RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo desta tese/dissertação será disponibilizado somente a partir de 24/04/2027

At the author's request, the full text of this thesis/dissertation will not be available online until April 24, 2027

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu

FACULDADE DE CIÊNCIAS AGRONÔMICAS
UNESP | BOTUCATU

EDGAR SIMKEVIC MARTINS DORNELAS

MODELOS DE APRENDIZADO DE MÁQUINA NO ESTUDO DA INFLUÊNCIA CLIMÁTICA SOBRE A PRODUTIVIDADE DE SOJA

Botucatu
2026

EDGAR SIMKEVIC MARTINS DORNELAS

MODELOS DE APRENDIZADO DE MÁQUINA NO ESTUDO DA INFLUÊNCIA CLIMÁTICA SOBRE A PRODUTIVIDADE DE SOJA

Dissertação apresentada à Faculdade de Ciências Agronômicas da Unesp Câmpus de Botucatu, para obtenção do título de Mestre em Agricultura.

Orientador(a): Maria Márcia Pereira Sartori

Botucatu

2026

D713m

Dornelas, Edgar
Modelos de aprendizado de máquina no estudo da influência climática sobre a produtividade de soja / Edgar Dornelas. -- Botucatu, 2026
50 p. : tabs., mapas

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências Agronômicas, Botucatu
Orientadora: Maria Márcia Pereira Sartori

1. Soja. 2. Aprendizado do computador. 3. Temperatura. 4. Produtividade agrícola. 5. Mudanças climáticas. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Dados fornecidos pelo autor(a).

CERTIFICADO DE APROVAÇÃO

Título:

MODELOS DE APRENDIZADO DE MÁQUINA NO ESTUDO DA INFLUÊNCIA CLIMÁTICA SOBRE A PRODUTIVIDADE DE SOJA

AUTOR: EDGAR SIMKEVIC MARTINS DORNELAS

ORIENTADORA: MARIA MÁRCIA PEREIRA SARTORI

Aprovado como parte das exigências para obtenção do Título de Mestre em Agronomia (Agricultura), pela Comissão Examinadora:

Pesquisadora Dr.ª MARIA MÁRCIA PEREIRA SARTORI (Participaçao Virtual)
Produção Vegetal / UNESP / Câmpus de Botucatu - FCA

Documento assinado digitalmente
RENATO FERNANDES CANTAO
Data: 07/05/2026 19:37:31-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. RENATO FERNANDES CANTÃO (Participaçao Virtual)
Física, Química e Matemática / Universidade Federal de São Carlos

Documento assinado digitalmente
LUCIANO BARBOSA
Data: 06/05/2026 17:52:32-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. LUCIANO BARBOSA (Participaçao Virtual)
Biodiversidade e Bioestatística / UNESP / Câmpus de Botucatu - IBB

Botucatu, 24 de abril de 2026.

Aos meus amados pais,

José Geraldo e Luciana,

dedico

AGRADECIMENTOS

Agradeço primeiramente a Deus, por conceder força, sabedoria e perseverança ao longo de toda a realização deste trabalho. Expresso minha profunda gratidão aos meus pais, pelo amor incondicional, apoio constante e pelos valores que sempre me guiaram.

Registro também meus sinceros agradecimentos à minha orientadora, Maria Marcia Pereira Sartori, pela dedicação, orientação científica, paciência e confiança depositadas durante o desenvolvimento desta pesquisa. Estendo minha gratidão aos professores e funcionários da UNESP – Faculdade de Ciências Agronômicas (FCA), cujo profissionalismo e auxílio foram fundamentais para a concretização deste estudo.

A todos que contribuíram direta ou indiretamente para esta jornada, deixo o meu sincero agradecimento.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

