UNIVERSIDADE ESTADUAL PAULISTA – UNESP

CÂMPUS DE JABOTICABAL

MODELOS DE MACHINE LEARNING PARA ESTIMAÇÃO DE PRODUTIVIDADE DE SOJA E EUCALIPTO NO CERRADO BRASILEIRO

Valter Barbosa dos Santos

Engenheiro Agrônomo

2024

UNIVERSIDADE ESTADUAL PAULISTA – UNESP

CÂMPUS DE JABOTICABAL

MODELOS DE MACHINE LEARNING PARA ESTIMAÇÃO DE PRODUTIVIDADE DE SOJA E EUCALIPTO NO CERRADO BRASILEIRO

Valter Barbosa dos Santos

Orientador: Prof. Dr. Glauco de Souza Rolim

Tese apresentada à Faculdade de Ciências Agrárias e Veterinárias – Unesp, Câmpus de Jaboticabal, como parte das exigências para a obtenção do título de Doutor em Agronomia (Ciência do Solo).

2024

S237m

Santos, Valter Barbosa dos
Modelos de machine learning para estimação de produtividade de soja e eucalipto no Cerrado brasileiro / Valter Barbosa dos Santos. -- Jaboticabal, 2024
78 p.

Tese (doutorado) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientador: Glauco de Souza Rolim

1. Agricultura. 2. Floresta. 3. Agrometeorologia. 4. Machine leaning. I. Título.

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

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA TESE: MODELOS DE MACHINE LEARNING PARA ESTIMAÇÃO DE PRODUTIVIDADE DE SOJA E EUCALIPTO NO CERRADO BRASILEIRO

AUTOR: VALTER BARBOSA DOS SANTOS

ORIENTADOR: GLAUCO DE SOUZA ROLIM

Aprovado como parte das exigências para obtenção do Título de Doutor em Agronomia (Ciência do Solo), pela Comissão Examinadora:

Prof. Dr. GLAUCO DE SOUZA ROLIM (Participação Virtual)
Departamento de Engenharia e Ciências Exatas / FCAV UNESP Jaboticabal

Profa. Dra. KAMILA CUNHA DE MENESES (Participação Virtual)
Coordenação de Zootecnia / Universidade Federal do Maranhão (U

gov.br Documento assinado digitalmente
KAMILA CUNHA DE MENESES
Data: 11/09/2024 12:25:10-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. ALEXANDRE DAL PAI (Participação Virtual)
Departamento de Bioprocessos e Biotecnologia / FCA UNESP Botucatu

gov.br Documento assinado digitalmente
ALEXANDRE DAL PAI
Data: 16/09/2024 18:27:28-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. GUSTAVO ANDRÉ DE ARAÚJO SANTOS (Participação Virtual)
Centro de Ciências de Chapadinha / Universidade Federal do Maranhão (UFMA)

gov.br Documento assinado digitalmente
GUSTAVO ANDRE DE ARAUJO SANTOS
Data: 11/09/2024 14:38:57-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. SALVADOR BOCCALETTI RAMOS (Participação Virtual)
Departamento de Ciências Exatas / FCAV UNESP Jaboticabal

Salvador Boccaletti Ramos

Jaboticabal, 09 de maio de 2024

DADOS CURRICULARES DO AUTOR

Valter Barbosa dos Santos- Filho de José Carlos dos Santos e Severina Barbosa (*In memoriam*). Nasceu em São José de Ribamar, Maranhão, no dia 22 de fevereiro de 1990. Técnico em Agropecuária pela Escola Agrotécnica Federal do Maranhão (2009), cursou engenharia Agronômica na Universidade Estadual do Maranhão-UEMA, câmpus Paulo VI, de São Luís-MA, de 2010 a 2017. Bolsista de iniciação cientifica pela PIBIC/UEMA em 2014-2015, atuando principalmente nos seguintes temas: manejo de cultivos agrícolas, potássio, adubação orgânica, agricultura familiar, Fertilidade do Solo. Trabalhou como servidor público, concursado no cargo de Fiscal Ambiental em 2013-2018. Em março de 2018, ingressou no Curso de Mestrado em Agronomia (Ciência do Solo), na Faculdade de Ciências Agrárias e Veterinárias – UNESP, desenvolvendo pesquisa sobre modelagem e inteligência artificial para aplicação no setor agrícola. Em março de 2020, foi aprovado no Curso de Doutorado em Agronomia (Ciência do Solo), mantendo a linha pesquisa em modelagem e Inteligência artificial. É integrante do grupo de pesquisa: “Group of Agrometeorological Studies” (GAS), da Unesp – Câmpus de Jaboticabal.

Debaixo do céu há momento para tudo, e tempo certo para cada coisa.

Eclesiastes 3,1

DEDICO

A Deus, pela dadiva da vida pois sem ele eu não teria traçado o meu caminho e feito a minha escolha pela Agronomia.

Aos meus pais, José Carlos e Severina Barbosa **(***In memoriam*) pelo apoio e confiança, mesmo distantes. Ao meu filho pela companhia nessa jornada.

OFEREÇO

À minha família, pelo incentivo e força nessa caminhada.

E ao professor Glauco Rolim pelo excepcional papel de orientador que desempenha na Unesp-Jaboticabal.

AGRADECIMENTOS

Agradeço primeiramente a **Deus**, pela dadiva da vida pois sem ele eu não teria traçado o meu caminho e feito a minha escolha pela Agronomia.

A minha **família** pelo amor incondicional, especialmente meus pais e meus irmãos, acreditando, apoiando e confiado em mim, aos meus sobrinhos e a toda família Moreno e Barbosa dos Santos.

Ao meu filho Samuel pelo amor e por me acompanhar nessa jornada.

Ao meu orientador professor Glauco de Souza Rolim, por sua amizade, dedicação pelos seus ensinamentos que me possibilitou realizar mais essa etapa da minha vida.

A professora Drª. Ana Maria Silva de Araújo, pelos ensinamentos, amizade, disposição, e paciência, por ter acreditado no meu potencial desde 2014 quando começamos a trabalhar juntos, me despertando o lado científico das ciências agronômicas através das orientações em projetos de iniciação científica entre os anos de 2014 a 2015.

Ao professor Gener Tadeu Pereira e ao meu amigo José Reinaldo Moraes membros da banca de qualificação pelas importantes sugestões.

Ao Grupo de Pesquisa em Agrometeorologia da Unesp – GAS, pelo recebimento no grupo e pelos conhecimentos compartilhados.

Aos funcionários do Departamento de Ciências Exatas, Maria José Servidone Trizólio, Shirley Aparecida Martineli de Sousa, Adriana Elisabete Takakura, por me receberem bem no departamento, e pelo carinho.

Ao programa de Pós-Graduação em Agronomia (Ciência do Solo), pela oportunidade em cursar o mestrado.

A todos os meus amigos que contribuíram direta e indiretamente nessa caminhada.

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

SUMÁRIO

MODELOS DE MACHINE LEARNING PARA ESTIMAÇÃO DE PRODUTIVIDADE DE SOJA E EUCALIPTO NO CERRADO BRASILEIRO

RESUMO – As mudanças ocorridas no Cerrado promoveram grandes desafios para o bioma. Em geral, a conservação do Cerrado tem sido conduzida por meio de políticas nacionais e locais. Essas políticas consideram a variação cultural e socioeconômica entre os municípios do Cerrado proporcionando avanços tecnológicos tanto na área florestal quanto na agricultura, tornando o Brasil líder mundial na produção de soja. Buscando reduzir a pressão do desmatamento e a manutenção da biodiversidade, ocorreu a introdução do eucalipto no Cerrado. Dessa forma, avaliou-se diferentes modelos de machine learning para estimação de produtividade de soja para o sul do Maranhão, com até um mês antecedência. Os resultados mostraram que o algoritmo Random Forest - RF atinge a maior precisão e acurácia, com R^2 de 0,81, RMSE de 176,93 kg ha^{-1} e tendência (EME) de 1,99 kg ha^{-1}. Por outro lado, o algoritmo Suport vector machine kernel RBF - SVM_RBF apresentou o menor desempenho com R^2 de 0,74, RMSE de 213,58 kg ha^{-1} e EME de 15,06 kg ha^{-1}. Em um segundo estudo buscamos estimar o volume de madeira de eucalipto no cerrado brasileiro utilizando técnicas de machine learning e apenas dados climáticos como entrada dos modelos, abrangendo diferentes idades de crescimento em dois períodos do ano entre janeiro e junho e entre julho e dezembro. Os modelos apresentaram ótimos resultados na estimativa do volume de madeira. O modelo Random Forest apresentou as melhores métricas durante o treinamento e teste com, R^2= 0,93 e RMSE = 18,36 m^3ha^{-1} para o modelo janeiro-junho e R^2= 0,92 e RMSE = 19,52 m^3ha^{-1} para o modelo de julho-dezembro.

