

UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL

**MODELAGEM DA PRODUTIVIDADE DE CANA-DE-AÇÚCAR A PARTIR DE
ALGORITMOS DE APRENDIZADO DE MÁQUINA**

LUCAS DE OLIVEIRA GONÇALVES

Orientador: Prof. Dr. Alan Rodrigo Panosso

Trabalho apresentado à Faculdade de Ciências Agrárias e Veterinárias - UNESP, Câmpus de Jaboticabal, para graduação em ENGENHARIA AGRÔNOMICA.

Jaboticabal - SP

nono/2023

FICHA CATALOGRÁFICA

G635m Gonçalves, Lucas de Oliveira
Modelagem da produtividade de cana-de-açúcar a partir de algoritmos de aprendizado de máquina / Lucas de Oliveira Gonçalves. -- Jaboticabal, 2023
36 p. : il., fotos

Trabalho de conclusão de curso (Bacharelado - Engenharia Agrônômica) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientador: Alan Rodrigo Panosso

1. Inteligência artificial. 2. Modelagem de produção. 3. Tomada de decisão. 4. Árvore de decisão. 5. Random forest. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.



UNIVERSIDADE ESTADUAL PAULISTA
CÂMPUS DE JABOTICABAL



DEPARTAMENTO: Engenharia e Ciências Exatas.

CERTIFICADO DE APROVAÇÃO TRABALHO DE INICIAÇÃO CIENTÍFICA

TÍTULO: Modelagem da produtividade de cana-de-açúcar a partir de algoritmos de aprendizado de máquina




ACADÊMICO: Lucas de Oliveira Gonçalves

CURSO: Engenharia Agrônômica

ORIENTADOR (ES): Prof. Dr. Alan Rodrigo Panosso


Aprovado e corrigido de acordo com as sugestões da Banca Examinadora

BANCA EXAMINADORA:

	(Nomes)	(Assinaturas)
Presidente	Prof. Dr. Alan Rodrigo Panosso	
Membro	Dr.ª Ludhanna Marinho Veras	
Membro	MSc. Kleve Freddy Ferreira Canteral	

Jaboticabal 3 / 8 / 2023.

Aprovado por ato "ad referendum" do Conselho do Departamento em 3 / 8 / 2023.


Prof. Dr. Danisio Prado Munari
Chefe de Departamento de Ensino

AGRADECIMENTOS

Agradeço à minha família, em especial aos meus pais, Marisa Aparecida de Oliveira Gonçalves e Paulo César Gonçalves, por todos os ensinamentos, conselhos e, principalmente por estarem ao meu lado em todas as minhas decisões e ideias desde pequeno. Essa conquista também é de vocês.

Aos professores da Unesp de Jaboticabal por todos os ensinamentos compartilhados de forma tão clara, completa e motivadora. Agradeço em especial, o professor Dr. Alan Rodrigo Panosso que desde 2019, é meu orientador com grandes conselhos e ensinamentos essenciais para a realização deste trabalho e também à professora Dra. Mara Cristina Pessôa da Cruz por todos os ensinamentos, conselhos, oportunidades de trabalho e conversas no laboratório.

À grande República Abatedouro por ter sido coautora de grandes conquistas pessoais e profissionais desde nossa convivência em 2018, na qual sou muito realizado de fazer parte e, que sempre esteve ao meu lado, sobretudo no meu intercâmbio na França. Hoje tenho a felicidade de chamá-la de segunda família.

Ao meu amigo Ezequias Correia pela participação na organização deste trabalho e por todas as experiências compartilhadas durante a graduação. Suas contribuições foram de grande relevância para a qualidade do trabalho final.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq - pelo auxílio financeiro para desenvolvimento desta pesquisa em um primeiro momento em abril de 2019 e de um segundo momento em abril de 2021.

Por fim, à esta incrível e grandiosa instituição de ensino “Alma Mater, a Faculdade de Ciências Agrárias e Veterinárias da UNESP/FCAV, câmpus de Jaboticabal, que sem a oportunidade que me foi concedida, eu não poderia ter tido tal crescimento pessoal e profissional. Eu fico honrado com a concessão do intercâmbio na França que agregou tanto à minha vida.

ÍNDICE

RESUMO:	1
ABSTRACT:	2
1. INTRODUÇÃO	3
2. REVISÃO DE LITERATURA	5
2.1 A cultura da cana-de-açúcar	5
2.2 A importância da cana-de-açúcar no Brasil	6
2.3 Machine Learning (ML)	8
3. MATERIAL E MÉTODOS	11
3.1 Caracterização e amostragem da área de estudo	11
3.2 Análise dos dados	14
4. RESULTADOS E DISCUSSÃO	15
5. CONCLUSÕES	24
6. REFERÊNCIAS	25

MODELAGEM DA PRODUTIVIDADE DE CANA-DE-AÇÚCAR A PARTIR DE ALGORITMOS DE APRENDIZADO DE MÁQUINA

RESUMO:

A cana-de-açúcar é uma das culturas de maior interesse do mercado agrícola brasileiro. Algoritmos de aprendizado de máquina têm sido cada vez mais empregados para modelagens agronômicas e podem fornecer informações que contribuam para melhorar o planejamento, gestão e, conseqüentemente, aumentar a produção agrícola. Dessa forma, o objetivo foi criar modelos estatísticos por meio de técnicas de aprendizado de máquina, incluindo as estratégias de aplicação, para a previsão da produção da cana-de-açúcar em toneladas de cana por hectare (TCH) no Estado de São Paulo. Os dados experimentais foram coletados nos anos de 2016, 2017 e 2018, sendo provenientes de duas unidades produtivas, localizadas nas intermediações dos municípios de Potirendaba e Catanduva, SP. Foram utilizadas 20, 12 e 10 mil observações para os anos de 2016, 2017 e 2018, respectivamente. Os dados foram padronizados com o objetivo de identificar os atributos do solo e manejos específicos, que são variáveis independentes e, que melhor explicam a variabilidade espaço temporal do rendimento da cana-de-açúcar em toneladas de cana por hectare (TCH). As modelagens foram realizadas por meio do modelo de árvores de decisão e *Random forest*. O algoritmo *Random forest* apresentou melhor desempenho preditivo para a cultura de cana-de-açúcar quando comparado ao modelo de árvore de decisão. Em termos numéricos, para o *Random forest*, obteve-se R^2 superior à 56% e RMSE inferior a 18,2 t ha⁻¹, enquanto que para a árvore de decisão, o R^2 obtido foi superior à 43% e RMSE inferior à 23,4 t ha⁻¹. As variáveis foram avaliadas quanto ao nível de importância entre elas na produção de cana-de-açúcar. A variável corte apresentou maior nível de importância na predição da cultura, no mínimo 3,5 vezes superior, em relação às outras variáveis de manejo subsequentes como variedade, idade e mês. Além disso, o açúcar total recuperável (ATR), bem como os níveis de potássio (K⁺), fósforo disponível (P) e matéria orgânica do solo (MOS), também se apresentaram inferiores à nível de importância para a predição em relação às demais variáveis.

Palavras-chave: inteligência artificial; modelagem de produção; tomada de decisão; árvore de decisão; *Random forest*.

MODELING SUGARCANE PRODUCTIVITY USING MACHINE LEARNING

ABSTRACT:

Sugarcane is one of the crops of most interest to the Brazilian agricultural market. Machine learning algorithms have been increasingly used for agronomic modeling and can provide information to help improve planning and management and, consequently, increase agricultural production. Therefore, the objective was to create statistical models using machine learning techniques, including application strategies, to estimate sugarcane production in tons of cane per hectare (TCH) in the state of São Paulo. The experimental data was collected in 2016, 2017 and 2018 from two production units located in the middle of the municipalities of Potirendaba and Catanduva, SP. The study used 20,000, 12,000 and 10,000 observations for 2016, 2017 and 2018, respectively. The data was standardized in order to identify the soil attributes and specific managements that are independent variables and that best explain the spatiotemporal variability of sugarcane yields in tons of cane per hectare (TCH). The modeling was conducted using decision trees and *Random forest*. The *Random forest* algorithm showed better predictive performance for the sugarcane crop when compared to the decision tree model. In numerical terms, the *Random forest* obtained an R² of more than 56% and an RMSE of less than 18.2 t ha⁻¹, while the decision tree obtained an R² of more than 43% and an RMSE of less than 23.4 t ha⁻¹. The variables were evaluated according to their level of importance in sugarcane production. The variable cut showed the highest level of importance in predicting the crop, at least 3.5 times higher than the other subsequent management variables such as variety, age and month. In addition, the total recoverable sugar (ATR), as well as the levels of potassium (K⁺), available phosphorus (P) and soil organic matter (MOS), were also lower in terms of importance for prediction than the other variables.

Keywords: artificial intelligence; production modeling; decision making; decision tree; *Random forest*.

1. INTRODUÇÃO

O Brasil é um dos países de maior destaque do setor sucroenergético e grande parte dessa evidência é dada pela cana-de-açúcar (*Saccharum spp.*) que possui grande influência em termos sociais, econômicos e ambientais (CONSTANTINO, 2022). Essa *commodity* está presente majoritariamente no Estado de São Paulo e, com expansões desde o início do século 21, para outras regiões como por exemplo, a região Centro-Oeste, em especial para Goiás e Mato Grosso do Sul (GRANCO et al., 2015; CALDAS et al., 2018; VEDANA et al., 2019). Em âmbito mundial, o Brasil está entre as maiores potências de produção, sendo a estimativa de produção da safra 2022/23, segundo o terceiro levantamento, de 598,3 milhões de toneladas, com grande participação do Estado de São Paulo que é o maior produtor de cana-de-açúcar do Brasil (CONAB, 2022). É importante ressaltar também que o Brasil contribui diretamente para o segundo Objetivo do Desenvolvimento Sustentável das Nações Unidas, o de fome zero e agricultura sustentável.

Atualmente emprega-se alta tecnologia no cultivo da cana-de-açúcar e a resposta é uma alta produtividade, que neste aspecto, participam com grande influência a adubação e a inoculação de bactérias (FERNANDES, 2022). É constante o esforço empregado para que seja possível o aumento da produção de biomassa, o rápido crescimento e a tolerância a pragas e doenças (AMARAL et al., 2019) e, além disso, manter a característica benéfica da planta em relação à captura do dióxido de carbono (CO₂) da atmosfera por meio das altas taxas fotossintéticas (KORMANN, 2022).

