

RAFAEL MARTINS DE CARVALHO ARAUJO

APRENDIZADO DE MÁQUINA PARA MODELAGEM DO
CRESCIMENTO E PRODUÇÃO EM POVOAMENTOS DE *EUCALYPTUS*
SPP

Dissertação apresentada ao Programa de Pós-Graduação em
Agricultura e Informações Geoespaciais da Universidade
Federal de Uberlândia, Campus Monte Carmelo, como parte
das exigências para obtenção do título de “Mestre”.

Orientador

Prof. Dr. Murillo Guimarães Carneiro

Coorientador

Prof. Dr. Alvaro Augusto Vieira Soares

MONTE CARMELO
MINAS GERAIS - BRASIL

2023

RAFAEL MARTINS DE CARVALHO ARAUJO

APRENDIZADO DE MÁQUINA PARA MODELAGEM DO
CRESCIMENTO E PRODUÇÃO EM POVOAMENTOS DE *EUCALYPTUS*
SPP

Dissertação apresentada ao Programa de Pós-Graduação em
Agricultura e Informações Geoespaciais da Universidade
Federal de Uberlândia, Campus Monte Carmelo, como parte
das exigências para obtenção do título de “Mestre”.

Banca examinadora

Prof. Dr. Luiz Gustavo Almeida Martins – Faculdade de Computação (UFU)

Prof. Dr. Diogo Nepomuceno Cosenza – Instituto de Ciências Agrárias (UFV)

Prof. Dr. Murillo Guimarães Carneiro

Faculdade de Computação (UFU)

(Orientador)

MONTE CARMELO

MINAS GERAIS - BRASIL

2023

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

A663 Araujo, Rafael Martins de Carvalho, 1997-
2023 Aprendizado de Máquina para Modelagem do Crescimento e
Produção em Povoamentos de Eucalyptus spp [recurso
eletrônico] / Rafael Martins de Carvalho Araujo. - 2023.

Orientador: Murillo Guimarães Carneiro.
Coorientador: Alvaro Augusto Vieira Soares.
Dissertação (Mestrado) - Universidade Federal de
Uberlândia, Pós-graduação em Agricultura e Informações
Geoespaciais.

Modo de acesso: Internet.

Disponível em: <http://doi.org/10.14393/ufu.di.2023.393>

Inclui bibliografia.

Inclui ilustrações.

1. Agronomia. I. Carneiro, Murillo Guimarães ,1988-,
(Orient.). II. Soares, Alvaro Augusto Vieira ,1986-,
(Coorient.). III. Universidade Federal de Uberlândia.
Pós-graduação em Agricultura e Informações Geoespaciais.
IV. Título.

CDU: 631

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:
Gizele Cristine Nunes do Couto - CRB6/2091
Nelson Marcos Ferreira - CRB6/3074



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
 Coordenação do Programa de Pós-Graduação em Agricultura e Informações
 Geoespaciais
 Rodovia LMG 746, Km 01, s/nº, Bloco 1AMC, Sala 1A202, Monte Carmelo-MG, CEP 38.500-000
 Telefone: (34) 3810-1033 - ppgaig@iciag.ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Agricultura e Informações Geoespaciais				
Defesa de:	Dissertação de Mestrado Acadêmico				
Data:	04/07/2023	Hora de início:	8h30	Hora de encerramento:	11h55
Matrícula do Discente:	32012AIG012				
Nome do Discente:	Rafael Martins de Carvalho Araujo				
Título do Trabalho:	Aprendizado de Máquina para Modelagem do Crescimento e Produção em Povoamentos de <i>Eucalyptus spp</i>				
Área de concentração:	Informações geoespaciais e tecnologias aplicadas à produção agrícola				
Linha de pesquisa:	Aplicações e desenvolvimento de métodos em informações geoespaciais				
Projeto de Pesquisa de vinculação:	Inteligência artificial aplicada à agricultura e informações geoespaciais				

Reuniu-se na sala virtual a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Agricultura e Informações Geoespaciais, assim composta: Professores Doutores: Diogo Nepomuceno Cosenza (Universidade Federal de Viçosa); Luiz Gustavo Almeida Martins (Universidade Federal de Uberlândia) e Murillo Guimarães Carneiro (Universidade Federal de Uberlândia), orientador do candidato. Iniciando os trabalhos o presidente da mesa, Dr. Murillo Guimarães Carneiro, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovado

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Murillo Guimarães Carneiro, Professor(a) do Magistério Superior**, em 04/07/2023, às 15:28, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Luiz Gustavo Almeida Martins, Professor(a) do Magistério Superior**, em 04/07/2023, às 15:55, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Diogo Nepomuceno Cosenza, Usuário Externo**, em 14/07/2023, às 16:41, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4624876** e o código CRC **E13E3F9A**.

*À minha irmã Gabriela
Aos meus pais Jair e Walda*

AGRADECIMENTOS

Agradeço, primeiramente, a Deus por estar aqui e poder desfrutar de todo o processo; ao meu orientador, prof. Dr. Murillo Guimarães Carneiro, pelo incentivo e confiança no desenvolvimento desse projeto, mesmo havendo incertezas e dificuldades ao longo do processo; e a todos os demais colaboradores, amigos e professores da instituição, em especial, ao meu coorientador prof. Dr. Álvaro Augusto Vieira pelo conhecimento repassado e à companhia Celulose Nipo-Brasileira S.A (CENIBRA) por ceder os dados necessários a pesquisa.

BIOGRAFIA

Rafael Martins de Carvalho Araujo nasceu em Monte Carmelo, Minas Gerais, Brasil, 1997. Graduado em Engenharia de Agrimensura e Cartográfica pela Universidade Federal de Uberlândia (UFU), Minas Gerais, Brasil, 2019. Seus interesses de pesquisa incluem o desenvolvimento de algoritmos para a estimativa de produção a partir de variáveis agrícolas e de sensoriamento remoto.

SUMÁRIO

RESUMO	i
ABSTRACT	ii
1. INTRODUÇÃO	13
2. OBJETIVOS	14
3. TRABALHOS RELACIONADOS	14
4. MATERIAL E DESCRIÇÃO DOS MODELOS	16
4.1 Conjunto de dados	16
4.2 Modelo de Clutter	16
4.3 Modelos de aprendizado de máquina	18
4.3.1 Redes Neurais Artificiais	18
4.3.2 Máquina de Vetor de Suporte	20
4.3.3 Floresta aleatória	21
4.3.4 K-Vizinhos mais Próximos	21
4.4 Dados de sensoriamento remoto	22
4.4.1 Índice de Vegetação por Diferença Normalizada	22
4.4.2 Razão Simples	23
4.4.3 Índice de Vegetação Aprimorado	23
4.4.4 Índice de Resistência Atmosférica Visível	23
4.4.5 Índice de Vegetação de Diferença Normalizada Verde.....	24
4.4.6 Índice de Vegetação Ajustado do Solo Transformado	24
5 METODOLOGIA DE PROCESSAMENTO E ANÁLISE	25
5.1 Modelagem com variáveis tradicionais	25
5.2 Obtenção de variáveis de sensoriamento remoto	26
5.3 Medidas de desempenho	27
5.4 Alternativas de análises de composição dos modelos	27
6 RESULTADOS	29
6.1 Análise 1: estimativa do volume presente	29

6.2 Análise 2: estimativa do volume presente com adição de dados de sensoriamento remoto.....	29
6.3 Análise 3: estimativa do volume futuro	30
6.4 Análise 4: estimativa do volume futuro com adição de dados de cadastro do inventário.....	31
7 CONCLUSÕES	32
REFERÊNCIAS	33

ARAÚJO, Rafael Martins de Carvalho. Aprendizado de Máquina para Modelagem do Crescimento e Produção em Povoamentos de *Eucalyptus spp.* 2023, 38 p. Dissertação (Mestrado em Agricultura e Informações Geoespaciais) - Universidade Federal de Uberlândia, Campus Monte Carmelo, Minas Gerais, Brasil. Disponível em: <http://doi.org/10.14393/ufu.di.2023.393>

