

**LUCAS SÉRGIO DE SOUSA LOPES**

**USO DE APRENDIZADO DE MÁQUINA INTERPRETÁVEL NA AVALIAÇÃO DA  
PRODUTIVIDADE DE POVOAMENTOS DE EUCALIPTO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para obtenção do título de *Doctor Scientiae*.

Orientador: Helio Garcia Leite  
Coorientador: Carlos Pedro Boechat Soares

**VIÇOSA - MINAS GERAIS  
2024**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

L864u  
2024  
Lopes, Lucas Sérgio de Sousa, 1994-  
    Uso de aprendizado de máquina interpretável na avaliação  
da produtividade de povoamentos de eucalipto / Lucas Sérgio de  
Sousa Lopes. – Viçosa, MG, 2024.  
    1 tese eletrônica (111 f.): il. (algumas color.).

Inclui apêndices.

Orientador: Hélio Garcia Leite.

Tese (doutorado) - Universidade Federal de Viçosa,  
Departamento de Engenharia Florestal, 2024.

Referências bibliográficas: f. 80-83.

DOI: <https://doi.org/10.47328/ufvbbt.2024.167>

Modo de acesso: World Wide Web.

1. Silvicultura. 2. Eucalipto. 3. Redes neurais.  
4. Aprendizado de máquina . I. Leite, Hélio Garcia, 1960-.  
II. Universidade Federal de Viçosa. Departamento de Engenharia  
Florestal. Doutorado em Ciência Florestal. III. Título.

CDD 22. ed. 634.973766

**LUCAS SÉRGIO DE SOUSA LOPES**

**USO DE APRENDIZADO DE MÁQUINA INTERPRETÁVEL NA AVALIAÇÃO DA  
PRODUTIVIDADE DE POVOAMENTOS DE EUCALIPTO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência Florestal, para obtenção do título de *Doctor Scientiae*.

APROVADA: 24 de junho de 2024

Assentimento:

  
\_\_\_\_\_  
Lucas Sérgio de Sousa Lopes  
Autor

  
\_\_\_\_\_  
Helio Garcia Leite  
Orientador

*Dedico este trabalho e todos os outros à minha  
amada e eterna avó Maria (in memoriam)*

## AGRADECIMENTOS

A Deus.

Ao meu amor maior, a pessoa que mais torceu e vibrou por essa realização e que segue cuidando de mim, agora de outro plano, minha avó Maria (*in memoriam*).

Ao meu grande exemplo de vida e de ser humano, meu avô Luiz, por me inspirar e me guiar no caminho do bem e da justiça.

À minha mãe, Lucia Marcia, pelo infinito amor e cuidado infinitos, por ser o meu exemplo de força e dedicação, por cuidar de mim independente de qualquer distância.

Ao meu pai, Mario Sérgio, por todo o amor, suporte, preocupação e cuidado ao longo de toda vida.

À minha tia Luciane, por sempre me incentivar e cuidar de mim e por todo o amor e carinho.

Aos meus irmãos Paulo, Nalanda e Lunna, tias Leyla e Lucirene e primas Juliana, Jamile, Luanda e Rafael.

Aos meus sobrinhos Maria e o pequeno Henry, que em breve está conosco.

Ao meu orientador e amigo, Prof. Helio Garcia Leite, pelos valiosos ensinamentos, confiança, pelo exemplo de humildade ao ensinar e por acreditar em mim.

Ao Prof. Carlos Boechat e ao Dr. Daniel Binoti, pelas valorosas contribuições para este e outros trabalhos ao longo da minha jornada acadêmica.

Ao Dr. Leonardo Pires pelo fornecimento dos dados utilizados nessa pesquisa e pela disponibilidade em sanar dúvidas e prestar esclarecimentos.

Ao professor e amigo Dr. Rafael Rode pela participação em todas as bancas da minha vida acadêmica e pelo incentivo a seguir pelo caminho da pesquisa.

À amiga Daniela Pauletto pelas inúmeras parcerias acadêmicas e pelos inúmeros arrimos emocionais prestados.

As amigas Luana Bernardes e Jessica Sabrina, pela amizade e apoio desde o início do doutorado, me ouvindo e aconselhando.

Ao meu amigo Luis Ricardo, um irmão que Viçosa me deu e que foi um grande suporte na reta final do doutorado.

Aos meus compadres e afilhado querido,IVALDO, JIANNE e MATHEUS, por serem minha família em Viçosa e meu suporte acadêmico e emocional ao longo da pós-graduação.

À minha “irmã” de orientação e grande amiga, profa. Gabriela Cristina, por toda amizade, acolhimento e valorosas contribuições durante a jornada acadêmica e de vida.

Aos amigos Gustavo e Bruna, companheiros de jornada e perrengues da vida acadêmica, sempre solícitos e prontos para ajudar.

Aos meus colegas de “república” Marcos Liziero e Giowana Ribeiro pela amizade e pela companhia durante o doutorado.

Aos meus colegas de trabalho da “salinha” do manejo Gian Marco, Leonardo Fardin, Mathaus Messias, Mindaye, Ana Carol e Aline Araujo pelas parcerias e trocas acadêmicas.

Aos amigos Illana, Hiago, Theo, Lucas Gomes, Walter, Thaynara e Andressa que me acompanharam na reta final do doutorado tornando Viçosa mais leve.

Aos funcionários do Departamento de Engenharia Florestal, em especial, Alexandre, Dilson e Francisco, pela eficiência e atenção sempre.

Aos amigos e membros da VascoUAI pelos momentos essenciais de descontração durante o doutorado, mesmo que sofrendo pelo Vasco.

À Universidade Federal de Viçosa, pela oportunidade de realizar a pós-graduação.

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

*“E eu inda sou bem moço  
pra tanta tristeza.  
Deixemos de coisas,  
cuidemos da vida,  
senão chega à morte  
ou coisa parecida,  
e nos arrasta moço  
sem ter visto a vida  
ou coisa parecida aparecida”.*  
(Belchior)

## RESUMO

LOPES, Lucas Sérgio de Sousa, D.Sc., Universidade Federal de Viçosa, junho de 2024. **Uso de aprendizado de máquina interpretável na avaliação da produtividade de povoamentos de eucalipto**. Orientador: Helio Garcia Leite. Coorientador: Carlos Pedro Boechat Soares

As ferramentas de inteligência artificial, como os algoritmos de aprendizado de máquina e as redes neurais artificiais, evoluíram rapidamente a partir de ideias baseadas nas teorias do aprendizado comportamental e cognitivo, até alcançarem os princípios do aprendizado de máquina interpretável. Técnicas *post-hoc* de interpretabilidade podem ser úteis para modelar a produtividade de povoamentos florestais, possibilitando compreender as relações entre as variáveis preditoras e a produção volumétrica na idade de corte. O objetivo desta tese foi aplicar e avaliar diferentes tipos de técnicas de aprendizado de máquina interpretável para prever a produtividade média aos 7 anos ( $IMA_7$ ) de povoamentos de eucalipto no norte de Minas Gerais. Para tal, a tese foi estruturada em três capítulos: I) Revisão de literatura sobre aprendizado de máquina; II) Técnicas *post-hoc* de aprendizado de máquina para prever a produtividade de povoamentos de eucalipto; III) Interpretabilidade de variáveis preditoras da produtividade de povoamentos de eucalipto utilizando perturbação de redes neurais artificiais. Utilizou-se uma base de dados composta por 320 talhões localizados em povoamentos de clones de híbridos de *Eucalyptus urophylla* x *Eucalyptus grandis* em Minas Gerais. Os algoritmos *random forest*, *gradient boosting machine*, árvore de decisão e redes neurais artificiais (RNA) foram treinados para generalizar a produtividade aos 7 anos em função de 304 variáveis preditoras classificadas em silviculturais, ambientais e de manejo. Técnicas de interpretabilidade pós-hoc e de perturbação de RNA foram aplicadas para identificar as variáveis com maiores importâncias relativas. As técnicas post-hoc permitiram visualizar os efeitos das variáveis de maior importância para o  $IMA_7$ : altitude  $9,7 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$  (38%), teor de argila no solo  $4,9 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$  (19,3%), teor de matéria orgânica  $2,4 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$  (9,5%) e dias úmidos na idade de um ano do povoamento  $1,7 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$  (6,7%). A ampliação de bancos de dados a partir da simulação de variáveis e adição de novas variáveis gerou importâncias relativas distintas pela perturbação de RNA. Todas as técnicas de interpretabilidade de modelos de

aprendizado de máquina incluíram a altitude, o teor de matéria orgânica e o teor de argila como variáveis de maior importância relativa para prever o *IMA*<sub>7</sub>.

Palavras-chave: Tratos silviculturais; Perturbação de redes neurais artificiais; Eucalipto-productividade; *Gradient boosting machine*; *Random forest*; Árvore de decisão; Redes neurais artificiais.

## ABSTRACT

LOPES, Lucas Sérgio de Sousa, D.Sc., Federal University of Viçosa, June 2024.  
**Using interpretable machine learning to assess eucalypt stands productivities.**  
Adviser: Helio Garcia Leite. Co-advisor: Carlos Pedro Boechat Soares

Artificial intelligence tools, such as machine learning algorithms and artificial neural networks, have rapidly evolved from ideas based on behavioral and cognitive learning theories to the principles of interpretable machine learning. Post-hoc interpretability techniques can be useful for modeling the productivity of forest stands, enabling an understanding of the relationships between predictor variables and volumetric production at the cutting age. The objective of this thesis was to apply and evaluate different types of interpretable machine learning techniques to predict the average productivity at 7 years ( $IMA_7$ ) of eucalypt stands in northern Minas Gerais. To this end, the thesis was structured into three chapters: I) Literature review on machine learning; II) Post-hoc machine learning techniques to predict the productivity of eucalypt stands; III) Interpretability of predictor variables of eucalypt stands productivity using artificial neural network perturbation. A database composed of 320 plots located in hybrid *Eucalyptus urophylla* x *Eucalyptus grandis* clone stands in Minas Gerais was used. The random forest (RF), gradient boosting machine (GBM), decision tree (CART), and artificial neural network (ANN) algorithms were trained to generalize productivity at 7 years based on 304 predictor variables classified as silvicultural, environmental, and management. Post-hoc interpretability techniques and ANN perturbation were applied to identify the variables with the highest relative importance. The post-hoc techniques allowed visualizing the effects of the most important variables for  $IMA_7$ : altitude  $9.7 \text{ m}^3\text{ha}^{-1}\text{year}^{-1}$  (38%), soil clay content  $4.9 \text{ m}^3\text{ha}^{-1}\text{year}^{-1}$  (19.3%), organic matter content  $2.4 \text{ m}^3\text{ha}^{-1}\text{year}^{-1}$  (9.5%), and humid days at one year of age of the stand  $1.7 \text{ m}^3\text{ha}^{-1}\text{year}^{-1}$  (6.7%). The expansion of databases through variable simulation and the addition of new variables generated distinct relative importance through ANN perturbation. All machine learning model interpretability techniques included altitude, organic matter content, and clay content as variables of greatest relative importance for predicting  $IMA_7$ .

Keywords: Silvicultural treatments; Perturbation of artificial neural networks; Eucalyptus-productivity; Gradient boosting machine; Random forest; Decision tree; Artificial neural networks.

## SUMÁRIO

1. INTRODUÇÃO .....	13
2. REVISÃO DE LITERATURA .....	17
2.1. Conceitos de aprendizado de máquina .....	17
2.2. Evolução do aprendizado de máquina .....	19
2.3. Aprendizado de máquina interpretável .....	25
2.4. Técnicas de interpretabilidade .....	27
2.5. Aprendizado de máquina interpretável na mensuração florestal .....	30
REFERÊNCIAS .....	31
CAPÍTULO 1: Técnicas <i>post-hoc</i> de aprendizado de máquina interpretável para predição da produtividade de povoamentos de eucalipto .....	37
1. INTRODUÇÃO .....	39
2. MATERIAIS E MÉTODOS .....	41
3. RESULTADOS E DISCUSSÃO .....	53
4. CONCLUSÕES .....	61
REFERÊNCIAS .....	62
CAPÍTULO 2: Interpretabilidade de variáveis preditoras da produtividade de povoamentos de eucalipto utilizando perturbação de redes neurais artificiais .....	66
1. INTRODUÇÃO .....	68
2. MATERIAIS E MÉTODOS .....	70
3. RESULTADOS E DISCUSSÃO .....	75
4. CONCLUSÕES .....	80
REFERÊNCIAS .....	80
CONCLUSÕES GERAIS .....	83
APÊNDICE A .....	84
APÊNDICE B .....	96
APÊNDICE C .....	104

## 1. INTRODUÇÃO

Um dos elementos essenciais do manejo florestal é a predição ou projeção do crescimento e da produção dos povoamentos, objetivando quantificar os recursos disponíveis em diferentes períodos do tempo, como idades atual e futura (Campos e Leite, 2017; Davis e Johnson (1987). A prognose do crescimento e da produção geralmente é realizada a partir de modelos baseados em abordagens distintas, denominados de modelos de crescimento e produção (Campos e Leite, 2017).

Os modelos de crescimento e produção tradicionais podem ser classificados em modelos em nível de povoamento total, distribuição diamétrica ou classe de tamanho, e modelos em nível de árvore individual, geralmente baseados em métodos estatísticos, como a regressão (Burkhardt e Tomé, 2012). No entanto, outras abordagens têm sido difundidas, como os modelos ecofisiológicos e as ferramentas de inteligência computacional, como os algoritmos de aprendizado de máquinas (Oliveira Neto et al., 2022; Silva et al., 2023; Smethurst et al., 2020).

A modelagem de atributos florestais com ferramentas de aprendizado de máquina no Brasil teve algumas das primeiras aplicações no final da década de 2000, sendo voltadas principalmente para obtenção de estimativas tão precisas quanto as obtidas por modelos de regressão usuais (Gorgens et al., 2006; Silva et al., 2009). Desde então, a utilização de algoritmos baseados em aprendizagem de máquina tem sido incorporada as rotinas das empresas florestais para predição de variáveis como o volume comercial (Cordeiro et al., 2015), altura total de árvores (Miranda et al., 2022), crescimento e produção volumétrica (Alvares et al., 2023; Santos et al., 2023) e outras variáveis.

A modelagem do crescimento e da produção de povoamentos equiâneos com algoritmos de aprendizado de máquina apresenta características importantes, como a maior facilidade de introdução de variáveis categóricas, redução de custos de coleta de dados em inventários florestais, ganhos em exatidão das predições e outras (Binoti et al., 2013; Lopes et al., 2020; Rocha et al., 2024). No entanto, ainda são escassas as aplicações de ferramentas de aprendizado de máquina voltadas para integrar o potencial preditivo dos algoritmos com a necessidade de interpretabilidade e de conhecimento das relações desempenhadas entre as variáveis preditoras e as estimativas (Binoti et al., 2022; Jones et al., 2022).

A crescente demanda por matéria-prima tem aumentado as preocupações do setor florestal com relação às mudanças climáticas e em como seus efeitos podem impactar na produtividade dos povoamentos (Binoti et al., 2022; Florêncio et al., 2022). Além disso, a necessidade de incluir modelos com estimativas consistentes e com grande exatidão para auxiliar no planejamento estratégico dos empreendimentos tem motivado a busca por ferramentas de modelagem cada vez mais complexas (Alvares et al., 2023).

As ferramentas de aprendizado de máquina interpretável têm sido utilizadas nos mais diversos domínios, possibilitando avaliar o relacionamento e os efeitos das variáveis preditoras em relação à variável de saída (Garnaik et al., 2022; Hu et al., 2023; Ryo, 2022). No entanto, tais ferramentas não têm sido utilizadas na predição de atributos de florestas.

Diante do exposto, o objetivo geral dessa tese foi testar a eficiência de diferentes tipos de técnicas de aprendizado de máquina interpretável para a predição da produtividade média aos 7 anos de povoamentos de eucalipto no norte de Minas Gerais, considerando variáveis silviculturais operacionais, ambientais e de manejo. De maneira específica, objetivou-se analisar os efeitos e os comportamentos das principais variáveis preditoras a partir de técnicas *post-hoc* de interpretabilidade de algoritmos de aprendizado de máquina e técnica de perturbação de redes neurais artificiais.

## REFERÊNCIAS

- ALVARES, C. A.; CEGATTA, Í. R.; SCOLFORO, H. F.; MAFIA, R. G. Decision-Tree Application to Predict and Spatialize the Wood Productivity Probabilities of Eucalyptus Plantations. **Forests**, 14, n. 7, p. 1334, 2023.
- BINOTI, D. H. B.; BINOTI, M. L. M. S.; LEITE, H. G.; SILVA, A. Redução dos custos em inventário de povoamentos equiâneos. **Revista Brasileira de Ciências Agrárias - Brazilian Journal of Agricultural Sciences**, 8, n. 1, p. 125–129, 2013.
- BINOTI, D. H. B.; LEITE, H. G.; ANDRADE, V. C. L. et al. Importância relativa das variáveis preditoras no processo de modelagem da produtividade florestal. **Advances in Forestry Science**, 9, n. 4, p. 1867–1873, 2022.
- BURKHART, H. E.; TOMÉ, M. Modeling Forest Stand Development. **Modeling Forest Trees and Stands**. p. 233–244, 2012.
- CAMPOS, J. C. C.; LEITE, H. G. **Mensuração florestal: perguntas e respostas**. 5. ed. Viçosa, MG: Editora UFV, 2017. 636 p.

CORDEIRO, M. A.; PEREIRA, N. N. DE J.; BINOTI, D. H. B.; BINOTI, M. L. M. DA S.; LEITE, H. G. Estimativa do volume de *Acacia mangium* utilizando técnicas de redes neurais artificiais e máquinas vetor de suporte. **Pesquisa Florestal Brasileira**, 35, n. 83, p. 255, 2015.

DAVIS, L. S.; JOHNSON, K. N. **Forest Management**. McGraw-Hill, 1987.

FLORENCIO, G. W. L.; MARTINS, F. B.; FAGUNDES, F. F. A. Climate change on Eucalyptus plantations and adaptive measures for sustainable forestry development across Brazil. **Industrial Crops and Products**, 188, n. August, 2022.

GARNAIK, S.; SAMANT, P. K.; MANDAL, M. et al. Untangling the effect of soil quality on rice productivity under a 16-years long-term fertilizer experiment using conditional random forest. **Computers and Electronics in Agriculture**, 197, n. April, p. 106965, 2022.

GÖRGENS, Eric Bastos. **Estimação do volume de árvores utilizando redes neurais artificiais**. 2006. 94 f. Dissertação (Mestrado em Ciência Florestal) – Departamento de Engenharia Florestal, Universidade Federal de Viçosa, Viçosa, MG, 2006.

HU, S.; XIONG, C.; CHEN, P.; SCHONFELD, P. Examining nonlinearity in population inflow estimation using big data: An empirical comparison of explainable machine learning models. **Transportation Research Part A: Policy and Practice**, 174, n. June, p. 103743, 2023.

IBGE. Produção da Extração Vegetal e da Silvicultura 2020. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9105-producao-da-extracao-vegetal-e-da-silvicultura.html>>. Acesso em 18 dez. 2023.

JONES, E. J.; BISHOP, T. F. A.; MALONE, B. P.; et al. Identifying causes of crop yield variability with interpretive machine learning. **Computers and Electronics in Agriculture**, 192, p. 106632, 2022.

LOPES, L. S. S.; RODE, R.; PAULETTO, D. et al. Uso de regressão e redes neurais artificiais na estimativa do volume de *Khaya ivorensis*. **Revista Ciência da Madeira - RCM**, 11, n. 2, p. 74–84, 2020.

MEDEIROS, R. A.; DE PAIVA, H. N.; NOGUEIRA, G. S.; LEITE, H. G. **Planejamento de experimentos com espécies florestais**. Editora UFV, Universidade Federal de Viçosa, 2022.

MIRANDA, E. N.; BARBOSA, B. H. G.; SILVA, S. H. G. et al. Variable selection for estimating individual tree height using genetic algorithm and random forest. **Forest Ecology and Management**, 504, p. 119828, 2022.

OLIVEIRA NETO, R. R.; LEITE, H. G.; GLERIANI, J. M. et al. Estimation of Eucalyptus productivity using efficient artificial neural network. **European Journal of Forest Research**, 141, n. 1, p. 129–151, 2022.

ROCHA, S. J. S. S.; TORRES, C. M. M. E.; VILLANOVA, P. H.; et al. Machine learning methods: Modeling net growth in the Atlantic Forest of Brazil. **Ecological Informatics**, 81, n. July 2023, 2024.

RYO, M. Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. **Artificial Intelligence in Agriculture**, 6, p. 257–265, 2022.

SANTOS, J. S.; MENDONÇA, A. R.; GONÇALVES, F. G. et al. Predicting eucalyptus plantation growth and yield using Landsat imagery in Minas Gerais, Brazil. **Ecological Informatics**, v. 75, p. 102120, 2023.

SILVA, M. L. M.; BINOTI, D. H. B.; GLERIANI, J. M.; LEITE, H. G. Adjustment of the Schumacher and Hall model and application of artificial neural networks to estimate volume of eucalypt trees. **Revista Arvore**, 33, n. 6, p. 1133–1139, 2009.

SILVA, G. C. C.; NEVES, J. C. L.; MARCATTI, G. E. et al. Improving 3-PG calibration and parameterization using artificial neural networks. **Ecological Modelling**, 479, n. August 2022, p. 110301, 2023.

SMETHURST, P. J.; VALADARES, R. V.; HUTH, N. I.; et al. Generalized model for plantation production of *Eucalyptus grandis* and hybrids for genotype-site-management applications. **Forest Ecology and Management**, v. 469, n. January, p. 118164, 2020.

## 2. REVISÃO DE LITERATURA

### 2.1. Conceitos de aprendizado de máquina

O aprendizado de máquina (*machine learning*) consiste no conjunto de técnicas e ferramentas que compõem um campo da inteligência artificial. Enquanto a inteligência artificial dedica-se a construção de sistemas computacionais que possam realizar tarefas que requerem inteligência humana, o aprendizado de máquina objetiva a criação e implementação de máquinas e sistemas inteligentes dotados da capacidade de aprender (Faceli et al., 2011).

Dessa maneira, o aprendizado de máquina é um dos elementos essenciais na construção de sistemas inteligentes, juntamente como o processamento de linguagem natural, raciocínio automatizado e as formas de representação do conhecimento, sendo responsável por adaptar o sistema a novas condições e a detectar e generalizar padrões (Norvig; Russell, 2013).

O rápido avanço tecnológico vivenciado pela Ciência da Computação na segunda metade do século XX e o caráter multidisciplinar torna difícil a definição de um conceito único e abrangente sobre aprendizado de máquina na literatura existente. Concorda-se que a aprendizagem automática está principalmente preocupada com a identificação de padrões a partir de dados e com a utilização desses padrões para fazer previsões sobre dados não vistos (Vieira et al., 2020).

O aprendizado de máquina pode ser entendido como uma parte da ciência da computação que se debruça sobre as questões inerentes à construção de programas de computador que aprendam com o acúmulo de experiências (Mitchell, 1997). Esse acúmulo muitas vezes é denominado de aprendizagem, estando submetido a leis fundamentais (Mitchell, 2006). Para Mitchell (1997, p. 2) a definição formal de aprendizado de máquina pode ser estabelecida como: "Diz-se que um programa de computador aprende com a experiência E para alguma classe de tarefas T e medida de desempenho P, se seu desempenho em tarefas em T, conforme medido por P, melhora com a experiência E."

O aprendizado de máquina é um conjunto de técnicas que melhora o desempenho do sistema aprendendo com a experiência por meio de métodos computacionais (Molnar, 2024). A experiência é representada nos sistemas computacionais pelos dados, sendo o principal objetivo do aprendizado de máquina

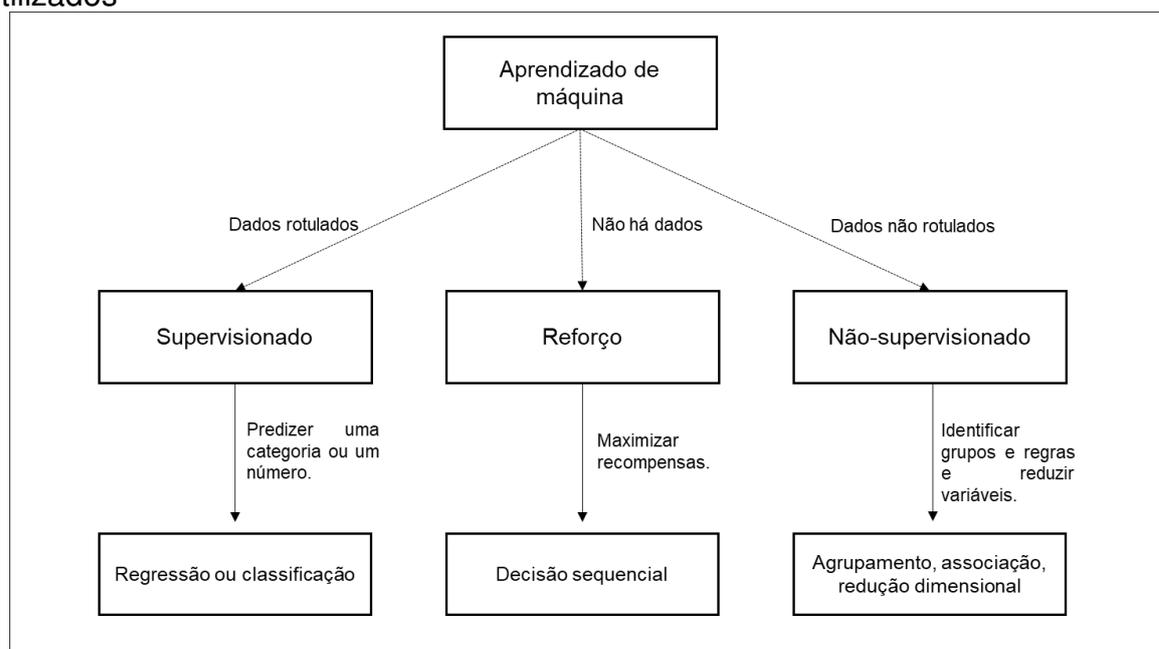
desenvolver algoritmos e modelos com capacidade de aprender a partir desses dados (Boehmke e Greenwell, 2019; Zhou, 2021)

O aprendizado de máquina está na interseção da ciência da computação e da estatística. São formuladas teorias estatístico-computacionais, além de ser usado para desvendar padrões em grandes conjuntos de dados, que são difíceis de analisar com abordagens estatísticas mais convencionais, contribuindo com a indústria de mineração de dados (Adadi e Berrada, 2018; Galvão e Marin, 2009).

O aprendizado de máquina clássico geralmente é dividido em aprendizado supervisionado e não-supervisionado. No supervisionado, os dados têm a característica de serem pré-categorizado por rótulos. Quando não há nenhum tipo de pré categorização, tem-se o chamado aprendizado não-supervisionado (Vieira, 2020).

Recentemente, muitos autores têm incluído o aprendizado por reforço como um tipo especial de aprendizado de máquina (Zhou, 2021; Kaelbling et al., 1996). Nesse caso não há necessidade de um conjunto de dados pré-existente. Os dados são obtidos por uma máquina, a partir de interações com o ambiente, buscando maximizar as recompensas (Arulkumaran et al., 2017; Ris-Ala, 2023). A Figura 1 ilustra esquematicamente a classificação do aprendizado de máquina.

Figura 1: Classificação das técnicas de aprendizado de máquina e dos tipos de dados utilizados



Fonte: Adaptado de Ris-Ala (2023, p. 19).

## 2.2. Evolução do aprendizado de máquina

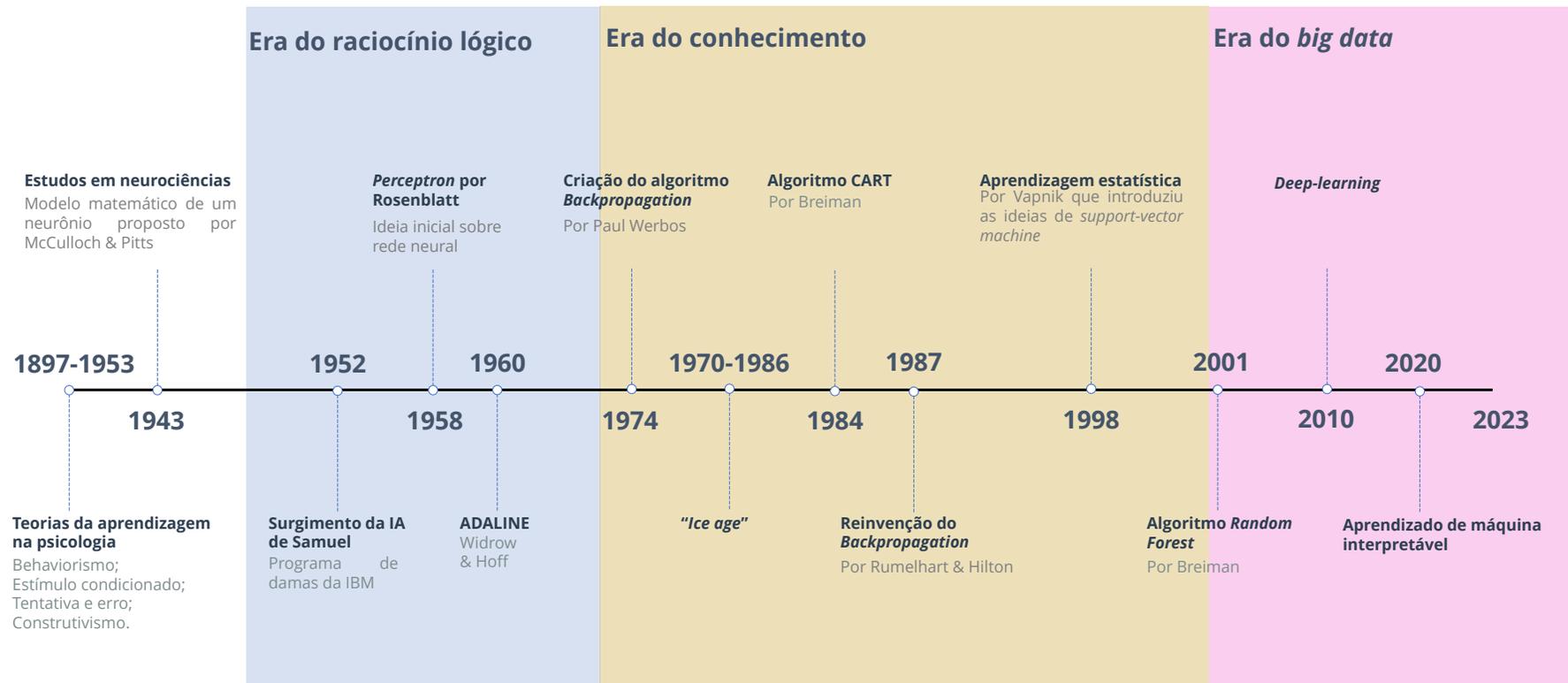
O aprendizado de máquina tem ganhado espaço na vida cotidiana dos seres humanos, na sociedade moderna, estando presente em uma grande diversidade de tarefas que vão, por exemplo, desde a identificação de spams em caixas de entrada de correios eletrônicos até no desbloqueio de smartphones com reconhecimento facial e muitas outras tarefas (Carvalho et al., 2019; Dada et al., 2019; Kavita e Chhillar, 2022).