Nesse contexto, novas tecnologias e ferramentas são utilizadas na agricultura e estão inseridas na nova era de dados, como por exemplo o uso do

aprendizado de máquina (*Machine Learning* - ML), área importante da inteligência artificial e que faz parte do que chamamos hoje de agricultura 5.0 (REICHSTEIN et al., 2019). Em conjunto com o aprendizado de máquina, os agricultores podem utilizar dados que facilitem a tomada de decisão, para resolução de problemas relacionados à gestão da água, gestão de solos e gestão de culturas, sendo que esta, em especial, inclui a previsão do rendimento, detecção de plantas entre outros (CRAVERO, et al., 2022).

O ML é uma área de grande importância dentro da Inteligência Artificial (IA) e da ciência de dados. Sales (2023) define *Machine Learning* como sendo previsões alcançadas a partir da análise dos dados retirados de fatos pretéritos, considerando que o contexto em que os dados foram inseridos não mude. Segundo ela, a máquina é ensinada a detectar padrões que tendem a se repetir futuramente. Ele pode ainda ser definido como um campo de pesquisa estatística para o treinamento de algoritmos computacionais que transformam um conjunto de dados a fim de maximizar a classificação, predição, agrupamento, como também descobrir novos padrões em um conjunto de dados-alvo (REICHSTEIN et al., 2019).

Júnior et al. (2022) veem o *Machine Learning* como uma ferramenta importante na tomada de decisão para o monitoramento de pragas e doenças na cultura da soja. Eles utilizaram a ferramenta para realizar o reconhecimento das pragas lagarta da soja e falsa-medideira, a partir da análise de imagens de folhas coletadas. Segundo eles, a tecnologia de drones para captar as imagens à campo e uma interface de usuário seriam de grande importância para tornar o monitoramento da lavoura mais completo e assertivo (JÚNIOR et al., 2022).

Diante do exposto, esse trabalho teve como objetivo criar modelos estatísticos por meio de técnicas de aprendizado de máquina e compará-los, utilizando estratégias de aplicação, para realizar a previsão da produção da cana-de-açúcar em toneladas de cana por hectare (TCH) no Estado de São Paulo.

2. REVISÃO DE LITERATURA

2.1 A cultura da cana-de-açúcar

A cana-de-açúcar (*Saccharum* spp) possui grande importância para o Brasil, é uma das principais culturas agrícolas do país, sendo uma importante fonte de matéria-prima para a produção de açúcar e álcool. Ela é uma planta da família *Poaceae*, subfamília *Panicoideae*, tribo *Andropogoneae*, subtribo *Saccharinae* e ao gênero *Saccharum* L. (LUCCHESI, 2001). Atualmente no Brasil e no mundo, são desenvolvidas e cultivadas seis espécies comerciais, oriundas deste gênero que porta uma grande variabilidade genética (EMBRAPA, 2015). A cana-de-açúcar pertence ao grupo de plantas C4 que são aquelas que formam compostos orgânicos com quatro carbonos durante a fotossíntese, processo este com alta taxa de ocorrência e, apresenta eficiência na utilização e resgate de CO₂ (gás carbônico) da atmosfera (SEGATO et al., 2006).

A cana-de-açúcar, por ser uma planta com metabolismo C4, possui uma maior taxa fotossintética do que as plantas C3, o que indica que alta incidência de radiação solar, dias longos, com temperaturas elevadas e com boa disponibilidade de água no solo levam a um maior perfilhamento e, conseqüentemente, a produtividades mais elevadas. Por outro lado, em

condições de dias curtos, com baixa irradiância solar e temperaturas amenas, o perfilhamento é menos intenso (MONTEIRO, 2011) e as canas se tornam mais longas e finas (RODRIGUES, 1995), mesmo sob condições de boa disponibilidade de água no solo.

Desde a introdução há 500 anos de cultivo da cultura no período colônia, o setor canavieiro desenvolveu-se enormemente. A cana-de-açúcar deixou de ser matéria-prima voltada apenas à produção de açúcar e tornou-se via de produção de álcool e energia expandindo, assim, sua utilização. Ela é a matéria-prima com o melhor rendimento na produção de etanol e, quando comparada aos combustíveis fósseis, está presente como uma opção de combustível menos poluente, além de ser renovável (CARDOZO, 2012).

A cana-de-açúcar é classificada como uma planta semi-perene e possui um ciclo de cultivo que se estende por aproximadamente cinco anos (LEITE et al., 2008). A classificação é dada devido ao fato de a cultura permanecer, em média, 5 anos sendo submetida à cortes no campo, após a ocorrência da operação de plantio. Segundo Doorenbos; Kassan (1979) a cana-de-açúcar é uma cultura típica de climas tropicais e subtropicais que é cultivada sob condições de sequeiro em regiões úmidas, como também, sob irrigação em regiões sub-úmidas com latitudes entre 35° N e 35° S.

2.2 A importância da cana-de-açúcar no Brasil

Segundo CONAB (2022), a produção de cana-de-açúcar no Brasil realizada na safra 2022/23, foi de 598,3 milhões de toneladas, com uma grande participação da região Centro-Sul, que se estima contribuir com 539,6 milhões

de toneladas na produção deste período, sendo superior em 2,8 % em relação à safra anterior de 2021/22. Na safra de 2022/23 a área estimada para o cultivo da cana-de-açúcar é de 8,3 milhões de hectares. Houve redução de 37,7 mil hectares de produção de cana-de-açúcar devido às excelentes cotações do milho e da soja e, portanto, uma perda de espaço para as culturas anuais, mesmo havendo boas cotações no preço do açúcar e do etanol (CONAB, 2022).

Incentivos governamentais como o advento do Programa Nacional do Álcool – PROÁLCOOL criado pelo governo federal (Decreto 76593 de 14/11/1975) culminou no surgimento de um novo ciclo de pesquisa que suportou a expansão da cultura no país, ocupando novas áreas nas regiões tradicionalmente produtoras, como o Estado de São Paulo e, também, expandindo para regiões de cerrado (LANDELL et al., 2003a).

Centros de pesquisa e suporte à produção de cana-de-açúcar foram criados a fim de acompanharem a expansão e os novos desafios advindos desse cenário da cultura, visto que é necessário o constante lançamento de novas variedades e avanço com melhoramento genético, principalmente pelo curto tempo de vida útil de uma variedade de cana-de-açúcar (REFERÊNCIA). Deu-se início aos programas de melhoramento genético do Instituto Agrônomo de Campinas (IAC), Cooperativa de Produtores de Cana-de-Açúcar, Açúcar e Álcool do Estado de São Paulo (COPERSUCAR) e Rede Interuniversitária para o Desenvolvimento do Setor Sucroenergético (RIDESA), instituições estas que surgiram, no Brasil, no começo do século passado e que possuem como premissa a obtenção de novas variedades (LANDELL et al., 2003b). Atualmente,

o programa do Centro de Tecnologia Canavieira (CTC) ocupa a posição que antes era do programa de melhoramento da COPERSUCAR.

2.3 Machine Learning (ML)

A inteligência artificial (IA) permite que tarefas características da inteligência humana sejam realizadas através da assimilação de informações de um algoritmo, compreendendo reconhecer objetos e sons, contextualizar a linguagem, aprender com o ambiente e resolver problemas. É possível dizer que a IA reflete no potencial de tomada da melhor decisão possível por uma máquina referente à quantidade de informação disponível e habilidade de adaptação à diferentes situações (KUHN; JOHNSON, 2013; CHOLLET; ALLAIRE, 2017).

Em 1950 surgiram as primeiras aplicações da Inteligência Artificial, utilizada na área da ciência da computação, objetivando automatizar tarefas intelectuais que no geral são executadas pelo homem. De acordo com Chollet e Allaire (2017), até o ano de 1989, havia um entendimento de que a IA poderia, em níveis do homem, ser compreendida e alcançada através da programação de grande conjunto de regras explícitas e, possibilitar assim, a manipulação do conhecimento. Essa visão ficou conhecida como IA simbólica e, mesmo sendo capaz de solucionar muito dos problemas de ordem lógica bem definida, não apresentou eficiência na solução de problemas mais complexos e confusos, como exemplo a classificação de imagens, o reconhecimento de linguagem natural e, também a tradução. A explicação para essa funcionalidade é dada quando são analisados os dados responsáveis para a compreensão dos problemas. A IA simbólica não confere resultados em descobrir regras explícitas

pois, essas necessitam ser previamente programadas e, os dados são processados em conformidade à essas regras (KUHN; JOHNSON, 2013; BRUCE; BRUCE, 2017; CHOLLET; ALLAIRE, 2017).

No ano de 1959, o americano Arthur Samuel, pioneiro em jogos de computador e inteligência artificial, definiu o aprendizado de máquina ou Machine Learning (ML) como um campo de estudo que permite aos computadores a capacidade de aprender sem serem expressamente programados (PHU NGUYEN; LEE, 2017). Esse aprendizado de máquina é uma das maiores áreas dentro da inteligência artificial e da ciência de dados e, como exemplo de utilização, em 2017 ele foi utilizado para uma partida de xadrez em que, inicialmente, foi disponibilizado para a máquina apenas as regras do jogo. Posteriormente ela foi capaz de aprender com os erros da partida anterior e adaptar-se de forma a demonstrar um nível superior quando comparado à partida anterior.

O aprendizado de máquina pode ser definido, também, como um subconjunto da inteligência artificial e, que é diretamente relacionado com algoritmos e métodos capazes de aprender a partir de um referenciamento com um banco de dados. Comparativamente ao processo cognitivo dos humanos, os fundamentos são idênticos, porém o modo de cognição é diferente. Uma determinada resposta pode ser aperfeiçoada por meio do erro, ou seja, a experiência conduz ao aprendizado (AZEVEDO, 2018). De uma outra forma, BRUCE; BRUCE (2017) acrescentam que a própria máquina aprende as regras para a tomada de decisão.

Os autores Hastie et al., (2001); Bishop, (2006) e Tarca et al., (2007) definem o objetivo principal do aprendizado de máquina como a inferência de propriedades gerais na distribuição de dados com um treinamento de alguns dados prévio. Sendo definido automaticamente as regras da aprendizagem mecânica através do algoritmo com o objetivo de discriminar as classes que podem, assim, serem aplicadas ao conjunto completo de dados. Esse tipo de aprendizagem é denominado por supervisionada.