RESUMO

O setor florestal, especialmente o de florestas plantadas, é um dos mais importantes pilares do agronegócio brasileiro. As florestas plantadas brasileiras se destacam pela sua alta produtividade. No planejamento da produção madeireira, é imprescindível a capacidade de se estimar a produção, tanto no presente quanto, e sobretudo, no futuro. Estudos no ramo da biometria desenvolveram importantes abordagens e modelos que estimam de forma acurada a produção, os quais são baseados em técnicas de regressão linear e não linear. Atualmente, algoritmos de aprendizado de máquinas têm sido testados e os resultados obtidos se mostram muito promissores. Dessa forma, o presente trabalho teve como objetivo avaliar algoritmos de aprendizado de máquina para estimar a produção presente e futura de plantios de eucalipto. Foram avaliados os seguintes algoritmos de aprendizado de máquina: Redes Neurais Artificiais (ANN), Floresta aleatória (RF), Máquina de Vetor de Suporte (SVM), e K-Vizinhos mais Próximos (KNN). Esses métodos foram comparados ao modelo de Clutter, um sistema de equações lineares tradicionalmente empregado pela indústria florestal na prognose da produção. Os modelos foram ajustados em quatro análises florestais, medidas em 1.280 parcelas permanentes entre os anos de 2013 e 2019, no estado de Minas Gerais. Os modelos de aprendizado de máquina foram ajustados usando dados biométricos de campo e derivadas de sensoriamento remoto, como os índices de vegetação. Em geral, os modelos de aprendizagem de máquina tiveram erros de estimativa menores que os do modelo de Clutter, destacando-se o método SVM. Além disso, o uso de índices de vegetação no modelo SVM melhorou a acurácia das estimativas de produção corrente. O uso de variáveis tradicionais em conjunto com as de cadastro do inventário florestal no modelo SVM melhorou a acurácia das estimativas de produção futura. Dessa forma, as análises mostraram que os modelos de aprendizado de máquina são bons estimadores de produção quando associados a dados biométricos de campo e derivados de sensoriamento remoto.

Palavras-chave: Manejo Florestal; Produção madeireira; Aprendizado de Máquina; Índices de Vegetação

ARAUJO, Rafael Martins de Carvalho. Machine Learning for Modeling Growth and Production in *Eucalyptus spp.* Stands. 2023, 38 p. Dissertation (Master's in Agriculture and Geospatial Information) - Federal University of Uberlândia, Monte Carmelo Campus, Minas Gerais, Brazil. Available at: <http://doi.org/10.14393/ufu.di.2023.393>

ABSTRACT

The forestry sector, especially that of planted forests, is one of the most crucial pillars of Brazilian agribusiness. Brazilian planted forests stand out for their high productivity. In the planning of timber production, it is essential to have the ability to estimate production both in the present and, particularly, in the future. Studies in the field of biometrics have developed important approaches and models that accurately estimate production, especially based on linear and non-linear regression techniques. Currently, machine learning algorithms have been tested, and the results have proven to be very promising. Thus, this study aimed to evaluate machine learning algorithms for estimating present and future production in eucalyptus plantations. The following machine learning algorithms were assessed: Artificial Neural Networks (ANN), Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). These methods were compared to the Clutter model, a system of linear equations traditionally used by the forestry industry for production prognosis. The models were fitted in four forest analyses, measured in 1280 permanent plots between 2013 and 2019 in the state of Minas Gerais. Machine learning models were adjusted using field biometric data and derivatives of remote sensing, such as vegetation indices. In general, machine learning models had lower estimation errors than the Clutter model, with SVM standing out. Additionally, the use of vegetation indices in the SVM model improved the accuracy of current production estimates. The use of traditional variables together with forest inventory registration variables in the SVM model enhanced the accuracy of future production estimates. Thus, the analyses demonstrated that machine learning models are good production estimators when associated with field biometric data and remote sensing derivatives.

Key-words: Forest Management; Timber Production; Machine Learning; Vegetation Indices

1. INTRODUÇÃO

O setor florestal é um dos principais pilares do agronegócio brasileiro e é destaque na produção florestal mundial. O Brasil foi o maior produtor mundial de carvão vegetal em 2020 e o principal exportador de celulose no ano de 2021 (Iba, 2022). As florestas plantadas no Brasil ocupam atualmente 9,93 milhões de hectares, sendo que 75,8% delas é composta por eucalipto; ao passo que o restante delas fica entre plantios de pinheiros (19,4%) e outras espécies (4,8%), como seringueira, acácia, teca e paricá. Os principais estados produtores de eucalipto em área plantada são Minas Gerais (30%), Mato Grosso do Sul (14%), São Paulo (13%), Bahia (8%), Rio Grande do Sul (8%) e Paraná (6%), que juntos atingiram uma produtividade média de 38,9 m³ ha⁻¹ ano⁻¹ em 2022 (Iba, 2022).

O progresso notável na produção florestal brasileira tem como pano de fundo o avanço nas pesquisas em práticas silviculturais e de melhoramento genético (Gonçalves *et al.* 2013). Não menos importante, avanços na mensuração e no monitoramento florestal, bem como na prognose da produção, permitiram o refinamento das tomadas de decisão no planejamento florestal.

O uso de tecnologias como o sensoriamento remoto e, mais recentemente, a inteligência artificial, surgem como importantes suportes da mensuração florestal, tanto para a determinação de estoque atual como para a estimativa da produção futura. Modelos robustos utilizam como preditores variáveis derivadas de medições do procedimento de inventário florestal, com possibilidade de inclusão de outras variáveis, como variáveis edafoclimáticas, variáveis derivadas de sensoriamento remoto e variáveis de cadastro das unidades de manejo. Diversas pesquisas têm sido conduzidas de forma a avaliar diferentes abordagens de modelagem na tentativa de se obter estimativas mais acuradas da produção. Exemplos de pesquisas são: Penido (2020), com a avaliação de três modelos tradicionais, o Modelo de Povoamento (MP) Clutter, o Modelo de Distribuição Diamétrica (MDD) e o Modelo de Árvore Individual (MAI); Dai (2021), com modelos que incluem índices de vegetação (IV) derivados de sensoriamento remoto; Cordeiro (2022), com o uso de Redes Neurais Artificiais (ANN) e da Máquina de Vetor de Suporte (SVM); entre outros dispostos na literatura, todos com o mesmo objetivo, isto é, estimar a produção de eucalipto com precisão.

2. OBJETIVOS

Diante do potencial dessas aplicações, o presente estudo tem por objetivo desenvolver modelos de aprendizado de máquina capazes de estimar a produção de eucalipto com alta precisão, no que se refere a dados de um inventário florestal contínuo de *Eucalyptus Spp.*

Para tal, são apresentados os seguintes objetivos específicos:

- Avaliar a acurácia dos modelos de aprendizado de máquina em relação aos modelos de Clutter para estimar a produção presente e futura de povoamentos de eucalipto;
- Avaliar o ganho de acurácia dos modelos preditivos com a inclusão de variáveis de cadastro e derivadas de sensoriamento remoto como preditoras do modelo.

3. TRABALHOS RELACIONADOS

Nos últimos 20 anos diversas pesquisas foram desenvolvidas com base em métodos de aprendizado de máquina para aprimorar a modelagem biométrica das florestas. Exemplos de uso desses modelos incluem: a estimativa de variáveis dendrométricas diversas (Binoti, 2012); a estimativa de produção (Gorgens, 2009); a probabilidade de sobrevivência de uma árvore (Reis, 2018); o quanto inserção de variáveis fora do inventário melhoram as previsões dos modelos (Freitas, 2020).

Essa gama de possibilidades faz com que o número de estudos seja cada vez maior, sendo possível dividi-los em duas categorias: aqueles voltados apenas para a estimativa de uma variável categórica em um modelo de produção, como a altura, e aqueles que buscam apenas aperfeiçoar o processo de estimação de produção (volume) por meio de outras abordagens.

Um exemplo da primeira categoria é o estudo de Binoti (2012), que procurou a estimativa da altura das árvores por meio de ANN, para dados de inventários florestais contínuos de povoamentos clonais de eucalipto, como diâmetro com casca, altura dominante e idade, localizados no Sul da Bahia. Como resultado, o autor concluiu que a ANN permitiu a diminuição da quantidade de dados coletados para a calibragem dos modelos e, conseqüentemente, reduziu o custo de inventário, isso sem que houvesse perda na precisão dos modelos.

Um outro exemplo é o estudo de Were (2015), que comparou a performance dos modelos Floresta aleatória (RF), Regressão por Vetores de Suporte (SVR) e ANN para mapear

a variabilidade de estoques de carbono em solo orgânico (SOC, do inglês *Soil Organic Carbon*) na reserva florestal de Mau, no Leste do Quênia. O estudo mostrou que os modelos superestimaram os estoques de SOC com um desempenho semelhante entre o SVR e a ANN.

Reis (2018) propôs estimar a probabilidade de sobrevivência e de mortalidade de árvores individuais em uma floresta com colheita seletiva usando ANN para apoiar decisões silviculturais na gestão da Floresta Nacional de Tapajós, oeste do estado do Pará. O estudo mostrou que a ANN estima a sobrevivência individual das árvores com mais precisão, mas isso não ocorre com a mortalidade, que é um evento mais raro do que a sobrevivência.