A concepção de máquinas inteligentes passa obrigatoriamente pelo conceito de aprendizagem. Para a psicologia, o processo de aprendizagem se refere a obtenção de conhecimento a partir de experiências adquiridas, habilidades, valores e atitudes (Tabile; Jacometo, 2017). De acordo com essa abordagem, o conceito de aprendizado pode ser definido como a mudança comportamental resultante do treino ou da experiência (Giusta, 2013). A conceituação clássica proposta por Skinner (1974) define que aprendizado é expresso a partir das mudanças que o sujeito produz no ambiente. Essa conceituação permite compreender a aprendizagem como uma ferramenta adaptativa que possibilita generalizar os comportamentos aprendidos.

Uma parte significativa das ideias e fundamentos que impulsionaram os avanços da inteligência artificial e do aprendizado de máquina emergiram a partir dos conceitos de aprendizado e das teorias advindas da psicologia e da neurociência, no começo do século XX (Zhou, 2021; Vieira et al, 2020). Uma parte considerável dessas teorias se dedicava a tentar explicar e sistematizar o processo de aprendizado humano, cognitivamente e fisiologicamente (Faceli et al., 2011). Se o objetivo era que as máquinas pudessem aprender, passou a ser essencial e lógico começar a partir do entendimento de como o aprendizado é construído no cérebro e nos comportamentos humanos (Fradkov, 2020; Vieira et al., 2020; Ris-Ala, 2023).

A evolução do conhecimento sobre aprendizado de máquina pode ser dividida em três eras ao longo dos últimos 70 anos: a era do raciocínio lógico (1952 – 1973), do conhecimento (1974 – 2000) e do “*big data*” (2001 - atual). Em cada um desses períodos históricos, o conhecimento e os principais avanços empreendidos foram baseados em uma premissa específica. Uma linha do tempo detalhada é apresentada na Figura 2.

Figura 2: Linha do tempo evolutiva do aprendizado de máquina destacando as diferentes eras



Fonte: Autor (2024) adaptado de Zhou, 2021

Durante a era do raciocínio lógico foram desenvolvidos os primeiros sistemas inteligentes e as primeiras redes neurais artificiais (RNA). As descobertas desse período são fortemente baseadas em analogias ao sistema neural a partir das neurociências (Zhou, 2021). Grande parte dos estudos desenvolvidos nessa época foram marcados pelo simbolismo, na tentativa de representar o pensamento e o aprendizado de maneira matemática e mecânica, com forte natureza multidisciplinar, englobando às ciências da computação, medicina, biologia, matemática e estatística, sendo o neurônio matemático de McCulloch e Pitts (1943) o marco inicial desse período (Zilio e Carrara, 2010; Fradkov, 2020).

Além do neurônio matemático, outros três marcos principais da era do raciocínio lógico foram: o surgimento da inteligência artificial de Samuel em 1952; a criação do *perceptron* de Rosenblatt em 1958 e a concepção da rede neural artificial ADALINE por Widrow e Holf em 1960 (Rosenblatt, 1958). Esses achados formam a base teórica inicial para o desenvolvimento das redes neurais artificiais e são fundamentados no conexionismo vigente na neurociência (Haykin, 2001).

Em 1958, Frank Rosenblatt idealizou o *perceptron*, baseado nas contribuições de McCulloch e Pitts sobre o modelo de neurônio artificial matemático. O *perceptron* é conhecido como a rede neural artificial de maior simplicidade, possuindo caráter binário e funcionando como um classificador linear que resulta em um hiperplano (Rosenblatt, 1958). Os pesos entre as camadas são atualizados a partir da regra de Hebb. Apesar de suas limitações, o *perceptron* de Rosenblatt foi fundamental para o avanço e refinamento dos modelos subsequentes de redes neurais artificiais e para a integração entre as teorias neurofisiológicas e a inteligência artificial (Russell e Norvig, 2013). Na década de 1960, Widrow & Hoff foram os responsáveis pela criação das redes ADALINE (*adaptive linear element*). A rede ADALINE se diferenciava do *perceptron* por possuir uma regra delta para calcular o erro, visando a sua minimização (Fradkov, 2020).

Além das teorias a nível de organização neural, no campo das ciências cognitivas, outras teorias sobre o aprendizado humano também impactaram o aprendizado de máquina no século XX. O behaviorismo foi a corrente de maior relevância para as ideias iniciais sobre aprendizado de máquinas na era do raciocínio lógico (Vieira, 2020).

O behaviorismo surgiu no século XIX com John B. Watson (*behavior* = comportamento) e tem como principal contribuição a ideia de que o estudo da

psicologia humana deve se basear em observações de comportamentos mensuráveis, com forte caráter experimental, priorizando uma menor ênfase na subjetividade, geralmente representada por emoções e sentimentos (Zilio e Carrara, 2010; Watson, 1913). No entanto, foi só a partir da década de 1920 que essa corrente ganhou maior destaque e aceitação dentro da psicologia. Entre as principais teorias de aprendizagem surgidas no espectro behaviorista e que encontraram reflexo no aprendizado de máquinas estão o estímulo condicionado (Pavlov, 1927); aprendizagem por tentativa e erro (Thorndike e Woodworth, 1901); condicionamento operante (Skinner, 1958) e a teoria construtivista (Bruner et al., 1956).

O estímulo condicionado ou estímulo-resposta é uma teoria desenvolvida por Pavlov em 1920. Para Pavlov, o reflexo resultante de um estímulo condicionado, ou seja, um reflexo condicionado, é a unidade básica do aprendizado, tal reflexo condicionado é o produto de uma associação entre estímulos neutros, incondicionados e condicionados (Staddon, 2001). A nível cerebral, o processo resume-se a ter um estímulo condicionado, que cria sinapses duradouras, assim garantindo o aprendizado (Miltenberger, 2016).

O psicólogo e neurologista Donald Hebb propôs em 1949 uma regra de aprendizado a nível neural que posteriormente ficou conhecida por regra de aprendizado hebbiano e consiste em “quando um axônio de um neurônio A está próximo o suficiente para excitar uma célula B e repetidamente participa do disparo de B, então ocorre um processo metabólico em uma das células ou em ambas de forma que a eficiência de A em contribuir para o disparo de B é aumentada.”

Tanto o estímulo condicionado como a regra de Hebb reverberaram positivamente nas primeiras concepções sobre redes neurais artificiais durante a era do raciocínio lógico. As duas teorias descrevem, em níveis comportamentais e cerebrais, o processo de aprendizado humano, destacando o reflexo condicionado e as sinapses como unidades básicas de aprendizado (Vieira, 2020). A regra de Hebb foi fundamental para a criação dos primeiros modelos de neurônios artificiais com capacidade de aprendizado, onde o conhecimento é representado pelo peso das sinapses entre os neurônios (Haykin, 2001). É nessa interseção que as descobertas de Pavlov, Hebb e McCulloch e Pitts se encontram, servindo de base teórica para o denominado conexionismo que resultou nas RNA.

A teoria da aprendizagem por tentativa e erro (conhecida também como lei do efeito) de Edward Thorndike (1901) e o condicionamento operante de Skinner (1958)

foram fundamentais para elaboração das bases conceituais do tipo de aprendizado de máquina denominado de aprendizado por reforço (Ris-Ala, 2023). Para Thorndike a solução de uma determinada tarefa passa pela execução de diversas tentativas, quando a tentativa se aproxima da solução o sistema é recompensado.

Posteriormente, Skinner conceituou o condicionamento operante, que preconiza que um dado comportamento vai ocorrer com maior ou menor frequência a partir da consequência gerada por ele (Skinner, 1974). Dessa forma, se um comportamento gera uma consequência positiva, as probabilidades do comportamento se repetir são maiores, gerando um reforço, de maneira contrária, se a consequência for negativa, o comportamento será inibido, gerando uma punição (Nilza Micheletto e Sérgio, 1993).

Em 1952 Arthur Samuel apresentou o programa de damas da IBM que consistia em um sistema capaz de analisar partidas de damas, identificando as jogadas boas e ruim e aprender com tais jogadas de maneira autônoma, conseguindo em pouco tempo vencer jogos contra humanos (Wiederhold e McCarthy, 1992). Nota-se a semelhança entre a forma de aprendizado do programa com as ideias teóricas desenvolvidas por Skinner e Thorndike, como a busca de soluções através de mecanismos de tentativa e erro e punições e recompensas e o princípio do aprendizado por reforço.

A partir da teoria construtivista de Bruner (1956) o processo de aprendizado pode ser entendido com um mecanismo ativo, de construção de conceitos e ideias, baseados em informações e experiências anteriores. Segundo Bruner os seres humanos são capazes de extrair informações e atributos de objetos e caracterizá-los, gerando regras iniciais que vão se expandindo a partir de novos exemplos (Bruner et al., 2017).

O construtivismo reverberou no aprendizado de máquina de duas maneiras. Primeiramente, através da analogia ao aprendizado de máquina supervisionado, onde as observações são caracterizadas pela presença de rótulos (dados rotulados), permitindo que a categorização ocorra dentro do treinamento (Faceli et al., 2011; Vieira, 2020). Além disso, a teoria de Bruner aborda também a capacidade de generalização do aprendizado a partir dos exemplos vistos. Em aprendizado de máquina, a capacidade de generalização corresponde a predição e/ou classificação da variável alvo para novos dados, independentes dos dados de treinamento (Haykin, 2001; Molnar, 2024).

A partir da década de 1970, ocorreu uma mudança significativa no campo do aprendizado de máquina. A chamada "era do conhecimento" teve início, abandonando as ideias de associação e simbolismo que prevaleciam anteriormente durante o período do raciocínio lógico. Nessa nova era, o objetivo dos cientistas da computação passou a ser o desenvolvimento de algoritmos de aprendizado de máquina fundamentados em lógicas específicas, com ênfase na generalização, isto é, a criação de algoritmos capazes de adquirir conhecimentos (Zhou, 2021).

Os eventos mais significativos dessa fase incluem a criação do algoritmo *backpropagation* por Paul Werbos em 1974, a proposta inicial das árvores de classificação e regressão por Breiman em 1984, e a reinvenção do *backpropagation*, que marcou o fim da chamada "era do gelo" das RNA, por Rumelhart e Hinton em 1987 (Graupe, 2013; Gurney, 1997). A conclusão desse período é marcada pelo desenvolvimento da aprendizagem estatística por Vapnik em 1998 (Vapnik, 1998)

Entre as décadas de 1950 e 1960 o número de abordagens conexionistas, com utilização de RNA para tratar problemas linearmente separáveis, era crescente. No entanto, em 1969 os pesquisadores Marvin Minsky e Seymour Papert teorizaram que as RNA de camada única não seriam capazes de tratar problemas com complexidade não linear, exigindo que redes de múltiplas camadas fossem adotadas (Braga et al., 2007; Minsky e Papert, 1969). Contudo, tais redes necessitariam de algoritmos complexos para ajuste dos pesos sinápticos, esbarrando nas limitações computacionais da época e na inexistência desses algoritmos naquele momento (Russell e Norvig, 2013). Esse paradigma marca o começo da chamada *ice age* das RNA, um período que compreende até 1986, onde as pesquisas com RNA foram desencorajadas (Galushkin, 2007).

Em 1974, Paul Werbos concebeu a ideia inicial do algoritmo de retropropagação do erro, conhecido como *backpropagation*, visando o treinamento de RNA de múltiplas camadas (Werbos, 1990, 2009). No entanto, a concepção inicial de retropropagação proposta por Werbos só foi reconhecida posteriormente, através da reformulação realizada por Rumelhart, Hinton e Williams em 1986 (Rumelhart et al., 1986). A ampliação do *backpropagation* na década de 1980 incluiu um maior detalhamento do funcionamento do algoritmo e aumento da eficiência no cálculo das derivadas parciais dos erros, responsáveis pelas atualizações dos pesos (Braga et al., 2007). A implementação do *backpropagation* permitiu o aumento do número de problemas tratáveis com RNA, visto que as redes com múltiplas camadas passaram

a ser mais comuns (Gurney, 1997).

Ainda durante a era do conhecimento, novos algoritmos largamente difundidos no aprendizado de máquina foram concebidos, como as árvores de decisão e as máquinas de vetor de suporte (*support vector machine*). Ambos os algoritmos aprendem as tarefas de classificação e aproximação de funções (Molnar, 2024). A árvore de decisão foi proposta inicialmente por John Quinlan, numa tentativa de entender o processo de aprendizado humano baseando-se em uma estratégia de dividir para conquistar em 1962 (Zhou, 2021). Posteriormente em 1984, Breiman et al. (1984) desenvolveram a base teórica e experimental das árvores de decisão para classificação e regressão, denominadas de *classification and regression tree* – CART. O marco final da era do conhecimento é a concepção do aprendizado estatístico, atribuído a Vladimir Vapnik em 1998 e representado pelos algoritmos denominados de máquina de vetor de suporte (*support vector machine*) comumente utilizados até os dias atuais (Vapnik, 1998; Molnar, 2024).

A partir dos anos 2000, iniciou-se a era do *big data* no aprendizado de máquina, impulsionada pelo crescimento exponencial de dados disponíveis e pelo avanço das tecnologias da informação e de ferramentas de análise cada vez mais complexas (Zhou, 2021). O conexionismo voltou a ter evidência a partir do aprendizado profundo, também conhecido como redes neurais artificiais profundas (*deeping learning*), empregado em aplicações como o YOLO (*You only look once*) para detecção de objetos e outras tarefas (LeCun et al., 2015; Shafiee et al., 2017). Além disso, a disponibilidade de algoritmos e técnicas mais complexas, como o *random forest*, que engloba as árvores de decisão e ferramentas de amostragem como o *bootstrap* e o *bagging*, marca o início de um período em que os cientistas de dados começam a demandar atenção aos métodos que permitem interpretar e explicar os resultados dos modelos (Breiman, 2001); Molnar, 2024).

### **2.3. Aprendizado de máquina interpretável**

As últimas décadas foram marcadas pela expansão da aplicação de técnicas de aprendizado de máquina para os mais variados domínios, como a medicina, direito, ciências políticas e outras áreas, visando tratar problemas de grande complexidade e com alto risco associado na tomada de decisão (Carvalho et al., 2019; Rudin, 2019).

Essa expansão tem motivado uma busca mais intensa pela interpretação dos modelos e das previsões geradas por aprendizado de máquina (Miller, 2019).

A resolução de problemas de alto risco com ferramentas de inteligência computacional tem sido amplamente criticada desde as primeiras aplicações, principalmente por não apresentar características de interpretabilidade e explicabilidade de forma simples, como outros métodos, como por exemplo os modelos de regressão (Hakkoum et al., 2024; Murdoch et al., 2019). Por esse motivo é comum encontrar na literatura especializada os termos “modelos de caixa preta” ou “*black box*” associados aos algoritmos de *machine learning* (Ribeiro et al., 2016). A denominação “caixa preta” se refere ao fato de a modelagem via aprendizado de máquina focar principalmente no desempenho preditivo da variável resposta, negligenciando as inferências sobre aspectos como a natureza, dinâmica e processos do problema modelado (Hu et al., 2023; Molnar et al., 2020).

As críticas aos algoritmos de aprendizado de máquina motivaram o avanço das pesquisas com métodos e técnicas de interpretabilidade de modelos baseados em inteligência artificial. Embora não possua uma definição matemática, é comumente conceituada como o grau em que um ser humano pode compreender a causa de uma determinada decisão (Miller, 2019).

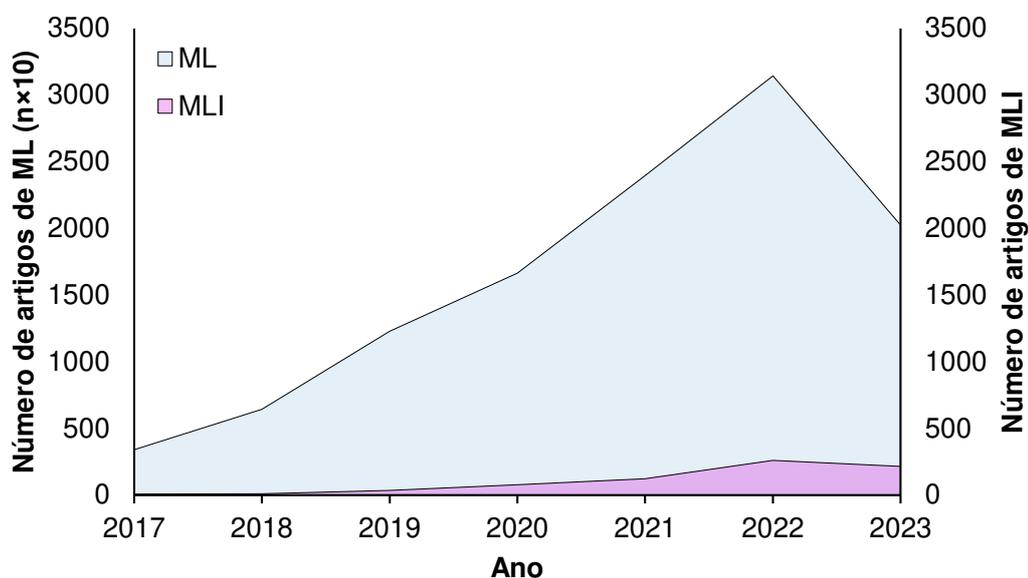
O aprendizado de máquina interpretável (*explainable machine learning* – XAI) pode ser definido como um campo do aprendizado de máquina que engloba um conjunto de métodos e modelos que possibilitam tornar o comportamento e as previsões dos sistemas compreensíveis para os seres humanos (Molnar, 2024). O XAI permite que além das previsões, seja gerado também o conhecimento contido nos dados e aprendido pelos modelos de aprendizado de máquina (Murdoch et al., 2019).

O conceito de interpretabilidade, no contexto da inteligência artificial, se refere a capacidade de explicar o resultado de um modelo de aprendizado de máquina, visando a compreensão das previsões realizadas pelo algoritmo e das relações desempenhadas entre as variáveis (Rudin et al., 2022). Tal conceito depende da disponibilidade e capacidade humana de compreensão e aceitação das explicações geradas a partir do modelo (Molnar et al., 2020). Tal conceito não inclui a decomposição do sistema de aprendizado de máquina de modo a explicar o funcionamento completo de um determinado modelo (Apley; Zhu, 2020). Tais ideias perfazem o conceito de explicabilidade, que corresponde a descrição em detalhes da lógica e do funcionamento de um sistema de previsão baseado em aprendizado de

máquina (Carvalho et al., 2019; Marcinkevičs e Vogt, 2023). No entanto, é limitada pela complexidade dos algoritmos e modelos utilizados, visto que quanto mais complexos os algoritmos, mais difícil será a explicação das previsões obtidas (Lisboa et al., 2023).

Embora o aprendizado de máquina interpretável seja um campo essencial para construção de modelos de previsão com maior transparência e menores riscos associados nas previsões obtidas, ainda é uma área pouco pesquisada quando comparada ao aprendizado de máquina clássico. O número de publicações com aplicações de aprendizado de máquina convencional é aproximadamente 160 vezes maior do que o número de artigos publicados com técnicas de aprendizado interpretável, conforme mostrado na Figura 3. Observa-se que somente a partir do ano de 2020 o número de artigos voltados para investigação da interpretação em modelos de inteligência artificial começou a crescer de forma expressiva na ciência. Essas tendências demonstram a grande lacuna no conhecimento sobre dentro do aprendizado de máquina.

Figura 3: Número de artigos publicados sobre “aprendizado de máquina” (*machine learning*) e “aprendizado de máquina interpretável” (*explainable machine learning*) nos últimos sete anos, na plataforma Scopus



Fonte: Autor (2024).

#### 2.4. Técnicas de interpretabilidade

As principais técnicas podem ser classificadas em duas classes: modelos inerentemente interpretáveis e interpretação *post-hoc* (Apley e Zhu, 2020; Molnar, 2024).

Os modelos inerentemente interpretáveis ou também denominados modelos intrinsecamente interpretáveis são aqueles que possuem estrutura simples e que permite a interpretação sem a utilização de técnicas adicionais (Lisboa et al., 2023; Murdoch et al., 2019). São exemplos de modelos interpretáveis as regressões lineares, as árvores de decisão, redes neurais artificiais e outros (Molnar, 2020; Molnar, 2024). A particularidade dessas técnicas reside na possibilidade de interpretação baseada na própria construção do algoritmo, explicitando a sua lógica interna e a forma de obtenção dos resultados, se aproximando de explicações completas dos resultados (Carvalho et al., 2019). Porém, esbarram na limitação de aplicabilidade restrita a problemas mais simples (Hakkoum et al., 2024).

As técnicas *post-hoc* são consideradas independentes dos modelos e consistem em explicar as saídas dos modelos de aprendizado de máquina, com especial atenção para as variáveis preditoras e os relacionamentos desempenhados entre elas e as variáveis preditas (Ribeiro et al., 2016). A denominação *post-hoc* indica que tais ferramentas são aplicadas após o treinamento dos algoritmos, podendo ser utilizadas inclusive em modelos intrinsecamente interpretáveis (Apley e Zhu, 2020; Molnar et al., 2020). Essas técnicas são particularmente úteis para interpretar os resultados obtidos, no entanto, não fornecem detalhes sobre o processo de treinamento dos algoritmos e sobre a lógica de funcionamento (Murdoch et al., 2019). As maiores possibilidades de aplicação tornaram as técnicas *post-hoc* mais difundidas.

As principais técnicas *post-hoc* são: gráficos de dependência parcial (PDP); curvas de expectativas condicionais individuais (ICE); gráficos de efeitos locais acumulados (ALE) e importância da variável/recurso.

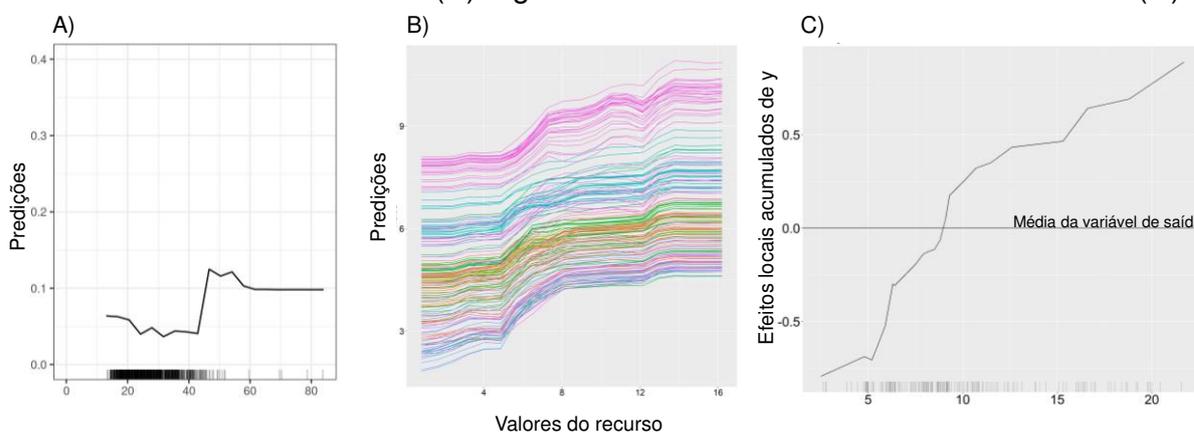
A importância de variável foi introduzida por Breiman (2001) juntamente com o algoritmo *random forest* e se baseia no princípio de permutação. Para obter a importância de uma dada variável, seu valor é permutado enquanto os valores das demais variáveis permanecem os mesmos. Quanto maior for o erro gerado com a variável permutada, maior a importância da variável para as previsões do modelo (Greenwell e Boehmke, 2019).

Os gráficos de dependência parcial (PDP) demonstram o efeito marginal que uma ou mais variáveis têm na predição da variável de saída do modelo (Figura 4A). A principal contribuição desses gráficos é a demonstração do tipo de relação desempenhada entre as variáveis preditoras e a variável predita, indicando se a relação é linear, monotônica, não linear ou desempenha outros comportamentos (Apley e Zhu, 2020). De forma simplificada, a dependência parcial demonstra o comportamento médio da previsão do modelo quando o  $j$ -ésimo recurso é alterado (Molnar, 2024). No entanto, demonstra limitações para variáveis altamente correlacionadas.

As curvas de expectativas condicionais individuais (ICE) consistem na previsão para cada instância quando o  $j$ -ésimo recurso é alterado (Molnar, 2024). Ou seja, o PDP é o comportamento médio das instâncias que são representadas pelos ICE. Os gráficos ICE detêm grande importância pois permitem observar se há presença de relações heterogêneas nos dados e se os modelos são capazes de prever tendências distintas para instâncias diferentes (Figura 4B) (Apley e Zhu, 2020).

Os gráficos de efeitos locais acumulados (ALE) contornam a limitação apresentada pelos PDP quando há presença de variáveis altamente correlacionadas (Figura 4C) (Rudin, 2019). Consistem em descrever como os recursos impactam, de forma média, a predição obtida pelo modelo. No ALE, diferentemente dos gráficos PDP, o efeito do recurso é avaliado a partir de um pequeno intervalo entre os valores e não entre os valores fixos (Apley e Zhu, 2020). Essa particularidade permite maior imparcialidade na avaliação dos recursos.

Figura 4: Gráfico de dependência parcial – PDP (A), gráfico de curvas de expectativas condicionais individuais – ICE (B) e gráfico de efeitos locais acumulados – ALE (C)



Fonte: Autor (2024) adaptado de Molnar (2024).

## 2.5. Aprendizado de máquina interpretável na mensuração florestal

Apesar do número expressivo de publicações sobre emprego de métodos de aprendizado de máquinas dos últimos anos, para predições de variáveis na mensuração florestal como volume comercial (Binoti et al., 2014; Lopes et al., 2020); crescimento e produção de povoamentos equiâneos (Casas et al., 2023; Santos et al., 2023) e inequiâneos (Reis et al., 2019; Rocha et al., 2020); biomassa e carbono (Ferraz et al., 2014); *taper* (Tavares Júnior et al., 2021); modelagem processual (Silva et al., 2023) e muitas outras, não foram encontrados estudos incluindo análises *post-hoc* como uma etapa adicional aos modelos preditivos.

Embora alguns estudos realizados na área florestal incorporem a importância da variável por permutação e/ou perturbação (Binoti et al., 2022; Rocha et al., 2024), não foram encontrados registros de aplicações na literatura das demais técnicas *post-hoc* na construção de modelos de predição baseados em aprendizado de máquina, em mensuração florestal. No entanto, o interesse pelo entendimento das relações desempenhadas entre as variáveis preditoras e preditas em problemas de mensuração florestal modelados com aprendizado de máquina já tem sido abordado em pesquisas desde a década passada.

Freitas (2020) utilizou as técnicas de Olden, Garson, Lekprofile e outras para avaliação de sensibilidade das variáveis preditoras da produtividade de eucalipto, a partir de predições por redes neurais artificiais. Binoti et al., (2022) obteve a importância relativa de variáveis silviculturais, climáticas e de manejo utilizadas na previsão da produtividade de eucalipto ao final da idade de rotação para povoamentos localizados em Minas Gerais, com técnicas de perturbação de RNA. Oliveira Neto et al. (2022) incorporou técnicas de estatística multivariada, como análise de componentes principais (PCA), para diminuição do número de variáveis preditoras, constatando que a utilização das variáveis ambientais compondo grandes conjuntos de variáveis preditoras fornece resultados mais precisos.

A adoção das técnicas *post-hoc* na modelagem florestal pode acrescentar diversas vantagens comparativas em relação aos métodos tradicionais, como os modelos de regressão clássicos, bem como, aos próprios modelos construídos com aprendizado de máquina. Além de assegurar a tomada de decisão com mais segurança, pode acrescentar maior realismo biológico aos modelos utilizados para modelar a produção e o crescimento das florestas (Ryo, 2022; Ryo et al., 2021), permite a simulação de cenários com mudanças climáticas e seus impactos na produtividade dos

povoamentos (Hu et al., 2023), economicamente pode auxiliar a quantificação de custos e receitas marginais, bem como, simular padrões de análises de sensibilidade (Carvalho et al., 2019; Murdoch et al., 2019).

## REFERÊNCIAS

ADADI, A.; BERRADA, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). **IEEE Access**, 6, p. 52138–52160, 2018.

GALUSHKIN, A. I. **Neural Networks Theory by Alexander I. Galushkin (z-lib.org)**. 2007.

APLEY, D. W.; ZHU, J. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, 82, n. 4, p. 1059–1086, 2020.

ARULKUMARAN, K.; DEISENROTH, M. P.; BRUNDAGE, M.; BHARATH, A. A. Deep Reinforcement Learning: A Brief Survey. **IEEE Signal Processing Magazine**, 34, n. 6, p. 26–38, 2017.

BINOTI, D. H. B.; BINOTI, M. L. M. S.; LEITE, H. G. Configuração de Redes Neurais Artificiais para Estimação do Volume de Árvores. **Revista Ciência da Madeira - RCM**, 5, n. 1, p. 58–67, 2014.

BINOTI, D. H. B.; LEITE, H. G.; ANDRADE, V. C. L. et al. Importância relativa das variáveis preditoras no processo de modelagem da produtividade florestal. **Advances in Forestry Science**, 9, n. 4, p. 1867–1873, 2022.

BOEHMKE, B.; GREENWELL, B. **Hands-On Machine Learning with R**. Chapman and Hall/CRC, 2019.

BRAGA, A. DE P.; LUDERMIR, T. B.; CARVALHO, A. C. P. **Redes neurais artificiais: teoria e aplicações**. LTC, 2007.

BREIMAN, L. Random forests. **Machine Learning**, 45, n. 1, p. 5–32, 2001.

BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. **Classification And Regression Trees**. 1º ed. New York: Routledge, 1984.

BRUNER, J. S.; GOODNOW, J. J.; AUSTIN, G. A. **A study of thinking**. Oxford, England: John Wiley and Sons, 1956.

BRUNER, J. S.; GOODNOW, J. J.; AUSTIN, G. A. **A Study of Thinking**. Routledge, 2017.

CARVALHO, D. V.; PEREIRA, E. M.; CARDOSO, J. S. Machine Learning Interpretability: A Survey on Methods and Metrics. **Electronics**, 8, n. 8, p. 832, 2019.

CASAS, G. G.; ISMAIL, Z. H.; LIMEIRA, M. M. C.; SILVA, A. A. L.; LEITE, H. G. Automatic Detection and Counting of Stacked Eucalypt Timber Using the YOLOv8 Model. **Forests**, 14, n. 12, p. 2369, 2023.

DADA, E. G.; BASSI, J. S.; CHIROMA, H. et al. Machine learning for email spam filtering: review, approaches and open research problems. **Heliyon**, 5, n. 6, p. e01802, 2019.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. C. **Inteligência artificial: uma abordagem de máquina**, Rio de Janeiro: LTC, 2011.

FERRAZ, A. S.; SOARES, V. P.; SOARES, C. P. B. et al. Estimativa do estoque de biomassa em um fragmento florestal usando imagens orbitais. **Floresta e Ambiente**, 21, n. 3, p. 286–296, 2014.

FLORÊNCIO, G. W. L.; MARTINS, F. B.; FAGUNDES, F. F. A. Climate change on Eucalyptus plantations and adaptive measures for sustainable forestry development across Brazil. **Industrial Crops and Products**, 188, n. August, 2022.

FRADKOV, A. L. Early History of Machine Learning. **IFAC-PapersOnLine**, 53, n. 2, p. 1385–1390, 2020.

FREITAS, E. C. S.; PAIVA, H. N.; NEVES, J. C. L. et al. Modeling of eucalyptus productivity with artificial neural networks. **Industrial Crops and Products**, 146, n. February, p. 112149, 2020.

GALVÃO, N. D.; MARIN, H. DE F. Técnica de mineração de dados: uma revisão da literatura. **Acta Paulista de Enfermagem**, 22, n. 5, p. 686–690, 2009. Escola Paulista de Enfermagem, Universidade Federal de São Paulo.

GIUSTA, A. DA S. Concepções de aprendizagem e práticas pedagógicas. **Educação em Revista**, 29, n. 1, p. 20–36, 2013. Faculdade de Educação da Universidade Federal de Minas Gerais.

