MESTRADO PROFISSIONAL

EM ECONOMIA

EXPLORANDO O POTENCIAL DO MACHINE LEARNING NA PRODUÇÃO AGRÍCOLA BRASILEIRA: UMA ANÁLISE EXPERIMENTAL DA PRECISÃO DE ALGORITMOS NA PREVISÃO DA PRODUÇÃO DE MILHO, SOJA, CAFÉ E CANA-DE-AÇÚCAR

RAFAEL VIANA VALLE LINS DE ALBUQUERQUE

Brasília-DF, 2025

RAFAEL VIANA VALLE LINS DE ALBUQUERQUE

EXPLORANDO O POTENCIAL DO MACHINE LEARNING NA PRODUÇÃO AGRÍCOLA BRASILEIRA:
UMA ANÁLISE EXPERIMENTAL DA PRECISÃO DE ALGORITMOS NA PREVISÃO DA PRODUÇÃO DE MILHO, SOJA, CAFÉ E CANA-DE-AÇÚCAR

Dissertação apresentada ao Programa de Pós Graduação em Economia, do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, como requisito parcial para obtenção do grau de Mestre.

Orientador
Professor Doutor Mathias Schneid Tessmann

Brasília-DF 2025

RAFAEL VIANA VALLE LINS DE ALBUQUERQUE

EXPLORANDO O POTENCIAL DO MACHINE LEARNING NA PRODUÇÃO AGRÍCOLA BRASILEIRA:
UMA ANÁLISE EXPERIMENTAL DA PRECISÃO DE ALGORITMOS NA PREVISÃO DA PRODUÇÃO DE MILHO, SOJA, CAFÉ E CANA-DE-AÇÚCAR

Dissertação apresentada ao Programa de Pós Graduação em Economia, do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, como requisito parcial para obtenção do grau de Mestre.

Aprovado em 30 / 05 / 2025

Banca Examinadora

Prof. Dr. Mathias Schneid Tessmann - Orientador

Prof. Dr. Danny de Castro Soares

Prof. Dr. Marcelo de Oliveira Passos

A345e Albuquerque, Rafael Viana Valle Lins de
Explorando o potencial do machine learning na produção agrícola brasileira: uma análise experimental da precisão de algoritmos na previsão da produção de milho, soja, café e cana-de-açúcar/ Rafael Viana Valle Lins de Albuquerque. – Brasília: Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, 2025.

51 p.
Inclui bibliografia.

Dissertação – Instituto Brasileiro de
Ensino, Desenvolvimento e Pesquisa – IDP, Curso de Mestrado Profissional em Economia, Brasília, 2025.
Orientador: Prof. Dr. Mathias Schneid Tessmann.

1. Inteligência Artificial. 2. Machine Learning. 3. Alocação de recursos. I. Título.

CDD 338.5443

Ficha catalográfica elaborada pela Biblioteca Ministro Moreira Alves
Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa

RESUMO

Este estudo investiga o potencial de algoritmos de Machine Learning (ML) para prever com precisão a produção agrícola brasileira, utilizando modelos treinados com dados climáticos, ambientais e históricos de produção, com foco nas culturas de café, cana-de-açúcar, soja e milho. Além disso, compara o desempenho de algoritmos como Random Forests, Gradient Boosting, redes neurais e Regressão Linear, com o objetivo de realizar uma análise comparativa dos métodos de ML frente a modelos lineares tradicionais. O trabalho também avalia o uso de ferramentas open-source, explorando possibilidades de democratização do acesso a soluções baseadas nessa abordagem. Os resultados sugerem que os algoritmos de ML podem oferecer algumas vantagens em relação aos modelos tradicionais, particularmente na identificação de relações mais complexas entre variáveis. Embora tenham apresentado indícios de maior precisão preditiva, é importante interpretar esses achados com cautela, considerando as limitações do estudo e a necessidade de validações adicionais. Ademais, os insights fornecidos podem subsidiar a formulação de políticas públicas, orientando decisões sobre alocação de recursos, estruturação de incentivos e estratégias de mitigação das mudanças climáticas.

Palavras-chave: Aprendizado de Máquina, Agricultura, Inteligência Artificial, Big Data Agrícola.
Classificação JEL: Q16, C45, C53, Q12, O33

ABSTRACT

This study investigates the potential of Machine Learning (ML) algorithms to accurately predict Brazilian agricultural production, using models trained with climatic, environmental and historical production data, focusing on coffee, sugarcane, soybean and corn crops. It also compares the performance of algorithms such as Random Forests, Gradient Boosting, Neural Networks and Linear Regression, with the aim of carrying out a comparative analysis of ML methods against traditional linear models. The project also evaluates the use of open-source tools, exploring possibilities for democratizing access to solutions based on this approach. The results suggest that ML algorithms may offer some advantages over traditional models, particularly in identifying more complex relationships between variables. Although they showed signs of greater predictive accuracy, it is important to interpret these findings with caution, considering the limitations of the study and the need for further validation. Furthermore, the insights provided can support the formulation of public policies, guiding decisions on resource allocation, incentive structuring and climate change mitigation strategies.

JEL Classification: Q16, C45, C53, Q12, O33

LISTA DE ABREVIATURAS E SIGLAS

IBGE	Instituto Brasileiro de Geografia e Estatística
Inmet	Instituto Nacional de Meteorologia
MAE	Mean Absolute Error
ML	Machine Learning
PAM	Produção Agrícola Municipal
RMSE	Root Mean Square Error
Sidra	Sistema IBGE de Recuperação Automática
UF	Unidade Federativa

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO


1

1 INTRODUÇÃO

A agricultura desempenha um papel vital na economia brasileira, contribuindo significativamente para o abastecimento interno e a exportação de produtos. No entanto, a eficiência e a produtividade desse setor estão sujeitas a uma série de desafios, como as mudanças climáticas, a variabilidade nas condições de cultivo e a dinâmica do mercado global. Nesse contexto, a capacidade de prever com precisão a produção agrícola torna-se essencial não apenas para os produtores, mas também para o planejamento estratégico econômico e para formulação de políticas públicas voltadas para o desenvolvimento sustentável do agronegócio brasileiro.

O Brasil, reconhecidamente um dos maiores e mais importantes produtores agropecuários do mundo, apresenta, além de uma participação massiva na economia agrícola internacional, um enorme potencial para a ampliação de sua produção (Saath e Fachinello, 2018), mas para fazer jus a esse potencial de forma sustentável, é necessária uma sólida base de análises e planejamentos precisos.

Os avanços tecnológicos no campo da inteligência artificial e ciência de dados democratizaram o acesso a algoritmos e modelos avançados de estatística computacional e abriram portas para o desenvolvimento e aplicação dessas ferramentas em contextos variados, tendo obtido resultados muito satisfatórios no setor agropecuário (em especial na produção de grãos) de países como Estados Unidos (Shahhosseini, M., Hu, G., Huber, I. et al) e Índia (Rashid et al., 2021).

As técnicas avançadas de machine learning oferecem uma oportunidade única para aprimorar a previsão da produção agrícola e, por consequência, auxiliar na tomada de decisões por parte dos produtores e na elaboração de políticas orientadas por dados. Ao integrar grandes volumes de dados históricos de safras, mercado, informações meteorológicas, e variáveis socioeconômicas, é possível desenvolver modelos preditivos robustos capazes de identificar padrões e tendências relevantes para esse tão importante setor da economia brasileira.

Este trabalho explora o potencial do machine learning na previsão da produção agrícola no Brasil, com o objetivo de fornecer insights valiosos tanto para o produtor como para a formulação e implementação de estratégias direcionadas ao setor agrícola. Por meio da análise integrada de múltiplas fontes de dados da produção brasileira (milho, café, cana-de-açúcar e soja), foi desenvolvido e comparado o desempenho de modelos preditivos que possam contribuir para a gestão mais eficiente dos recursos agrícolas, a mitigação de riscos e a promoção da sustentabilidade ambiental e econômica no contexto rural brasileiro.

A produção agrícola, especialmente de culturas como o milho e a soja, desempenha um papel crucial na economia brasileira. No entanto, além de ser um importante produtor agrícola mundial, alimentando mais de meio bilhão de pessoas no mundo apenas com sua produção de grãos (Contini e Aragão,2021), o Brasil também carrega a grande responsabilidade de proteger o seu inestimável patrimônio ambiental e simultaneamente reparar suas profundas feridas sociais. Equilibrar essas responsabilidades em um território tão vasto e diverso é uma tarefa complexa por si só, e demanda ainda mais atenção quando se leva em consideração o paradoxo de um país que mesmo produzindo alimento para bilhões de pessoas, não consegue ainda garantir a segurança alimentar de uma parcela considerável da própria população.

É possível aliar o crescimento produtivo e econômico com a preservação dos bens naturais e uma alimentação digna para todos, mas para isso é preciso encontrar soluções que permitam difundir a compreensão e maximizar a eficiência dos meios produtivos de forma a reduzir a utilização de ações tradicionais mais agressivas, como o uso desenfreado de agrotóxicos e o desmatamento de novas terras.

A revolução digital que estamos vivendo traz consigo ferramentas e tecnologias que nos permitem desenvolver soluções disruptivas para os mais diversos desafios humanos. Dentre as tecnologias emergentes mais disruptivas, os métodos de aprendizado de máquina e inteligência artificial tem se destacado com excelentes resultados. O uso dessas técnicas na agricultura, como para a previsão de produção e quantificação de fatores influentes, encontrou sucesso em países de diferentes contextos socioeconômicos e geográficos.