Já para a segunda categoria, mais comum às empresas florestais e ao objetivo deste estudo, pode-se mencionar alguns outros exemplos, como o de Alcantara (2012), que buscou modificações em modelos tradicionais, como o de Clutter (1963), com o objetivo de modelar o crescimento em nível de povoamento de clones de eucalipto na região de Minas Gerais. Nesse estudo, foram utilizados dados de inventários florestais contínuos, como índice de local, altura dominante e idade, para a estimativa da área basal e do volume mediante cinco equações derivadas do modelo original de Clutter. Os resultados indicaram a eficiência das equações, concluindo que o índice de local, na equação de projeção da área basal, pode ser substituído pela altura dominante na mesma idade sem perda de eficiência e de exatidão.

Gorgens *et al.* (2009) procuraram modelos mais robustos, como a ANN, para estimar com eficiência o volume em nível de povoamento de clones de eucalipto. No seu estudo, eles utilizaram dados de inventário florestal de quatro empresas brasileiras, como diâmetro na altura do peito (DAP), altura total (HT), cinco locais e duas espécies, ambos representados por variáveis binárias, na qual testaram 5 tipos de normalização e de equalização de dados e 4 tipos de arquiteturas para a rede. Eles concluíram que as redes que apresentaram resultados estaticamente equivalentes foram construídas com duas camadas intermediárias, com função de ativação exponencial; as redes com melhores resultados apresentaram dados normalizados e equalizados pelas variações do método de interpolação linear. Em síntese, os autores viram que a modelagem por redes neurais foi viável.

Finalizando essa categoria, temos o estudo de Freitas *et al.* (2020), que buscaram utilizar uma ANN para estimar a produtividade de eucalipto no estado de Minas Gerais. Eles utilizaram dados de 507 povoamentos de inventários florestais contínuos, como o tipo de clone, espaçamento, genótipo, temperatura e outros, como práticas silviculturas. Para modelar as redes, eles utilizaram a ANN com neurônios, variando nas camadas ocultas, de acordo com cada configuração, algoritmos de treinamento (*Backpropagation*, *Resilient Propagation (Rprop +)* e *Quick Propagation*) e funções de ativação (logística, tangente hiperbólica e linear), totalizando 87.500 redes. Ademais, para a avaliação do erro, empregaram a raiz quadrada do erro-médio

(RMSE). Os autores concluíram que as 10 melhores redes tiveram correlação acima de 85% entre os dados observados e estimados e RMSE menor que 15% para estimar a produtividade. Notaram, também, que as variáveis que tiveram maior importância na estimativa da produtividade foram genótipo, espaçamento, argila, matéria orgânica, capacidade de troca catiônica (CTC), chuva, temperatura, déficit de água e performance de fertilização durante a rotação.

Esses estudos mostram a evolução das pesquisas sobre modelagem de dados florestais no Brasil, onde modelos de aprendizado de máquina e tradicionais são frequentemente utilizados para as estimativas de variáveis relacionadas ao processo de crescimento e de produção do eucalipto. Entretanto, existe um número menor de pesquisas com um maior aprofundamento em investigar um número maior de modelos de aprendizado de máquina para a modelagem da produção, utilizando como variáveis preditoras as de cadastro do inventário florestal e derivadas de sensoriamento remoto em conjunto com as já usadas nos modelos tradicionais, o que abre novos caminhos a serem analisados e testados com vistas ao melhoramento da acurácia no processo de estimativa da produção.

4. MATERIAL E DESCRIÇÃO DOS MODELOS

4.1 Conjunto de Dados

A base de dados utilizada no estudo foi concedida pela Celulose Nipo-Brasileira S/A (Cenibra) e é referente à 1.280 unidades amostrais (parcelas) distribuídas em 507 talhões de *Eucalyptus spp* localizados na região do Vale do Rio doce, Leste do estado de Minas Gerais. Os dados foram coletados em inventários florestais contínuos, com medições anuais entre 2013 e 2019, e abrangem idades de aproximadamente 1,5 a 10 anos.

4.2 Modelo de Clutter

Os modelos mais utilizados de crescimento e produção em povoamentos florestais são aqueles de densidade variável (Campos; Leite, 2017). Um dos modelos mais usuais é o de Clutter (1963).

Desenvolvido no ano de 1963, o modelo de Clutter é composto por um sistema de equações (1). Uma equação tem como variável a resposta a área basal do povoamento e a outra equação cuja variável resposta é o volume. A equação de área basal permite a projeção da área

basal para uma idade de interesse que, por sua vez, é utilizada na equação de volume para a estimação do volume nessa idade.

$$\begin{cases} LnV_2 = \beta_0 + \beta_1 \left(\frac{1}{I_2}\right) + \beta_2 S + \beta_3 LnB_2 + \varepsilon \\ LnB_2 = LnB_1 \left(\frac{I_1}{I_2}\right) + \alpha_0 \left(1 - \frac{I_1}{I_2}\right) + \alpha_1 \left(1 - \frac{I_1}{I_2}\right) S + \varepsilon \end{cases} \quad (1)$$

em que: I_1 é a idade presente, em meses; I_2 , é a idade futura, em meses; B_1 , é a área basal na idade I_1 , em $m^2 ha^{-1}$; B_2 , é a área basal na idade I_2 , em $m^2 ha^{-1}$; S , é o índice de local; V_2 , é o volume na idade I_2 em $m^3 ha^{-1}$; α_i e β_i , são os parâmetros do modelo; e ε é o erro aleatório, em que $\varepsilon \sim N(0, \sigma^2)$.

O índice de local é uma métrica indicadora da capacidade produtiva e é dado pela altura dominante, que consiste na média aritmética da altura das 100 árvores mais grossas uma idade de referência (Assman, 1970). Essa idade é comumente estabelecida como seis ou sete anos para plantios de eucalipto (Campos; Leite, 2017). Para a estimação do índice de local, normalmente, utiliza-se o método da curva-guia, que consiste no ajuste e na manipulação algébrica da equação (2), gerando a equação (3) (Campos; Leite, 2017).

$$LnHd = \beta_0 + \beta_1 \left(\frac{1}{I}\right) + \varepsilon \quad (2)$$

$$LnS = LnHd - \beta_1 \left(\frac{1}{I} - \frac{1}{I_{ref}}\right) \quad (3)$$

em que: S é o índice de local em metros; Hd é a altura dominante em metros, I é a idade em meses; I_{ref} a idade de referência, geralmente adotado o valor de 72 meses e β_i , são os parâmetros do modelo; e ε é o erro aleatório, em que $\varepsilon \sim N(0, \sigma^2)$.

Uma das formas de ajuste do modelo de Clutter é a substituição do termo lnB_2 da equação de volume pela equação de área basal. Assim, essa nova equação (4) é ajustada pelo método dos mínimos quadrados ordinários (Campos; Leite, 2017). Após o ajuste, para que se possa desmembrar a equação (4) e atingir o sistema em sua forma original, os parâmetros da equação de área basal são obtidos por $\alpha_0 = \beta_4/\beta_3$ e $\alpha_1 = \beta_5/\beta_3$.

$$LnV_2 = \beta_0 + \beta_1 \left(\frac{1}{I_2}\right) + \beta_2 S + \beta_3 LnB_1 \left(\frac{I_1}{I_2}\right) + \beta_4 \left(1 - \frac{I_1}{I_2}\right) + \beta_5 \left(1 - \frac{I_1}{I_2}\right) S + \varepsilon \quad (4)$$

em que: I_1 é a idade presente, em meses; I_2 , é a idade futura, em meses; B_1 , é a área basal na idade I_1 , em $m^2 ha^{-1}$; S , é o índice de local na idade I_1 ; V_2 , é o volume na idade I_2 em $m^3 ha^{-1}$; β_i , são os parâmetros do modelo; e ε é o erro aleatório, em que $\varepsilon \sim N(0, \sigma^2)$.

4.3 Modelos de Aprendizado de Máquina

Segundo Brink *et al.* (2017), o Aprendizado de Máquina são modelos capazes de reproduzir padrões de dados de forma automatizada a partir de um conjunto de dados previamente apresentado. Esses sistemas seguem formulações algorítmicas que, em geral, têm a capacidade de modelar dados complexos por meio de aproximação de funções, podendo aceitar dados preditores variados de entrada e sem fazer, necessariamente, pré-suposições sobre a distribuição deles (Maxwell *et al.*, 2018). Alguns algoritmos são mais utilizados, como ANN, SVM, Floresta aleatória (RF) e K-Vizinhos mais Próximos (KNN), devido à sua alta capacidade de agregar muitas variáveis de entrada e, como saída, retornar uma resposta de forma precisa e pontual (Rex, 2019).