Palavras chaves: agricultura, floresta, agrometeorologia, machine leaming

MACHINE LEARNING MODELS FOR ESTIMATING SOYBEAN AND EUCALYPTUS PRODUCTIVITY IN THE BRAZILIAN CERRADO

ABSTRACT – The changes that occurred in the Cerrado promoted major challenges for the biome. In general, Cerrado conservation has been driven through national and local policies. These policies consider the cultural and socioeconomic variation between the municipalities of the Cerrado, providing technological advances in both forestry and agriculture, making Brazil a world leader in soybean production. Seeking to reduce the pressure of deforestation and maintain biodiversity, eucalyptus was introduced into the Cerrado. In this way, different ML models were evaluated to predict soybean productivity for the south of Maranhão, up to one month in advance. The results showed that the RF algorithm achieves the highest precision and accuracy, with R^2 of 0.81, RMSE of 176.93 kg ha^{-1} and trend (EME) of 1.99 kg ha^{-1}. On the other hand, the SVM_RBF algorithm presented the lowest performance with R^2 of 0.74, RMSE of 213.58 kg ha^{-1} and EME of 15.06 kg ha^{-1}. In a second study, we sought to estimate the volume of eucalyptus wood in the Brazilian cerrado using machine learning techniques and only climate data as model inputs, covering different growth ages in two periods of the year between January and June and between July and December. The models showed excellent results in estimating the volume of wood. The Random Forest model presented the best metrics during training and testing with R^2= 0.93 and RMSE = 18.36 m^3ha^{-1} for the January-June model and R^2= 0.92 and RMSE = 19.52 m^3ha^{-1} for the July-December model.

Keywords: agriculture, forest, agrometeorology, machine leaming

LISTA DE ABREVIATURAS E SIGLAS

IA – Inteligência Artificial

ML – Machine learning

RNAs – Redes Neurais Artificiais

DNNs – Redes Neurais Profundas

CNNs- Redes Neurais Convolucionais

LSTMs - Long Short Term Memory (Memória de longo prazo)

MLP - Multilayer Perceptron

GEE – Gases do Efeito Estufa

RUE - Eficiência no uso da radiação

T – Temperatura do ar (ºC)

P – Precipitação Pluviométrica (mm)

AWC – Capacidade de água disponível (mm)

CET – Evapotranspiração da cultura (mm)

AET – Evapotranspiração real (mm)

STO – Armazenamento de água no solo (mm)

DEF – Déficit Hídrico (mm)

EXC – Excedente Hídrico (mm)

BH – Balanço Hídrico

CV —Validação Cruzada

RF – Random Forest

SVM – Suport Vector Machine (Máquina de vetores de suporte)

XGBoost – Extreme Gradient Boosting

LISTA DE FIGURAS

independentes para o período de janeiro a junho; b) Resultado do modelo de regressão após stepwise backward, com 47 variáveis para o período de janeiro a junho; c) Ajuste inicial do modelo de regressão com 71 variáveis independentes para o período de julho a dezembro; d) Resultado do modelo de regressão após stepwise backward, com 45 variáveis para o período de julho a dezembro.

Figura 15 (Capítulo 3) - Teste dos modelos Random forest – RF, Regressão Linear Multipla – MLR, Suporte Vector Machine – SVM e XGBoost Regressor – XGB, com conjunto de dados climáticos entre janeiro e julho da série histórica de 12 anos, considerando florestas entre 2 e 7 anos de idade na região do Cerrado – Mato Grosso do Sul, Brasil.

Figura 16 (Capítulo 3) – Teste dos modelos Random forest – RF, Regressão Linear Multipla – MLR, Suporte Vector Machine – SVM e XGBoost Regressor – XGB, com conjunto de dados climáticos entre janeiro e julho da série histórica de 12 anos, considerando florestas entre 2 e 7 anos de idade na região do Cerrado – Mato Grosso do Sul, Brasil.

Figura 17 (Capítulo 3) - Volume estimado ao longo do crescimento do eucalipto, considerando os valores médios para cada idade do conjunto de teste dos modelos (janeiro-junho) Random forest – RF, Regressão Linear Multipla – MLR, Suporte Vector Machine – SVM e XGBoost Regressor – XGB, para a região do Cerrado – Mato Grosso do Sul, Brasil.

Figura 18 (Capítulo 3) - Valores médios de volume total com casca por idade utilizados no conjunto de teste dos modelos seguido dos resultados médios estimados com avaliação dos resíduos entre o estimado e observado para o modelo de janeiro-junho.

Figura 19 (Capítulo 3) – Comportamento da curva do volume estimado ao longo do crescimento do eucalipto, considerando os valores médios para cada idade do conjunto de teste dos modelos (julho-dezembro) Random forest – RF, Regressão Linear Multipla – MLR, Suporte Vector Machine – SVM e XGBoost Regressor – XGB, para a região do Cerrado – Mato Grosso do Sul, Brasil.

Figura 20 (Capítulo 3) - Valores médios de volume total com casca por idade utilizados no conjunto de teste dos modelos seguido dos resultados médios estimados com avaliação dos resíduos entre o estimado e observado para o modelo de julho-dezembro.

CAPÍTULO 1 – Considerações gerais

1.1 Introdução

O Cerrado brasileiro é o segundo maior bioma da América do Sul, cobrindo originalmente aproximadamente 2 milhões de km^2 (Glatzle et al., 2024). É um ecossistema heterogêneo formado por um mosaico de ambientes abertos e fisionomias de vegetação fechada, composta por pastagens, savanas e florestas estacionais (Costa et al., 2020). No entanto, este bioma passou por importantes transformações nas últimas cinco décadas (Hunke et al., 2015).

O avanço antrópico sobre o bioma Cerrado está intrinsecamente ligado ao desenvolvimento da região a qual ele ocupa no território nacional (Ruas et al., 2022). O Cerrado é considerado uma grande região biogeográfica que possui riqueza sociocultural e ecológica inestimável, apresenta uma rica biodiversidade e um mosaico de fitofisionomias que se aglutinam sob relações de interdependência entre clima, solo, relevo e vegetação (de Souza et al., 2023).

A expansão da agricultura e o uso de tecnologias no Cerrado geraram benefícios socioeconômicos inegáveis: aumento da oferta dos produtos agrícolas, ganhos na produtividade da agricultura, diversificação das economias locais e aumento da renda de municípios, e melhorias sociais em várias localidades (Klink et al., 2005).

As transformações ocorridas no Cerrado trouxeram grandes desafios para o bioma que sofreu com a fragmentação de habitats, mitigação da biodiversidade, degradação de ecossistemas, desequilíbrios no ciclo do carbono e possível modificações climáticas regionais (Klink et al., 2005).

Ao mesmo tempo, esse avanço foi parte essencial do crescimento da economia brasileira, que até a década de 1960 ainda importava alimentos, como milho, arroz, cereais e carne de frango. Atualmente o país é um ativo exportador de alimentos, principalmente carnes e grãos (Carneiro Filho et al., 2016), com destaque para a soja produzida no Cerrado, ocupando a primeira posição no ranking exportador global e o segundo em produção cujo avanços tecnológicos foram o principal motivo que tornou isso possível (Magalhães et al., 2020).