No entanto, prever com precisão a produção agrícola é um desafio devido à complexidade dos fatores envolvidos. Os algoritmos de aprendizado de máquina tem emergido nas últimas décadas como uma ferramenta para esse e outros fins, mas ainda há muito o que se explorar sobre a aplicação dessas tecnologias em meio as peculiaridades do contexto brasileiro. Outro fator a ser considerado é a viabilidade da democratização dessas soluções, para garantir igualdade tecnológica a pequenos e médios produtores. Nesse contexto, surge a questão: até que ponto a aplicação de algoritmos de *machine learning* pode melhorar a precisão das previsões de produção agrícola no Brasil, e como esses algoritmos se comparam em termos de eficiência?

O trabalho validou a hipótese de que algoritmos de *machine learning* podem prever com precisão a produção agrícola brasileira, por meio da construção e treinamento de modelos a partir de dados climáticos, ambientais e históricos na produção nacional de soja e milho. Também compara o desempenho de modelos criados em diferentes algoritmos, de modo a estimar os algoritmos mais adequados ao contexto do cenário descrito. Além disso, ao longo do desenvolvimento aplicou-se o uso de ferramentas e linguagens *open source*, de modo a averiguar possibilidades de acesso democrático às soluções baseadas nesse tipo de abordagem.

A relevância empírica deste trabalho reside na aplicação de técnicas avançadas de ML na previsão da produção agrícola, especificamente no contexto da cultura de grãos no Brasil. Com o aumento da demanda global por alimentos e a necessidade de uma agricultura mais eficiente e sustentável, exploram-se aqui novas abordagens que podem aprimorar a precisão das previsões agrícolas.

A contribuição pessoal deste estudo está na oportunidade de adquirir conhecimentos avançados em ML aplicada à agricultura, além de desenvolver habilidades analíticas e de pesquisa. Academicamente, este trabalho busca contribuir com o acervo na literatura da área interdisciplinar de contato da agronomia, economia e ciência de dados, gerando novas perspectivas sobre a eficácia das técnicas de ML na agricultura brasileira.

Além disso, ao impulsionar a eficiência na previsão da produção agrícola, o trabalho pode contribuir para a estabilidade econômica e o desenvolvimento rural, beneficiando diretamente agricultores, consumidores e toda a cadeia produtiva do setor agrícola.

Assim, o presente trabalho investiga a utilização do ML para previsão de produção agrícola. Para isso, é mensurada a eficácia dos algoritmos de *random forests*, *gradient boosting*, e redes neurais profundas (DNN) em comparação com a Regressão Linear na previsão das produções de milho, soja, café e cana-de-açúcar no Brasil. Os resultados sugerem que os algoritmos apresentam um desempenho superior e um grande potencial de aplicação.

Este trabalho está estruturado em quatro seções, além desta introdução. A próxima seção apresenta o referencial teórico; a seção três detalha a metodologia, incluindo os dados e os modelos utilizados; a seção quatro expõe e discute os resultados; e a seção cinco traz as conclusões e implicações práticas.


2

2 REFERENCIAL TEÓRICO

Iniciando com os trabalhos anteriores que exploraram a aplicação de ML em contextos semelhantes, Saaed e Wang (2019), utilizaram modelos para previsão da produção de lavouras em condados norte-americanos, detalhando o funcionamento do algoritmo Random Forest. Este algoritmo utilizou amostragem aleatória de dados para construir múltiplas árvores de decisão, cujos resultados foram combinados para gerar previsões robustas. Avaliaram a eficácia do algoritmo RF na previsão da produção de trigo, milho e batata. O artigo concluiu que o Random Forest foi útil para a previsão de produtividade agrícola, oferecendo precisão e robustez em diferentes escalas. A técnica pode apoiar a formulação de políticas agrícolas e melhorar a resiliência do setor a eventos climáticos extremos, e é um dos métodos explorados no presente trabalho.

Franco (2019) apresentou um sistema de predição baseado em um dispositivo embarcado, analisando a viabilidade de utilizar algoritmos de aprendizagem para realizar o processamento dos dados sem a necessidade de um computador. Isso lhe permitiu obter a distribuição de sistemas independentes capazes de diagnosticar ocorrências e condições das máquinas em campo.

Xiong et al. (2015) apresentaram um método combinado para previsão intervalar dos preços futuros de commodities agrícolas. O objetivo foi melhorar a precisão e a confiabilidade das previsões de preços, fundamentais para a gestão de riscos e a tomada de decisões em mercados agrícolas. A abordagem utilizou técnicas de aprendizado de máquina e modelos estatísticos para gerar previsões em forma de intervalos, considerando as incertezas inerentes aos mercados futuros. A previsão intervalar forneceu uma faixa provável para os preços futuros em vez de valores únicos, o que permite aos tomadores de decisão compreenderem melhor os riscos e incertezas das previsões. O método utilizoi uma combinação de técnicas de aprendizado de máquina, como redes neurais e máquinas de suporte vetorial (SVM), com métodos estatísticos tradicionais, como modelos autorregressivos. Os autores concluíram que a previsão intervalar com métodos combinados de ML consistiu em uma abordagem eficaz para lidar com as incertezas nos preços futuros de commodities agrícolas. A técnica

ofereceu uma ferramenta prática de planejamento e mitigação de riscos para gestores, investidores e formuladores de políticas que operam em mercados voláteis.

Lima et al. (2018) exploraram a avaliação de modelos de predição e previsão construídos por algoritmos de aprendizado de máquina em lote e em fluxo de dados. O estudo apresentou o processo de avaliação de modelos para previsão e predição, construídos utilizando algoritmos de aprendizado de máquina, destacando a importância da sistematização do processo de avaliação de modelos no contexto de cidades inteligentes.

Van Klompenburg, Kassahun e Catal (2020), em uma revisão sistemática, abordaram o potencial da previsão de produtividade agrícola com ML na Índia. Analisaram as principais abordagens, métodos e desafios relacionados ao uso de aprendizado de máquina nesse campo, fornecendo uma visão abrangente das tendências e lacunas de pesquisa. Apontaram que variáveis como clima, manejo do solo, práticas agrícolas e características das plantas foram fundamentais para as previsões. Destacaram também que combinações de dados heterogêneos, como imagens de satélite, sensores IoT e dados históricos, podem ser cada vez mais integradas para melhorar a qualidade das previsões. Os autores concluíram que o aprendizado de máquina tem potencial para transformar a previsão de produtividade agrícola, permitindo maior precisão e eficiência no planejamento agrícola. Contudo, ressaltaram que a plena realização desse potencial depende de avanços na coleta de dados, melhoria na interpretabilidade dos modelos e soluções que abordem a variabilidade regional e climática.

Silva et al. (2020) analisaram a aplicação de técnicas de ML na engenharia de produção. O estudo explorou as bases do ENEGEP, SIMPEP, ConBRepro e SBPO para quantificar e analisar os artigos que utilizaram técnicas de ML em áreas da engenharia de produção nos últimos cinco anos, identificando tendências e lacunas na literatura.

Silva (2020) analisou a aplicação de ML na previsão de ações da B3, explorando a predição de movimentos nos preços do mercado de ações e destacando a importância de técnicas de ML na análise de séries temporais financeiras.

Pinheiro et al. (2021) exploraram a aplicação de ML no setor agrícola, com ênfase no setor sementeiro, discutindo como tecnologias

emergentes estão transformando práticas tradicionais e otimizando processos produtivos. A pesquisa destacou que a agricultura enfrenta desafios globais, como aumento da demanda por alimentos, limitação de recursos naturais e necessidade de sustentabilidade, os quais podem ser mitigados com o uso de modelos de ML.

Nosratabadi et al. (2020) propuseram modelos híbridos de ML, combinando redes neurais artificiais com algoritmos de otimização, para prever a produtividade agrícola com maior precisão. O modelo ANN-GWO apresentou melhor desempenho na predição de rendimentos agrícolas.

Bassine et al. (2023) realizaram uma revisão crítica sobre as aplicações recentes de ML, sensoriamento remoto e Internet das Coisas (IoT) na predição de safras, destacando como essas tecnologias podem melhorar a eficiência e a sustentabilidade das práticas agrícolas.

Cunha, Silva e Avegliano (2023) propuseram uma abordagem abrangente para a previsão de produtividade que combina modelos baseados em inteligência artificial e simulação de culturas. Os autores desenvolveram uma solução para calibrar modelos de simulação em larga escala, criando um modelo substituto que assegura execuções mais rápidas sem comprometer a precisão.

Gupta et al. (2023) implementaram seis modelos de regressão para prever a produtividade agrícola em 37 países em desenvolvimento ao longo de 27 anos, utilizando dados multivariados. O estudo identificou que o modelo de regressão Random Forest apresentou um coeficiente de determinação (r^2) de 0,94, indicando alta precisão preditiva.

Pathak et al. (2023) analisaram diferentes modalidades de entrada e modelos de ML na previsão de produtividade agrícola em níveis de campo e subcampo. O estudo destacou a importância da granularidade dos dados no aumento da precisão preditiva.

Kallenberg et al. (2023) investigaram a predição de produtividade de batata utilizando uma abordagem híbrida de meta-modelagem. O estudo combinou modelos de crescimento de culturas com redes neurais convolucionais, resultando em previsões mais precisas quando comparadas a abordagens puramente baseadas em dados.

Souza e Oliveira (2024) realizaram uma análise comparativa das técnicas de ML, como Regressão Linear, Árvores de Decisão e Support Vector Machines. O estudo apresentou uma visão geral de cada algoritmo, descrevendo seus princípios fundamentais, vantagens e desvantagens, e os tipos de problemas para os quais são mais adequados.