4.3.1 Redes Neurais Artificiais (ANN)

As redes neurais artificiais (ANN, do inglês *Artificial Neural Network*) segundo Haykin (2001), em sua forma geral, são como uma máquina projetada para modelar a forma como o cérebro humano realiza uma tarefa, utilizando-se de uma rede composta por células computacionais denominadas neurônios ou unidades de processamento, que, em um processo de aprendizagem, tem a função de armazenar conhecimento experimental e torná-lo disponível para o uso.

O neurônio é uma unidade de processamento primordial para a operação de uma rede neural. Como ilustrado pela Figura 1, ele é composto basicamente por cinco elementos: as sinapses ou elos de conexão, caracterizados por um peso ou força própria; um somador, para somar os sinais de entrada; uma função de ativação, para restringir a amplitude da saída de um neurônio; o bias, que tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação e a saída, que pode ser a entrada para outro neurônio ou para o resultado final (Haykin, 2001).

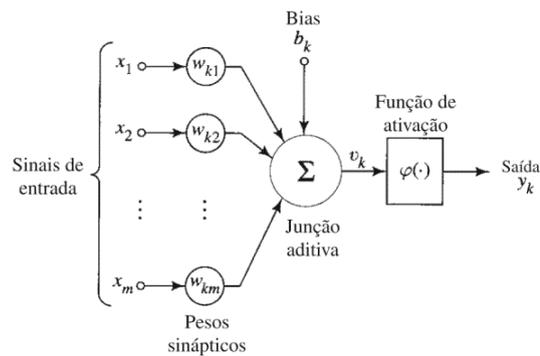


FIGURA 1 – Representação de um neurônio

Fonte: Haykin (2001).

Uma importante classe das redes neurais conhecida como Perceptron multicamadas (MLP, do inglês *Multilayer Perceptron*), tomou destaque pela sua arquitetura com uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída, com o sinal se propagando para frente, camada por camada, como na Figura 2.

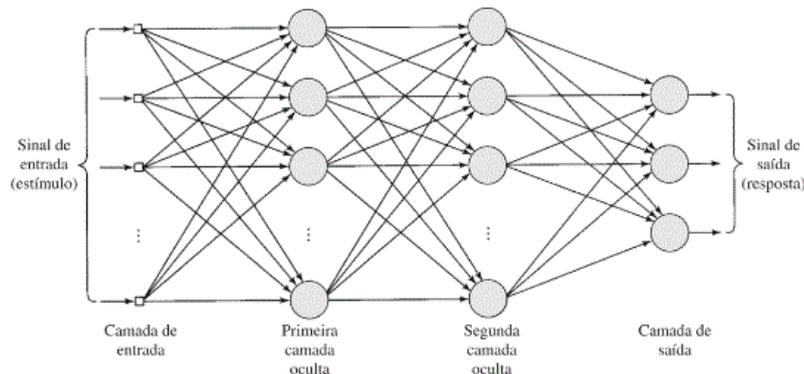


FIGURA 2 – Representação de uma MLP

Fonte: Haykin (2001).

Ela tem sido frequentemente aplicada na resolução de problemas difíceis com a ajuda do treinamento supervisionado e do algoritmo de retropropagação de erro (*error backpropagation*). O aprendizado por retropropagação é definido em dois passos: o primeiro, para frente, no qual o vetor de entrada é aplicado aos nós da rede e se propaga para frente, camada por camada, por meio de sinais funcionais, com os pesos sinápticos fixos, produzindo uma resposta real da rede. Durante a volta, no sentido para trás, os pesos sinápticos são ajustados e o resultado real é subtraído do desejado, produzindo sinais de erro que são trazidos e

corrigindo os pesos sinápticos, de modo que a resposta da rede se mova mais perto da resposta desejada, como na Figura 3.

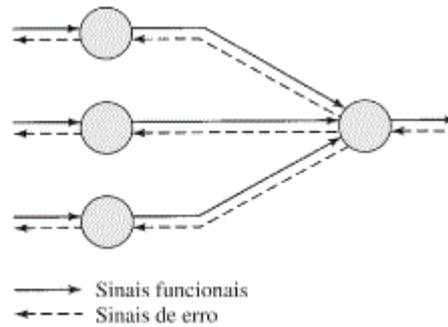


FIGURA 3 – Representação do princípio de retropropagação

Fonte: Haykin (2001).

4.3.2 Máquina de Vetor de Suporte (SVM)

A máquina de vetor de suporte (SVM, do inglês *Support Vector Machine*) é um método semelhante à regressão linear, no qual a equação da linha é denominada hiperplano. Os pontos dos lados próximos ao plano são chamados de vetores de suporte (SVR, do inglês *Support Vector Regressor*), usados para definir o plano (Figura 4). A ideia por trás da SVR é tentar ajustar a melhor linha dentro de um valor limiar (distância entre o hiperplano e a linha de fronteira). Para uma regressão não linear, a solução é mapear os dados para um espaço de maior dimensão, no qual uma função *Kernel* transforma e realiza essa separação linear (Carneiro, 2017).

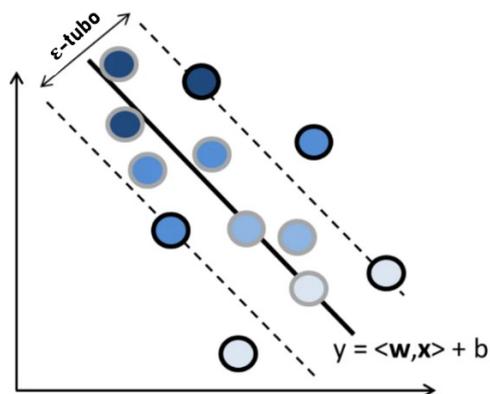


FIGURA 4 – Exemplo de uma SVR

Fonte: Adaptado de Rodriguez (2022).

4.3.3 Floresta aleatória (RF)

A Floresta aleatória (RF, do inglês *Random Forest*) ajusta várias árvores de decisão usando amostras dos dados de treinamento e das variáveis preditoras (Figura 5). O valor final estimado pelo modelo é a média das estimativas de cada árvore de regressão individual, (Galiano et al, 2015). Dessa forma, há um aumento na precisão da previsão devido à robustez em enfrentar pequenas variações nos dados de entrada (Breiman, 2001).

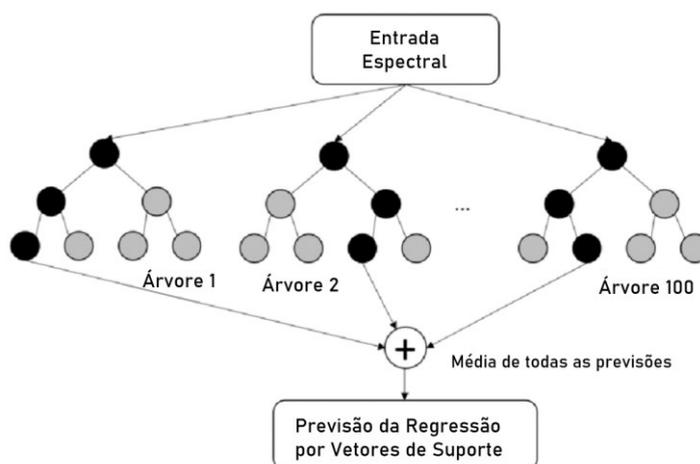


FIGURA 5 – Exemplo de uma floresta aleatória

Fonte: Adaptado de Nishikawa (2023).

4.3.4 K-Vizinhos mais Próximos (KNN)

A técnica de aprendizado por K-Vizinhos mais Próximos (KNN, do inglês *K-Nearest Neighbors*) trata-se basicamente de encontrar uma porção de dados (vizinhos) mais próxima segundo uma medida de distância entre os indivíduos observados e o que se quer estimar. A estimativa do modelo é dada com base na média de seus vizinhos mais próximos, que pode variar em função do número k de vizinhos e dos pesos aplicados a eles (Figura 6) (Pedregosa et al., 2011).