O Cerrado brasileiro compreende 1297 municípios com diferentes históricos de ocupação em 12 estados brasileiros Bahia, Distrito Federal, Goiás,

Maranhão, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Paraná, Piauí, Rondônia, São Paulo, Tocantins. Além disso, as características urbanísticas, socioeconômicas, climáticas e geológicas variam entre os municípios e com o tempo (Campolina, 2019). A conservação desse bioma tem sido conduzida por meio de políticas nacionais e locais (Eloy et al., 2017), que podem visar diferentes aspectos do desmatamento em diferentes regiões. Considerando a variação climática, cultural e socioeconômica entre os municípios do Cerrado, é provável que a importância dos fatores associados ao desmatamento seja distinta em um contexto mais regional (Lima et al., 2018).

Além da sua importância intrínseca em termos de biodiversidade, o cerrado tropical é também essencial no sistema climático da Terra, responsável por 21% da evapotranspiração global (Miralles et al., 2011). Estudos recentes indicam uma taxa de conversão de 5.000 km^2 ano^{-1}, principalmente de pastagens e bosques em relevo plano e suave, que foram convertidos para terras agrícolas (Ferreira et al., 2016). No entanto, a produção agrícola na região não é realizada durante todo o ano devido ao acentuado déficit hídrico entre junho e setembro, quando o solo fica exposta ou coberta por resíduos de culturas de verão (Neto et al., 2010; Maia et al., 2022).

Um segundo fator antrópico que afeta o sistema climático no Cerrado brasileiro é a alta concentração de partículas de aerossóis na atmosfera causada pela queima de biomassa em incêndios florestais. A queima é uma prática antiga e difundida em todo o Cerrado brasileiro e é usado tanto por comunidades tradicionais quanto por pecuaristas (Eloy et al., 2018). Além disso, o fogo é usado para remover resíduos de culturas de plantios comerciais (Pivello, 2011; Garcia et al., 2021).

No total, mais de 68 mil incêndios são detectados anualmente por satélites em todo o Cerrado brasileiro, dos quais mais de 80% ocorrem entre julho e outubro (INPE, 2020). Portanto, conciliar interesses entre agronegócio e conservação de recursos naturais do Cerrado é um dos principais desafios do Brasil (Magalhães et al., 2020).

1.2 Revisão de literatura

1.2.1 Panorama da soja no cerrado

Nos últimos vinte anos, o crescimento da produção de soja no Brasil atingiu enormes proporções (Loayza et al., 2023). O Brasil se tornou líder na produção mundial de soja, ultrapassando os Estados Unidos e consolidando-se como um dos maiores produtores agrícolas do mundo (Jesus, 2023).

A nível regional, a cultura da soja impulsionou o crescimento dos estados do Sul e Sudeste, transformou o Centro-Oeste em uma região de crescimento robusto e expandiu-se pelos cerrados nordestinos da Bahia, Piauí e Maranhão e, também, nortista, representado pelo estado do Tocantins; demonstrando ser umas das maiores impulsionadoras do crescimento econômico em vastas áreas (Benevides; Staback, 2023).

Uma área de expansão agrícola que tem contribuído fortemente com o agronegócio brasileiro é a região do MATOPIBA, acrônimo referente à área de intersecção dos estados do Maranhão, Tocantins, Piauí e Bahia com área de 73.173.485 hectares (Pereira; Pauli 2016), correspondendo aproximadamente a 1,3 vezes a área da França (Barbosa dos Santos et al., 2021), envolvendo 337 municípios. Foi instituído pelo Decreto Presidencial nº 8.447, de 2015, o Plano de Desenvolvimento Agrícola do MATOPIBA (Pereira; Pauli 2016)

Esta fronteira agrícola no bioma cerrado responde por grande parte da produção brasileira de grãos, especialmente a soja (EMBRAPA, 2019). Uma fronteira agrícola é definida como uma região dominada por vegetação natural que começou a enfrentar intensa ocupação da terra relacionada à agricultura (Araújo et al., 2019).

Apesar de toda a atividade econômica, existe grande preocupação na preservação ambiental. O MATOPIBA possui 50 unidades de conservação federais, estaduais e municipais (7,2 milhões de ha) e 23 terras indígenas (3,6 milhões de ha) (Araújo et al., 2019). O Brasil já demonstrou por meio de uma série de iniciativas e ações políticas que inibir o desmatamento é possível (West e Fearnside, 2021).

O Brasil possui diversas iniciativas relacionadas à produção mais sustentável (Toloi et al., 2021), como o aumento da produtividade, o aproveitamento de áreas degradadas pela pecuária extensiva, a adoção de

práticas econômica e ambientalmente sustentáveis, como a integração da pecuária, e, por fim, a integração lavoura-pecuária (Oliveira et al., 2010).

Essas ações resultam em redução de insumos químicos e CO_2 emissões por meio da redução do uso de combustível e mão de obra agrícola, potencializando a mitigação de gases de efeito estufa e o sequestro de carbono (De Freitas; Landers, 2014).

1.2.2 Panorama do eucalipto no cerrado

No Brasil a área de árvores plantadas totalizou 9,94 milhões de hectares desse total, o eucalipto representa cerca de 7,6 milhões de hectares (ou 76%) sendo assim a espécie florestal mais cultivada. Devido ao seu alto potencial de adaptação em muitas regiões, propagação clonal e rápido crescimento (Castro et al. 2016) o eucalipto está localizado principalmente nas regiões Sudeste e Centro-Oeste do país, com destaque para Minas Gerais (29%), Mato Grosso Sul (15%) e São Paulo (13%) (IBÁ, 2023).

Visando a redução da pressão do desmatamento em áreas nativas e manutenção da biodiversidade, incentivos governamentais juntamente com pesquisas incentivadas pelas empresas florestais, deram início a estudos sobre a adaptabilidade do gênero Eucalyptus, já que o benefício econômico que a espécie trazia era o foco principal da produção e, com isso, foi surgindo uma visão social em relação aos benefícios dessa espécie (Ruas et al., 2022).

A silvicultura no Brasil é uma das mais avançadas do mundo, com a maior produtividade média mundial estimada em 32,7 $m^3\ ha^{-1}\ ano^{-1}$ (com casca), com uma duração média do ciclo de 6,7 anos (IBÁ 2023). Esse sucesso se deve aos ganhos proporcionados pelos programas de melhoramento e pelos avanços tecnológicos nas práticas silviculturais, como adubação e controle de doenças e plantas daninhas (Castro et al. 2016).

O eucalipto se tornou um recurso essencial para suprir a demanda por madeira em diversos setores industriais, como o de papel e celulose, painéis de madeira, energia, carvão vegetal para a indústria siderúrgica, postes, postes de vedação e, mais recentemente, madeira serrada (Flores et al. 2018). Apesar da grande adaptabilidade, variações nas condições ambientais e na qualidade do local e pelo seu alto potencial de utilização como matéria prima em diversas

áreas da indústria, o eucalipto requer um manejo eficiente para o aprimoramento da produtividade (Da Cunha, 2021).

No passado, as plantações de eucalipto concentravam-se nas regiões Sul e Sudeste do Brasil, principalmente no bioma Mata Atlântica (Reis et al. 2017; André et al., 2021). Atualmente, buscam-se novas regiões de plantações, que tenham menor valor de terra, maiores incentivos fiscais e maior flexibilidade em termos de infraestrutura e logística, com o objetivo de aumentar a competitividade no mercado internacional. Consequentemente, ao longo dos anos, as plantações se expandiram para outras regiões, principalmente áreas rurais já antropizadas no bioma Cerrado (Fernandes et al. 2016; Oliveira et al., 2020).