Nos últimos anos foram conduzidos vários trabalhos utilizando técnicas geoestatísticas no Brasil almejando compreender a variabilidade espacial da emissão de CO₂ do solo (FCO₂), conferindo em informações que possam beneficiar áreas agrícolas, principalmente na escolha de práticas agrícolas mais sustentáveis para a mitigação das emissões de gases de efeito estufa (GEE) (LA SCALA et al., 2000; LA SCALA et al., 2009; PANOSSO et al., 2009; PANOSSO et al., 2012; TEIXEIRA et al., 2013; BICALHO et al., 2014; LEON et al., 2014; MANTOVANELLI et al., 2016; DA CUNHA et al., 2018; TAVANTI et al., 2020).

SANTOS et al. (2022) aplicaram o ML para realizar uma predição da gestão da água na cultura do lúpulo. Dentro desse contexto de agricultura de precisão, essa ferramenta pode trazer avanços interessantes quanto à gestão da irrigação sobretudo em ambientes de plantios de safras orgânicas e mais complexas, sendo capaz de analisar dados em tempo real e prever recomendações de irrigação.

Hammer (2020), analisando dados de produtividade, manejo agrícola e dados meteorológicos de 18 usinas de cana-de-açúcar no estado de São Paulo, ordenou as variáveis que mais responsivas à produtividade de cana-de-açúcar,

de acordo com a sua importância, e desenvolveu modelos matemáticos, com base em algoritmos de aprendizado de máquina (*Random forest*, árvore de decisão, XgBoost e Support Vector Machine). O autor observou que dentre as variáveis que afetam a produtividade da cana, o número de cortes foi o que teve maior peso, seguido da idade do corte e de variáveis ambientais relativas às condições meteorológicas e dos ambientes de produção (solo). Em adição, observou que os modelos gerados tiveram desempenho mediano, no entanto, superiores ao modelo da zona agroecológica da FAO.

A utilização da Inteligência Artificial é difundida atualmente em vários campos da ciência, principalmente nas áreas da saúde, da indústria e ambiental (GE et al., 2017; CHIAVEGATTO FILHO et al., 2018; DI MININ et al., 2018; SANTOS et al., 2019). É notório o sucesso da aplicação dessas técnicas que tornam a IA uma ferramenta constante no cotidiano da sociedade contemporânea, como exemplo, quando inserida aos dados de mídia social, que permite vasta capacidade de investigação do comportamento humano (DI MININ et al., 2018).

3. MATERIAL E MÉTODOS

3.1 Caracterização e amostragem da área de estudo

Os dados utilizados neste trabalho foram oriundos de duas unidades produtivas de cana-de-açúcar referentes a um banco de dados contendo informações agrícolas e adquiridos a partir de uma parceria entre a universidade, representada pela UNESP, iniciativa privada e a usina produtora de cana-de-

açúcar, localizada no interior do estado de São Paulo, com um total de aproximadamente 53,7 mil há.

Os dados utilizados no experimento foram coletados entre os anos de 2016 e 2018 a partir de uma parceria entre CSME (FCAV/UNESP), Athenas Agrícola e a usina COFCO Internacional, totalizando mais de 50 mil pontos de observação (Figura 1) e são referentes às unidades produtoras de cana-de-açúcar do noroeste paulista da usina COFCO Internacional nos municípios de Catanduva e Potirendaba.

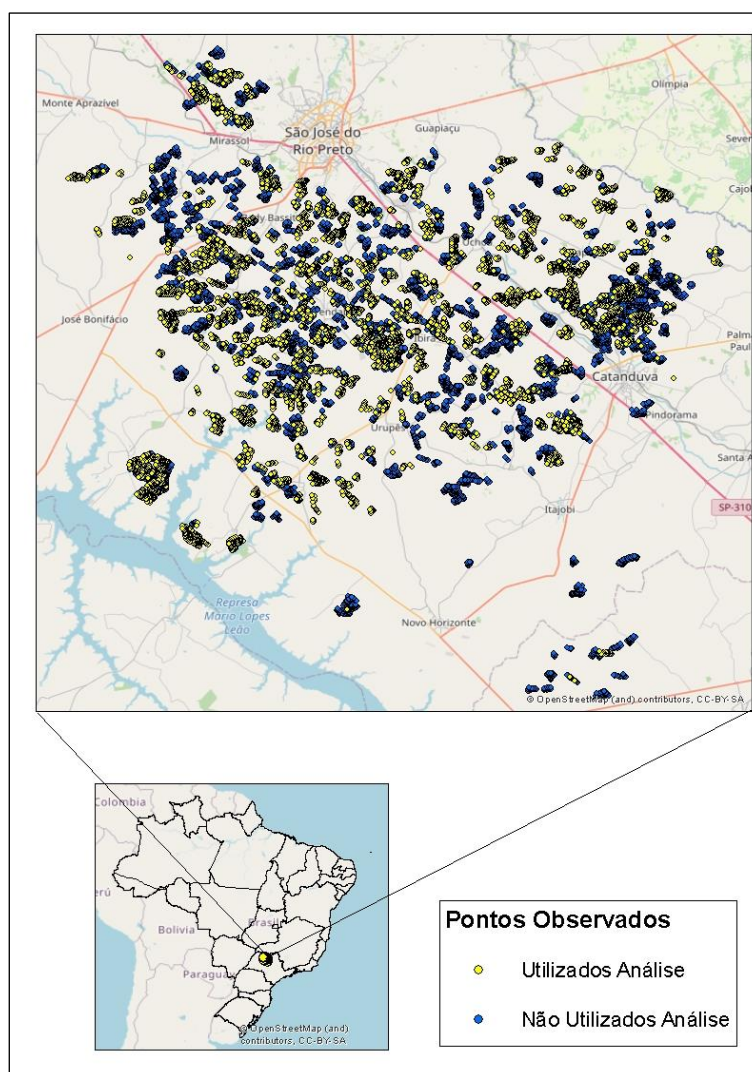


Figura 1 - Áreas de estudo localizadas nas intermediações dos municípios de Potirendaba e Catanduva, SP.

Como complemento, os dados referentes ao ano de 2016 foram divididos em 70% para o aprendizado de máquina (treinamento do modelo) e 30% para o teste (validação do modelo). Para o ano de 2018 foi realizado, também um treinamento, com os dados do conjunto 2016 mais 2017 e, realizou-se o mesmo procedimento relativo à 70% designado para o aprendizado de máquina e 30% para o teste. A Figura 2 representa a divisão entre as fases de treinamento e teste dos algoritmos. No eixo Y observa-se o número de observações consideradas e, no eixo X, as variáveis de manejo da cultura e os atributos químicos do solo para o ano de 2016.

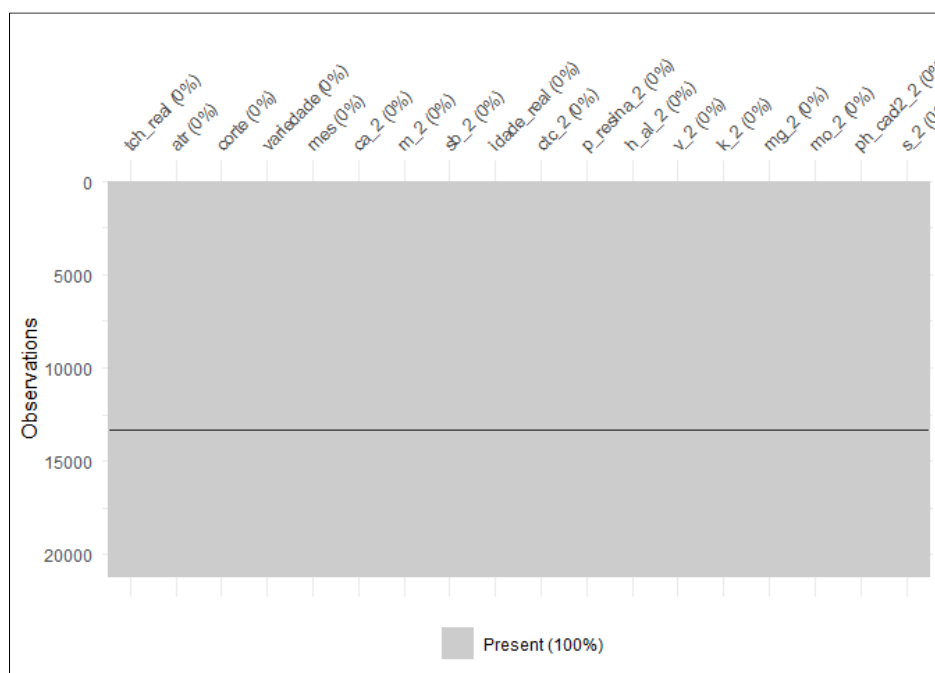


Figura 2. Variáveis de manejo da cultura e atributos químicos do solo observadas em área produtiva de cana-de-açúcar, na região de Catanduva e Potirendaba – SP em 2016.

As amostras de solo foram coletadas na densidade de 1 ponto a cada 3 hectares, nas camadas de 0,00 a 0,10 m e de 0,10 a 0,20 m, submetidas a análises laboratoriais para a determinação dos atributos químicos do solo:

matéria orgânica do solo (MOS), pH, fósforo (P), potássio (K⁺), cálcio (Ca²⁺), magnésio (Mg²⁺), acidez potencial (H + Al), soma de bases (SB), capacidade de troca catiônica (CTC) e saturação p bases (V). As amostras foram retiradas anualmente durante as safras de 2016 a 2018.

O clima da região foi classificado de acordo com o Sistema Internacional de Köppen, tropical úmido (Aw), com estações chuvosas no verão e seca no inverno com médias anuais de: precipitação 1.400 mm, temperatura de 22,3 °C e umidade relativa do ar entre 70 e 80%. A vegetação primária da região foi classificada como floresta pluvial estacional e cerrado, com principal uso agrícola da cultura da cana-de-açúcar, com mais de 10 anos em sistema de colheita mecanizada, no sistema cana crua. O solo das áreas experimentais foi classificado como Argissolo Vermelho-Amarelo eutrófico textura média/argilosa, conforme Santos et al. (2013).

3.2 Análise dos dados

O banco de dados foi composto por 20, 12 e 10 mil observações referentes aos anos de 2016, 2017 e 2018, respectivamente, totalizando 42 mil observações. O processamento de dados foi realizado mediante a padronização dos valores das variáveis no programa R utilizando a seguinte fórmula:

$$Z = \frac{x_i - \bar{x}}{\sigma}$$

em que Z é o conjunto de dados com média igual a 0 e desvio padrão igual a 1 e o x_i corresponde ao valor da amostra. Dessa forma, todas as variáveis seguem uma distribuição normal padrão. Após a padronização das variáveis foi utilizado o programa R para identificar os atributos do solo e manejos específicos

(variáveis independentes) que melhor explicaram a variabilidade espaço temporal do rendimento da cana-de-açúcar (TCH – tonelada de cana por hectare), por meio de algoritmos de árvore de decisão e *Random forest*.