Gomes (2024) discute como algoritmos de Deep Learning estão sendo empregados na identificação precoce de ameaças, permitindo intervenções rápidas e eficazes no manejo agrícola.

Maazallahi et al. (2024) analisaram a predição de produtividade agrícola na Índia, de 1997 a 2020, focando em diversos cultivos e fatores ambientais chave. Os modelos de ML, particularmente Naïve Bayes e Random Forest, demonstraram alta eficácia na previsão de rendimentos agrícolas.

Tabela 1 – Revisão de Literatura

Autores	Metodologia	Objeto do Estudo	Conclusões
Xiong et al. (2015)	ML + modelos estatísticos	Previsão de preços futuros de commodities	Método combinado melhora a confiabilidade das previsões em mercados voláteis
Lima et al. (2018)	ML em lote e fluxo	Avaliação de modelos preditivos	Importância da sistematização no processo de avaliação de modelos
Franco (2019)	Dispositivos embarcados com ML	Diagnóstico de máquinas agrícolas	Viabilidade de análise independente sem computador central
Saaed e Wang (2019)	Random Forest	Previsão de produção agrícola em lavouras norte-americanas	Alta precisão e robustez; útil para políticas agrícolas
Silva et al. (2020)	Revisão bibliográfica	Aplicações de ML na engenharia de produção	Identificação de tendências e lacunas na literatura
Silva (2020)	Modelos de ML	Previsão de ações da B3	Relevância de ML para séries temporais financeiras

Nosratabadi et al. (2020)	ML híbrido (ANN + otimização)	Previsão de produtividade agrícola	Modelo ANN-GWO teve melhor desempenho
Van Klompenburg et al. (2020)	Revisão sistemática	Previsão de produtividade agrícola na Índia	ML tem alto potencial; desafios em coleta e variabilidade de dados
Pinheiro et al. (2021)	ML + Visão computacional	Setor agrícola (sementes)	Transformação de práticas tradicionais e promoção da sustentabilidade
Bassine et al. (2023)	Revisão crítica (ML, IoT, sensoriamento remoto)	Predição de safras	Tecnologias promovem eficiência e sustentabilidade
Cunha et al. (2023)	IA + simulação de culturas	Previsão de produtividade agrícola	Modelo substituto eficiente e preciso
Gupta et al. (2023)	Modelos de regressão	Previsão em 37 países	"Random Forest com r^2=0. 94, alta precisão preditiva"
Pathak et al. (2023)	ML com diferentes entradas	Previsão agrícola em diferentes escalas	Granularidade dos dados aumenta precisão
Kallenberg et al. (2023)	Meta-modelagem híbrida	Produtividade de batata	Combinação com redes neurais aumenta precisão
Souza e Oliveira (2024)	Comparação de técnicas de ML	Análise de algoritmos ML	Fornece visão geral das abordagens e modelos de ML
Gomes (2024)	Deep Learning	Detecção precoce de pragas/doenças	Permite intervenções rápidas e eficazes
Maazallahi et al. (2024)	Naïve Bayes e Random Forest	Produtividade agrícola na Índia (1997–2020)	Alta eficácia em diversas condições

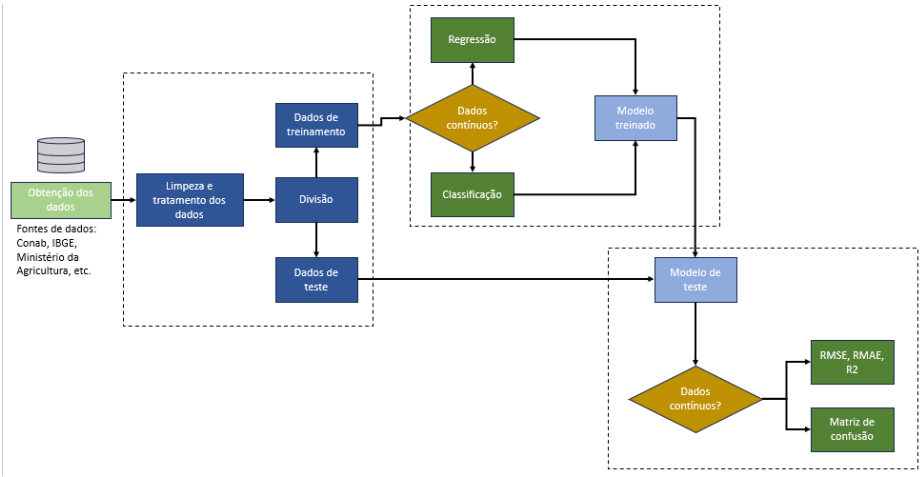
Fonte: Elaborado pelo autor.


3

3 METODOLOGIA

Nesta seção, serão detalhados os métodos e procedimentos que serão empregados no estudo para avaliar a eficiência de machine learning na previsão da produção de café, cana-de-açúcar, milho e soja no Brasil, incluindo a construção, avaliação e comparação de modelos de machine learning. A Figura 1 apresenta uma o fluxograma de projetos com Machine Learning:

Figura 1 – Fluxo de Projeto de Machine Learning



Fonte: Elaborado pelo autor

3.1 DADOS

Identificação e obtenção de dados históricos confiáveis e relevantes relacionados à produção de milho e soja no Brasil, a partir de fontes de dados abertos. Para séries temporais de clima foram obtidos dados do Instituto Nacional de Meteorologia (INMET), e para os dados de produção para as *commodities* escolhidas (milho, café, soja e cana-de-açúcar), foram obtidos dados da Produção Agrícola Municipal (PAM), publicados pelo IBGE, por meio do sistema SIDRA. Foram consideradas variáveis climáticas importantes como umidade, radiação solar, temperatura, pressão atmosférica, entre outros.

Construiu-se o conjunto de dados final, unindo os dados de clima e produção. As informações de produto e área foram usadas na criação da coluna-alvo, que descreve a produção (em toneladas) por área (em

hectares). As colunas do conjunto de dados final estão especificadas na Tabela 1.

Tabela 2 – Lista final de colunas e atributos	
Coluna	**Data type**
UF	object
Year	int64
VENTO_VELOCIDADE HORARIA (m/s)	float64
PRECIPITAÇÃO TOTAL_ HORÁRIO (mm)	float64
PRESSAO ATMOSFERICA AO NIVEL DA ESTACAO_ MEDIA (mB)	float64
PRESSÃO ATMOSFERICA MAX. (AUT) (mB)	float64
PRESSÃO ATMOSFERICA MIN. (AUT) (mB)	float64
RADIACAO GLOBAL (KJ/m²)	float64
TEMPERATURA DO AR - BULBO SECO HORARIA (°C)	float64
TEMPERATURA DO PONTO DE ORVALHO (°C)	float64
TEMPERATURA MÁXIMA (AUT) (°C)	float64
TEMPERATURA MÍNIMA (AUT) (°C)	float64
TEMPERATURA ORVALHO MAX. (AUT) (°C)	float64
TEMPERATURA ORVALHO MIN. (AUT) (°C)	float64
UMIDADE REL. MAX. (AUT) (%)	float64
UMIDADE REL. MIN. (AUT) (%)	float64
UMIDADE RELATIVA DO AR MEDIA (%)	float64
VENTO_DIREÇÃO HORARIA (gr) (° (gr))	float64
VENTO_RAJADA MAXIMA (m/s)	float64
Produto	object
prod/area(Ton/ha2)	float64

Fonte: Elaborada pelo autor

A seleção de atributos, ou *feature selection*, é uma etapa crucial em trabalhos que lidam com grandes volumes de dados, devido ao impacto direto que exerce na eficiência, interpretabilidade e

desempenho do modelo ou da análise. Isso permitiu também identificar os atributos menos importantes e eliminá-los do treinamento dos modelos, garantindo uma maior precisão final.

3.2 MODELOS DE PREVISÃO

Para este estudo, foram selecionados os algoritmos de *Random Forest* e *Gradient Boosting* para serem comparados com a regressão linear simples.

3.2.1 REGRESSÃO LINEAR

Regressão Linear é um dos métodos mais simples e amplamente utilizados para prever valores numéricos. Esse método assume que há uma relação linear entre as variáveis independentes (*predictors*) e a variável dependente (*target*). O modelo tenta ajustar uma linha reta que melhor represente os dados.

O modelo estima os coeficientes ($\beta 0, \beta 1, ..., \beta$ N) de uma equação linear:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta nXn + \epsilon \quad (1)$$

Onde:

Y: Variável dependente (valor previsto).

Xn: Variáveis independentes.

ϵ: Erro ou ruído.

Esse método é amplamente utilizado por sua simplicidade e facilidade de interpretação, sendo ideal para dados com relações lineares claras. No entanto, é limitado para lidar com dados não linearmente separáveis e sensível a outliers, que podem distorcer os resultados.

3.2.2 RANDOM FOREST

O *Random Forest* é um algoritmo de aprendizado supervisionado que combina múltiplas árvores de decisão para melhorar a precisão e a robustez das previsões. Este método utiliza a técnica de *bagging* (*Bootstrap Aggregating*) em árvores de decisão. O *Bootstrap Aggregator* é uma técnica geral de *ensemble learning*

(aprendizado em “conjunto”) em ML, na qual múltiplos modelos são aplicados, e os resultados combinados são considerados no cálculo da previsão final (BREIMAN, 1996). Durante a construção de cada árvore de decisão, um subconjunto aleatório de variáveis é considerado para dividir os nós, o que reduz a correlação entre as árvores. Para prever, o algoritmo agrega os resultados das árvores: na regressão, calcula a média das previsões, enquanto na classificação, escolhe o voto majoritário.