GRAUPE, D. **Principles of Artificial Neural Networks**. 3rd ed. WORLD SCIENTIFIC, 2013.

GURNEY, K. **An Introduction to Neural Networks**. USA: Taylor & Francis, Inc., 1997.

HAKKOUM, H.; IDRI, A.; ABNANE, I. Global and local interpretability techniques of supervised machine learning black box models for numerical medical data. **Engineering Applications of Artificial Intelligence**, 131, p. 107829, 2024.

HAYKIN, S. **Redes Neurais: Princípios e Prática**. Bookman Editora, 2001.

HU, T.; ZHANG, X.; BOHRER, G. et al. Crop yield prediction via explainable AI and interpretable machine learning: Dangers of black box models for evaluating climate change impacts on crop yield. **Agricultural and Forest Meteorology**, 336, p. 109458, 2023.

KAEHLING, L. P.; LITTMAN, M. L.; MOORE, A. W. Reinforcement learning: a survey. **J. Artif. Int. Res.**, 4, n. 1, p. 237–285, 1996. El Segundo, CA, USA: AI Access Foundation.

KAVITA; CHHILLAR, R. S. Human Face Recognition and Age Estimation with Machine Learning: A Critical Review and Future Perspective. **International Journal of Electrical and Computer Engineering Systems**, 13, n. 10, p. 945–952, 2022.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, 521, n. 7553, p. 436–444, 2015.

LISBOA, P. J. G.; SARALAJEW, S.; VELLIDO, A.; FERNÁNDEZ-DOMENECH, R.; VILLMANN, T. The coming of age of interpretable and explainable machine learning models. **Neurocomputing**, 535, p. 25–39, 2023.

LOPES, L. S. S.; RODE, R.; PAULETTO, D.; et al. Uso de regressão e redes neurais artificiais na estimativa do volume de Khaya ivorensis. **Revista Ciência da Madeira - RCM**, 11, n. 2, p. 74–84, 2020.

MARCINKEVIČS, R.; VOGT, J. E. Interpretable and explainable machine learning: A methods-centric overview with concrete examples. **WIREs Data Mining and Knowledge Discovery**, 13, n. 3, 2023.

MILLER, T. Explanation in artificial intelligence: Insights from the social sciences. **Artificial Intelligence**, 267, p. 1–38, 2019.

MILTENBERGER, R. G. **Behavior modification: Principles and procedures**, 6th ed. Boston, MA, US: Cengage Learning, 2016.

MINSKY, M.; PAPERT, S. **Perceptrons; an Introduction to Computational Geometry**. MIT Press, 1969.

MITCHELL, T. **The Discipline of Machine Learning**, 2006.

MITCHELL, T. M. **Machine Learning**. McGraw-Hill, 1997.

MOLNAR, C.; CASALICCHIO, G.; BISCHL, B. Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges. **Communications in Computer and Information Science**, v. 1323, n. 01, p. 417–431, 2020.

MOLNAR, C. **Interpretable Machine Learning: A Guide for Making Black Box Models Explainable**. Leanpub, 2024. Disponível em: <<https://christophm.github.io/interpretable-ml-book/>>

MURDOCH, W. J.; SINGH, C.; KUMBIER, K.; ABBASI-ASL, R.; YU, B. Definitions, methods, and applications in interpretable machine learning. **Proceedings of the National Academy of Sciences**, 116, n. 44, p. 22071–22080, 2019.

MICHELETTO, N.; SÉRIO, T. M. A. P. Homem: objeto ou sujeito para skinner? **Temas em Psicologia**, n. 1, p. 11–21, 1993.

NORVIG, P.; RUSSELL, S. **Inteligência Artificial**. ELSEVIER EDITORA, 2013.

OLIVEIRA NETO, R. R.; LEITE, H. G.; GLERIANI, J. M. et al. Estimation of Eucalyptus productivity using efficient artificial neural network. **European Journal of Forest Research**, 141, n. 1, p. 129–151, 2022.

PAVLOV, I. P. **Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex**. Oxford, England: Oxford Univ. Press, 1927.

REIS, L. P.; DE SOUZA, A. L.; DOS REIS REIS, P. C. M. et al. Modeling of tree recruitment by artificial neural networks after wood harvesting in a forest in eastern amazon rain forest. **Ciencia Florestal**, 29, n. 2, p. 583–594, 2019.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. Model-Agnostic Interpretability of Machine Learning. **ArXiv**, v. abs/1606.0, 2016.

RIS-ALA, R. **Fundamentos de Aprendizagem por Reforço**. Rafael Ris-Ala, 2023.

ROCHA, S. J. S. S. DA; TORRES, C. M. M. E.; VILLANOVA, P. H.; et al. Drought effects on carbon dynamics of trees in a secondary Atlantic Forest. **Forest Ecology and Management**, 465, p. 118097, 2020.

ROCHA, S. J. S. S.; TORRES, C. M. M. E.; VILLANOVA, P. H.; et al. Machine learning methods: Modeling net growth in the Atlantic Forest of Brazil. **Ecological Informatics**, 81, n. July 2023, 2024.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, 1958. US: American Psychological Association.

RUDIN, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. **Nature Machine Intelligence**, 1, n. 5, p. 206–215, 2019.

RUDIN, C.; CHEN, C.; CHEN, Z.; et al. Interpretable machine learning: Fundamental principles and 10 grand challenges. **Statistics Surveys**, 16, n. none, 2022.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, 323, n. 6088, p. 533–536, 1986.

RYO, M. Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. **Artificial Intelligence in Agriculture**, 6, p. 257–265, 2022.

RYO, M.; ANGELOV, B.; MAMMOLA, S.; et al. Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. **Ecography**, 44, n. 2, p. 199–205, 2021.

SANTOS, J. S.; MENDONÇA, A. R. DE; GONÇALVES, F. G.; et al. Predicting eucalyptus plantation growth and yield using Landsat imagery in Minas Gerais, Brazil. **Ecological Informatics**, 75, p. 102120, 2023.

SHAFIEE, M. J.; CHYWL, B.; LI, F.; WONG, A. Fast YOLO: A Fast You Only Look Once System for Real-time Embedded Object Detection in Video. **Journal of Computational Vision and Imaging Systems**, 3, 2017.

SILVA, G. C. C.; NEVES, J. C. L.; MARCATTI, G. E.; et al. Improving 3-PG calibration and parameterization using artificial neural networks. **Ecological Modelling**, 479, n. August 2022, p. 110301, 2023.

SKINNER, B. F. DIAGRAMMING SCHEDULES OF REINFORCEMENT. **Journal of the Experimental Analysis of Behavior**, v. 1, n. 1, p. 67–68, 1958.

SKINNER, B. F. **About behaviorism**. Oxford, England: Alfred A. Knopf, 1974.

STADDON, J. E. R. **The New Behaviorism: Mind, Mechanism, and Society**. Psychology Press, 2001.

TABILE, A. F.; JACOMETO, M. C. D. Fatores influenciadores no processo de aprendizagem: um estudo de caso. **Revista Psicopedagogia**, 34, p. 75–86, 2017.

TAVARES JÚNIOR, I. S.; DE SOUZA, J. R. M.; LOPES, L. S. S.; et al. Machine learning and regression models to predict multiple tree stem volumes for teak. **Southern Forests: a Journal of Forest Science**, 83, n. 4, p. 294–302, 2021.

THORNDIKE, E. L.; WOODWORTH, R. S. The influence of improvement in one mental function upon the efficiency of other functions. (I). **Psychological Review**, 8, p. 247–261, 1901.

VAPNIK, V. The Support Vector Method of Function Estimation. In: J. A. K. Suykens; J. Vandewalle (Orgs.); **Nonlinear Modeling: Advanced Black-Box Techniques**. p. 55–85, 1998. Boston, MA: Springer US.

VIEIRA, S.; LOPEZ PINAYA, W. H.; MECHELLI, A. Introduction to machine learning. **Machine Learning**. p.1–20, 2020. Elsevier.

WATSON, J. B. Psychology as the behaviorist views it. **Psychological Review**, 20, n. 2, p. 158–177, 1913.

WERBOS, P. J. Backpropagation through time: what it does and how to do it. **Proceedings of the IEEE**, 78, n. 10, p. 1550–1560, 1990.

WERBOS, P. J. Intelligence in the brain: A theory of how it works and how to build it. **Neural Networks**, 22, n. 3, p. 200–212, 2009.

WIDROW, B.; HOFF, M. E. **Adaptive switching circuits**, 1960 IRE WESCON Convention Record, New York: IRE, p. 96-104, 1960.

WIEDERHOLD, G.; MCCARTHY, J. Arthur Samuel: Pioneer in Machine Learning. **IBM Journal of Research and Development**, 36, n. 3, p. 329–331, 1992.

ZHOU, Z.-H. **Machine Learning**. Singapore: Springer Singapore, 2021.

ZILIO, D.; CARRARA, K. Mentalismo e explicação do comportamento: aspectos da crítica behaviorista radical à ciência cognitiva. **Acta Comportamentalia: Revista Latina de Análisis del Comportamiento**, 16, n. 3, 2010.

## **CAPÍTULO 1: Técnicas *post-hoc* de aprendizado de máquina interpretável para predição da produtividade de povoamentos de eucalipto**

### RESUMO:

Os objetivos do estudo foram testar algoritmos de aprendizado de máquina para prever a produtividade média de eucalipto ao final da rotação e analisar o efeito das principais variáveis preditoras a partir de técnicas *post-hoc* de aprendizado de máquina interpretável. Utilizou-se um banco de dados com 320 talhões localizados em povoamentos de eucalipto no norte de Minas Gerais. Três algoritmos de aprendizado de máquina foram treinados (*Random Forest*, *Gradient Boosting Machine* e Árvore de Decisão) para prever a produtividade média do eucalipto aos 7 anos ( $IMA_7$ ) com base em 304 variáveis preditoras. As variáveis preditoras foram categorizadas em silviculturais (operacionais, edáficas e de manejo) e ambientais (agroclimáticas e complementares climáticas), e diferentes composições de variáveis foram adotadas para o treinamento. As técnicas de interpretabilidade, como gráficos de dependência parcial (PDP), curvas de expectativa condicional individual (ICE) e gráfico de efeitos locais acumulados (ALE), foram aplicadas no algoritmo com melhor desempenho estatístico. Os algoritmos *Random Forest* e *Gradient Boosting Machine* apresentaram valores mais adequados para as estatísticas de avaliação de treinamento e generalização, prevendo o  $IMA_7$  com maior precisão. As variáveis mais importantes para a predição da produtividade aos 7 anos foram altitude, teor de matéria orgânica do solo, teor de argila e número de dias úmidos com um ano de idade. A altitude teve o maior efeito sobre o  $IMA_7$  ( $9,7 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$ ), beneficiando os povoamentos localizados entre 860 m e 940 m com um efeito positivo sobre a produtividade. As variáveis edáficas, teor de argila e teor de matéria orgânica no solo, exerceram um efeito total de  $4,9 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$  e  $2,4 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$  sobre o  $IMA_7$ , respectivamente. As variáveis silviculturais operacionais não foram importantes na predição da produtividade. As técnicas de interpretabilidade de aprendizado de máquina *post-hoc* são eficientes para identificar e quantificar as relações entre as variáveis preditoras de maior importância relativa e a produtividade média aos sete anos de idade em povoamentos de eucalipto na região de estudo.

Palavras-chave: Altitude. Modelagem. Silvicultura.

## CHAPTER 1: Post-hoc interpretable machine learning techniques for predicting eucalypt stands productivity

### ABSTRACT:

The objectives of the study were to test machine learning algorithms to predict the average productivity of eucalypt at the end of the rotation and to analyze the effect of the main predictor variables using post-hoc interpretable machine learning techniques. A database with 320 plots located in eucalypt stands in northern Minas Gerais was used. Three machine learning algorithms were trained (Random Forest, Gradient Boosting Machine, and Decision Tree) to predict the average productivity of eucalypt at 7 years ( $IMA_7$ ) based on 304 predictor variables. The predictor variables were categorized into silvicultural (operational, edaphic, and management) and environmental (agroclimatic and complementary climatic), and different compositions of variables were adopted for training. Interpretability techniques such as partial dependence plots (PDP), individual conditional expectation (ICE) curves, and accumulated local effects (ALE) plots were applied to the algorithm with the best statistical performance. The Random Forest and Gradient Boosting Machine algorithms presented more suitable values for training and generalization evaluation statistics, predicting  $IMA_7$  with greater accuracy. The most important variables for predicting productivity at 7 years were altitude, soil organic matter content, clay content, and the number of humid days at one year of age. Altitude had the greatest effect on  $IMA_7$  ( $9.7 \text{ m}^3\text{ha}^{-1}\text{year}^{-1}$ ), benefiting stands located between 860 m and 940 m with a positive effect on productivity. The edaphic variables, clay content and soil organic matter content, had a total effect of  $4.9 \text{ m}^3\text{ha}^{-1}\text{year}^{-1}$  and  $2.4 \text{ m}^3\text{ha}^{-1}\text{year}^{-1}$  on  $IMA_7$ , respectively. Operational silvicultural variables were not important in predicting productivity. Post-hoc machine learning interpretability techniques are efficient for identifying and quantifying the relationships between the most important predictor variables and the average productivity at seven years of age in eucalypt stands in the study region.

Key-words: Altitude. Modeling. Silviculture.

## 1. INTRODUÇÃO

A prognose da produção florestal é um dos elementos essenciais do manejo florestal, sendo indispensável para a tomada de decisão e manutenção da sustentabilidade da produção dos reflorestamentos (Campos e Leite, 2017). A determinação da produção dos povoamentos em diferentes idades é realizada, na maioria das vezes, utilizando modelos denominados como modelos de crescimento e produção. Tais modelos podem ser classificados de acordo com o nível de informação gerada, em modelos de povoamento total, modelos de distribuição de diâmetros e/ou classe de tamanho e modelos de árvore individual (Burkhardt; Tomé, 2012).

Os modelos em nível de povoamento utilizam variáveis que representam todo o povoamento para expressar o crescimento e a produção, como idade ( $I$ ), área basal ( $B$ ), índice de sítio ( $S$ ) e outras (Davis; Johnson, 1987). Esses modelos podem ser subdivididos em dois tipos: modelos de densidade variável que incluem a densidade como um componente dinâmico para projeção e predição da produção futura ( $Y_2 = f(I_1, I_2, S, B_1, B_2)$ ), e em modelos do tipo normal, baseados em relações funcionais entre produção vs. idade, sítio ou área (Clutter, 1983; Guera et al., 2019; Oliveira et al., 2009; Santos et al., 2017).

Nos modelos de distribuição de diâmetros, a variável resposta é o número de árvores por hectare, área basal ou volume, em cada classe de diâmetro, nas idades atual e futura (Burkhardt, 2008; Campos e Leite, 2017). Esse tipo de modelo é geralmente utilizado juntamente com equações individuais de altura, volume e *taper*, resultando na estimativa da produção e do sortimento de múltiplos produtos (Binoti et al., 2017; Leite et al., 2005). Os modelos classificados como modelos de árvores individuais utilizam a árvore como unidade básica de modelagem, projetando o incremento de variáveis dendrométricas como o diâmetro a 1,3 m de altura ( $d_{1.3}$ ), altura total, volume e mortalidade (Liu e Ashton, 1995; Sharma et al., 2019).

Os modelos tradicionais de crescimento e produção são em sua maioria baseados em modelagem estatística e técnicas de regressão (Campos e Leite, 2017). Apesar da grande adoção desses modelos para realização da prognose de florestas equiâneas, com a difusão dos métodos de inteligência computacional, nas últimas décadas outras técnicas também começaram a ser difundidas com o mesmo fim, como é o caso dos algoritmos de aprendizado de máquina (Binoti et al., 2022).

Em um primeiro momento, as pesquisas se concentraram em demonstrar que a aplicação de métodos de aprendizado de máquina para prognose da produção de florestas gerava previsões com exatidão igual ou superior aos métodos tradicionais (Gorgens et al., 2009; Silva et al., 2009). No entanto, atualmente, é necessário ir além dessa premissa e buscar compreender como esses algoritmos aprendem e como o relacionamento entre as variáveis pode afetar as previsões dos modelos, acrescentando interpretação dos resultados obtidos sem comprometer o desempenho preditivo (Bai et al., 2024; Garnaik et al., 2022).

A compreensão das relações estabelecidas entre as variáveis preditoras e a previsão da produção florestal é essencial para estabelecer um planejamento estratégico consistente e alinhado dentro do manejo florestal (Binoti et al., 2022). A investigação acerca do relacionamento entre as variáveis e a produção dos povoamentos de eucalipto utilizando aprendizado de máquina já foi abordada utilizando diferentes técnicas como análise de sensibilidade (Freitas et al., 2020) perturbação de variáveis (Binoti et al., 2022), métodos multivariados (Oliveira Neto et al., 2022) e outras. No entanto, a aplicação de técnicas de interpretabilidade de aprendizado de máquina como os gráficos de dependência parcial (PDP); curvas de expectativas condicionais individuais (ICE) e gráficos de efeitos locais acumulados (ALE), ainda não foi testada para investigar as relações entre variáveis ambientais, edafoclimáticas, silviculturais e operacionais na produtividade de florestas equiâneas.

A interpretabilidade de modelos de aprendizado de máquina tem sido uma questão amplamente debatida dentro da inteligência artificial e seus domínios, visto que a possibilidade de maior compreensão das respostas geradas pelos algoritmos gera maior segurança e transparência na tomada de decisão (Lisboa et al., 2023; Rudin et al., 2022). Para o manejo florestal, especificamente, o uso de métodos interpretáveis pode resultar em ganhos expressivos em termos de realismo biológico, na simulação de cenários que incluem mudanças climáticas e operacionais e seus impactos na produtividade, na redução de custos e na avaliação da sensibilidade dos indicadores econômicos.

Neste estudo foram desenvolvidos e avaliados modelos de aprendizado de máquina para prever a produtividade de eucalipto aos 7 anos. As variáveis preditoras utilizadas incluíram elementos do clima e características edáficas, silviculturais e operacionais. Foram utilizadas técnicas de aprendizado de máquina interpretáveis para identificar e caracterizar as relações entre as variáveis preditoras

de maior importância e a produtividade. Os objetivos do estudo foram testar algoritmos de aprendizado de máquina para prever a produtividade média de eucalipto ao final da rotação aos sete anos e analisar o efeito das principais variáveis preditoras utilizando de técnicas *post-hoc* de aprendizado de máquina interpretável.

## 2. MATERIAIS E MÉTODOS

### 2.1. Área de estudo e dados utilizados

Os dados utilizados neste estudo foram provenientes de inventários florestais contínuos conduzidos em povoamentos de clones de híbridos de *Eucalyptus urophylla* X *Eucalyptus grandis* localizados em onze municípios do norte do estado de Minas Gerais, Brasil. A vegetação da região é caracterizada com áreas de Cerrado, Campo Cerrado e Campo de Altitude (IGA, 2012). A classificação climática de Köppen é de clima tropical úmido (Aw), com presença de um período de inverno seco e verão com chuvas (Alvares et al., 2014). A temperatura média anual na região varia entre 22 e 24 °C e os índices pluviométricos entre 500 e 1.200 mm anuais (Alvares et al., 2014; Carneiro, 2003). A altitude nos povoamentos amostrados variou entre 750 m e 1020 m acima do nível do mar, com altitude média de 902,8 m.

Os dados foram provenientes de 832 parcelas, distribuídas em 320 talhões de aproximadamente 25 ha. As parcelas permanentes foram mensuradas em no mínimo duas ocasiões e no máximo sete, no período entre 2010 e 2022. Para a obtenção da produtividade média aos 7 anos ( $IMA_7$ ), foram projetadas as últimas medições de cada um dos talhões a partir de equações de crescimento e produção específicas para a idade de rotação, correspondente a 7 anos.

### 2.2. Tipos de variáveis

As variáveis preditoras foram classificadas em dois grupos baseados nas fontes de obtenção: variáveis silviculturais e variáveis ambientais. As estatísticas descritivas das variáveis preditoras e os coeficientes de correlação entre as variáveis e o  $IMA_7$  estão nos APÊNDICE A e APÊNDICE B, respectivamente.

#### Variáveis silviculturais

O grupo de variáveis silviculturais inclui as variáveis obtidas a partir das intervenções silviculturais aplicadas aos povoamentos e registradas quantitativa ou qualitativamente. Tais características foram subdivididas em operacionais; edáficas e de manejo (Tabela 1). As variáveis operacionais correspondem aos números de dias adiantados ou atrasados em relação a data de referência para execução da atividade. Nos talhões onde a atividade foi realizada na data de referência, os registros das variáveis operacionais correspondem a zero.

Para os talhões onde a atividade foi executada na data correta, foram realizadas simulações, baseando-se na distribuição observada da variável. A Figura 1 demonstra as distribuições observadas das variáveis silviculturais operacionais e as distribuições simuladas. Durante a etapa de simulação dos valores, foram estabelecidos valores mínimos e máximos, a partir da distribuição observada, visando evitar a simulação de valores extremos. As variáveis foram simuladas a partir de uma distribuição normal, utilizando a função “*rnorm*” do *software* R, de forma que a média e o desvio-padrão das simulações correspondessem as estatísticas das variáveis observadas. A simulação desses valores visou permitir o treinamento dos algoritmos de aprendizado de máquina com todas as instâncias completas, possibilitando a aplicação das técnicas de interpretabilidade.

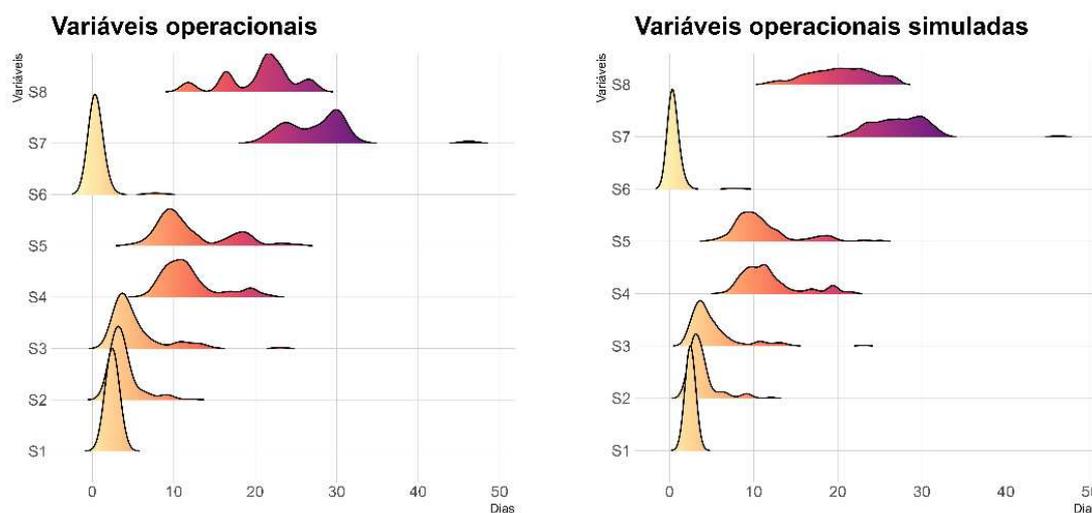
Tabela 1: Médias e desvio padrão das variáveis preditoras silviculturais utilizadas para prever a produtividade média aos sete anos em povoamentos de eucalipto no norte de Minas Gerais

Tipo	Sigla	Variável	Descrição	Média	Desvio-padrão
Operacionais	S1	Capina mecânica em linha total com pré-emergente	Dias em relação à data de referência*	2,48	± 0,32
	S2	Primeira adubação de manutenção		3,78	± 1,71
	S3	Primeira capina manual em linha total		5,09	± 3
	S4	Segunda adubação de manutenção		11,83	± 3,42
	S5	Segunda capina manual em linha total		11,10	± 3,71
	S6	Subsolagem com fosfatagem		0,54	± 0,93
	S7	Terceira adubação de manutenção		27,33	± 3,45
	S8	Terceira capina manual em linha total		20,37	± 3,72
Edáficas	CLAY	Teor de argila	Porcentagem	30,45	± 16,64
	ORGMATTER	Teor de matéria orgânica	Porcentagem	30,45	± 1,76
	P	Quantidade de adubação de Fósforo	kg/ha	81,11	± 16,19
	N	Quantidade de adubação de Nitrogênio	kg/ha	54,75	± 21,32
	K	Quantidade de adubação de Potássio	kg/ha	188,13	± 46,54
Manejo	PlantArea	Área útil por planta	m <sup>2</sup> por planta	9,98	± 1,18
	GEN	Genótipo	Identificação do clone	8 genótipos	

\* N<sup>o</sup> de dias para realização da atividade, em relação ao momento do plantio: S1 = 90 dias; S2 = 90 dias; S3 = 366 dias; S4 = 180 dias; S5 = -30 dias; S6 = 360 dias; S7 = 720; S8 = 60 dias.

Fonte: Autor (2024).

Figura 1: Distribuições observadas e simuladas das variáveis silviculturais operacionais utilizadas na predição da produtividade média de povoamentos de eucalipto aos 7 anos (*IMA<sub>7</sub>*)



Fonte: Autor (2024).

### Variáveis ambientais

As variáveis ambientais foram obtidas do Projeto de Previsão de Recursos Energéticos Mundiais (NASA POWER), utilizando API em linguagem de programação R (Sparks, 2024). As variáveis preditoras ambientais foram classificadas e denominadas em dois tipos: agroclimáticas (*Agroclimatology* no sistema NASA POWER) e climáticas complementares. Foram coletados 27 atributos em escala diária para cada parcela de inventário, considerando o intervalo de anos entre 2010 e 2022. As variáveis agroclimáticas foram classificadas segundo o sistema do NASA POWER em variáveis de fluxo solar, temperatura e fluxo infravermelho, umidade e precipitação, vento e pressão e propriedades de solo (Tabela 2). Foram obtidos 3,5 milhões de registros para cada variável agroclimática. Posteriormente, esses dados foram compatibilizados por talhão, com médias e totais mensais, e por ano correspondente a idade do povoamento, onde o ano de implantação é equivalente ao ano 0 e a idade de rotação é igual a 7 anos. A compatibilização por ano aumenta o número de recursos do banco de dados, uma vez que cada atributo ambiental passa a ser representado por oito variáveis (do ano 0 ao ano 7). Foram obtidas 224 variáveis agroclimáticas. As altitudes de cada talhão também foram obtidas pelo NASA POWER, no entanto, constitui apenas uma variável, sem necessidade de compatibilização por ano.

Tabela 2: Descrição das variáveis agroclimáticas preditoras coletadas no sistema NASA POWER para prever a produtividade média de povoamentos de eucalipto no norte de Minas Gerais

Tipo de variável	Sigla	Variável	Unidade
Altitude	ALTITUDE	Altitude em m acima do nível do mar	m
Fluxo solar	ALLSKY_SFC_UVA	Irradiância UVA de toda a superfície do céu	W/m <sup>2</sup>
	ALLSKY_SFC_UVB	Irradiância UVB de toda a superfície do céu	W/m <sup>2</sup>
	ALLSKY_SFC_UV_INDEX	Índice UV de toda a superfície do céu	Adimensional
	ALLSKY_SFC_SW_DWN	Irradiância descendente de ondas curtas da superfície do céu	MJ/m <sup>2</sup> /dia
	ALLSKY_KT	Índice de clareza de insolação de todo o céu	Adimensional
	TOA_SW_DWN	Irradiância descendente de ondas curtas no topo da atmosfera	MJ/m <sup>2</sup> /dia
	CLRSKY_SFC_SW_DWN	Irradiância descendente de ondas longas da superfície do céu claro	W/m <sup>2</sup>
	ALLSKY_SFC_LW_DWN	Irradiância descendente de ondas longas da superfície do céu	W/m <sup>2</sup>
	ALLSKY_SFC_PAR_TOT	Radiação fotossinteticamente ativa (PAR) total de toda a superfície do céu	W/m <sup>2</sup>
	CLRSKY_SFC_PAR_TOT	Radiação fotossinteticamente ativa (PAR) total de toda a superfície do céu claro	W/m <sup>2</sup>
Temperatura e fluxo infravermelho	T2M	Temperatura em 2 m	°C
	T2M_MAX	Temperatura máxima em 2 m	°C
	T2M_MIN	Temperatura mínima em 2 m	°C
	T2MWET	Temperatura no bulbo úmido em 2 m	°C
	T2MDEW	Ponto de orvalho/geada em 2 m	°C
	T2M_RANGE	Faixa de variação diária de temperatura em 2 m	°C
	TS	Temperatura da superfície da terra	°C
Umidade e precipitação	QV2M	Umidade específica a 2 m	(g/kg)
	RH2M	Umidade relativa a 2 metros	%
	PRECTOTCORR	Precipitação corrigida	mm
	PRECTOTCORR A	Precipitação acumulada anual	mm
	WS2M	Velocidade do vento em 2 m	ms <sup>-1</sup>

Tipo de variável	Sigla	Variável	Unidade
	TQV	Coluna total de água precipitável	kg m <sup>-2</sup>
	WS10M	Velocidade do vento em 10 m	ms <sup>-1</sup>
Propriedades do solo	GWETPROF	Porcentagem de umidade do solo do perfil	%
	GWETTOP	Porcentagem de umidade do solo (até 5 cm)	%
	GWETROOT	Porcentagem de umidade do solo na zona radicular (até 100 cm)	%

Fonte: Autor (2024).

Além das variáveis agroclimáticas, foram introduzidas no conjunto de dados de variáveis preditoras denominadas de *climáticas complementares*. Tais variáveis resumem, em número de dias ou em porcentagem, condições climáticas específicas. Foram calculadas oito variáveis a partir da temperatura a 2 m (T2M); Umidade relativa a 2 m (RH2M) e precipitação corrigida (PRECTOTCORR). Para estabelecer os critérios de dias quentes, frios, úmidos e secos obteve-se primeiramente o percentil 90% e 10% dos dados diários anuais. Os registros diários de temperatura e umidade que atendiam aos critérios de inclusão estabelecidos pelos percentis eram contabilizados (Tabela 3). A compatibilização dessas variáveis correspondeu ao total de dias no ano com os critérios atendidos.

A precipitação corrigida foi utilizada para gerar os números de dias com ocorrência de precipitação, onde o critério utilizado foi o de precipitação em mm > 0. Para obtenção da porcentagem de chuva no período seco (considerado para o estado de Minas Gerais os meses entre junho e setembro), utilizou-se a relação entre o total de precipitação no período seco / total de precipitação anual x 100. O número de dias de temperatura ótima foi contabilizado a partir da faixa de variação de °C ≥ 18 - °C ≤ 22 para T2M, conforme indicado por Queiroz et al. (2020).

Tabela 3: Descrição das variáveis ambientais complementares preditoras para predição da produtividade média de povoamentos de eucalipto aos 7 anos (IMA<sub>7</sub>)

Variável	Sigla	Critério
Número de dias quentes	DQ	°C ≥ percentil 90% anual
Número de dias frios	DF	°C ≤ percentil 10% anual
Número de dias de temperatura ótima	D_otima	°C ≥ 18 - °C ≤ 22
Número de dias úmidos	DU	°C ≥ percentil 90% anual
Número de dias secos	DS	°C ≤ percentil 10% anual
Número de dias com precipitação	DC	mm > 0
Número de dias sem precipitação	DSC	mm ≤ 0
% de chuva no período seco	Chuva%seco	$\frac{\text{precipitação no período seco}}{\text{precipitação total anual}} \times 100$

Fonte: Autor (2024).