4. RESULTADOS E DISCUSSÃO

As variáveis de manejo (número de cortes, idade, variedade e mês), bem como a variável ATR (açúcar total recuperável), que está no âmbito da qualidade, são as mais correlacionadas com a produtividade da cana-de-açúcar na área de estudo para o período observado de 2016 a 2018. Além disso, optou-se em escolher a segunda camada do solo analisada (0,10 a 0,20 m), como fonte de dados.

O modelo avaliado para o ano de 2016, foi satisfatório e, superior ao modelo utilizado para o ano de 2018, que considerou a previsão dos dois últimos anos para que seja realizada a predição tanto com árvore de decisão, quanto com o algoritmo *Random forest*. Porém para ambos os períodos da análise, houve redução de desempenho do modelo devido à influência de duas nuvens distintas de pontos observados, bem definidas, que podem representar as duas usinas produtoras de cana-de-açúcar em questão.

As Figuras 3 e 4, representam o processo de aprendizado do modelo, sendo que representam, respectivamente, os modelos de árvore de decisão e de *Random forest*, como ferramentas de previsão da produção da cultura para o ano de 2016. Nota-se que ambos os modelos possuem altos Coeficientes de Determinação (R^2), 0,84 para o modelo de árvore de decisão e, 0,93 para o de *Random forest*, resultando em satisfatória explicação de grande parte da

variabilidade dos dados. É possível observar, também, que o modelo de *Random forest* demonstrou menor dispersão dos pontos referentes aos dados observados quando comparado ao modelo de árvore de decisão.

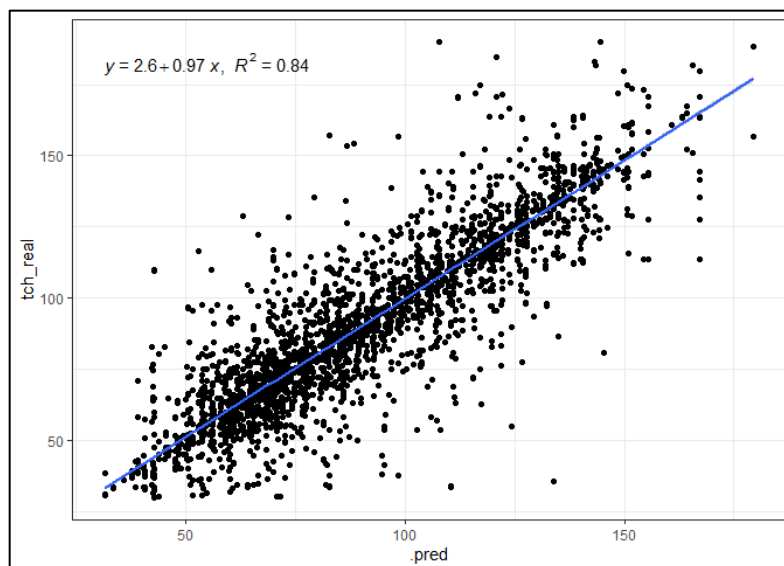


Figura 3. Processo de aprendizado do modelo de árvore de decisão utilizado para estimar produção de cana-de-açúcar nas regiões de Catanduba e Potirendaba – SP, em 2016.

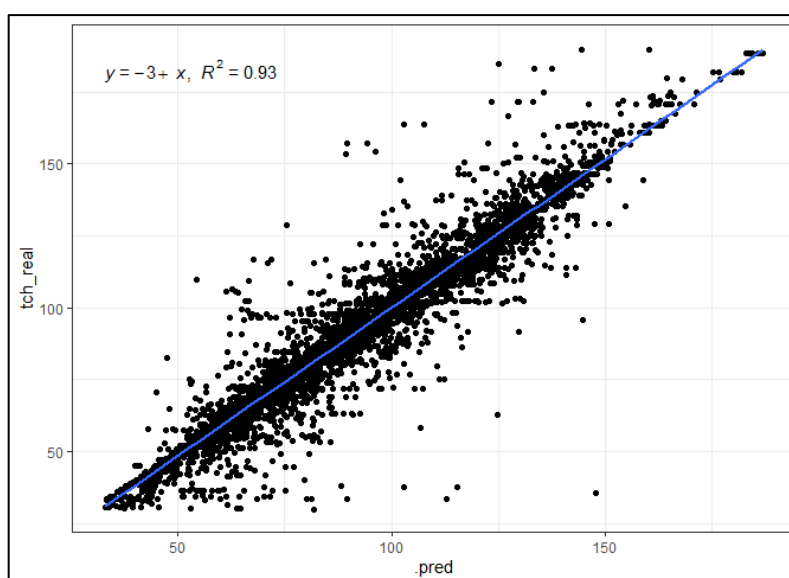


Figura 4. Processo de aprendizado do modelo de *Random forest* utilizado para estimar produção de cana-de-açúcar nas regiões de Catanduba e Potirendaba – SP, em 2016.

A Figura 5 representa a previsão dos dados reais referentes à produção em toneladas de cana por hectare (TCH) para o ano de 2017, obtida a partir dos dados de 2016. Observa-se uma semelhança com a distribuição normal e, portanto, define zona de produção média que está próxima de 100 toneladas de cana por hectare.

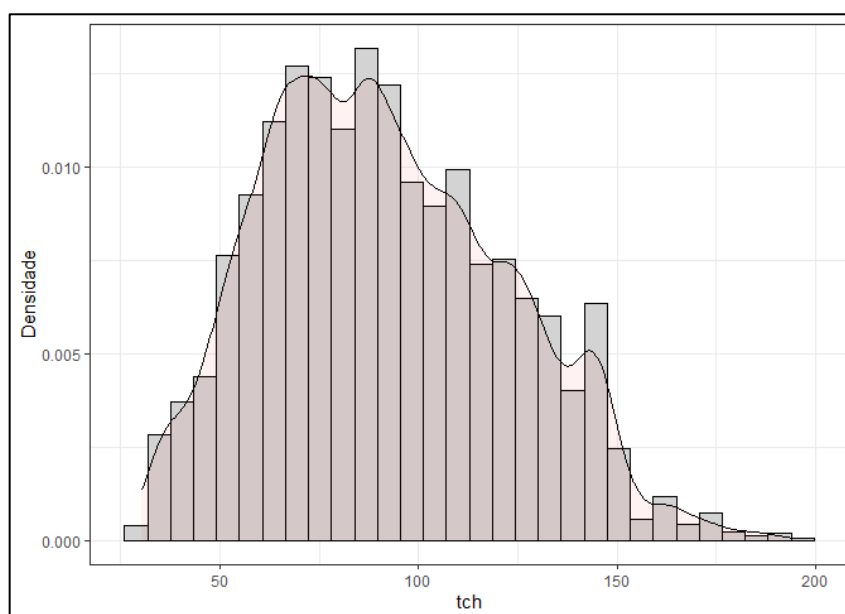


Figura 5. Previsão dos dados reais da produção em toneladas de cana-de-açúcar por hectare (TCH) nas regiões de Catanduva e Potirendaba-SP, em 2017.

As Figuras 6 e 7 representam o nível de importância de cada variável do modelo bem como a previsão real dos pontos observados, respectivamente, a partir dos dados de 2016, por meio da utilização do modelo de árvore de decisão. Nota-se que a variável corte (variável de manejo) apresenta-se, no mínimo, 3,5 vezes mais responsável por explicar a produção de TCH que outras variáveis presentes, como a variedade, idade do canavial, ATR e mês, como também, os atributos químicos do solo. Foi observado um poder de previsão R^2 não tão

expressivo, de aproximadamente 59 %, como também, um RMSE próximo de 21 t ha⁻¹, que pode ser explicado pela presença de duas nuvens de dados bem definidas, referentes às duas usinas produtoras de cana-de-açúcar no mesmo modelo utilizado.

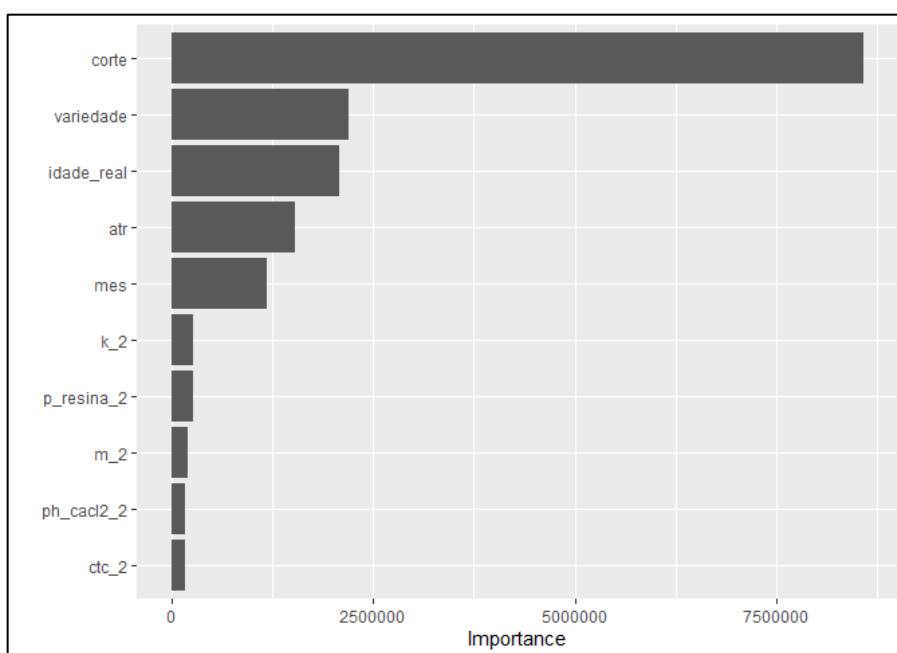


Figura 6. Nível de importância das variáveis para estimar a produção de cana-de-açúcar nas regiões de Catanduva e Potirendaba - SP pelo modelo de árvore de decisão. Atr = açúcar total recuperável, k_2 = nível de potássio na segunda camada do solo, p_resina_2 = nível de fósforo resina na segunda camada do solo, m_2 = acidez potencial na segunda camada do solo, ph_cacl2_2 = leitura do pH em cloreto de cálcio na segunda camada do solo, ctc_2 = capacidade de troca de cátions na segunda camada do solo.