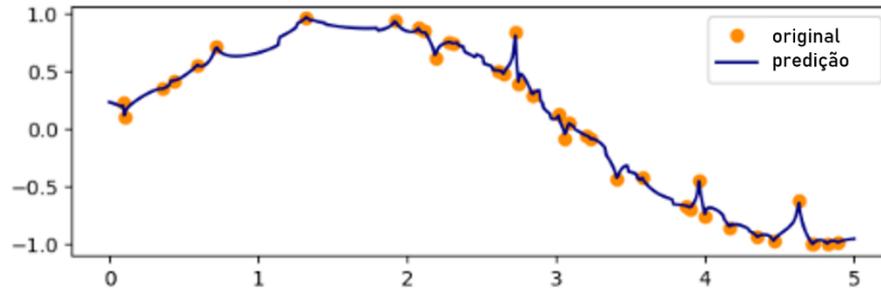


FIGURA 6 - Regressão KNN com ($k = 5$, pesos = distância)

Fonte: Adaptado de Pedregosa *et al.* (2011).

4.4 Dados de sensoriamento remoto

O sensoriamento remoto tem por finalidade adquirir dados de um objeto sem tocá-lo. Essa aquisição é feita por meio de sensores acoplados a plataformas espaciais, aéreas ou terrestres. Já os sensores podem registrar informações das regiões do visível, infravermelho, ultravioleta e micro-ondas do espectro eletromagnético ou gerar nuvem de pontos tridimensionais (Jensen, 2009). A análise das informações adquiridas pode ser feita tanto de forma visual como por meio do processamento digital de imagens, onde dentro desta última temos os Índices de Vegetação (IV), que conseguem relacionar os valores de reflectância espectral a variáveis de interesse florestal.

Entre os vários índices dispostos na literatura, os mais utilizados no setor florestal são aqueles que associam melhores estimativas para volume e biomassa, como o Índice de Vegetação por Diferença Normalizada (NDVI), Razão Simples (SR), Índice de Vegetação Aprimorado (EVI), Índice de Resistência Atmosférica Visível (VARI), Índice de Vegetação de Diferença Normalizada Verde (GNDVI), Índice de Vegetação Ajustado do Solo Transformado (TSAVI), os quais são ideais para mapeamento da cobertura vegetal e suas relações.

4.4.1 Índice de Vegetação por Diferença Normalizada (NDVI)

Proposto por Rouse *et al.* (1973), o Índice de Vegetação por Diferença Normalizada (NDVI, do inglês *Normalized Difference Vegetation Index*), equação (5), é um dos índices a mais populares devido a sua correlação com resposta as variações de vigor da vegetação.

$$NDVI = \frac{(NIR-Red)}{(NIR+Red)} \quad (5)$$

em que: *NIR* é a reflectância no infravermelho próximo e *Red* é a reflectância no vermelho.

4.4.2 Razão Simples (SR)

Desenvolvido por Jordan (1969), a Razão Simples (SR, do inglês *Ratio Simple*), equação (6), está entre os índices mais simples conhecidos para estimar a quantidade de vegetação, devido a redução dos efeitos da atmosfera e topografia.

$$SR = \frac{NIR}{Red} \quad (6)$$

em que: *NIR* é a reflectância no infravermelho próximo e *Red* é a reflectância no vermelho.

4.4.3 Índice de Vegetação Aprimorado (EVI)

Proposto por Huete *et al.* (1997), o Índice de Vegetação Aprimorado (EVI, do inglês *Enhanced Vegetation Index*), equação (7), busca realçar o sinal da vegetação em áreas com maior sensibilidade a fitomassa, por meio da diminuição das influências atmosféricas e de solo.

$$EVI = G \left\{ \frac{(NIR-Red)}{[NIR+C1(Red-C2(Blue+L))]} \right\} \quad (7)$$

em que: *G* é o fator de ganho; *NIR* é a reflectância no infravermelho próximo; *Red* é a reflectância no vermelho; *Blue* é a reflectância no azul; *C1* é o coeficiente de ajuste para o efeito de aerossóis da atmosfera na banda do vermelho; *C2* é o coeficiente de ajuste para o efeito de aerossóis da atmosfera na banda do azul; e *L* o fator de ajuste do solo.

4.4.4 Índice de Resistência Atmosférica Visível (VARI)

Desenvolvido por Gitelson *et al.* (2002), o Índice de Resistência Atmosférica Visível (VARI, do inglês *Visible Atmospherically Resistant Index*), equação (8), é usado para estimar a fração de vegetação de forma quantitativa, com apenas a faixa visível do espectro eletromagnético.

$$VARI = \frac{(Green - Red)}{(Green + Red - Blue)} \quad (8)$$

em que: *Green* é a reflectância no verde; *Red* é a reflectância no vermelho; e *Blue* é a reflectância no azul.

4.4.5 Índice de Vegetação de Diferença Normalizada Verde (GNDVI)

Proposto por Gitelson *et al.* (1996), o Índice de Vegetação de Diferença Normalizada Verde (GNDVI, do inglês *Green Normalized Difference Vegetation Index*), equação (9), é um índice usado para estimar a atividade fotossintética de uma vegetação de forma a determinar a absorção de água e hidrogênio no dossel da planta.

$$GNDVI = \frac{(NIR - Green)}{(NIR + Green)} \quad (9)$$

em que: *NIR* é a reflectância no infravermelho próximo e *Green* é a reflectância no verde.

4.4.6 Índice de Vegetação Ajustado do Solo Transformado (TSAVI)

Desenvolvido por Baret *et al.* (1989), o Índice de Vegetação Ajustado do Solo Transformado (TSAVI, do inglês *Transformed Soil Adjusted*), equação (10), é um índice que minimiza os efeitos as influências do brilho do solo, assumindo que a linha do solo tem uma inclinação e interceptação arbitrárias.

$$TSAVI = \frac{[s*(NIR - s*Red - a)]}{[(a*NIR + Red - a*s + X*(1+s^2))]} \quad (10)$$

em que: *s* é a declividade do solo; *NIR* é a reflectância no infravermelho próximo; *Red* é a reflectância no vermelho; *a* é o intercepto da linha do solo; e *X* é o fator de ajuste visando minimizar os ruídos do solo.

5 METODOLOGIA DE PROCESSAMENTO E ANÁLISE

Os modelos de crescimento e produção e os algoritmos de aprendizado de máquinas foram ajustado/treinado de duas formas: a) contendo variáveis do povoamento, tradicionalmente utilizadas na modelagem florestal, o que foi chamado de modelagem tradicional; e b) adicionando variáveis de sensoriamento remoto ou de cadastro com o propósito de checar se haveria melhora na capacidade preditiva dos modelos.

5.1 Modelagem com variáveis tradicionais

Para a arquitetura do problema, optou-se pela criação de um código único em linguagem *Python*, contendo todos os algoritmos, a fim de modelar os dados do inventário florestal a partir de uma base de dados consistida.

Levando em conta a modelagem tradicional, a equação de Hd (2), que gerara o modelo de sítio, equação (3), e o sistema de equações ajustado de Clutter, equação (4), foram inseridos no código, sendo os ajustes dos coeficientes realizados por simulações com validação cruzada, a fim de avaliar a acurácia dos modelos.

No mesmo código, para parte dos modelos de aprendizado de máquina, foram inseridos os algoritmos segundo a plataforma de Pedregosa *et al.* (2011), Scikit-learn, com as RNA, RF, SVM, KNN e o algoritmo Grid Search para realizar a seleção de parâmetros dos modelos, utilizando as mesmas variáveis usadas para o modelo de Clutter.

Já no algoritmo RNA, foram utilizados como parâmetro número máximo de iterações = 400, taxa de aprendizado inicial = {0,001, 0,01, 0,05}, neurônios por camada oculta = {(10,),(50,),(100,),(500,),(10,10,),(50,10,),(50,50,)} para duas camadas ocultas.

Na SVM, foram utilizados os parâmetros função kernel = função de base radial, gama = $\{2^{-10}, 2^{-8}, \dots, 2^2\}$, regularização C = $\{2^{-2}, 2^0, \dots, 2^{10}\}$.

No algoritmo RF, foi utilizado o parâmetro, número de árvores = $\{2^1, 2^2, \dots, 2^{10}\}$

E no KNN, foi utilizado o parâmetro número de vizinhos = {1,3,5, ..., 21}

Os ajustes das equações e o treinamento dos algoritmos de aprendizado de máquinas, bem como sua avaliação, foram realizados com validação cruzada. Para tal, a base de dados foi dividida em 5 partes e foram realizadas 5 execuções de treinamento e teste em cada uma delas,

considerando uma porção diferente dos dados para teste e as demais para treinamento dos modelos.

5.2 Obtenção de variáveis de sensoriamento remoto

Por meio da base de dados do inventário florestal, no que se refere ao período entre os anos de 2013 e 2019, selecionou-se um satélite que tivesse o maior número de imagens da área de estudo que contemplassem o período de 6 anos. O satélite escolhido foi o Landsat 8 com sensor Operador de Imagem Terrestre (OLI), lançado em fevereiro de 2013, com 8 bandas espectrais, resolução espacial de 30 metros nas bandas do costal, visível, infravermelho e cirrus, e 15 metros na banda do pancromático, com resolução temporal de 16 dias, resolução radiométrica de 12 bits e faixa imageada de 185 km (Zanter, 2016).