Nos últimos dez anos, a área de plantação de eucalipto nessas novas fronteiras florestais aumentou 85%, ou 810.000 ha (Reis et al. 2017, IBÁ 2019). Consequentemente, é fundamental readaptar as práticas silviculturais considerando as novas condições edáficas e climáticas, com especial atenção à seleção e desenvolvimento de novos clones de eucalipto que sejam tolerantes à seca e outros fatores de estresse na região do Cerrado (Oliveira et al., 2020).

Nessa direção, o aumento das áreas ocupadas por florestas de eucalipto pode representar aumento do sequestro de carbono e redução das emissões de GEE (Pereira-Silva et al., 2021). Em estudo realizado por Teodoro et al. (2024) no estado do Mato Grosso do Sul no Brasil, a expansão dos plantios de eucalipto sobre pastagens sugere um aumento no sequestro de carbono (Teodoro et al., 2024).

Nesse contexto, a restauração e conservação dos ecossistemas tropicais do Brasil podem contribuir significativamente para a expansão do estoque de carbono, fornecendo mecanismos flexíveis para atingir as metas climáticas e reduzir as emissões de gases de efeito estufa em diversos biomas, em especial o Cerrado Brasileiro (Barros et al., 2023).

O Eucalipto, é uma importante árvore de rotação curta, espécie que fixa CO_2 atmosférico em biomassa e sequestra carbono a uma taxa mais rápida em comparação a outras espécies florestais de curta rotação (Behera et al., 2020).

Estudo realizado por Chauhan et al. (2009), mostram que o armazenamento de C do caule da árvore (4,20 t ha^{-1}) e o total de C armazenamento (9,36 t ha^{-1}) foram registrados em Eucalyptus tereticornis.

Nesse contexto, as áreas de florestas plantadas podem atuar como agentes sequestradores de carbono, e isso enfatiza a grande importância do manejo adequado para a mitigação do CO_2 e emissões para a atmosfera (Teodoro et al., 2024).

1.2.3 Modelagem da produtividade

É cada vez mais comum o uso de modelos que realizam a estimativa de crescimento e que avaliam o desenvolvimento de cultivos, dessa forma contribuindo para previsão da produtividade dos cultivos e tornando compreensível os fatores que estão envolvidos nas diferentes respostas ao ambiente (Anar et al., 2019).

O emprego de modelos matemáticos e técnicas de machine learning tem se mostrado promissor para modelar as complexas inter-relações florestais (De Oliveira Neto, 2022), proporcionando “insights” valiosos para a gestão florestal. Os modelos de crescimento e produção podem ajudar a simular o crescimento das dimensões das árvores (altura e diâmetro) para prever a produtividade florestal em diferentes níveis (De oliveira et al., 2021).

O uso de técnicas de otimização baseadas em modelos matemáticos pode melhorar o planejamento dos plantios. Diversos modelos baseados em processos ecofisiológicos têm sido utilizados para estimar a produtividade. (Gou et al., 2021)

Existem vários modelos que realizam a estimativa de produtividade da soja e demais cultivos anuais, dentre os mais usados cita-se os modelos mecanísticos DSSAT – “Decision Support System for Agrotechnology Transfer”, que é um sistema de suporte à tomada de decisão, ele integra informações da cultura, solos, clima e manejo de para simular sistemas agrícolas (Jones et al., 2003), e o modelo AQUACROP-FAO que é um modelo de simulação de culturas de sequeiro, focando em áreas onde a água é um fator limitante (Steduto et al., 2009).

Dentre os modelos ecofisiológicos para simulação do desenvolvimento e produtividade do eucalipto, destaca-se o 3-PG, que permite predizer o potencial produtivo da floresta em função das variáveis ambientais e das práticas de manejo (De Freitas et al., 2020).

O APSIM (Agricultural Production Systems sIMulator) é um modelo de simulação de sistemas agrícolas que é usado para prever a produtividade de várias culturas em diferentes condições ambientais e de manejo agrícola (Huth et al., 2001). Ele considera uma ampla gama de processos biológicos e físicos, incluindo crescimento de culturas, interações com o solo, manejo de água e nutrientes, entre outros (Holzworth et al., 2014).

O Forest-DNDC é um modelo de simulação que combina o modelo DNDC (Denitrification-Decomposition) com processos específicos de ecossistemas florestais. Ele é usado principalmente para avaliar os fluxos de carbono, nitrogênio e gases de efeito estufa em florestas e ecossistemas terrestres (Li et al., 2000).

O modelo FAO é um modelo matemático-fisiológico que simula a fotossíntese bruta em uma etapa de tempo diária de acordo com o mecanismo de fixação de carbono e adaptação climática da planta. A produtividade potencial (Yp) é simulada de acordo com as interações do genótipo com a radiação solar, fotoperíodo e temperatura do ar. Posteriormente, o Yp é penalizado pelo déficit hídrico em diferentes períodos da rotação do eucalipto e de acordo com a intensidade do déficit hídrico acumulado (Elli et al., 2019).

Apesar de generalistas, esses modelos são complexos, e necessitam um grande número de parâmetros, o que deixa a sua aplicação mais desafiadora (De Freitas et al., 2020). A complexidade do ambiente florestal recomenda que a interação entre plantas e fatores ambientais seja analisada em conjunto, pois o todo é maior que a soma de suas partes (Billings, 1952). Para considerar todos os atributos que definem uma floresta simultaneamente, devemos reconhecer não apenas sua dinâmica, mas também como eles mudam como uma entidade ao longo do tempo (Elli et al., 2019).

Para construir modelos com simulem as interações não lineares complexas entre variáveis de um sistema, as pesquisas têm utilizado uma ampla gama de métodos numéricos, matemáticos ou estatísticos (Lek e Guégan 1999). No entanto essas técnicas convencionais, têm dificuldade em modelar esses comportamentos complexos e não lineares, mas com o surgimento das técnicas de inteligência artificial como as redes neurais artificiais (RNA), torna-se possível a modelagem empírica de tais sistemas com alta acurácia (Gue et al., 2020).

O aprendizado de máquina, que é uma área da Inteligência Artificial (IA) com foco no aprendizado, é uma abordagem prática que pode fornecer melhor previsão de produtividade com base em várias variáveis independentes ou features. O aprendizado de máquina pode determinar padrões e correlações e descobrir conhecimento a partir de conjuntos de dados. Os modelos precisam ser treinados usando conjuntos de dados, onde os resultados são representados com base na experiência passada. O modelo de estimação é construído usando várias features e, como tal, os parâmetros dos modelos são determinados usando dados históricos durante a fase de treinamento. Para a fase de teste, parte dos dados históricos que não foram usados para treinamento é usada para fins de avaliação de desempenho (Van Klompenburg et al., 2020).

Há diversos algoritmos de aprendizagem de máquina disponíveis, por exemplo, o Random Forest (RF) que é um algoritmo que ajusta várias árvores de decisão de várias subamostras do conjunto de dados e usa uma árvore média para melhorar a precisão preditiva e controlar o ajuste excessivo (Breiman, 2000). A Redes Neurais Artificiais (RNA) que é um algoritmo de aprendizagem supervisionada, ele aprende uma função de treinamento em um conjunto de dados para fornecer uma ou mais saídas. (Basheer; Hajmeer, 2000). O Support vector machines (SVM) que é um algoritmo de ML avançado que funciona separando vetores de suporte à distância máxima usando um hiperplano (Tehrany et al., 2015). O XGBoost é um algoritmo baseado no aumento de árvores de decisão, que usa uma expressão eficiente de segunda ordem. Esse modelo é generalizável e evita o “overfitting” e “underfitting” das estimações (Chen e Guestrin, 2016; De Souza Diniz et al., 2023).

Neste sentido, a seleção do melhor modelo para estimar a produtividade de lavouras e florestas depende principalmente dos dados de entrada, das informações disponíveis para sua calibração e avaliação, e do nível de detalhamento e precisão almejado pelo usuário (Pasquel et al. 2022).

1.3 Objetivo geral

O objetivo principal deste trabalho foi testar e avaliar diferentes modelos de machine learning para previsão da produtividade de soja e estimação do volume de madeira de eucalipto em diferentes idades de crescimento.