### 2.3. Processamento e análise dos dados

O pré-processamento adotado para o treinamento dos algoritmos incluiu a padronização das variáveis quantitativas (Equação 1). As variáveis qualitativas, como genótipo, foram transformadas em variáveis *dummy*, correspondentes aos 8 genótipos presentes no banco de dados.

$$Z_1 = (x_i - \bar{x})/\sigma \quad (1)$$

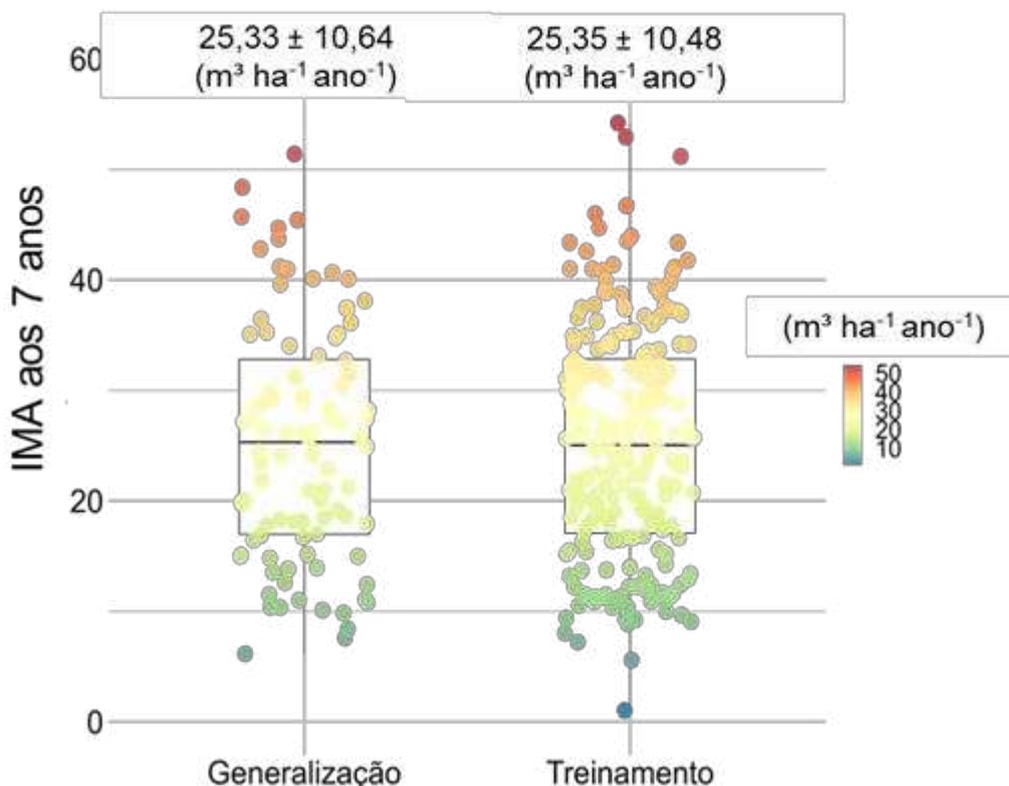
Em que:  $Z_1$  = Valores padronizados;  $x_i$  = valor da  $i$ -ésima observação;  $\bar{x}$  = média dos valores observados;  $\sigma$  = desvio-padrão.

O conjunto de dados foi separado em conjunto de treinamento e validação, composto por 70% dos talhões (224) e conjunto de generalização com 30% (96) dos dados (Figura 2). Foram treinados três algoritmos de aprendizado de máquina: *random forest* (RF), árvore de decisão (CART) e *Gradient Boosting Machine* (GBM).

Os parâmetros testados durante a etapa de treinamento do RF foram o número de árvores (*n**tree*, número de variáveis preditoras  $\times$  10) e o número de variáveis de divisão (*m**try*, testados entre 2 e o número de variáveis preditoras). Para o GBM, foram testados o número de árvores variando entre 50 e 500 e utilizada a taxa de aprendizagem = 0,1. Para o algoritmo CART, foram otimizados os parâmetros de taxas de poda (CP) e o número de nós internos. As recomendações dos parâmetros de treinamento foram obtidas em Boehmke e Greenwell (2019).

Os parâmetros dos algoritmos foram otimizados durante o treinamento, com processo padronizado utilizando validação cruzada (*k-fold cross-validation*) com três repetições (*repeats* = 3) e dez partições (*number* = 10). O particionamento do conjunto de dados e os treinamentos dos algoritmos foram realizados no software R com a biblioteca *Caret* (Kuhn, 2008).

Figura 2: Estatísticas descritivas (média e desvio-padrão) para a produtividade média de povoamentos de eucalipto aos 7 anos ( $IMA_7$ ) nos conjuntos de treinamento e generalização



Fonte: Autor (2024).

A avaliação do desempenho preditivo dos algoritmos de aprendizado de máquina foi realizada a partir de estatísticas de precisão e exatidão e, também, análise gráfica de resíduos, conforme indicado por Campos e Leite (2017), considerando tanto o conjunto de treinamento quanto o conjunto de generalização. Foram utilizadas as estatísticas: coeficiente de correlação entre o  $IMA_7$  observado e o  $IMA_7$  predito pelos algoritmos ( $r_{\hat{y}y}$ ), raiz quadrada do erro quadrático médio em termos absolutos (RQEM) e relativos (RQEM%) e média das diferenças absolutas (MAE). Além disso, foram construídos e interpretados gráficos conjuntos de valores observados e correspondentes valores preditos e de distribuição das diferenças entre valores preditos e correspondentes valores observados (Tabela 4).

Tabela 4: Estatísticas de avaliação do treinamento e da generalização para algoritmos de aprendizado de máquina na predição da produtividade média de povoamentos de eucalipto

Estatística	Estimador
Coefficiente de correlação entre o $IMA_7$ observado e $IMA_7$ predito	$r_{\hat{y}y} = \frac{(\sum_i^n (IMA_{7i} - \overline{IMA}_{7i})(\widehat{IMA}_{7i} - \overline{\widehat{IMA}}_7))^2}{\sqrt{\sum_i^n (IMA_{7i} - \overline{IMA}_{7i})^2 \times \sum_i^n (\widehat{IMA}_{7i} - \overline{\widehat{IMA}}_7)^2}}$
Raiz quadrada do erro quadrático médio ( $m^3ha^{-1}ano^{-1}$ )	$RQEM = \sqrt{\frac{\sum_{i=1}^n (IMA_{7i} - \widehat{IMA}_{7i})^2}{n}}$
Raiz quadrada do erro quadrático médio relativo (%)	$RQEM (\%) = 100 \overline{IMA}_7^{-1} \sqrt{\frac{\sum_{i=1}^n (IMA_{7i} - \widehat{IMA}_{7i})^2}{n}}$
Média das diferenças absolutas ( $m^3ha^{-1}ano^{-1}$ )	$MAE = \frac{\sum_{i=1}^n  IMA_{7i} - \widehat{IMA}_{7i} }{n}$
Resíduo ( $m^3ha^{-1}ano^{-1}$ )	Resíduo = $\widehat{IMA}_{7i} - IMA_{7i}$

$IMA_{7i}$ : Produtividade anual ( $m^3ha^{-1}ano^{-1}$ ) na idade final de rotação de 7 anos observada no talhão;  $\overline{IMA}_{7i}$ : Média da produtividade anual ( $m^3ha^{-1}ano^{-1}$ ) na idade final de rotação de 7 anos observada;  $\widehat{IMA}_{7i}$ : Produtividade anual predita ( $m^3ha^{-1}ano^{-1}$ ) na idade final de rotação de 7 anos observada no talhão; n: número de observações.  
Fonte: Autor (2024).

O algoritmo de aprendizado com a melhor performance preditiva, segundo as estatísticas e a análise gráfica dos resíduos, foi selecionado para a etapa de aplicação das técnicas de interpretabilidade. Primeiramente, a importância relativa das variáveis foi obtida utilizando a permutação simples e simulações. No caso da importância com

simulação, adotou-se o número de 500 simulações de Monte Carlos utilizando o MAE como métrica.

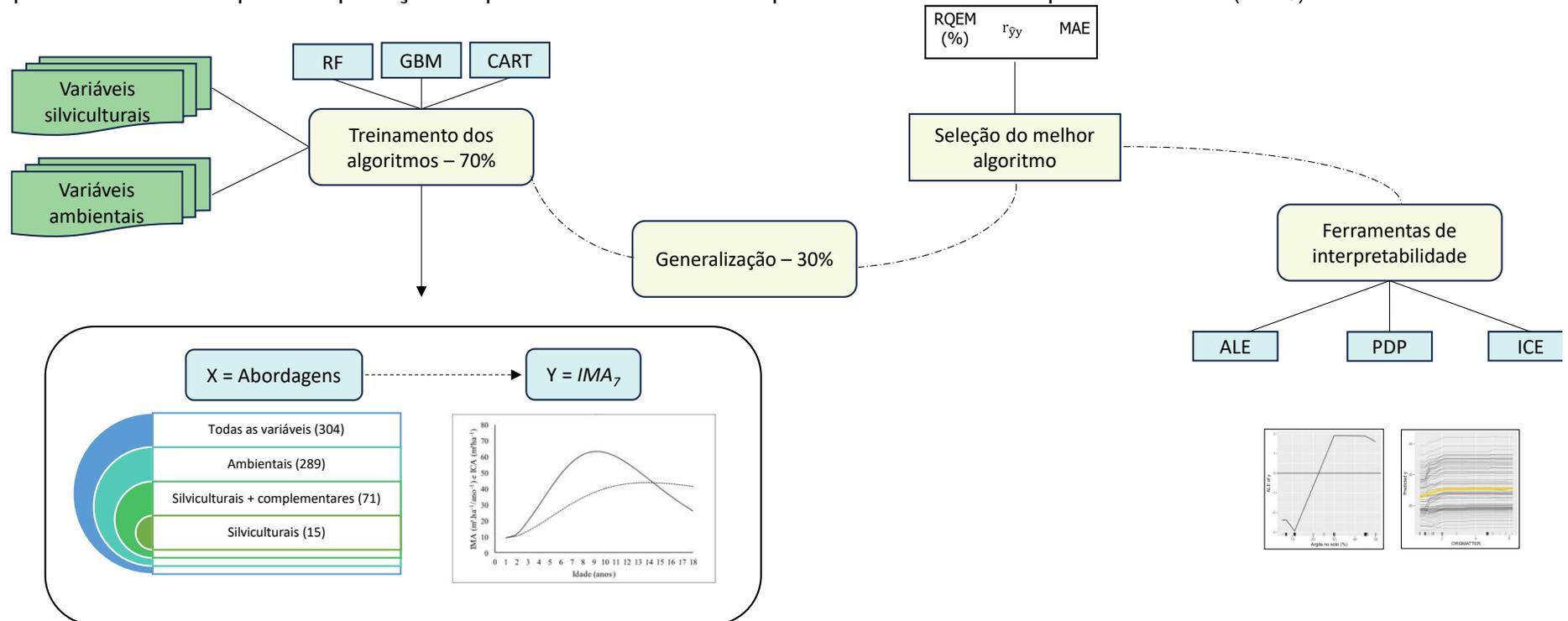
As duas variáveis com maiores valores de importância, considerando todas as simulações, foram escolhidas para gerar os gráficos de dependência parcial, curvas de expectativa condicional individual e gráfico de efeitos locais acumulados para as abordagens que apresentaram os melhores indicadores estatísticos de generalização. Para gerar as estimativas de importância das variáveis foi utilizada a biblioteca vip (Greenwell; Boehmke, 2020) e para obtenção dos gráficos de interpretabilidade o pacote “iml” (Molnar, 2018), ambos no *software* R. As técnicas foram aplicadas somente nos dados de generalização, conforme recomendado por Molnar (2020).

#### **2.4. Abordagens de treinamento**

Foram utilizadas quatro abordagens de treinamento dos algoritmos de aprendizado de máquina, cada uma com um conjunto de variáveis preditoras. A primeira abordagem incluiu as 304 variáveis preditoras (ambientais e silviculturais). A segunda consistiu no treinamento dos algoritmos somente com as 289 variáveis ambientais como preditores da produtividade média. Na terceira abordagem foi utilizado um conjunto de variáveis silviculturais e variáveis ambientais complementares, totalizando 71 variáveis. Por último, a quarta abordagem foi composta somente pelos preditores silviculturais (15).

Um resumo ilustrativo das etapas de construção do modelo de aprendizado de máquina interpretável é apresentado na Figura 3.

Figura 3: Resumo metodológico das etapas de treinamento, generalização e avaliação da interpretabilidade para modelos de aprendizado de máquina na predição da produtividade média de povoamentos de eucalipto aos 7 anos ( $IMA_7$ )



Fonte: Autor (2024).

### 3. RESULTADOS E DISCUSSÃO

Os desempenhos preditivos dos três algoritmos testados para previsão da produtividade de eucalipto estão na Tabela 5. As árvores de decisão apresentaram os menores valores de  $r_{\hat{y}y}$  e os maiores valores de RQEM (%) para todas as abordagens e conjuntos de dados. As árvores de decisão são algoritmos que se baseiam na ideia de particionar os dados usando regras de decisão (Alvares et al., 2023; Breiman et al., 1984). Embora sejam mais simples e de fácil interpretação, as árvores de decisão apresentam baixo poder preditivo em relação aos algoritmos mais recentes (Boehmke; Greenwell, 2019).

O algoritmo GBM resultou nos melhores desempenhos preditivos, com valores inferiores de RQEM (%) e maiores coeficientes de correlação entre os valores de  $IMA_7$  preditos e observados, nas abordagens com todas as variáveis e com as variáveis ambientais. Nas abordagens restantes (somente com as variáveis silviculturais e variáveis silviculturais + climáticas complementares), o algoritmo RF gerou previsões com menores valores de erros (RQEM e RQEM%) e maiores  $r_{\hat{y}y}$ . No entanto, os valores de MAE para as previsões do RF foram menores em todas as abordagens, considerando a generalização.

Os algoritmos com melhores desempenhos preditivos variaram conforme o número de variáveis presentes no espaço de recursos de previsão. Nas abordagens que utilizaram maiores números de variáveis preditoras, o algoritmo GBM apresentou melhores estatísticas de avaliação de treinamento e generalização. Ao passo que para as abordagens que incluíam números de preditores menores, o desempenho do RF foi superior aos demais algoritmos.

Os algoritmos RF e GBM partem do princípio da ideia das árvores de decisão, mas incorporam procedimentos de treinamento (como amostragem *bootstrap*, *boosting* e *bagging*) para aumentar a robustez das previsões (Friedman, 2001; Nayak et al., 2022). Além disso, o GBM utiliza uma estratégia de treinamento sequencial, construindo as árvores de decisão de forma progressiva para reduzir os erros de previsão nas observações que inicialmente possuíam maiores erros (Boehmke; Greenwell, 2019). Essas características podem atribuir maior potencial preditivo para bancos de dados com grandes números de variáveis (Bui et al., 2021; Wongchai et al., 2022).

O RF consiste em conjuntos de árvores de decisão criadas com randomização de variáveis e técnicas de *bagging* e tem sido amplamente utilizado para gerar previsões de diversos fenômenos dentro da mensuração florestal, apresentando boa performance preditiva tanto para tarefas de aproximação de funções, quanto de classificação (Lima et al., 2022; Rocha et al., 2024). Além disso, permite boa interpretabilidade em comparação a outros algoritmos, como redes neurais profundas (Breiman, 2001; Ryo, 2022).

Tabela 5: Estatísticas de avaliação para as quatro abordagens preditivas dos algoritmos de aprendizado de máquina testados para a previsão da produtividade média de povoamentos de eucalipto aos 7 anos ( $IMA_7$ )

Abordagem	Conjunto	Algoritmo	RQEM	RQEM (%)	$r_{\hat{y}y}$	MAE
Todas as variáveis	Generalização	RF	70,274	27,74	0,57	52,459
		Cart	74,380	29,36	0,52	60,265
		GBM	69,839	27,57	0,57	53,021
	Treinamento	RF	27,014	10,66	0,94	20,015
		Cart	53,086	20,94	0,74	41,234
		GBM	36,238	14,30	0,89	27,099
Variáveis ambientais	Generalização	RF	71,179	28,10	0,55	53,568
		Cart	73,453	29,00	0,52	57,314
		GBM	70,723	27,92	0,56	53,930
	Treinamento	RF	48,584	19,17	0,78	35,525
		Cart	56,361	22,24	0,71	42,005
		GBM	50,494	19,92	0,77	37,296
Variáveis silviculturais + complementares	Generalização	RF	64,123	25,31	0,65	48,115
		Cart	75,321	29,73	0,50	59,598
		GBM	67,307	26,57	0,61	53,173
	Treinamento	RF	31,082	12,26	0,92	23,263
		Cart	53,051	20,93	0,74	41,554
		GBM	46,086	18,18	0,81	34,275
Variáveis silviculturais	Generalização	RF	71,512	28,23	0,54	54,944
		Cart	82,784	32,68	0,42	64,260
		GBM	76,472	30,19	0,48	61,033
	Treinamento	RF	28,375	11,20	0,94	22,208
		Cart	52,239	20,61	0,75	41,821
		GBM	48,401	19,10	0,80	36,773

RF: *Random forest*; Cart: árvore de decisão e GBM: *Gradient Boosting Machine*.  
Fonte: Autor (2024).

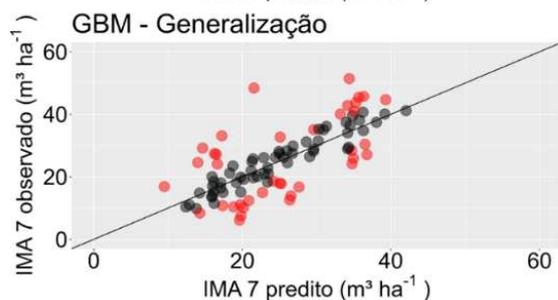
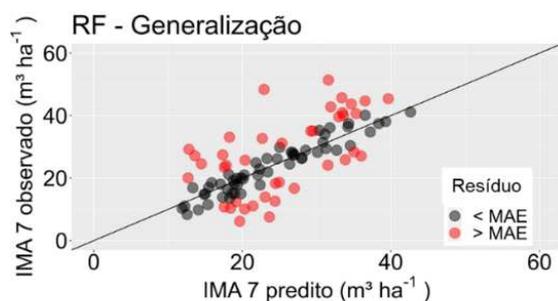
Dentre as quatro abordagens testadas, os melhores resultados preditivos foram obtidos no conjunto de dados composto somente das variáveis silviculturais e

ambientais complementares (Tabela 5). Os erros relativos de generalização para predição da produtividade média nessa abordagem variaram entre 25,31 e 29,73%. A redução de variáveis agroclimáticas por meio de variáveis complementares se mostrou uma estratégia de redução de variáveis interessante e que não compromete a exatidão das predições. O ganho em exatidão propiciado por essa abordagem pode estar relacionado a redução da redundância dos atributos causada por variáveis preditoras altamente correlacionadas, como as diferentes medidas de temperaturas e variáveis de fluxo solar (Boussaada et al., 2018; Bouzgou; Gueymard, 2017; Chan et al., 2022).

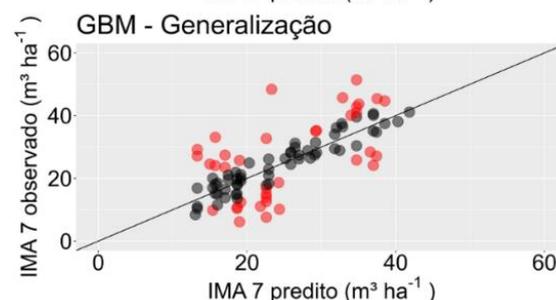
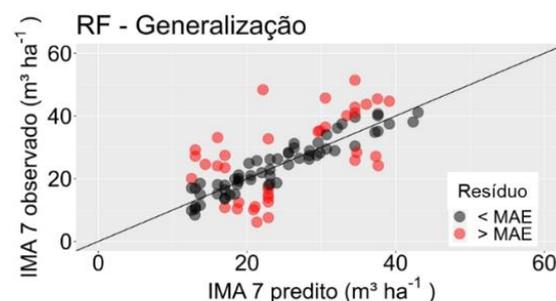
Os gráficos de  $IMA_7$  predito x  $IMA_7$  observado, para os algoritmos com as melhores estatísticas de avaliação são apresentados na Figura 4. As predições utilizando somente as variáveis silviculturais demonstram valores com maiores dispersões em torno da reta de 45°, bem como, maior número de resíduos > MAE.

Figura 4: Dispersões entre produtividades médias observadas e previstas ( $IMA_7$ ), em povoamentos de eucalipto, utilizando dados de generalização dos algoritmos *random forest* e *gradient boosting machine*

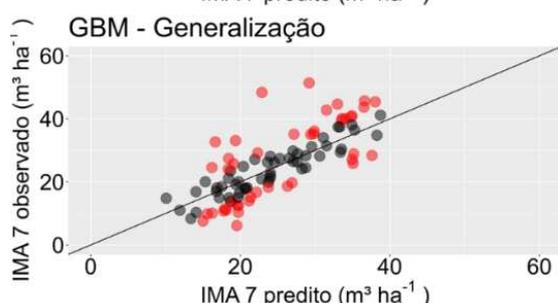
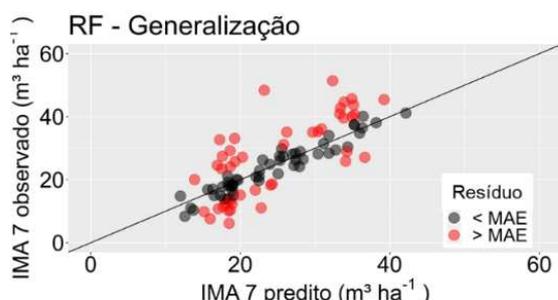
Todas as variáveis



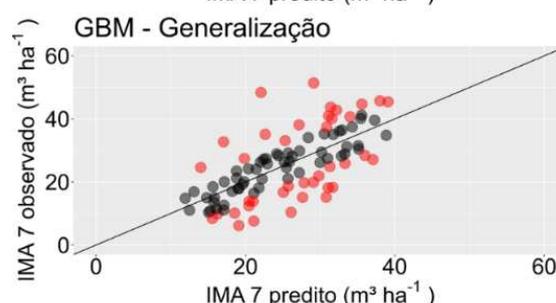
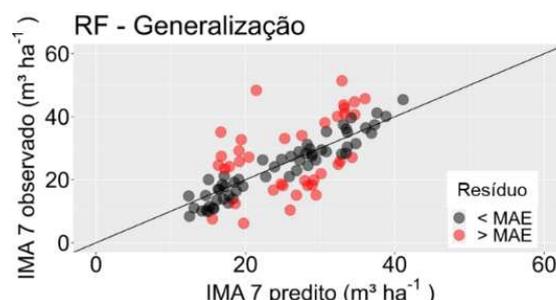
Ambientais



Silviculturais + climáticas complementares



Silviculturais



Fonte: Autor (2024).

Na Figura 5 pode ser observada a importância relativa das variáveis mais expressivas, dentro de cada abordagem de treinamento testada, considerando as previsões do algoritmo *random forest*. A altitude foi a variável preditora com maiores valores de importância relativa para a previsão da produtividade média de eucalipto em todas as abordagens que foi incluída. A altitude pode ser considerada um elemento ambiental relevante para definir a produtividade de reflorestamento, visto que é

responsável por estabelecer relacionamentos complexos com variáveis ambientais como a temperatura, radiação solar e precipitação (Sondermann et al., 2022). Essas relações são essenciais para criação de microclimas específicos que podem interferir diretamente nas condições de crescimento dos povoamentos, ocasionando em diferentes padrões ao longo do ciclo (Rody et al., 2016). Além disso, dentro dos modelos de aprendizado de máquina, a variável altitude pode estar servindo com uma *proxy* para os municípios que compõem o banco de dados, captando as diferenças topográficas existentes.

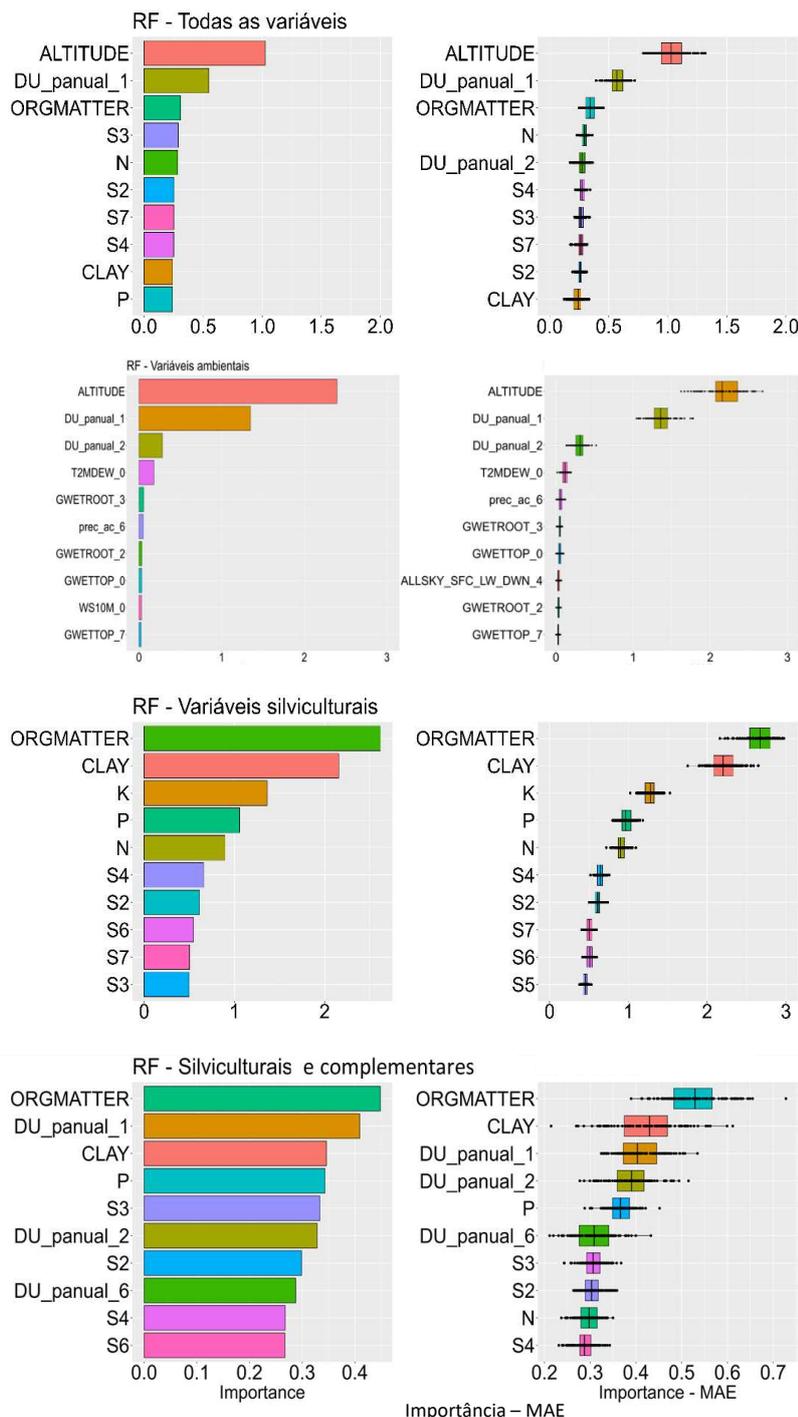
As variáveis silviculturais operacionais demonstraram pouca importância para a produtividade dos povoamentos de eucalipto, obtendo destaque somente na abordagem que incluía apenas variáveis silviculturais. No entanto, é importante ressaltar que tais variáveis têm importância econômica dentro dos planejamentos hierárquicos das empresas florestais, onde o atraso ou adiantamento das operações pode acarretar alterações significativas nos custos operacionais (Binoti et al., 2022). Dessa forma, é interessante analisar a importância desempenhada por essas variáveis de maneira conjunta com indicadores econômicos e financeiros.

O teor de matéria orgânica no solo (ORGMATTER) obteve os maiores valores de importância relativa para as abordagens que consideravam as variáveis silviculturais, demonstrando fortes relações com a produtividade dos povoamentos de eucalipto. O teor de argila (CLAY) também apresentou valores relativamente altos de importância relativa. As variáveis edáficas relacionadas ao solo, como argila e matéria orgânica, apresentam relevante contribuição para a produção de madeira de eucalipto, visto que estão associadas a processos físicos e químicos do solo, como a retenção de água, formação da estrutura, disponibilidade de nutrientes e outros (Gava e Gonçalves, 2008; Santos e Reichert, 2022). Essas variáveis já demonstraram, em estudos anteriores, possuir elevado poder preditivo dentro de algoritmos de aprendizado de máquina para predição da produtividade de eucalipto (Freitas et al., 2020).

O número de dias úmidos (DU\_p anual) nas idades iniciais dos povoamentos (1 e 2 anos) foi de grande importância relativa nas abordagens com variáveis ambientais e silviculturais. A umidade do ar é uma variável de grande relevância para o crescimento de plantas, dada a relação estreita com os processos fisiológicos como a respiração (Elli et al., 2020). Essa importância é elevada nas fases iniciais de

estabelecimento dos povoamentos em que a demanda por água e nutrientes é maior e o metabolismo é mais acelerado (Yang et al., 2022).

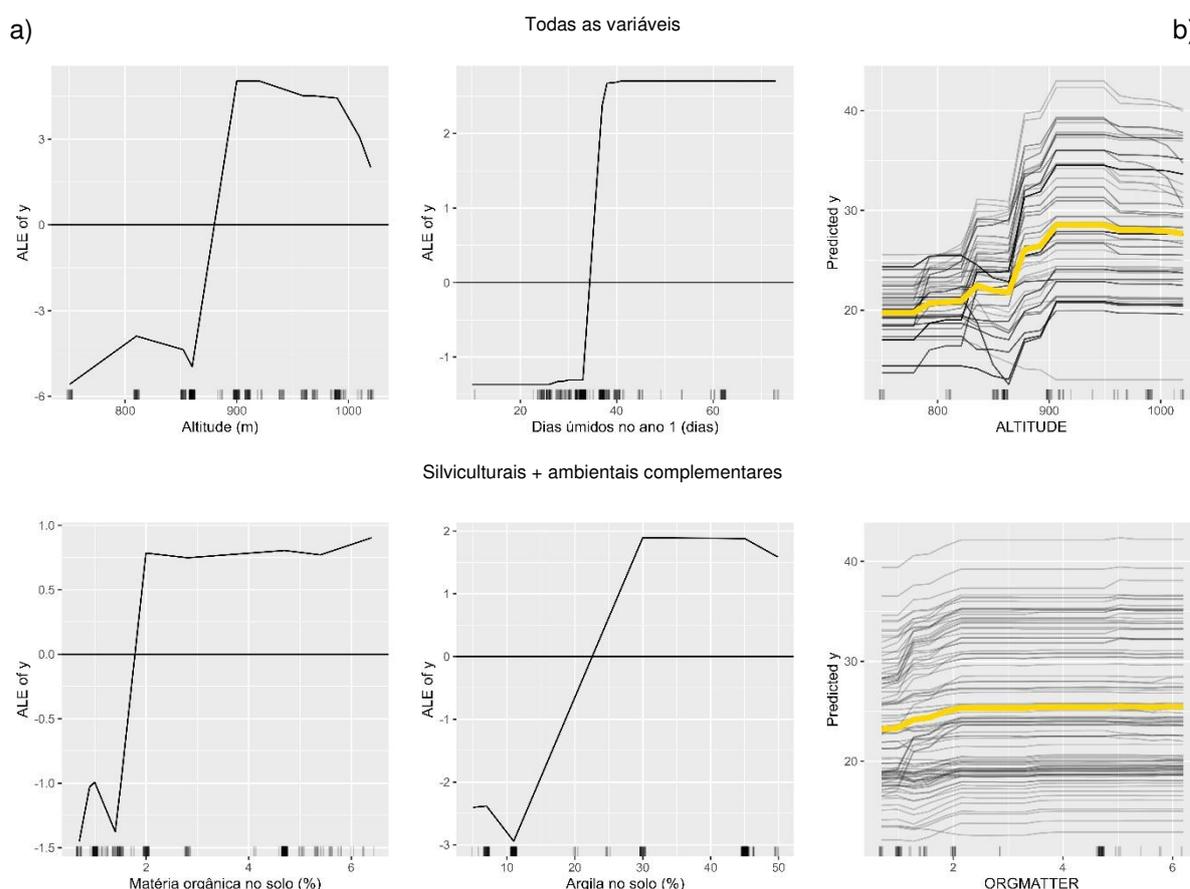
Figura 5: Importância relativa das variáveis preditoras utilizadas na predição da produtividade média de povoamentos de eucalipto aos 7 anos de idade ( $IMA_7$ ), pelo algoritmo *random forest*, em quatro diferentes abordagens de treinamento



Fonte: Autor (2024).