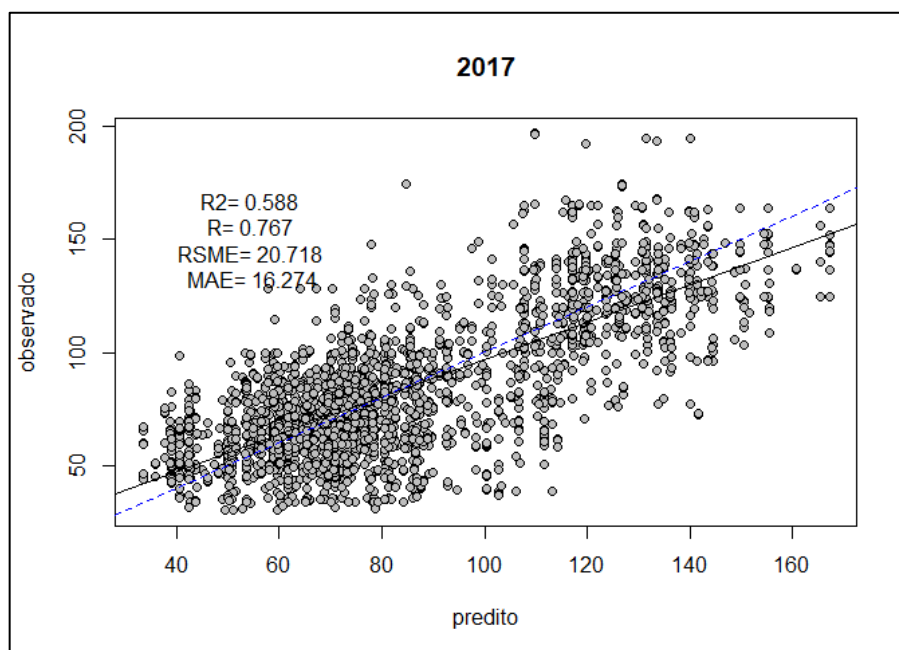


Figura 7. Previsão real dos pontos observados para estimar a produção de cana-de-açúcar, nas regiões de Catanduva e Potirendaba - SP pelo modelo de árvore de decisão.

As Figuras 8 e 9 representam a mesma situação, para o mesmo período, porém utiliza os algoritmos de *Random forest* para o estudo. Nota-se que o corte, para esse modelo, também possui importância semelhante ao resultado obtido a partir do modelo anterior, porém a idade do canavial, para esse caso, apresenta maior relevância para a previsão da cultura em relação a variedade, por exemplo. Em relação ao modelo anterior, nota-se um aumento significativo no R^2 , de aproximadamente 67%, como também redução no RMSE, de aproximadamente 18 t ha^{-1} e, menor dispersão dos pontos dispostos no gráfico.

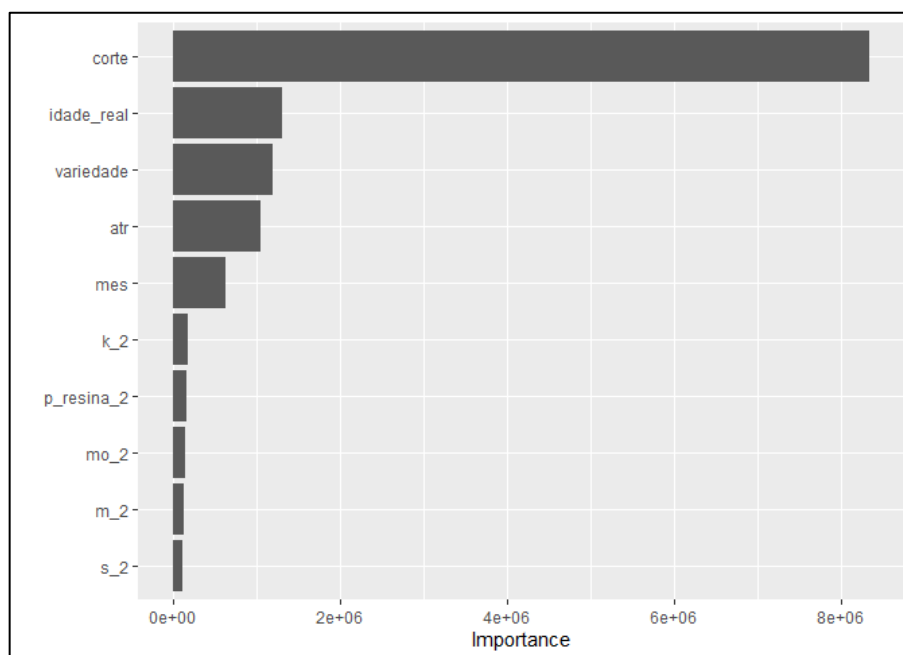


Figura 8. Nível de importância das variáveis utilizando algoritmos de *Random forest* para estimar a produção de cana-de-açúcar nas regiões de Catanduva e Potirendaba - SP. Atr = açúcar total recuperável, k_2 = nível de potássio na segunda camada do solo, p_resina_2 = nível de fósforo resina na segunda camada do solo, mo_2 = nível de matéria orgânica na segunda camada do solo, m_2 = acidez potencial na segunda camada do solo, s_2 = nível de enxofre na segunda camada do solo.

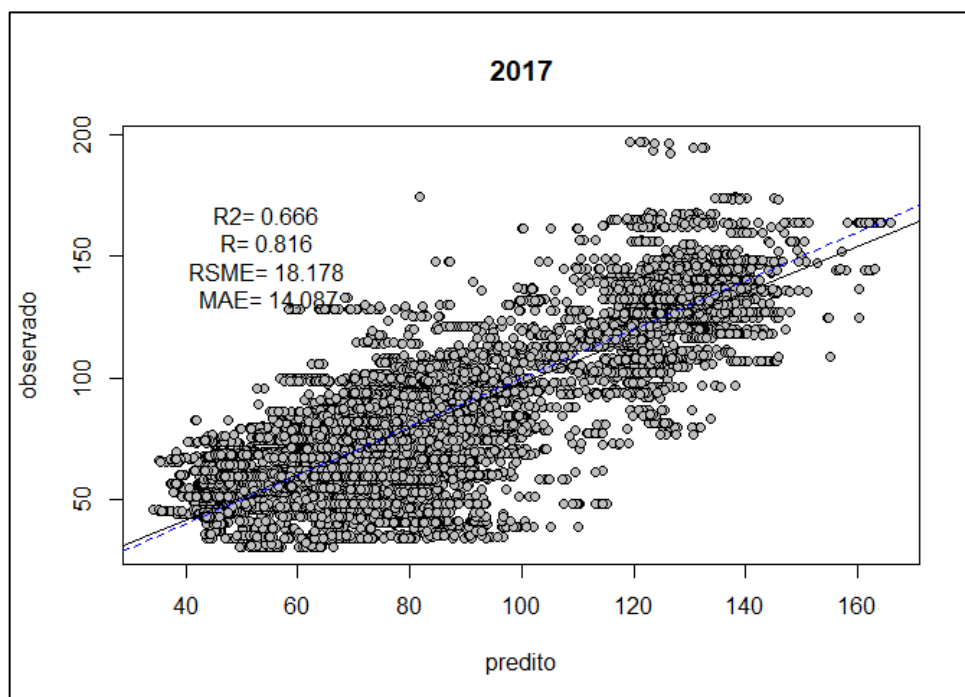


Figura 9. Previsão real dos pontos utilizando algoritmos de *Random forest* para estimar a produção de cana-de-açúcar, nas regiões de Catanduva e Potirendaba - SP.

Para o ano de 2018, também foram detalhadas as variáveis que obtiveram maiores importâncias na previsão da cultura e a previsão real dos pontos observados. As Figuras 10 e 11, representam a análise por meio de árvores de decisão, em que a grande diferença observada, tange à uma perda considerável na predição, segundo o R^2 de 43% e um erro RMSE de 23 t ha⁻¹, aproximadamente.

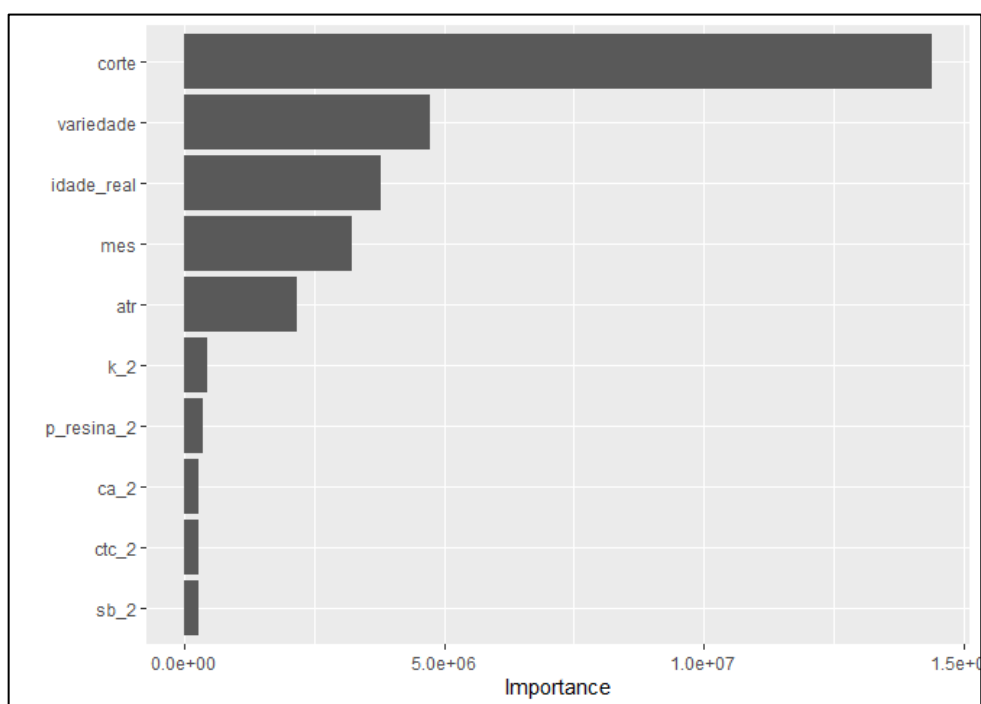


Figura 10. Nível de importância das variáveis utilizando a árvore de decisão para estimar a produção de cana-de-açúcar nas regiões de Catanduva e Potirendaba - SP.