O objetivo deste estudo foi avaliar o desempenho de modelos de aprendizado de máquina na predição da produtividade da soja (*Glycine max* L.) a partir de variáveis agronômicas e climáticas, considerando cenários de mudanças climáticas. Foram analisadas 411 observações com variáveis de desenvolvimento e componentes de rendimento, associadas a dados climáticos, utilizando os modelos M5P, Random Forest, Random Tree e Regressão Linear. O método de seleção de atributos empregado foi o CfsSubsetEval com busca GreedyStepwise, e a validação dos modelos foi realizada por meio de cross-validation 10-fold. O modelo M5P apresentou o melhor desempenho, com R = 0,99, RMSE = 302,35 e MAE = 143,22, demonstrando elevada capacidade preditiva. Observou-se que a temperatura mínima exerce influência negativa significativa sobre a produtividade da soja, indicando que o aumento das temperaturas noturnas pode comprometer o rendimento da cultura. As projeções climáticas dos cenários RCP 4.5 e RCP 8.5 revelaram tendência de aumento das anomalias de temperatura mínima até 2100, com maior intensidade no cenário RCP 8.5. Conclui-se que o uso de técnicas de aprendizado de máquina integradas a dados climáticos constitui ferramenta promissora para a previsão de produtividade e para o planejamento agrícola frente às mudanças climáticas.

Palavras-chave: soja; aprendizado de máquina; temperatura mínima; RCP; produtividade agrícola.

ABSTRACT

The objective of this study was to evaluate the performance of machine learning models in predicting soybean (*Glycine max* L.) yield based on agronomic and climatic variables, considering future climate change scenarios. A total of 411 observations were analyzed, combining developmental and yield component variables with climatic data, using the M5P, Random Forest, Random Tree, and Linear Regression models. The attribute selection method employed was CfsSubsetEval with GreedyStepwise search, and model validation was performed through 10-fold cross-validation. The M5P model showed the best performance, with com R = 0,99, RMSE = 302,35 and MAE = 143,22, demonstrating high predictive accuracy. The results indicated that minimum temperature has a significant negative influence on soybean yield, suggesting that increasing nighttime temperatures may compromise crop productivity. Climate projections for RCP 4.5 and RCP 8.5 scenarios revealed a trend of increasing minimum temperature anomalies up to 2100, with greater intensity under RCP 8.5. It is concluded that the use of machine learning techniques integrated with climatic data is a promising tool for yield prediction and agricultural planning in the context of climate change.

Keywords: soybean; machine learning; minimum temperature; RCP; agricultural yield.

LISTA DE ILUSTRAÇÕES

Figura 1 - Localização das áreas experimentais utilizadas no estudo.....25

Figura 2 - Correlação entre variáveis agronômicas, climáticas e produtividade da soja 35

Figura 3 - Árvore de decisão gerada pelo algoritmo M5P para a predição da produtividade da soja.....38

Figura 4 - Dispersão entre produtividades observadas e estimadas pelo modelo M5P.....39

Figura 5 - Distribuição dos resíduos do modelo M5P. À esquerda, dispersão dos resíduos em relação aos valores previstos; ao centro, histograma dos resíduos; à direita, gráfico Q–Q.....40

Figura 6 - Anomalia média da temperatura mínima do mês mais frio (°C) na região de estudo entre 2025 e 2100……..….41

Figura 7 - Anomalia da temperatura mínima do mês mais frio (°C) para os cenários RCP 4.5 e RCP 8.5 nos anos de 2025, 2030 e 2035……..42

Figura 8 - Anomalia da temperatura mínima do mês mais frio (°C) para os cenários RCP 4.5 e RCP 8.5 no ano de 2100……43

LISTA DE TABELAS

Tabela 1 – Desempenho dos modelos preditivos testados.....31
Tabela 2 – Coeficientes dos modelos lineares gerados pelo algoritmo M5P para a predição da produtividade da soja.....37

SUMÁRIO

1	INTRODUÇÃO.....	19
2	REVISÃO DE LITERATURA.....	22
3	MATERIAL E MÉTODOS.....	24
3.1	Base de dados.....	24
3.2	Seleção de atributos.....	26
3.3	Modelos de Machine Learning testados……..	28
3.3.1	M5P……	28
3.3.2	Regressão Linear……...	29
3.3.3	Random Forest……...	29
3.3.4	Random Tree……...	30
3.4	Avaliação dos modelos……	30
4	RESULTADOS E DISCUSSÃO….....	31
4.1	Desempenho dos modelos testados.....	31
4.2	Análise comparativa dos algorítimos.....	32
4.3	Caracterização dos dados	33
4.4	Modelo gerado……	36
4.5	Análise de resíduos……..……...	39
4.6	Análise dos dados climáticos……..……	40
5	CONCLUSÕES.....	46
	REFERÊNCIAS…….....	48

INTRODUÇÃO

A soja (*Glycine max* L.) destaca-se como uma das culturas mais relevantes para a agricultura global, desempenhando papel estratégico na produção de proteína vegetal, óleo e subprodutos destinados às cadeias alimentícia e industrial. No Brasil, a cultura ocupa posição de destaque absoluto, consolidando o país como o maior produtor e exportador mundial de soja (CONAB, 2024). Para além de sua importância econômica, a soja tem impacto social e ambiental significativo, contribuindo para a geração de empregos, o desenvolvimento regional e a segurança alimentar.