As ferramentas de interpretabilidade utilizadas: gráficos de dependência parcial (PDP); curvas de expectativas condicionais individuais (ICE); gráficos de efeitos locais acumulados (ALE) foram aplicadas nas duas variáveis com os maiores valores de importância relativa em cada uma das duas abordagens que obtiveram melhores desempenhos preditivos (todas as variáveis e variáveis silviculturais + ambientais complementares). Os gráficos de interpretabilidade estão na Figura 6.

Figura 6: Técnicas de interpretabilidade utilizadas para avaliar as variáveis de maior importância relativa em duas abordagens de treinamento para predição da produtividade média de povoamentos de eucalipto aos 7 anos de idade ( $IMA_7$ ), na região Norte de Minas Gerais. Em a) gráficos de efeitos locais acumulados (ALE) e em b) gráficos de dependência parcial (PDP) e curvas de expectativas condicionais individuais (ICE)



Fonte: Autor (2024).

Nos gráficos de ALE a amplitude das curvas de interação entre a variável predita e a variável preditora indica o tamanho do efeito da interação entre essas

variáveis. Dessa forma, variáveis com maiores importâncias para a predição da produtividade média aos 7 anos apresentam maiores amplitudes de curva ALE. Os gráficos de efeitos locais acumulados (Figura 6A), considerando todas as variáveis, demonstram que o efeito total da variável altitude na produtividade de eucalipto equivale a aproximadamente  $9,7 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$ . Há efeito positivo da altitude na faixa de variação de 860 m até 940 m. A partir desse limiar de altitude o  $IMA_7$  apresenta leve tendência a diminuição. Uma análise do PDP e das curvas ICE (Figura 6B) permite observar a estabilização dos valores de produtividade acima da altitude de 860 m, com  $IMA_7$ , em aproximadamente  $26 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$ . O número de dias úmidos durante o primeiro ano dos povoamentos implica em um efeito total no  $IMA_7$  de  $1,7 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$  (Figura 6A) demonstrou. Observou-se no gráfico ALE que o efeito positivo na produtividade de eucalipto ( $1,1 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$ ) é obtido nos anos em que há pelo menos 30 dias com umidade relativa do ar alta o suficiente para que o dia seja considerado úmido.

O teor de matéria orgânica e o teor de argila, foram as responsáveis pelos maiores efeitos totais no  $IMA_7$  na abordagem que incluí variáveis silviculturais e climáticas complementares (Figura 6). O efeito total do teor de argila no  $IMA_7$  foi de aproximadamente  $4,9 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$ , como demonstrado na Figura 6A. Povoamentos implantados em solos com teores de argila acima dos 22% obtiveram ganhos expressivos na produtividade média aos 7 anos. Uma leve tendência de diminuição do  $IMA_7$  é percebida em locais cujo teor de argila ultrapassa os 45%. O teor de matéria orgânica, considerada pela análise das importâncias relativas por permutação a variável de maior importância, obteve um efeito menor na produtividade média aos 7 anos, equivalente a  $2,4 \text{ m}^3\text{ha}^{-1}\text{ano}^{-1}$ . O gráfico de PDP revela uma tendência constante para essa variável, principalmente em povoamentos com solos com teores de matéria orgânica superiores a 1,8% (Figura 6B).

As curvas de expectativas condicionais individuais para ambas as abordagens demonstraram variabilidade nas predições dos talhões individualmente (Figura 6B). Esse efeito é desejado em algoritmos de aprendizado de máquina pois possibilitam gerar predições com diferentes tendências individualmente, conferindo maior confiabilidade aos modelos de predição (Ryo et al., 2022).

Diversos estudos que utilizaram métodos de avaliação dos efeitos de variáveis na predição da produtividade de eucalipto com aprendizado de máquina já foram realizados, no entanto, esses estudos indicam somente quais preditores apresentam

maior importância ou resultam em ganhos de previsão. Binoti et al. (2022) aplicaram a perturbação de variáveis em redes neurais artificiais para previsão da produção volumétrica aos 7 anos, identificando a precipitação acumulada aos 5 anos de idade como a variável mais relevante. Variáveis edáficas como altitude, teores de argila e de matéria orgânica, e variáveis de manejo, como o espaçamento, aumentaram a exatidão da produtividade estimada por RNA para povoamentos de eucalipto em diferentes regiões de Minas Gerais (Freitas et al., 2020). Na presente abordagem, identificou-se que tanto as variáveis edáficas quanto variáveis relacionadas às características ambientais que influenciam em processos fisiológicos das plantas, como a umidade relativa, fornecem maiores contribuições aos algoritmos de aprendizado de máquina testados. Além disso, identificou-se, por meio das técnicas de interpretabilidade *post-hoc*, o quanto estas variáveis influenciam na produtividade média aos 7 anos, em termos de  $m^3ha^{-1}ano^{-1}$ , diferentemente de outras técnicas de interpretabilidade baseadas somente nos ganhos de exatidão, que não possibilitam quantificar os efeitos das alterações das quantidades dos recursos nas variáveis de resposta (Adadi e Berrada, 2018; Ryo et al., 2022).

#### 4. CONCLUSÕES

Os algoritmos *Random forest* e *Gradient Boosting Machine* são indicados para modelagem da produtividade de eucalipto.

A altitude foi a variável com maior influência nas previsões do  $IMA_7$  para os povoamentos de eucalipto no norte de Minas Gerais, possivelmente como uma proxy.

Os teores de matéria orgânica e argila tem forte efeito na produtividade média de povoamentos de eucalipto, aos 7 anos de idade, nos locais onde se encontravam os povoamentos desse estudo.

As técnicas de aprendizado de máquina *post-hoc*, como gráficos de dependência parcial, curvas de expectativa condicional individual e gráfico de efeitos locais acumulados, são eficientes para identificar as relações entre as variáveis preditoras de maior importância relativa e a produtividade média ao final da rotação de povoamentos de eucalipto.

## REFERÊNCIAS

- ADADI, A.; BERRADA, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). **IEEE Access**, 6, p. 52138–52160, 2018.
- ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; et al. Köppen's climate classification map for Brazil, **Meteorologische Zeitschrift**, 22, n. 6, p. 711–728, 2013.
- ALVARES, C. A.; CEGATTA, Í. R.; SCOLFORO, H. F.; MAFIA, R. G. Decision-Tree Application to Predict and Spatialize the Wood Productivity Probabilities of Eucalyptus Plantations. **Forests**, 14, n. 7, p. 1334, 2023.
- BAI, T.; WANG, X.-S.; HAN, P. F. Controls of groundwater-dependent vegetation coverage in the yellow river basin, china: Insights from interpretable machine learning. **Journal of Hydrology**, 631, p. 130747, 2024.
- BINOTI, D. H. B.; LEITE, H. G.; DA SILVA, M. L. M. Sistema computacional para ajuste de funções densidade de probabilidade. **Floresta e Ambiente**, 24, 2017.
- BINOTI, D. H. B.; LEITE, H. G.; ANDRADE, V. et al. Importância relativa das variáveis preditoras no processo de modelagem da produtividade florestal. **Advances in Forestry Science**, 9, n. 4, p. 1867–1873, 2022.
- BOEHMKE, B.; GREENWELL, B. **Hands-On Machine Learning with R**. Chapman and Hall/CRC, 2019.
- BOUSSAADA, Z.; CUREA, O.; REMACI, A.; CAMBLONG, H.; MRABET BELLAAJ, N. A. Nonlinear Autoregressive Exogenous (NARX) Neural Network Model for the Prediction of the Daily Direct Solar Radiation. **Energies**, 11, n. 3, p. 620, 2018.
- BOUZGOU, H.; GUEYMARD, C. A. Minimum redundancy – Maximum relevance with extreme learning machines for global solar radiation forecasting: Toward an optimized dimensionality reduction for solar time series. **Solar Energy**, 158, p. 595–609, 2017.
- BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.
- BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. **Classification And Regression Trees**. 1º ed. New York: Routledge, 1984.
- BUI, Q.-T.; CHOU, T.-Y.; HOANG, T.-V. et al. Gradient Boosting Machine and Object-Based CNN for Land Cover Classification. **Remote Sensing**, 13, n. 14, p. 2709, 2021.
- BURKHART, H. Modelling Growth and Yield for Intensively Managed Forests. **Journal of forest and environmental science**, 24, n. 3, p. 119–126, 2008.
- BURKHART, H. E.; TOMÉ, M. Modeling Forest Stand Development. **Modeling Forest Trees and Stands**. p. 233–244, 2012. Dordrecht: Springer Netherlands.
- CAMPOS, J. C. C.; LEITE, H. G. **Mensuração florestal: perguntas e respostas**. 5. ed. Viçosa, MG: Editora UFV, 2017. 636 p.

CARNEIRO, M. F. B. Região Norte de Minas: caracterização geográfica e a organização espacial – breves considerações. **Revista Cerrados**, 1, n. 1, p. 91–106, 2003.

CHAN, J. Y.-L.; LEOW, S. M. H.; BEA, K. T.; et al. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. **Mathematics**, 10, n. 8, p. 1283, 2022.

CLUTTER, J. L. **Timber Management: A Quantitative Approach**. Wiley, 1983.

DAVIS, L. S.; JOHNSON, K. N. **Forest Management**. McGraw-Hill, 1987.

ELLI, E. F.; SENTELHAS, P. C.; BENDER, F. D. Impacts and uncertainties of climate change projections on Eucalyptus plantations productivity across Brazil. **Forest Ecology and Management**, 474, p. 118365, 2020.

FLORÊNCIO, G. W. L.; MARTINS, F. B.; FAGUNDES, F. F. A. Climate change on Eucalyptus plantations and adaptive measures for sustainable forestry development across Brazil. **Industrial Crops and Products**, 188, n. August, 2022.

FREITAS, E. C. S. DE; PAIVA, H. N. DE; NEVES, J. C. L.; MARCATTI, G. E.; LEITE, H. G. Modeling of eucalyptus productivity with artificial neural networks. **Industrial Crops and Products**, 146, n. February, p. 112149, 2020. Elsevier.

FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. **The Annals of Statistics**, 29, n. 5, 2001.

GARNAIK, S.; SAMANT, P. K.; MANDAL, M.; et al. Untangling the effect of soil quality on rice productivity under a 16-years long-term fertilizer experiment using conditional random forest. **Computers and Electronics in Agriculture**, 197, p. 106965, 2022.

GAVA, J. L.; GONÇALVES, J. L. DE M. Soil attributes and wood quality for pulp production in plantations of Eucalyptus grandis clone. **Scientia Agricola**, 65, n. 3, p. 306–313, 2008.

GORGENS, E. B.; LEITE, H. G.; SANTOS, H. DO N.; GLERIANI, J. M. Estimação do volume de árvores utilizando redes neurais artificiais. **Revista Árvore**, 33, n. 6, p. 1141–1147, 2009.

GREENWELL, B. M.; BOEHMKE, B. C. Variable Importance Plots—An Introduction to the vip Package. **R Journal**, 12, n. 1, p. 343–366, 2020.

GUERA, O. G. M.; SILVA, J. A. A. DA; FERREIRA, R. L. C.; LAZO, D. A. Á.; MEDEL, H. B. Alternatives to Growth and Yield Prognosis for Pinus caribaea var. caribaea Barrett & Golfari. **Floresta e Ambiente**, v. 26, n. 4, 2019.

IGA – INSTITUTO DE GEOCIÊNCIAS APLICADAS. Vegetação do estado de Minas Gerais 2012. Disponível em: <<https://www.mg.gov.br/pagina/geografia>>. Acesso em 17 de mar. 2024.

KUHN, M. Building predictive models in R using the caret package. **Journal of Statistical Software**, v. 28, n. 5, p. 1–26, 2008.

LEITE, H. G.; NOGUEIRA, G. S.; CAMPOS, J. C. C.; SOUZA, A. L. DE; CARVALHO, A. Avaliação de um modelo de distribuição diamétrica ajustado para povoamento de *Eucalyptus* sp. submetidos a desbaste. **Revista Árvore**, 29, n. 2, p. 271–280, 2005.

LIMA, E. D. S.; DE SOUZA, Z. M.; OLIVEIRA, S. R. D. M.; MONTANARI, R.; FARHATE, C. V. V. Random Forest Model To Predict the Height of *Eucalyptu*. **Engenharia Agrícola**, 42, n. SpecialIssue, p. 12–20, 2022.

LISBOA, P. J. G.; SARALAJEW, S.; VELLIDO, A.; FERNÁNDEZ-DOMENECH, R.; VILLMANN, T. The coming of age of interpretable and explainable machine learning models. **Neurocomputing**, 535, p. 25–39, 2023.

LIU, J.; ASHTON, P. S. Individual-based simulation models for forest succession and management. **Forest Ecology and Management**, 73, n. 1–3, p. 157–175, 1995.

MOLNAR, C. iml: An R package for Interpretable Machine Learning. **Journal of Open Source Software**, 3, n. 26, p. 786, 2018.

MOLNAR, C. **Interpretable Machine Learning**. Leanpub, 2020.

NAYAK, H. S.; SILVA, J. V.; PARIHAR, C. M. et al. Interpretable machine learning methods to explain on-farm yield variability of high productivity wheat in Northwest India. **Field Crops Research**, 287, n. February, 2022.

OLIVEIRA, M. L. R. DE; LEITE, H. G.; NOGUEIRA, G. S.; CAMPOS, J. C. C. Modelagem e prognose em povoamentos não desbastados de clones de eucalipto. **Revista Árvore**, 33, n. 5, p. 841–852, 2009.

OLIVEIRA NETO, R. R.; LEITE, H. G.; GLERIANI, J. M.; STRIMBU, B. M. Estimation of *Eucalyptus* productivity using efficient artificial neural network. **European Journal of Forest Research**, 141, n. 1, p. 129–151, 2022.

QUEIROZ, T. B.; CAMPOE, O. C.; MONTES, C. R.; et al. Temperature thresholds for *Eucalyptus* genotypes growth across tropical and subtropical ranges in South America. **Forest Ecology and Management**, 472, p. 118248, 2020.

ROCHA, S. J. S. S. DA; TORRES, C. M. M. E.; VILLANOVA, P. H.; et al. Machine learning methods: Modeling net growth in the Atlantic Forest of Brazil. **Ecological Informatics**, 81, n. July 2023, 2024.

RODY, Y. P.; RIBEIRO, A.; OLIVEIRA, A. S. DE; LEITE, F. P. Modeling and simulation of relief influence on *eucalyptus* forests: interaction between solar irradiance and productivity. **Revista Árvore**, 40, n. 2, p. 319–327, 2016.

RUDIN, C.; CHEN, C.; CHEN, Z.; et al. Interpretable machine learning: Fundamental principles and 10 grand challenges. **Statistics Surveys**, 16, p. 1–85, 2022.

RYO, M. Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. **Artificial Intelligence in Agriculture**, 6, p. 257–265, 2022.

SANTOS, A. C. A.; SILVA, S.; LEITE, H. G.; CRUZ, J. P. DA. Influência da variabilidade edafoclimática no crescimento de clones de eucalipto no Nordeste baiano. **Pesquisa Florestal Brasileira**, 37, n. 91, p. 259, 2017.

SANTOS, K. F. DOS; REICHERT, J. M. Best tillage practices for eucalyptus growth and productivity: A review on the Brazilian experience. **Revista Brasileira de Ciência do Solo**, 46, 2022.

SHARMA, R. P.; VACEK, Z.; VACEK, S.; KUČERA, M. Modelling individual tree height–diameter relationships for multi-layered and multi-species forests in central Europe. **Trees**, 33, n. 1, p. 103–119, 2019.

SILVA, M. L. M.; BINOTI, D. H. B.; GLERIANI, J. M.; LEITE, H. G. Adjustment of the Schumacher and Hall model and application of artificial neural networks to estimate volume of eucalypt trees. **Revista Arvore**, 33, n. 6, p. 1133–1139, 2009.

SPARKS, A. nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. **Journal of Open Source Software**, 3, n. 30, p. 1035, 2018.

SONDERMANN, M.; CHOU, S. C.; LYRA, A. et al. Climate change projections and impacts on the eucalyptus plantation around the Doce River basin, in Minas Gerais, Brazil. **Climate Services**, 28, p. 100327, 2022.

WONGCHAI, W.; ONSREE, T.; SUKKAM, N.; PROMWUNGKWA, A.; TIPPAYAWONG, N. Machine learning models for estimating above ground biomass of fast growing trees. **Expert Systems with Applications**, 199, p. 117186, 2022.

YANG, G.; DENG, Y.; LAN, P.; et al. Estimation of evapotranspiration in Eucalyptus plantation and mixed forests based on air temperature and humidity. **Forest Ecology and Management**, 504, p. 119862, 2022.

## **CAPÍTULO 2: Interpretabilidade de variáveis preditoras da produtividade de povoamentos de eucalipto utilizando perturbação de redes neurais artificiais**

### RESUMO:

O presente estudo identifica, compara e avalia as mudanças nas importâncias relativas de variáveis preditoras da produtividade de povoamentos de eucalipto ao final da rotação, utilizando uma base de dados ampliada em relação ao trabalho de Binoti et al. (2022). Os objetivos foram de verificar o efeito da quantidade de variáveis preditoras na predição da produtividade média de povoamentos de eucalipto e de avaliar a eficiência da ampliação de base de dados por simulação, na validação de modelos de redes neurais artificiais. Uma base de dados composta inicialmente por 320 talhões de eucalipto utilizadas por Binoti et al. (2022) foi aumentada com a inclusão de novas variáveis ambientais e silviculturais, totalizando 304 variáveis, incluindo atrasos ou adiantamentos nas operações silviculturais. Redes neurais artificiais foram treinadas para prever a produtividade média aos 7 anos ( $IMA_7$ ) com duas composições distintas de variáveis preditoras (25 e 304 variáveis). Perturbações nas entradas das redes quantificaram a importância relativa de cada variável. O estudo identificou diferenças nas variáveis de maior importância relativa em relação às RNA treinadas por Binoti et al. (2022) e entre as duas abordagens testadas. A precipitação acumulada até um ano de idade foi o preditor mais relevante com 24% de importância relativa na abordagem com 25 variáveis. Na base de dados ampliada, o genótipo resultou em 5% de importância relativa. Houve conformidade de 14 variáveis entre as 25 mais importantes nas duas abordagens, destacando os preditores genótipo, espaçamento, altitude e teor de argila. A inclusão de variáveis adicionais afetou a exatidão do  $IMA_7$  predito e a importância relativa das variáveis.

Palavras-chave: Aprendizado de máquina. IMA. Manejo Florestal. Silvicultura.

## **CHAPTER 2: Interpretability of predictor variables for eucalypt stands productivity using artificial neural network perturbation**

### **ABSTRACT:**

The present study identifies, compares, and evaluates changes in the relative importance of predictor variables for the productivity of eucalypt stands at the end of the rotation, using an expanded database in relation to the work of Binoti et al. (2022). The objectives were to verify the effect of the number of predictor variables on the prediction of the average productivity of eucalypt stands and to evaluate the efficiency of expanding the database through simulation in validating artificial neural network models. A database initially composed of 320 eucalypt stands used by Binoti et al. (2022) was increased with the inclusion of new environmental and silvicultural variables, totaling 304 variables, including delays or advances in silvicultural operations. Artificial neural networks were trained to predict the average productivity at 7 years ( $IMA_7$ ) with two different compositions of predictor variables (25 and 304 variables). Perturbations in the network inputs quantified the relative importance of each variable. The study identified differences in the variables of greatest relative importance compared to the ANN trained by Binoti et al. (2022) and between the two tested approaches. The accumulated precipitation up to one year of age was the most relevant predictor with 24% relative importance in the approach with 25 variables. In the expanded database, genotype resulted in 5% relative importance. There was conformity of 14 variables among the 25 most important in both approaches, highlighting the predictors genotype, spacing, altitude, and clay content. The inclusion of additional variables affected the accuracy of the predicted  $IMA_7$  and the relative importance of the variables.

Key-words: Machine Learning, IMA, Forest Management, Silviculture.

## 1. INTRODUÇÃO

O crescimento e a produção de povoamentos equiâneos é dependente de processos complexos relacionados a diversos fatores e elementos, de ordem biológica, silvicultural, ambiental, edáfica e de manejo (Billings, 1968; Campos, 1970). Parte da complexidade é representada pelas variáveis utilizadas na construção de determinados modelos de crescimento e produção (Binoti et al., 2022). Esses modelos são essenciais para predição ou projeção dos estoques de crescimento e de colheita de madeira nas diferentes idades dos povoamentos, sendo um componente de um dos elementos essenciais do manejo florestal (Campos e Leite, 2017).

Diferentes abordagens têm sido utilizadas para estimar a produção futura de povoamentos de eucalipto, incluindo modelos de regressão (Dias et al., 2005; Reis et al., 2022), modelos ecofisiológicos (Silva et al., 2023) e algoritmos de inteligência computacional, como redes neurais artificiais e técnicas de aprendizado de máquinas (Binoti et al., 2022; Casas et al., 2022).

As ferramentas de aprendizado de máquina têm sido utilizadas para prognose em povoamentos equiâneos, dadas as características dessas técnicas tais como facilidade operacional de treinamento de algoritmos, capacidade de modelagem de relações não lineares, exatidão e precisão das predições geradas, possibilidade de simulação de condições adversas, como mudanças climáticas, e outras vantagens (Alcântara et al., 2018; Casas et al., 2022; Oliveira Neto et al., 2022; Souza et al., 2019).

Uma das técnicas de inteligência computacional mais utilizada para prever o crescimento e a produção em plantios são as redes neurais artificiais – RNA (Casas et al., 2023; Freitas et al., 2020; Oliveira Neto, 2022; Silva et al., 2015). As RNA podem ser definidas como sistemas de processamento distribuídos de forma paralela, formados por unidades básicas denominadas de neurônios matemáticos, inspirados nas redes neurais do sistema nervoso de seres vivos (Braga et al., 2007).

Além do bom desempenho preditivo, as RNA permitem o estudo das variáveis preditoras e de seus impactos nas predições geradas pelas redes, por meio de diferentes ferramentas de interpretabilidade (Molnar, 2024). Essa possibilidade é de grande interesse para a pesquisa e para as empresas florestais que adotam RNA na predição da produção, visto que possibilitam avaliar os impactos de decisões

silviculturais e de manejo, bem como, efeitos de variáveis edáficas e de elementos climáticos (Binoti et al., 2022).

Algumas ferramentas já foram testadas na mensuração florestal para identificar os efeitos de variáveis preditoras da produtividade de povoamentos equiâneos com RNA, como o algoritmo de Garson, análise de sensibilidade global (Freitas et al., 2020) e importância relativa da variável por perturbação (Binoti et al., 2022) e a inversão de redes neurais (Oliveira Neto, 2022). Conforme esses últimos autores, a inversão de uma rede neural é feita quando se deseja definir combinações de níveis de diferentes variáveis de entrada para obtenção de uma saída fixa. Por exemplo, diferentes combinações de variáveis de manejo que resultam em uma mesma produtividade ao final da rotação.

Oliveira Neto (2022) utilizou a meta-heurística *Simulated Annealing* (SA) para inverter uma RNA treinada para predição da produtividade média de povoamentos de eucalipto aos 6 e 7 anos de idade. A SA foi eficiente para inversão da rede e consequente definição da importância de diferentes variáveis de entrada, na produtividade média de povoamentos de eucalipto, distribuídos por todo o estado de Minas Gerais, ao final da rotação florestal. O estudo permitiu a definição de diferentes combinações de níveis de variáveis de entrada para se obter produtividades fixadas para 6 e 7 anos. Duas variáveis importantes na predição da produtividade foram o espaçamento de plantio e o genótipo (clone), isto para as condições climáticas e ambientais dos locais onde se encontravam os povoamentos.

A obtenção da importância relativa das variáveis por meio da perturbação consiste na introdução de dados de entrada deliberadamente alterados, chamados de ruído branco, para uma determinada variável (Olden et al., 2004). A mudança ocasionada pelo ruído branco nos erros de predição da RNA reflete a importância da variável. Binoti et al. (2022) utilizou este método para gerar as importâncias relativas de 25 variáveis preditoras da produtividade de povoamentos de eucalipto ao final da idade de rotação, na região Norte de Minas Gerais.

Para este capítulo do presente estudo ampliamos a base de dados utilizada por Binoti et al. (2022) com a inclusão de mais uma medição das parcelas permanentes utilizadas pelos autores (medição de 2022). Além disso, simulamos valores para algumas das variáveis preditoras e acrescentamos novas variáveis de entrada. O escopo desse capítulo se baseia na comparação e avaliação das mudanças ocasionadas nas importâncias relativas das variáveis mediante a ampliação da base

de dados entre os dois estudos. Os objetivos foram de verificar o efeito da quantidade de variáveis preditoras na predição da produtividade média de povoamentos de eucalipto e de avaliar a eficiência da ampliação de base de dados por simulação, na validação de modelos de redes neurais artificiais.

## 2. MATERIAIS E MÉTODOS

Os dados utilizados neste estudo foram provenientes de inventários florestais contínuos conduzidos em povoamentos de clones de híbridos de *Eucalyptus urophylla* X *Eucalyptus grandis* localizados em onze municípios do norte do estado de Minas Gerais, Brasil. A vegetação da região é caracterizada com áreas de Cerrado, Campo Cerrado e Campo de Altitude (IGA, 2012). A classificação climática de Köppen é de clima tropical úmido (Aw), com presença de um período de inverno seco e verão com chuvas (Alvares et al., 2014). A temperatura média anual na região varia entre 22 e 24 °C e os índices pluviométricos entre 500 e 1.200 mm anuais (Alvares et al., 2014; Carneiro, 2003). A altitude nos povoamentos amostrados variou entre 750 m e 1020 m acima do nível do mar, com altitude média de 902,8 m

A base de dados utilizada consiste em 320 talhões de reflorestamentos de clones de híbridos de clones de híbridos de *Eucalyptus urophylla* X *Eucalyptus grandis* distribuídos em onze municípios localizados no norte do estado de Minas Gerais, Brasil. Em média, cada talhão dispõe de três parcelas, com pelo menos duas medições. Em comparação a base de dados utilizada por Binoti et al. (2022) houve acréscimo de novas medições, realizadas entre 2021 e 2022 e, também, de novas variáveis, tanto mensuradas quanto simuladas. As variáveis consideradas em Binoti et al. (2022) estão relacionadas na Tabela 1.

Tabela 1: Variáveis consideradas para predição da produtividade florestal aos 7 anos ( $IMA_7$ ) em Binoti et al. (2022)

Variável	Descrição
Capina Mecânica em Linha Total com Pré-Emergente	Dias até data de referência
Primeira Adubação de Manutenção (Ref. 6 meses)	Dias até data de referência
Primeira Capina Manual em Linha Total	Dias até data de referência
Segunda Adubação de Manutenção (Ref. 12 meses)	Dias até data de referência
Segunda Capina Manual em Linha Total	Dias até data de referência
Subsolagem com Fosfatagem	Dias até data de referência
Terceira Adubação de Manutenção (Monitoramento)	Dias até data de referência
Terceira Capina Manual em Linha Total	Dias até data de referência
Genótipo	Identificação do clone
Teor de argila	Porcentagem
Teor de matéria orgânica	Porcentagem
Altitude	Metros acima do nível do mar
Área útil por planta	m <sup>2</sup> por planta
Quantidade de adubação de Fósforo	kg/ha
Quantidade de adubação de Nitrogênio	kg/ha
Quantidade de adubação de Potássio	kg/ha
Precipitação até um ano de idade (Idade $\leq 1$ ano)	mm
Precipitação com um ano de idade	mm
Precipitação com dois anos de idade	mm
Precipitação com três anos de idade	mm
Precipitação com quatro anos de idade	mm
Precipitação com cinco anos de idade	mm
Precipitação com seis anos de idade	mm
Precipitação com sete anos de idade	mm
Porcentagem de chuva no período seco	%

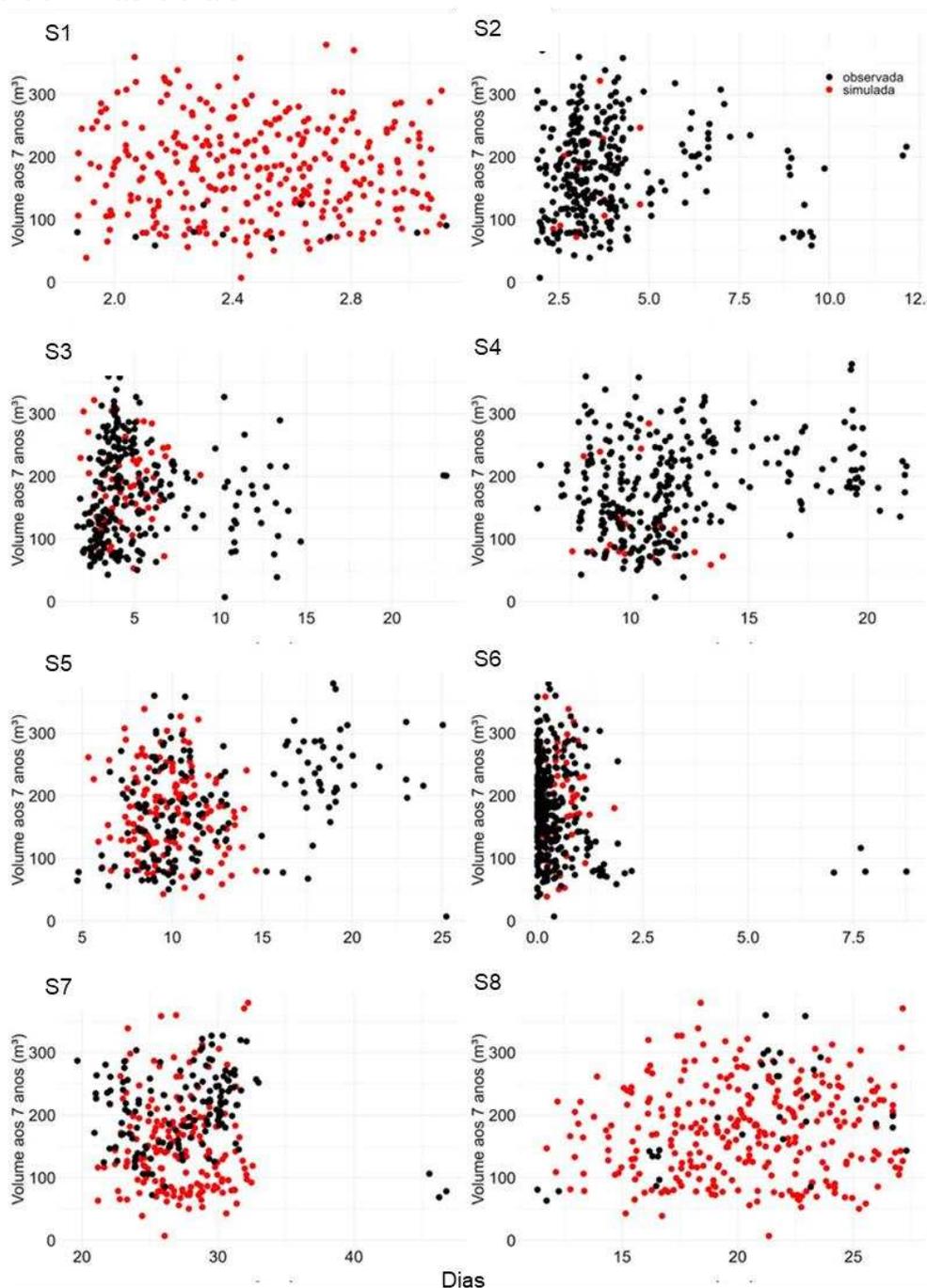
A variável predita foi a produtividade média aos 7 anos ( $IMA_7$ ), que foi a rotação regulatória definida para esta pesquisa. O conjunto de variáveis preditoras incluiu variáveis silviculturais, ambientais e de manejo. As variáveis silviculturais e seus detalhamentos estão na Tabela 2. A realização da operação silvicultural na referência correta, implica no valor zero para a variável silvicultural. Foram simulados os valores das variáveis silviculturais nos talhões onde houve realização no período estipulado. As simulações foram realizadas levando em consideração a distribuição observada da variável e utilizando valores mínimos e máximos, com intuito de evitar simulações de *outliers*. A Figura 1 mostra a dispersão observada e simulada das variáveis silviculturais operacionais.