Para a aquisição das imagens, foi utilizada a plataforma Earth Explorer da USGS (Serviço Geológico dos Estados Unidos) (Tabela 1). Na Tabela 1, vemos que os três primeiros meses de 2013 não possuem imagens de satélite, em função da data do seu lançamento. Nos demais meses, até o final do ano de 2019, obteve-se imagens, entretanto, a visibilidade delas não é satisfatória. As melhores imagens derivam de julho, agosto e setembro, com porcentagem de nuvens abaixo de 13%, na maioria dos meses, para todos os talhões avaliados.

TABELA 1- Organização dos dados geoespaciais

%nuvens	2013	2014	2015	2016	2017	2018	2019
janeiro	x	49.59	43.25	31.43	14.82	0.17	29.35
fevereiro	x	50.01	19.85	2.76	50.81	80.46	10.14
março	x	65.26	72.16	42.45	4.09	79.03	58.97
abril	95.22	69.56	50.31	34.69	45.24	35.68	7.27
maio	75.56	21.86	83.98	53.77	61.85	1.49	11.84
junho	73.23	93.20	79.80	49.33	4.28	10.11	41.39
julho	12.23	50.99	57.11	58.22	40.45	1.43	2.45
agosto	2.00	0.00	2.59	0.00	12.77	86.56	0.01
setembro	54.01	75.67	0.00	49.22	68.56	0.97	16.64
outubro	22.80	78.30	27.27	96.41	87.42	38.00	77.91
novembro	x	75.92	19.41	87.39	36.66	42.31	56.19
dezembro	75.38	47.52	44.11	95.63	43.72	0.43	71.44

Foram escolhidas aquelas imagens com porcentagem de nuvens abaixo de 13%, sendo realizadas as correções geométrica, radiométrica e atmosférica pelo *software* ENVI, buscando, como produto, a imagem em nível de reflectância.

Posteriormente, buscou-se casar as datas das imagens de satélite com as datas de medições em campo do inventário florestal. Para aqueles em que a data de medição não fosse igual a data imagem de satélite, buscou-se uma data de medição mais próxima possível.

De posse dos limites dos talhões, os arquivos shapefile (.shp) foram convertidos para centróide por meio do software ArcGis, em função da geometria não uniforme dos talhões, o que auxiliou na extração dos índices de vegetação da “seção 4.4”, no *software* ENVI, os quais, posteriormente, foram populados em uma base de dados consistida.

5.3 Medidas de desempenho

Como medida de acurácia dos modelos, foi utilizada a média dos erros absolutos (MAE), equação (11), e a raiz quadrada do erro médio quadrático (RMSE), equação (12):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (12)$$

Em que: y_i = a variável dependente obtida do inventário florestal; \hat{y}_i = a variável dependente da estimativa dos modelos; n = o número de observações.

5.3 Alternativas de análises de composição dos modelos

Buscando avaliar o desempenho dos modelos, foram feitas quatro análises com vistas a estimar o volume de madeira por hectare, nas quais o número da análise, seu objetivo e as variáveis preditoras e seu quantitativo podem ser vistos nos Quadros 1 e 2 a seguir.

QUADRO 1. Análises e seus objetivos

Análise	Objetivo	Variáveis preditoras
1	Estimativa do volume presente (Clutter)	Inventário: idade (meses), área basal (G), índice de sítio (S) e volume total com casca (VTCC).
2	Estimativa do volume presente com índices de sensoriamento remoto (algoritmos de aprendizado de máquinas)	Inventário: idade (meses), área basal (G) e índice de sítio (S) e volume total com casca (VTCC). Sensoriamento remoto: NDVI, SR, VARI, TSAVI, EVI, NGDVI.
3	Estimativa do volume futuro (Clutter)	Inventário: idade atual (meses), área basal (G), índice de sítio (S), volume total com casca (VTCC), idade futura (meses) e área basal futura (G2).
4	Estimativa do volume futuro com variáveis de cadastro (melhor algoritmo de aprendizado de máquinas da análise 2)	Inventário: idade atual (meses), área basal (G), altura dominante (Hd), volume total com casca (VTCC), idade futura (meses) e área basal futura (G2). Variáveis adicionais: talhão, clone, percentual de sobrevivência na idade atual, média aritmética do diâmetro na altura do peito (DAP), média aritmética da altura.

QUADRO 2 – Número de amostras e variáveis preditoras

Análise	Objetivo	Amostras	Variáveis preditoras
1	Estimativa do volume presente (Clutter)	5853	4
2	Estimativa do volume presente com índices de sensoriamento remoto (algoritmos de aprendizado de máquinas)	135	11
3	Estimativa do volume futuro (Clutter)	3405	6
4	Estimativa do volume futuro com variáveis de cadastro (melhor algoritmo de aprendizado de máquinas da análise 2)	3405	12

6 RESULTADOS

6.1 Análise 1: estimativa do volume presente

Buscou-se estimar o volume presente utilizando os algoritmos da “seção 5.1” e as variáveis preditoras descritas no Quadro 1. Em relação ao desempenho preditivo dos modelos, foi utilizado a média dos erros absolutos (MAE), equação (11), e a raiz quadrada do erro quadrático médio (RMSE), equação (12), com seus respectivos desvios padrões. Os resultados para a estimativa do volume presente são apresentados na Tabela 2.

TABELA 2. Erro da estimativa da produção volumétrica presente ($m^3 ha^{-1}$)

Algoritmo	MAE	RMSE
Clutter	7,38 ± 0,07	10,36 ± 0,65
KNN	8,43 ± 0,23	12,13 ± 0,73
RF	7,80 ± 0,10	10,83 ± 0,62
MLP	9,05 ± 1,15	11,89 ± 1,35
SVM	7,13 ± 0,06	9,95 ± 0,65

Observa-se que o modelo de Clutter apresentou valores de estimativa de erro melhores que os algoritmos de aprendizado de máquina, ficando atrás apenas do algoritmo SVM, que obteve os melhores resultados para a estimativa de volume presente, com MAE, $7,13 m^3 ha^{-1}$, e RMSE $9,95 m^3 ha^{-1}$.

Isso aponta a razão de esse modelo ser o mais utilizado pelas empresas florestais no Brasil, o que é explicado pelo seu menor grau de complexidade e por atender aos requisitos de planejamento de produção de madeira (Castro, 2013b).

6.2 Análise 2: estimativa do volume presente com a adição de dados de sensoriamento remoto

Dado que a SVM gerou estimativas com a menor taxa de erro, essa foi escolhida para o a análise 2, que consistiu na adição de variáveis de sensoriamento remoto como preditores do volume presente. As variáveis utilizadas como entrada foram as mesmas utilizadas na seção anterior (tradicionais), com o acréscimo dos índices de vegetação NDVI, DVI, VARI, TSAVI, EVI, NGDVI, individualmente e todos juntos.

Os resultados obtidos são mostrados na Tabela 3, na qual as combinações de entrada são descritas no nome da análise.

TABELA 3. Erro da estimativa produção volumétrica presente considerando índices de vegetação e dados de inventário florestal (m^3ha^{-1})

Variáveis preditoras	MAE	RMSE
Tradicionais	4,30 ± 0,59	6,11 ± 1,12
Tradicionais+ NDVI	4,21 ± 0,66	5,99 ± 1,14
Tradicionais+ DVI	4,28 ± 1,19	5,94 ± 0,44
Tradicionais+ VARI	4,07 ± 0,53	5,62 ± 0,58
Tradicionais+ TSAVI	4,28 ± 0,62	6,05 ± 1,12
Tradicionais+ EVI	4,29 ± 0,21	5,96 ± 0,13
Tradicionais+ NGDVI	4,16 ± 0,58	5,66 ± 0,64
Tradicionais+ All	4,24 ± 0,41	5,87 ± 0,41

Os índices de vegetação adicionados às variáveis tradicionais como preditores melhoraram o erro na estimativa do volume presente. O destaque se deu para o índice VARI, que resultou no menor erro de predição, com redução do MAE para $4,07 m^3ha^{-1}$ e do RMSE com $5,62 m^3ha^{-1}$, que perfazem reduções percentuais de aproximadamente 5% e 8%, respectivamente. Isso pode ser explicado pela capacidade de esse índice detectar o estresse da planta em função da quantidade de verde, mensurado em cima do desenvolvimento da cultura em determinada área (Mappa, 2022).