REFERÊNCIAS

Anar, M. J., Lin, Z., Hoogenboom, G., Shelia, V., Batchelor, W. D., Teboh, J. M., & Khan, M. (2019). Modeling growth, development and yield of Sugarbeet using DSSAT. **Agricultural systems**, 169, 58-70.

André, J. L., Oliveira, R. D. S., Sette Jr, C. R., Alfenas, A. C., Zauza, E. Â. V., de Siqueira, L., & Novaes, E. (2021). Wood volume of Eucalyptus clones established under different spacings in the Brazilian Cerrado. **Forest Science**, 67(4), 478-489.

Araújo MLS, Sano EE, Bolfe ÉL, Santos JRN, dos Santos JS, Silva FB (2019) Spatiotemporal dynamics of soybean crop in the Matopiba region, Brazil (1990–2015). **Land Use Policy**, v. 80, p. 57-67.

Barbosa dos Santos, V., Santos, A. M. F. D., & Rolim, G. D. S. (2021). Estimation and forecasting of soybean yield using artificial neural networks.**Agronomy Journal**,*113*(4), 3193-3209.

Barros, F. D. V., Lewis, K., Robertson, A. D., Pennington, R. T., Hill, T. C., Matthews, C., & Rowland, L. (2023). Cost-effective restoration for carbon sequestration across Brazil's biomes. **Science of The Total Environment**, *876*, 162600.

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. **Journal of microbiological methods**, 43(1), 3-31.

Behera, L., Ray, L. I., Ranjan Nayak, M., & Mehta, A. (2020). Carbon sequestration potential of Eucalyptus spp.: A review. **E-Planet**, *18*(1), 79-84.

Benevides, R., & Staback, D. (2023). PERFIL LOCACIONAL DA SOJA: UM ESTUDO DAS MESORREGIÕES PARANAENSES PARA OS ANOS DE 2000, 2010 E 2020. **Informe Gepec**, 27(2).

Billings, W. Dwight. The environmental complex in relation to plant growth and distribution. **The Quarterly Review of Biology**, v. 27, n. 3, p. 251-265, 1952.

Binkley, D., & Fisher, R. F. (2019). *Ecology and management of forest soils*. John Wiley & Sons.

Binkley, D., Campoe, O. C., Alvares, C., Carneiro, R. L., Cegatta, Í., & Stape, J. L. (2017). The interactions of climate, spacing and genetics on clonal Eucalyptus plantations across Brazil and Uruguay**. Forest Ecology and Management**, *405*, 271-283.

Breiman, L. (2000). Randomizing outputs to increase prediction accuracy. **Machine Learning**, 40, 229-242.

Campolina, B. (2019). Population growth and urbanization in the Brazilian Cerrado. **Innovation with Spatial Impact: Sustainable Development of the Brazilian Cerrado**, 163188.

Carneiro Filho, A. R. N. A. L. D. O., & Costa, K. A. R. I. N. E. (2016). A expansão da soja no cerrado. Caminhos para a ocupação territorial, uso do solo e produção sustentável. São Paulo, **Agroicone**, p1-30.

Carneiro Filho, A., & Costa, K. (2016). The expansion of soybean production in the Cerrado: Paths to sustainable territorial occupation, land use and production. *INPUT-AGROICONE (https://www.inputbrasil.org/wpcontent/uploads/2016/11/The-expansion-of-soybean-production-in-the-Cerrado_Agroicone_INPUT.pdf). Acesso em* 27de mar de 2024

Carvalho FMV, Marco-Júnior P, Ferreira LG. 2009. The Cerrado into-pieces: habitat fragmentation as a function of landscape use in the savannas of central Brazil. **Biological Conservation** 142: 1329–1403

Castro, C. A. D. O., Resende, R. T., Bhering, L. L., & Cruz, C. D. (2016). Brief history of Eucalyptus breeding in Brazil under perspective of biometric advances. **Ciência Rural**, *46*, 1585-1593.

Chauhan, S.K., Gupta, N., Ritu, S., Yadav and Chauhan, R. 2009. Biomass and carbon allocation in different parts of agroforestry tree species. **Indian Forester** 135(7): 981-993

Costa, A. N., Bartimachi, A., Vasconcelos, H. L., Bruna, E. M., & Vieira-Neto, E. H. (2020). Annual litter production in a Brazilian Cerrado woodland savanna. **Southern Forests: a Journal of Forest Science**, *82*(1), 65-69.

da Cunha, Thammi Queuri Gomes et al. Eucalyptus expansion in Brazil: Energy yield in new forest frontiers. **Biomass and Bioenergy**, v. 144, p. 105900, 2021.

de Freitas, E. C. S., de Paiva, H. N., Neves, J. C. L., Marcatti, G. E., & Leite, H. G. (2020). Modeling of eucalyptus productivity with artificial neural networks. **Industrial Crops and Products**, *146*, 112149.

de Freitas, P. L., & Landers, J. N. (2014). The transformation of agriculture in Brazil through development and adoption of zero tillage conservation agriculture. **International Soil and Water Conservation Research**, 2(1), 35-46.

de Oliveira Neto, Ricardo Rodrigues et al. Estimation of Eucalyptus productivity using efficient artificial neural network. **European Journal of Forest Research**, v. 141, n. 1, p. 129-151, 2022.

de Oliveira, Bruno Rodrigues et al. Eucalyptus growth recognition using machine learning methods and spectral variables. **Forest Ecology and Management**, v. 497, p. 119496, 2021.

de Souza, R. F., Oliveira, G. R., Freitas, E. G., Pinheiro, A. C., & de Souza, R. N. (2023). Agricultura no Cerrado e impactos ambientais decorrentes. ***Observatório de la economía latinoamericana***, *21*(12), 25068-25081.

Elli, E. F., Sentelhas, P. C., de Freitas, C. H., Carneiro, R. L., & Alvares, C. A. (2019). Intercomparison of structural features and performance of Eucalyptus simulation models and their ensemble for yield estimations. **Forest ecology and management**,*450*, 117493.

Eloy, L., Aubertin, C., Toni, F., Lúcio, S. L. B., & Bosgiraud, M. (2017). On the margins of soy farms: traditional populations and selective environmental policies in the Brazilian Cerrado. **Soy, Globalization, and Environmental Politics in South America** (pp. 244-266). Routledge.

Eloy, L., Schmidt, I. B., Borges, S. L., Ferreira, M. C., & Dos Santos, T. A. (2019). Seasonal fire management by traditional cattle ranchers prevents the spread of wildfire in the Brazilian Cerrado. **Ambio**, 48, 890-899.

Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA (2019) MATOPIBA. Disponível em:<https://www.embrapa.br/tema-matopiba>. Acesso em: 27 mar. 2024

Fernandes GW, Coelho MS, Machado RB, Ferreira ME, Aguiar LMS, Dirzo R, Scariot A and Lopes CR (2016) Afforestation of savannas: an impending ecological disaster. **Natureza & Conservação** 14: 146151.

Ferreira, M. E., Ferreira, L. G., Latrubesse, E. M., & Miziara, F. (2016). Considerations about the land use and conversion trends in the savanna environments of Central Brazil under a geomorphological perspective. **Journal of Land Use Science**, 11(1), 33–47.

Flores, T. B., Alvares, C. A., Souza, V. C., & Stape, J. L. (2018). *Eucalyptus in Brazil: climatic zoning and identification guide*. Piracicaba, Brazil: IPEF.

Garcia, L. C., Szabo, J. K., de Oliveira Roque, F., Pereira, A. D. M. M., da Cunha, C. N., Damasceno-Júnior, G. A., & Ribeiro, D. B. (2021). Record-breaking wildfires in the world's largest continuous tropical wetland: Integrative fire management is urgently needed for both biodiversity and humans. **Journal of environmental management**, 293, 112870.

Glatzle, S., de Almeida, R. G., Pereira Barsotti, M., Bungenstab, D. J., Giese, M., Macedo, M. C. M., & Asch, F. (2024). Integrated Land-Use Systems Contribute to Restoring Water Cycles in the Brazilian Cerrado Biome. **Land**, *13*(2), 221.

