57
- [2] Surendran, U.; Ramesh, V.; Jayakumar, M.; Marimuthu, S.; Sridevi, G. Improved sugarcane productivity with tillage and trash management practices in semi-arid tropical agro ecosystem in India Soil Tillage Res., 158, 2016.pg 10-21.
- [3] INMAN-BAMBER, N.G.; BONNETT, G.D.; SPILLMAN, M.F.; HEWITT, M.L.; JACKSON, J. Increasing sucrose accumulation in sugarcane by manipulating leaf extension and photosynthesis with irrigation. Australian Journal of Agricultural Research, v.59, p.13-26, 2008.
- [4] MACHADO, R.S.; RIBEIRO, R.V.; MARCHIORI, P.E.R.; MACHADO, D.F.S.P.; MACHADO, E.C.; LANDELL, M.G.A. Respostas biométricas e fisiológicas ao déficit hídrico em cana-de-açúcar em diferentes fases fenológicas. Pesquisa Agropecuária Pesquisa Agropecuária Brasileira, Brasília. v.44, n.12, p.1575-1582, 2009.1582, 2009.
- [5] IBGE. Produção Agrícola Municipal – PAM. 2019. Disponível em: <<https://sidra.ibge.gov.br>>. Acesso em: 10 ago. 2021.
- [6] FAO. The Future of Food and Agriculture-Trends and Challenges, 2017.
- [7] IPCC Summary for policymakers. Masson-Delmotte, V. Zhai, P.; Pörtner, H.O. Roberts, D. Skea, J. Shukla, P.; Pirani, A.; Moufouma-Okia, W. Péan, C.; Pidcock, R. Connors, S.; Matthews, J. Chen, Y. Zhou, X.; Gomis, M.; Lonnoy, E.; Maycock, T.; Tignor, M.; Waterfield, T. (Eds.), Global Warming of 1.5 °C. An IPCC Special Report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. World Meteorological Organization, Geneva, Switzerland, 2018, 32pg.
- [8] Holzman, M.E.; Rivas, R.; Piccolo, M.C. Estimating soil moisture and the relationship with crop yield using surface temperature and vegetation index. Int. J. Appl. Earth Obs. Geoinf., 28, 2014, pg. 181-192. 10.1016/j.jag.2013.12.006
- [9] Bocca, F.F.; Rodrigues, L.H.A. Arraes, N.A.M. When do I want to know and why? Different demands on sugarcane yield predictions. Agric. Syst., 135, 2015, pg. 48-56.
- [10] Everingham, Y.L.; Muchow, R.C.; Stone, N.G.; Inman-Bamber, A. Singels, C.N. Bezuidenhout Enhanced risk management and decision-making capability across the sugarcane industry value chain based on seasonal climate forecasts Agric. Syst., 74, 2002, pg. 459-477. 10.1016/S0308-521X (02)00050-1
- [11] Basso, B., Cammarano, D., Carfagna, E. Review of Crop Yield Forecasting Methods and Early Warning Systems, in: Report Presented to First Meeting of the Scientific Advisory Committee of the Global Strategy to Improve Agricultural and Rural Statistics FAO Headquarters, Rome, Italy. 2, 2013, pg. 1-56. 10.1017/CBO9781107415324.004
- [12] Picoli, M.C. A Estimativa da produtividade da cana-de-açúcar utilizando agregados de redes neurais artificiais: estudo de caso Usina Catanduva. 2007. 90p. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2007.
- [13] Singla, S.K., Garg, R.D., Dubey, O.P. Ensemble machine learning methods to estimate the sugarcane yield based on remote sensing information. Revue d'Intelligence Artificielle, v. 34, n.6, 2020, pg. 731-743. <https://doi.org/10.18280/ria.340607>
- [14] Cechin-Júnior, C. Johann, J. A. Antunes, J. F. G.; Deppe, F. Sugarcane mapping in Paraná State Brazil using MODIS EVI images. International Journal of Advanced Remote Sensing and GIS, v.9, n.1, pg. 3205-3221, 2020.
- [15] Verma, A.M.; GaRG, P. K.; Prasad, K. S. H.; Dadhwal, V. K.; Dubey, S. K.; Kumar, A. Sugarcane Yield Forecasting Model Based on Weather Parameters. Sugar Tech. v.23, n.1, 2021, pg.158–166.
- [16] Hammer, R. G.; Sentelhas, P. C.; Mariano, J. C. Q. Sugarcane yield prediction through data mining and crop simulation models. Sugar Tech., 22, 2020, pg. 216-225. 10.1007/s12355-019-00776-z.
- [17] Luciano, A. C. S.; Picoli, M. C. A.; Duft, D. G.; Rocha, J. V.; Leal, M. R. L. V.; Maire, G. Empirical model for fore-

casting sugarcane yield on a local scale in Brazil using Landsat imagery and random forest algorithm. *Computers and Electronics in Agriculture*. 184, 2021, 106063. <https://doi.org/10.1016/j.compag.2021.106063>.