Isso demonstra a potencialidade que os índices de vegetação têm em extrair informações intrínsecas de determinada cultura e, mesmo com uma baixa quantidade de amostras, conseguir agregar uma melhor previsão quando associados a modelos robustos, como a SVM.

6.3 Análise 3: estimativa do volume futuro

Para a estimativa do volume futuro, ou seja, em uma idade diferente da idade atual, duas variáveis componentes do modelo de Clutter foram adicionadas aos modelos de aprendizado de máquinas: a idade futura e a área basal futura. Os resultados para a estimação do volume futuro são apresentadas na Tabela 4.

TABELA 4. Erro da estimativa de produção volumétrica futura ($m^3 ha^{-1}$)

Algoritmo	MAE	RMSE
Clutter	28,22 ± 2,19	42,52 ± 3,76
KNN	29,71 ± 1,19	44,82 ± 3,39
RF	27,92 ± 1,97	42,49 ± 2,98
MLP	25,30 ± 1,28	39,27 ± 2,51
SVM	24,78 ± 1,70	39,36 ± 3,94

Os algoritmos de aprendizado de máquinas, exceto o KNN, tiveram melhor desempenho que o modelo de Clutter nas estimativas para o volume futuro. O SVM se mostrou o melhor para o MAE, com erro de $24,78 m^3 ha^{-1}$, e a MLP o melhor para o RMSE, com erro de $39,27 m^3 ha^{-1}$.

No geral, houve pequena discrepância dos indicadores de desempenho entre os modelos. Castro (2013a) ao comparar uma MLP ao modelo de Clutter, também reportou melhor desempenho daquela sobre esta.

6.4 Análise 4: estimativa do volume futuro com adição de dados de cadastro do inventário

Nessa análise, foram realizadas novas simulações cujas variáveis utilizadas como entrada foram as mesmas utilizadas na seção anterior, acrescidas das variáveis de cadastro e de outras variáveis do inventário, mas que geralmente não são empregadas como preditores nos modelos de crescimento e produção em nível de povoamento mais tradicionais: Talhão, Clone, Sobrevivência, média aritmética do DAP e média aritmética da altura, que foram adicionadas individualmente e em conjunto. As variáveis adicionais do inventário eram aquelas referentes à idade atual, ou seja, não à idade da prognose. Os resultados de desempenho das variações nas adições dessas variáveis são mostrados na Tabela 5.

TABELA 5. Erro da estimativa de volume de produção futuro considerando variáveis de cadastro do inventário florestal ($m^3 ha^{-1}$)

Variáveis predictoras	MAE	RMSE
Tradicionais	25,70 ± 1,78	40,33 ± 3,76
Tradicionais+ Talhão	25,50 ± 1,83	40,17 ± 3,82
Tradicionais+ Clone	24,94 ± 1,96	39,57 ± 4,03
Tradicionais+ Sobrev	24,78 ± 1,70	39,36 ± 3,94
Tradicionais+ \overline{DAP}	26,09 ± 1,90	40,74 ± 3,65
Tradicionais+ \overline{Ht}	25,46 ± 1,87	40,22 ± 3,83
Tradicionais+ Todas adicionais	24,18 ± 1,72	38,95 ± 3,67

A inserção de novas variáveis do inventário promoveu melhora na estimativa do volume futuro. A combinação das variáveis tradicionais mais todos as de cadastro resultou nos menores valores de MAE e RMSE, sendo 24,18 $m^3 ha^{-1}$ e 38,95 $m^3 ha^{-1}$, correspondentes a uma melhoria de aproximadamente 6,0% e 3,4%, respectivamente. Isso demonstra que os algoritmos de aprendizado de máquina, como a SVM, conseguem ter predições mais acuradas, tendo os mesmos atributos do modelo tradicional de Clutter.

7 CONCLUSÕES

Não houve discrepância na acurácia entre o modelo de Clutter e os modelos de aprendizado de máquinas quanto à capacidade de estimar o volume presente de madeira. Nessa situação, mantendo os mesmos preditores, o modelo de SVM foi o que teve melhor desempenho. Para predição da produção volumétrica futura, os modelos de aprendizado de máquina foram capazes de melhorar, ainda que modestamente, as estimativas quando comparados ao tradicional de Clutter, com destaque novamente para a SVM. Vale ressaltar que tal investigação foi publicada em Araujo et al. (2023).

A inserção dos índices de vegetação no modelo SVM para estimar o volume presente repercutiu em ligeira melhora das estimativas em comparação com as variáveis tradicionais, com destaque para o VARI, com a melhor estimativa.

A inserção de variáveis adicionais de cadastro e do inventário no modelo da SVM para a estimativa da produção futura contribui para uma pequena melhora da acurácia dos modelos.

Espera-se, futuramente, buscar melhorias nos modelos apresentados, usando de técnicas mais robustas de aprendizado de máquina, como as Redes Neurais Convolucionais, assim como utilizar outras fontes de dados geográficos.

REFERÊNCIAS

ALCÂNTARA, A. E. M. **Alternativas de modelagem para projeção do crescimento de eucalipto em nível de povoamento**. 2012. 52f. 2012. Dissertação de Mestrado. Universidade Federal de Viçosa, Viçosa, MG.

ALCÂNTARA, A. E. M. **Redes neurais artificiais para prognose do crescimento e da produção de povoamentos de eucalipto em Minas Gerais**. 2015. 43f. 2015. Tese de Doutorado. Universidade Federal de Viçosa, Viçosa, MG.

ALMEIDA, A. C. **Metodologia Integrada Utilizando Sensoriamento Remoto e Redes Neurais Artificiais na Quantificação do Potencial de Biomassa Florestal na Amazônia**. 135 f. Tese (Doutorado em Sistemas de Energia) - Instituto de Tecnologia, Universidade Federal do Pará, Belém, 2008.

ARAÚJO, R. M. C. et. al. Modeling of growth and production of Eucalyptus spp stands using supervised machine learning. In: **Anais do XX Encontro Nacional de Inteligência Artificial e Computacional**. SBC, 2023. p. 1-14. ISSN 2763-9061. DOI: <https://doi.org/10.5753/eniac.2023.232893>

ASSAMANN, E. **The principles of Forest yield study**. New York: Pergamon Press, 1970. 506p. DOI: <https://doi.org/10.1016/C2013-0-01587-3>

BARET, F; GUYOT, G. Potentials and limits of vegetation indices for LAI and APAR assessment. **Remote sensing of environment**, v. 35, n. 2-3, p. 161-173, 1991. DOI: [https://doi.org/10.1016/0034-4257\(91\)90009-U](https://doi.org/10.1016/0034-4257(91)90009-U)

BINOTI, M. L. M. S. **Emprego de redes neurais artificiais em mensuração e manejo florestal**. 2012. 130 f. 2012. Tese de Doutorado. Tese (Doutorado em Ciência Florestal) -Universidade Federal de Viçosa, Viçosa, MG.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. DOI: <https://doi.org/10.1023/A:1010933404324>

BRINK, H; RICHARDS, J. W.; FETHEROLF, M. **Real-world machine learning**. Manning, 400p. 2017.

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

CARNEIRO, M. G. **Redes complexas para classificação de dados via conformidade de padrão, caracterização de importância e otimização estrutural**. 2017. Tese de Doutorado. Universidade de São Paulo. DOI: [10.11606/T.55.2017.tde-01022017-100223](https://doi.org/10.11606/T.55.2017.tde-01022017-100223)

CASTRO, R. V. O. et al. Individual growth model for Eucalyptus stands in Brazil using artificial neural network. **ISRN Forestry**, v. 2013, 2013a. DOI: <https://doi.org/10.1155/2013/196832>

CASTRO, R. V. O. et al. Crescimento e produção de plantios comerciais de eucalipto estimados por duas categorias de modelos. **Pesquisa Agropecuária Brasileira**, v. 48, n. 3, p.287-295, 2013b. DOI: <https://doi.org/10.1590/S0100-204X2013000300007>

CLUTTER, J. L. Compatible growth and yield for loblolly pine. **Forest Science**, v.9, n.3, p.354-371, 1963.