Contudo, a produtividade da soja é fortemente condicionada por fatores genéticos, agronômicos e ambientais, sendo a variabilidade climática um dos principais limitantes. Mudanças nos padrões de temperatura, precipitação e disponibilidade hídrica têm afetado a estabilidade produtiva de culturas agrícolas em todo o mundo. Em particular, o aumento da temperatura mínima noturna tem sido apontado como um fator crítico, por elevar a taxa de respiração vegetal, reduzir a eficiência fotossintética e comprometer o balanço energético das plantas, resultando em perdas de produtividade (Lobell; Gourdji, 2012).

As mudanças climáticas globais, amplamente documentadas pelo Painel Intergovernamental sobre Mudanças Climáticas (IPCC, 2022), têm se intensificado nas últimas décadas, exigindo novas abordagens para mitigar seus impactos e adaptar sistemas produtivos. Nesse contexto, a aplicação de modelos preditivos baseados em aprendizado de máquina (*machine learning*) tem se mostrado uma ferramenta promissora, ao permitir a análise integrada de variáveis agronômicas e climáticas, identificando padrões complexos e relações não lineares que influenciam o rendimento das culturas (Kuhn; Johnson, 2013).

Entre os algoritmos de aprendizado supervisionado, o M5P apresenta vantagens relevantes para aplicações agrícolas. Esse método combina a segmentação dos dados por meio de árvores de decisão com a construção de modelos lineares nos nós terminais, permitindo simultaneamente alta precisão e concistência (Quinlan, 1992; Witten *et al*., 2016). Diversos estudos têm demonstrado sua eficácia na predição de produtividade de culturas como trigo (Nawar *et al*., 2017), milho (Yassue *et al*., 2020) e soja (Pereira *et al*., 2021), reforçando seu potencial como ferramenta de apoio à tomada de decisão.

A integração de modelos preditivos com projeções climáticas futuras permite avaliar de forma prospectiva o comportamento das culturas frente a diferentes cenários de emissões de gases de efeito estufa. Esses cenários são representados pelos *Representative Concentration Pathways* (RCPs), que constituem trajetórias de concentração de gases de efeito estufa desenvolvidas pelo IPCC para estimar o forçamento radiativo até o ano de 2100 (Van Vuuren *et al*., 2011; IPCC, 2014). O forçamento radiativo representa o balanço entre a energia solar recebida pela Terra e a energia devolvida ao espaço, expresso em watts por metro quadrado (W/m²), sendo diretamente associado ao aquecimento global.

Entre os principais cenários, o RCP 4.5 e o RCP 8.5 têm sido amplamente utilizados em estudos agrícolas e climáticos. O RCP 4.5 representa um cenário intermediário, no qual as emissões de gases de efeito estufa atingem seu pico em torno de 2040 e posteriormente diminuem, resultando em um forçamento radiativo de aproximadamente 4,5 W/m² até o final do século. Esse cenário pressupõe a implementação de políticas de mitigção e o uso de tecnologias mais limpas.

Por outro lado, o RCP 8.5 representa o cenário mais pessimista, caracterizado por um aumento contínuo das emissões ao longo do século XXI, levando a um forçamento radiativo de cerca de 8,5 W/m² em 2100. Nesse contexto, observa-se um aquecimento global acentuado, com projeções de elevações significativas nas temperaturas mínimas e máximas, alterações no regime de chuvas e intensificação de eventos climáticos extremos, como estiagens e ondas de calor. Essas condições representam ameaças diretas à estabilidade produtiva da agricultura tropical e subtropical.

Estudos recentes têm demonstrado que, sob o RCP 8.5, o aumento das temperaturas noturnas é mais pronunciado, o que pode reduzir a duração dos estádios fenológicos e alterar a distribuição geográfica de áreas aptas ao cultivo da soja (Zilli *et al*., 2021). Por sua vez, o RCP 4.5 tende a apresentar impactos mais moderados, permitindo certa margem de adaptação mediante o uso de cultivares tolerantes e práticas de manejo aprimoradas.