Para regressão, o valor predito é a média das previsões de todas as árvores individuais:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} \hat{y}_t \tag{2}$$

Para classificação, o valor predito é determinado pela maioria dos votos das árvores:

$$\hat{y} = \arg\max_k \left(\frac{1}{T}\sum_{t=1}^{T} I(\hat{y}_t = k) \right) \tag{3}$$

Onde $I\left(\widehat{y^{(t)}} = k\right)$ é a função indicadora que retorna 1 se a t-ésima árvore previu a classe k, e 0 caso contrário. $\sum_{t=1}^{T} I\left(\widehat{y^{(t)}} = k\right)$Representa o número total de votos que a classe k recebeu de todas as árvores, e *argmax* identifica a classe k com o maior número de votos.

O objetivo é selecionar a melhor divisão para um nó, maximizando o ganho de informação ou minimizando a impureza. Para classificação, a impureza é frequentemente medida usando o índice de Gini:

$$G(s) = \sum_{i=1}^{k} p_i(1 - p_i) \tag{4}$$

Onde:

k é o número de classes.

pi é a proporção de amostras da classe *i* no nó atual.

Para regressão, a métrica de divisão é geralmente o erro quadrático médio (MSE):

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 \quad (5)$$

Durante a construção de cada árvore, o algoritmo seleciona aleatoriamente um subconjunto de variáveis *m* de todas as *M* disponíveis, com *m*<*M*. Para classificação, é feito:

$$m = \sqrt{M} \quad (6)$$

E para regressão:

$$m = \frac{M}{3} \quad (7)$$

Cada árvore é construída a partir de um conjunto de treinamento gerado por *bootstrap sampling*, com *n* amostras retiradas aleatoriamente com reposição do conjunto de dados original. A importância de uma variável *Xj* pode ser medida pela redução acumulada na impureza em todos os nós que usam *Xj*:

$$\text{VI}(X_j) = \sum_{t=1}^{T}\sum_{s \in S_j} \Delta I(s,t) \quad (8)$$

Onde:

Sj é o conjunto de nós que usam *Xj* em todas as árvores.

ΔI(*s*,*t*) é a redução de impureza causada pela divisão *s* no nó da árvore *t*.

Esse algoritmo é robusto contra *overfitting* e funciona bem em dados complexos, não linearmente separáveis, ou de alta dimensionalidade. No entanto, é computacionalmente intensivo, especialmente em conjuntos de dados muito grandes. O Random Forest é amplamente usado em tarefas como previsão de produtividade agrícola, diagnóstico médico e detecção de fraudes, devido à sua precisão e capacidade de lidar com dados heterogêneos.

3.2.3 GRADIENT BOOSTING

O *Gradient Boosting* é altamente eficaz em dados não linearmente separáveis, proporcionando alta precisão preditiva. Sua flexibilidade em suportar funções de perda personalizadas o torna ideal

para cenários complexos. Contudo, é computacionalmente mais caro e propenso a *overfitting* se não configurado adequadamente.

O algoritmo começa com uma previsão inicial, geralmente um valor constante, como a média dos valores da variável dependente para regressão ou a classe majoritária para classificação. Em cada iteração, o algoritmo calcula os erros residuais, ou seja, a diferença entre os valores observados (reais) e as previsões do modelo atual. Esses erros representam o que o modelo ainda precisa aprender. Um novo modelo, chamado de *modelo fraco*, tenta capturar os padrões que o modelo anterior não conseguiu, sendo ajustado para prever os erros residuais.

Esse processo é repetido diversas vezes, e a cada iteração, um novo modelo corrige os erros residuais restantes. As previsões do novo modelo são combinadas com as previsões das iterações anteriores, utilizando uma taxa de aprendizado (η), que controla o impacto do modelo atual nas previsões finais, evitando ajustes excessivos nos primeiros passos. O modelo final é a soma ponderada de todos os modelos fracos. Devido ao processo iterativo e sequencial, essa abordagem é computacionalmente intensiva e propensa a *overfitting*. No entanto, a quantidade de variações e o potencial de customização fazem desse algoritmo uma ferramenta muito utilizada nos projetos de ML.

Quando baseado em Árvores de Decisão, *Gradient Boosting* é um modelo aditivo que constrói o preditor *FM(x)* de forma iterativa, adicionando *M* árvores de decisão:

$$F_M(x) = F_0(x) + \sum_{m=1}^{M} \nu h_m(x; \theta_m) \qquad (9)$$

Onde *F0(x)* é modelo inicial (frequentemente a média dos valores para regressão ou log-odds para classificação), *hm(x;θm)* é árvore de decisão ajustada na iteração, ν é a taxa de aprendizado (*learning rate*), expressa em um valor entre 0 e 1 ($0 < \nu \leq 1$), e, por fim, *M* representa o número total de árvores.

No início, o modelo *F0(x)* é configurado para minimizar a função de perda *L(y,F)* em relação a *y*. Isso fornece uma aproximação inicial da resposta. Para regressão, é feito:

$$F_0(x) = \arg\min_c \sum_{i=1}^{N} L(y_i, c) \tag{10}$$

Para classificação (usando *log-odds* para duas classes):

$$F_0(x) = \frac{1}{2}\ln\left(\frac{p}{1-p}\right) \tag{11}$$

Os resíduos (ou pseudo-resíduos) na m-ésima iteração representam o gradiente negativo da função de perda em relação às previsões atuais. Esses resíduos indicam a direção em que o modelo deve ser ajustado para melhorar:

$$r_{im} = -\left[\frac{\partial L\big(y_i, F(x_i)\big)}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)} \tag{12}$$

Então, uma nova árvore é calculada para os resíduos, minimizando a perda:

$$\theta_m = \arg\min_{\theta} \sum_{i=1}^{N} L\big(y_i, F_{m-1}(x_i) + \nu h(x_i; \theta)\big) \tag{13}$$

E então, o modelo é atualizado adicionando a contribuição da nova árvore ajustada *hm(x)*, multiplicada pela taxa de aprendizado ν:

$$F_m(x) = F_{m-1}(x) + \nu h_m(x; \theta_m) \tag{15}$$

3.2.4 REDES NEURAIS DENSAS (DNN)

As redes neurais densas ou profundas (do inglês, *Deep Neural Networks* — DNN) são um tipo de modelo de aprendizado de máquina inspirado no funcionamento do sistema nervoso natural. Elas são compostas por camadas de neurônios artificiais (perceptrons) organizadas em três níveis principais: camada de entrada (*input layer)*, camadas ocultas (*hidden layers*) e camada de saída (*output layer*). Cada neurônio realiza uma combinação linear das entradas recebidas, seguida de uma transformação não linear, geralmente por meio de uma função de ativação como a ReLU (*Rectified Linear Unit*).

A ideia central das DNNs é que, ao empilhar múltiplas camadas, o modelo é capaz de capturar padrões complexos e hierárquicos nos dados, representando relações não lineares entre as variáveis de entrada e a variável alvo. Isso as torna especialmente úteis em

problemas onde a relação entre entrada e saída não pode ser facilmente descrita por modelos lineares tradicionais.

Na arquitetura implementada no presente trabalho, foi implementada uma camada de entrada, que recebe os dados que correspondem aos atributos relevantes para a produtividade. O número de neurônios nesta camada é definido pela quantidade de atributos utilizados. com dimensão correspondente ao número de variáveis preditoras.

As três camadas ocultas densamente conectadas, com 128, 64 e 32 neurônios, respectivamente, são camadas intermediárias que processam as informações da camada anterior e as transformam em representações mais abstratas. A função de ativação ReLU (*Rectified Linear Unit*) foi utilizada em cada camada oculta para introduzir não linearidade no modelo.

A camada de saída, por sua vez, fornece a predição da produtividade agrícola. Neste caso, possui um único neurônio, que representa a produtividade estimada da cultura em questão.

Durante o treinamento, cada observação do conjunto de dados passa pela rede neural, que ajusta seus pesos internos por meio de um processo de retropropagação (*backpropagation*), minimizando a função de erro — neste caso, o erro quadrático médio (MSE). O otimizador utilizado foi o algoritmo *Adam*, com uma taxa de aprendizado de 0.001. A função de perda utilizada foi o erro quadrático médio (MSE), que quantifica a diferença entre as predições do modelo e os valores reais de produtividade. A métrica de avaliação utilizada foi o erro absoluto médio (MAE), que fornece uma medida da precisão das predições.

Os dados foram previamente normalizados com a técnica *StandardScaler*, o que garante que todas as variáveis possuam média zero e desvio padrão unitário. Essa etapa é crucial para melhorar a estabilidade e a velocidade de convergência do modelo. A base foi dividida em conjuntos de treino (80%) e teste (20%), com parte do treino (20%) sendo usada como validação durante o processo de treinamento. O modelo foi inicialmente treinado com os dados da cultura do milho, sendo posteriormente aplicado às culturas de soja, café e cana-de-açúcar, utilizando a mesma arquitetura para fins de padronização e comparabilidade entre os resultados obtidos.

A escolha da arquitetura da rede, função de ativação, algoritmo de otimização e função de perda foi baseada em práticas recomendadas e experimentos preliminares.

3.3 AVALIAÇÃO DE DESEMPENHO

3.3.1 ERRO QUADRÁTICO MÉDIO (RMSE)

O Erro Quadrático Médio (RMSE**)** é uma métrica amplamente utilizada para avaliar o desempenho de modelos de regressão, medindo o erro médio entre os valores observados (yi) e os valores previstos (ŷi). Uma característica distintiva do RMSE é sua penalização mais severa para erros maiores, devido ao uso de erros elevados ao quadrado. Sua fórmula é expressa como:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (16)$$

Onde n é o número total de observações, *y* são os valores reais e $\hat{y}$ os valores previstos.

