
**Modelo de Predição do Crescimento e Produção
para Plantios de Tachi-branco baseado em
Autômato Celular e Aprendizado de Máquina**

Pablo Henrique de Freitas



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Uberlândia
2023

Pablo Henrique de Freitas

**Modelo de Predição do Crescimento e Produção
para Plantios de Tachi-branco baseado em
Autômato Celular e Aprendizado de Máquina**

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Luiz Gustavo Almeida Martins

Coorientador: Murillo Guimarães Carneiro

Uberlândia

2023

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

F866
2023

Freitas, Pablo Henrique de, 1997-
Modelo de Predição do Crescimento e Produção para
Plantios de Tachi-branco baseado em Autômato Celular e
Aprendizado de Máquina [recurso eletrônico] : / Pablo
Henrique de Freitas. - 2023.

Orientador: Luiz Gustavo Almeida Martins.
Coorientador: Murillo Guimarães Carneiro.
Dissertação (Mestrado) - Universidade Federal de
Uberlândia, Pós-graduação em Ciência da Computação.
Modo de acesso: Internet.
Disponível em: <http://doi.org/10.14393/ufu.di.2023.562>
Inclui bibliografia.

1. Computação. I. Martins, Luiz Gustavo Almeida, 1974-,
(Orient.). II. Carneiro, Murillo Guimarães, 1988-,
(Coorient.). III. Universidade Federal de Uberlândia.
Pós-graduação em Ciência da Computação. IV. Título.

CDU: 681.3

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
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Os abaixo assinados, por meio deste, certificam que leram e recomendam para a Faculdade de Computação a aceitação da dissertação intitulada **Modelo de Predição do Crescimento e Produção para plantios de Tachi-branco baseado em Autômato Celular e Aprendizado de Máquina** por **Pablo Henrique de Freitas** como parte dos requisitos exigidos para a obtenção do título de Mestre em Ciência da Computação.

Uberlândia, ____ de _____ de ____

Orientador: _____

Prof. Dr. Luiz Gustavo Almeida Martins
Universidade Federal de Uberlândia

Coorientador: _____

Prof. Dr. Murillo Guimarães Carneiro
Universidade Federal de Uberlândia

Banca examinadora:

Prof. Dr. Álvaro Augusto Vieira Soares
Universidade Federal de Uberlândia

Prof^a. Dra. Danielli Araújo Lima
Instituto Federal do Triângulo Mineiro

Dedico este trabalho aos meus amados pais, esposa e filho, cujo amor incondicional, apoio incansável e inspiração constante tornaram possível a realização deste sonho acadêmico. Cada conquista minha é um reflexo do investimento que fizeram em minha vida. A vocês, minha gratidão eterna.

Agradecimentos

Agradeço, primeiramente a Deus, que com sua maravilhosa Graça, concedeu-me tudo o que foi necessário para passar por esses 2 anos e meio de mestrado com êxito. Toda minha gratidão a Ele! À minha mãe Rosilene e ao meu pai Paulo, por todo apoio, amor, educação, ensinamentos e sustento. Sem todos os esforços e sacrifícios feitos por eles, nada disso seria possível. A minha esposa Mariana pela paciência, apoio incondicional e sustento nas noites de estudo e pesquisa. Ao meu professor de matemática do Ensino Fundamental e Médio, Alessandro Moreira, que me fez gostar de matemática e me apresentou a Estatística. Graças às suas dicas, orientações e apoio, pude descobrir e me apaixonar pela minha amada profissão. Aos professores Luiz Gustavo e Murillo Carneiro, pelas orientações no desenvolvimento do trabalho e pesquisa.

*“Porquanto é o SENHOR quem concede sabedoria, e da sua boca procedem a
inteligência e o discernimento.”
(Provérbios 2:6)*

Resumo

A dinâmica das plantações florestais tem sido amplamente estudada com aplicações de simulação computacional. Autômatos celulares (AC) é uma técnica capaz de modelar estados futuros com base em um conjunto de regras de transição. No entanto, essa construção não é simples, muitas vezes exigindo conhecimento técnico do processo por meio de anos de pesquisa científica. Técnicas de aprendizado de máquina podem ser aplicadas nesse contexto, facilitando a construção desses simuladores. De acordo com esse cenário, este trabalho apresenta um modelo de simulação baseado em autômatos celulares probabilísticos capazes de estimar a evolução da produção de madeira ao longo do período de manejo. Diferente de outros trabalhos da literatura, a construção da regra de transição do AC é baseada exclusivamente em dados históricos de uma plantação de Tachi-branco, uma espécie de floresta manejada. Modelos de regressão linear e logística são aplicados para aprender e representar as regras de transição locais do autômato e simular sua evolução. A abordagem proposta baseada em AC foi capaz de prever o comportamento futuro das plantações nas áreas monitoradas com erros em torno de 4%, confirmando o potencial do uso de aprendizado de máquina na descoberta de regras de transição para modelos precisos baseados em autômatos celulares.

Palavras-chave: Autômato Celular. Aprendizado de Máquina. Predição da produção. Florestas plantadas.

Abstract

The dynamics of forest plantations have been widely studied with computational simulation applications. Cellular automata (CA) is a technique capable of modelling future states based on a set of transition rules. However, this construction is not simple, often requiring technical knowledge of the process through years of scientific research. Machine learning techniques can be applied in this context, facilitating the construction of these simulators. According to this scenario, this work presents a simulation model based on probabilistic cellular automata capable of estimating the evolution of wood production throughout the management period. Unlike other works in the literature, the construction of the AC transition rule is based exclusively on historical data from a Tachi-branco plantation, a managed forest species. Linear and logistic regression models are applied to learn and represent the local transition rules of the automaton and simulate its evolution. The proposed CA-based approach was able to predict the future behavior of plantations in the monitored areas with errors around 4%, confirming the potential of using machine learning in discovering transition rules for precise models based on cellular automata.

Keywords: Cellular Automata. Machine Learning. Production prediction. Managed Forest.

Lista de ilustrações

Figura 1 – Árvore de Tachi-branco (Fonte: arquivo pessoal de Álvaro Soares) ¹ . . .	21
Figura 2 – Tipos de dimensionalidade dos autômatos celulares. (FRAGA; OLIVEIRA; MARTINS, 2020)	27
Figura 3 – Tipos de formatos de célula dos autômatos celulares. (FRAGA; OLIVEIRA; MARTINS, 2020)	28
Figura 4 – Tipos de bordas dos autômatos celulares.	28
Figura 5 – Tipos de vizinhança dos autômatos celulares.	29
Figura 6 – Exemplos de raios para os tipos de vizinhança usados em autômatos celulares 2D.	29
Figura 7 – Evolução de um autômato celular elementar.	31
Figura 8 – Evolução de um reticulado no Jogo da Vida considerando quatro instantes de tempo.	31
Figura 9 – Ilustração teórica de um modelo MLP de uma camada.	33
Figura 10 – Ilustração teórica de um modelo Random Forest.	34
Figura 11 – Ilustração teórica de um modelo XGB. Disponível em < https://www.researchgate.net/figure/general-architecture-of-XGBoost_fig3335483097 >	34
Figura 12 – Curva logística probabilística.	36
Figura 13 – Local do experimento com a plantação de Tachi-branco o leste da Amazônia, no estado do Pará. Disponível em (LIMA et al., 2022)	46
Figura 14 – Visão geral do processo de treinamento e validação do modelo de predição proposto.	48
Figura 15 – Exemplo da evolução de um reticulado do autômato celular, representando uma parcela da plantação.	48
Figura 16 – Processo de evolução do reticulado do autômato celular durante a simulação.	51
Figura 17 – Matriz de confusão do modelo de regressão logística para previsão do estado da árvore.	55

Figura 18 – Valor dos coeficientes e importância de cada variável no modelo de mortalidade.	56
Figura 19 – Comparação entre os índices de competição das árvores vivas e mortas.	57
Figura 20 – Valores reais vs. preditos pelos modelos de regressão linear.	58
Figura 21 – Comparativo entre os dados reais e simulados pelo modelo AC para as medições da variável DEQ.	60
Figura 22 – Comparativo entre os dados reais e simulados pelo modelo AC para as medições da variável Altura.	60
Figura 23 – Comparativo entre os dados reais e simulados pelo modelo AC para as medições do volume total de madeira (m ³).	61
Figura 24 – Comparativo entre os dados reais e simulados pelo modelo AC para as contagem da quantidade de árvores vivas.	62
Figura 25 – Comparativo entre os dados reais e simulados pelo modelo AC para as medições da variável DEQ em relação a variável N ^o de fustes.	63
Figura 26 – Comparativo entre os dados reais e simulados pelo modelo AC para as medições da variável Altura em relação a variável N ^o de fustes.	63
Figura 27 – Comparativo entre os dados reais e simulados pelo modelo AC para as medições do volume total de madeira (m ³) em relação a variável N ^o de fustes.	63
Figura 28 – Comparativo entre os dados reais e simulados pelo modelo AC para as contagem da quantidade de árvores vivas.	64

Lista de tabelas

Tabela 1 – Características dos modelos desenvolvidos nos trabalhos correlatos. . .	43
Tabela 2 – Acurácia dos modelos de aprendizado de máquina em cada rodada de treinamento (em %).	53
Tabela 3 – Média das métricas dos modelos de aprendizado de máquina (em %). .	53
Tabela 4 – Desempenho do modelos de aprendizado de máquina e do autômato celular em cada rodada de simulação.	59
Tabela 5 – Teste t-studente para comparação entre as curvas baseado em seus valores médios.	61

Lista de siglas

Sumário

1	INTRODUÇÃO	15
1.1	Motivação	17
1.2	Objetivos	18
1.3	Organização da Dissertação	19
2	FUNDAMENTAÇÃO TEÓRICA	20
2.1	A espécie Tachi-branco	20
2.2	Modelos de Crescimento das Árvores	22
2.2.1	Funções de Crescimento	23
2.2.2	Modelos de Árvores Individuais	23
2.2.3	Índices de Competição	25
2.3	Autômatos celulares	26
2.3.1	Breve Histórico	26
2.3.2	Espaço Celular	27
2.3.3	Estado e Vizinhança	28
2.3.4	Regra de Transição	30
2.4	Modelos de Aprendizado de Máquina	32
2.4.1	Rede Neural Multi-layer Perceptron (MLP)	32
2.4.2	Random Forest	33
2.4.3	XGB - Extreme Gradient Boosting	33
2.4.4	Modelos de Regressão	35
3	REVISÃO DA LITERATURA	40
4	MEPACAM - MODELO DE ESTIMAÇÃO DA PRODUÇÃO DE MADEIRA BASEADO EM AUTÔMATO CELULAR E APRENDIZAGEM DE MÁQUINA	44
4.1	Estudo de Caso	44

4.2	Construção do Modelo	45
4.3	Espaço Celular	47
4.4	Regra de Transição	49
5	EXPERIMENTOS E ANÁLISE DOS RESULTADOS	52
5.1	Análise dos Modelos de Aprendizado de Máquina	52
5.1.1	Estimação de Mortalidade das Árvores	54
5.1.2	Estimação das Medidas Dendométricas	57
5.2	Desempenho do Modelo baseado em Autômato Celular	58
5.3	Análise da Dinâmica de Crescimento Resultante	59
5.3.1	Espaçamento do Plantio	59
5.3.2	Número de fustes vivos	62
6	CONCLUSÃO	65
6.1	Trabalhos Futuros	66
6.2	Produção Bibliográfica Originada da Pesquisa	67
	REFERÊNCIAS	68

Introdução

A silvicultura vêm se tornando um importante pilar da economia brasileira, com as florestas de manejo já ocupando grandes áreas plantadas e em ampla expansão. Para apoiar a tomada de decisões no manejo florestal, as empresas monitoram o crescimento e a produção das florestas a partir de inventários florestais. Tradicionalmente, o inventário florestal é uma tarefa de custo elevado e que demanda um tempo considerável para as medições *in loco* das árvores. Uma alternativa é a adoção de modelos capazes de fornecer uma projeção rápida e precisa do crescimento e produção nas parcelas monitoradas com base em seus dados de plantio, crescimento e mortalidade. No entanto, esses modelos devem lidar com padrões estocásticos da dinâmica florestal, o que complica seu design (REIS et al., 2018). Portanto, o estudo de estratégias para construir modelos de alta precisão para simular o crescimento das árvores e estimar a produção futura de áreas plantadas é um tópico de pesquisa atual e de grande relevância.

Árvores de Eucalipto (gênero *Eucalyptus*) é o gênero florestal mais plantado no Brasil e manejado intensivamente para a obtenção de alguns produtos como, por exemplo, polpa celulósica, papel, madeira para marcenaria e geração de energia (STAPE; RYAN; BINKLEY, 2004; HIGA; WREGGE, 2010). As plantações de Eucalipto são a fonte primária de biomassa para fins energéticos no Brasil, entretanto, não há plantações suficientes na região norte, estimulando assim pesquisas de novas alternativas para suprir a demanda local (RODRIGUES; JUNIOR, 2019).

O reflorestamento com *Tachigali vulgaris* (*T. vulgaris*), comumente conhecido como Tachi Branco (RODRIGUES et al., 2020; SILVA et al., 2021) é uma potencial fonte de energia renovável para o abastecimento da região Amazônica Brasileira. O plantio de espécies nativas traz muitas vantagens, como eliminar o período de adaptação das espécies ao ambiente (ORELLANA et al., 2018), mas também desafios, como promover uma exploração sustentável que afete minimamente o bioma amazônico. Tachi Branco é uma espécie promissora para florestas bioenergéticas na região amazônica por ser adaptada às condições ambientais regionais, apresentar rápido crescimento e produzir madeira de qualidade para os fins desejados (SILVA et al., 2021).

Em relação ao estudo de técnicas relacionadas ao manejo florestal, a modelagem do crescimento e produção de florestas comerciais apresenta-se bem desenvolvida, principalmente no manejo de eucalipto. Um povoamento refere-se à população de árvores em uma área específica que foi plantada ou cultivada para a produção sustentável de madeira ou outros produtos florestais. Muitas pesquisas e estudos foram desenvolvidos, considerando modelos em nível de povoamento e de distribuição diamétrica. Os modelos em nível de povoamento são utilizados para se obter estimativas do crescimento e produção por unidade de área (CAMPOS; LEITE, 2006), onde a predição do volume por hectare é obtida através de variáveis sobre o plantio, como, por exemplo, idade, índice local, área basal, entre outros. Já os modelos de distribuição diamétrica predizem o número de árvores e a produção por hectare para cada classe de diâmetro (HYINK; MOSER, 1983). Neste caso, a estimativa da produção total é obtida através da soma de todas as classes de diâmetro. Nessa técnica, é utilizada uma função densidade de probabilidade para descrever a distribuição dos diâmetros para a idade atual e futura, sendo a função Weibull a mais utilizada (RENNOLLS; GEARY; ROLLINSON, 1985).

Por outro lado, são poucos os modelos de crescimento em nível de árvore individual no Brasil (MENDES et al., 2006). Existem alguns estudos onde modelos de crescimento em nível de árvore individual têm sido desenvolvidos para algumas espécies de árvores, como Canela Preta (DURLO, 2001), Cedro e Araucária (DELLA-FLORA, 2001), e eucalipto (CASTRO, 2011)(MARTINS, 2011). Também existem estudos utilizando modelos de árvores individuais para o crescimento de florestas nativas pós desmatamento e queimadas (BANDINI; PAVESI, 2004). Entretanto, no Brasil, esse tipo de modelo não é usual na estimação da produção em reflorestamentos, principalmente em pesquisas que buscam prever o volume total de madeira produzido em plantações de Tachi Branco.

Nos modelos de árvores individuais, cada árvore é considerada e estudada, levando-se em consideração suas próprias características individuais e também características locais, como as variáveis de seus vizinhos e condições de competição e produção (HASENAUER, 2000). A estrutura de projeção do crescimento nesse nível são mais complexas, pois é necessário representar as interações entre os indivíduos nos sub-modelos de crescimento e mortalidade que compõem o modelo de predição (BIGING; DOBBERTIN, 1992).

Dado que nos modelos de árvores individuais a estimativa dos dados de cada indivíduo é qualificado pela interação entre vizinhos, percebe-se que autômatos celulares (AC) têm grande potencial para a simulação do crescimento florestal nesse nível de detalhe, se mostrando simples e eficaz na estimação e evolução dos efeitos competitivos entre árvores vizinhas, bem como suas influências no crescimento e mortalidade das árvores monitoradas. Com o emprego de modelos de predição baseados em autômatos celulares, a partir de apenas uma medição inicial de uma parcela do plantio, é possível fazer estimativas precisas acerca das medidas dendométricas das árvores, bem como da produção individual e total das parcela em tempos futuros (ex: no tempo de corte estabelecido). Entretanto,

a construção desses modelos de estimação, bem como a configuração adequada dos seus parâmetros, a fim de representar a dinâmica observada nos dados do inventário florestal, não é uma tarefa trivial. Esse processo costuma demandar um tempo considerável e exigir uma grande expertise sobre o fenômeno modelado.

Nesse contexto, este trabalho visa desenvolver um modelo de árvores individuais baseado em autômato celular para representar espacialmente a dinâmica do crescimento das árvores ao longo dos anos e estimar, de forma precisa, o volume de produção de uma floresta plantada de Tachi Branco. Nesse processo, pretende-se utilizar técnicas de aprendizado de máquina para construir a regra de transição do AC a partir dos dados históricos de inventários florestais, possibilitando que o modelo se adapte às características da área monitorada e, conseqüentemente, reduza o erro na predição do volume produzido na época do corte das árvores.

1.1 Motivação

Tradicionalmente, estimar a produção de uma área de manejo é baseado em inventários florestais que demandam um tempo considerável na medição de árvores e possui um custo elevado. Visando tornar esse processo mais ágil e barato, diversas pesquisas têm investigado a adoção de modelos para predizer o volume total de madeira produzida.

Com o emprego de autômatos celulares nesse tipo de abordagem, a partir de apenas uma medição inicial de uma parcela do plantio, é possível fazer estimativas precisas acerca das medidas dendométricas das árvores, bem como da produção total de madeira dessa parcela em tempos futuros, por exemplo, no momento do corte. Os ACs são capazes de representar comportamentos dinâmicos complexos a partir de computações simples e interações locais entre seus agentes (computação emergente). Entretanto, a definição de uma regra de transição adequada para reproduzir o comportamento desejado nem sempre é uma tarefa fácil, que costuma demandar um tempo considerável e um conhecimento especializado prévio sobre o fenômeno modelado. Nesse contexto, o uso de técnicas de aprendizado de máquina se torna uma alternativa interessante, pois permite mapear os padrões escondidos nos dados históricos dos inventários florestais. Então, os modelos de regressão gerados podem ser usados como regra de transição do AC para representar as mudanças nos atributos das árvores (estado das células) ao longo do tempo. Além de facilitar o processo de construção da regra de transição do AC, a adoção de métodos de aprendizado de máquina provê modelos mais genéricos, possibilitando sua adaptação a diferentes regiões e espécimes.

Geralmente, técnicas mais complexas são empregadas na construção de modelos de alta precisão. Entretanto, os modelos gerados costumam ser do tipo "caixa-preta", onde o conhecimento usado na tomada de decisão fica intrínseco, dificultando o seu entendimento pelo usuário (FARIA; NASCIMENTO; MARTINS, 2021). Por outro lado, técnicas que

visam modelos mais interpretáveis, tipicamente são menos precisas. Portanto, encontrar uma solução equilibrada entre precisão e interpretabilidade é um desafio de pesquisa. Na literatura, várias pesquisas têm empregado regressões lineares para prever as variáveis dendométricas de cada árvore e modelos baseados rede neural artificial ou regressão logística para estimar a probabilidade de mortalidade de cada árvore no próximo instante de tempo da simulação. Neste trabalho, também foram utilizadas as regressões linear e logística, devido ao baixo custo computacional de treinamento e a capacidade de gerar regras de transição simples e que facilitam a interpretação pelos especialistas, sem comprometer significativamente a precisão das estimações do modelo desenvolvido.

Nosso modelo baseado em autômato celular foi aplicado para estimar a produção de reflorestamentos de “Tachi Branco”. Essa espécie é nativa da região amazônica e vem sendo bastante investigada devido ao seu potencial de produção de biomassa para suprir a demanda energética local. Além de estar adaptada às condições ambientais da região, esse tipo de árvore apresenta um rápido crescimento e produz uma madeira de qualidade para fins bioenergéticos (SILVA et al., 2021).

1.2 Objetivos

O objetivo principal desta pesquisa é a construção de um modelo de árvores individuais baseado em autômato celular, que seja capaz de representar a dinâmica de crescimento e interações das árvores em uma floresta de “Tachi Branco”, de modo a prover estimações precisas do volume total de madeira produzida. Nesse sentido, o trabalho foi conduzido de modo a atingir os seguintes objetivos específicos:

- Desenvolvimento de um modelo de árvore individual baseado em um autômato celular probabilístico para prever com precisão a produção de madeira ao longo do ciclo de corte.
- Criação de uma nova forma de calcular o índice de competição utilizado no modelo de estimação, considerando apenas as relações locais entre árvores vizinhas, caracterizando, assim, o princípio de interação local que deve estar presente em modelos baseados em autômato celular.
- Adoção de uma abordagem baseada em aprendizado de máquina para a construção automática e adaptativa da regra de transição usada no autômato celular a partir de dados históricos de inventário florestal, sem a necessidade de conhecimento prévio especializado. Nesse processo, foram analisadas diferentes técnicas de aprendizado de máquina a fim de escolher aquela com o melhor custo-benefício entre a precisão das estimações, o custo de treinamento e a simplicidade dos modelos gerados, que é uma característica importante para a regra de transição de um AC.