Tabela 2: Padrões operacionais para as principais práticas silviculturais aplicadas aos talhões em povoamentos de eucalipto no Norte de Minas Gerais

Atividade	Sigla	Nº de dias para realização da atividade, em relação ao momento do plantio
Primeira adubação de manutenção	S1	90
Primeira capina manual em linha total	S2	90
Segunda adubação de manutenção	S3	366
Segunda capina manual em linha total	S4	180
Subsolagem com fosfatagem	S5	-30
Terceira capina manual em linha total	S6	360
Terceira adubação de manutenção	S7	720
Capina mecânica em linha com pré-emergente	S8	60

Fonte: Binoti et al. (2022).

Figura 1: Dispersões observadas e simuladas das variáveis silviculturais operacionais predictoras da produtividade florestal aos 7 anos ( $IMA_7$ ) de povoamentos de eucalipto no Norte de Minas Gerais



S1 = Capina mecânica em linha total com pré-emergente; S2 = Primeira adubação de manutenção; S3 = Primeira capina manual em linha total; S4 = Segunda adubação de manutenção; S5 = Segunda capina manual em linha total; S6 = Subsolagem com fosfatagem; S7 = Terceira adubação de manutenção; S8 = Terceira capina manual em linha total.

Fonte: Autor (2024).

Além das variáveis silviculturais operacionais, também foram obtidas as precipitações desde o ano de plantio dos povoamentos até a idade de corte (7 anos), obtidas no sistema NASA POWER por meio de API em linguagem R (Sparks, 2024) e variáveis relacionadas as propriedades físicas do solo e adubação (Tabela 2). Além dessas variáveis, foram incorporadas novas variáveis ambientais não contidas em Binoti et al. (2022), também obtidas no NASA POWER, bem como, variáveis ambientais complementares. A lista completa das variáveis e estatísticas descritivas estão no APÊNDICE A.

Dessa forma, foram construídos dois bancos de dados, o primeiro com 25 variáveis preditoras utilizadas por Binoti et al. (2022) acrescidas de simulações. No segundo banco de dados foram acrescidas 279 novas variáveis ambientais, compatibilizadas com o ano de plantio até a idade de rotação de 7 anos, perfazendo um conjunto de 304 variáveis preditoras. Cada um desses bancos de dados foi utilizado para treino de uma rede neural artificial com as seguintes especificações: arquitetura *multilayer perceptron* com uma camada oculta de 12 neurônios, algoritmo de treinamento *resilient propagation*, RPROD+ e função de ativação do tipo sigmoide nas camadas de entrada e de saída. Foram considerados 300 ciclos, com validação cruzada *k-folder* com quatro grupos e 30 execuções. As redes foram treinadas no *software* R com o pacote *nnet* (Ripley e Venables, 2023). Os resultados foram expressos por meio de gráficos de erros relativos percentuais em função da produtividade média aos 7 anos. Foram calculados os coeficientes de correlação entre as produtividades médias aos 7 anos preditas e observadas ( $r_{\hat{y}y}$ ) e raiz quadrada do erro quadrático médio relativo (RMSE%), conforme indicado por Campos e Leite (2017).

Cada variável foi perturbada 100 vezes, sendo a importância relativa obtida pela média das perturbações. A métrica utilizada na obtenção da importância relativa foi a raiz quadrada do erro quadrático médio (RMSE). Por conta da numerosa quantidade de variáveis preditoras no conjunto de dados definido após os acréscimos do presente estudo, optou-se por selecionar somente as 25 variáveis com maiores importâncias relativas para interpretação e discussão dos resultados. No entanto, os valores de importância relativa de todas as variáveis podem ser consultados no APÊNDICE B. As duas abordagens foram comparadas com as variáveis com maiores importâncias relativas obtidas por Binoti et al. (2022).

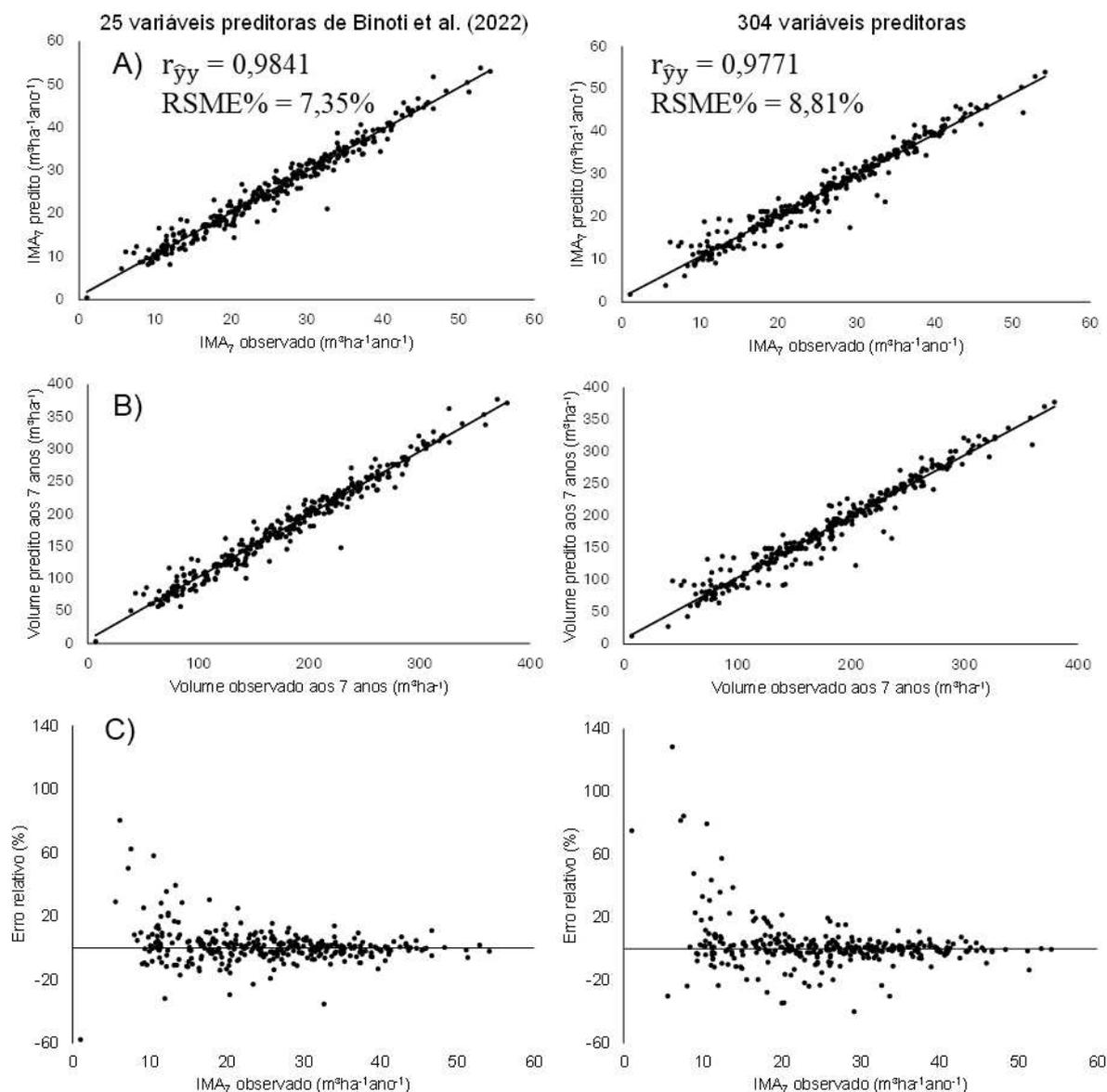
O efeito da introdução das variáveis com maiores importâncias relativas, segundo o resultado das perturbações, foi expresso por meio do incremento no coeficiente de correlação linear entre os valores preditos e observados para o  $IMA_7$ . Para construção dos gráficos, foram treinadas cinco redes com número de neurônios na camada de entrada variando de 1 (variável com maior valor de importância relativa) até 25 (todas as variáveis), utilizando as mesmas especificações da RNA perturbada. A ordem de inclusão das variáveis no processo de treinamento seguiu a ordem decrescente de importância relativa. A média do coeficiente de correlação linear entre os valores preditos e observados para o  $IMA_7$  das cinco redes treinadas foi utilizado para construir os gráficos.

### 3. RESULTADOS E DISCUSSÃO

A produtividade aos 7 anos foi obtida com as duas abordagens de treinamento propostas, usando conjuntos de dados com 25 e 304 variáveis preditoras. Os resultados das predições estão expressos em produtividade média (Figura 2A) e volume total na idade de 7 anos (Figura 2B). A RNA com as mesmas variáveis utilizadas por Binoti et al. (2022) e variáveis silviculturais simuladas, apresentou valor de  $r_{\hat{y}y}$  superior ao obtido anteriormente ( $r_{\hat{y}y} = 0,9466$ ) e o valor de RMSE% de 5,6%, considerando o  $IMA_7$  como variável resposta. A introdução de novas variáveis resultou em uma RNA com  $r_{\hat{y}y}$  inferior e maior RMSE%, com tendência a superestimação da produtividade de povoamentos menores valores de volume observado (Figura 12C).

A utilização de um número maior de variáveis preditoras pode incorrer em redes com maior complexidade computacional de treinamento, exigindo adequações na estrutura da RNA durante a etapa de treino, visando captar de forma mais adequada a complexidade do fenômeno modelado (Braga et al., 2007).

Figura 2: Análise gráfica da predição da produção volumétrica na idade de corte (7 anos) para povoamentos de eucalipto no Norte de Minas Gerais. (A) relação entre produtividade média aos 7 anos observadas e preditas pelas redes neurais artificiais; (B) relação entre volume total aos 7 anos observados e preditos pelas redes neurais artificiais e (C) erros relativos percentuais para as predições de produtividade média aos 7 anos.



Fonte: Autor (2024).

As importâncias relativas obtidas pelas perturbações das redes para as variáveis predictoras das duas abordagens utilizadas estão na Figura 3. A ampliação da base de dados e o aumento do número de variáveis predictoras ocasionaram diferenças relevantes nos valores de importância relativa para as 25 variáveis no processo de predição da produtividade aos sete anos, tanto no que se refere às duas

abordagens testadas neste estudo, quanto a abordagem adotada por Binoti et al. (2022).

A variável com maior valor de importância relativa nas RNA treinadas por Binoti et al. (2022) foi a precipitação aos cinco anos de idade, seguida pela precipitação aos quatro anos. Para a base de dados ampliada, a variável precipitação também apresentou maior importância relativa (24%), no entanto, para a idade inferior a um ano (Figura 3). Considerando todas as 304 variáveis do conjunto de predição, a maior importância relativa foi obtida pelo genótipo (5%). Tais diferenças entre a ordenação das variáveis mais importantes podem ser atribuídas, no caso das abordagens com o mesmo número de variáveis preditoras, aos efeitos da ampliação da base de dados por meio da simulação das variáveis silviculturais e da inclusão de novas medições, que acrescem maior intervalo de anos, impactando nas variáveis ambientais coletadas.

O genótipo é uma variável categórica que demonstra bastante importância para a modelagem do crescimento e da produção de povoamentos florestais, tanto com técnicas de regressão, onde serve como fator de estratificação de equações, quanto como variável de entrada em algoritmos de aprendizado de máquina (Freitas et al. 2020; Tavares Júnior et al., 2019). A introdução do genótipo como variável categórica em redes neurais artificiais para predição da produtividade média resultam em ganhos significativos de exatidão (Freitas et al. 2020). Segundo Silva et al. (2015) o genótipo, em muitos casos, é capaz de representar outras variáveis ambientais locais que influenciam na produção de volume nos povoamentos, como elementos edáficos e de manejo, ocasionando em maior importância relativa.

No caso de bancos de dados com grandes números de recursos preditores, como a abordagem com 304 variáveis, geralmente há presença de colinearidade e forte redundância entre as variáveis e observações (Boehmke e Greenwell, 2019). Essas características podem ter contribuído para as diferenças entre os ordenamentos de variáveis nas abordagens testadas. Uma maneira eficiente para reduzir esta quantidade de variáveis é o emprego de análise de componentes principais (PCA), conforme feito por Oliveira Neto et al. (2022). No presente estudo tal método multivariado não foi utilizado uma vez que pelo objetivo definido era necessário ter a rede treinada com todas as novas variáveis.

Quatro variáveis silviculturais relacionadas a operações realizadas nos talhões estão presentes entre as 25 com maiores importâncias relativas na abordagem de 304

variáveis. A simulação de número de dias de atraso ou adiantamento na realização dessas atividades pode estar associado ao aumento da importância desses recursos. Essas variáveis podem ser úteis na análise econômica de cenários silviculturais e avaliação de impactos na produtividade média dos povoamentos, auxiliando o planejamento estratégico dos empreendimentos florestais (Binoti et al., 2022; Telles et al., 2021).

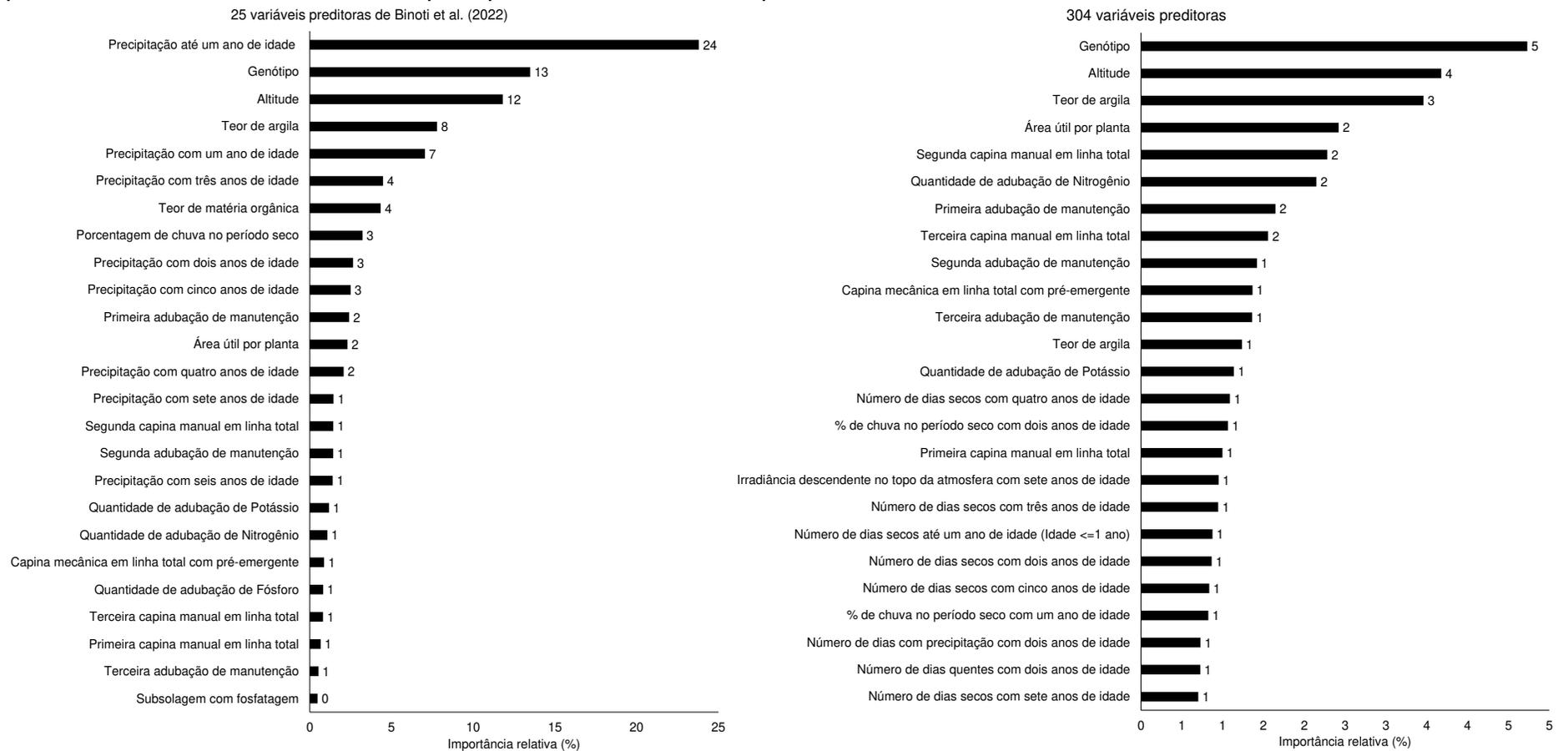
Nas três abordagens avaliadas neste estudo, houve forte presença de variáveis relacionadas a dinâmica da precipitação (precipitação em mm ao longo dos anos, porcentagem de chuva no período seco, dias com presença de precipitação > 0 mm) entre os dez preditores com maiores importâncias relativas.

A importância dessas variáveis está em consonância com o realismo biológico do processo de crescimento de plantas, visto que o regime de precipitação é essencial para disponibilidade de água e para os processos fisiológicos de desenvolvimento das árvores de eucalipto (Binkley et al., 2017; Scolforo et al., 2019). Além disso, a inclusão de variáveis relacionadas a elementos climáticos como a precipitação possibilita realizar simulações de regimes de chuvas, fornecendo maior robustez para a modelagem do crescimento e da produção em cenários de mudanças climáticas (Florêncio et al., 2022).

Ao comparar as duas abordagens com 25 variáveis preditoras e a abordagem ampliada, nota-se que 14 variáveis são comuns entre os preditores de maior importância relativa, como altitude, espaçamento (área útil por planta), teores de matéria orgânica e de argila no solo e outras. A utilização dessas variáveis em modelos de aprendizado de máquina revela não somente ganhos preditivos, mas também tornam o processo de predição mais realista (Alcântara et al., 2020; Alvares et al., 2023; Binoti et al., 2015; Freitas et al., 2020). A maior parte dessas variáveis representam prescrições de manejo ou características edáficas de fácil obtenção a partir de análises de solos. Dessa maneira, devem ser priorizadas na composição de bancos de dados para modelagem da produtividade de povoamentos de eucalipto.

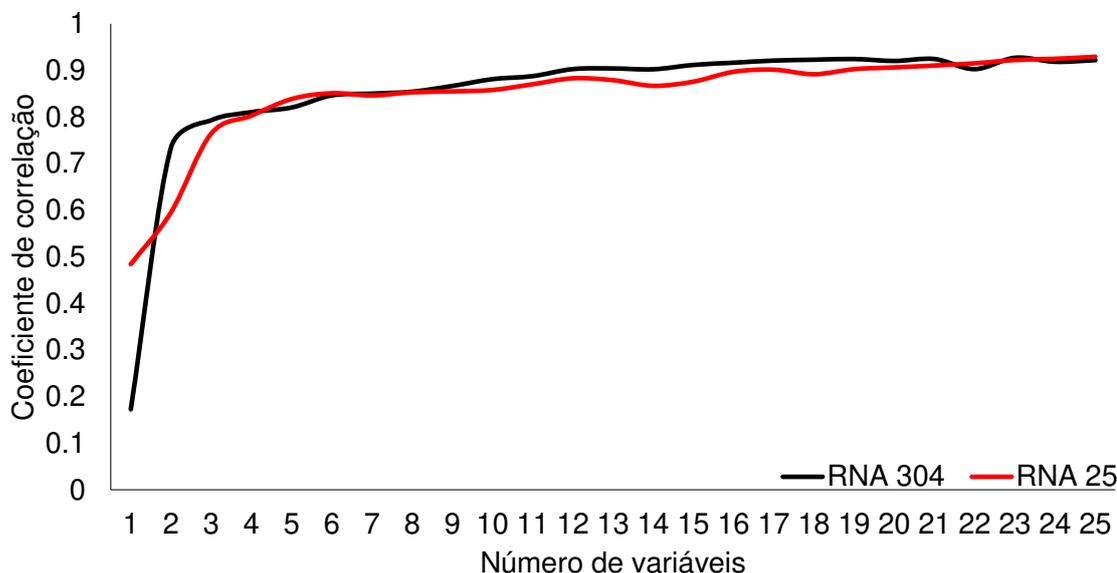
O incremento no coeficiente de correlação linear entre os valores preditos e observados para a produtividade média aos 7 anos nas duas abordagens utilizadas pode ser visualizado na Figura 4. Assim como em Binoti et al. (2022), foram observadas tendências de estabilização dos valores de  $r_{\hat{y}y}$  a partir da inclusão da sexta variável na entrada das RNA. No entanto, essa tendência foi menos evidente na abordagem com 304 variáveis preditoras.

Figura 3: Importância relativa das variáveis preditoras obtidas por perturbação de RNA para as duas abordagens de predição da produtividade média aos sete anos para povoamentos de eucalipto no norte de Minas Gerais



Fonte: Autor (2024).

Figura 4: Variação do coeficiente de correlação linear entre valores preditos e observados para a produtividade média aos 7 anos de eucalipto no norte de Minas Gerais pelas duas abordagens de treinamento das RNA em função da adição das variáveis predictoras de maior importância relativa



Fonte: Autor (2024)

#### 4. CONCLUSÕES

A ampliação dos conjuntos de dados de variáveis predictoras em relação aos dados utilizados por Binoti et al. (2022), por meio da inclusão de variáveis e de novas medições, bem como, simulação de valores, resulta em diferenças na importância relativa das variáveis predictoras da produtividade de povoamentos de eucalipto aos 7 anos.

A ampliação de base de dados silviculturais por meio de simulação permite uma validação eficiente de modelos de redes neurais artificiais para predição da produtividade de povoamentos de eucalipto.

Variáveis relacionadas à dinâmica de precipitação (precipitação em mm ao longo dos anos, porcentagem de chuva no período seco e dias com presença de precipitação) devem ser incluídas como predictoras da produtividade de povoamentos de eucalipto, visto suas importâncias relativas elevadas em todas as abordagens.

#### REFERÊNCIAS

ALCÂNTRA, A. E. M.; SANTOS, A. C. A.; SILVA, M. L. M. et al. Use of artificial neural networks to assess yield projection and average production of eucalyptus stands. **African Journal of Agricultural Research**, 13, n. 42, p. 2285–2297, 2018.

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; et al. Köppen's climate classification map for Brazil, **Meteorologische Zeitschrift**, 22, n. 6, p. 711–728, 2013.

ALVARES, C. A.; CEGATTA, Í. R.; SCOLFORO, H. F.; MAFIA, R. G. Decision-Tree Application to Predict and Spatialize the Wood Productivity Probabilities of Eucalyptus Plantations. **Forests**, 14, n. 7, p. 1334, 2023.

BILLINGS, W. D. **Las plantas y el ecosistema**. H. Herman Sucesores, 1968.

BINKLEY, D.; CAMPOE, O. C.; ALVARES, C.; et al. The interactions of climate, spacing and genetics on clonal Eucalyptus plantations across Brazil and Uruguay. **Forest Ecology and Management**, 405, n. July, p. 271–283, 2017.

BINOTI, D. H. B.; LEITE, H. G.; ANDRADE, V. C. L. et al. Importância relativa das variáveis preditoras no processo de modelagem da produtividade florestal. **Advances in Forestry Science**, 9, n. 4, p. 1867–1873, 2022.

BOEHMKE, B.; GREENWELL, B. **Hands-On Machine Learning with R**. Chapman and Hall/CRC, 2020.

BRAGA, A. P.; LUDERMIR, T. B.; CARVALHO, A. C. P. L. F. **Redes neurais artificiais: teoria e aplicações**. LTC, 2007.

CAMPOS, J. C. C. Principais fatores do meio que afetam o crescimento das árvores. **FLORESTA**, 2, n. 3, 1970.

CAMPOS, J. C. C.; LEITE, H. G. **Mensuração florestal: perguntas e respostas**. 5. ed. Viçosa, MG: Editora UFV, 2017. 636 p.

CASAS, G. G.; FARDIN, L. P.; SILVA, S. et al. Improving Yield Projections from Early Ages in Eucalypt Plantations with the Clutter Model and Artificial Neural Networks. **Pertanika Journal of Science and Technology**, 30, n. 2, p. 1257–1272, 2022.

CASAS, G. G.; ISMAIL, Z. H.; LIMEIRA, M. M. C. et al. Automatic Detection and Counting of Stacked Eucalypt Timber Using the YOLOv8 Model. **Forests**, 14, n. 12, 2023.

CARNEIRO, M. F. B. Região Norte de Minas: caracterização geográfica e a organização espacial – breves considerações. **Revista Cerrados**, 1, n. 1, p. 91–106, 2003.

DIAS, A. N.; LEITE, H. G.; SILVA, M. L. et al. Avaliação financeira de plantações de eucalipto submetidas a desbaste. **Revista Árvore**, 29, n. 3, p. 419–429, 2005.

FLORÊNCIO, G. W. L.; MARTINS, F. B.; FAGUNDES, F. F. A. Climate change on Eucalyptus plantations and adaptive measures for sustainable forestry development across Brazil. **Industrial Crops and Products**, 188, n. August, 2022.

FREITAS, E. C. S.; PAIVA, H. N.; NEVES, J. C. L. et al. Modeling of eucalyptus productivity with artificial neural networks. **Industrial Crops and Products**, 146, n. February, p. 112149, 2020.

IGA – INSTITUTO DE GEOCIÊNCIAS APLICADAS. Vegetação do estado de Minas Gerais 2012. Disponível em: <<https://www.mg.gov.br/pagina/geografia>>. Acesso em 17 de mar. 2024.

MOLNAR, C. **Interpretable Machine Learning: A Guide for Making Black Box Models Explainable**. Leanpub, 2024. Disponível em: <<https://christophm.github.io/interpretable-ml-book/>>

OLDEN, J. D.; JOY, M. K.; DEATH, R. G. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. **Ecological Modelling**, 178, n. 3, p. 389–397, 2004.

OLIVEIRA NETO, R. R.; LEITE, H. G.; GLERIANI, J. M. et al. Estimation of Eucalyptus productivity using efficient artificial neural network. **European Journal of Forest Research**, 141, n. 1, p. 129–151, 2022.

REIS, T.; SILVA, S.; JUNIOR, I. T. et al. Assessing alternatives for growth and yield modelling in eucalypt stands. **Southern Forests**, 84, n. 1, p. 34–43, 2022.

RIPLEY, B.; VENABLES, W. **Package ‘nnet’**. Package Version 7.3-19, 2023.

SCOLFORO, H. F.; MCTAGUE, J. P.; BURKHART, H. et al. Modeling whole-stand survival in clonal eucalypt stands in Brazil as a function of water availability. **Forest Ecology and Management**, v. 432, n. October 2018, p. 1002–1012, 2019.

SILVA, M. L. M. B.; LEITE, H. G.; BINOTI, D. H. B. et al. Prognose em nível de povoamento de clones de eucalipto empregando redes neurais artificiais. **Cerne**, 21, n. 1, p. 97–105, 2015.

SILVA, G. C. C.; NEVES, J. C. L.; MARCATTI, G. E. et al. Improving 3-PG calibration and parameterization using artificial neural networks. **Ecological Modelling**, 479, n. August 2022, p. 110301, 2023.

SOUZA, G. S. A.; SOARES, V. P.; LEITE, H. G. et al. Multi-sensor prediction of Eucalyptus stand volume: A support vector approach. **ISPRS Journal of Photogrammetry and Remote Sensing**, 156, n. August, p. 135–146, 2019.

SPARKS, A. nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. **Journal of Open Source Software**, 3, n. 30, p. 1035, 2018.

TAVARES JÚNIOR, I. S.; ROCHA, J. E. C.; EBLING, A. A. et al. Artificial neural networks and linear regression reduce sample intensity to predict the commercial volume of eucalyptus clones. **Forests**, 10, n. 3, 2019.

TELLES, L. A. A.; ARROYO, J. E. C.; BINOTI, D. H. B. et al. When, where and what cultivate: An optimization model for rural property planning. **Journal of Cleaner Production**, v. 290, 2021.

## CONCLUSÕES GERAIS

A produtividade média de povoamentos de eucalipto aos 7 anos de idade pode ser predita com exatidão em função de variáveis silviculturais e ambientais utilizadas como entradas para o treinamento os algoritmos de aprendizado de máquina *random forest*, *gradiente boosting machine* e redes neurais artificiais.

A ampliação de bases de dados com simulações e a adição de novas variáveis causam alterações nas variáveis de maior importância relativa. No entanto, há concordância entre todas as técnicas testadas neste estudo, indicando que as variáveis altitude, teor de matéria orgânica e teor de argila no solo são as de maior relevância dentro do processo de modelagem do *IMA*<sub>7</sub> na área de estudo.

As técnicas de interpretabilidade para modelos de aprendizado de máquina apresentam grande potencial de utilização em problemas de mensuração florestal, principalmente em contextos em que há interesse em associar os ganhos em exatidão com a possibilidade de interpretação e quantificação de recursos.

A incorporação da interpretabilidade como uma etapa adicional ao treinamento e avaliação de algoritmos de aprendizado de máquina resulta em maior eficiência na modelagem da produtividade de eucalipto, auxiliando na construção de modelos que resultem em realismo biológico, inclusive com possibilidade de simulação de efeitos de mudanças climáticas.