Gue, I. H. V., Ubando, A. T., Tseng, M. L., & Tan, R. R. (2020). Artificial neural networks for sustainable development: a critical review**. Clean Technologies and Environmental Policy**, 22, 1449-1465.

Guo, Y., Zhao, H., Zhang, S., Wang, Y., & Chow, D. (2021). Modeling and optimization of environment in agricultural greenhouses for improving cleaner and sustainable crop production. **Journal of Cleaner Production**, 285, 124843.

Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., & Keating, B. A. (2014). APSIM–evolution towards a new generation of agricultural systems simulation**. Environmental Modelling & Software**, *62*, 327-350.

Hunke, P., Mueller, E. N., Schröder, B., & Zeilhofer, P. (2015). The Brazilian Cerrado: assessment of water and soil degradation in catchments under intensive agricultural use. **Ecohydrology**, *8*(6), 1154-1180.

Huth, N. I., Snow, V. O., & Keating, B. A. (2001, December). Integrating a forest modelling capability into an agricultural production system modelling environment-current applications and future possibilities. In *Proceedings of the International Congress on Modelling and Simulation, Canberra, Australia* (pp. 1895-1900).

IBÁ - Indústria Brasileira de Árvores (2019). Relatório Anual 2019. IBÁ, Brasília, 80p.

IBÁ - Indústria Brasileira de Árvores (2023). Relatório Anual 2023. IBÁ, Brasília, 91p.

INPE. (2020). Banco de Dados de queimadas. Retrieved March 3, 2019. Disponível em: http://queimadas.dgi.inpe.br/queimadas/bdqueimadas.

Jesus, F. R. (2023). A Expansão do agronegócio e o desenvolvimento socioeconômico de municípios da nova fronteira agrícola (Matopiba): uma análise de 2000 e 2010. 110 f. Dissertação (Mestrado em Economia Rural) - Programa de Pós-Graduação em Economia Rural, Universidade Federal do Ceará.

Klink, C. A., & Machado, R. B. (2005). Conservation of the Brazilian cerrado. **Conservation biology**, *19*(3), 707-713.

Lek, S., & Guégan, J. F. (1999). Artificial neural networks as a tool in ecological modelling, an introduction. **Ecological modelling**, 120(2-3), 65-73.

Li, C., Aber, J., Stange, F., Butterbach-Bahl, K., & Papen, H. (2000). A process-oriented model of N2O and NO emissions from forest soils: 1. Model development. **Journal of Geophysical Research: Atmospheres**, *105*(D4), 4369-4384.

Lima, T.C., Ribeiro, S.C., Soares-Filho, B., 2018. Integrating econometric and spatially explicit dynamic models to simulate land use transitions in the Cerrado biome. In: Olmedo, M.T.C., Paegelow, M., Mas, J.F., Escobar, F. (Eds.), Geomatic Approaches for Modeling Land Change Scenarios. Springer, Switzerland, pp. 399–417.

Loayza, A. C. V., dos Reis, M. V. S., de Jesus, F. R., Ipolito, A. L. M., & Ribeiro, I. G. (2023). Evolução dos indicadores da produção de soja no Matopiba. **Observatório de la economía latinoamericana**,*21*(12), 27824-27845.

Magalhães, I. B., de Paula Pereira, A. S. A., Calijuri, M. L., do Carmo Alves, S., dos Santos, V. J., & Lorentz, J. F. (2020). Brazilian Cerrado and Soy moratorium: Effects on biome preservation and consequences on grain production. **Land use policy**, 99, 105030.

Magalhães, I. B., de Paula Pereira, A. S. A., Calijuri, M. L., do Carmo Alves, S., dos Santos, V. J., & Lorentz, J. F. (2020). Brazilian Cerrado and Soy moratorium:

Effects on biome preservation and consequences on grain production. **Land use policy**, *99*, 105030.

Maia, S. M. F., de Souza Medeiros, A., dos Santos, T. C., Lyra, G. B., Lal, R., Assad, E. D., & Cerri, C. E. P. (2022). Potential of no-till agriculture as a nature-based solution for climate-change mitigation in Brazil. **Soil and Tillage Research**, 220, 105368.

Ministério do Meio Ambiente (MMA), 2019. O Bioma Cerrado. Disponível em: https://www.mma. gov.br/biomas/cerrado.

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., & Dolman, A. J. (2011). Magnitude and variability of land evaporation and in the central United States attributed to agricultural intensification**. Geophysical Research Letters**, 45(3), 1586–1594.

Neto, M. S., Scopel, E., Corbeels, M., Cardoso, A. N., Douzet, J.-M., Feller, C., Piccolo, M. D. C., Cerri, C. C., & Bernoux, M. (2010). Soil carbon stocks under no-tillage mulch-based cropping systems in the Brazilian Cerrado: An on-farm synchronic assessment. **Soil and Tillage Research**, 110(1), 187–195.

Oliveira, A.; Oliveira, J.J.; Hirakuri, M.H. Desenvolvimento, Mercado, Rentabilidade da Soja Brasileira. 2010. Disponível em: https://www.embrapa.br/buscadepublicacoes//publicacao/854125/desenvolvimento-mercado-rentabilidade-da-soja-brasileira (acesso em 27 de mar de 2024).

Oliveira, R. D. S., Ribeiro, C. V. G., Neres, D. F., Porto, A. C. D. M., Ribeiro, D., Siqueira, L. D., & Novaes, E. (2020). Evaluation of genetic parameters and clonal selection of Eucalyptus in the Cerrado region. **Crop Breeding and Applied Biotechnology**, *20*, e29982031.

Oliveira-Filho AT, Ratter, JT. (2002). Vegetation Physiognomies and Woody Flora of the Cerrado Biome. In Oliveira, P.S. and Marquis, R.J. (eds.), The Cerrados of Brazil: Ecology and Natural History of a Neotropical Savanna. New York: Columbia University Press. pp 91–120

Pasquel, D., Roux, S., Richetti, J., Cammarano, D., Tisseyre, B., & Taylor, J. A. (2022). A review of methods to evaluate crop model performance at multiple and changing spatial scales.**Precision Agriculture**, *23*(4), 1489-1513.

Pereira LI, Pauli L (2016) O processo de estrangeirização da terra e expansão do agronegócio na região do Matopiba. Campo-Território: **Revista de Geografia Agrária**, v. 11, n. 23 jul.

Pereira-Silva, E. F., Gardon, F. R., Hardt, E., Keller, V. C., & dos Santos, R. F. (2021). Carbon ecosystem services and cellulose income from natural and commercial forests in the Brazilian savanna. **Forest Ecology and Management**, *499*, 119582.

Pivello, V. R. (2011). The use of fire in the Cerrado and Amazonian rainforests of Brazil: Past and present. **Fire Ecology**, 7(1), 24–39.

Reis CAF, Talone Neto A, Brunckhorst A, Moreira JMMAP, Pereira AV and Moraes AC (2017) Cenário do setor de florestas plantadas no estado de Goiás. Embrapa Florestas, Colombo, 80p.

Ruas, A. C. P., & Schettino, S. (2022). Ainda, sobre os paradigmas associados ao cultivo do eucalipto no cerrado. **Pesquisas agrárias e ambientais**. Nova Xavantina, MT: Pantanal Editora, 2022. v. 11.

Soares-Filho, B., Rajão, R., Macedo, M., Carneiro, A., Costa, W., Coe, M., & Alencar, A. (2014). Cracking Brazil's forest code. **Science**, 344(6182), 363-364.

Tehrany, M. S., Pradhan, B., Mansor, S., & Ahmad, N. (2015). Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. **Catena**, 125, 91-101.

Teodoro, P. E., Rossi, F. S., Teodoro, L. P. R., Santana, D. C., Ratke, R. F., de Oliveira, I. C., & da Silva Junior, C. A. (2024). Soil CO2 emissions under different land-use managements in Mato Grosso do Sul, Brazil. **Journal of Cleaner Production**, *434*, 139983.