- ❑ Construção de modelos de regressão interpretáveis, permitindo compreender os atributos relevantes para as previsões e, conseqüentemente, sua validação por especialistas.
- ❑ Validação da abordagem proposta na estimação do volume total de madeira produzida em parcelas de “Tachi Branco”, comparando os resultados obtidos com aqueles registrados nos inventários florestais de plantações na região norte do Brasil.
- ❑ Análise da influência do espaçamento de plantio e da quantidade de fustes da árvore nas predições, verificando se o modelo é capaz de mapear a distinção das capacidades produtivas com base nessas características.

1.3 Organização da Dissertação

O restante desta dissertação está organizada da seguinte forma:

- ❑ Capítulo 2: introduz os principais conceitos referentes aos modelos de estimação e crescimento de árvores, bem como descreve os princípios e o funcionamento básico dos autômatos celulares e das técnicas de aprendizado de máquina utilizadas na construção da regra de transição do AC.
- ❑ Capítulo 3: apresenta os trabalhos encontrados na literatura que possuem alguma relação com a pesquisa desenvolvida, quer seja na investigação de autômatos celulares para representar modelos de crescimento de vegetação ou no uso de aprendizado de máquina para estimar a mortalidade ou a produção de madeira em florestas de manejo.
- ❑ Capítulo 4: descreve a abordagem desenvolvida em nossa proposta, detalhando as configurações e estratégias de modelagem aplicadas para atingir os objetivos propostos.
- ❑ Capítulo 5: apresenta os resultados dos experimentos realizados com o modelo desenvolvido, bem como as respectivas análises descritivas e de desempenho dessa abordagem.
- ❑ Capítulo 6: discute as principais conclusões obtidas a partir dos experimentos e desenvolvimento do trabalho, bem como algumas propostas para a continuidade da pesquisa.

Fundamentação Teórica

O presente trabalho consiste em utilizar um autômato celular, cuja dinâmica é ajustada a partir de métodos de aprendizagem de máquina, para prever e evolução da produção de madeira em florestas de manejo. Basicamente, esse modelo determina quais árvores irão sobreviver a cada passo de tempo e estima o seu crescimento.

Neste capítulo, é apresentado um embasamento teórico sobre os métodos e modelos utilizados nesta pesquisa. Inicialmente, é dada uma introdução sobre modelos de estimação do crescimento de árvores. Em seguida, são apresentados os conceitos básicos referentes aos autômatos celulares, englobando as definições de espaço celular, vizinhança e regra de transição, além das principais características dos ACs determinísticos e probabilísticos. Por fim, as técnicas de aprendizado de máquina usadas na construção da regra de transição do AC, a saber: regressão linear múltipla e regressão logística, são descritas.

2.1 A espécie Tachi-branco

O tachi-branco (*Tachigali vulgaris*) é uma espécie de árvore nativa da Amazônia, que ocorre no Brasil, Guiana, Suriname e Venezuela. É uma árvore de porte médio a grande, podendo atingir até 40 metros de altura. A copa é densa e piramidal, e a casca é lisa e cinza.

O Tachi-branco é uma espécie pioneira, que se desenvolve em áreas abertas, como clareiras e capoeiras. É uma espécie tolerante a solos pobres e ácidos, e tem um rápido crescimento. A madeira do tachi-branco é de boa qualidade, e é utilizada na construção civil, na fabricação de móveis e na produção de carvão vegetal. A espécie também é importante para a fauna, pois fornece alimento e abrigo para diversas espécies de animais.¹

A capacidade de adaptação do taxi-branco a condições desfavoráveis de fertilidade do solo, seu rápido crescimento, produção abundante de serapilheira e habilidade para fixar

¹ Foto retirada da reportagem do Portal G1, publicada em 02/11/2021 (disponível em: <https://g1.globo.com/mg/triangulo-mineiro/noticia/2021/11/02/docentes-do-curso-de-engenharia-florestal-da-ufu-desenvolvem-pesquisas-sobre-o-tachi-branco.ghtml>)

nitrogênio são atributos que o qualificam para a recuperação de áreas degradadas, bem como para o enriquecimento de capoeiras e a integração em sistemas agroflorestais na região amazônica. Além disso, em plantios homogêneos, revela-se uma opção viável para a produção de carvão vegetal, conforme indicado por estudos anteriores.

Comparativamente à madeira de eucalipto, a madeira do taxi-branco apresenta características semelhantes em termos de poder calorífico da lenha e do carvão, bem como nos rendimentos de peso e volume no processo de carbonização. Sua adaptação à xenogamia destaca a importância dos insetos, especialmente apoídeas e dípteras, na reprodução da planta. O odor adocicado das flores, perceptível a uma distância considerável, serve como mecanismo fundamental de atração para os polinizadores.

A espécie, tipicamente melitófila não seletiva, oferece néctar e pólen, sendo este último o recurso mais buscado pelos visitantes e polinizadores ao longo do dia. Com um alto índice de frutificação em condições naturais, o taxi-branco se destaca como uma espécie crucial para a recuperação da cobertura vegetal em áreas afetadas pela ação humana. Em suma, suas características biológicas e ecológicas o posicionam como uma valiosa contribuição para a sustentabilidade ambiental e econômica na região de interesse.



Figura 1 – Árvore de Tachi-branco (Fonte: arquivo pessoal de Álvaro Soares)¹

2.2 Modelos de Crescimento das Árvores

O crescimento de uma árvore é influenciado pela sua capacidade de expansão, associada ao seu potencial biológico, sua atividade fotossintética, seu metabolismo e absorção de nutrientes; bem como pelas restrições impostas por fatores externos, como competição, restrição de recursos e doenças, ou por mecanismos de autoregulação e envelhecimento (GOES et al., 2013).

Os modelos de crescimento desempenham um papel crucial na previsão das produções futuras e na exploração de diversas opções de manejo florestal, permitindo aos pesquisadores e gestores entender as consequências imediatas das decisões da gestão florestal sobre o crescimento das árvores e a produção das áreas exploradas. Tais modelos devem prever a dinâmica florestal ao longo do tempo, considerando crescimento, mortalidade, reprodução e outras mudanças sob diferentes cenários de gestão ou influências ambientais. Eles são ferramentas importantes para o planejamento e controle de políticas de manejo florestal (VANCLAY, 1994).

A classificação dos modelos de crescimento varia de acordo com seus objetivos e níveis de complexidade, podendo consistir em equações simples ou em sistemas de submodelos inter-relacionados que formam um sistema de simulação. Todos os modelos devem ser biologicamente realistas, baseando-se em informações empíricas e conhecimento dos processos de crescimento, para refletir as variações nas condições ecológicas da região de interesse (PORTE; BARTELINK; DREYFUS, 2001). A seleção de um modelo depende dos objetivos específicos e da utilidade prática desejada, reconhecendo que a validade do modelo está alinhada com seus propósitos e os objetivos dos potenciais usuários. A abordagem de modelagem deve ser orientada pelos dados, recursos e conhecimentos disponíveis, resultando em um modelo objetivo, bem documentado e acessível (GOES et al., 2013).

Os modelos de crescimento são classificados com base nas variáveis utilizadas para simular o processo de crescimento. Essa simulação pode ser direcionada para o comportamento de cada árvore individualmente ou para o comportamento do povoamento como um todo, a partir da extrapolação para uma unidade de área. Os modelos de árvore consideram parâmetros individuais de cada árvore monitorada, como diâmetro e altura, e podem ser dependentes ou independentes da distância entre árvores (AVERY; BURKHART, 2015). Por outro lado, os modelos de povoamento descrevem os parâmetros que regem o comportamento geral do povoamento, como área basal e altura dominante, sendo mais simples, robustos e adequados para prever o crescimento do povoamento como um todo. No contexto deste trabalho, optou-se por desenvolver um modelo de árvore individual independente da distância, visando explorar as relações locais entre as árvores de uma parcela e, assim, simular com mais precisão os diferentes cenários presentes no inventário florestal de Tachi Branco investigado.

2.2.1 Funções de Crescimento

O estudo da evolução das variáveis dendrométricas ao longo do tempo requer dados de pelo menos dois períodos sucessivos, que podem ser obtidos de parcelas permanentes, ensaios com delineamento, inventários florestais contínuos ou análises de tronco (GOES et al., 2013). Equações de produção podem ser derivadas da integração de equações de crescimento, permitindo também a determinação da produção por meio da projeção sequencial de crescimento e acumulação dos incrementos ao longo dos anos (VANCLAY, 1994).

A seleção de funções de crescimento é crucial para a estimação da produção florestal (TOMÉ; PAULO, 2009). Na modelagem de organismos vivos, incluindo o ciclo de vida das árvores, as funções de crescimento seguem um curva sigmoideal, assíntota e com ponto de inflexão inicial, a qual é caracterizada por um início exponencial relativamente rápido; seguido por uma fase de estabilidade linear, na qual o fator crescimento é mantido por períodos iguais de tempo; e termina com um período onde o crescimento não é relevante (GOES et al., 2013). O formato dessa curva depende da espécie e das variáveis modeladas e, embora o crescimento seja influenciado pelas variações no ambiente, ele costuma apresentar um padrão estável a longo prazo. Portanto, é essencial buscar por equações que representem os padrões de crescimento presentes nos dados dos inventários florestais, identificando as tendências relevantes e minimizando o efeito de eventos acidentais.

A forma diferencial das equações de crescimento, com o incremento temporal como variável dependente e a idade da árvore como variável independente, facilita a compreensão do processo de crescimento e a decomposição das equações em componentes de expansão e declínio. Funções de crescimento diversas são utilizadas na modelagem florestal, mas muitas podem ser expressas em duas formas básicas: uma componente anabólica proporcional ao logaritmo da idade (*LTD*, *log-time decline*) ou diretamente proporcional à idade (*TD*, *time decline*).

No presente trabalho, as funções de crescimento obtidas serão obtidas a partir do treinamento de modelos de regressão linear, partindo do pressuposto que podemos definir modelos a partir de dados históricos sem conhecimento prévio do pesquisador ou modelos já previamente definidos na literatura da área.

2.2.2 Modelos de Árvores Individuais

Nos modelos em nível de árvores individuais, a abordagem considera cada árvore como a unidade fundamental para a modelagem, oferecendo detalhes sobre a dinâmica do crescimento do povoamento em termos de seus elementos individuais. Esses modelos consistem em submodelos, sendo os mais importantes aqueles que estimam a probabilidade de morte/sobrevivência e o crescimento das variáveis dendrométricas, como altura e diâmetro, de cada árvore. Outras características, como dimensões das copas, também

podem ser consideradas, bem como a previsão de entrada de novas árvores no modelo, quando no contexto de florestas naturais (HASENAUER, 2000).

Considerando os modelos de crescimento para florestas de manejo, a previsão do crescimento dimensional em uma idade futura pode ser derivada a partir de valores atuais de cada variável e de fatores como índice de competição, idade, índice de local, entre outros. Também é possível adotar a modelagem do crescimento com base no potencial de crescimento, muitas vezes avaliado usando árvores em crescimento livre de competição ou aquelas que são dominantes. O crescimento de árvores em condições de competição é determinado por meio de uma função de modificação do crescimento potencial, que reduz o crescimento das árvores em função da disputa com outras árvores pelos recursos (interação local). Para essa função, índices de competição e variáveis do povoamento são usados para ajustar o crescimento das árvores sujeitas à competição (VANCLAY, 1994).

Segundo Goes et al. (2013), para uma completa modelagem do crescimento de florestas, além de mapear o crescimento das variáveis dendométricas, também é necessário representar as alterações em sua composição, tanto em relação ao ingresso de novos exemplares, quanto ao desaparecimento de árvores por mortalidade ou ações silvícolas. A estimativa da mortalidade em modelos de crescimento e produção florestal é uma tarefa desafiadora (HAMILTON, 1986; YANG; TITUS; HUANG, 2003), influenciada por fatores como competição, genética, envelhecimento e fenômenos atmosféricos (ex: seca, geada e vento) (PEET; CHRISTENSEN, 1987); (CAMPOS; LEITE, 2006). A probabilidade de mortalidade ou sobrevivência de uma árvore é frequentemente utilizada para estimativas (SOMERS et al., 1980), sendo comumente modelada em relação a índices de competição (HASENAUER; MERKL; WEINGARTNER, 2001) e variáveis dimensionais e de povoamento (BUCHMAN; PEDERSON; WALTERS, 1983). As funções Weibull e Logística, assim como o modelo proposto por (WEST, 1981), são amplamente empregadas nesse tipo de análise (SCHNEIDER et al., 2005). Recentemente, redes neurais artificiais também têm sido aplicadas para melhorar a precisão das estimativas de mortalidade, especialmente em povoamentos inequiâneos (MERKL; HASENAUER, 1998; WEINGARTNER; MERKL; HASENAUER, 2000; HASENAUER; MERKL; WEINGARTNER, 2001).

A maioria das abordagens relacionadas aos modelos em nível de árvores individuais adota equações lineares e não-lineares para estimar as variáveis em seus submodelos (MARTIN; EK, 1984; TOMÉ; BURKHART, 1989; HYNYNEN, 1995; MABVURIRA; MIINA, 2002; FOX; BI; ADES, 2007; MONTY; LEJEUNE; RONDEUX, 2008; TOMÉ; PAULO, 2009). Diversos modelos de regressão potencial para a modelagem de variáveis dimensionais das árvores (altura, diâmetro e volume) são apresentados por (ZEIDE, 1993), que discute a aplicabilidade flexível de cada um deles na modelagem de árvores individuais. Para conduzir a modelagem em nível de árvores individuais, é essencial a existência de medições repetidas das dimensões das árvores ao longo do tempo. Portanto, é imperativo dispor de dados provenientes de parcelas permanentes, que proporcionam

um histórico completo do desenvolvimento do povoamento, com rigoroso controle de identificação (UZOH; OLIVER, 2008).

2.2.3 Índices de Competição

A competição é caracterizada pelo impacto negativo no crescimento de exemplares da área monitorada, devido à disputa local entre árvores vizinhas pelos recursos disponíveis, como luz, água e nutrientes (KEDDY, 1989). O índice de competição é usado para quantificar a intensidade competitiva e pode ser determinado por abordagens independentes ou dependentes da distância. Índices independentes da distância (IID) diferem dos dependentes (IDD) pelo fato de não incorporarem medidas espaciais na avaliação da competição. Uma variedade de índices é aplicada para diferentes tipos de povoamentos, espécies e estágios de desenvolvimento (DANIELS; BURKHART; CLASON, 1986; TOMÉ; BURKHART, 1989). Estudos têm demonstrado que não há um índice de competição claramente superior aos demais. A seleção do índice mais adequado dependerá das informações disponíveis e da avaliação de cada índice nas estimativas dos submodelos. Resultados divergentes quanto ao desempenho dos índices em florestas com formação semelhante têm sido observados (BIGING; DOBBERTIN, 1992).

As principais métricas de competição que não dependem da distância foram detalhadas por (GLOVER; HOOL, 1979):

$$IID_1 = \frac{dap_i^2}{\bar{D}^2} \quad (1)$$

$$IID_2 = \frac{Ht_i}{\bar{H}t} \quad (2)$$

$$IID_3 = \frac{dap_i^2 Ht_i}{\bar{D}^2 \bar{H}t} \quad (3)$$

$$IID_4 = \frac{dap_i^2}{q^2} \quad (4)$$

Em que o dap_i é o diâmetro com casca (DAP) da árvore-referência, medido a 1,30 metros do solo; \bar{D} é a média aritmética dos diâmetros das árvores; Ht_i é a altura total da árvore-referência; $\bar{H}t$ é a altura média das árvores e q é o diâmetro quadrático. As medidas de diâmetro são geralmente dadas em centímetros, enquanto que as alturas normalmente estão em metros.

Neste trabalho, é proposto um cálculo do índice de competição considerando apenas as interações na vizinhança de cada árvore, em conformidade com o princípio de localidade presente em um autômato celular. Essa estratégia diverge das equações tradicionais apresentadas acima e em outros trabalhos, as quais consideram todas as árvores da parcela na computação dos índices de competição.

2.3 Autômatos celulares

Em problemas que envolvem a modelagem e simulação de sistemas dinâmicos complexos e computacionalmente caros, os autômatos celulares vem sendo largamente utilizados como alternativa aos sistemas de equações diferenciais (TOFFOLI, 1984). Tipicamente, um autômato celular é um sistema dinâmico e discreto no tempo, espaço e estado, capaz de representar e simular comportamentos complexos baseados em um conjunto de regras simples, tornando mais fácil o entendimento do modelo e o comportamento evolutivo do fenômeno modelado (TOFFOLI; MARGOLUS, 1987).

2.3.1 Breve Histórico

John von Neumann e Stanislaw Ulam, dois renomados matemáticos, apresentaram a teoria de autômatos celulares pela primeira vez em um trabalho conjunto, na década de 1940. Ulam trabalhava com o crescimento de cristais utilizando uma rede bidimensional simples e, von Neumann estudava a auto-reprodução. Ulam recomendou a von Neumann adotar uma equação matemática semelhante à que ele utilizava e, desta forma, descobriram que uma auto-reprodução simulada poderia ser descrita por meio de um algoritmo simples (NEUMANN; BURKS et al., 1966).

O modelo de autômato celular mais conhecido atualmente é o “Jogo da Vida”, apresentado por John Horton Conway na década de 1970 (CONWAY et al., 1970). Esse modelo reproduz através de algumas regras simples, alterações em um agrupamento de células, as quais cada uma delas é apontada como morta ou nascida de acordo com sua vizinhança.

Os trabalhos de Wolfram na década de 80, que investigaram o comportamento dinâmico dos Autômatos Celulares (ACs) e foram publicados em várias ocasiões (1983a, 1983b, 1984a, 1984b, 1988), tornaram-se uma referência fundamental para todos os estudos subsequentes nessa área. Embora a regra de transição do Life seja notavelmente simples, consistindo em uma estrutura bidimensional com uma vizinhança de nove células, isso não permitiu uma análise completa do espaço de regras. Uma das contribuições mais significativas de Wolfram, que impactou significativamente a direção das pesquisas nessa área, foi a demonstração de que até mesmo os modelos mais simples de ACs - aqueles unidimensionais, binários e com uma vizinhança formada por apenas três células - eram capazes de gerar padrões intrigantes e comportamentos emergentes. Além disso, havia apenas 256 regras possíveis para ACs desse tipo, o que possibilitou uma análise exaustiva de todas elas. Um dos resultados mais notáveis dos estudos de Wolfram (1984a) foi a introdução de um esquema de classificação que ele propôs, categorizando os ACs com base em seus comportamentos dinâmicos. Mais tarde, em 2002, Wolfram lançou o muito aguardado livro *A New Kind of Science*, no qual ele fundamentou suas ideias polêmicas com base em experimentos que ele próprio conduziu com ACs.

Desde então, várias pesquisas têm investigado o uso de autômatos celulares para representar sistemas e fenômenos complexos nas mais diferentes áreas, tais como: robótica autônoma (SOUZA; LIMA, 2020), sistemas epidemiológicos (GOMES, 2023), criptografia (SILVA; SOARES; LIMA, 2016), propagação de incêndios (LIMA; LIMA, 2014), dinâmica de pedestres (GONÇALVES; LIMA, 2020), entre outros.

2.3.2 Espaço Celular

Os autômatos celulares são constituídos por um conjunto de células que são organizadas espacialmente em um reticulado N -dimensional (ex: 1D ou linear, 2D ou planar, 3D, etc.), conforme exemplificado na Figura 2. Nessa organização, cada célula de um autômato celular está conectada àquelas que estão dentro de um raio r de distância, formando sua vizinhança.

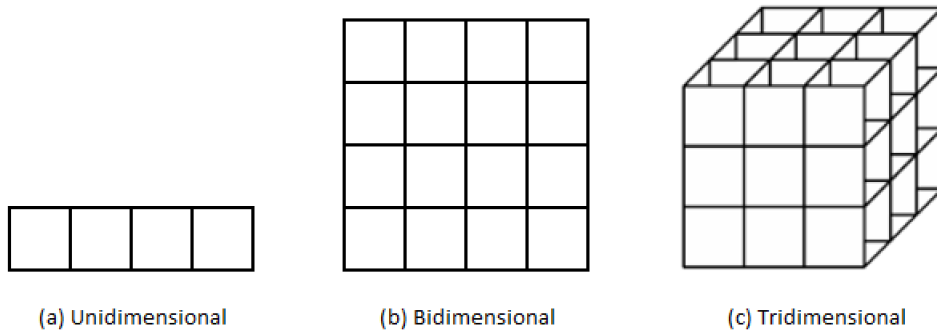


Figura 2 – Tipos de dimensionalidade dos autômatos celulares. (FRAGA; OLIVEIRA; MARTINS, 2020)

Em outras palavras, o reticulado de um AC, também conhecido como espaço celular, é composto por um conjunto de células distribuídas espacialmente e seguindo uma mesma estrutura de conexão com seus vizinhos mais próximos. Cada célula possui estados pré-definidos (morte ou vida, por exemplo) e a atualização desses estados ocorre a cada interação no tempo em um intervalo discreto, de acordo com a regra de transição que representa a dinâmica modelada. Essa regra é aplicada em todo o reticulado e, o estado de cada célula no instante de tempo $t + 1$ é calculado com base em seu próprio estado e nos estados das células vizinhas no tempo t . Uma célula pode ter vários tipos de formato, conforme ilustrado na Figura 3. Entretanto, todas as células de um reticulado devem ter o mesmo formato.