CORDEIRO, M. A. et al. Estimativas volumétricas em povoamentos de eucalipto utilizando máquinas de vetores de suporte e redes neurais artificiais. **Madera y bosques**, v. 28, n. 1, 2022. DOI: <https://doi.org/10.21829/myb.2022.2812252>

COSENZA, D. N. et al. Site classification for eucalypt stands using artificial neural network based on environmental and management features. **Cerne**, v. 23, n. 3, p. 310-320, 2017. DOI: <https://doi.org/10.1590/01047760201723032352>

DAI, P. V. S. et al. Estimativa de volume de madeira baseada em índices de vegetação. **Scientia Forestalis**, v. 49, n. 129, p. e3301, 2021. DOI: <https://doi.org/10.18671/scifor.v49n129.06>

CAMPOS, A. S. et al. Análises bibliométricas sobre a produção e tecnologia de celulose e papel de eucalipto. In: **9º Congresso Florestal Brasileiro**. 2022. p. 365-368. DOI: <https://doi.org/10.55592/CFB.2022.5287494>

FREITAS, E. C. S. et al. Modeling of eucalyptus productivity with artificial neural networks. **Industrial Crops and Products**, v. 146, p. 112149, 2020. DOI: <https://doi.org/10.1016/j.indcrop.2020.112149>

FERRAZ, A.S. et al. Estimativa do estoque de biomassa em um fragmento florestal usando imagens orbitais. **Floresta e Ambiente**, v. 21, n. 3, p. 286-296, 2014. DOI: <https://doi.org/10.1590/2179-8087.052213>

GALIANO, V. R. et al. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. **Ore Geology Reviews**, v. 71, p. 804-818, 2015. DOI: <https://doi.org/10.1016/j.oregeorev.2015.01.001>

GARCIA, S. C. **O uso de árvores de decisão na descoberta de conhecimento na área da saúde**. 87 f. 2003. Dissertação de Mestrado em Ciências da Computação. Universidade Federal do Rio Grande do Sul, Porto Alegre. 2003.

GITELSON, A. A, et al. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. **Remote sensing of Environment**, v. 58, n. 3, p. 289-298, 1996. DOI: [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)

GITELSON, A. A. et al. Novel algorithms for remote estimation of vegetation fraction. **Remote sensing of Environment**, v. 80, n. 1, p. 76-87, 2002. DOI: [https://doi.org/10.1016/S0034-4257\(01\)00289-9](https://doi.org/10.1016/S0034-4257(01)00289-9)

GOETZ, J. N. et al. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. **Computers & geosciences**, v. 81, p. 1-11, 2015. DOI: <https://doi.org/10.1016/j.cageo.2015.04.007>

GONÇALVES, M. J. L. et al. Integrating genetic and silvicultural strategies to minimize abiotic and biotic constraints in Brazilian eucalypt plantations. **Forest ecology and management**, v. 301, p. 6-27, 2013. <https://doi.org/10.1016/j.foreco.2012.12.030>

GORGENS, E. B. et al. Estimação do volume de árvores utilizando redes neurais artificiais. **Revista Árvore**, v. 33, n. 6, p. 1141-1147, 2009. DOI: <https://doi.org/10.1590/S0100-67622009000600016>

HAYKIN, S. **Redes Neurais: Princípios e Práticas**. [S.l.]: Bookman, 2001. v. 2.

HUETE, A. R. et al. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. **Remote sensing of environment**, v. 59, n. 3, p. 440-451, 1997. DOI: [https://doi.org/10.1016/S0034-4257\(96\)00112-5](https://doi.org/10.1016/S0034-4257(96)00112-5)

IBA. Relatório Ibá 2022. Relatório. Brasília, 2022. Relatório. Mimeografado.

JORDAN, C. F. Derivation of leaf-area index from quality of light on the forest floor. **Ecology**, v. 50, n. 4, p. 663-666, 1969. DOI: <https://doi.org/10.2307/1936256>

JENSEN, J. R. **Sensoriamento Remoto do Ambiente: uma perspectiva em recursos terrestres**. São José dos campos: Parêntese, 2009. 598p.

LEITE, H. G.; CAMPOS, J. C. C. **Mensuração Florestal: Perguntas e respostas**. 2017.

MALINOVISK. **A Produtividade das Florestas Plantadas de Eucalipto**. Youtube, 23 jul. 2020. Disponível em: <https://www.youtube.com/watch?v=YA51Ee_RMkc> Acesso em: 16 ago. 2021.

MAPPA, **Índices de Vegetação: IFV, VARI, NDVI e NDRE. 2022**. Disponível em: <<https://mappa.ag/blog/indices-de-vegetacao-ifv-vari-ndvi-mare-ndre/#:~:text=Índice%20Resistente%20à%20Atmosfera%20na,cultura%20em%20uma%20determinada%20área.>>. Acesso em: 15 fev. 2023.

MAXWELL, A E.; WARNER, T A.; FANG, F. Implementation of machine-learning classification in remote sensing: An applied review. **International journal of remote sensing**, v. 39, n. 9, p. 2784-2817, 2018. DOI: <https://doi.org/10.1080/01431161.2018.1433343>

NASCIMENTO, R. F. F. et al. O algoritmo Support Vector Machines (SVM): avaliação da separação ótima de classes em imagens CCD-CBERS-2. **Anais do XIV Simpósio Brasileiro de Sensoriamento Remoto [CDROM]**, p. 25-30, 2009.

NISHIKAWA, Hitoshi et al. Dry Matter Yield and Nitrogen Content Estimation in Grassland Using Hyperspectral Sensor. **Remote Sensing**, v. 15, n. 2, p. 419, 2023. DOI: <https://doi.org/10.3390/rs15020419>

PAGLIARUSSI, M. S. **O ensino do modelo clássico de regressão linear por meio de simulação de Monte Carlo**. **Revista de Contabilidade e Organizações**, [S. l.], v. 12, p. e152100, 2018. DOI: <https://doi.org/10.11606/issn.1982-6486.rco.2018.152100>

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

PENIDO, T. M. A. et al. Modelos de crescimento e produção para a estimativa volumétrica em povoamentos comerciais de eucalipto. **Scientia Forestalis**, v. 48, n. 128, p. e3340, 2020. DOI: <https://doi.org/10.18671/scifor.v48n128.06>

REIS, L. P. et al. Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest. **Ecological Engineering**, v. 112, p.140-147, 2018. DOI: <https://doi.org/10.1016/j.ecoleng.2017.12.014>

REIS, L. P. et al. Modeling of tree recruitment by artificial neural networks after wood harvesting in a forest in eastern Amazon rain forest. **Ciência Florestal**, v. 29, n. 2, p. 583-594, 2019. DOI: <https://doi.org/10.5902/1980509825808>

REX, F, E. **Estimativas do estoque e dinâmica de biomassa acima do solo utilizando diferentes abordagens estatísticas e dados LiDAR em floresta tropical**. 2019. 119f. Dissertação de Mestrado. Universidade Federal do Paraná, Curitiba, PR.

RODRÍGUEZ, P, R.; BAJORATH, J. Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. *Journal of Computer-Aided Molecular Design*, v. 36, n. 5, p. 355-362, 2022. DOI: <https://doi.org/10.1007/s10822-022-00442-9>

ROUSE, J.W. et al. Monitoring vegetation systems in the great plains with ERTS. In: Third ERTS Symposium, Proceedings, **NASA SP-351**, NASA, Washington,DC, v. 1, p. 309-317, 1973.

SALAKEN, S. M. et al. Seeded transfer learning for regression problems with deep learning. **Expert Systems with Applications**, v. 115, p. 565-577, 2019. DOI: <https://doi.org/10.1016/j.eswa.2018.08.041>

SANQUETTA, C. R. et al. Volume estimation of *Cryptomeria japonica* logs in southern Brazil using artificial intelligence models. **Southern Forests: a Journal of Forest Science**, v. 80, n.1, p. 29-36, 2018. DOI: <https://doi.org/10.2989/20702620.2016.1263013>

SANTOS, D. E. C. d. **Regressão e Deep Learning Aplicados em Modelagem de Árvore Individual para Eucalipto**. Dissertação (Mestrado). Programa de Pós-Graduação em Ciência Florestal, Universidade Federal de Viçosa, Viçosa, 2021. DOI: <https://doi.org/10.47328/ufvbbt.2021.033>

SOARES, C.P.B.; LEITE, H.G.; CAMPOS, J.C.C. Um novo modelo de crescimento e produção. **Revista Árvore**, v.25, n.2, p.265-270. 2001.

TAVARES JÚNIOR. I. S. et al. Artificial neural networks and linear regression reduce sample intensity to predict the commercial volume of eucalyptus clones. **Forests**, v. 10, n. 3, p. 268, 2019. DOI: <https://doi.org/10.3390/f10030268>

WERE, K. et al. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. **Ecological Indicators**, v. 52, p. 394-403, 2015. DOI: <https://doi.org/10.1016/j.ecolind.2014.12.028>

ZANTER, K. Landsat 8 (L8) data users handbook. **Landsat Science Official Website**, v. 33, 2016.