Toloi, M. N. V., Bonilla, S. H., Toloi, R. C., Silva, H. R. O., & Nääs, I. D. A. (2021). Development indicators and soybean production in Brazil. **Agriculture**, 11(11), 1164.

Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. **Computers and Electronics in Agriculture**,*177*, 105709.

West, T. A., & Fearnside, P. M. (2021). Brazil’s conservation reform and the reduction of deforestation in Amazonia. **Land use policy**, *100*, 105072.

CAPÍTULO 2: Algoritmos de aprendizado de máquina para previsão de produtividade de soja no Cerrado brasileiro.

Resumo – Quantificar os efeitos adversos do clima, principalmente em área de expansão agrícola como a região Sul do Maranhão, pertencente a fronteira de grãos do MATOPIBA, é uma importante ferramenta para o planejamento e gerenciamento das atividades nas lavouras de soja. O objetivo deste trabalho é avaliar diferentes modelos de ML para previsão de produtividade de soja para o sul do Maranhão, com até um mês antecedência. O estudo consiste em 13 locais no Sul do Maranhão. Dados meteorológicos mensais foram coletados na plataforma NASA-POWER e de produtividade de soja na base do SIDRA/IBGE entre os anos de 2008 a 2017 permitindo assim o cálculo do armazenamento de água no solo (STO), excedente hídrico (EXC), evapotranspiração real (ETa), Déficit hídrico (DEF) e evapotranspiração das culturas (ETc).Os modelos de ML avaliados foram, Random Forest (RF), Redes neurais artificiais (RNA), Support vector machines Base Radial (SVM_RBF), modelo linear (SVM_LIN) e regressão polinomial (SVM_POLY). Como método de avaliação do desempenho dos modelos, utilizou-se a validação cruzada, obtendo-se o valor médio de precisão pelo R2 e acurácia pelo RMSE. Os resultados mostraram que o algoritmo RF atinge a maior precisão e acurácia, com R^2 de 0,81, RMSE de 176,93 kg ha^{-1} e tendência (EME) de 1,99 kg ha^{-1}. Por outro lado, o algoritmo SVM_RBF apresentou o menor desempenho com R^2 de 0,74, RMSE de 213,58 kg ha^{-1} e EME de 15,06 kg ha^{-1}. Os valores de produtividade média previstos pelos modelos ficaram dentro do esperado para a região, que possui valor médio histórico de 2.730 kg ha^{-1}. Todos os modelos apresentaram precisão, acurácia e tendência aceitáveis, o que possibilita a utilização de todos os algoritmos avaliados na previsão da produtividade da cultura da soja, observando as particularidades da região a serem estudada, além de ser uma ferramenta útil para o planejamento agrícola e tomada de decisões em regiões produtoras de soja, como o Cerrado brasileiro.

Palavras-chave: machine learning, modelo de cultivo, agrometeorologia, *soybean*

2.1 Introdução

A soja é um produto agrícola amplamente consumido em todo o mundo, podendo ser usado no consumo humano, animal e na produção de biocombustíveis (Alambert et al., 2019). A produção mundial de soja foi superior a 347 milhões de toneladas, no ano /safra 2017/18, quando foram plantados mais de 126 milhões de hectares (USDA, 2018). O Brasil, com uma produção de 107 milhões de toneladas, foi responsável por 30% de toda soja produzida no mundo (CONAB, 2018).

No Brasil, a região do MATOPIBA (Maranhão, Tocantins, Piauí e Bahia), caracteriza uma importante área de expansão crescente de grãos, representando 14% da produção brasileira de soja (CONAB, 2018). produção de soja aumentou nos quatro Estados que compõem a região do MATOPIBA, com incrementos de 26,4% na Bahia, 23,1% no Piauí, 13,5% no Maranhão e 20,2% em Tocantins (CONAB, 2018).

No Maranhão, 73% da produção de soja se concentra na região Sul do estado (IBGE, 2019), compondo parte do cerrado brasileiro. No entanto, essa região sobre grande influência da Amazônia maranhense, alterando as condições de clima e consequentemente previsões e estimativas de produtividade deste local.

O clima é um dos principais fatores causadores da queda na produtividade da cultura da soja (Sentelhas et al., 2015), contendo características ambientais como temperatura, precipitação pluvial, radiação solar e outros elementos meteorológicos que são controladas pela natureza (Shine et al., 2018). Portanto, melhorar a tomada de decisão com base em modelos de previsão, tornam-se uma importante ferramenta para o planejamento e gerenciamento da atividade agrícola (Moraes et al., 2020).

A melhor forma de sintetizar quantitativamente os efeitos do clima sobre a produtividade agrícola, é por meio de algoritmos de seleção de variáveis. A calibração e teste dos seus parâmetros com base no conjunto de características adicionadas, fazem uma previsão precisa das safras agrícolas (Gopal e Bhargavi, 2019).

Como resultado, os modelos baseados em dados climáticos ganharam aplicações importantes para estimativas e previsões em áreas agrícolas, utilizando métodos de aprendizagem automática (ML) (Mbangiwa e Mabhaudhi,

2019). Algoritmos de aprendizado de máquina, como floresta aleatória, redes neurais e máquina de vetores de suporte, foram usados com sucesso para prever a produtividade das culturas (Alghamdi et al, 2019).

Estudos como o de Cai et al., (2019) integraram várias fontes de dados para prever a produção de trigo na Austrália de 2000 a 2014, os autores usaram o método de regressão conhecido LASSO, como referência e três métodos principais de aprendizado de máquina (máquina de vetores de suporte, floresta aleatória e rede neural) para construir vários modelos empíricos para previsão de produtividade, e confirmaram que a combinação de dados climáticos e de satélite pode alcançar alto desempenho de previsão com R^2 de 0,75.

Sakamoto (2020) usou random forest para estimar a produtividade de milho e soja dos Estados Unidos (EUA), obtendo elevada precisão. Schwalbert et al., (2020) fizeram a previsão da produtividade da soja usando Random forest e redes neurais, Alves et al., (2018) usaram redes neurais artificiais, Michelon et al., (2017) usaram support vector machine.

Embora sejam ferramentas robusta, os modelos de aprendizado de máquina ainda são pouco utilizados em estudos de cultivo de soja, principalmente no Brasil. Portanto, o presente estudo tem como objetivo avaliar diferentes modelos de machine learning para previsão de produtividade de soja para o sul do Maranhão com até um mês de antecedência.

2.2 Material e métodos

2.2.1 Localização e caracterização da área de estudo

A área de estudo consiste em locais da região Sul do estado do Maranhão, área da fronteira agrícola do MATOPIBA, uma das mais importantes áreas de produção agrícola do Brasil, responsável por mais de 11% da safra brasileira (Tabela 1 e Figura 1).

Tabela 1 – Descrição geográfica das regiões produtoras de soja na região Sul do Maranhão.