Como o reticulado possui tamanho finito, as células das regiões de bordas (fronteiras), que delimitam o espaço celular, demandam uma estratégia diferenciada para a determinação da vizinhança considerada na aplicação da regra de transição. Basicamente, existem três formas diferentes de definição da vizinhança, a saber: borda periódica, cir-

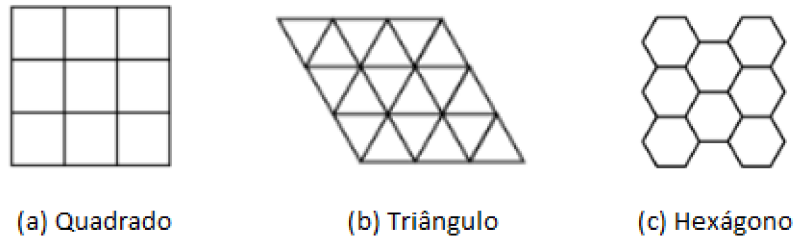


Figura 3 – Tipos de formatos de célula dos autômatos celulares. (FRAGA; OLIVEIRA; MARTINS, 2020)

cular ou cíclica (Figura 4a); borda reflexiva (Figura 4b); e borda fixa (Figura 4c). Nas bordas periódicas, considera-se que as extremidades do reticulado estão conectadas. Na borda reflexiva, considera-se os estados das células das extremidades para completar a vizinhança. Já na borda fixa, adota-se um estado pré-definido, representado na Figura 4c por um ‘X’, como valor para as células complementares da vizinhança.

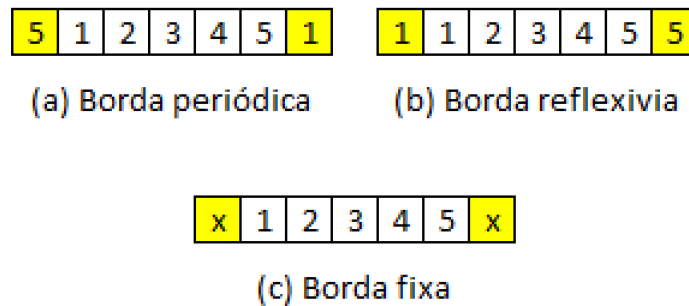


Figura 4 – Tipos de bordas dos autômatos celulares.

2.3.3 Estado e Vizinhança

Em um Autômato Celular, cada célula possui um estado definido dentro de um conjunto de possíveis valores e, a cada passo de tempo discreto, os estados de todas as células são evoluídos de acordo com a regra de transição do AC, com base em seu estado atual e de seus vizinhos. Uma célula é considerada vizinha de outra quando elas possuem uma proximidade física pré-definida, lógica ou arbitrária. Os principais tipos de vizinhança usados na evolução de autômatos celulares bidimensionais são: vizinhança de von Neumann e de Moore (SARKAR, 2000). A Figura 5 mostra dois reticulados, considerando esses dois tipos de vizinhança. A vizinhança de von Neumann é constituída pela célula

central e pelas 4 células vizinhas localizadas na vertical e horizontal, conforme ilustrado na Figura 5a. Por outro lado, a vizinhança de Moore é composta pelas oito células no entorno da célula central, ou seja, as células da vertical e horizontal, como na vizinhança de von Neumann, mais as quatro células da diagonal, conforme mostrado na Figura 5b.

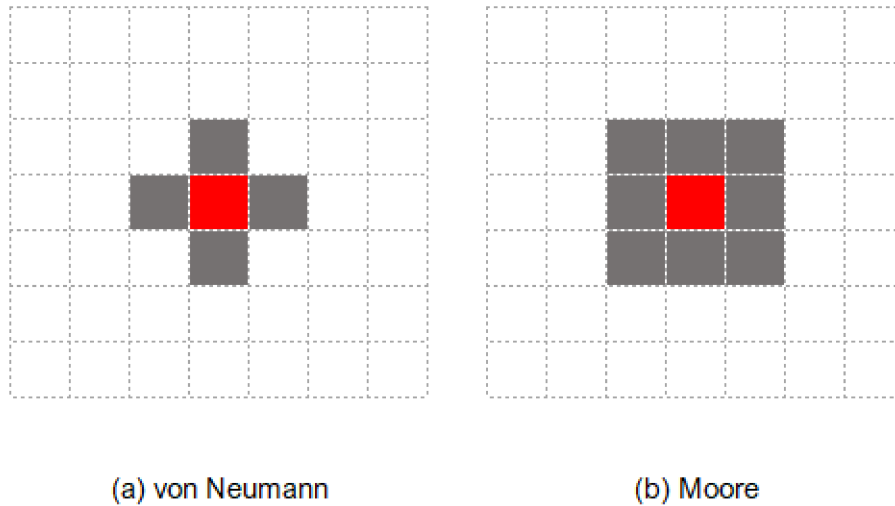


Figura 5 – Tipos de vizinhança dos autômatos celulares.

Outro fator importante na definição da vizinhança é o seu raio de abrangência (r), o qual define a distância da vizinhança em função do número de células que serão consideradas. A Figura 6 mostra a diferença nas vizinhanças (células cinzas) de uma célula central (vermelha), considerando os raios 1 e 2 nos dois tipos de vizinhança usados em ACs bidimensionais (von Neumann e Moore).

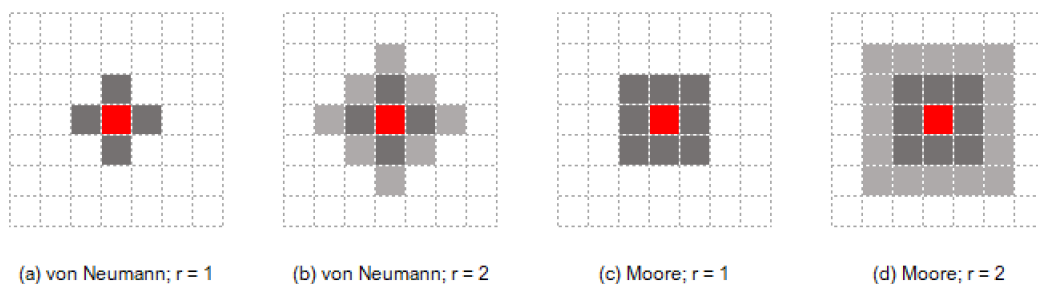


Figura 6 – Exemplos de raios para os tipos de vizinhança usados em autômatos celulares 2D.

2.3.4 Regra de Transição

A atualização dos estados das células de um reticulado ocorre por meio da aplicação da regra de transição a cada intervalo de tempo discreto (instante t) durante a evolução do AC.

Uma das formas de modelar a regra de transição de um AC é por meio de uma tabela, similar a uma tabela-verdade, que representa todas as combinações possíveis dos estados de uma vizinhança. Esse método de representação é frequentemente aplicado em autômatos celulares elementares, ou seja, aqueles compostos por reticulados unidimensionais que operam com estados binários e adotam vizinhança de raio 1. Como cada uma das células podem assumir apenas 2 valores (estados binários) e a vizinhança é formada por 3 células (célula central e suas duas vizinhas imediatas), existem 8 combinações possíveis de vizinhança ($2^3 = 8$). Para cada uma dessas combinações, é determinado o valor da célula central no instante $t + 1$. Portanto, existem 256 regras de transição que podem ser aplicadas nesse AC elementar ($2^8 = 256$).

A Figura 7 exemplifica a evolução de um autômato celular elementar. Nesse exemplo, o reticulado do AC é formado por 13 células que podem assumir o estado 0 (branco) ou 1 (preto). Nota-se que, no reticulado inicial ($t = 0$) ilustrado na Figura 7b, apenas a célula mais ao centro (posição 7) está no estado 1 e sua vizinhança, considerando raio 1, é destacada em vermelho. A figura também apresenta a evolução temporal do reticulado por 7 passos de tempo (Figura 7c), considerando a regra de transição ilustrada na Figura 7a.

A quantidade de vizinhanças possíveis cresce consideravelmente com o aumento do raio, do número de estados e, principalmente, da dimensionalidade do reticulado do AC. Esse crescimento também eleva o custo computacional (memória e tempo de execução) da representação por tabela, sendo usual a adoção de regras de transição totalísticas. Uma regra de transição é considerada totalística quando apenas a quantidade de células em determinado estado é levada em consideração no cálculo do estado da célula central no instante seguinte. Um exemplo notável de autômato celular totalístico é o Jogo da Vida (Alife). A regra de transição desse autômato é definida pelo seguinte conjunto de regras totalísticas:

- ❑ Se uma célula possui exatamente três vizinhos vivos, ela se torna uma célula viva na próxima geração.
- ❑ Se uma célula possui dois vizinhos vivos, ela permanece no mesmo estado na próxima geração.
- ❑ Caso contrário (menos de 2 ou mais de 3 vizinhos vivos), a célula morre na próxima geração.

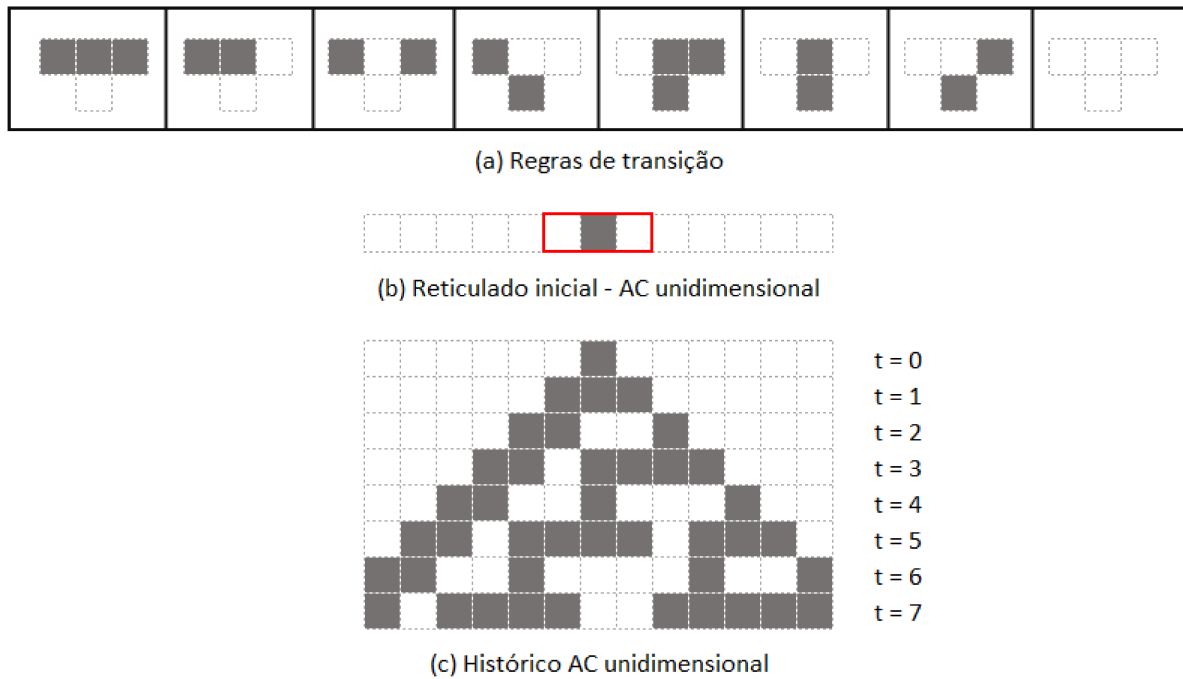


Figura 7 – Evolução de um autômato celular elementar.

Apesar de sua regra de transição simples, esse autômato é capaz de representar uma variedade de comportamentos que simulam a dinâmica de organismos vivos. A Figura 11 ilustra a evolução de uma rede utilizando essas regras por 5 etapas do autômato celular, onde as células vivas estão destacadas em preto e as células mortas em branco.

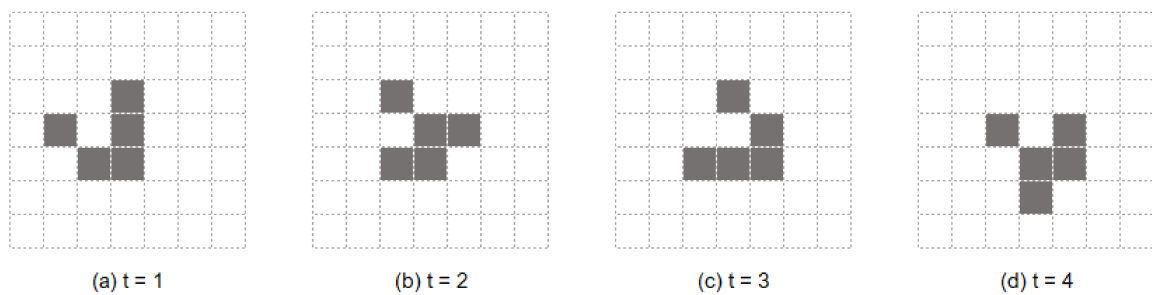


Figura 8 – Evolução de um reticulado no Jogo da Vida considerando quatro instantes de tempo.

Um autômato celular é probabilístico quando as equações que compõem sua regra de transição consideram probabilidades. Esse tipo de abordagem é particularmente útil na representação de fenômenos que lidam com fatores estocásticos, como ocorre na estimação da mortalidade das árvores pelo modelo de crescimento.

2.4 Modelos de Aprendizado de Máquina

Nesta seção, é apresentado um breve resumo teórico dos modelos de aprendizado de máquina testados e analisados neste trabalho.

Aprendizado de máquina (Machine Learning, em inglês) é um campo da inteligência artificial que se concentra no desenvolvimento de algoritmos e modelos que permitem que os computadores aprendam padrões a partir de dados. Em vez de programar explicitamente instruções para realizar uma tarefa, os modelos de Machine Learning são treinados usando dados para aprender como realizar a tarefa.

Os modelos podem ser divididos entre supervisionados e não supervisionados, onde nos modelos supervisionados, o algoritmo é treinado em um conjunto de dados rotulado, onde cada exemplo no conjunto de dados possui uma resposta desejada. O objetivo é aprender uma função que mapeia entradas para saídas, de modo que o modelo possa fazer previsões ou classificações em novos dados não rotulados. Já nos modelos não supervisionados o algoritmo é treinado em um conjunto de dados sem rótulos, e o objetivo principal é explorar a estrutura ou padrões subjacentes nos dados.

Além disso, existem diferentes tipos de modelos com propósitos diferentes. Os algoritmos de classificação são usados quando a variável de saída é uma categoria, como "sim" ou "não", "spam" ou "não spam", ou rótulos de classes como "cachorro", "gato" ou "pássaro". O modelo aprende a associar padrões nas entradas aos rótulos de classe correspondentes. Por outro lado, os modelos de regressão são usados quando a variável de saída é um valor contínuo, e em vez de prever categorias, o modelo prevê um número específico. Por exemplo, prever o preço de uma casa com base em características como número de quartos, localização, etc.

2.4.1 Rede Neural Multi-layer Perceptron (MLP)

A Rede Neural *Multi-layer Perceptron* (MLP) é um modelo de machine learning que pertence à categoria de redes neurais artificiais. É uma arquitetura composta por múltiplas camadas de neurônios, cada uma conectada à próxima. Cada neurônio em uma camada está conectado a todos os neurônios na camada seguinte, formando uma rede de camadas interconectadas.

A MLP é capaz de aprender padrões complexos e realizar tarefas de classificação e regressão. Durante o treinamento, os pesos das conexões entre os neurônios são ajustados para minimizar a diferença entre as saídas previstas e os rótulos reais dos dados de treinamento. Isso é geralmente feito usando algoritmos de otimização, como o gradiente descendente (HAYKIN, 2001).

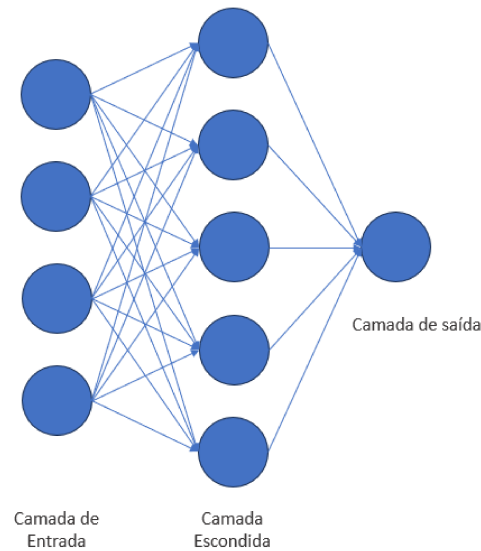


Figura 9 – Ilustração teórica de um modelo MLP de uma camada.

2.4.2 Random Forest

O modelo *Random Forest* ou Floresta Aleatória, é um modelo de machine learning que pertence à categoria de aprendizado ensemble, combinando múltiplos modelos para melhorar o desempenho preditivo. Ele é baseado em árvores de decisão, onde várias árvores são construídas durante o treinamento e suas previsões são combinadas para produzir uma saída final.

Cada árvore de decisão na Random Forest é construída a partir de uma amostra aleatória dos dados de treinamento e em diferentes conjuntos de características. Esse processo de aleatoriedade ajuda a reduzir a sobreajuste (overfitting) e aumentar a robustez do modelo. Durante a previsão, as saídas das árvores individuais são combinadas através de votação (classificação) ou média (regressão) para obter uma resposta final (BREIMAN, 2001).

A Figura 10 apresenta uma ilustração didática do funcionamento do processo de previsão de um modelo de Floresta Aleatória.

2.4.3 XGB - Extreme Gradient Boosting

O XGB (*Extreme Gradient Boosting*) é um modelo de machine learning baseado em árvores de decisão e pertence à família de algoritmos de gradient boosting. Ele é projetado para otimizar a eficiência e o desempenho do boosting, combinando várias árvores de decisão fracas para formar um modelo robusto e preciso.

O XGBoost utiliza um processo iterativo em que cada árvore de decisão subsequente

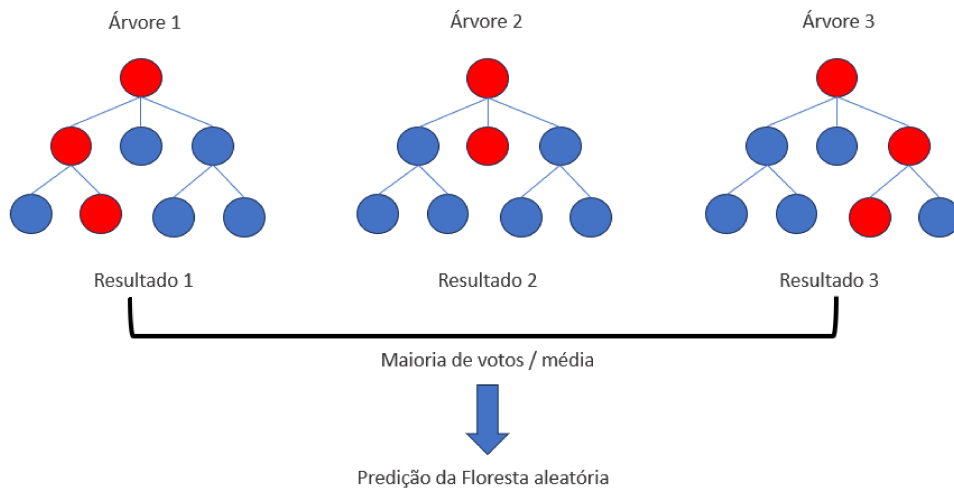


Figura 10 – Ilustração teórica de um modelo Random Forest.

corrige os erros dos modelos anteriores. Ele introduz a regularização para evitar o overfitting e otimiza uma função de perda específica, incorporando um termo de regularização nos parâmetros. Além disso, o XGBoost oferece suporte a recursos como processamento paralelo, o que melhora significativamente a eficiência computacional.

Esse modelo é amplamente utilizado em competições de ciência de dados e machine learning devido à sua eficácia em uma variedade de tarefas, como classificação, regressão e ranking (CHEN; GUESTRIN, 2016).

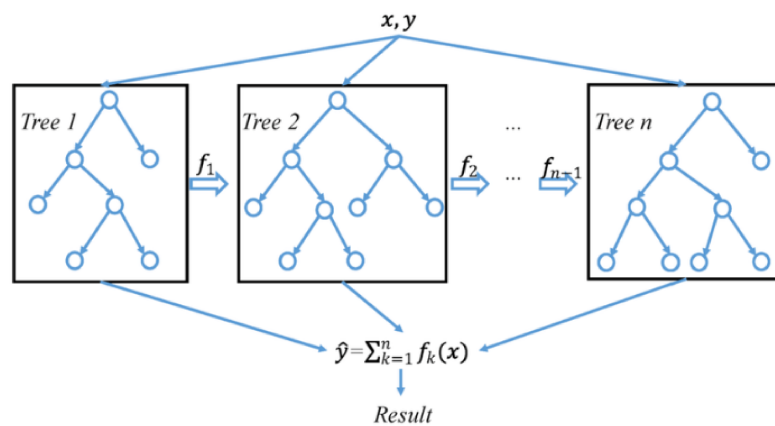


Figura 11 – Ilustração teórica de um modelo XGB. Disponível em <https://www.researchgate.net/figure/A-general-architecture-of-XGBoost_fig3_35483097>

2.4.4 Modelos de Regressão

Nos modelos de regressão em geral, estuda-se a relação de uma variável dependente Y , também conhecida como variável resposta, com outras variáveis independentes (X_1, X_2, \dots, X_p) , em que a média de Y é definida por $E(Y|X)$, sendo X o vetor que contém as variáveis preditoras.