O RMSE tem como vantagem a manutenção das unidades da variável dependente, o que permite uma interpretação direta do erro médio em termos absolutos. Isso torna a métrica particularmente útil em contextos onde é necessário quantificar a magnitude dos erros de previsão de maneira intuitiva.

3.3.2 ERRO ABSOLUTO MÉDIO (MAE)

O *Mean Absolute Error* (MAE), ou Erro Absoluto Médio, é uma métrica amplamente utilizada em problemas de regressão, destinada a avaliar a precisão de modelos preditivos. O MAE calcula o erro médio absoluto entre os valores observados (*yi*) e os valores previstos (*ŷi*), independentemente da direção do erro (se positivo ou negativo). A fórmula para o cálculo do MAE é representada por:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (17)$$

Onde *n* é o número total de observações. Diferentemente do Erro Quadrático Médio (RMSE), o MAE não eleva os erros ao quadrado,

o que reduz sua sensibilidade a outliers. Isso faz com que o MAE seja uma métrica robusta para cenários onde grandes discrepâncias não devem dominar a avaliação do desempenho do modelo. Além disso, sua simplicidade facilita a interpretação, permitindo compreender o erro médio em termos absolutos.

3.3.3 COEFICIENTE DE VARIAÇÃO

O Coeficiente de Determinação (R^2) é uma métrica estatística que avalia a proporção da variabilidade total nos dados que é explicada por um modelo de regressão. Ele quantifica o quão bem os valores previstos ($\hat{y}i$) se ajustam aos valores reais (yi) em relação à média dos valores reais ($\bar{y}$). Sua fórmula é expressa como:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (18)$$

Onde n é o número total de observações. O valor do R^2 varia de 0 a 1, sendo que um valor de 1 indica que o modelo explica completamente a variabilidade dos dados, enquanto um valor de 0 indica que o modelo não explica nenhuma variabilidade além do que seria alcançado utilizando apenas a média. Em alguns casos, valores negativos podem surgir, indicando que o modelo apresenta um desempenho inferior ao de uma previsão baseada apenas na média.


4

4 RESULTADOS

Antes da seleção de atributos, os resultados foram os explicitados na Tabela 3. O modelo de redes neurais foi implementado posteriormente, e, portanto, aparecerá na próxima tabela.

Tabela 3 – Avaliação de Desempenho dos Modelos Aplicados

Produto	Model	RMSE	MAE	R^2
Cafe	Linear Regression	1.459508	0.407333	-6.947099
Cafe	Random Forest	0.402573	0.297540	0.395376
Cafe	Gradient Boosting	0.423114	0.318230	0.332103
Soja	Linear Regression	0.423712	0.305385	0.044304
Soja	Random Forest	0.418185	0.304091	0.089164
Soja	Gradient Boosting	0.420465	0.305187	0.058985
Milho	Linear Regression	1.492011	1.217133	0.406193
Milho	Random Forest	1.306512	1.023686	0.544668
Milho	Gradient Boosting	1.374407	1.093795	0.461140
Cana	Linear Regression	12.347638	9.758570	0.250682
Cana	Random Forest	9.912176	6.998399	0.517123
Cana	Gradient Boosting	10.586907	7.911737	0.449146

Fonte: Elaborado pelo autor.

Com a seleção de atributos, percebeu-se uma falta de correlação das colunas relacionadas ao vento (velocidade e direção) e umidade relativa do ar (devido, provavelmente, ao uso dos sistemas de irrigação,

que mantém a umidade das áreas de lavoura em um nível mais estável do que nos pontos de medição). A eliminação desses atributos no treinamento e execução do modelo resultou nos resultados demonstrados na Tabela 4.

Tabela 4 – Resultados Ajustados

Produto	Model	RMSE	MAE	R^2
Cafe	Linear Regression	1.370358	0.393253	-6.005888
Cafe	Random Forest	0.094444	0.027430	0.966723
Cafe	Gradient Boosting	0.166922	0.127353	0.896051
Cafe	DNN	0.0490	0.0558	0.971801
Soja	Linear Regression	0.384844	0.278231	0.211676
Soja	Random Forest	0.118731	0.053311	0.924965
Soja	Gradient Boosting	0.247510	0.175666	0.673922
Soja	DNN	0.0535	0.1732	0.957410
Milho	Linear Regression	1.424937	1.141022	0.458382
Milho	Random Forest	0.480828	0.283067	0.938329
Milho	Gradient Boosting	0.721061	0.542400	0.861310
Milho	DNN	0.0493	0.1639	0.961501
Cana	Linear Regression	12.287023	9.823559	0.258021
Cana	Random Forest	2.805696	0.763012	0.961312
Cana	Gradient Boosting	4.516239	3.203102	0.899757
Cana	DNN	0.0478	0.1638	0.973002

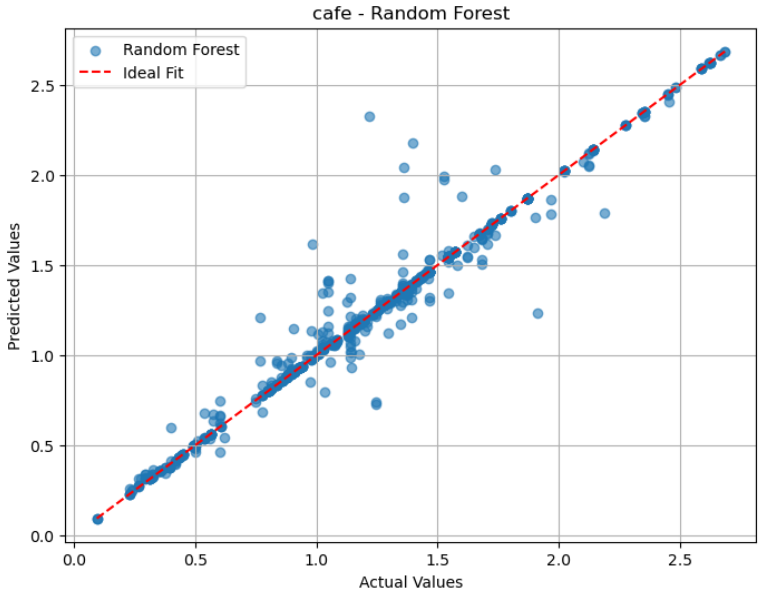
Fonte: Elaborado pelo autor.

Os resultados da Tabela 4 mostram que os modelos de aprendizado de máquina, especialmente o *random forest* e o *gradient*

boosting, apresentaram melhor desempenho em relação à regressão linear para todas as culturas avaliadas.

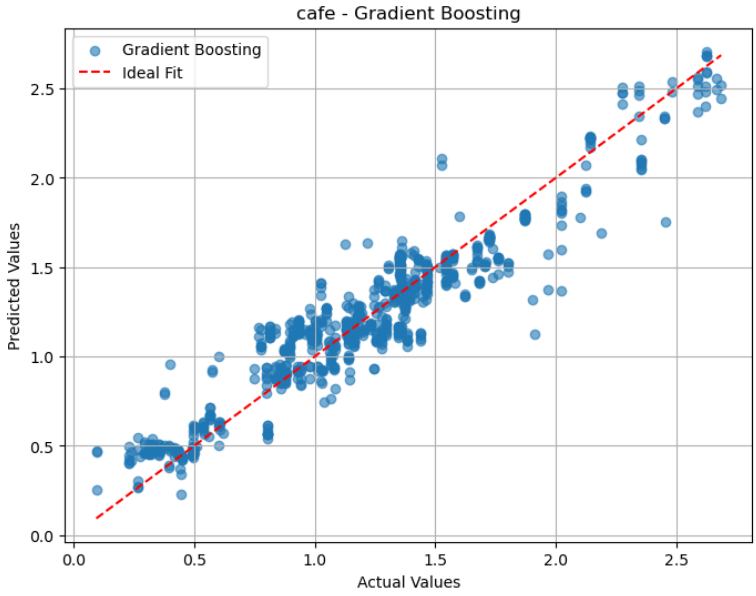
Para o café, o *Random Forest* obteve um RMSE de apenas 0.094 e um R^2 de 0.97, indicando alta precisão e capacidade de explicação da variabilidade nos dados. De forma semelhante, o *Gradient Boosting* também demonstrou resultados consistentes, com um R^2 de 0.90. Esses valores contrastam com o desempenho da Regressão Linear, cujo RMSE foi 1.37 e o R^2 negativo (-6.00), sugerindo que este modelo não é adequado para capturar a complexidade das variáveis que impactam a produtividade do café. As DNN's, no entanto, apresentaram um desempenho consideravelmente superior, com um R^2 de 0.9718.

Figura 2 – Desempenho do modelo de Random Forest para a produção de café



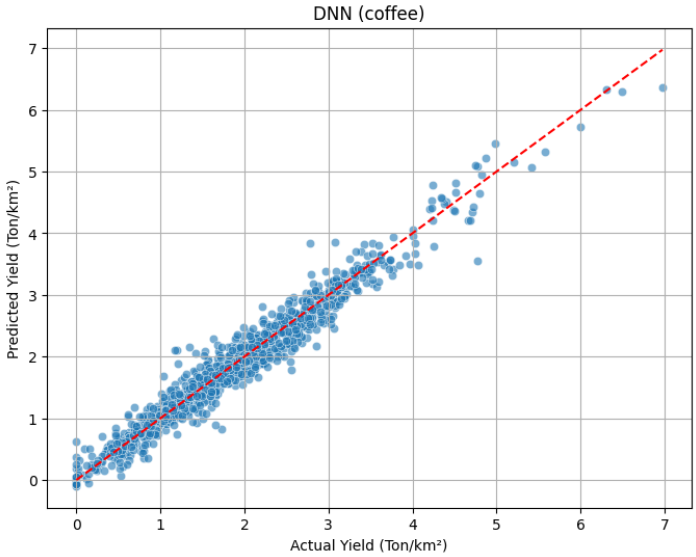
Fonte: Elaborado pelo autor.