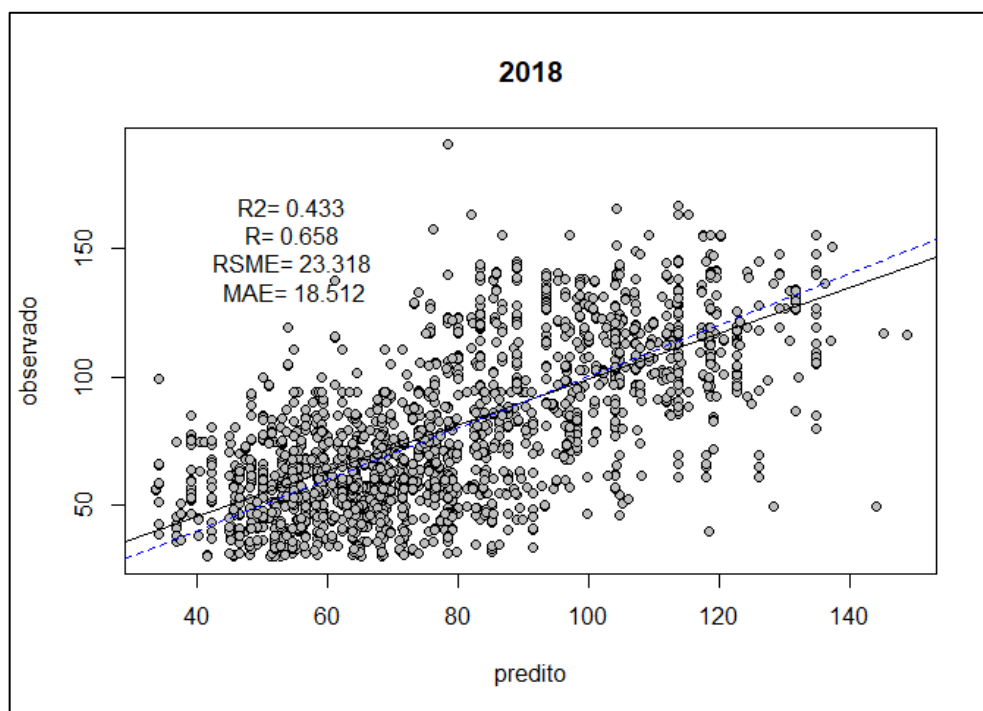


Figura 11. Previsão real dos pontos observados utilizando a árvore de decisão para estimar a produção de cana-de-açúcar, nas regiões de Catanduva e Potirendaba - SP.

Para o mesmo período, realizou-se o estudo com os algoritmos de *Random forest*. Nas Figuras 12 e 13, foi observado mudança na importância de algumas variáveis em relação à análise anterior, como por exemplo, maior contribuição da idade real e do ATR, como também, a participação da MOS, fator importante para explicar a produção da cultura. É visto também a mesma perda na capacidade de predição da produção do modelo, representada pelo R^2 de 56% e erro RMSE de 20, aproximadamente.

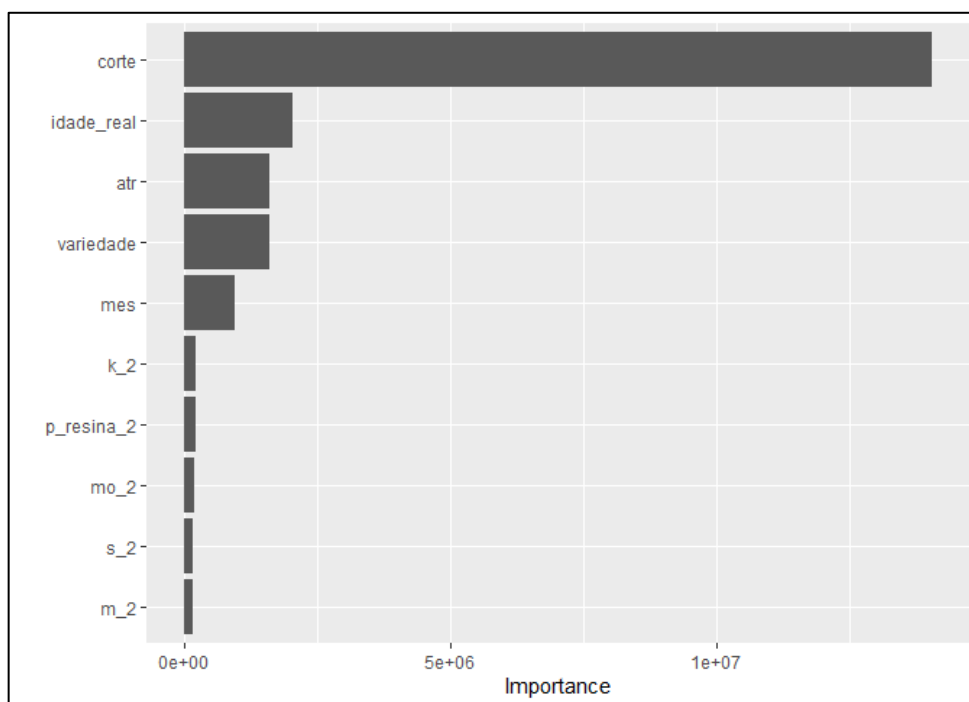


Figura12. Nível de importância das variáveis do modelo utilizando algoritmos de *Random forest* para estimar a produção de cana-de-açúcar nas regiões de Catanduba e Potirendaba - SP.

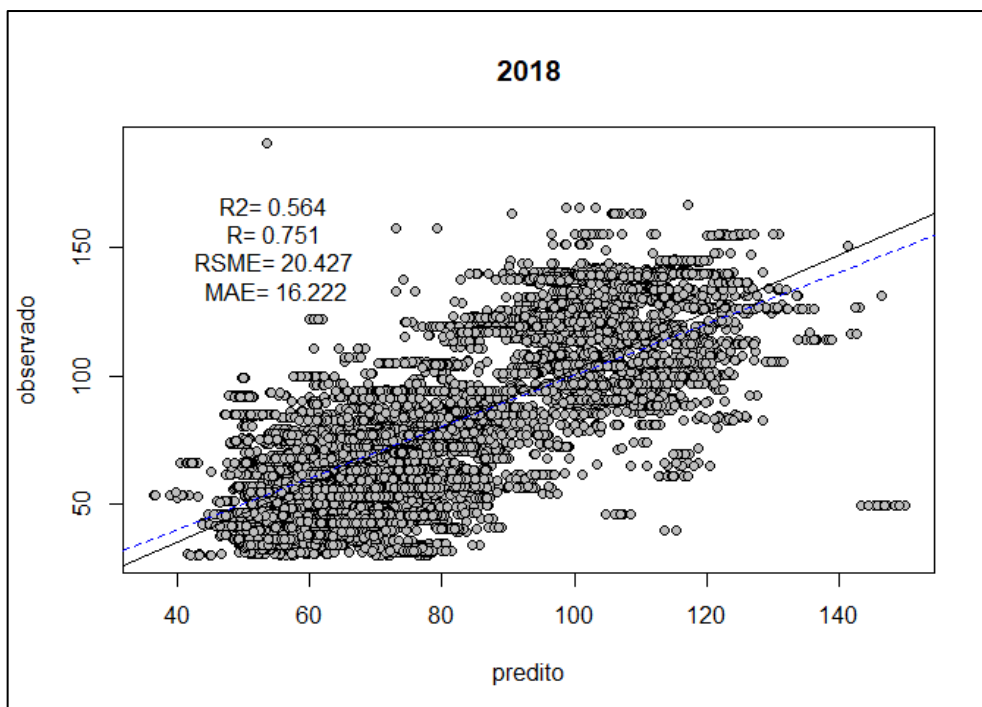


Figura 13. Previsão real dos pontos observados utilizando algoritmos de *Random forest* para estimar a produção de cana-de-açúcar, nas regiões de Catanduba e Potirendaba - SP.

5. CONCLUSÕES

Os resultados indicam que o algoritmo *Random forest* apresentou desempenho preditivo superior ao algoritmo de árvore de decisão para os anos avaliados (2016, 2017 e 2018).

Foi observado uma redução da eficiência preditiva de ambos os modelos utilizados para o ano de 2018, em comparação com os anos de 2016 e 2017. Estudos e análises devem ser feitos, porém, pode-se dizer que talvez seja necessário analisar cada ano separadamente, ou no máximo, considerar até um ano antecedente para a modelagem, como também, realizar a modelagem para cada usina separadamente, a fim de reduzir nuvens de dados distintas dispostas nos resultados.

Essa abordagem se mostrou promissora para suporte à tomada de decisão dentro das usinas e unidade produtoras, visto o bom funcionamento e desempenho dos modelos utilizados para treinamento. Contudo, para obtenção de modelos ainda mais assertivos, incentivamos que os novos estudos com enfoque na predição da produtividade de cana-de-açúcar, levem em consideração as informações de até no máximo 1 ano anterior.

Além disso, a melhor capacidade preditiva dos modelos gerados, a partir do banco de treinamento de um ano em comparação ao ano de 2016 e 2017, indica a grande influência do manejo da cultura no processo de modelagem, sendo necessário, portanto, uma análise frequente dos resultados, culminando em uma contínua atualização dos modelos de aprendizado. Assim faz-se necessário a existência de um setor atrelado e, responsável unicamente, para o

banco de dados, visto a notória complexidade e ampla utilização das análises de modelagem observadas, como também, a grande dependência de qual e quantos anos antecedentes estão sendo considerados para a modelagem.