Os dois tipos de modelos mais conhecidos e utilizados são os de regressão Linear e Logística (HOSMER; LEMESHOW, 2000; MONTGOMERY; PECK; VINING, 2021). Nos modelos lineares, a variável dependente é contínua e a estimação dos parâmetros é feita pelo método dos mínimos quadrados, enquanto que, nos modelos logísticos, a variável resposta é sempre categórica e a estimação é feita pelo método da máxima verossimilhança. Além disso, os modelos logísticos não necessitam de algumas pressuposições que os modelos lineares exigem, como por exemplo, a normalidade dos erros e a homogeneidade da variância.

2.4.4.1 Regressão Logística

Os modelos de regressão logística (HOSMER; LEMESHOW, 2000) são comumente utilizados em problemas de classificação, em situações que a variável dependente é de natureza dicotômica ou binária, enquanto as variáveis independente podem ser categóricas ou não. Por meio dessa técnica, também é possível estimar a probabilidade associada à ocorrência de um determinado evento em função do conjunto de variáveis explicativas.

Na regressão logística binária, a variável resposta Y tem distribuição Bernoulli, em que $\pi(x) = P(\text{Sucesso}) = P(Y = 1|X)$. A variável Y só assume dois valores, 0 ou 1, sendo 1 a ocorrência do evento de interesse com probabilidade $\pi(x)$ e 0 a não ocorrência do evento com probabilidade $[1 - \pi(x)]$. Nesse modelo, as variáveis independentes X_k , sendo $k = 1, 2, \dots, p$, podem ser discretas ou contínuas.

O modelo de regressão logística é definido pela expressão:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} \quad (5)$$

onde $\pi(x) = E(Y|X)$; p é o número de variáveis independentes consideradas no modelo ajustado; e o preditor linear é $\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$.

Os coeficientes $\beta_0, \beta_1, \dots, \beta_p$ são estimados pelo método da máxima verossimilhança a partir do conjunto de dados, sendo que os valores de obtidos de $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ maximizam o logaritmo da função de verossimilhança.

Pode-se observar na Figura 12 que a curva logística para a relação de uma variável Y com uma variável x_k tem um comportamento probabilístico no formato de “S”, fato esse que é algo característico dos modelos de regressão logística com coeficiente positivo (HOSMER; LEMESHOW, 2000). Há também a curva logística com formato de “S” invertido, sendo o oposto da Figura 12, que ocorre em casos onde o coeficiente β é negativo.

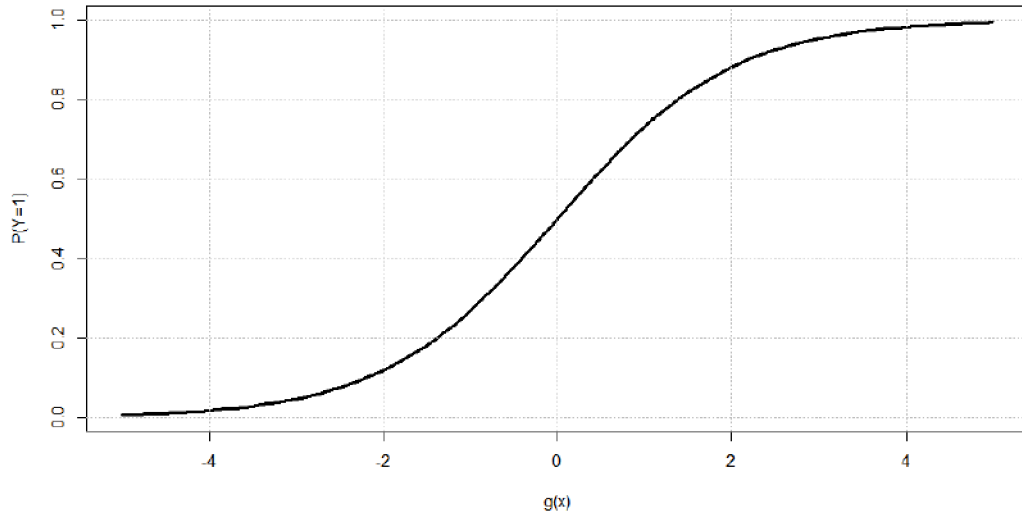


Figura 12 – Curva logística probabilística.

A interpretação dos parâmetros de um modelo de regressão logística é feita comparando a probabilidade de sucesso com a probabilidade de fracasso, utilizando a função denominada *odds ratio* ou razão de chances (OR). Calcula-se essa função a partir da função *odds*, onde:

$$g(x) = \frac{\pi(x)}{[1 - \pi(x)]} = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_i} \quad (6)$$

Desta forma, tomando dois valores distintos de uma variável explicativa, $x_{k(j+1)}$ e x_{kj} , temos:

$$OR = \frac{g(x_{k(j+1)})}{g(x_{kj})} \quad (7)$$

E, aplicando o logaritmo, obtemos:

$$\ln(OR) = \ln \left[\frac{g(x_{k(j+1)})}{g(x_{kj})} \right] = \ln [g(x_{k(j+1)})] - \ln [g(x_{kj})] = \beta_p (x_{k(j+1)} - x_{kj}) \quad (8)$$

Assim, considerando os valores das variáveis explicativas como $x_{j+1} - x_j = 1 \text{und.}$, temos que:

$$\ln(OR) = \ln(e^{\beta_k}) = \beta_k \quad (9)$$

Desta maneira, tomando como referência os valores de uma variável explicativa x_i , podemos obter a chance de sucesso de um grupo em relação a outro utilizando a função *odds ratio*, em que cada um desses dois grupos possuem valores diferentes da mesma

variável x_i , como por exemplo, homens comparados as mulheres, doentes comparados a saudáveis, entre outros.

Em um contexto de regressão logística, o Odds Ratio é frequentemente calculado para avaliar a associação entre uma variável independente e a probabilidade do evento ocorrer. Um OR igual a 1 indica que não há associação, valores acima de 1 indicam uma associação positiva, e valores abaixo de 1 indicam uma associação negativa.

Na interpretação prática, se o Odds Ratio for 2, isso significa que a chance de o evento ocorrer é duas vezes maior no grupo com a presença da variável independente em comparação com o grupo de referência. A interpretação do Odds Ratio é valiosa para entender o impacto relativo de diferentes variáveis na probabilidade do evento de interesse, contribuindo assim para a interpretação e tomada de decisões em contextos de análise de dados binários.

$$OR(\beta_k) = e^{\beta_k} \quad (10)$$

Assim, quando $\beta_p > 0$ então $OR > 1$ e, conseqüentemente, as chances de sucesso dos indivíduos x_{j+1} são maiores que as dos indivíduos x_j . Por outro lado, quando $\beta_p < 0$ então $OR < 1$, e assim as chances de sucesso dos indivíduos x_{j+1} são menores que as dos indivíduos x_j .

A estimação dos parâmetros β_i do modelo de regressão logística é feita pelo método da máxima verossimilhança. Os valores de $\hat{\beta}_i$ que maximizam o logaritmo da função de verossimilhança são os estimadores de máxima verossimilhança de β_i . Considerando uma amostra independente (x_i, y_i) de tamanho n , em que x_i é o valor da variável independente da i -ésima observação e y_i é o valor da variável resposta dicotômica, $Y_i \text{ Ber}(\pi_i)$, isto é, Y tem distribuição de probabilidade Bernoulli. A distribuição de probabilidade de Y_i é dada por:

$$f(y_i, \pi_i) = \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (11)$$

Dado que as observações Y_i são independentes, a função de verossimilhança é dada pela seguinte expressão:

$$L(\beta) = \prod_{i=1}^n f(y_i, \pi_i) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (12)$$

onde, $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$ é o vetor dos parâmetros do modelo, e y_i é 0 e 1.

Deseja-se determinar os estimadores $\hat{\beta}_i$ que maximiza $L(\beta)$ por meio da máxima verossimilhança. Portanto, aplicando o logaritmo na expressão (12) tem-se:

$$l(\beta) = \ln(L(\beta)) = \ln \left(\prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \right) = \sum_{i=1}^n \left[y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) + \ln(1 - \pi_i) \right] \quad (13)$$

Substituindo π_i e $1 - \pi_i$ da equação (13) por suas respectivas expressões, tem-se:

$$l(\beta) = \sum_{i=1}^n \left[y_i \ln(\beta_0 + \beta_1 x_i + \dots + \beta_p x_i) + \ln \left(\frac{1}{1 + \exp(\beta_0 + \beta_1 x_i + \dots + \beta_p x_i)} \right) \right] \quad (14)$$

$$l(\beta) = \sum_{i=1}^n [y_i \ln(\beta_0 + \beta_1 x_i + \dots + \beta_p x_i) - \ln(1 + \exp(\beta_0 + \beta_1 x_i + \dots + \beta_p x_i))] \quad (15)$$

Derivando a expressão (15) em relação a cada parâmetro do modelo, serão obtidas $p+1$ equações de máxima verossimilhança para os $p+1$ parâmetros que deseja-se estimar. Essas equações podem ser expressas por:

$$\sum_{i=1}^n y_i \ln[\pi(x_i)] = 0 \quad (16)$$

e

$$\sum_{i=1}^n x_{ij} \ln[y_i - \pi(x_i)] = 0 \quad (17)$$

As equações de verossimilhança para a regressão logística, são não-lineares em β , logo, necessita-se de métodos especiais para solucioná-las. Tais métodos são iterativos e requerem auxílio computacional disponíveis na maioria dos softwares estatísticos, sendo o mais comum deles o método de estimação de Newton-Raphson.

2.4.4.2 Regressão Linear

Os modelos de regressão linear são ferramentas amplamente utilizadas para modelar e entender a relação entre variáveis dependentes e independentes. Essa abordagem estatística é amplamente empregada em diversas áreas, como ciências sociais, economia, engenharia, biologia e muitas outras, para analisar e prever como uma variável responde a mudanças em outras variáveis (MONTGOMERY; PECK; VINING, 2021).

Em um modelo de regressão linear, a variável dependente é expressa como uma combinação linear das variáveis independentes, juntamente com um termo de erro aleatório. A forma mais simples é a regressão linear simples, em que uma única variável independente é utilizada para prever a variável dependente. No entanto, os modelos de regressão linear podem ser estendidos para incluir múltiplas variáveis independentes, conhecidos como regressão linear múltipla. A equação geral de um modelo de regressão linear é:

$$Y = \beta_0 + \beta_1 \times_1 + \beta_2 \times_2 + \dots + \beta_n \times_n + \epsilon \quad (18)$$

onde, Y é a variável dependente que estamos tentando prever, X_1, X_2, \dots, X_n são as variáveis independentes, $\beta_1, \beta_2, \dots, \beta_n$ são os coeficientes que representam o efeito das variáveis independentes na variável dependente e ϵ é o erro aleatório, que captura o componente não explicado pelo modelo.

A determinação dos coeficientes $\beta_1, \beta_2, \dots, \beta_n$ é feita através de técnicas de ajuste, sendo a mais comum o método dos mínimos quadrados, que visa minimizar a soma dos quadrados das diferenças entre os valores observados e os valores previstos pelo modelo.

O método dos mínimos quadrados é uma técnica fundamental na estatística e na análise de regressão, usada para estimar os coeficientes de um modelo de regressão linear, a fim de encontrar a “melhor” linha de ajuste que minimize a soma dos quadrados das diferenças entre os valores observados e os valores previstos pelo modelo. Isso cria uma linha que melhor se ajusta aos dados disponíveis.

O procedimento dos mínimos quadrados envolve as seguintes etapas:

- **Formulação do modelo:** inicialmente, é necessário estabelecer o modelo de regressão linear que relaciona as variáveis independentes e dependentes, conforme a equação geral mencionada anteriormente.
- **Cálculo das previsões:** com o modelo definido, são calculadas as previsões para a variável dependente utilizando as variáveis independentes e os coeficientes a serem estimados.
- **Cálculo dos resíduos:** os resíduos são as diferenças entre os valores observados e os valores previstos. Ou seja, $\epsilon_i = y_i - \hat{y}_i$, onde ϵ_i é o resíduo do i –ésimo ponto, y_i é o valor observado e \hat{y}_i .
- **Minimização da soma dos quadrados dos resíduos:** o objetivo dessa etapa é minimizar a soma dos quadrados dos resíduos, ou seja, encontrar os coeficientes que tornam essa soma tão pequena quanto possível. Matematicamente, isso é representado por:

$$\sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (19)$$

Para minimizar a soma dos quadrados dos resíduos, os coeficientes são ajustados de forma a encontrar os valores que minimizam essa soma. Isso pode ser feito através de cálculos de derivadas parciais e equações normais. Uma vez que os coeficientes são estimados, eles fornecem informações sobre o impacto relativo das variáveis independentes na variável dependente. É importante avaliar o ajuste do modelo aos dados, usando métricas como o coeficiente de determinação (R^2).

Revisão da Literatura

As empresas vêm investindo tempo e recursos no desenvolvimento de modelos capazes de estimar o crescimento das florestas, a fim de auxiliá-los no planejamento. Vários trabalhos vêm investigando o uso de autômatos celulares e técnicas de aprendizagem de máquina na construção de modelos de estimação precisos.

Pesquisas têm utilizado AC para construir modelos capazes de prever o crescimento florestal. Bandini e Pavesi Bandini e Pavesi (2004) argumentam que o crescimento e comportamento das árvores plantadas em determinadas áreas depende de vários fatores, como: luz solar, chuvas, temperaturas e como as árvores competem entre si pelos recursos naturais. Eles utilizam um modelo discreto de autômato celular bidimensional e investigam várias espécies de árvores como carvalho e pinheiros, localizados aos pés dos alpes italianos. Os autores fizeram as simulações do modelo considerando vizinhanças de Von Neumann e Moore e utilizando variáveis como água, luz, nitrogênio e potássio. As variáveis individuais das árvores consideradas foram: tamanho, quantidade de recursos necessárias para a árvore continuar a viver, quantidade de recursos (volume/fruto), sementes produzidas. As simulações reproduzindo populações de carvalho e pinheiros que vivem no sopé dos alpes italianos mostraram resultados qualitativamente semelhantes a estudos de caso reais.

Em (GARCÍA-DURO et al., 2018), os autores utilizam um modelo discreto baseado em autômato celular bidimensional que simula a reconstrução de uma floresta de árvores de diferentes espécies pós incêndio. Eles fizeram as simulações com a vizinhança de Moore e consideraram variáveis como: estrutura da planta, comportamento pós incêndio, topografia, clima, ciclo das plantas, comportamento de dispersão. O valor médio da população pré-fogo cerca de 3 anos após o fogo na validação do subconjunto foi de $88,1 \pm 2,7$ (média \pm desvio padrão) e o erro quadrado foi de 5,16. O erro das simulações de validação foi relativamente baixo tendo em conta a grande variação dos dados.

Vários estudos também investigaram o uso de técnicas de aprendizado de máquina na construção de modelos de estimativa precisos. Modelos de crescimento individual são propostos em (TÉO, 2022), os quais são baseados em métodos de equações diferenciais e

regressão logística para realizar previsões do crescimento e produção florestal de plantios desbastados e não desbastados de *Pinus taeda* L. na região Meio Oeste do estado de Santa Catarina. As equações foram utilizadas para prever as variáveis dendrométricas (diâmetro e altura) e probabilidade de mortalidade das árvores. O autor estudou as árvores de 4 a 35 anos de idade com o objetivo geral de criar equações para estimar várias variáveis dendrométricas e biológicas das árvores, mas sem fazer relação entre elas.

Martins Martins (2011) apresenta uma abordagem de simulação que utiliza modelos de regressão para estimar a mortalidade e o crescimento de árvores em uma floresta de eucaliptos. Nessa abordagem, as equações (Lundqvist-Korf e West) foram utilizadas para prever as variáveis dendrométricas (diâmetro e altura) e a probabilidade de mortalidade das árvores. Nesse trabalho foram realizadas projeções de 36 a 72 meses, utilizando índices de competição (IC) na composição das equações para prever a mortalidade, sendo testados vários tipos diferentes de IC presentes na literatura, a fim de se obter o que traz maior ganho de precisão nos modelos. O autor avaliou vários tipos de regressões e escolheu aquela que se mostrou mais eficiente nos dados da empresa. Entretanto, em vez de estimar o número de árvores vivas real vs. estimadas, o trabalho compara apenas os resultados do volume de madeira total do plantio e os resultados de diagnóstico do modelo. O modelo de árvore individual resultante estimou o volume do povoamento com um viés em torno de 7% na classe de produtividade alta, 4% na classe de produtividade média e 1% na classe de produtividade baixa.

Castro (CASTRO et al., 2013) utiliza redes neurais artificiais para simulação do crescimento de uma floresta de eucaliptos. Diferentes índices de competição foram avaliados na composição das equações para prever a mortalidade e os valores das variáveis dendrométricas usadas na estimação do volume de madeira com casca (VTCC), considerando projeções anuais de 24 a 72 meses. Na construção do modelo, as árvores do inventário foram previamente divididas em 3 grupos de acordo com sua capacidade de produção (alta, média e baixa) e submodelos de previsão baseados em regressão e rede neural foram treinados a fim de identificar a configuração com melhor acurácia para cada grupo específico. Além disso, emprega-se as médias aritméticas das medidas dendrométricas de todas as árvores da parcela no cálculo dos índices de competição, descaracterizando o princípio da localidade esperado em um autômato celular. Segundo o autor, a mortalidade foi estimada corretamente para 37%, 42% e 38% das árvores nas classes de capacidade produtiva alta, média e baixa, respectivamente, na idade final de projeção. Ele também ressalta que, mesmo apresentando um grau de acerto menor de 50% quanto às árvores que morreram durante o período de projeção, foi possível estimar sem tendência o número de árvores vivas por hectare nas parcelas selecionadas para validação, uma vez que a estimativa de mortalidade nesse tipo de modelo não depende da exatidão individual para cada árvore.

Nossa abordagem é semelhante àquela apresentada em (ASTORE; GUIMARÃES; JUNIOR, 2022), a qual os autores defendem e utilizam técnicas de aprendizado de máquina

para a construção das regras de transição baseados exclusivamente em dados históricos. Naquele trabalho, foi aplicado modelos de séries temporais nebulosas (FTS - *Fuzzy Time Series*) para aprender e representar as regras de transição do AC. A metodologia proposta foi validada através da construção de um modelo de autômato celular que simule a dinâmica populacional do vetor responsável pela doença de Chagas. Segundo os autores, a abordagem foi capaz de prever o comportamento futuro da quantidade de insetos com erros em torno de 12%, confirmando o potencial das técnicas de aprendizado de máquina aplicadas na construção das regras de transição.

Portanto, nosso modelo pode ser visto como uma mescla das abordagens apresentadas em Castro (CASTRO et al., 2013) e Astore (ASTORE; GUIMARÃES; JUNIOR, 2022), o qual combina autômato celular e técnicas de aprendizado de máquina para construir um modelo de árvore individual preciso a partir dos dados dos inventários florestais. Além disso, por não termos encontrado na literatura nenhum trabalho com aplicação de AC no estudo da evolução e produção de plantações da espécie *Tachigali vulgaris*, o modelo será construído com base nos dados estudados por Lima (LIMA et al., 2022).

A Tabela 1 apresenta um resumos dos principais pontos acerca das abordagens investigadas na literatura, destacando suas características em relação às técnicas empregadas na predição das medidas dendométricas e da mortalidade; ao tipo do problema e da espécie modelada; e ao índice de competição utilizado. Lembrando que, em todos os trabalhos estudados, o índice de competição utilizado considera todas as árvores da parcela, ou seja, um árvore que está na fronteira tem a mesma influência sobre suas vizinhas imediatas e nos exemplares que estão na extremidade oposta da parcela. Por outro lado, em nossa abordagem, apenas as interações locais são consideradas no computo do índice de competição, respeitando um dos princípios de um autômato celular.

Tabela 1 – Características dos modelos desenvolvidos nos trabalhos correlatos.

Trabalhos	(BANDINI; PAVESI, 2004)	(GARCÍA-DURO et al., 2018)	(TÉO, 2022)	(MARTINS, 2011)	(CASTRO et al., 2013)	Nossa Abordagem
Predição das medidas	Equação matemática	Não utiliza	Equação de Lundqvist-Korf	Regressão Linear	Equação de Lundqvist-Korf	Regressão Linear
Função de Mortalidade	Não utiliza	Equações pré-definidas de probabilidade	Regressão Logística	Modelo de West	Rede Neural (MLP)	Regressão Logística
Espécie de árvore	Carvalho e Pinheiro	Diversas espécies	<i>Pinus taeda L.</i>	Eucalipto	Eucalipto	Tachi-Branco
Contexto do problema	Crescimento de florestas naturais	Crescimento de florestas naturais pós incêndio	Crescimento de floresta plantada	Crescimento de floresta plantada	Crescimento de floresta plantada	Crescimento de floresta plantada
Índice de competição (IC)	Não utiliza	Não utiliza	Área basal (global)	Vários, sendo área basal o melhor (global)	Vários, sendo área basal o melhor (global)	<i>IID</i> ₃ (local)
Utiliza AC	Sim	Sim	Não	Não	Não*	Sim
Obtenção da regra do AC	Manual (empírica)	Manual (empírica)	Não se aplica	Não se aplica	Não se aplica	Automática (ML)
Atributos considerados nos modelos	Recursos naturais (luz solar, chuva) e nº de sementes	Germinação de sementes	Desbaste, altura, diâmetro, idade e índices de competição	Classe produtiva, altura, diâmetro, índices de competição e taxa de crescimento	Classe produtiva, altura, diâmetro, índices de competição, idade e índice de sítio	Altura, diâmetro, índice de competição, idade, nº de fustes e espaçamento de plantio

* Os autores alegam usar AC, mas consideram todas as árvores da parcela no cálculo do IC, contrariando o princípio de vizinhança e computação local característico de um AC.