[18] Everingham, Y., Sexton, J., Skocaj, D., Inman-Bamber, G. Accurate prediction of sugarcane yield using a random forest algorithm. *Agron. Sustent. Dev.*, 36, 2016, 10.1007 / s13593-016-0364-z.

[19] Kamir, E.; Waldner, F.; Hochman, Z. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *Journal of Photogrammetry and Remote Sensing*, v.160, 2020, pg. 124-135.

[20] Schwalbert, R.A.; Amado, T.; Corassa, G.; Pott, L.P.; Prasad, P.V.V.; Ciampittia, I.A. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology*. v.284, 2020, p. 07886.

[21] Shendryk, Y.; Davy, R.; Thorburn, P. Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning. *Field Crops Research*. v.260, 2021, p.107984. <https://doi.org/10.1016/j.fcr.2020.107984>

[22] Köppen W, Geiger R (1928) *Klimate der Erde*. Justus Perthes, Gotha. 1928. Wall-Map 150 cm x 200 cm.

[23] RSTUDIO. Studio Team. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. 2020. URL <http://www.rstudio.com/rstudio>

[24] Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation system in the great plains with ERTS Earth Resources Technology Satellite-1 Symposium, 2. Washington, D.C., Proceeding,1, NASA. Goddard Space Flight Center, Washington, D.C., 1973, pg.309-317.

[25] BRUCE, P; BRUCE, A. *Estatística Prática para Cientistas de Dados: 50 conceitos essenciais*. 1. ed. Rio de Janeiro: Alta Books, 2019.

[26] Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random forests for land cover classification *Pattern Recognit. Lett.*, 27, 2006, pg. 294-300. 10.1016/j.patrec.2005.08.011.

[27] Breiman, L. Random forests. *Mach. Learn.*, 45, 2001, pg. 5-32. 10.1023/A:1010933404324.

[28] Charoen-Ung, P. Sugarcane Yield Grade Prediction using Random Forest and Gradient Boosting Tree Techniques.15th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2018, pg.1-6.

[29] Bocca, F. F.; Rodrigues, L. H. A. The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Computers and Electronics in Agriculture*, v. 128, 2016, p. 67-76.

[30] Pelóia, P.R., F.F. Bocca, and L.H.A. Rodrigues. 2019. Identification of patterns for increasing production with decision trees in sugarcane mill data. *Scientia Agricola* 76. 2019, pg.281-289.

[31] Ramburan, S.; Wettergreen, T.; Berru, S.D.; Shongwe, B. Genetic, environmental and management contributions to ratoon decline in sugarcane. *Field Crops Research*, Amsterdam, v. 146, 2013, pg.105-112.

[32] Arnt, W. R. Desempenho de variedades de cana-de-açúcar em duas épocas de colheita no pontal do Paranapanema. Orientador: 2016. 55p. Dissertação (Mestrado) – Agronomia. Universidade Federal da Grande Dourados. Dourados-MT.

[33] Rudorff, B.F.T.; Batista, G.T. Yield estimation of sugarcane based on agrometeorological spectral models. *Remote Sens. Environ.*, 33, 1990, pg. 183-192, 10.1016/0034-4257(90)90029-L

[34] Silleos, N.G.; Alexandridis, T.K.; Gitas, I.Z.; Perakis, K. Vegetation indices: advances made in biomass estimation and vegetation monitoring in the last 30 years. *Geocarto Int.*, 21, 2006, pg. 21-28. 10.1080/10106040608542399

[35] Pinheiro Lisboa, I.; Melo Damian, J. Roberto Cherubin, M. Silva Barros, P.P. Ricardo Fiorio, P. Cerri, C.C. Eduardo Pellegrino Cerri, C. Prediction of Sugarcane Yield Based on NDVI and Concentration of Leaf-Tissue Nutrients in Fields Managed with Straw Removal. *Agronomy* 8, 2018, pg.196. <https://doi.org/10.3390/agronomy8090196>

[36] Mulianga, B.; Begue, A.; Simoes, M. Todoroff, P. Forecasting regional sugarcane yield based on time integral and spatial aggregation of MODIS NDVI. *Remote Sens.*, v.5, n.5, 2013, pg. 2184-2199. 10.3390/rs5052184.

[37] Fernandes, J.L.; Esquerdo, J.C.D.; Favilla, N.F. Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble. *Taylor & Francis*, v.38, n.16, 2017, pg. 4631-4644.