Diante desse panorama, o presente trabalho teve como objetivo aplicar o algoritmo M5P na predição da produtividade da soja a partir de variáveis agronômicas e climáticas, bem como analisar os efeitos das mudanças projetadas nas temperaturas mínimas sob os cenários RCP 4.5 e RCP 8.5 até o ano de 2100. Ao integrar dados observacionais e projeções futuras, este estudo busca contribuir para a compreensão

dos impactos potenciais das mudanças climáticas sobre a produtividade da soja e subsidiar o desenvolvimento de estratégias adaptativas para o setor agrícola.

5 CONCLUSÕES

O presente estudo demonstrou o elevado potencial do algoritmo M5P na predição da produtividade da soja, integrando variáveis agronômicas e climáticas em um modelo de alta precisão e robusto. O modelo alcançou desempenho robusto (R^2 = 0,99; RMSE = 302,35; MAE = 143,22), indicando sua eficácia na representação das relações entre características agronômicas e fatores ambientais. A inclusão da variável climática BIO06 (temperatura mínima do mês mais frio) foi determinante para a acurácia das estimativas, evidenciando a forte influência da temperatura noturna sobre o rendimento da cultura.

Os resultados revelaram coeficientes negativos associados à temperatura mínima em três dos dez modelos do M5P (LM 8, 9 e 10), impactando 69,23% dos casos e indicando que sua elevação reduz a produtividade. De acordo com os coeficientes médios obtidos (aproximadamente -8,35 kg ha^{-1} por °C, podendo chegar a -130,86 kg ha^{-1} em condições mais sensíveis), um acréscimo projetado de 4 °C na temperatura mínima até o ano de 2100 sob o cenário RCP 8.5 poderá resultar em uma perda estimada entre 13 e 520 kg ha^{-1} na produtividade média da soja. Essa redução representa um impacto expressivo, considerando o papel econômico e social da cultura nas regiões produtoras do sul do Brasil.

As projeções de temperatura mínima indicaram tendência de aquecimento contínuo, com intensificação após a década de 2050, especialmente no cenário RCP 8.5. Tal comportamento sugere que, sem a adoção de estratégias de mitigação e adaptação, a produtividade da soja poderá ser significativamente comprometida até o final do século. Esses resultados reforçam a necessidade de políticas agrícolas voltadas à seleção de cultivares tolerantes ao calor, aprimoramento do manejo e integração de dados climáticos em sistemas de previsão.

Conclui-se, portanto, que a combinação de algoritmos de aprendizado de máquina, como o M5P, com projeções climáticas detalhadas representa uma abordagem promissora para a avaliação de riscos e para o planejamento agrícola em cenários de aquecimento global. O presente trabalho contribui para a compreensão quantitativa dos efeitos das mudanças climáticas sobre a produtividade da soja e destaca a importância do monitoramento contínuo da temperatura mínima como variável crítica para a sustentabilidade da produção agrícola.

O presente trabalho teve como objetivo analisar as relações entre variáveis meteorológicas, temperatura média, mínima e máxima, e precipitação, e a produtividade agrícola, empregando métodos de modelagem estatística e de aprendizado de máquina, com ênfase nos algoritmos M5P e Random Forest. Os resultados permitiram identificar padrões relevantes entre fatores climáticos e rendimento das culturas, reforçando a importância do monitoramento ambiental e do uso de ferramentas preditivas na agricultura moderna. Apesar dos avanços obtidos, algumas dificuldades foram evidenciadas ao longo do estudo. A primeira refere-se à ausência de séries históricas completas das variáveis meteorológicas da área de produção, além disso, a complexidade das interações não lineares entre clima e produtividade representou um desafio adicional, demandando ajustes sucessivos nos modelos a fim de mitigar overfitting e ampliar a acurácia preditiva. Observou-se também que a variabilidade espacial intrínseca às áreas agrícolas nem sempre pode ser integralmente capturada por modelos construídos apenas com dados meteorológicos. Tais desafios destacam a necessidade de aprimoramentos metodológicos. Sugere-se, para estudos posteriores, a incorporação de dados de solo, manejo, fertilização e índices fenológicos, de modo a ampliar a capacidade explicativa dos modelos. A integração de diferentes fontes de informação, como imagens de satélite, dados de sensoriamento remoto e informações de campo, pode contribuir significativamente para melhorar o desempenho dos algoritmos de aprendizado de máquina. Recomenda-se, ainda, testar arquiteturas de modelos mais avançadas, como redes neurais profundas e métodos híbridos que combinem abordagens paramétricas e não paramétricas.