MEPACAM - Modelo de Estimação da Produção de Madeira baseado em Autômato Celular e Aprendizagem de Máquina

O modelo de predição proposto utiliza um autômato celular probabilístico com o objetivo de representar a dinâmica de crescimento das plantações de Tachi Branco na região norte do Brasil, estimando com precisão o volume total de madeira produzido ao longo de cada do desenvolvimento das árvores. Dado que a estimativa de produção é influenciada pelas características individuais das árvores e pela interação com suas vizinhas, os autômatos celulares têm grande potencial para simular o crescimento florestal nesse nível de detalhe.

A descrição dos componentes do modelo de AC proposto, bem como o conjunto de dados utilizado como estudo de caso neste trabalho, são apresentados a seguir.

4.1 Estudo de Caso

O Tachi Branco surge como alternativa ao eucalipto para o fornecimento de madeira para a geração de energia na região amazônica. Ela é promissora para florestas energéticas porque está adaptada às condições ambientais da região e apresenta crescimento rápido e qualidade adequada de madeira (LIMA et al., 2022). Analisamos os dados de plantios experimentais com diferentes espaçamentos de plantio, possibilitando avaliar o efeito desses espaçamentos na ocorrência de bifurcações de tronco, mortalidade das árvores, qualidade da madeira, entre outros fatores. Os dados utilizados neste estudo são provenientes de um experimento pertencente à empresa Jari Celulose S.A., fruto de uma parceria entre a empresa e a Embrapa Amazônia Oriental. O experimento está localizado na região do Baixo Amazonas, em Monte Dourado, distrito do município de Almeirim-PA.

O mesmo conjunto de dados também foi utilizado nesta pesquisa, o qual consiste de um inventário florestal de uma plantação de Tachi Branco localizada na Amazônia Oriental, o qual foi fornecido por uma empresa privada. O delineamento experimental foi em blocos casualizados composto por três blocos, onde cada bloco foi dividido em seis parcelas de 60×51 metros (3.060 m^2), totalizando aproximadamente 5,6 hectares. Seis espaçamentos de plantio distintos (4,5, 6,0, 7,5, 9,0, 10,5 e 12 m^2 por planta) foram distribuídos aleatoriamente entre as parcelas de cada bloco. As medições das árvores foram realizadas em uma área de medição interna, composta por 49 plantas (dispostas em uma matriz 7×7) localizadas na região central de cada parcela. Neste estudo foram utilizados dados de 8 medições realizadas nos anos de 2013 a 2021, com exceção de 2017.. Portanto, o conjunto de dados utilizado neste estudo é composto por 882 árvores (49 árvores nas 18 subdivisões de parcelas) e 7.056 amostras (8 medições para cada árvore). É importante salientar que, as 882 árvores não estão presentes em todas as medições, pois com a ocorrência da mortalidade ao longo dos anos a quantidade de árvores vivas a cada medição vai diminuindo, portanto, as 882 árvores são referentes ao total de árvores no início do experimento.

Para estimar a evolução do volume de madeira das árvores ao longo do tempo, duas medições são utilizadas: altura (Ht) e diâmetro equivalente (DEQ). O diâmetro da árvore é tipicamente medido a 1,30 metros acima do solo, o que é chamado de diâmetro à altura do peito (DAP) (CASTRO, 2011). Para árvores bifurcadas, o DAP dos troncos distintos são consolidados em um único valor, como segue:

$$DEQ = \sqrt{DAP_1^2 + DAP_2^2 + \dots + DAP_n^2} \quad (20)$$

Além dos dados normalmente usados em modelos baseados em árvores individuais, como idade (em meses), altura, diâmetro, volume e mortalidade (viva ou morta) das árvores, também foram empregados atributos relacionados ao espaçamento de plantio e ao número de fustes por árvore. O impacto dessas novas características nas medições das árvores é analisado na Seção 5.3, justificando seu uso no treinamento dos modelos. Outros atributos também foram definidos para o posicionamento espacial das árvores, permitindo que o AC lide com iterações locais. Esse posicionamento consiste em identificar a parcela e a posição cardinal de cada árvore na área de medição interna (arranjo 7×7).

4.2 Construção do Modelo

O processo de construção do modelo de predição baseado em autômato celular começa com a criação da base de dados a partir do inventário florestal usado neste estudo. Nessa etapa, os dados de cada parcela monitorada passam por um processo de limpeza e preparação (ex: cálculo do DEQ e posicionamento espacial no reticulado) para uso nos modelos de regressão usados como regra de transição do autômato celular.

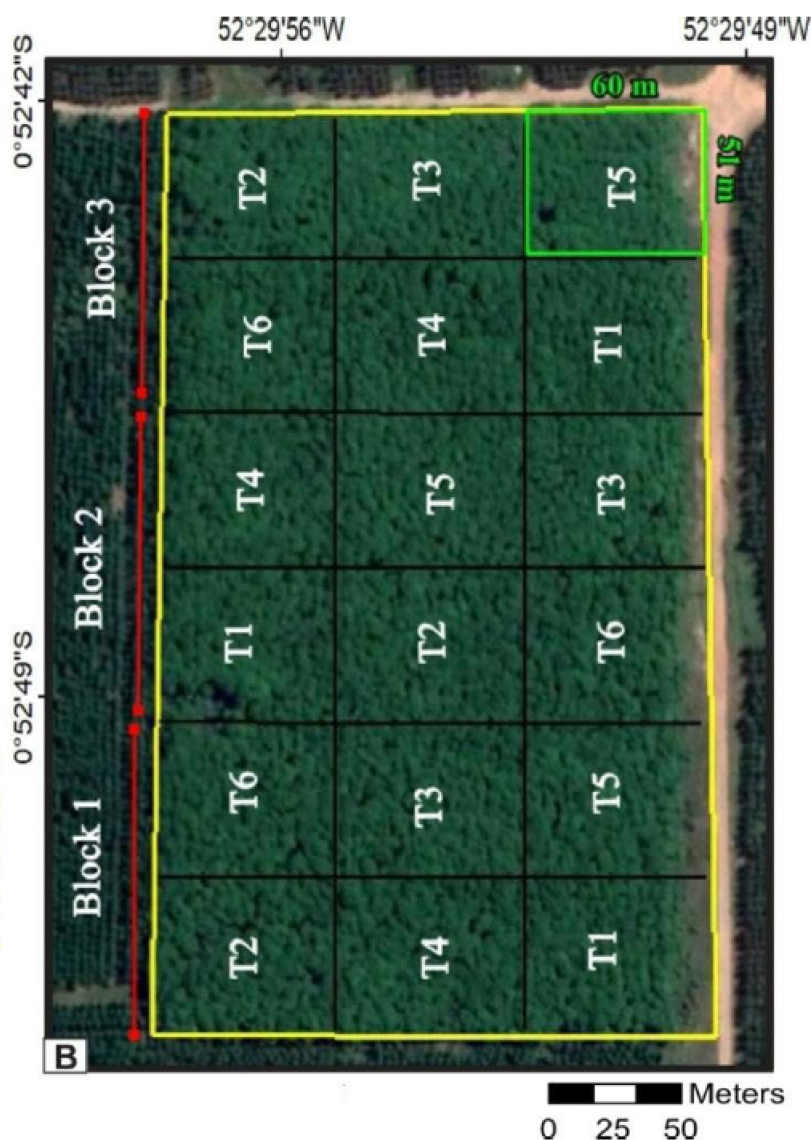


Figura 13 – Local do experimento com a platagem de Tachi-branco o leste da Amazônia, no estado do Pará. Disponível em (LIMA et al., 2022)

Em seguida, são realizados dez ciclos do processo de construção que consistem na separação dos dados, no treinamento dos modelos de regressão e na simulação do crescimento das árvores usando o modelo de autômato celular resultante.

Em cada ciclo, as parcelas presentes na base de dados são divididas aleatoriamente em dois conjuntos de dados: treinamento e teste. 80% das parcelas são destinadas para o treinamento dos modelos de regressão usando técnicas de aprendizado de máquina supervisionado, enquanto o restante dos dados (20% das parcelas) são usados na avaliação das estimações realizadas durante as simulações com o modelo de autômato celular resultante.

Durante o treinamento dos modelos de regressão, adotou-se a validação cruzada (*10-fold crossvalidation*), que particiona o conjunto de treinamento de forma aleatória em 10

partes. Dentre as partições geradas, nove são empregadas no aprendizado dos modelos (treinamento em si), enquanto a parte restante é usada para avaliar o desempenho do modelo durante o ajuste dos hiperparâmetros (validação). Esse processo se repete até que todas as 10 partições sejam usadas na etapa de validação dos modelos. Ao final, o valor médio dos parâmetros obtidos nos treinamentos é usado no ajuste dos modelos.

Como as classes do conjunto de dados são desbalanceadas, ou seja, há poucos registros de árvores mortas em relação as vivas, antes do treinamento do modelo de regressão logística, foi aplicada uma técnica de sobreamostragem (*oversampling*) para o balanceamento dessas classes. Essa técnica consiste em replicar, de forma aleatória, algumas amostras reais da classe minoritária no conjunto de dados, até que as duas classes estejam com quantidades similares (HE; MA, 2013).

Com os modelos de aprendizado de máquina treinados, suas respectivas equações são utilizadas como regra de transição do modelo de predição baseado em autômato celular, sendo o modelo de regressão logística para estimar a probabilidade de sobrevivência das árvores de uma parcela e as regressões lineares para predizer as respectivas alturas e diâmetros (DEQ). Todos os modelos se baseiam nos valores atuais das variáveis monitoradas (instante t) para predizer seus valores futuros no próximo passo do autômato celular (instante $t + 1$). Considerando a periodicidade entre as medidas existentes no inventário florestal, cada passo do autômato celular modelado corresponde aproximadamente a um ano de crescimento das árvores da parcela. O modelo de autômato celular é então aplicado para estimar o volume total de madeira produzido pelas árvores de cada parcela do conjunto de validação, ao longo do ciclo de corte simulado ($t = 8$). Os resultados simulados a cada passo de tempo do AC são comparados com os respectivos dados reais do conjunto de teste, a fim de avaliar a qualidade das estimações e a precisão do modelo.

Como mencionado anteriormente, o ciclo descrito acima se repete por 10 vezes, proporcionando aferir o desempenho médio do modelo de predição. A repetição do ciclo visa evitar uma análise enviesada da abordagem proposta, de modo que a minimizar a interferência da divisão das parcelas entre os conjuntos de treinamento e teste na avaliação geral do modelo de predição. A Figura 14 ilustra todo o processo de simulação empregado neste trabalho.

4.3 Espaço Celular

O espaço celular utilizado em nosso modelo consiste em um reticulado bidimensional que representa uma parcela da floresta plantada, como mostrado na Figura 15. Para representar a distribuição espacial das árvores de acordo com o inventário utilizado, o reticulado do autômato celular foi estruturado em um arranjo 7×7 de células, onde cada célula representa uma árvore e sua posição na parcela.

O estado de uma célula armazena os dados da árvore no instante t . Ele é representado

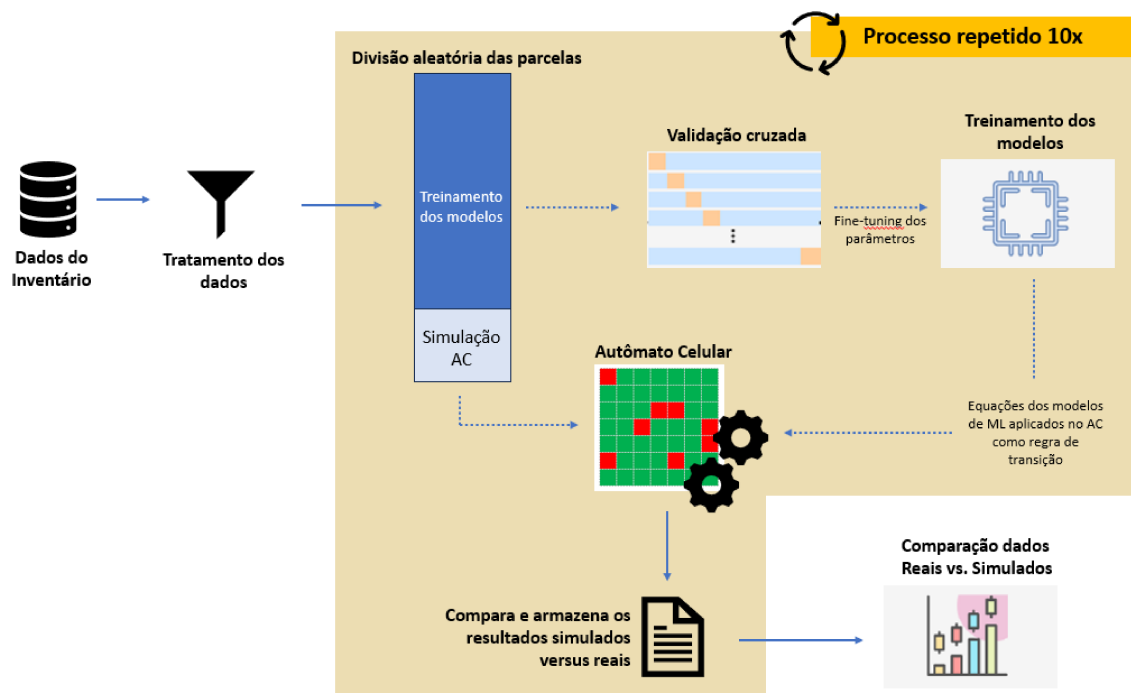


Figura 14 – Visão geral do processo de treinamento e validação do modelo de predição proposto.

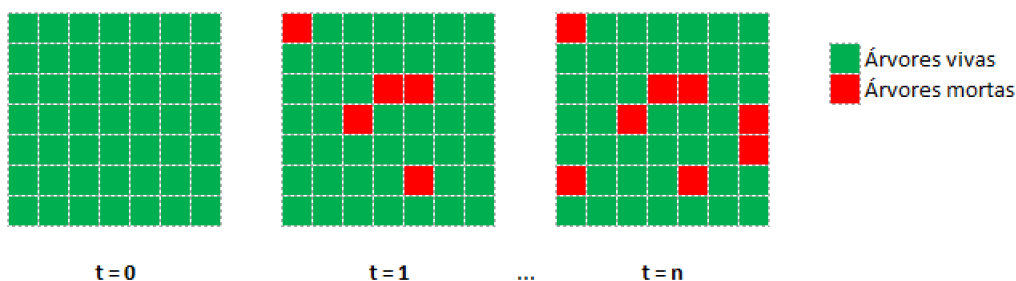


Figura 15 – Exemplo da evolução de um reticulado do autômato celular, representando uma parcela da plantação.

como uma tupla formada pelos atributos: idade (em meses), espaçamento de plantio, número de fustes, diâmetro equivalente (DEQ), altura (Ht), volume de madeira (Vw) e status de mortalidade (morto ou vivo). O reticulado inicial do autômato celular representa os dados das árvores aferidos na primeira medição da parcela monitorada, geralmente realizada por volta de 24 meses após o plantio.

4.4 Regra de Transição

Durante a evolução do autômato celular, uma regra de transição atualiza simultaneamente os estados de todas as células, onde os valores do reticulado no instante t são usados no cálculo dos estados das células no instante $t + 1$. As células são atualizadas com base em seu estado atual e no estado de suas vizinhas. A vizinhança da célula é determinada previamente em relação à sua proximidade física, lógica ou arbitrária, sendo que as vizinhanças de von Neumann e Moore são as mais comumente usadas em grades de AC 2D (SARKAR, 2000).

A regra de transição usada em nosso modelo emprega a vizinhança de Moore e consiste de três etapas: cálculo do índice de competição, previsão de mortalidade e estimativa das medições dendrométricas e do volume de madeira para as árvores presentes no reticulado. Para as árvores (células) das bordas do reticulado, adotou-se a técnica de interpolação para atribuir os valores dos vizinhos, isto é, calculou-se o valor dos vizinhos imaginários nas bordas com base nos valores médios das árvores de todo o reticulado. Essa abordagem é especialmente útil quando a média dos valores da vizinhança é uma aproximação razoável para o valor esperado das células nas bordas.

A competição entre árvores é uma característica relevante em modelos baseados em árvores individuais. Essa interação local é naturalmente modelada no AC por meio do conceito de vizinhança presente em sua regra de transição. Em nossa abordagem, o índice de competição (IC) de cada célula é calculado considerando a vizinhança de Moore e raio 1, como segue:

$$IC = \frac{DEQ^2 Ht}{\overline{DEQ}^2 \overline{Ht}} \quad (21)$$

onde DEQ e Ht são, respectivamente, o diâmetro equivalente e a altura (em metros) da árvore de referência (célula central); e \overline{DEQ} e \overline{Ht} são os valores médios de diâmetro equivalente e altura de suas vizinhas.

Este IC foi escolhido após o treinamento de 4 modelos de regressão linear usando os 4 índices diferentes, sendo que o modelo que utiliza o *IID3* foi o que se apresentou ligeiramente com o melhor desempenho.

Vale ressaltar que, ao contrário de outros trabalhos na literatura que consideram todas as árvores na parcela para calcular o IC (CASTRO, 2011), nossa abordagem considera apenas as árvores adjacentes (raio 1). O índice de competição é usado como entrada para os modelos de regressão, juntamente com os outros dados das árvores.

Um modelo de regressão logística é aplicado para gerar a probabilidade de mortalidade, ou seja, as chances de uma árvore estar morta no próximo passo do AC ($t + 1$), com base em seu estado e índice de competição atual (instante t). Para cada árvore, essa probabilidade de mortalidade é comparada a um número gerado aleatoriamente entre 0

e 1. Se a probabilidade for maior, o status de mortalidade da célula é 0 (morta). Caso contrário, seu status de mortalidade é 1 (viva).

Para cada célula viva, dois modelos de regressão linear são executados para estimar a altura e o diâmetro futuros da árvore (Ht e DEQ), também com base em seu estado e no índice de competição atual. Os valores estimados são então usados para calcular o volume de madeira da árvore (Vw), usando a mesma fórmula empregada pela empresa na preparação do inventário florestal, como segue:

$$Vw = e^{-9.1646+1.9121 \cdot \ln(DEQ)+0.7514 \cdot \ln(Ht)} \quad (22)$$

O uso desse tipo de fórmula é bastante comum tanto na academia quanto na indústria para calcular o volume de madeira em inventários florestais, uma vez que ela retorna valores altamente precisos e apropriados para cada tipo de plantio (SILVA et al., 2009).

Uma vez que o novo estado de todas as células é calculado, o reticulado do AC está devidamente atualizado e o processo é repetido. O número de passos do AC usados na simulação é determinado de acordo com o período de tempo desejado (em anos) para o ciclo de corte. Portanto, em cada passo do AC, 12 meses são adicionados à idade de cada árvore do reticulado. Neste trabalho, o ciclo de corte foi estabelecido em 8 anos. Considerando que a primeira medida disponível no inventário corresponde ao terceiro ano de plantio (reticulado inicial), durante as simulações, esse reticulado é evoluído por sete passos de tempo do autômato celular ($t = 7$).

A Figura 16 mostra as etapas da evolução do reticulado do autômato celular durante o processo de simulação usando a abordagem proposta.

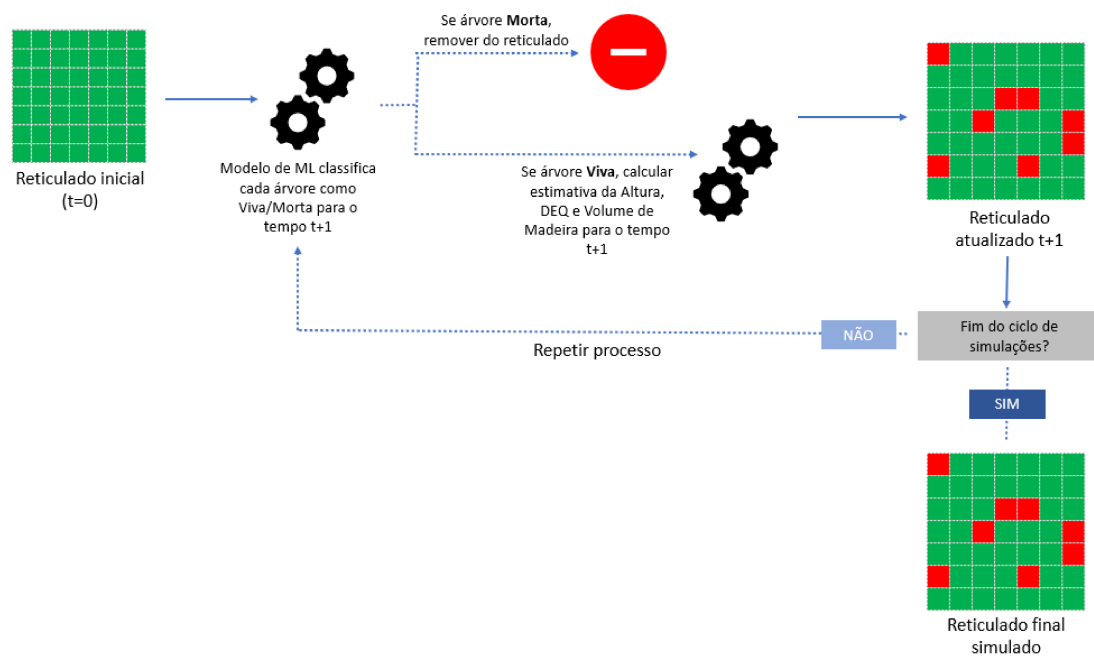


Figura 16 – Processo de evolução do reticulado do autômato celular durante a simulação.