## APÊNDICE A

Tabela: Descrição do conjunto ampliado de variáveis preditoras da produtividade florestal aos 7 anos ( $IMA_7$ ) de povoamentos de eucalipto no Norte de Minas Gerais

Sigla	Variável	Unidade	Média	Desvio-padrão
prec_0	Precipitação até um ano de idade (Idade <=1 ano)	mm	673,24	244,86
prec_1	Precipitação com um ano de idade	mm	770,27	252,14
prec_2	Precipitação com dois anos de idade	mm	851,67	262,94
prec_3	Precipitação com três anos de idade	mm	738,65	161,56
prec_4	Precipitação com quatro anos de idade	mm	780,85	173,63
prec_5	Precipitação com cinco anos de idade	mm	809,58	182,21
prec_6	Precipitação com seis anos de idade	mm	976,90	159,76
prec_7	Precipitação com sete anos de idade	mm	958,58	191,68
prec_ac_0	Precipitação acumulada até <= um ano	mm	673,24	244,86
prec_ac_1	Precipitação acumulada até um ano de idade	mm	1443,51	410,01
prec_ac_2	Precipitação acumulada até dois anos de idade	mm	2295,18	534,48
prec_ac_3	Precipitação acumulada até três anos de idade	mm	3033,83	504,40
prec_ac_4	Precipitação acumulada até quatro anos de idade	mm	3814,68	477,99
prec_ac_5	Precipitação acumulada até cinco anos de idade	mm	4624,26	576,28
prec_ac_6	Precipitação acumulada até seis anos de idade	mm	5601,16	563,09
prec_ac_7	Precipitação acumulada até sete anos de idade	mm	6559,74	618,51
ALLSKY_SFC_UVA_0	Irradiância UVA de toda a superfície do céu até um ano de idade (Idade <=1 ano)	W/m <sup>2</sup>	14,26	0,70
ALLSKY_SFC_UVA_1	Irradiância UVA de toda a superfície do céu com um ano de idade	W/m <sup>2</sup>	14,24	0,59
ALLSKY_SFC_UVA_2	Irradiância UVA de toda a superfície do céu com dois anos de idade	W/m <sup>2</sup>	14,16	0,58
ALLSKY_SFC_UVA_3	Irradiância UVA de toda a superfície do céu com três anos de idade	W/m <sup>2</sup>	14,21	0,49
ALLSKY_SFC_UVA_4	Irradiância UVA de toda a superfície do céu com quatro anos de idade	W/m <sup>2</sup>	14,29	0,58
ALLSKY_SFC_UVA_5	Irradiância UVA de toda a superfície do céu com cinco anos de idade	W/m <sup>2</sup>	14,23	0,62
ALLSKY_SFC_UVA_6	Irradiância UVA de toda a superfície do céu com seis anos de idade	W/m <sup>2</sup>	13,87	0,51
ALLSKY_SFC_UVA_7	Irradiância UVA de toda a superfície do céu com sete anos de idade	W/m <sup>2</sup>	14,03	0,58

Sigla	Variável	Unidade	Média	Desvio-padrão
ALLSKY_SFC_UVB_0	Irradiância UVB de toda a superfície do céu até um ano de idade (Idade <=1 ano)	W/m <sup>2</sup>	0,41	0,02
ALLSKY_SFC_UVB_1	Irradiância UVB de toda a superfície do céu com um ano de idade	W/m <sup>2</sup>	0,41	0,02
ALLSKY_SFC_UVB_2	Irradiância UVB de toda a superfície do céu com dois anos de idade	W/m <sup>2</sup>	0,41	0,02
ALLSKY_SFC_UVB_3	Irradiância UVB de toda a superfície do céu com três anos de idade	W/m <sup>2</sup>	0,41	0,01
ALLSKY_SFC_UVB_4	Irradiância UVB de toda a superfície do céu com quatro anos de idade	W/m <sup>2</sup>	0,42	0,02
ALLSKY_SFC_UVB_5	Irradiância UVB de toda a superfície do céu com cinco anos de idade	W/m <sup>2</sup>	0,42	0,02
ALLSKY_SFC_UVB_6	Irradiância UVB de toda a superfície do céu com seis anos de idade	W/m <sup>2</sup>	0,41	0,02
ALLSKY_SFC_UVB_7	Irradiância UVB de toda a superfície do céu com sete anos de idade	W/m <sup>2</sup>	0,41	0,02
ALLSKY_SFCUVINDEX_0	Índice UV de toda a superfície do céu até um ano de idade (Idade <=1 ano)	Adimensional	2,11	0,11
ALLSKY_SFCUVINDEX_1	Índice UV de toda a superfície do céu com um ano de idade	Adimensional	2,16	0,10
ALLSKY_SFCUVINDEX_2	Índice UV de toda a superfície do céu com dois anos de idade	Adimensional	2,15	0,10
ALLSKY_SFCUVINDEX_3	Índice UV de toda a superfície do céu com três anos de idade	Adimensional	2,12	0,07
ALLSKY_SFCUVINDEX_4	Índice UV de toda a superfície do céu com quatro anos de idade	Adimensional	2,17	0,11
ALLSKY_SFCUVINDEX_5	Índice UV de toda a superfície do céu com cinco anos de idade	Adimensional	2,21	0,11
ALLSKY_SFCUVINDEX_6	Índice UV de toda a superfície do céu com seis anos de idade	Adimensional	2,15	0,11
ALLSKY_SFCUVINDEX_7	Índice UV de toda a superfície do céu com sete anos de idade	Adimensional	2,17	0,11
QV2M_0	Umidade específica a 2 m até um ano de idade (Idade <=1 ano)	(g/kg)	11,57	0,25
QV2M_1	Umidade específica a 2 m com um ano de idade	(g/kg)	11,69	0,36
QV2M_2	Índice UV de toda a superfície do céu com dois anos de idade	(g/kg)	11,72	0,35
QV2M_3	Índice UV de toda a superfície do céu com três anos de idade	(g/kg)	11,60	0,33
QV2M_4	Índice UV de toda a superfície do céu com quatro anos de idade	(g/kg)	11,85	0,25
QV2M_5	Índice UV de toda a superfície do céu com cinco anos de idade	(g/kg)	11,97	0,37
QV2M_6	Índice UV de toda a superfície do céu com seis anos de idade	(g/kg)	12,00	0,44
QV2M_7	Índice UV de toda a superfície do céu com sete anos de idade	(g/kg)	11,86	0,25
RH2M_0	Umidade relativa a 2 m até um ano de idade (Idade <=1 ano)	%	62,50	5,52
RH2M_1	Umidade relativa a 2 m com um ano de idade	%	61,36	5,43
RH2M_2	Umidade relativa a 2 m com dois anos de idade	%	62,29	4,34
RH2M_3	Umidade relativa a 2 m com três anos de idade	%	62,00	3,70

Sigla	Variável	Unidade	Média	Desvio-padrão
RH2M_4	Umidade relativa a 2 m com quatro anos de idade	%	61,76	3,54
RH2M_5	Umidade relativa a 2 m com cinco anos de idade	%	63,02	3,84
RH2M_6	Umidade relativa a 2 m com seis anos de idade	%	65,01	3,61
RH2M_7	Umidade relativa a 2 m com sete anos de idade	%	64,20	3,33
WS2M_0	Velocidade do vento em 2 m até um ano de idade (Idade <=1 ano)	ms <sup>-1</sup>	2,20	0,22
WS2M_1	Velocidade do vento em 2 m com um ano de idade	ms <sup>-1</sup>	2,19	0,21
WS2M_2	Velocidade do vento em 2 m com dois anos de idade	ms <sup>-1</sup>	2,24	0,24
WS2M_3	Velocidade do vento em 2 m com três anos de idade	ms <sup>-1</sup>	2,24	0,25
WS2M_4	Velocidade do vento em 2 m com quatro anos de idade	ms <sup>-1</sup>	2,16	0,23
WS2M_5	Velocidade do vento em 2 m com cinco anos de idade	ms <sup>-1</sup>	2,20	0,25
WS2M_6	Velocidade do vento em 2 m com seis anos de idade	ms <sup>-1</sup>	2,15	0,25
WS2M_7	Velocidade do vento em 2 m com sete anos de idade	ms <sup>-1</sup>	2,05	0,22
TQV_0	Coluna total de água precipitável até um ano de idade (Idade <=1 ano)	kg m <sup>-2</sup>	29,43	1,02
TQV_1	Coluna total de água precipitável com um ano de idade	kg m <sup>-2</sup>	30,19	1,31
TQV_2	Coluna total de água precipitável com dois anos de idade	kg m <sup>-2</sup>	30,01	1,16
TQV_3	Coluna total de água precipitável com três anos de idade	kg m <sup>-2</sup>	29,62	1,54
TQV_4	Coluna total de água precipitável com quatro anos de idade	kg m <sup>-2</sup>	30,45	1,07
TQV_5	Coluna total de água precipitável com cinco anos de idade	kg m <sup>-2</sup>	30,29	1,18
TQV_6	Coluna total de água precipitável com sete anos de idade	kg m <sup>-2</sup>	30,40	1,46
TQV_7	Coluna total de água precipitável com seis anos de idade	kg m <sup>-2</sup>	30,04	0,91
ALLSKY_SFCSWDWN_0	Irradiância descendente de ondas curtas da superfície do céu até um ano de idade	MJ/m <sup>2</sup> /dia	20,01	1,09
ALLSKY_SFCSWDWN_1	Irradiância descendente de ondas curtas da superfície do céu com um ano de idade	MJ/m <sup>2</sup> /dia	19,96	0,97
ALLSKY_SFCSWDWN_2	Irradiância descendente de ondas curtas da superfície do céu com dois anos de idade	MJ/m <sup>2</sup> /dia	19,87	0,98
ALLSKY_SFCSWDWN_3	Irradiância descendente de ondas curtas da superfície do céu com três anos de idade	MJ/m <sup>2</sup> /dia	19,96	0,79
ALLSKY_SFCSWDWN_4	Irradiância descendente de ondas curtas da superfície do céu com quatro anos de idade	MJ/m <sup>2</sup> /dia	20,06	0,89
ALLSKY_SFCSWDWN_5	Irradiância descendente de ondas curtas da superfície do céu com cinco anos de idade	MJ/m <sup>2</sup> /dia	19,98	0,94
ALLSKY_SFCSWDWN_6	Irradiância descendente de ondas curtas da superfície do céu com seis anos de idade	MJ/m <sup>2</sup> /dia	19,49	0,85
ALLSKY_SFCSWDWN_7	Irradiância descendente de ondas curtas da superfície do céu com sete anos de idade	MJ/m <sup>2</sup> /dia	19,78	0,95

Sigla	Variável	Unidade	Média	Desvio-padrão
ALLSKY_KT_0	Índice de clareza de insolação de todo o céu até um ano de idade (Idade <=1 ano)	Adimensional	0,58	0,03
ALLSKY_KT_1	Índice de clareza de insolação de todo o céu com um ano de idade	Adimensional	0,58	0,03
ALLSKY_KT_2	Índice de clareza de insolação de todo o céu com dois anos de idade	Adimensional	0,58	0,03
ALLSKY_KT_3	Índice de clareza de insolação de todo o céu com três anos de idade	Adimensional	0,58	0,02
ALLSKY_KT_4	Índice de clareza de insolação de todo o céu com quatro anos de idade	Adimensional	0,59	0,03
ALLSKY_KT_5	Índice de clareza de insolação de todo o céu com cinco anos de idade	Adimensional	0,59	0,03
ALLSKY_KT_6	Índice de clareza de insolação de todo o céu com seis anos de idade	Adimensional	0,57	0,03
ALLSKY_KT_7	Índice de clareza de insolação de todo o céu com sete anos de idade	Adimensional	0,58	0,03
TOA_SW_DWN_0	Irradiância descendente de ondas curtas no topo da atmosfera até um ano de idade	MJ/m <sup>2</sup> /dia	34,49	0,08
TOA_SW_DWN_1	Irradiância descendente de ondas curtas no topo da atmosfera com um ano de idade	MJ/m <sup>2</sup> /dia	34,50	0,08
TOA_SW_DWN_2	Irradiância descendente de ondas curtas no topo da atmosfera com dois anos de idade	MJ/m <sup>2</sup> /dia	34,49	0,08
TOA_SW_DWN_3	Irradiância descendente de ondas curtas no topo da atmosfera com três anos de idade	MJ/m <sup>2</sup> /dia	34,48	0,08
TOA_SW_DWN_4	Irradiância descendente de ondas curtas no topo da atmosfera com quatro anos de idade	MJ/m <sup>2</sup> /dia	34,48	0,08
TOA_SW_DWN_5	Irradiância descendente de ondas curtas no topo da atmosfera com cinco anos de idade	MJ/m <sup>2</sup> /dia	34,49	0,08
TOA_SW_DWN_6	Irradiância descendente de ondas curtas no topo da atmosfera com seis anos de idade	MJ/m <sup>2</sup> /dia	34,49	0,09
TOA_SW_DWN_7	Irradiância descendente de ondas curtas no topo da atmosfera com sete anos de idade	MJ/m <sup>2</sup> /dia	34,48	0,09
CLRSKY_SFCSWDWN_0	Irradiância descendente de ondas longas da superfície do céu claro até um ano de idade	W/m <sup>2</sup>	24,87	0,14
CLRSKY_SFCSWDWN_1	Irradiância descendente de ondas longas da superfície do céu claro com um ano de idade	W/m <sup>2</sup>	24,83	0,11
CLRSKY_SFCSWDWN_2	Irradiância descendente de ondas longas da superfície do céu claro com dois anos de idade	W/m <sup>2</sup>	24,88	0,16
CLRSKY_SFCSWDWN_3	Irradiância descendente de ondas longas da superfície do céu claro com três anos de idade	W/m <sup>2</sup>	24,90	0,16
CLRSKY_SFCSWDWN_4	Irradiância descendente de ondas longas da superfície do céu claro com quatro anos de idade	W/m <sup>2</sup>	24,83	0,16
CLRSKY_SFCSWDWN_5	Irradiância descendente de ondas longas da superfície do céu claro com cinco anos de idade	W/m <sup>2</sup>	24,85	0,14
CLRSKY_SFCSWDWN_6	Irradiância descendente de ondas longas da superfície do céu claro com seis anos de idade	W/m <sup>2</sup>	24,80	0,14
CLRSKY_SFCSWDWN_7	Irradiância descendente de ondas longas da superfície do céu claro com sete anos de idade	W/m <sup>2</sup>	24,80	0,14
WS10M_0	Velocidade do vento em 10 m até um ano de idade (Idade <=1 ano)	ms <sup>-1</sup>	3,39	0,24
WS10M_1	Velocidade do vento em 10 m com um ano de idade	ms <sup>-1</sup>	3,38	0,23
WS10M_2	Velocidade do vento em 10 m com dois anos de idade	ms <sup>-1</sup>	3,44	0,29
WS10M_3	Velocidade do vento em 10 m com três anos de idade	ms <sup>-1</sup>	3,45	0,30

Sigla	Variável	Unidade	Média	Desvio-padrão
WS10M_4	Velocidade do vento em 10 m com quatro anos de idade	ms <sup>-1</sup>	3,33	0,25
WS10M_5	Velocidade do vento em 10 m com cinco anos de idade	ms <sup>-1</sup>	3,39	0,28
WS10M_6	Velocidade do vento em 10 m com seis anos de idade	ms <sup>-1</sup>	3,32	0,28
WS10M_7	Velocidade do vento em 10 m com sete anos de idade	ms <sup>-1</sup>	3,19	0,24
T2M_0	Temperatura em 2 m até um ano de idade (Idade <=1 ano)	°C	23,51	1,43
T2M_1	Temperatura em 2 m com um ano de idade	°C	24,03	1,35
T2M_2	Temperatura em 2 m com dois anos de idade	°C	23,72	1,02
T2M_3	Temperatura em 2 m com três anos de idade	°C	23,64	1,04
T2M_4	Temperatura em 2 m com quatro anos de idade	°C	24,06	0,95
T2M_5	Temperatura em 2 m com cinco anos de idade	°C	23,85	0,90
T2M_6	Temperatura em 2 m com seis anos de idade	°C	23,31	0,85
T2M_7	Temperatura em 2 m com sete anos de idade	°C	23,32	0,89
T2M_MAX_0	Temperatura máxima em 2 m até um ano de idade (Idade <=1 ano)	°C	30,36	1,58
T2M_MAX_1	Temperatura máxima em 2 m com um ano de idade	°C	30,85	1,50
T2M_MAX_2	Temperatura máxima em 2 m com dois anos de idade	°C	30,55	1,12
T2M_MAX_3	Temperatura máxima em 2 m com três anos de idade	°C	30,52	1,12
T2M_MAX_4	Temperatura máxima em 2 m com quatro anos de idade	°C	30,91	1,12
T2M_MAX_5	Temperatura máxima em 2 m com cinco anos de idade	°C	30,72	1,06
T2M_MAX_6	Temperatura máxima em 2 m com seis anos de idade	°C	30,04	0,94
T2M_MAX_7	Temperatura máxima em 2 m com sete anos de idade	°C	30,08	1,11
T2MWET_0	Temperatura no bulbo úmido em 2 m até um ano de idade (Idade <=1 ano)	°C	19,06	0,66
T2MWET_1	Temperatura no bulbo úmido em 2 m com um ano de idade	°C	19,41	0,64
T2MWET_2	Temperatura no bulbo úmido em 2 m com dois anos de idade	°C	19,26	0,53
T2MWET_3	Temperatura no bulbo úmido em 2 m com três anos de idade	°C	19,14	0,62
T2MWET_4	Temperatura no bulbo úmido em 2 m com quatro anos de idade	°C	19,52	0,51
T2MWET_5	Temperatura no bulbo úmido em 2 m com cinco anos de idade	°C	19,49	0,49
T2MWET_6	Temperatura no bulbo úmido em 2 m com seis anos de idade	°C	19,23	0,56
T2MWET_7	Temperatura no bulbo úmido em 2 m com sete anos de idade	°C	19,14	0,51

Sigla	Variável	Unidade	Média	Desvio-padrão
T2MDEW_0	Ponto de orvalho/geada em 2 m até um ano de idade (Idade <=1 ano)	°C	14,61	0,38
T2MDEW_1	Ponto de orvalho/geada em 2 m com um ano de idade	°C	14,79	0,54
T2MDEW_2	Ponto de orvalho/geada em 2 m com dois anos de idade	°C	14,81	0,54
T2MDEW_3	Ponto de orvalho/geada em 2 m com três anos de idade	°C	14,63	0,52
T2MDEW_4	Ponto de orvalho/geada em 2 m com quatro anos de idade	°C	14,98	0,40
T2MDEW_5	Ponto de orvalho/geada em 2 m com cinco anos de idade	°C	15,12	0,51
T2MDEW_6	Ponto de orvalho/geada em 2 m com seis anos de idade	°C	15,15	0,61
T2MDEW_7	Ponto de orvalho/geada em 2 m com sete anos de idade	°C	14,96	0,40
T2M_RANGE_0	Faixa de variação diária de temperatura em 2 m até um ano de idade (Idade <=1 ano)	°C	12,66	0,50
T2M_RANGE_1	Faixa de variação diária de temperatura em 2 m com um ano de idade	°C	12,68	0,51
T2M_RANGE_2	Faixa de variação diária de temperatura em 2 m com dois anos de idade	°C	12,65	0,42
T2M_RANGE_3	Faixa de variação diária de temperatura em 2 m com três anos de idade	°C	12,75	0,50
T2M_RANGE_4	Faixa de variação diária de temperatura em 2 m com quatro anos de idade	°C	12,74	0,54
T2M_RANGE_5	Faixa de variação diária de temperatura em 2 m com cinco anos de idade	°C	12,72	0,56
T2M_RANGE_6	Faixa de variação diária de temperatura em 2 m com seis anos de idade	°C	12,38	0,45
T2M_RANGE_7	Faixa de variação diária de temperatura em 2 m com sete anos de idade	°C	12,50	0,55
T2M_MIN_0	Temperatura mínima em 2 m até um ano de idade (Idade <=1 ano)	°C	17,70	1,14
T2M_MIN_1	Temperatura mínima em 2 m com um ano de idade	°C	18,17	1,09
T2M_MIN_2	Temperatura mínima em 2 m com dois anos de idade	°C	17,90	0,84
T2M_MIN_3	Temperatura mínima em 2 m com três anos de idade	°C	17,78	0,87
T2M_MIN_4	Temperatura mínima em 2 m com quatro anos de idade	°C	18,17	0,75
T2M_MIN_5	Temperatura mínima em 2 m com cinco anos de idade	°C	18,00	0,70
T2M_MIN_6	Temperatura mínima em 2 m com seis anos de idade	°C	17,66	0,71
T2M_MIN_7	Temperatura mínima em 2 m com sete anos de idade	°C	17,58	0,66
PS_0	Pressão de superfície até um ano de idade (Idade <=1 ano)	kPa	92,75	0,76
PS_1	Pressão de superfície com um ano de idade	kPa	92,76	0,75
PS_2	Pressão de superfície com dois anos de idade	kPa	92,77	0,76
PS_3	Pressão de superfície com três anos de idade	kPa	92,77	0,74

Sigla	Variável	Unidade	Média	Desvio-padrão
PS_4	Pressão de superfície com quatro anos de idade	kPa	92,76	0,75
PS_5	Pressão de superfície com cinco anos de idade	kPa	92,77	0,75
PS_6	Pressão de superfície com seis anos de idade	kPa	92,76	0,74
PS_7	Pressão de superfície com sete anos de idade	kPa	92,74	0,75
GWETPROF_0	Porcentagem de umidade do solo do perfil até um ano de idade (Idade <=1 ano)	%	0,61	0,09
GWETPROF_1	Porcentagem de umidade do solo do perfil com um ano de idade	%	0,60	0,09
GWETPROF_2	Porcentagem de umidade do solo do perfil com dois anos de idade	%	0,60	0,09
GWETPROF_3	Porcentagem de umidade do solo do perfil com três anos de idade	%	0,59	0,09
GWETPROF_4	Porcentagem de umidade do solo do perfil com quatro anos de idade	%	0,60	0,10
GWETPROF_5	Porcentagem de umidade do solo do perfil com cinco anos de idade	%	0,60	0,09
GWETPROF_6	Porcentagem de umidade do solo do perfil com seis anos de idade	%	0,61	0,09
GWETPROF_7	Porcentagem de umidade do solo do perfil com sete anos de idade	%	0,61	0,09
GWETTOP_0	Porcentagem de umidade do solo (até 5 cm) até um ano de idade (Idade <=1 ano)	%	0,48	0,05
GWETTOP_1	Porcentagem de umidade do solo (até 5 cm) com um ano de idade	%	0,47	0,06
GWETTOP_2	Porcentagem de umidade do solo (até 5 cm) com dois anos de idade	%	0,48	0,05
GWETTOP_3	Porcentagem de umidade do solo (até 5 cm) com três anos de idade	%	0,47	0,06
GWETTOP_4	Porcentagem de umidade do solo (até 5 cm) com quatro anos de idade	%	0,48	0,06
GWETTOP_5	Porcentagem de umidade do solo (até 5 cm) com cinco anos de idade	%	0,49	0,06
GWETTOP_6	Porcentagem de umidade do solo (até 5 cm) com seis anos de idade	%	0,51	0,06
GWETTOP_7	Porcentagem de umidade do solo (até 5 cm) com sete anos de idade	%	0,51	0,06
GWETROOT_0	Porcentagem de umidade do solo na zona radicular (até 100 cm) até um ano de idade	%	0,58	0,07
GWETROOT_1	Porcentagem de umidade do solo na zona radicular (até 100 cm) com um ano de idade	%	0,58	0,07
GWETROOT_2	Porcentagem de umidade do solo na zona radicular (até 100 cm) com dois anos de idade	%	0,58	0,07
GWETROOT_3	Porcentagem de umidade do solo na zona radicular (até 100 cm) com três anos de idade	%	0,57	0,08
GWETROOT_4	Porcentagem de umidade do solo na zona radicular (até 100 cm) com quatro anos de idade	%	0,57	0,08
GWETROOT_5	Porcentagem de umidade do solo na zona radicular (até 100 cm) com cinco anos de idade	%	0,58	0,08
GWETROOT_6	Porcentagem de umidade do solo na zona radicular (até 100 cm) com seis anos de idade	%	0,59	0,08
GWETROOT_7	Porcentagem de umidade do solo na zona radicular (até 100 cm) com sete anos de idade	%	0,59	0,08

Sigla	Variável	Unidade	Média	Desvio-padrão
TS_0	Temperatura da superfície da terra até um ano de idade (Idade <=1 ano)	°C	24,62	1,53
TS_1	Temperatura da superfície da terra com um ano de idade	°C	25,15	1,45
TS_2	Temperatura da superfície da terra com dois anos de idade	°C	24,82	1,03
TS_3	Temperatura da superfície da terra com três anos de idade	°C	24,78	1,07
TS_4	Temperatura da superfície da terra com quatro anos de idade	°C	25,18	1,00
TS_5	Temperatura da superfície da terra com cinco anos de idade	°C	24,92	0,95
TS_6	Temperatura da superfície da terra com seis anos de idade	°C	24,29	0,85
TS_7	Temperatura da superfície da terra com sete anos de idade	°C	24,30	0,99
ALLSKY_SFC_LW_DWN_0	Irradiância descendente de ondas longas da superfície do céu até um ano de idade (Idade <=1 ano)	W/m <sup>2</sup>	374,39	5,17
ALLSKY_SFC_LW_DWN_1	Irradiância descendente de ondas longas da superfície do céu com um ano de idade	W/m <sup>2</sup>	378,41	4,54
ALLSKY_SFC_LW_DWN_2	Irradiância descendente de ondas longas da superfície do céu com dois anos de idade	W/m <sup>2</sup>	377,31	3,80
ALLSKY_SFC_LW_DWN_3	Irradiância descendente de ondas longas da superfície do céu com três anos de idade	W/m <sup>2</sup>	375,75	3,74
ALLSKY_SFC_LW_DWN_4	Irradiância descendente de ondas longas da superfície do céu com quatro anos de idade	W/m <sup>2</sup>	377,30	2,71
ALLSKY_SFC_LW_DWN_5	Irradiância descendente de ondas longas da superfície do céu com cinco anos de idade	W/m <sup>2</sup>	376,15	2,70
ALLSKY_SFC_LW_DWN_6	Irradiância descendente de ondas longas da superfície do céu com seis anos de idade	W/m <sup>2</sup>	374,98	2,68
ALLSKY_SFC_LW_DWN_7	Irradiância descendente de ondas longas da superfície do céu com sete anos de idade	W/m <sup>2</sup>	374,87	2,80
ALLSKY_SFC_PAR_TOT_0	Radiação fotossinteticamente ativa total de toda a superfície do céu até um ano de idade	W/m <sup>2</sup>	106,50	5,79
ALLSKY_SFC_TOT_1	Radiação fotossinteticamente ativa total de toda a superfície do céu com um ano de idade	W/m <sup>2</sup>	106,32	5,00
ALLSKY_SFC_TOT_2	Radiação fotossinteticamente ativa total de toda a superfície do céu com dois anos de idade	W/m <sup>2</sup>	105,68	4,96
ALLSKY_SFC_TOT_3	Radiação fotossinteticamente ativa total de toda a superfície do céu com três anos de idade	W/m <sup>2</sup>	106,09	4,07
ALLSKY_SFC_TOT_4	Radiação fotossinteticamente ativa total de toda a superfície do céu com quatro anos de idade	W/m <sup>2</sup>	106,82	4,68
ALLSKY_SFC_TOT_5	Radiação fotossinteticamente ativa total de toda a superfície do céu com cinco anos de idade	W/m <sup>2</sup>	106,39	5,02
ALLSKY_SFC_TOT_6	Radiação fotossinteticamente ativa total de toda a superfície do céu com seis anos de idade	W/m <sup>2</sup>	103,69	4,44
ALLSKY_SFC_TOT_7	Radiação fotossinteticamente ativa total de toda a superfície do céu com sete anos de idade	W/m <sup>2</sup>	105,16	4,91
CLRSKY_SFC_TOT_0	Radiação fotossinteticamente ativa total de toda a superfície do céu claro até um ano de idade	W/m <sup>2</sup>	131,29	1,00
CLRSKY_SFC_TOT_1	Radiação fotossinteticamente ativa total de toda a superfície do céu claro com um ano de idade	W/m <sup>2</sup>	131,16	0,75
CLRSKY_SFC_TOT_2	Radiação fotossinteticamente ativa total de toda a superfície do céu claro com dois anos de idade	W/m <sup>2</sup>	131,24	0,80
CLRSKY_SFC_TOT_3	Radiação fotossinteticamente ativa total de toda a superfície do céu claro com três anos de idade	W/m <sup>2</sup>	131,24	0,87

Sigla	Variável	Unidade	Média	Desvio-padrão
CLRSKY_SFC_TOT_4	Radiação fotossinteticamente ativa total de toda a superfície do céu claro com quatro anos de idade	W/m <sup>2</sup>	131,07	1,03
CLRSKY_SFC_TOT_5	Radiação fotossinteticamente ativa total de toda a superfície do céu claro com cinco anos de idade	W/m <sup>2</sup>	131,06	0,91
CLRSKY_SFC_TOT_6	Radiação fotossinteticamente ativa total de toda a superfície do céu claro com seis anos de idade	W/m <sup>2</sup>	130,54	0,70
CLRSKY_SFC_TOT_7	Radiação fotossinteticamente ativa total de toda a superfície do céu claro com sete anos de idade	W/m <sup>2</sup>	130,41	0,74
DQ_0	Número de dias quentes até um ano de idade (Idade <=1 ano)	Contagem	38,70	20,34
DQ_1	Número de dias quentes com um ano de idade	Contagem	40,78	21,83
DQ_2	Número de dias quentes com dois anos de idade	Contagem	42,44	23,43
DQ_3	Número de dias quentes com três anos de idade	Contagem	39,23	19,85
DQ_4	Número de dias quentes com quatro anos de idade	Contagem	40,95	21,27
DQ_5	Número de dias quentes com cinco anos de idade	Contagem	40,03	20,95
DQ_6	Número de dias quentes com seis anos de idade	Contagem	37,54	18,75
DQ_7	Número de dias quentes com sete anos de idade	Contagem	37,56	18,11
DF_0	Número de dias frios até um ano de idade (Idade <=1 ano)	Contagem	31,42	28,05
DF_1	Número de dias frios com um ano de idade	Contagem	33,35	27,48
DF_2	Número de dias frios com dois anos de idade	Contagem	34,61	25,88
DF_3	Número de dias frios com três anos de idade	Contagem	33,48	27,76
DF_4	Número de dias frios com quatro anos de idade	Contagem	34,35	26,10
DF_5	Número de dias frios com cinco anos de idade	Contagem	32,74	23,26
DF_6	Número de dias frios com seis anos de idade	Contagem	30,66	23,71
DF_7	Número de dias frios com sete anos de idade	Contagem	35,48	25,19
T_otima_0	Número de dias de temperatura ótima até um ano de idade (Idade <=1 ano)	Contagem	103,93	62,77
T_otima_1	Número de dias de temperatura ótima com um ano de idade	Contagem	81,30	58,76
T_otima_2	Número de dias de temperatura ótima com dois anos de idade	Contagem	83,51	49,49
T_otima_3	Número de dias de temperatura ótima com três anos de idade	Contagem	91,76	48,86
T_otima_4	Número de dias de temperatura ótima com quatro anos de idade	Contagem	75,00	39,88
T_otima_5	Número de dias de temperatura ótima com cinco anos de idade	Contagem	77,93	32,87
T_otima_6	Número de dias de temperatura ótima com seis anos de idade	Contagem	98,19	39,82
T_otima_7	Número de dias de temperatura ótima com sete anos de idade	Contagem	100,19	45,60