6. REFERÊNCIAS

- ABERNETHY, M. **Mineração de dados com Weka, parte 2: Classificação e armazenamento em cluster.** Disponível em <<http://www.ibm.com/developerworks/br/opensource/library/os-weka2/>> Acesso em 13 mar. 2021.
- ALLEN, R.G.; PEREIRA, L.S.; RAES, D.; SMITH, M. **Crop evapotranspiration: guidelines for computing crop water requirements.** Rome: FAO Irrigation and Drainage Paper, 56, p. 300, 1998.
- AMARAL, M. A. C.; COELHO, R. D.; JOSÉ, J. V.; OLIVEIRA, A. S. **Temperatura base e taxa de crescimento de oito variedades de cana-de-açúcar,** v. 5, 2019.
- ANDERSON, D. L.; PORTIER, K. M.; OBREZA, T. A.; COLLINS, M. E.; PITTS, D. J. **Tree Regression Analysis To Determine Effects Of Soil Variability On Sugarcane Yields.** Soil Sci. Soc. Am. J., v. 63, n. 3, p. 592–600, 1999.
- AZEVÊDO, L.L.A. **Métodos estatísticos em aprendizado de máquinas para problemas de classificação.** Dissertação (Mestrado em Estatística) - Universidade de Brasília, 2018.
- BATISTA, G. H. A. P. A. **Pré-processamento de dados em aprendizado de máquina supervisionado.** 2003. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, University of São Paulo, São Carlos, 2003.
- BICALHO, E. S.; PANOSSO, A. R.; TEIXEIRA, D. D. B.; MIRANDA, J. G. V.; PEREIRA, G. T.; LA SCALA, N. **Spatial variability structure of soil CO₂ emission and soil attributes in a sugarcane area.** Agriculture Ecosystems & Environment, v. 189, n., p. 206-215, 2014.
- BISHOP, C. M. **Pattern Recognition and Machine Learning.** Springer Verlag, New York, 2006.
- BLACKBURN, F. **Sugarcane.** New York: Longman, p. 414, 1984.
- BOCCA, F. F. **Produtividade de cana-de-açúcar: caracterização dos contextos de decisão e utilização de técnicas de mineração de dados para modelagem.** Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Agrícola, Campinas, SP, 2014.

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, p. 67-76, 2016.

BOLSONELLO, J. E. K., FREDDO, A. R., GALLINA, A. L., ELIAS, F., STARIKOFF, K. R. **Aplicação do aprendizado de máquina sobre a qualidade da produção agroecológica de leite.** Anais da IX Jornada de Iniciação Científica e Tecnológica - IX JIC, v. 1 n. 9, 2019.

BRUCE, P.; BRUCE, A. **Practical Statistics for Data Scientists.** United States of America: O'Reilly Media, Inc., 562 p., 2017.

CALDAS, M. M.; GTANCO, G.; BISHOP, C.; KASTENS, J.; BROWN, J. C. **Effects of sugarcane ethanol expansion in the brazilian Cerrado: land use response in the new frontier.** In: QIN, Z.; MISHRA, U.; HASTINGS, A. (Ed.). **Bioenergy and Land Use Change.** Hoboken: J. Wiley, p.39-51, 2018.

CÂMARA, G.M.S. **Ecofisiologia da cultura da cana-de-açúcar.** In: Produção de cana-de-açúcar. Piracicaba: FEALQ, p. 31-64, 1993.

CARDOZO, N. P. **Modelagem da maturação da cana-de-açúcar em função de variáveis meteorológicas.** Dissertação (Mestrado em Ciências) – Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2012.

CARDOZO, N.P.; SENTELHAS, P.C. **Climatic effects on sugarcane ripening under the influence of cultivars and crop age.** Scientia Agricola, Piracicaba, v. 70, p. 449-456, 2013.

CASAGRANDE, A. A. **Crescimento da cana-de-açúcar.** Stab, Açúcar, Álcool e Subprodutos, Piracicaba, v.14, n.5, p.7-8, 1996.

CHIAVEGATTO FILHO, A.D.P.; DOS SANTOS, H.G.; DO NASCIMENTO, C.F.; MASSA, K.; KAWACHI, I. **Overachieving Municipalities in Public Health: A Machine-learning Approach.** Epidemiology, v. 29, n. 6, p. 836-840, 2018.

CHOLLET, F.; ALLAIRE, J.J. **Deep Learning with R,** 341 p, 2017.

COELHO, A. P.; BETTIOL, J. V. T.; DALRI, A. B.; FISCHER, J. A.; DE FARIA, R. T.; PALARETTI, L. F. **Application of artificial neural networks in the prediction of sugarcane juice Pol.** Revista Brasileira De Engenharia Agricola E Ambiental, v. 23, n. 1, p. 9-15, 2019.

CONAB, Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de cana-de-açúcar.** Brasília. Safra 2022/23. Acesso em: 05 fev. 2023.

CONSTANTINO, H. S. **Controle biológico de nematoides na cana-de-açúcar: revisão sistemática.** Dissertação para obtenção do título de Mestre em Agronomia, Universidade Estadual de Maringá, 2022.

CRAVERO, A.; PARDO, S.; SEPÚLVEDA, S.; MUÑOZ, L. **Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review.** Agronomy, p. 748, v. 12, 2022. DOI: <https://doi.org/10.3390/agronomy12030748>

DA CUNHA, J.M.; CAMPOS, M.C.C.; GAIO, D.C.; DE SOUZA, Z.M.; SOARES, M.D.R.; DA SILVA, D.M.P.; SIMÕES, E.L. **Spatial variability of soil respiration in Archaeological Dark Earth areas in the Amazon**. *Catena*, v. 162, n., p. 148-156, 2018.

DI MININ, E.; FINK, C.; TENKANEN, H.; HIIPPALA, T. **Machine learning for tracking illegal wildlife trade on social media**. *Nature Ecology & Evolution*, v. 2, n. 3, p. 406-407, 2018.

DOORENBOS, J.; KASSAM, A. H. **Yield response to water**. Rome: FAO, Irrigation and Drainage Paper, p. 179, 1979.

DOORENBOS, J.; KASSAM, A.H. **Yield response to water, Irrigation and Drainage Paper No. 33**. FAO, Rome, 1979.

EMBRAPA, Empresa Brasileira de Pesquisa Agropecuária. **Documentos 200, melhoramento genético de cana-de-açúcar**. Tabuleiros Costeiros. 2015. Acesso em: 03 fev. 2023.

EVERINGHAM, Y. L.; SMYTH, C. W.; INMAN-BAMBER, N. G. **Ensemble data mining approaches to forecast regional sugarcane crop production**. *Agricultural and Forest Meteorology*, v. 149, n. 3, p. 689-696, 2009.

EVERINGHAM, Y.; SEXTON, J.; SKOCAJ, D.; INMAN-BAMBER, G. **Accurate prediction of sugarcane yield using a *Random forest* algorithm**. *Agronomy for Sustainable Development*, v. 36, n. 2, p., 2016.

FANG, C.; MONCRIEFF, J.B.; GHOLZ, H.L.; CLARK, K.L. **Soil CO₂ efflux and its spatial variation in a Florida slash pine plantation**. *Plant and Soil*, Dordrecht, v. 205, n. 2, p. 135-146, 1998.

FAO Chapter 5. Sugar, In: FAO (Ed.), *OECD-FAO AGRICULTURAL OUTLOOK 2018-2027*: FAO, 2018. p.

FERNANDES, G. C. **Nutrição, produtividade e qualidade tecnológica da segunda soqueira de cana-de-açúcar coinoculada com bactérias promotoras de crescimento associada à adubação fosfatada**. Dissertação para obtenção do título de Mestre em Agronomia, Unesp Campus de Ilha Solteira, 2022.

GE, Z.; SONG, Z.; DING, S.X.; HUANG, B. **Data Mining and Analytics in the Process Industry: The Role of Machine Learning**. *IEEE Access*, v. 5, n., p. 20590-20616, 2017.

GRANCO, G.; CALDAS, M. M.; BERGTOLD, J. S.; SANT'ANNA, A. C. **Exploring the policy and social factors fueling the expansion and shift of sugarcane production in the Brazilian Cerrado**. *GeoJournal*, v. 82, p. 63-80, 2015.

HAMMER, R. G.; SENTELHAS, P. C.; MARIANO, J. C. Q. **Sugarcane Yield Prediction Through Data Mining and Crop Simulation Models**. *Sugar Tech*, v.22, p.216-225, 2020.

HAN, J.; KAMBER, M.; PEI, J. **Data mining concepts and techniques**. USA: Elsevier, 2012. HOSHGOFTAAR, GOLAWALA AND HULSE, 2007. **An empirical study of learning from imbalanced data using *Random forests***. In 19th IEEE international conference on tools with artificial intelligence, 2, pp. 310-317.

- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. Springer Verlag, New York, 2001.
- HAYKIN, S. **Redes neurais: princípios e prática**. Bookman, Porto Alegre, v.2, 2001.
- ISAACS, E. H.; SRIVASTAVA, R. M. **Applied geostatistics**. Nova York: Oxford University Press, p. 561, 1989.
- JÚNIOR, W. P. S.; SILVEIRA B.; NORONHA, K. **Utilização de IA para Controle de Pragas na Agricultura**. Curso de engenharia de software, Anápolis, 2022.
- KAAB, A.; SHARIFI, M.; MOBLI, H.; NABAVI-PELESARAEI, A.; CHAU, K. W. **Combined life cycle assessment and artificial intelligence for prediction of output energy and environmental impacts of sugarcane production**. Science of the Total Environment, v. 664, n., p. 1005-1019, 2019.
- KAISER, H. F. **The varimax criterion for analytic rotation in factor-analysis**. Psychometrika, v. 23, n. 3, p. 187-200, 1958.
- KING, N. J.; MUNGOMERY, R. W.; HUGUES, C. G. **Manual of cane growing**. New York, Elsevier, 1965.
- KUHN, M.; JOHNSON, K. **Applied predictive modeling**. 2013.
- LA SCALA, N.; MARQUES, J.; PEREIRA, G.T.; CORA, J.E. **Short-term temporal changes in the spatial variability model of CO₂ emissions from a Brazilian bare soil**. Soil Biology & Biochemistry, v. 32, n. 10, p. 1459-1462, 2000.
- LA SCALA, N.; PANOSSO, A.R.; PEREIRA, G.T.; GONZALEZ, A.P.; MIRANDA, J.G.V. **Fractal dimension and anisotropy of soil CO₂ emission in an agricultural field during fallow**. International Agrophysics, v. 23, n. 4, p. 353-358, 2009.
- LANDELL, M.G.A.; CAMPANA, M.P.; FIGUEIREDO, P.; SILVA, M.A. **Programa de melhoramento genético: estratégias para o desenvolvimento de novas variedades de cana-de-açúcar**. In: LANDELL M.G.A.; VASCONCELOS, A.C.M.; DINARDOMIRANDA, L.L. (Coords.) WORKSHOP, 2, Programa Cana IAC. Instituto Agrônomo de Campinas. p. 7-10. 2003a.
- LANDELL, M.G.A.; FIGUEIREDO, P.; VASCONCELOS, A.C.M. **O estado da arte da pesquisa em cana-de-açúcar na região Centro-sul do Brasil**. In: RIFIB, Catanduva, SP, p. 1-9, 2003b.
- LEITE, G.M.V.; ANDRADE, L.A.B.; GARCIA, J.C.; ANJOS, I.A. **Efeitos de fontes e doses de silicato de cálcio no rendimento agrícola e na qualidade tecnológica da cana-de-açúcar, cultivar SP80-1816**. Ciência e Agrotecnologia, Lavras, v. 32, n. 4, p. 1120-1125, 2008.
- LEON, E.; VARGAS, R.; BULLOCK, S.; LOPEZ, E.; PANOSSO, A.R.; LA SCALA JR, N. **Hot spots, hot moments, and spatio-temporal controls on soil CO₂ efflux in a water-limited ecosystem**. Soil Biology and Biochemistry, v. 77, n. 0, p. 12-21, 2014.
- LUCCHESI, A. A. SUGARCANE (In Brazilian). In: CASTRO, P. R. C.; KLUGE, R. A. (Ed.). **Ecophysiology of extractive crops: sugarcane, rubber, coconut, oil palm and olive**. Piracicaba: Cosmópolis Stoller do Brasil. v. 1, p. 13-45, 2001.