Experimentos e Análise dos Resultados

Neste capítulo, são apresentados os experimentos conduzidos durante nossa pesquisa para validar a eficiência do modelo proposto. Inicialmente, verificou-se a qualidade dos modelos de regressão usados como regra de transição do AC. Nesse processo, além de avaliar os desempenhos individuais de cada modelo utilizado em nossa abordagem, também foi comparado o uso de diferentes técnicas de aprendizado de máquina na estimação da probabilidade de mortalidade das árvores. Em seguida, avaliou-se a capacidade do modelo baseado em autômato celular em prover estimações precisas do volume total de madeira das parcelas ao longo do ciclo de corte da floresta plantada monitorada. Por fim, analisou-se a influência dos novos atributos considerados em nossa abordagem (espaçamento de plantio e número de fustes vivos) na dinâmica de crescimento das árvores.

Os resultados e análises resultantes de cada um desses experimentos são apresentados a seguir.

5.1 Análise dos Modelos de Aprendizado de Máquina

No treinamento dos modelos de aprendizado de máquina, inicialmente o conjunto de dados foi dividido aleatoriamente em conjuntos de treinamento e teste, sendo 80% das amostras destinadas ao treinamento dos modelos e os 20% restantes para a avaliação de desempenho (teste). Em seguida, os modelos de aprendizado de máquina foram treinados usando validação cruzada com 10 partições para um ajuste não enviesado de seus hiperparâmetros. A equação de cada modelo foi inserida no AC como parte de sua regra de transição e o modelo de predição resultante foi avaliado usando apenas as parcelas presentes no conjunto de teste.

A escolha do modelo de regressão logística como técnica de estimação da probabilidade de mortalidade se deu após uma análise comparativa com outros tipos de modelos de aprendizado de máquina. Inicialmente, quatro técnicas de aprendizado de máquina foram previamente selecionadas para serem avaliadas como parte da regra de transição do AC,

a saber: Regressão Logística (RL), Rede Neural *Multi-layer Perceptron* (MLP), *Random Forest* (RF) e *Extreme Gradient Boosting* (XGB).

Utilizando a biblioteca PyCaret do Python ¹, foram realizados 10 treinamentos aleatórios para cada uma das técnicas a partir do conjunto de dados. A análise comparativa entre o desempenho dos modelos resultantes foi realizada com base na acurácia média dos modelos resultantes. Os valores individuais da métrica, considerando todas as rodadas de treinamento, também foram submetidas a uma análise de variância (ANOVA) (), a fim de verificar se as diferenças de desempenho observadas são estatisticamente significativas. A Tabela 3 mostra os resultados de acurácia obtidos para cada modelo nas 10 rodadas de treinamento, e a Tabela ?? apresenta as médias das 10 rodadas de treinamento das demais métricas dos modelos.

Tabela 2 – Acurácia dos modelos de aprendizado de máquina em cada rodada de treinamento (em %).

Rodada Treinamento	RL	MLP	RF	XGB
1	83,64	83,73	82,87	83,51
2	83,44	83,90	82,64	82,8
3	83,81	83,73	83,61	84,19
4	83,77	84,07	84,01	83,72
5	83,60	84,24	84,05	82,99
6	83,03	83,92	83,60	83,01
7	83,09	83,84	83,16	84,39
8	83,17	82,86	84,17	83,57
9	83,17	84,19	83,40	84,02
10	83,28	84,29	84,00	83,85
Média	83,40	83,88	83,56	83,61

Tabela 3 – Média das métricas dos modelos de aprendizado de máquina (em %).

Métrica	RL	MLP	RF	XGB
Acurácia	83,4	83,88	83,56	83,61
Precisão	83,87	84,15	84,36	85,42
Recall	88,48	88,92	87,13	87,2
F1-score	86,11	86,46	85,72	86,3

Pelo teste ANOVA, que compara as médias de acordo com suas variações, o *p-value* obtido foi de 0.1421, logo, considerando um grau de liberdade (erro) de 5%, aceita-se a hipótese nula do teste, ou seja, não há evidências estatísticas de diferenças significativas entre as médias de acurácia dos modelos. Considerando que não há diferenças significativas entre a acurácia dos modelos, optou-se pela utilização da regressão logística em nosso modelo, dadas as suas características de interpretabilidade de parâmetros, facilidade de treinamento, e menor custo de processamento.

¹ <https://pycaret.org/>

Como o modelo de regressão linear múltipla é o mais comum e interpretável dentre os modelos de regressões para variáveis contínuas, além de ser amplamente utilizados nos trabalhos correlatos da área por se adequar bem ao problema de estimativas lineares de crescimento e ter apresentado um excelente desempenho nos experimentos (R^2 acima de 0,9), optou-se por não comparar outras técnicas para a estimação das medidas das árvores. Entretanto, em trabalhos futuros, outras técnicas de aprendizado de máquina podem ser investigadas na tentativa de aprimorar as estimativas.

As análises dos desempenhos individuais dos modelos de regressão linear e logística adotados como regra de transição do AC são apresentadas a seguir.

5.1.1 Estimação de Mortalidade das Árvores

Um modelo de regressão logística foi treinado para estimar a probabilidade de sobrevivência das árvores monitoradas. Em outras palavras, o modelo determina se uma dada árvore da parcela irá sobreviver ou morrer no próximo passo do autômato celular. Nesse contexto, o modelo de regressão realiza uma tarefa de classificação e, portanto, a avaliação do seu desempenho foi baseada na matriz de confusão obtida em cada um dos ciclos de simulação, bem como nos valores médios das métricas acurácia e área sob a curva ROC (AUC - *Area Under the ROC Curve*).

A matriz de confusão é uma ferramenta importante para avaliar a qualidade de um modelo de classificação, fornecendo informações sobre seus acertos e erros em relação aos valores verdadeiros do conjunto de dados. Este quadro consiste em uma tabela que compara as previsões feitas por um modelo com os valores reais do conjunto de dados. Ela é chamada de matriz de confusão porque mostra o quão confuso ou preciso é o modelo em suas previsões, ao indicar quantos exemplos foram classificados corretamente e quantos foram classificados incorretamente. A matriz é organizada da seguinte forma:

- ❑ Verdadeiro Positivo (VP): Indica o número de exemplos que foram classificados corretamente como positivos pelo modelo.
- ❑ Verdadeiro Negativo (VN): Indica o número de exemplos que foram classificados corretamente como negativos pelo modelo.
- ❑ Falso Positivo (FP): Indica o número de exemplos que foram classificados incorretamente como positivos pelo modelo quando, na verdade, eram negativos. Também é conhecido como erro do Tipo I.
- ❑ Falso Negativo (FN): Indica o número de exemplos que foram classificados incorretamente como negativos pelo modelo quando, na verdade, eram positivos. Também é conhecido como erro do Tipo II.

Além disso, a acurácia é uma métrica que mede a proporção de previsões corretas feitas pelo modelo em relação ao número total de previsões. Ela é calculada dividindo o

número de previsões corretas (verdadeiros positivos e verdadeiros negativos) pelo número total de previsões. A AUC da curva ROC avalia a capacidade de um modelo em distinguir entre classes positivas e negativas, fornecendo uma medida global de desempenho que é menos sensível a desequilíbrios de classe.

A Figura 17 apresenta a matriz de confusão do modelo treinado em um dos ciclos de simulação, demonstrando que o modelo tem uma acurácia ajustada e balanceada entre as classes (0 para árvores mortas e 1 para vivas). Considerando os resultados de todos os 10 ciclos de simulação, os modelos treinados apresentaram uma acurácia média de 83,71%, um valor médio da AUC de 0,8367, precisão de 86,5%, recall de 78,9% e F1-score de 0,825. Isso indica que, de acordo com as métricas avaliadas, houve um bom ajuste dos hiper-parâmetros do modelo de regressão logística para a estimação da probabilidade de mortalidade das árvores presentes no conjunto de dados.

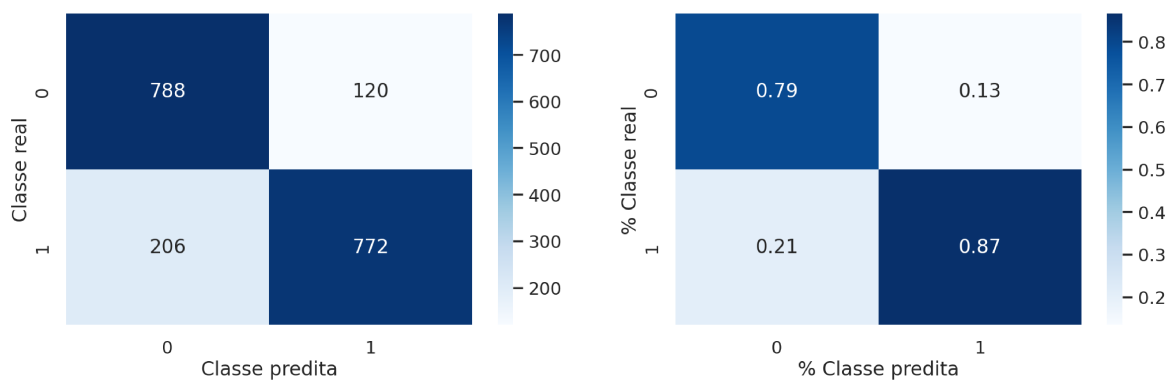


Figura 17 – Matriz de confusão do modelo de regressão logística para previsão do estado da árvore.

A Figura 18 apresenta os coeficientes médios das equações de regressão logística obtidas nas simulações de treinamento, além da importância de cada uma delas para o modelo, ou seja, uma medida da influência da variável na estimação da probabilidade da árvore estar viva no tempo futuro. Pode-se perceber que o índice de competição (*iid3*) tem uma influência quase 10 vezes maior que as outras variáveis no modelo. Como explicado anteriormente, o índice de competição é a variável do modelo que representa a interação local no AC, uma vez que considera os valores das árvores que são vizinhas imediatas da célula central (vizinhança de Moore com raio 1) em seu cálculo. Isso corrobora que a competição local entre as árvores pelos recursos disponíveis tem forte influência em suas chances de sobrevivência ao longo do tempo.

Em modelos lineares, como regressão linear e regressão logística utilizados neste trabalho, a importância das variáveis preditoras (*feature importance*) pode ser calculada com base nos coeficientes atribuídos a cada uma delas, ou seja, a magnitude dos coeficientes indica a importância relativa das variáveis. Variáveis com coeficientes maiores têm um impacto mais significativo nas previsões do modelo (MULLER; GUIDO, 2017).

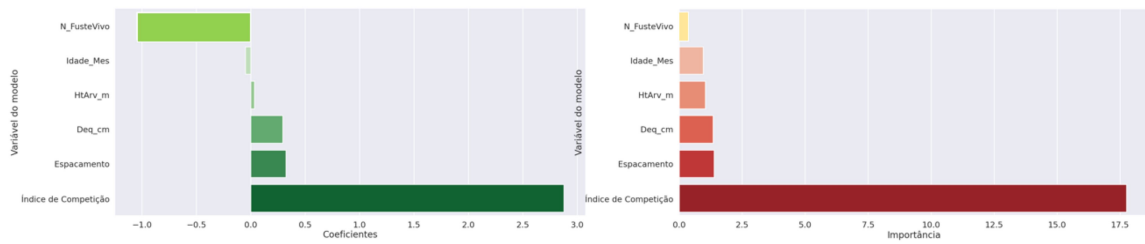


Figura 18 – Valor dos coeficientes e importância de cada variável no modelo de mortalidade.

Embora com uma relevância menor, as variáveis referentes ao espaçamento do plantio, diâmetro equivalente, altura e idade da árvore também são importantes para a probabilidade de mortalidade, sendo as duas primeiras um pouco mais importantes. A variável considerada no modelo que apresenta a menor influência na sobrevivência das árvores é o número de fustes vivos.

A Figura 19 compara entre as árvores que sobreviveram e morreram durante as simulações com o modelo, os valores do índice de competição de cada árvore em cada medição. Um boxplot, é uma representação visual que exibe a distribuição de um conjunto de dados por meio de cinco estatísticas principais: valor mínimo, primeiro quartil, mediana, terceiro quartil e valor máximo. Ele ajuda a identificar a dispersão e tendências dos dados, incluindo outliers, de forma concisa e informativa. Como pode ser observado, as árvores sobreviventes têm um índice de competição mediano maior em comparação com as árvores mortas. Esse efeito pode ser explicado pelo consumo de recursos naturais disponíveis, ou seja, quanto menor a árvore em relação às suas vizinhas adjacentes, menor será sua capacidade de conseguir os recursos naturais disponíveis naquela localidade, já que as árvores maiores ao seu redor consumirão mais recursos devido às suas dimensões.

O modelo de regressão logística treinado para prever a probabilidade de uma árvore permanecer viva no próximo passo de tempo ($P_{(t+1)}$) é representado na Equação 23:

$$P_{(t+1)} = \frac{e^{0.3279x_1 - 0.0588x_2 - 1.0833x_3 + 0.3267x_4 + 0.0495x_5 + 2.7069x_6}}{1 + e^{0.3279x_1 - 0.0588x_2 - 1.0833x_3 + 0.3267x_4 + 0.0495x_5 + 2.7069x_6}} \quad (23)$$

onde x_1 é o espaçamento usado no plantio da parcela (em metros); x_2 é a idade da árvore (em meses); x_3 é o número de fustes vivos; x_4 é o diâmetro equivalente (DEQ) da árvore; x_5 é a altura da árvore; e x_6 é o índice de competição da árvore. Essas variáveis x_i referem-se aos dados da árvore no instante t .

Os parâmetros ajustados pela regressão logística podem ser interpretados em termos de probabilidades. Por exemplo, o peso atribuído à idade (x_2) é $-0,0588$, indicando que para cada ano a mais na idade da árvore, a probabilidade dela permanecer viva diminui em 5,88%. Isso significa que à medida que os anos passam, as chances da árvore sobreviver diminuem. Outro exemplo é o parâmetro relacionado ao espaçamento do plantio (x_1). O

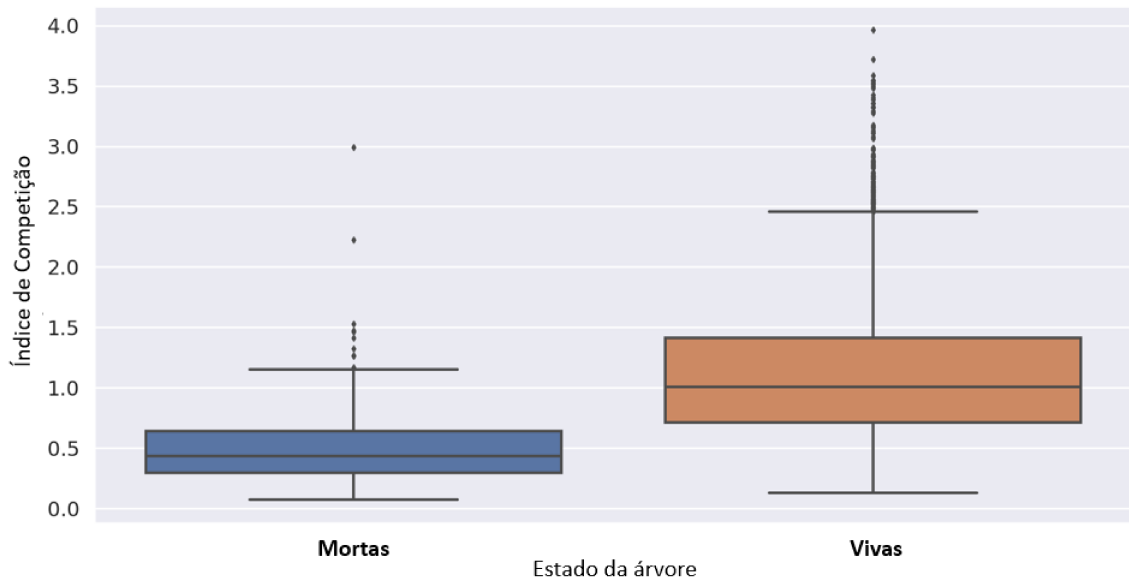


Figura 19 – Comparação entre os índices de competição das árvores vivas e mortas.

valor de 0,3279 pode ser interpretado da seguinte forma: para cada metro adicionado ao espaçamento entre as árvores, as chances da árvore sobreviver aumentam em 32,8%. Isso significa que uma árvore plantada no bloco de tratamento que adota um espaçamento de 4 metros entre as árvores tem uma chance 32,8% maior de sobreviver do que outra plantada em um bloco de tratamento com espaçamento de 3 metros.

5.1.2 Estimação das Medidas Dendométricas

Dois modelos de regressão linear foram treinados para a estimativa do diâmetro e da altura das árvores no próximo instante de tempo ($t + 1$). Ambos os modelos alcançaram bons valores em relação as métricas de desempenho adotadas (R^2 e $R^2_{ajustado}$), tendo o modelo de estimação da altura atingido um R^2 médio de 94% e $R^2_{ajustado}$ de 92,3%, enquanto que o modelo de estimação do DEQ atingiu um R^2 médio de 91,5% e $R^2_{ajustado}$ de 91,1%, em todas os 10 ciclos de simulação.

A Figura 20 mostra o bom ajuste do modelo ao apresentar os valores reais versus preditos pelo modelo para ambas as medidas dendométricas. Percebe-se que os valores seguem a tendência diagonal linear, com as variações normais em torno desta linha diagonal imaginária que define o ajuste perfeito do modelo.

As equações 24 e 25 representam os modelos de regressão linear treinados para prever a altura (H_t) e o diâmetro equivalente (DEQ) das árvores no tempo $t+1$, respectivamente.

$$H_{t(t+1)} = 0.5517x_1 + 0.0122x_2 + 0.0037x_3 - 0.0827x_4 + 0.9410x_5 + 1.0876x_6 \quad (24)$$

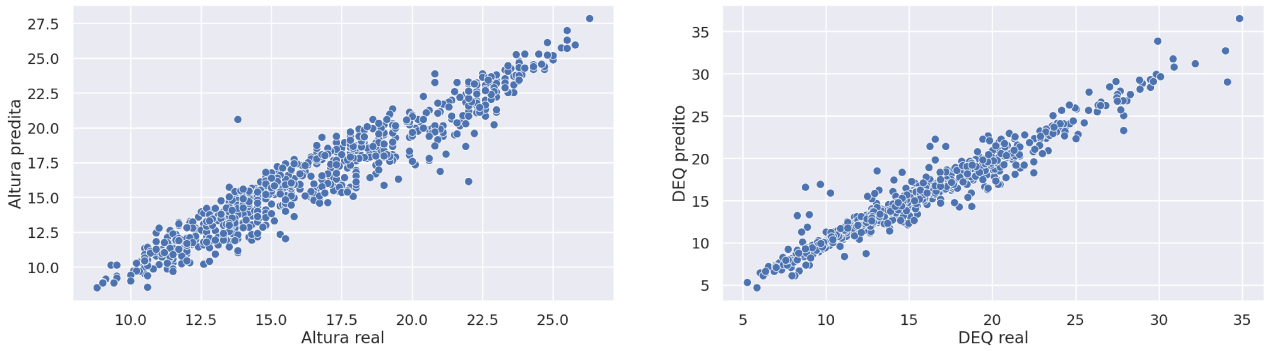


Figura 20 – Valores reais vs. preditos pelos modelos de regressão linear.

$$DEQ_{(t+1)} = 0.2202x_1 - 0.0140x_2 - 0.3795x_3 + 0.9899x_4 + 0.0744x_5 + 0.9296x_6 \quad (25)$$

onde x_1 é o espaçamento usado na parcela; x_2 é a idade da árvore; x_3 é o número de fustes vivos; x_4 é o diâmetro equivalente; x_5 é a altura; e x_6 é o índice de competição. Além disso, todas as variáveis x_i representam os respectivos valores da árvore no instante t .

5.2 Desempenho do Modelo baseado em Autômato Celular

Simulações foram realizadas usando o modelo baseado em autômato celular e considerando a primeira medição de cada parcela do conjunto de teste como reticulado inicial. A regra de transição obtida através dos modelos de aprendizado de máquina foi aplicada para simular a evolução do reticulado do AC ao longo dos anos do ciclo de plantio, até que o critério de parada fosse alcançado ($t = 8$). De maneira geral, o modelo baseado em AC obteve um desempenho satisfatório em ambas as métricas (número de árvores vivas e volume total de madeira) no final das simulações. A Tabela 4 apresenta os valores real e estimado dessas métricas e os respectivos erros dessas estimações, bem como os valores das métricas de desempenho dos modelos de regressão, considerando as 10 execuções do processo de simulação. A cada execução, a base do inventário é novamente particionada entre treinamento e teste, ocasionando as diferenças observadas nos valores reais apresentados na tabela.