Figura 3 – Desempenho do modelo de Gradient Boosting para a produção de café



Fonte: Elaborado pelo autor.

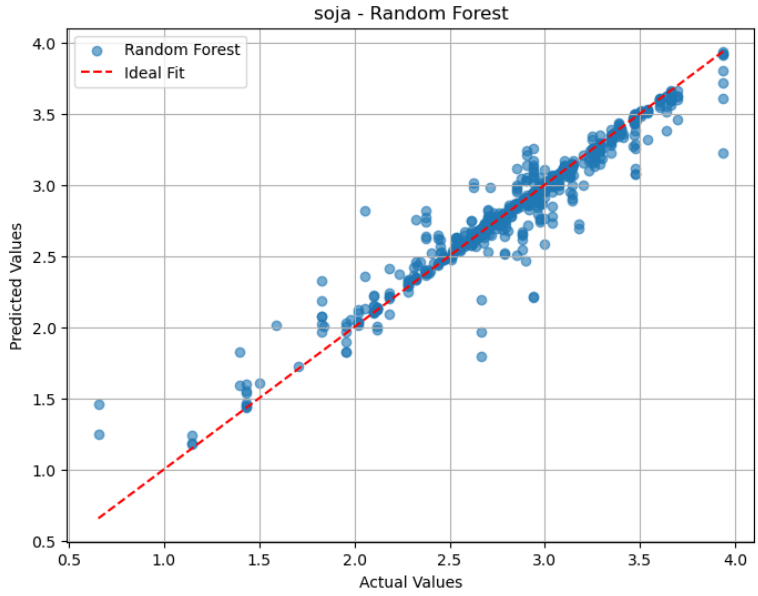
Figura 4 – Desempenho da DNN para a produção de café



Fonte: Elaborado pelo autor.

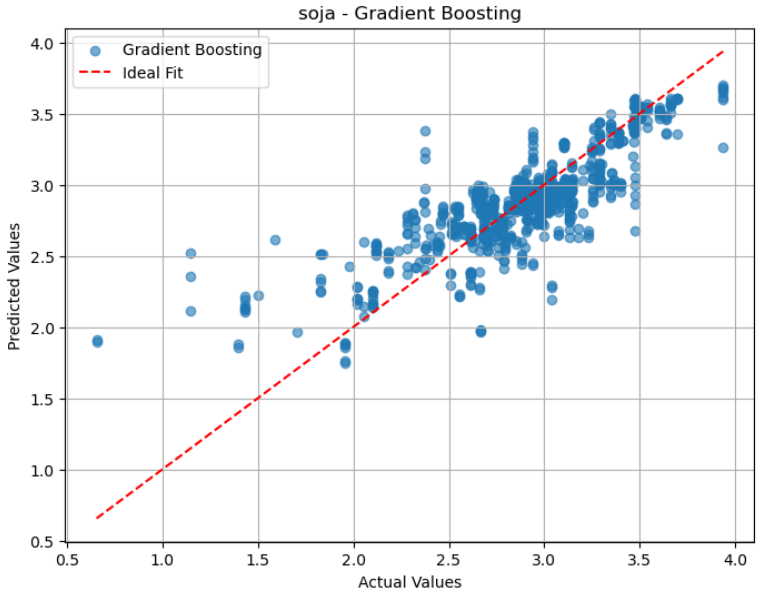
Para a soja, os modelos de aprendizado de máquina novamente superaram a regressão linear. O *Random Forest* alcançou um R^2 de 0.92 e um MAE de apenas 0.053, refletindo sua alta precisão na previsão da produtividade. O *Gradient Boosting*, embora ligeiramente inferior ao *Random Forest*, com R^2 de 0.67, ainda apresentou desempenho significativo. O modelo de DNN, no entanto, se destacou com um R^2 de 0.957. Já o modelo linear apresentou um R^2 muito baixo (0.21), demonstrando limitação para explicar a variabilidade nos dados da soja.

Figura 5 – Desempenho do modelo de Random Forest para a produção de soja



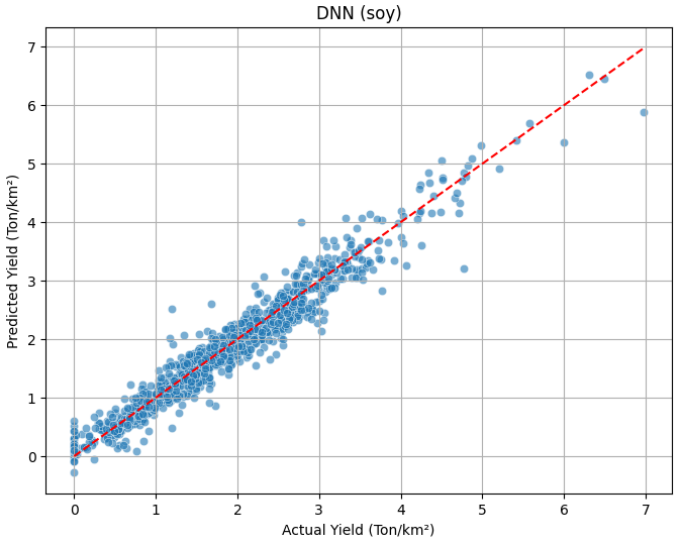
Fonte: Elaborado pelo autor.

Figura 6 – Desempenho do modelo de Gradient Boosting para a produção de soja



Fonte: Elaborado pelo autor.

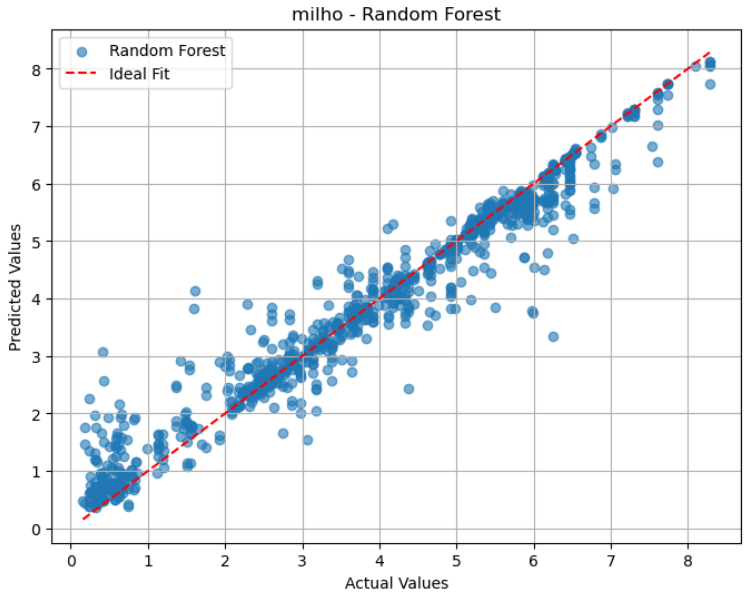
Figura 7 – Desempenho da DNN para a produção de soja



Fonte: Elaborado pelo autor.

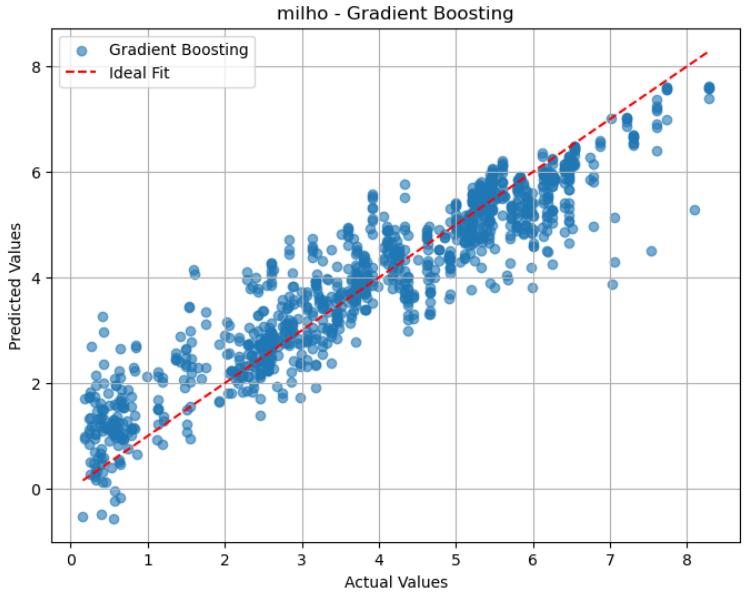
A análise dos resultados para o milho reforça a robustez dos modelos de aprendizado de máquina no contexto da produção agrícola, com o *Random Forest* alcançando um R^2 de 0.93 e um RMSE de 0.48, superando o *Gradient Boosting*, que obteve um R^2 de 0.86. A Regressão Linear, por outro lado, teve um desempenho mais fraco, com um R^2 de 0.45 e RMSE de 1.42, destacando novamente a superioridade dos modelos baseados em árvores para lidar com a complexidade da produção agrícola. No caso do milho, as DNNs também tiveram um desempenho notável, atingindo um R^2 de 0.9574.