MANTOVANELLI, B.C.; CAMPOS, M.C.C.; ALHO, L.C.; SILVA, P.C.; SILVA, D.A.P.; CUNHA, J.M.; SILVA, D.M.P.; SOARES, M.D.R. **Distribuição espacial da emissão de CO₂ e atributos do solo sob campo nativo na região de Humaitá, Amazonas.** Sociedade & Natureza, v. 28, n. 2, 2016.

MARIN, F.R.; PELLEGRINO, G.Q.; ASSAD, E.D.; PINTO, H.S.; ZULLO JUNIOR, J. Cana-de-açúcar. In: MONTEIRO, J.E.B.A. **Agrometeorologia dos cultivos: o fator meteorológico na produção agrícola.** Brasília: INMET, p. 111-130, 2009.

MASSRUHÁ, S. M. F. S.; MOURA, M. F.; LEITE, M. A. de A. **Análise de tendências da produção técnico-científica em tecnologias avançadas no agronegócio.** In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 10., 2015, Ponta Grossa. Uso de VANTs e sensores para avanços no agronegócio: anais. Ponta Grossa: Universidade Estadual de Ponta Grossa, 2015.

MITROVIC, S.; ANDREOLETTI, D.; AYOUB, O. **ChatGPT or human? Detect and explain. Explaining decisions of Machine Learning model for detecting short chatGPT-generated text.** Dalle Molle Institute for Artificial Intelligence and Information Systems and Networking Institute, Switzerland, 2023.

MONTEIRO, L.A. **Modelagem agrometeorológica como base para a definição de ambientes de produção para a cultura da cana-de-açúcar no estado de SP.** p. 121, 2011. Dissertação (Mestrado em Física do Ambiente Agrícola) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2011.

NAJAFI, B.; FAIZOLLAHZADEH ARDABILI, S.; SHAMSHIRBAND, S.; CHAU, K.-W.; RABCZUK, T. **Application of ANNs, ANFIS and RSM to estimating and optimizing the parameters that affect the yield and cost of biodiesel production.** Engineering Applications of Computational Fluid Mechanics, v. 12, n. 1, p. 611-624, 2018.

NCHES, G. M.; GRAZIANO MAGALHÃES, P. S.; JUNQUEIRA FRANCO, H. C. **Site-specific assessment of spatial and temporal variability of sugarcane yield related to soil attributes.** Geoderma, v. 334, n., p. 90-98, 2019.

OLIVEIRA, A.R.; ROESLER, V; LOCHPE, C; SCHIMIDT, M.I.; VIGO, A; BARRETO, S.M. **Comparison of machine-learning algorithms to build a predictive model for detecting undiagnosed diabetes – ELSA-Brasil: accuracy study.** São Paulo Med J., Porto Alegre, 2017.

OLIVEIRA, M. P. G. D.; BOCCA, F. F.; RODRIGUES, L. H. A. **From spreadsheets to sugar content modeling: A data mining approach.** Computers and Electronics in Agriculture, v. 132, n., p. 14-20, 2017.

OpenAI. **ChatGPT.** Accessed: 2023-06-14. 2023. url: <https://openai.com/blog/chatgpt/> (cit. on pp. 1, 16).

PANOSSO, A.R.; MARQUES, J.; PEREIRA, G.T.; LA SCALA, N. **Spatial and temporal variability of soil CO₂ emission in a sugarcane area under green and slash-and-burn managements.** Soil & Tillage Research, v. 105, n. 2, p. 275-282, 2009.

PANOSSO, A.R.; PERILLO, L.I.; FERRAUDO, A.S.; PEREIRA, G.T.; MIRANDA, J.G.V.; LA SCALA JR, N. **Fractal dimension and anisotropy of soil CO₂ emission in a**

mechanically harvested sugarcane production area. Soil & Tillage Research, v. 124, n., p. 8-16, 2012.

PEREIRA, A.R.; ANGELOCCI, L.R.; SENTELHAS, P.C. **Agrometeorologia: fundamentos e aplicações práticas.** Guaíba: Ed. Agropecuária, p. 478, 2002. .

PHU NGUYEN; LEE, Y.C. **An Application of Support Vector Machines to Customer Loyalty Classification of Korean Retailing Company Using R Language 1.** The Journal of Information Systems, v. 26, n. 4, p. 17-37, 2017.

R DEVELOPMENT CORE TEAM. **A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

RODRIGUES, J.D. **Fisiologia da cana-de-açúcar.** Botucatu: UNESP, p. 99, 1995.

RUFATTO, L. **Mineração de dados aplicada na agricultura - análise de insumos e safra.** 2015. Trabalho de Conclusão de Curso (Tecnologia em Análise e desenvolvimento de Sistemas) - Universidade Tecnológica Federal do Paraná, Medianeira, 2015.

SALES, A. D. R. **Inteligência Artificial (IA) à luz da teoria da decisão: um estudo sobre a utilização da IA em decisões judiciais.** Editora Dialética, 2023.

SANTOS, H. G.; JACOMINE, P. K. T.; ANJOS, L. H. C.; OLIVEIRA, V. A.; OLIVEIRA, J. B.; COELHO, M. R.; LUMBRERAS, J. F.; CUNHA, T. J. F. **Sistema brasileiro de classificação de solos.** Rio de Janeiro: Embrapa Solos, p. 353, 2013.

SANTOS, H.G.D.; NASCIMENTO, C.F.D.; IZBICKI, R.; DUARTE, Y.A.D.O.; CHIAVEGATTO FILHO, A.D.P. **Machine learning para análises preditivas em saúde: exemplo de aplicação para prever óbito em idosos de São Paulo, Brasil.** Cadernos de Saúde Pública, v. 35, n.7, p.2, 2019.

SANTOS, R. P.; BEKO, M.; LEITHARDT, V. R. Q. **Modelo de Machine Learning em Tempo Real para Agricultura de Precisão.** Instituto Federal do Paraná, 2022.

SEGATO, S.V. et al. **Aspectos fenológicos da cana-de-açúcar.** In: SEGATO, S.V. Atualização em produção de cana-de-açúcar. Piracicaba, p.19-36, 2006.

SEXTON, J.; EVERINGHAM, Y.; DONALD, D.; STAUNTON, S.; WHITE, R. **A comparison of non-linear regression methods for improved on-line near infrared spectroscopic analysis of a sugarcane quality measure.** Journal of near Infrared Spectroscopy, v. 26, n. 5, p. 297-310, 2018.

Silva, C. A. M. A. **Previsão do valor Brix: aplicação de algoritmos de Machine Learning.** Defesa de Mestrado, Universidade de Lisboa, 2021.

SOUZA, A.P.; GASPAR, M.; SILVA, E.A.; ULIAN, E.C.; WACLAWOSKY, A.J.; NISHIYAMA JR., M.Y.; Santos, R.V.; TEIXEIRA, M.M.; SOUZA, G.M.; BUCKERIDGE, M.S. **Elevated CO₂ increases photosynthesis, biomass and productivity, and modifies gene expression in sugarcane.** Plant, Cell & Environment, v.31, p.1116-1127, 2008.

TARCA, A.L.; CAREY, V.J.; CHEN, X.W.; ROMERO, R.; DRAGHICI, S. **Machine learning and its applications to biology.** PLOS Comput, Biol. 3, e116, 2007.

TAVANTI, R.F.R.; MONTANARI, R.; PANOSSO, A.R.; LA SCALA, N.; CHIQUITELLI NETO, M.; FREDDI, O.D.S.; PAZ GONZÁLEZ, A.; DE CARVALHO, M.A.C.; SOARES, M.B.; TAVANTI, T.R.; GALINDO, F.S. **What is the impact of pasture reform on organic carbon compartments and CO₂ emissions in the Brazilian Cerrado?** *Catena*, v. 194, n., p. 104702, 2020.

TEIXEIRA, D.B.; BICALHO, E.S.; CERRI, C.E.P.; PANOSSO, A.R.; PEREIRA, G.T.; LA SCALA JR, N. **Quantification of uncertainties associated with space-time estimates of short-term soil CO₂ emissions in a sugar cane area.** *Agriculture, Ecosystems and Environment*, v. 167, n., p. 33-37, 2013.

USDA, Departamento de Agricultura dos Estados Unidos. **Açúcar: Mercados e Comércio Mundial.** Novembro, 2021. Acesso em 10 fev. 2023.

VEDANA, R.; RODRIGUES, K. C. T. T.; PARRÉ, J. L.; SHIKIDA, P. F. A. **Distribuição espacial da produtividade de cana-de-açúcar no Brasil.** *Revista de Política Agrícola*, ano28, p. 121-133, 2019.

Vemprala, S., Bonatti, R., Bucker, A., & Kapoor, A. Chatgpt for robotics: Design principles and model abilities. **Microsoft Auton. Syst. Robot. Res**, v. 2, p. 20, 2023.

WEBSTER, R.; OLIVER, M. A. **Geostatistics for Environmental Scientists.** Oxford University Press, p. 315, 2009.

WEBSTER, R.; OLIVER, M. A. **Statistical methods in soil and land resource survey.** New York: Oxford University Press, p. 328, 1990.

WILEY, J. F. **R Deep Learning Essentials.** Birmingham, UK: Packt Publishing Ltda, 2016.

WOLFERT, S.; GE, L.; VERDOUW, C.; BOGAARDT, M.-J. **Big Data in Smart Farming – A review.** *Agricultural Systems*, v. 153, n., p. 69-80, 2017.

XU, M.; QI, Y. **Soil-surface CO₂ efflux and its spatial and temporal variations in a Young ponderosa pine plantation in northern California.** *Global Change Biology*, Oxford, v. 7, n. 6, p. 667-677, 2001.