Em média, o modelo estimou o volume total de madeira no tempo final do plantio com um erro médio de aproximadamente 4,2%, enquanto que a estimativa da quantidade total de árvores vivas apresentou um erro médio de aproximadamente 18,4%. É importante notar que, ao medir o erro dessas estimativas, a morte de cada árvore individual não é levada em consideração. Em outras palavras, o modelo não precisa “matar” exatamente a

Tabela 4 – Desempenho do modelos de aprendizado de máquina e do autômato celular em cada rodada de simulação.

Rodada de Simulação	Volume total (m^3)			Qtde. árvores vivas			Altura	DEQ	Mortalidade	
	Real	Estimado	Erro	Real	Estimado	Erro	R^2	R^2	Acurácia	AUC
1	44,2	41,4	6,3%	142	145	2,1%	94,2%	91,7%	82,9%	0,8290
2	37,8	37,4	1,1%	123	162	31,7%	93,8%	91,2%	83,2%	0,8324
3	39,0	37,2	4,6%	136	165	21,3%	93,7%	91,5%	84,8%	0,8465
4	42,0	41,5	1,2%	134	171	27,6%	94,4%	91,4%	82,9%	0,8281
5	26,0	23,4	10,0%	96	115	19,8%	93,6%	91,5%	86,1%	0,8603
6	33,1	29,6	10,6%	110	105	4,5%	93,9%	91,6%	82,1%	0,8207
7	34,3	36,3	5,8%	132	150	13,6%	94,1%	91,7%	83,6%	0,8357
8	20,1	20,0	0,5%	88	110	25,0%	94,3%	91,6%	85,1%	0,8500
9	29,0	29,1	0,3%	99	117	18,2%	93,5%	91,0%	84,0%	0,8393
10	27,9	27,4	1,8%	114	137	20,2%	94,0%	91,8%	82,5%	0,8250
Média	33,3	32,3	4,2%	117,4	137,7	18,4%	94,0%	91,5%	83,7%	0,8367

árvore que realmente morreu nos dados reais, mas estimar de forma precisa a quantidade que exemplares que permanecem vivos em cada parcela (CASTRO, 2011).

5.3 Análise da Dinâmica de Crescimento Resultante

Nesta seção, é apresentada uma análise exploratória, na qual se compara a evolução dos dados estimados pelo AC em relação as medidas reais presentes no inventário florestal, verificando, inclusive, o impacto dos novos atributos (espaçamento do plantio e o número de fustes vivos). Em outras palavras, a abordagem desenvolvida é avaliada no que tange a sua capacidade de capturar a dinâmica de crescimento presente nos dados do inventário, principalmente em relação as novas variáveis introduzidas nos modelos de regressão.

5.3.1 Espaçamento do Plantio

Inicialmente, avaliou-se a evolução dos valores médios das medidas dendrométricas (altura e DEQ), agrupadas de acordo com o espaçamento de plantio entre as árvores da parcela. Além de avaliar a influência do espaçamento na altura e no diâmetro das árvores, também é realizada uma análise comparativa entre as dinâmicas das curvas estimadas a partir do modelo de AC (gráfico à direita) e aquelas geradas a partir dos dados presentes no inventário florestal (gráfico à esquerda). Como pode ser observado, em ambos os casos, as dinâmicas de crescimento observadas nos dados reais foram mantidas nas estimações da simulação para os diferentes tipos de espaçamentos, indicando que, no geral, o modelo de predição foi capaz de reproduzir os padrões de crescimento dos dados do inventário ao longo dos anos.

Percebe-se que a altura e diâmetro das árvores possui uma relação direta com os espaçamentos utilizados em cada tratamento. Pode ser claramente visto na Figura 21 que quanto maior o espaçamento utilizado na plantação, maior é o diâmetro médio DEQ das árvores, comportamento que se repete para a variável altura (Figura 22). Esse fenômeno

pode novamente ser explicado pela disponibilidade de recursos naturais para cada árvore, afetando assim sua capacidade de crescimento, neste caso, relacionado ao diâmetro dos troncos. Um ponto a ser observado e investigado em trabalhos futuros, é o fato de que nos dados simulados, as curvas dos diferentes espaçamentos se distanciaram mais do que nos dados reais, além da tendência de crescimento ser ligeiramente diferente nos dados reais em relação aos simulados.

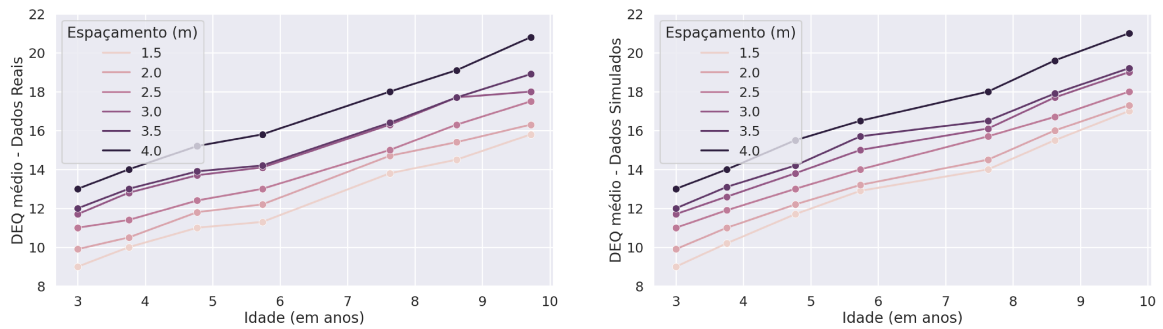


Figura 21 – Comparativo entre os dados reais e simulados pelo modelo AC para as medições da variável DEQ.

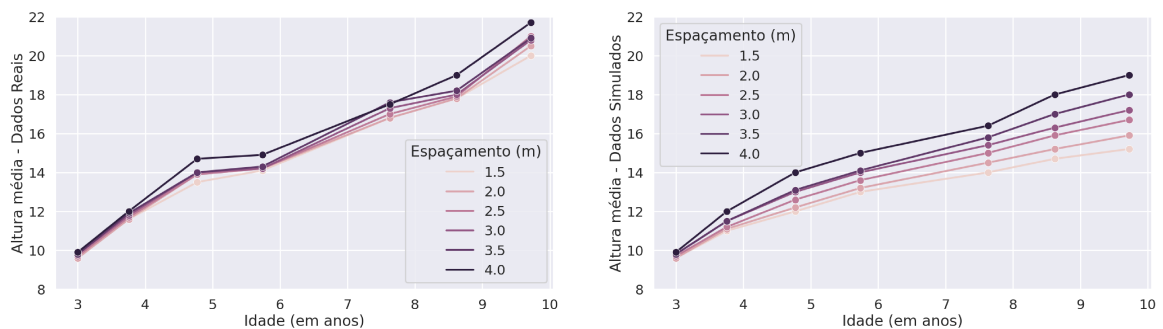


Figura 22 – Comparativo entre os dados reais e simulados pelo modelo AC para as medições da variável Altura.

Como visto anteriormente, duas métricas foram usadas para diagnosticar a qualidade do modelo: o número total de árvores e o volume total de madeira produzido na plantação. A Figura 23 apresenta a evolução do volume total de madeira produzido ao longo dos anos de plantio, considerando os dados reais do inventário florestal e aqueles estimados a partir do nosso modelo de AC. Como pode ser observado, o autômato celular foi capaz de representar com precisão a dinâmica do volume de madeira produzido.

Conforme apresentado na Figura 24, a evolução do número de árvores vivas também foi reproduzida, mostrando que o espaçamento utilizado em cada tratamento impacta a taxa de sobrevivência das árvores, com um espaçamento menor entre as árvores resultando em taxas de sobrevivência mais baixas. Isso pode ser explicado pela competição por recursos

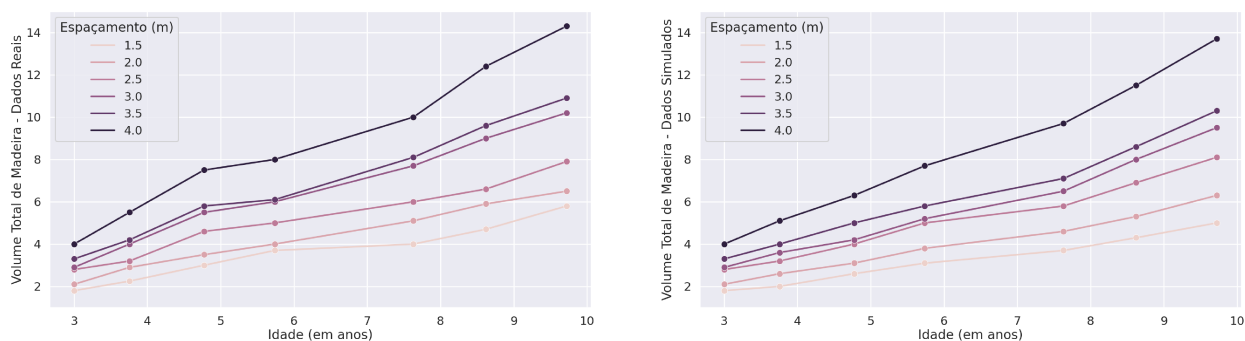


Figura 23 – Comparativo entre os dados reais e simulados pelo modelo AC para as medições do volume total de madeira (m³).

naturais disponíveis na área de plantio, ou seja, quanto mais árvores em um espaço, menos recursos naturais estão disponíveis para cada uma delas (minerais do solo, luz solar, água, etc.).

As curvas dos dados reais mostram uma estabilidade do número de árvores vivas até a 4^a medição, e um aumento acentuado na mortalidade a partir da 5^a medição. Já nos dados simulados, essa característica não foi reproduzida com precisão pelo modelo de regressão logística obtido a partir dos dados históricos, apesar do erro médio de estimativa na medição final ser próximo de 18%.

Foi realizado o teste t-student, que compara duas médias e verifica se a diferença entre elas é estatisticamente significativa. No nosso caso, comparamos a média das curvas reais versus simuladas. O *p-value* obtido nos testes realizados para cada métrica são apresentados na tabela 5.

Tabela 5 – Teste t-studente para comparação entre as curvas baseado em seus valores médios.

Métrica	<i>p-value</i>	Conclusão
Volume total de madeira	0,4812	Evidência inexistente contra H0
Qtd. Árvores vivas	0,0978	Evidência moderada contra H0
Altura média	0,0437	Evidência forte contra H0
DEQ médio	0,4933	Evidência inexistente contra H0

Considerando que a hipótese nula (H0) do teste é a de que não há diferenças significativas entre as médias, podemos concluir que as curvas das métricas Volume Total de Madeira e DEQ Médio não são estatisticamente diferentes, dado que o valor alto do *p-value* é favorável a H0. Por outro lado, a métrica Qtd. de árvores vivas possui evidência moderada contra H0, isto é, a diferença entre as curvas reais e simuladas é moderadamente significativa. Por fim, a métrica Altura média é a que apresenta o menor *p-value*, mostrando que a diferença entre as curvas reais e simulada é estatisticamente significativa.

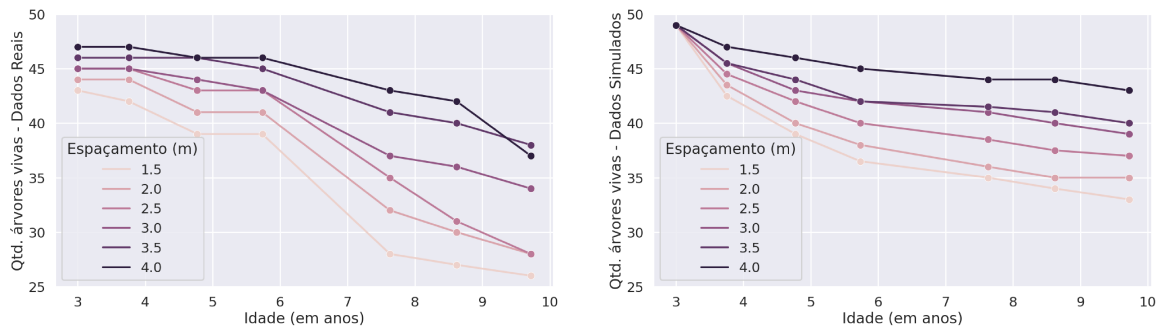


Figura 24 – Comparativo entre os dados reais e simulados pelo modelo AC para as contagem da quantidade de árvores vivas.

5.3.2 Número de fustes vivos

Nesse experimento é analisada a evolução média das medidas dendrométricas (altura e DEQ) de árvores agrupadas de acordo com a quantidade de fustes viva. Nessa análise, os valores médios estimados a cada passo de tempo do modelo de AC (gráfico à direita) são comparados com os respectivos dados reais presentes no inventário florestal (gráfico à esquerda).

De modo geral, o modelo reproduz parcialmente a dinâmica de crescimento dos dados do inventário ao longo dos anos. Entretanto, nota-se em, ambos os casos, existem pontos de divergência entre o padrão presente nos dados reais e o comportamento apresentado em algumas curvas geradas pelo modelo, principalmente em relação à altura das árvores.

Percebe-se que a altura e diâmetro das árvores possui uma relação direta com a quantidade de fustes vivos em cada árvore. Pode ser claramente visto na Figura 25 que quanto maior o nº de fustes da árvore, maior é o diâmetro médio DEQ das árvores, comportamento que se repete para a variável altura (Figura 26). Apesar dessa relação ser razoavelmente clara, percebe-se que o modelo não foi capaz de distinguir a tendência e distinção entre cada valor dessa variável, pois no gráficos dos dados simulados, a tendência de crescimento é linear enquanto que nos dados reais essa tendência é mais errática e sem padrão linear. Este é um ponto a ser observado e investigado em trabalhos futuros.

A Figura 27 apresenta a evolução do volume total de madeira produzido ao longo dos anos de plantio, considerando os dados reais do inventário florestal e aqueles estimados a partir do nosso modelo de AC em relação ao nº de fustes. Como pode ser observado, o autômato celular não foi capaz de representar totalmente e com precisão a dinâmica do volume de madeira produzido. A distinção e diferença entre as curvas é reproduzida nos dados simulados, exceto para os casos das curvas das árvores com nº de fustes = 1 e 3, que se cruzam nos dados reais mas seguem uma tendência diferente nos dados simulados.

Conforme apresentado na Figura 28, a evolução do número de árvores vivas também foi reproduzida, mostrando que o espaçamento utilizado em cada tratamento impacta a taxa de sobrevivência das árvores, onde árvores que possuem menos fustes possuem

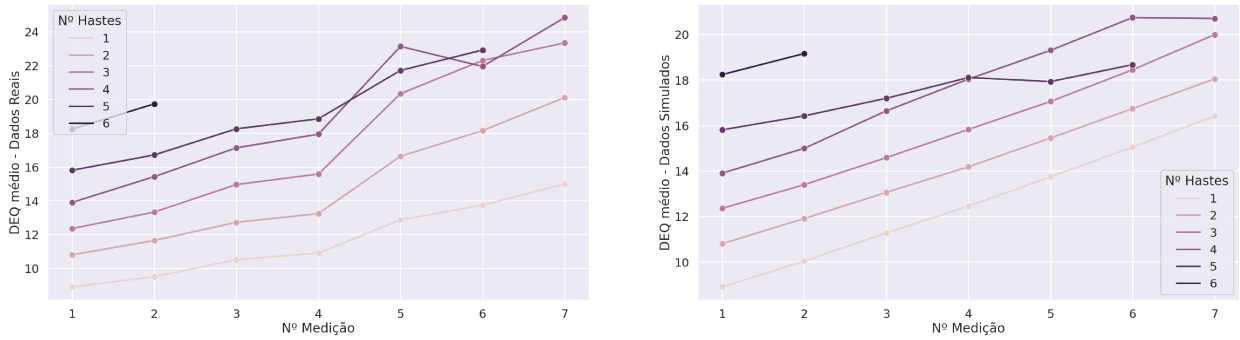


Figura 25 – Comparativo entre os dados reais e simulados pelo modelo AC para as medições da variável DEQ em relação a variável Nº de fustes.

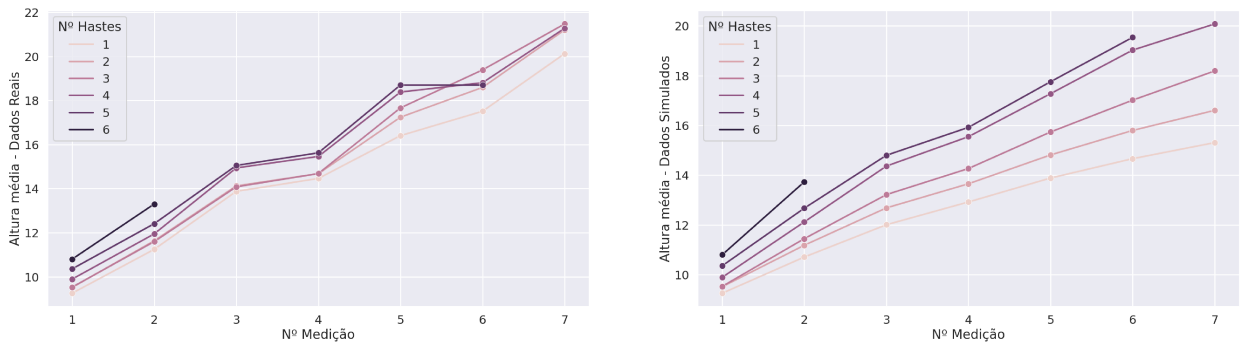


Figura 26 – Comparativo entre os dados reais e simulados pelo modelo AC para as medições da variável Altura em relação a variável Nº de fustes.

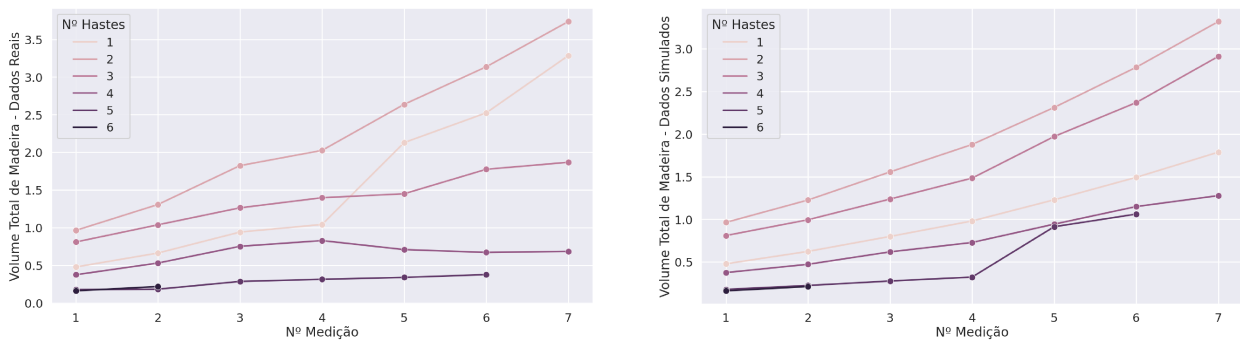


Figura 27 – Comparativo entre os dados reais e simulados pelo modelo AC para as medições do volume total de madeira (m³) em relação a variável Nº de fustes.

taxas de sobrevivência mais altas. Isso pode ser explicado pela competição por recursos naturais disponíveis no local de plantio da própria árvore, ou seja, quanto mais fustes em um pequeno espaço, menos recursos naturais estão disponíveis para cada uma delas evoluir e sobreviver. Apesar da distinção das curvas, a tendência não foi reproduzida com precisão para cada valor desta variável. Pode-se perceber que cada curva individualmente possui ritmos e tendências diferentes entre os dados reais e simulados

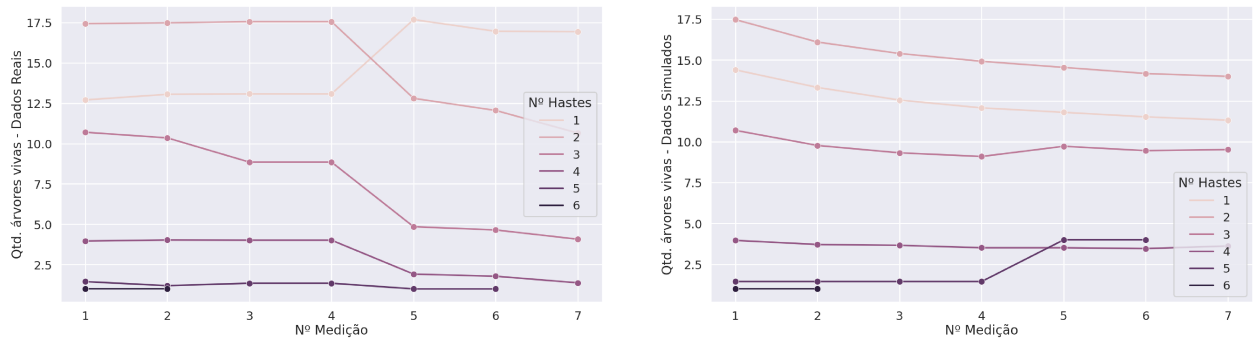


Figura 28 – Comparativo entre os dados reais e simulados pelo modelo AC para as contagem da quantidade de árvores vivas.