Figura 8 – Desempenho do modelo de Random Forest para a produção de milho



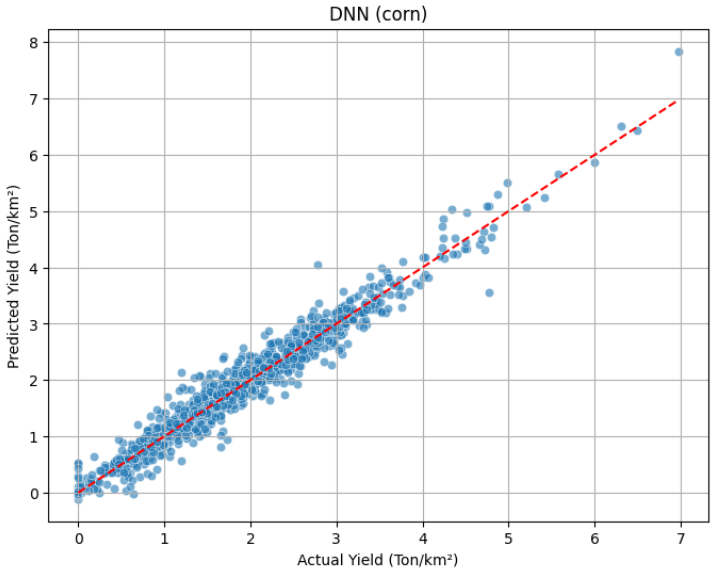
Fonte: Elaborado pelo autor.

Figura 9 – Desempenho do modelo de Gradient Boosting para a produção de milho



Fonte: Elaborado pelo autor.

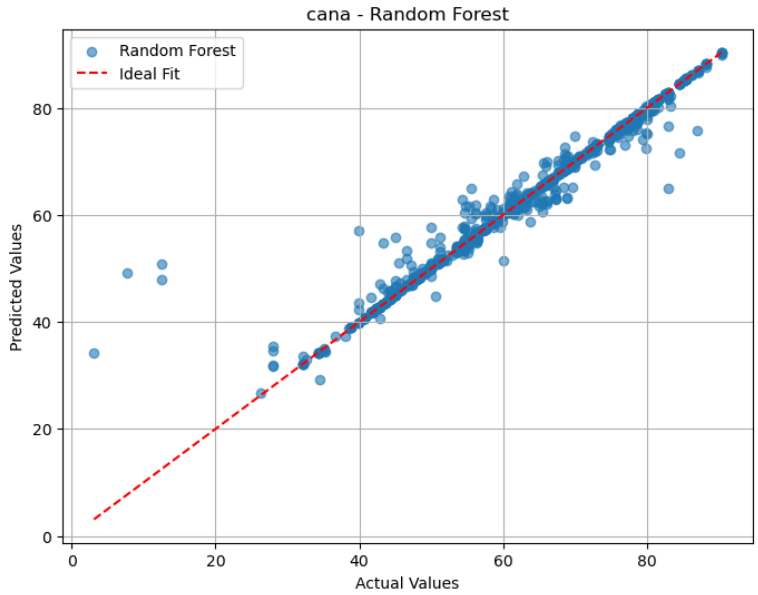
Figura 10 – Desempenho da DNN para a produção de milho



Fonte: Elaborado pelo autor.

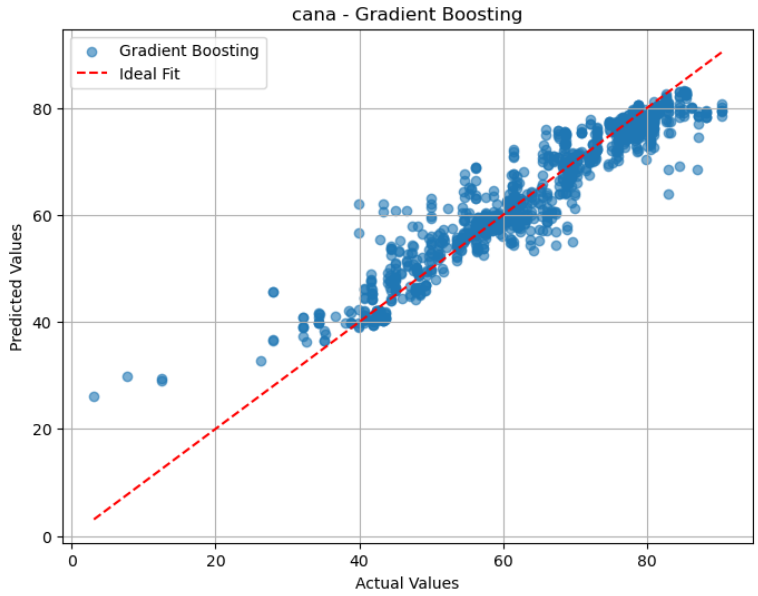
Por fim, para a cana-de-açúcar, o *Random Forest* demonstrou um desempenho notável, com um R^2 de 0.96 e um RMSE de 2.80, enquanto o *Gradient Boosting* obteve um R^2 de 0.90. Esses resultados indicam a capacidade desses modelos de lidar com a variabilidade nos dados de uma cultura altamente sensível a fatores climáticos e de manejo. As redes neurais obtiveram, mais uma vez, o melhor resultado, com um R^2 de 0.97. Em contraste, o modelo linear teve um R^2 de apenas 0.25 e um RMSE elevado de 12.28, mostrando pouca eficácia na modelagem da produtividade da cana.

Figura 11 – Desempenho do modelo de Random Forest para a produção de cana-de-açúcar



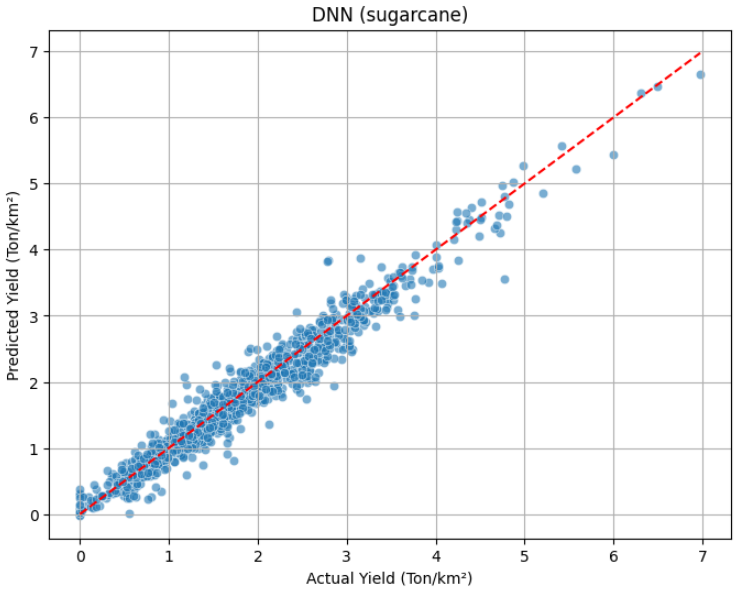
Fonte: Elaborado pelo autor.

Figura 12 – Desempenho do modelo de Gradient Boosting para a produção de cana-de-açúcar



Fonte: Elaborado pelo autor.

Figura 13 – Desempenho da DNN para a produção de cana-de-açúcar



Fonte: Elaborado pelo autor.

De forma geral, os resultados sugerem que algoritmos de aprendizado de máquina, incluindo *Random Forest* e *Gradient Boosting*, possuem potencial para melhorar as previsões de produtividade agrícola no Brasil, sendo capazes de capturar relações complexas entre variáveis climáticas, regionais e de manejo, e fornecendo previsões mais precisas e robustas em comparação com métodos lineares tradicionais. Os resultados também indicam que as redes neurais profundas (DNNs) apresentaram um desempenho superior aos demais algoritmos avaliados, demonstrando uma capacidade de modelar a complexidade dos dados agrícolas e gerar previsões com maior acurácia.


5

5 CONCLUSÃO

Este trabalho propôs explorar o uso do machine learning para prever a produção agrícola no Brasil, analisando dados de grãos como milho, café, cana-de-açúcar e soja. Foram avaliados modelos preditivos com potencial melhorar a gestão de recursos, mitigar riscos e promover sustentabilidade no setor rural brasileiro.

O trabalho reforça o potencial do uso de algoritmos de ML na previsão da produtividade agrícola no Brasil, destacando sua eficácia em lidar com as complexidades e variabilidades dos dados agrícolas. Os modelos de aprendizado de máquina avaliados, como o *random forests* e o *gradient boosting*, ofereceram vantagens significativas em relação à regressão linear, capturando relações não lineares e interações entre variáveis climáticas, regionais e de manejo. As redes neurais mostraram um potencial significativo, prevendo com mais de 90% de precisão. Esses resultados indicam que a aplicação de ML pode melhorar a precisão das previsões agrícolas, possibilitando um planejamento mais eficiente e estratégico no setor.

O desempenho satisfatório dos modelos, com apenas dados geográficos e climáticos, sugere que uma aplicação dessas abordagens aliada à mais atributos (como uso de fertilizantes, condições de mercado e cenário internacional) pode gerar modelos com precisão ainda maior, evidenciando o potencial dessa ferramenta no contexto agrícola.

O uso de Machine Learning na agricultura brasileira tem o potencial de gerar efeitos positivos no âmbito social e ambiental. Ao oferecer previsões mais precisas, as ferramentas de ML podem otimizar o uso de recursos como água, fertilizantes e energia, reduzindo o desperdício e minimizando impactos ambientais, como a contaminação do solo e a emissão de gases de efeito estufa. Do ponto de vista social, essas tecnologias podem aumentar a produtividade de pequenas e grandes propriedades agrícolas, contribuindo para a segurança alimentar, a redução da pobreza rural e a promoção de práticas agrícolas sustentáveis. Além disso, a disponibilização das bibliotecas de código livre consiste em um fator positivo para democratizar o acesso às ferramentas e resultados.