Sigla	Variável	Unidade	Média	Desvio-padrão
DU_panual_0	Número de dias úmidos até um ano de idade (Idade <=1 ano)	Contagem	34,00	11,22
DU_panual_1	Número de dias úmidos com um ano de idade	Contagem	36,19	10,58
DU_panual_2	Número de dias úmidos com dois anos de idade	Contagem	34,83	7,99
DU_panual_3	Número de dias úmidos com três anos de idade	Contagem	35,06	12,50
DU_panual_4	Número de dias úmidos com quatro anos de idade	Contagem	36,26	13,13
DU_panual_5	Número de dias úmidos com cinco anos de idade	Contagem	35,19	9,72
DU_panual_6	Número de dias úmidos com seis anos de idade	Contagem	33,35	12,38
DU_panual_7	Número de dias úmidos com sete anos de idade	Contagem	35,27	12,93
DS_panual_0	Número de dias secos até um ano de idade (Idade <=1 ano)	Contagem	39,43	17,81
DS_panual_1	Número de dias secos com um ano de idade	Contagem	39,90	21,86
DS_panual_2	Número de dias secos com dois anos de idade	Contagem	41,07	23,94
DS_panual_3	Número de dias secos com três anos de idade	Contagem	39,53	21,61
DS_panual_4	Número de dias secos com quatro anos de idade	Contagem	40,38	20,78
DS_panual_5	Número de dias secos com cinco anos de idade	Contagem	37,97	19,61
DS_panual_6	Número de dias secos com seis anos de idade	Contagem	36,57	16,49
DS_panual_7	Número de dias secos com sete anos de idade	%	35,82	14,02
chuva_seca_0	% de chuva no período seco até um ano de idade (Idade <=1 ano)	%	2,05	1,97
chuva_seca_1	% de chuva no período seco com um ano de idade	%	2,73	1,90
chuva_seca_2	% de chuva no período seco com dois anos de idade	%	3,07	1,68
chuva_seca_3	% de chuva no período seco com três anos de idade	%	2,58	1,88
chuva_seca_4	% de chuva no período seco com quatro anos de idade	%	3,34	1,93
chuva_seca_5	% de chuva no período seco com cinco anos de idade	%	3,65	2,45
chuva_seca_6	% de chuva no período seco com seis anos de idade	%	2,49	1,85
chuva_seca_7	% de chuva no período seco com sete anos de idade	%	4,08	2,02
DC_0	Número de dias com precipitação até um ano de idade (Idade <=1 ano)	Contagem	200,90	40,70
DC_1	Número de dias com precipitação com um ano de idade	Contagem	204,63	42,47
DC_2	Número de dias com precipitação com dois anos de idade	Contagem	203,01	41,27
DC_3	Número de dias com precipitação com três anos de idade	Contagem	203,18	35,80

Sigla	Variável	Unidade	Média	Desvio-padrão
DC_4	Número de dias com precipitação com quatro anos de idade	Contagem	211,15	26,06
DC_5	Número de dias com precipitação com cinco anos de idade	Contagem	226,36	26,93
DC_6	Número de dias com precipitação com seis anos de idade	Contagem	247,68	23,01
DC_7	Número de dias com precipitação com sete anos de idade	Contagem	248,96	16,05
DSC_0	Número de dias sem precipitação até um ano de idade (Idade <=1 ano)	Contagem	164,25	40,57
DSC_1	Número de dias sem precipitação com um ano de idade	Contagem	160,82	42,44
DSC_2	Número de dias sem precipitação com dois anos de idade	Contagem	162,36	41,41
DSC_3	Número de dias sem precipitação com três anos de idade	Contagem	161,86	35,81
DSC_4	Número de dias sem precipitação com quatro anos de idade	Contagem	153,99	26,00
DSC_5	Número de dias sem precipitação com cinco anos de idade	Contagem	139,10	26,72
DSC_6	Número de dias sem precipitação com seis anos de idade	Contagem	117,68	22,94
DSC_7	Número de dias sem precipitação com sete anos de idade	Contagem	116,07	16,05
S1	Capina Mecânica em Linha Total com Pré-Emergente	Dias ref.	2,48	0,32
S2	Primeira Adubação de Manutenção (Ref. 6 meses)	Dias ref.	3,78	1,71
S3	Primeira Capina Manual em Linha Total	Dias ref.	5,09	3,00
S4	Segunda Adubação de Manutenção (Ref. 12 meses)	Dias ref.	11,83	3,42
S5	Segunda Capina Manual em Linha Total	Dias ref.	11,10	3,71
S6	Subsolagem com Fosfatagem	Dias ref.	0,54	0,93
S7	Terceira Adubação de Manutenção (Monitoramento)	Dias ref.	27,33	3,45
S8	Terceira Capina Manual em Linha Total	Dias ref.	20,37	3,72
GEN	Genótipos		8 genótipos	
CLAY	Teor de argila	%	30,45	16,64
ORGMATTER	Teor de matéria orgânica	%	3,09	1,76
ALTITUDE	Altitude	m	902,82	68,90
PlantArea	Área útil por planta	m <sup>2</sup>	9,98	1,18
P	Quantidade de adubação de Fósforo	kg/ha	81,11	16,19
N	Quantidade de adubação de Nitrogênio	kg/ha	54,75	21,32
K	Quantidade de adubação de Potássio	kg/ha	188,13	46,54

Sigla	Variável	Unidade	Média	Desvio-padrão
IMA <sub>7</sub>	Produtividade média aos 7 anos	m <sup>2</sup> ha <sup>-1</sup> ano <sup>-1</sup>	25,34	10,52

### APÊNDICE B

Tabela: Coeficientes de correlação linear de Pearson ( $r_{\hat{y}y}$ ) e  $p$ -valor (95% de probabilidade) entre as variáveis preditoras e a produtividade média aos 7 anos ( $IMA_7$ ) de eucalipto no norte de Minas Gerais

Variável	$r_{\hat{y}y}$	$p$ -valor	Variável	$r_{\hat{y}y}$	$p$ -valor
prec_0	0,3528	0,0000	T2M_MIN_0	-0,2609	0,0000
prec_1	0,3040	0,0000	T2M_MIN_1	-0,4093	0,0000
prec_2	0,3122	0,0000	T2M_MIN_2	-0,3765	0,0000
prec_3	-0,0042	0,9402	T2M_MIN_3	-0,3175	0,0000
prec_4	-0,1177	0,0353	T2M_MIN_4	-0,2425	0,0000
prec_5	0,2614	0,0000	T2M_MIN_5	-0,3443	0,0000
prec_6	-0,0591	0,2923	T2M_MIN_6	-0,3510	0,0000
prec_7	0,0790	0,1584	T2M_MIN_7	-0,3101	0,0000
prec_ac_0	0,3528	0,0000	PS_0	-0,2208	0,0001
prec_ac_1	0,3977	0,0000	PS_1	-0,2087	0,0002
prec_ac_2	0,4586	0,0000	PS_2	-0,2123	0,0001
prec_ac_3	0,4846	0,0000	PS_3	-0,2059	0,0002
prec_ac_4	0,4686	0,0000	PS_4	-0,2046	0,0002
prec_ac_5	0,4714	0,0000	PS_5	-0,2053	0,0002
prec_ac_6	0,4657	0,0000	PS_6	-0,2015	0,0003
prec_ac_7	0,4484	0,0000	PS_7	-0,1999	0,0003
ALLSKY_SFC_UVA_0	-0,3511	0,0000	GWETPROF_0	-0,1992	0,0003

Variável	$r_{\hat{y}y}$	<i>p</i> -valor	Variável	$r_{\hat{y}y}$	<i>p</i> -valor
ALLSKY_SFC_UVA_1	-0,3079	0,0000	GWETPROF_1	-0,1366	0,0144
ALLSKY_SFC_UVA_2	-0,3134	0,0000	GWETPROF_2	-0,1899	0,0006
ALLSKY_SFC_UVA_3	-0,1661	0,0029	GWETPROF_3	-0,1759	0,0016
ALLSKY_SFC_UVA_4	-0,1021	0,0681	GWETPROF_4	-0,2336	0,0000
ALLSKY_SFC_UVA_5	-0,2747	0,0000	GWETPROF_5	-0,1930	0,0005
ALLSKY_SFC_UVA_6	-0,1844	0,0009	GWETPROF_6	-0,2310	0,0000
ALLSKY_SFC_UVA_7	-0,2113	0,0001	GWETPROF_7	-0,1842	0,0009
ALLSKY_SFC_UVB_0	-0,3280	0,0000	GWETTOP_0	-0,0065	0,9079
ALLSKY_SFC_UVB_1	-0,3194	0,0000	GWETTOP_1	0,2554	0,0000
ALLSKY_SFC_UVB_2	-0,3312	0,0000	GWETTOP_2	0,0263	0,6390
ALLSKY_SFC_UVB_3	-0,2091	0,0002	GWETTOP_3	0,0054	0,9227
ALLSKY_SFC_UVB_4	-0,0832	0,1378	GWETTOP_4	-0,2287	0,0000
ALLSKY_SFC_UVB_5	-0,2534	0,0000	GWETTOP_5	-0,0721	0,1981
ALLSKY_SFC_UVB_6	-0,2917	0,0000	GWETTOP_6	-0,2264	0,0000
ALLSKY_SFC_UVB_7	-0,2424	0,0000	GWETTOP_7	-0,0683	0,2230
ALLSKY_SFC_UV_INDEX_0	-0,3211	0,0000	GWETROOT_0	-0,0351	0,5312
ALLSKY_SFC_UV_INDEX_1	-0,3159	0,0000	GWETROOT_1	-0,0046	0,9340
ALLSKY_SFC_UV_INDEX_2	-0,3397	0,0000	GWETROOT_2	-0,0589	0,2938
ALLSKY_SFC_UV_INDEX_3	-0,2161	0,0001	GWETROOT_3	-0,0839	0,1340
ALLSKY_SFC_UV_INDEX_4	-0,0723	0,1969	GWETROOT_4	-0,1347	0,0159

Variável	$r_{\hat{y}y}$	<i>p</i> -valor	Variável	$r_{\hat{y}y}$	<i>p</i> -valor
ALLSKY_SFC_UV_INDEX_5	-0,2477	0,0000	GWETROOT_5	-0,0875	0,1181
ALLSKY_SFC_UV_INDEX_6	-0,3074	0,0000	GWETROOT_6	-0,1285	0,0215
ALLSKY_SFC_UV_INDEX_7	-0,2493	0,0000	GWETROOT_7	-0,0817	0,1449
QV2M_0	0,1958	0,0004	TS_0	-0,2622	0,0000
QV2M_1	0,2461	0,0000	TS_1	-0,4326	0,0000
QV2M_2	0,1473	0,0083	TS_2	-0,4024	0,0000
QV2M_3	0,1846	0,0009	TS_3	-0,3380	0,0000
QV2M_4	-0,1415	0,0113	TS_4	-0,1305	0,0195
QV2M_5	0,0993	0,0761	TS_5	-0,3362	0,0000
QV2M_6	-0,2195	0,0001	TS_6	-0,2639	0,0000
QV2M_7	0,3058	0,0000	TS_7	-0,2883	0,0000
RH2M_0	0,2873	0,0000	ALLSKY_SFC_LW_DWN_0	-0,1854	0,0009
RH2M_1	0,4313	0,0000	ALLSKY_SFC_LW_DWN_1	-0,4152	0,0000
RH2M_2	0,3909	0,0000	ALLSKY_SFC_LW_DWN_2	-0,2841	0,0000
RH2M_3	0,4180	0,0000	ALLSKY_SFC_LW_DWN_3	-0,3186	0,0000
RH2M_4	0,1300	0,0200	ALLSKY_SFC_LW_DWN_4	-0,1009	0,0715
RH2M_5	0,3277	0,0000	ALLSKY_SFC_LW_DWN_5	-0,1020	0,0684
RH2M_6	0,1243	0,0261	ALLSKY_SFC_LW_DWN_6	-0,2303	0,0000
RH2M_7	0,4180	0,0000	ALLSKY_SFC_LW_DWN_7	-0,3706	0,0000
WS2M_0	-0,1398	0,0123	ALLSKY_SFC_PAR_TOT_0	-0,3510	0,0000

Variável	$r_{\hat{y}y}$	<i>p</i> -valor	Variável	$r_{\hat{y}y}$	<i>p</i> -valor
WS2M_1	-0,0964	0,0853	ALLSKY_SFC_PAR_TOT_1	-0,3102	0,0000
WS2M_2	-0,1027	0,0665	ALLSKY_SFC_PAR_TOT_2	-0,3244	0,0000
WS2M_3	-0,1814	0,0011	ALLSKY_SFC_PAR_TOT_3	-0,1771	0,0015
WS2M_4	-0,0722	0,1978	ALLSKY_SFC_PAR_TOT_4	-0,1118	0,0457
WS2M_5	-0,0837	0,1350	ALLSKY_SFC_PAR_TOT_5	-0,2813	0,0000
WS2M_6	0,0323	0,5645	ALLSKY_SFC_PAR_TOT_6	-0,2082	0,0002
WS2M_7	-0,0272	0,6281	ALLSKY_SFC_PAR_TOT_7	-0,2296	0,0000
TQV_0	-0,1716	0,0021	CLRSKY_SFC_PAR_TOT_0	-0,3286	0,0000
TQV_1	-0,2672	0,0000	CLRSKY_SFC_PAR_TOT_1	-0,1842	0,0009
TQV_2	-0,0906	0,1057	CLRSKY_SFC_PAR_TOT_2	-0,2170	0,0001
TQV_3	-0,1493	0,0075	CLRSKY_SFC_PAR_TOT_3	-0,2266	0,0000
TQV_4	-0,3402	0,0000	CLRSKY_SFC_PAR_TOT_4	-0,2168	0,0001
TQV_5	-0,0811	0,1476	CLRSKY_SFC_PAR_TOT_5	-0,2680	0,0000
TQV_6	-0,3558	0,0000	CLRSKY_SFC_PAR_TOT_6	-0,0337	0,5482
TQV_7	-0,0441	0,4320	CLRSKY_SFC_PAR_TOT_7	-0,2192	0,0001
ALLSKY_SFC_SW_DWN_0	-0,3432	0,0000	DQ_0	-0,3033	0,0000
ALLSKY_SFC_SW_DWN_1	-0,2978	0,0000	DQ_1	-0,3482	0,0000
ALLSKY_SFC_SW_DWN_2	-0,3197	0,0000	DQ_2	-0,2290	0,0000
ALLSKY_SFC_SW_DWN_3	-0,1721	0,0020	DQ_3	-0,3182	0,0000
ALLSKY_SFC_SW_DWN_4	-0,0946	0,0910	DQ_4	-0,3404	0,0000

Variável	$r_{\hat{y}y}$	<i>p</i> -valor	Variável	$r_{\hat{y}y}$	<i>p</i> -valor
ALLSKY_SFC_SW_DWN_5	-0,2706	0,0000	DQ_5	-0,3455	0,0000
ALLSKY_SFC_SW_DWN_6	-0,1762	0,0016	DQ_6	-0,3380	0,0000
ALLSKY_SFC_SW_DWN_7	-0,2298	0,0000	DQ_7	-0,3358	0,0000
ALLSKY_KT_0	-0,3270	0,0000	DF_0	0,4133	0,0000
ALLSKY_KT_1	-0,2811	0,0000	DF_1	0,4279	0,0000
ALLSKY_KT_2	-0,3096	0,0000	DF_2	0,4105	0,0000
ALLSKY_KT_3	-0,1818	0,0011	DF_3	0,3905	0,0000
ALLSKY_KT_4	-0,1338	0,0166	DF_4	0,3963	0,0000
ALLSKY_KT_5	-0,2600	0,0000	DF_5	0,3530	0,0000
ALLSKY_KT_6	-0,1825	0,0010	DF_6	0,4113	0,0000
ALLSKY_KT_7	-0,2351	0,0000	DF_7	0,3524	0,0000
TOA_SW_DWN_0	-0,3003	0,0000	T_otima_0	0,3406	0,0000
TOA_SW_DWN_1	-0,2795	0,0000	T_otima_1	0,4232	0,0000
TOA_SW_DWN_2	-0,3089	0,0000	T_otima_2	0,4549	0,0000
TOA_SW_DWN_3	-0,2884	0,0000	T_otima_3	0,3895	0,0000
TOA_SW_DWN_4	-0,2794	0,0000	T_otima_4	0,2607	0,0000
TOA_SW_DWN_5	-0,2578	0,0000	T_otima_5	0,2648	0,0000
TOA_SW_DWN_6	-0,3114	0,0000	T_otima_6	0,3107	0,0000
TOA_SW_DWN_7	-0,2986	0,0000	T_otima_7	0,3625	0,0000
CLRSKY_SFC_SW_DWN_0	-0,2632	0,0000	DU_panual_0	0,4055	0,0000

Variável	$r_{\hat{y}y}$	<i>p</i> -valor	Variável	$r_{\hat{y}y}$	<i>p</i> -valor
CLRSKY_SFC_SW_DWN_1	-0,0474	0,3978	DU_panual_1	0,4464	0,0000
CLRSKY_SFC_SW_DWN_2	-0,1587	0,0044	DU_panual_2	0,4692	0,0000
CLRSKY_SFC_SW_DWN_3	-0,1496	0,0073	DU_panual_3	0,3826	0,0000
CLRSKY_SFC_SW_DWN_4	-0,1135	0,0425	DU_panual_4	0,3731	0,0000
CLRSKY_SFC_SW_DWN_5	-0,1824	0,0010	DU_panual_5	0,2488	0,0000
CLRSKY_SFC_SW_DWN_6	0,1574	0,0048	DU_panual_6	0,4554	0,0000
CLRSKY_SFC_SW_DWN_7	-0,2241	0,0001	DU_panual_7	0,4479	0,0000
WS10M_0	-0,1883	0,0007	DS_panual_0	-0,2823	0,0000
WS10M_1	-0,1162	0,0377	DS_panual_1	-0,3613	0,0000
WS10M_2	-0,1325	0,0178	DS_panual_2	-0,2388	0,0000
WS10M_3	-0,2259	0,0000	DS_panual_3	-0,2699	0,0000
WS10M_4	-0,1026	0,0669	DS_panual_4	-0,2249	0,0000
WS10M_5	-0,1170	0,0364	DS_panual_5	-0,3096	0,0000
WS10M_6	0,0247	0,6600	DS_panual_6	-0,3423	0,0000
WS10M_7	-0,0466	0,4063	DS_panual_7	-0,3281	0,0000
T2M_0	-0,2698	0,0000	%chuva_seca_0	0,0673	0,2302
T2M_1	-0,4246	0,0000	%chuva_seca_1	0,3154	0,0000
T2M_2	-0,3990	0,0000	%chuva_seca_2	0,2112	0,0001
T2M_3	-0,3468	0,0000	%chuva_seca_3	0,4447	0,0000
T2M_4	-0,1689	0,0024	%chuva_seca_4	0,1437	0,0101

Variável	$r_{\hat{y}y}$	<i>p</i> -valor	Variável	$r_{\hat{y}y}$	<i>p</i> -valor
T2M_5	-0,3517	0,0000	%chuva_seca_5	-0,0957	0,0873
T2M_6	-0,3079	0,0000	%chuva_seca_6	0,2129	0,0001
T2M_7	-0,3213	0,0000	%chuva_seca_7	-0,0516	0,3572
T2M_MAX_0	-0,2590	0,0000	DC_0	0,3597	0,0000
T2M_MAX_1	-0,4176	0,0000	DC_1	0,3685	0,0000
T2M_MAX_2	-0,3934	0,0000	DC_2	0,4155	0,0000
T2M_MAX_3	-0,3317	0,0000	DC_3	0,4399	0,0000
T2M_MAX_4	-0,0900	0,1080	DC_4	0,3122	0,0000
T2M_MAX_5	-0,3281	0,0000	DC_5	0,2918	0,0000
T2M_MAX_6	-0,2374	0,0000	DC_6	0,2098	0,0002
T2M_MAX_7	-0,2911	0,0000	DC_7	0,3291	0,0000
T2MWET_0	-0,2682	0,0000	DSC_0	-0,3610	0,0000
T2MWET_1	-0,3643	0,0000	DSC_1	-0,3660	0,0000
T2MWET_2	-0,3314	0,0000	DSC_2	-0,4167	0,0000
T2MWET_3	-0,2259	0,0000	DSC_3	-0,4398	0,0000
T2MWET_4	-0,2133	0,0001	DSC_4	-0,3131	0,0000
T2MWET_5	-0,2804	0,0000	DSC_5	-0,2899	0,0000
T2MWET_6	-0,3559	0,0000	DSC_6	-0,2151	0,0001
T2MWET_7	-0,1926	0,0005	DSC_7	-0,3294	0,0000
T2MDEW_0	0,0845	0,1316	S1	-0,0202	0,7184

Variável	$r_{\hat{y}y}$	<i>p</i> -valor	Variável	$r_{\hat{y}y}$	<i>p</i> -valor
T2MDEW_1	0,1980	0,0004	S2	0,0034	0,9515
T2MDEW_2	0,1068	0,0563	S3	0,0017	0,9755
T2MDEW_3	0,1523	0,0063	S4	0,2440	0,0000
T2MDEW_4	-0,1477	0,0081	S5	0,2587	0,0000
T2MDEW_5	0,0801	0,1527	S6	-0,1788	0,0013
T2MDEW_6	-0,2236	0,0001	S7	0,0156	0,7815
T2MDEW_7	0,2253	0,0000	S8	0,1320	0,0182
T2M_RANGE_0	-0,2253	0,0000	CLAY	0,4118	0,0000
T2M_RANGE_1	-0,3529	0,0000	ORGMATTER	0,4002	0,0000
T2M_RANGE_2	-0,2893	0,0000	ALTITUDE	0,4925	0,0000
T2M_RANGE_3	-0,1885	0,0007	PlantArea	-0,1399	0,0122
T2M_RANGE_4	0,1497	0,0073	P	-0,1023	0,0676
T2M_RANGE_5	-0,1861	0,0008	N	0,4162	0,0000
T2M_RANGE_6	0,0571	0,3089	K	0,1217	0,0295
T2M_RANGE_7	-0,2159	0,0001	-	-	-

### APÊNDICE C

Tabela: Importância relativa (%) obtida por perturbação de redes neurais artificiais para 304 variáveis preditoras da produtividade média aos 7 anos de eucalipto no norte de Minas Gerais

Variável	%	Variável	%
GEN	4,73	T2M_RANGE_2	0,19
ALTITUDE	3,68	GWETPROF_0	0,19
ORGMATTER	3,46	RH2M_7	0,19
PlantArea	2,42	T_otima_7	0,18
S5	2,28	DF_5	0,18
N	2,15	T_otima_1	0,18
S2	1,65	DU_panual_6	0,18
S8	1,56	PS_2	0,18
S4	1,42	WS10M_0	0,18
S1	1,37	CLRSKY_SFC_SW_DWN_2	0,18
S7	1,36	ALLSKY_SFC_LW_DWN_4	0,18
CLAY	1,24	PS_0	0,18
K	1,14	DS_panual_0	0,18
DS_panual_4	1,09	T2M_MAX_4	0,17
chuva_seca_2	1,07	WS10M_3	0,17
S3	1,00	ALLSKY_SFC_UV_INDEX_2	0,17
TOA_SW_DWN_7	0,95	RH2M_5	0,17

Variável	%	Variável	%
DS_panual_3	0,95	WS10M_4	0,17
DS_panual_1	0,87	prec_ac_3	0,17
DS_panual_2	0,87	prec_5	0,16
DS_panual_5	0,84	CLRSKY_SFC_SW_DWN_0	0,16
chuva_seca_1	0,82	ALLSKY_SFC_PAR_TOT_0	0,16
DC_2	0,73	T2M_RANGE_5	0,16
DQ_2	0,73	T2M_MIN_3	0,16
DS_panual_7	0,70	ALLSKY_SFC_UVB_7	0,16
S6	0,70	WS2M_4	0,16
chuva_seca_5	0,68	WS10M_1	0,15
DQ_4	0,68	prec_6	0,15
DSC_2	0,67	GWETPROF_7	0,15
DQ_1	0,66	GWETTOP_6	0,15
DS_panual_6	0,65	TQV_6	0,15
DC_1	0,65	TQV_5	0,15
P	0,64	PS_6	0,15
TOA_SW_DWN_5	0,63	ALLSKY_SFC_SW_DWN_7	0,15
TOA_SW_DWN_0	0,57	prec_ac_7	0,15
chuva_seca_6	0,56	GWETTOP_5	0,15
chuva_seca_4	0,54	T2M_2	0,14

Variável	%	Variável	%
TOA_SW_DWN_6	0,53	RH2M_4	0,14
T_otima_6	0,52	ALLSKY_SFC_UVA_7	0,14
DC_0	0,50	WS2M_5	0,14
T_otima_0	0,49	ALLSKY_KT_6	0,14
TOA_SW_DWN_4	0,49	ALLSKY_KT_1	0,14
DF_7	0,47	RH2M_6	0,14
ALLSKY_SFC_LW_DWN_7	0,47	T2MWET_3	0,14
GWETROOT_5	0,47	DC_6	0,14
prec_2	0,47	T2MDEW_5	0,14
ALLSKY_SFC_LW_DWN_6	0,47	T2M_0	0,14
DQ_5	0,47	T2M_RANGE_4	0,13
QV2M_2	0,46	ALLSKY_SFC_UVA_0	0,13
TOA_SW_DWN_2	0,45	TQV_2	0,13
ALLSKY_SFC_LW_DWN_0	0,45	ALLSKY_SFC_UVB_0	0,13
GWETROOT_2	0,44	QV2M_5	0,13
DU_panual_7	0,44	ALLSKY_KT_4	0,13
GWETROOT_6	0,43	ALLSKY_SFC_UV_INDEX_7	0,13
WS2M_0	0,42	T_otima_4	0,13
CLRSKY_SFC_PAR_TOT_2	0,41	DU_panual_2	0,13
ALLSKY_SFC_SW_DWN_2	0,41	QV2M_7	0,13

Variável	%	Variável	%
GWETROOT_4	0,41	WS2M_3	0,13
DSC_0	0,41	TQV_4	0,13
TOA_SW_DWN_1	0,41	GWETROOT_0	0,13
DC_3	0,41	T2M_RANGE_7	0,12
T2MDEW_2	0,40	T2MWET_5	0,12
DSC_1	0,40	GWETTOP_7	0,12
TOA_SW_DWN_3	0,39	ALLSKY_SFC_UVA_4	0,12
GWETPROF_6	0,39	T2M_MIN_5	0,12
chuva_seca_7	0,39	CLRSKY_SFC_PAR_TOT_6	0,12
ALLSKY_KT_3	0,39	ALLSKY_SFC_UVA_6	0,12
DQ_6	0,37	T2M_MAX_2	0,12
QV2M_1	0,37	WS10M_7	0,12
ALLSKY_SFC_LW_DWN_3	0,37	ALLSKY_SFC_UV_INDEX_3	0,12
chuva_seca_3	0,37	ALLSKY_SFC_UVA_1	0,12
GWETROOT_3	0,37	ALLSKY_SFC_UV_INDEX_4	0,12
GWETPROF_2	0,36	prec_ac_5	0,12
RH2M_0	0,36	ALLSKY_SFC_UVB_1	0,12
DSC_3	0,36	TS_5	0,12
ALLSKY_KT_2	0,36	T2M_RANGE_6	0,12
prec_ac_1	0,35	QV2M_3	0,12

Variável	%	Variável	%
prec_0	0,35	T2M_MIN_4	0,11
GWETROOT_7	0,35	T2MWET_6	0,11
GWETROOT_1	0,35	GWETTOP_3	0,11
DC_4	0,34	ALLSKY_KT_7	0,11
ALLSKY_SFC_LW_DWN_1	0,34	TQV_1	0,11
GWETTOP_2	0,33	GWETTOP_1	0,11
CLRSKY_SFC_SW_DWN_1	0,32	T2M_6	0,11
prec_ac_0	0,32	GWETTOP_0	0,11
ALLSKY_SFC_UVB_2	0,32	ALLSKY_SFC_UVB_4	0,11
RH2M_1	0,31	T2MWET_2	0,11
GWETPROF_1	0,31	ALLSKY_SFC_UVB_3	0,11
ALLSKY_SFC_PAR_TOT_2	0,31	T2M_MIN_2	0,10
GWETPROF_4	0,30	DF_6	0,10
DF_1	0,30	ALLSKY_SFC_SW_DWN_6	0,10
CLRSKY_SFC_SW_DWN_3	0,30	ALLSKY_SFC_PAR_TOT_6	0,10
T2M_MIN_0	0,30	DU_panual_1	0,10
CLRSKY_SFC_PAR_TOT_1	0,29	T2M_4	0,10
chuva_seca_0	0,29	ALLSKY_SFC_PAR_TOT_5	0,10
DQ_0	0,28	CLRSKY_SFC_SW_DWN_5	0,10
prec_ac_2	0,28	DU_panual_3	0,10

Variável	%	Variável	%
CLRSKY_SFC_PAR_TOT_4	0,28	DC_7	0,10
T2MDEW_0	0,28	TS_3	0,09
RH2M_2	0,28	T2M_3	0,09
TQV_3	0,27	T2M_MIN_7	0,09
TS_4	0,27	ALLSKY_SFC_PAR_TOT_7	0,09
CLRSKY_SFC_PAR_TOT_3	0,27	ALLSKY_SFC_PAR_TOT_1	0,09
GWETTOP_4	0,27	ALLSKY_SFC_LW_DWN_2	0,09
PS_7	0,27	ALLSKY_SFC_SW_DWN_4	0,09
ALLSKY_SFC_UVA_3	0,27	QV2M_4	0,09
T2M_MIN_6	0,27	WS10M_6	0,09
prec_ac_6	0,26	DSC_4	0,09
CLRSKY_SFC_PAR_TOT_0	0,26	T2MWET_4	0,08
PS_3	0,26	T_otima_2	0,08
T_otima_5	0,26	ALLSKY_KT_0	0,08
T2M_RANGE_0	0,26	ALLSKY_SFC_UV_INDEX_6	0,08
WS10M_2	0,26	TQV_7	0,08
TS_0	0,26	T2M_MAX_3	0,08
DF_3	0,26	T2MDEW_7	0,08
prec_1	0,26	DF_4	0,08
DQ_3	0,25	CLRSKY_SFC_SW_DWN_4	0,08

Variável	%	Variável	%
DU_panual_0	0,25	T2M_MAX_1	0,08
WS2M_1	0,25	ALLSKY_KT_5	0,08
ALLSKY_SFC_UV_INDEX_1	0,25	RH2M_3	0,08
DU_panual_4	0,25	T2MDEW_4	0,07
PS_5	0,25	TS_6	0,07
T_otima_3	0,24	ALLSKY_SFC_SW_DWN_0	0,07
GWETPROF_5	0,24	T2MWET_0	0,07
DQ_7	0,24	T2MWET_1	0,07
WS2M_6	0,23	TS_2	0,07
ALLSKY_SFC_UVA_2	0,23	TS_7	0,07
GWETPROF_3	0,23	DSC_5	0,07
ALLSKY_SFC_UVB_6	0,23	TQV_0	0,07
T2M_RANGE_3	0,23	T2M_MAX_6	0,07
ALLSKY_SFC_PAR_TOT_3	0,23	T2M_5	0,06
prec_4	0,22	T2M_MIN_1	0,06
prec_7	0,22	ALLSKY_SFC_SW_DWN_5	0,06
CLRSKY_SFC_PAR_TOT_7	0,22	T2MWET_7	0,06
CLRSKY_SFC_PAR_TOT_5	0,22	DSC_7	0,06
T2M_RANGE_1	0,21	ALLSKY_SFC_UV_INDEX_5	0,06
PS_1	0,21	ALLSKY_SFC_UV_INDEX_0	0,06

Variável	%	Variável	%
CLRSKY_SFC_SW_DWN_7	0,21	ALLSKY_SFC_UVB_5	0,06
PS_4	0,21	DU_panual_5	0,06
WS2M_2	0,21	ALLSKY_SFC_LW_DWN_5	0,06
CLRSKY_SFC_SW_DWN_6	0,21	DSC_6	0,06
T2MDEW_1	0,21	WS10M_5	0,06
QV2M_0	0,21	ALLSKY_SFC_PAR_TOT_4	0,05
ALLSKY_SFC_SW_DWN_3	0,21	T2M_1	0,05
T2M_MAX_0	0,20	ALLSKY_SFC_UVA_5	0,05
DF_2	0,20	TS_1	0,04
prec_ac_4	0,20	ALLSKY_SFC_SW_DWN_1	0,04
QV2M_6	0,20	T2M_MAX_5	0,04
WS2M_7	0,19	T2MDEW_6	0,03
DC_5	0,19	T2M_7	0,03
DF_0	0,19	T2M_MAX_7	0,02
T2MDEW_3	0,19	prec_3	0,01