Acredita-se que a adoção de valores consolidados em vez das medidas dendométricas individuais de cada haste, minimiza o seu impacto nas estimações do modelo. Por meio do gráfico de relevância apresentado na Figura 18, é possível observar que, isoladamente, a quantidade de fustes é o atributo que tem menor impacto na predição do modelo. Entretanto, no processo de treinamento dos modelos essa variável se mostrou estatisticamente significativa para a explicabilidade da decisão dos modelos. Isto é, embora o número de fustes seja o atributo de menor relevância entre aqueles considerados nas regressões, ele ainda tem uma relação significativa com a probabilidade de mortalidade das árvores e a evolução das suas variáveis dendométricas

Conclusão

Neste trabalho, foi desenvolvido um modelo de crescimento e produção de plantios puros de *Tachigali vulgaris* baseado em um autômato celular probabilístico que adota como regra de transição modelos de aprendizado de máquina treinados a partir de dados disponíveis no inventário florestal. Nosso modelo de AC foi capaz de reproduzir a dinâmica de crescimento de um plantio de Tachi-Branco, prevendo com uma precisão aceitável a quantidade de árvores vivas e o volume de madeira.

O uso da vizinhança do AC no cálculo das previsões do modelo, propiciou a representação das interações locais entre as árvores vizinhas do plantio. Como apresentado nos resultados experimentais, considerar as células vizinhanças no cálculo do índice de competição foi relevante para o desempenho dos modelos de aprendizado de máquina, melhorando a precisão de suas estimativas. Nos trabalhos estudados que adotam modelos de estimação baseados em árvores individuais, a definição da vizinhança não está explícita, nem se enquadra no princípio de computação local usualmente empregado em um autômato celular. Por outro lado, em nossa pesquisa é apresentada uma abordagem baseada em AC que adota a vizinhança de Moore e raio 1, que modela as interações entre a árvore de referência (célula central) e suas vizinhas adjacentes. Este é um ponto importante, já que espera-se que a disputa de uma árvore pelos recursos naturais disponíveis seja maior com suas vizinhas mais próximas, em vez de considerar que todas as árvores da parcela têm a mesma influência nessa competição. Portanto, acredita-se que a abordagem proposta representa de forma mais realística a competição local entre as árvores.

Os modelos de mortalidade (regressão logística) e de previsão das medidas dendométricas (regressão linear) conseguiram refletir de forma satisfatória a relação entre as variáveis relacionadas às características das árvores e o seu crescimento, apesar de não conseguirem reproduzir completamente o comportamento observado nos dados reais, principalmente em relação à quantidade de árvores vivas. Em contrapartida, a interpretabilidade desses modelos se mostraram alinhados com as características presentes nos dados do inventário, possibilitando a análise da dinâmica da plantação de forma fácil e didática.

A análise exploratória possibilitou evidenciar o impacto do tipo de espaçamento usado

no plantio e da quantidade de fustes vivos no crescimento das árvores, bem como a capacidade do modelo de predição proposto em capturar e representar essa influência por meio da regra de transição do autômato celular gerada a partir dos dados de inventário.

Em geral, a definição de regras de transição usando modelos de regressão obtidos por aprendizado de máquina é uma vantagem em relação àquelas definidas empiricamente. O uso de técnicas de aprendizado de máquina para treinar modelos com base nos dados de inventários florestais permite maior flexibilidade e adaptabilidade do modelo de predição, dado que o próprio modelo pode aprender os padrões e tendências implícitas nos dados históricos usados no treinamento, sem a necessidade de conhecimento prévio de um especialista. Isso possibilita a adoção desta abordagem de projeção como instrumento para investigar o manejo de novas espécies, inclusive de florestas nativas, bem como o crescimento em outras regiões de plantio.

Por fim, a previsão alcançada pelo modelo baseado em AC foi superior àquela encontrada na literatura para estimativas de mortalidade ou sobrevivência em modelos de árvores individuais (MONSERUD; STERBA, 1999; TAYLOR; MACLEAN, 2007; CASTRO, 2011). É importante observar que, até onde se sabe, não foi encontrado na literatura nenhum trabalho que aborde modelos de predição de mortalidade e crescimento para árvores de *Tachi-branco*. Portanto, nosso trabalho se baseia em abordagens aplicadas a outras espécies, como o eucalipto. Parte dos modelos de crescimento estudados não tratam a questão da mortalidade de árvores. Acredita-se que isso se deve ao fato da mortalidade de árvores ser um dos componentes mais difíceis e menos confiáveis de estimar, devido principalmente às incertezas das condições futuras do plantio. No entanto, a estimativa correta da mortalidade ajuda no planejamento florestal, pois permite estimar com mais precisão a produção futura de madeira, especialmente quando a mortalidade é alta, como observado no estudo de caso investigado neste trabalho, chegando próximo a 50%.

6.1 Trabalhos Futuros

Uma extensão natural desta pesquisa é o refinamento da estratégia de construção da regra de transição do AC, avaliando a adoção de outros métodos de aprendizado de máquina mais complexos, buscando aprimorar a qualidade das estimativas e aproximar as curvas de evolução simuladas à dinâmica das medidas presentes no inventário, principalmente em relação a quantidade de árvores vivas ao longo do ciclo de plantio.

Nos casos onde os modelos de regressão não representam precisamente as tendências de crescimento das medidas monitoradas ao longo do tempo, a aplicação de transformações lineares durante o pré-processamento da base de dados pode ser uma alternativa para aproximar as previsões do modelo ao comportamento observado nos dados reais.

Existe um problema operacional em relação a ocorrência de bifurcação do tronco das árvores em florestas plantadas, logo, é de extrema importância para os pesquisadores

da área florestal a previsibilidade acerca da ocorrência de bifurcação. Neste contexto, a criação de um modelo de machine learning para prever a probabilidade de ocorrência de bifurcação nas árvores pode ser abordado em trabalhos futuros. Vale salientar que o nosso trabalho possui limitações importantes, dado que o nosso modelo não prevê a ocorrência de bifurcação, o que impacta na acurácia das previsões das variáveis dendométricas e mortalidade.

Também é importante investigar o uso de novos atributos relacionados ao clima, recursos naturais e terreno nos modelos de regressão, bem como avaliar seus respectivos impactos na qualidade e precisão das estimações do modelo, principalmente em relação a previsão da mortalidade das árvores.

Além disso, a metodologia de construção do modelo AC proposto neste trabalho, pode ser usada como base para o estudo analítico e quantitativo de outras espécies de árvores emergentes.

Por fim, utilizando a biblioteca Streamlit do Python, um sistema simulador dinâmico pode ser implementado em trabalhos futuros, permitindo os pesquisadores fazer uploads de inventários de plantações e simulando em tempo real o crescimento da floresta, acompanhando visualmente a dinâmica de crescimento e as principais métricas, como volume de madeira, quantidade de árvores, entre outras medidas.

6.2 Produção Bibliográfica Originada da Pesquisa

A investigação realizada durante o mestrado resultou em um artigo que foi submetido ao evento científico 39th ACM/SIGAPP Symposium On Applied Computing (FREITAS; CARNEIRO; MARTINS, 2024), o qual descreve a abordagem proposta e apresenta as análises referentes à sua eficiência na previsão do volume total de madeira produzido em plantações de Tachi Branco e à comparação entre as dinâmicas de evolução dos dados presentes no inventário florestal e aquelas obtidas a partir das simulações com o modelo de previsão proposto.

Referências

- ASTORE, L. M.; GUIMARÃES, F. G.; JUNIOR, C. A. S. Automatic rule generation for cellular automata using fuzzy times series methods. In: SPRINGER. **Brazilian Conference on Intelligent Systems**. [S.l.], 2022. p. 268–282.
- AVERY, T. E.; BURKHART, H. E. **Forest measurements**. [S.l.]: Waveland Press, 2015.
- BANDINI, S.; PAVESI, G. A model based on cellular automata for the simulation of the dynamics of plant populations. 2004.
- BIGING, G. S.; DOBBERTIN, M. A comparison of distance-dependent competition measures for height and basal area growth of individual conifer trees. **Forest science**, Oxford University Press, v. 38, n. 3, p. 695–720, 1992.
- BREIMAN, L. Random forests. **Machine learning**, Springer, v. 45, p. 5–32, 2001.
- BUCHMAN, R. G.; PEDERSON, S. P.; WALTERS, N. R. A tree survival model with application to species of the great lakes region. **Canadian Journal of Forest Research**, NRC Research Press Ottawa, Canada, v. 13, n. 4, p. 601–608, 1983.
- CAMPOS, J. C. C.; LEITE, H. G. **Mensuração florestal: perguntas e respostas**. [S.l.]: UFV, 2006.
- CASTRO, R. V. O. Modelagem do crescimento em nível de árvores individuais utilizando redes neurais e autômatos celulares. Universidade Federal de Viçosa, 2011.
- CASTRO, R. V. O. et al. Individual growth model for eucalyptus stands in brazil using artificial neural network. **Int. Scholarly Research Notices**, 2013.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: **Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining**. [S.l.: s.n.], 2016. p. 785–794.
- CONWAY, J. et al. The game of life. **Scientific American**, v. 223, n. 4, p. 4, 1970.
- DANIELS, R. F.; BURKHART, H. E.; CLASON, T. R. A comparison of competition measures for predicting growth of loblolly pine trees. **Canadian journal of forest research**, NRC Research Press Ottawa, Canada, v. 16, n. 6, p. 1230–1237, 1986.

- DELLA-FLORA, J. **Modelos de crescimento para árvores singulares—Nectandra magapotamica (Spreng) Mez. e Ocotea puberula Ness. 2001. 60 f.** Tese (Doutorado) — Dissertação (Mestrado em Engenharia Florestal)—Universidade Federal de Santa . . . , 2001.
- DURLO, M. A. Relações morfométricas para cabralea canjerana (well.) mart. **Ciência Florestal**, SciELO Brasil, v. 11, p. 141–149, 2001.
- FARIA, T. P.; NASCIMENTO, M. Z. D.; MARTINS, L. G. A. Understanding the multiclass classification of lymphomas from simple descriptors. In: IEEE. **2021 International Conference on Computational Science and Computational Intelligence**. [S.l.], 2021. p. 1202–1208.
- FOX, J. C.; BI, H.; ADES, P. K. Spatial dependence and individual-tree growth models: I. characterising spatial dependence. **Forest Ecology and Management**, Elsevier, v. 245, n. 1-3, p. 10–19, 2007.
- FRAGA, L. M.; OLIVEIRA, G. M. B. de; MARTINS, L. G. A. Adjustment of an epidemiological cellular automata-based model using genetic algorithm. In: IEEE. **IEEE 32nd Int. Conference on Tools with Artificial Intelligence**. [S.l.], 2020. p. 589–594.
- FREITAS, P. H.; CARNEIRO, M. G.; MARTINS, L. G. A. Prediction of managed forest growth based on cellular automata and machine learning. In: **submitted to 39th ACM/SIGAPP Symposium On Applied Computing**. [S.l.: s.n.], 2024. p. 1–8.
- GARCÍA-DURO, J. et al. Hidden costs of modelling post-fire plant community assembly using cellular automata. In: SPRINGER. **13th Int. Conference on Cellular Automata for Research and Industry**. [S.l.], 2018. p. 68–79.
- GLOVER, G. R.; HOOL, J. N. A basal area ratio predictor of loblolly pine plantation mortality. **Forest Science**, Oxford University Press, v. 25, n. 2, p. 275–282, 1979.
- GOES, F. d. A. et al. **Modelação do crescimento da árvore individual de Eucalyptus globulus Labill.** Tese (Doutorado), 2013.
- GOMES, F. A. F. **Formulação de sistemas epidemiológicos por autômatos celulares: comparação com a teoria clássica e simulações para a COVID-19.** Tese (Doutorado) — [sn], 2023.
- GONÇALVES, D.; LIMA, D. A. Evacuação de pedestres em situação de pânico modelada através de autômatos celulares. **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**, v. 7, n. 1, 2020.
- HAMILTON, D. A. A logistic model of mortality in thinned and unthinned mixed conifer stands of northern idaho. **Forest Science**, Oxford University Press, v. 32, n. 4, p. 989–1000, 1986.
- HASENAUER, H. Princípios para a modelagem de ecossistemas florestais. **Ciência & Ambiente**, v. 20, p. 53–69, 2000.
- HASENAUER, H.; MERKL, D.; WEINGARTNER, M. Estimating tree mortality of norway spruce stands with neural networks. **Advances in Environmental Research**, Elsevier, v. 5, n. 4, p. 405–414, 2001.

- HAYKIN, S. **Redes neurais: princípios e prática**. [S.l.]: Bookman Editora, 2001.
- HE, H.; MA, Y. Imbalanced learning: foundations, algorithms, and applications. John Wiley & Sons, 2013.
- HIGA, R. C. V.; WREGE, M. S. Zoneamento climático de eucalyptus grandis para a região sul do brasil. Colombo: Embrapa Florestas, 2010., 2010.
- HOSMER, D. W.; LEMESHOW, S. Applied logistic regression. canada: Jhon wiley & sons. **Inc**, v. 200, 2000.
- HYINK, D. M.; MOSER, J. W. A generalized framework for projecting forest yield and stand structure using diameter distributions. **Forest Science**, Oxford University Press, v. 29, n. 1, p. 85–95, 1983.
- HYYNEN, J. Predicting the growth response to thinning for scots pine stands using individual-tree growth models. The Finnish Society of Forest Science and The Finnish Forest Research Institute, 1995.
- KEDDY, P. A. Effects of competition from shrubs on herbaceous wetland plants: a 4-year field experiment. **Canadian journal of Botany**, NRC Research Press Ottawa, Canada, v. 67, n. 3, p. 708–716, 1989.
- LIMA, D.; LIMA, H. Autômatos celulares estocásticos bidimensionais aplicados a simulação de propagação de incêndios em florestas homogêneas. In: SBC. **Anais do V Workshop de Computação Aplicada a Gestão do Meio Ambiente e Recursos Naturais**. [S.l.], 2014. p. 15–24.
- LIMA, M. D. R. et al. Tachigali vulgaris energy forests: understanding spacing, age, and stem type effects on tree growth patterns and wood density. **New Forests**, Springer, p. 1–23, 2022.
- MABVURIRA, D.; MIINA, J. Individual-tree growth and mortality models for eucalyptus grandis (hill) maiden plantations in zimbabwe. **Forest Ecology and Management**, Elsevier, v. 161, n. 1-3, p. 231–245, 2002.
- MARTIN, G. L.; EK, A. R. A comparison of competition measures and growth models for predicting plantation red pine diameter and height growth. **Forest Science**, Oxford University Press, v. 30, n. 3, p. 731–743, 1984.
- MARTINS, F. B. Modelagem de crescimento em nível de árvore individual para plantios comerciais de eucaliptos. Universidade Federal de Viçosa, 2011.
- MENDES, B. R. et al. Desenvolvimento de modelos de crescimento de árvores individuais fundamentado em equações diferenciais. **Cerne**, Universidade Federal de Lavras, v. 12, n. 3, p. 254–263, 2006.
- MERKL, D.; HASENAUER, H. Using neural networks to predict individual tree mortality. In: CITESEER. **Proceedings of the Int’l Conference on Engineering Applications of Neural Networks**. [S.l.], 1998. p. 10–12.
- MONSERUD, R. A.; STERBA, H. Modeling individual tree mortality for austrian forest species. **Forest Ecology and Management**, Elsevier, v. 113, n. 2-3, p. 109–123, 1999.

- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. [S.l.]: John Wiley & Sons, 2021.
- MONTY, A.; LEJEUNE, P.; RONDEUX, J. Individual distance-independent girth increment model for douglas-fir in southern belgium. **Ecological modelling**, Elsevier, v. 212, n. 3-4, p. 472–479, 2008.
- MULLER, A. C.; GUIDO, S. **Introduction to machine learning with Python**. [S.l.]: O'Reilly, 2017.
- NEUMANN, J. V.; BURKS, A. W. et al. Theory of self-reproducing automata. **IEEE Transactions on Neural Networks**, v. 5, n. 1, p. 3–14, 1966.
- ORELLANA, B. et al. Energetic density of tachigali vulgaris wood by diametric class in two experimental plants in the amazon. **Nativa: Pesquisas Agrárias e Ambientais**, Mato Grosso Federal University, v. 6, n. Especial, p. 773–781, 2018.
- PEET, R. K.; CHRISTENSEN, N. L. Competition and tree death. **Bioscience**, JSTOR, v. 37, n. 8, p. 586–595, 1987.
- PORTE, A.; BARTELINK, H.; DREYFUS, P. Links between mechanistic approaches and the modelling of heterogeneous stand dynamics. In: **Workshop of the REA (Reseau Ecophysiologie de l'Arbre/Tree Ecophysiology Network)**, Avignon, France, 2001. [S.l.: s.n.], 2001. p. 1–1.
- 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**, Elsevier, v. 112, p. 140–147, 2018.
- RENNOLLS, K.; GEARY, D.; ROLLINSON, T. Characterizing diameter distributions by the use of the weibull distribution. **Forestry: An International Journal of Forest Research**, Oxford University Press, v. 58, n. 1, p. 57–66, 1985.
- RODRIGUES, R. P. et al. Growth and mortality of tachigali vulgaris lg silva and hc lima in different planting spacing for biomass production. **Scientia Forestalis**, Instituto de Pesquisas e Estudos Florestais (IPEF), n. 128, 2020.
- RODRIGUES, T.; JUNIOR, A. B. Technological prospecting in the production of charcoal: A patent study. **Renewable and Sustainable Energy Reviews**, Elsevier, v. 111, p. 170–183, 2019.
- SARKAR, P. A brief history of cellular automata. **ACM Computing Surveys**, v. 32, n. 1, p. 80–107, 2000.
- SCHNEIDER, P. R. et al. Análise da mortalidade de acacia mearnsii de wild. **Ciência Florestal**, SciELO Brasil, v. 15, p. 137–143, 2005.
- SILVA, E. C.; SOARES, J. A.; LIMA, D. A. Autômatos celulares unidimensionais caóticos com borda fixa aplicados à modelagem de um sistema criptográfico para imagens digitais. **Revista de Informática Teórica e Aplicada**, v. 23, n. 1, p. 250–276, 2016.
- SILVA, M. L. M. d. et al. Ajuste do modelo de schumacher e hall e aplicação de redes neurais artificiais para estimar volume de árvores de eucalipto. **Revista Árvore**, SciELO Brasil, v. 33, p. 1133–1139, 2009.

- SILVA, M. O. dos S. et al. Variations in productivity and wood properties of amazonian tachi-branco trees planted at different spacings for bioenergy purposes. **Journal of Forestry Research**, Springer, v. 32, p. 211–224, 2021.
- SOMERS, G. L. et al. Predicting mortality with a weibull distribution. **Forest Science**, Oxford University Press, v. 26, n. 2, p. 291–300, 1980.
- SOUZA, N. L. B.; LIMA, D. A. Controlador de robôs autônomos usando autômatos celulares e funções contínuas. **EnPE**, v. 7, n. 1, 2020.
- STAPE, J. L.; RYAN, M. G.; BINKLEY, D. Testing the utility of the 3-pg model for growth of eucalyptusgrandis × urophylla with natural and manipulated supplies of water and nutrients. **Forest Ecology and Management**, Elsevier, v. 193, n. 1-2, p. 219–234, 2004.
- TAYLOR, S. L.; MACLEAN, D. A. Dead wood dynamics in declining balsam fir and spruce stands in new brunswick, canada. **canadian Journal of Forest research**, v. 37, n. 4, p. 750–762, 2007.
- TÉO, S. J. Modelagem do crescimento e produção de árvore individual, para pinus taeda l., na região meio oeste de santa catarina. **FLORESTA**, v. 52, n. 3, p. 436–456, 2022.
- TOFFOLI, T. Cellular automata as an alternative to (rather than an approximation of) differential equations in modeling physics. **Physica D: Nonlinear Phenomena**, Elsevier, v. 10, n. 1-2, p. 117–127, 1984.
- TOFFOLI, T.; MARGOLUS, N. **Cellular automata machines: a new environment for modeling**. [S.l.]: MIT press, 1987.
- TOMÉ, M.; BURKHART, H. E. Distance-dependent competition measures for predicting growth of individual trees. **Forest Science**, Oxford University Press, v. 35, n. 3, p. 816–831, 1989.
- TOMÉ, M.; PAULO, J. A. An individual tree growth model for juvenile cork oak stands in southern portugal. **Silva Lusitana**, Estação Florestal Nacional, p. 27–38, 2009.
- UZOH, F. C.; OLIVER, W. W. Individual tree diameter increment model for managed even-aged stands of ponderosa pine throughout the western united states using a multilevel linear mixed effects model. **Forest Ecology and Management**, Elsevier, v. 256, n. 3, p. 438–445, 2008.
- VANCLAY, J. K. **Modelling forest growth and yield: applications to mixed tropical forests**. [S.l.]: Cab International, 1994.
- WEINGARTNER, M.; MERKL, D.; HASENAUER, H. Improving tree mortality predictions of norway spruce stands with neural networks. In: **Symposium on Integration in Environmental Information Systems, Zell am See, Austria**. [S.l.: s.n.], 2000.
- WEST, P. Simulation of diameter growth and mortality in regrowth eucalypt forest of southern tasmania. **Forest Science**, Oxford University Press, v. 27, n. 3, p. 603–616, 1981.

YANG, Y.; TITUS, S. J.; HUANG, S. Modeling individual tree mortality for white spruce in alberta. **Ecological Modelling**, Elsevier, v. 163, n. 3, p. 209–222, 2003.

ZEIDE, B. Analysis of growth equations. **Forest science**, Oxford University Press, v. 39, n. 3, p. 594–616, 1993.