Em síntese, o trabalho contribuiu para o entendimento das relações entre clima e produtividade e demonstrou a utilidade de ferramentas de aprendizado de máquina na previsão agrícola. Entretanto, evidencia-se que a agricultura é um sistema complexo e multifatorial; portanto, a continuidade das pesquisas, com inclusão de novas variáveis, metodologias e escalas espaciais, é fundamental para o avanço do conhecimento e para o desenvolvimento de modelos ainda mais robustos e aplicáveis à realidade produtiva.

REFERÊNCIAS

CAILLON, F. *et al*. Global Bioclimatic Indicators from 1950 to 2100 Derived from CMIP5 Climate Projections. *PANGAEA*, **Bremen**, 2021. Disponível em: https://doi.org/10.1594/PANGAEA.936785. Acesso em: abr. 2025.

COMPANHIA NACIONAL DE ABASTECIMENTO – CONAB. *Acompanhamento da safra brasileira de grãos: safra 2023/2024 – Décimo levantamento*. **Brasília**: CONAB, 2024. Disponível em: https://www.conab.gov.br. Acesso em: 10 abr. 2025.

FAO – FOOD AND AGRICULTURE ORGANIZATION. *World Food and Agriculture – Statistical Yearbook 2023*. **Rome**: FAO, 2023.

GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. *Journal of Machine Learning Research*, **Cambridge**, v. 3, p. 1157–1182, 2003.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE – IPCC. *Climate Change 2014: Synthesis Report*. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the IPCC. **Geneva**: IPCC, 2014.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE – IPCC. *Climate Change 2022: Impacts, Adaptation and Vulnerability*. Contribution of Working Group II to the Sixth Assessment Report. **Cambridge**: Cambridge University Press, 2022.

KUHN, M.; JOHNSON, K. *Applied Predictive Modeling*. **New York**: Springer, 2013.

LOBELL, D. B.; GOURDJI, S. M. The influence of climate change on global crop productivity. *Plant Physiology*, **Bethesda**, v. 160, n. 4, p. 1686–1697, 2012. DOI: https://doi.org/10.1104/pp.112.208298.

NAWAR, S. *et al*. Machine learning modeling of within-field wheat yield variability using satellite imagery and topographic information. *Remote Sensing*, **Basel**, v. 9, n. 12, p. 1220, 2017.

PEREIRA, D. R. *et al*. Use of machine learning to predict soybean yield in different Brazilian environments. *Agricultural Systems*, **Amsterdam**, v. 190, p. 103103, 2021.

QUINLAN, J. R. Learning with continuous classes. In: *Australian Joint Conference on Artificial Intelligence*, 5., 1992, **Singapore**. *Proceedings...* **Singapore**: World Scientific, 1992. p. 343–348.

SENTELHAS, P. C.; GIL, L. F.; PEZZOPANE, J. R. M. *Agrometeorologia da soja no Brasil*. **Londrina**: Embrapa Soja, 2015. (Embrapa Soja. Documentos, 362).

TUKEY, J. W. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 1977.

VAN VUUREN, D. P. *et al*. The representative concentration pathways: an overview. *Climatic Change*, **Dordrecht**, v. 109, p. 5–31, 2011. DOI: https://doi.org/10.1007/s10584-011-0148-z.

WITTEN, I. H. *et al*. *Data Mining: Practical Machine Learning Tools and Techniques*. 4. ed. **Burlington**: Morgan Kaufmann, 2016.

YASSUE, R. M. *et al*. Estimating maize yield using the M5P model tree with meteorological and remote sensing data. *Precision Agriculture*, **Cham**, v. 21, p. 1120–1135, 2020.

ZILLI, J. E. *et al*. Climate change projections and their impacts on soybean production in Brazil. *Climatic Change*, **Dordrecht**, v. 164, n. 3, p. 1–17, 2021.