Adicionalmente, os insights gerados por algoritmos de ML podem apoiar a formulação de políticas públicas voltadas para o setor agrícola. Com base nas previsões de produtividade e na identificação de padrões regionais e temporais, gestores e formuladores de políticas podem tomar decisões mais informadas sobre incentivos agrícolas, alocação de recursos e estratégias para mitigar os impactos das mudanças climáticas na produção.

Assim, a incorporação de tecnologias de aprendizado de máquina no planejamento agrícola não apenas fortalece o setor produtivo, mas também contribui para um desenvolvimento econômico, social e ambiental mais equilibrado e sustentável no Brasil.

REFERÊNCIAS

REFERÊNCIAS

REFERÊNCIAS

BASSINE, F. Z.; EPULE, T. E.; KECHCHOUR, A.; CHEHBOUNI, A. **Recent applications of machine learning, remote sensing, and IoT approaches in yield prediction: a critical review**. ArXiv, 2023. Disponível em: https://arxiv.org/abs/2306.04566. Acesso em: 16 jun. 2024.

BREIMAN, L. **Bagging predictors**. Machine Learning, v. 24, n. 2, p. 123–140, ago. 1996.

CUNHA, R. L. F.; SILVA, B.; AVEGLIANO, P. B. **A comprehensive modeling approach for crop yield forecasts using AI-based methods and crop simulation models.** ArXiv, 2023.Disponível em: https://arxiv.org/abs/2306.10121. Acesso em: 16 jun. 2024.

DIAS, Marcos Vinícius Pretti. **Métodos Preditivos Espaço-Temporais e sua Aplicação na Agronomia.** Universidade Estadual de Londrina, 2024. Disponível em: https://sites.uel.br/dc/wp-content/uploads/2024/08/PROJETO_TCC_MARCOS_VINICIUS_PRETTI_DIAS.pdf. Acesso em: 10 dez. 2024.

FERREIRA, Lucas; ALMEIDA, Beatriz. **Machine Learning Aplicado na Predição da Qualidade Física de Grãos de Milho no Transporte.** Anais do Congresso Brasileiro de Engenharia Agrícola, 2022. Disponível em: https://conbea.org.br/anais/publicacoes/conbea-2022/anais-2022/ciencia-e-tecnologia-de-pos-colheita-ctp/3349-machine-learning-aplicado-na-predicao-da-qualidade-fisica-de-graos-de-milho-no-transporte/file. Acesso em: 16 dez. 2024.

FRANCO, I. T.; DE FIGUEIREDO, R. M. **Predictive Maintenance: an embedded system approach**. Journal of Control, Automation and Electrical Systems, Cham, v. 34, n. 1, p. 60–72, 2023. DOI: 10.1007/s40313-022

GOMES, J. A. **Machine Learning e Deep Learning: usos e aplicações na agricultura.** AgroAdvance, 2024. Disponível em: https://agroadvance.com.br/blog-machine-learning-deep-learning-agricultura/. Acesso em: 16 jun. 2024.

GUPTA, I.; AYALASOMAYAJULA, S.; SHASHIDHARA, Y.; KATARIA, A.; SHASHIDHARA, S.; KATARIA, K.; UNDURTI, A. **Innovations in Agricultural Forecasting: A Multivariate Regression Study on Global Crop Yield Prediction.** 2023. Preprint (estudo implementa seis modelos de regressão para prever produtividade agrícola em 37 países em desenvolvimento ao longo de 27 anos). Disponível em: arXiv. Acesso em: 26 jun. 2025.

IBGE, S. **Produção Agrícola Municipal**. 2019. https://sidra.ibge.gov.br/tabela/1612.

KALLENBERG, M. G. J.; MAESTRINI, B.; VAN BREE, R.; RAVENSBERGEN, P.; PYLIANIDIS, C.; VAN EVERT, F.; ATHANASIADIS, I. N. **Integrating process-based models and machine learning for crop yield prediction.** ArXiv, 2023. Disponível em: https://arxiv.org/abs/2307.13466. Acesso em: 16 jun. 2024.

LIMA, F.; SAMPAIO, I. G.; BERNARDINI, F.; PAES, A.; ANDRADE, E. O.; VITERBO, J. **Avaliação de modelos de predição e previsão construídos por algoritmos de aprendizado de máquina em problemas de cidades inteligentes**. In: TÓPICOS EM SISTEMAS DE INFORMAÇÃO: MINICURSOS SBSI 2019, 2019, Salvador. Anais... Salvador: SBC, 2019. p. 81–113. DOI: 10.5753/sbc.480.9.04.

MAAZALLAHI, A.; THOTA, S.; KONDABOINA, N. P.; MUKTINENI, V.; ANNEM, D.; ROKKAM, A. S.; AMINI, M. H.; SALARI, M. A.; NOROUZZADEH, P.; SNIR, E.; RAHMANI, B. ***Naïve Bayes and Random Forest for Crop Yield Prediction***. [s.l.]: arXiv, 23 abr. 2024. DOI: 10.21203/rs.3.rs-4345189/v1.

NOSRATABADI, S.; FELDE, I.; SZELL, K.; ARDABILI, S.; BESZEDES, B.; MOSAVI, A. **Comparative analysis of ANN-ICA and ANN-GWO for crop yield prediction**. *Preprint*, mar. 2020. Disponível em: repositório: arXiv ou ResearchGate.

PATHAK, D. K. et al. **Predicting Crop Yield with Machine Learning: An Extensive Analysis of Input Modalities and Models on a Field and Sub-Field Level.** In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2023, Pasadena, CA, USA. Anais... Pasadena: IEEE, Jul. 2023. DOI: 10.1109/IGARSS52108.2023.10282318

PINHEIRO, R. DE M. et al. **Inteligência artificial na agricultura com aplicabilidade no setor sementeiro**. Diversitas Journal, v. 6, n. 3, p. 2984–2995, 2021.

SAAED, K.; WANG, L. **Random forests for global and regional crop yield predictions**. Front. Plant Sci, v. 10, n. 621, p. 51–59, 2019.

SANTOS, Felipe Augusto; PEREIRA, Juliana. **Uso de Machine Learning no Manejo da Irrigação e Estimativa da Evapotranspiração de Referência**. Universidade de São Paulo, 2023. Disponível em: https://www.teses.usp.br/teses/disponiveis/11/11152/tde-01112023-182135/pt-br.php . Acesso em: 15 dez. 2024.

SAYWER, D. R. **Fluxo e refluxo da fronteira agrícola no brasil: ensaio de interpretação estrutural e espacial**. Revista Brasileira de Estudos de População, Revista Brasileira de Estudos de População, v. 1, p. 3–34, 2013.

SILVA, Ana Beatriz; COSTA, João Pedro; MENDES, Lucas. **Aplicações de Técnicas de Machine Learning nas Áreas da Engenharia de Produção**. Anais do Encontro Nacional de Engenharia de Produção, 2020. Disponível em: https://abepro.org.br/biblioteca/TN_STO_356_1840_42324.pdf . Acesso em: 14 dez. 2024.

SILVA, Eduardo. **Aplicações e Técnicas de Machine Learning na Agricultura**. Ciência e Dados, 2023. Disponível em: https://www.cienciaedados.com/aplicacoes-e-tecnicas-de-machine-learning-na-agricultura/ . Acesso em: 09 dez. 2024.

SILVA, Maria Clara; OLIVEIRA, Pedro Henrique. **Abordagem para predição de soja por índices de vegetação e modelos de Machine Learning**. Revista Brasileira de Engenharia Agrícola e Ambiental, v. 24, n. 4, p. 245-252, 2020. Disponível em: https://repositorio.unesp.br/bitstreams/6a6641ed-058c-4b49-95e2-ca4a3bdf4257/download . Acesso em: 09 dez. 2024.

SOUZA, Rafael; LIMA, Fernanda. **Desempenho de Algoritmos de Machine Learning na Estimativa de Nitrogênio do Feijão-Comum a partir da Leitura Indireta de Clorofila**. Anais do Congresso Brasileiro de Engenharia Agrícola, 2024. Disponível em: https://conbea.org.br/anais/publicacoes/conbea-2024/anais-2024/agricultura-digital-ad-3/4148-desempenho-de-algoritmos-de-machine-learning-na-estimativa-de-nitrogenio-do-feijao-comum-a-partir-da-leitura-indireta-de-clorofila/file . Acesso em: 10 dez. 2024.

STEWART J. I.; HAGAN, R. M.; PRUITT, W. **Production functions and predicted irrigation programs for principal crops as required for water resources planning and increased water use efficiency**. *Final report*. Washington DC, Department of Interior, p. 80, 1976.

VAN KLOMPENBURG, T.; KASSAHUN, A.; CATAL, C. **Crop yield prediction using machine learning: A systematic literature review**. Computers and Electronics in Agriculture, v. 177, p. 105709, out. 2020.

VASCONCELOS, Eduardo Silva; SILVA, Leandro Aureliano da. **Análise estatística e modelos de Machine Learning na produção agrícola brasileira: tendências temporais e eficiência produtiva ao longo de quatro décadas (1980-2019)**. Revista Contribuciones a las Ciencias Sociales, 2019. Disponível em: https://ojs.revistacontribuciones.com/ojs/index.php/clcs/article/view/8855. Acesso em: 08 dez. 2024.

XIONG, T. et al. **A combination method for interval forecasting of agricultural commodity futures prices**. Knowledge-Based Systems, v. 77, p. 92–102, mar. 2015.


idp
A ESCOLHA QUE
TRANSFORMA
O SEU CONHECIMENTO