Local	Latitude	Longitude	Altitude	Produtividade (Kg ha^{-1})
Estreito	6º33'38"	47º27'04"	153	2.364,5
Sucupira do Norte	6°28'37"	44°11'31"	480	2.820,5
Fortaleza dos Nogueiras	6°57'50"	46°10'37"	443	2.836,5
Pastos Bons	6°36'07"	44°04'37"	309	2.882,1
São Domingos do Azeitão	6°48'36"	44°38'42"	308	2.936,2
Loreto	7°05'02"	45°08'27"	193	2.654,7
São Raimundo das Mangabeiras	7°01'19	45°28'51"	225	2.691,0
Sambaíba	7°08'24"	45°20'45"	205	2.697,9
Riachão	7º21'43"	46º37'02"	383	2.723,2
Balsas	7°31'58"	46°02'09"	283	2.734,3
Carolina	7°19'58	47°28'08	148	2.603,5
Tasso Fragoso	8°28'30	45°44'34"	200	2.704,1
Alto Parnaíba	9º06'39"	45°55'48"	280	2.823,9

Fonte: IBGE, 2020

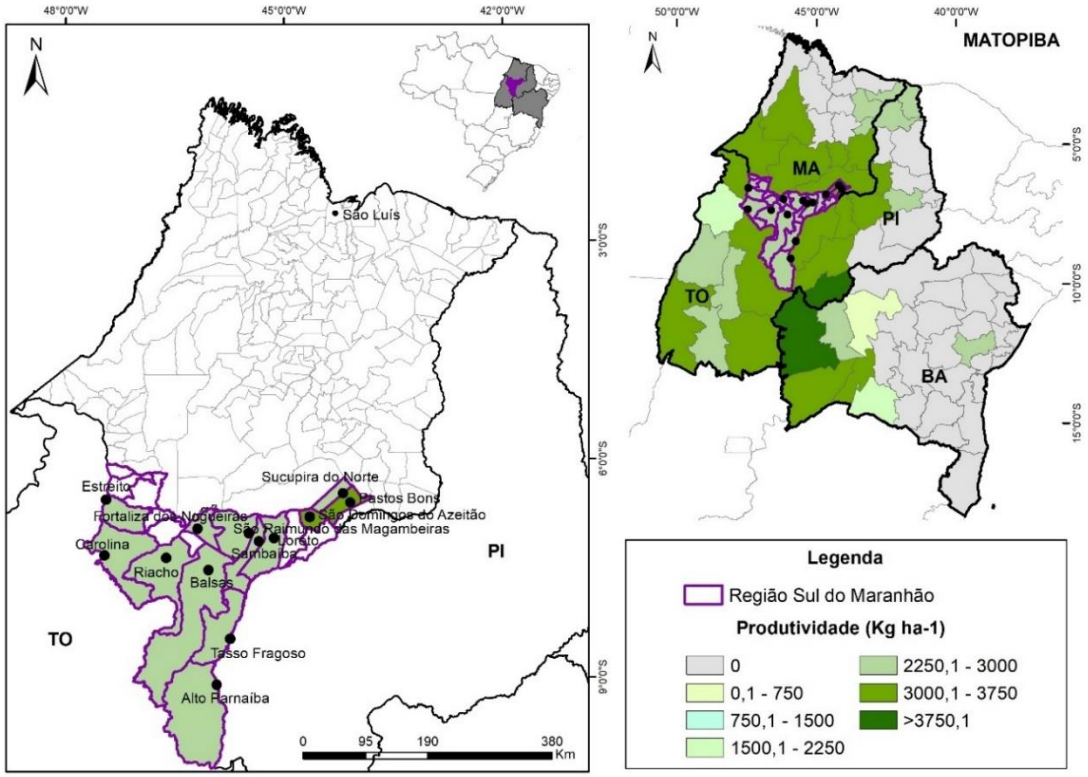

Figura 1 - Mapa de localização do MATOPIBA e região Sul do Maranhão.

A temperatura média da série histórica (Figura 2. A) para a região varia entre 26 e 34 °C; nota-se que a área de estudo apresenta temperaturas médias predominantemente amenas entre 26 e 28 °C. A soja adapta-se melhor a regiões

onde as temperaturas oscilam entre 20 e 30 °C. A precipitação média da série histórica (Figura 2. B) variou de aproximadamente 600 a 850 mm. Nota-se que o extremo sul da região apresenta os menores índices pluviométricos, o que influencia diretamente no resultado do balanço hídrico, principalmente no armazenamento de água no solo, que por sua vez influencia diretamente no desenvolvimento e produtividade da cultura.

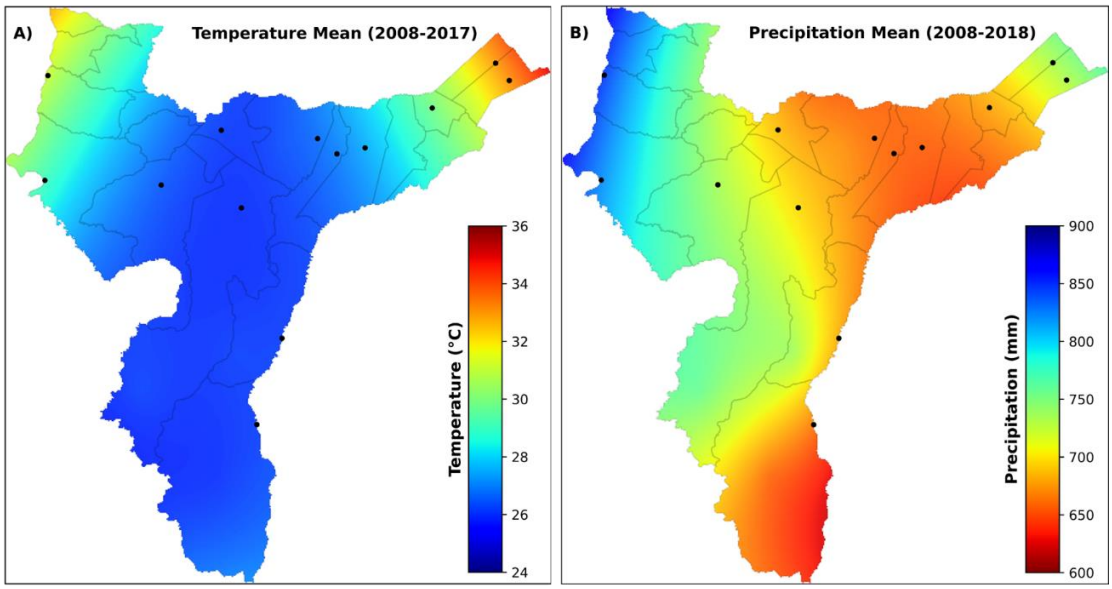

Figura 2 – Representação climática da área de estudo: (A) temperatura média (°C) da série histórica (2008–2017); (B) média de precipitação (mm) da série histórica (2008–2017).

Os dados de produtividade foram obtidos no Sistema de Recuperação Automática do Instituto Brasileiro de Geografia e Estatística – SIDRA/IBGE, na base de dados da Produção Agrícola Municipal – PAM, que investiga um conjunto de produtos das lavouras temporárias e permanentes do Brasil anualmente, entre os anos de 2008 e 2017 (IBGE,2019).

Os dados de produtividade foram ajustados conforme proposto por Prela-Pantano et al. (2011), a fim de eliminar a tendência tecnológica. Este ajustamento é necessário para minimizar os efeitos devidos a alterações no nível tecnológico empregado pelos produtores, obtendo assim a influência da variabilidade climática na produtividade.

2.2.2 Conjunto de dados

Foram coletados dados meteorológicos diários de precipitação pluvial (P, mm dia^{-1}), temperatura máxima e mínima do ar (T, °C), temperatura no ponto de orvalho (To, °C), umidade relativa (UR %), radiação solar no topo da atmosfera (Qo, Mj m^{-2} dia^{-1}), radiação solar global (Qg, Mj m^{-2} dia^{-1}) e velocidade média do vento (VVm, m s^{-1}) entre os anos de 2007 e 2017 para cada localidade.

Os dados meteorológicos foram obtidos na plataforma NASA-POWER (Stackhouse et al., 2017), utilizado a linguagem de programação em Python 3.6 por meio de uma interface de programação de aplicação - API, que fornece dados com resolução espacial de 1° e temporal variando dependendo do uso do produto de dados (Stackhouse et al., 2016). A evapotranspiração potencial de referência foi calculada pelo método de Camargo 1971 (Equações 1 e 2) e em seguida obteve-se a evapotranspiração de cultivo, multiplicando a evapotranspiração potencial com o coeficiente de cultivo kc segundo Evangelista (2017) (Tabela 2).

$$ETP = 0{,}01 \times \frac{Qo}{2{,}45} \times T \times NDA \quad (1)$$

$$ETc = ETP \times Kc \quad (2)$$

em que: ETP é a evapotranspiração potencial, Qo é a radiação solar no topo da atmosfera, T é a temperatura média, NDA é o número de dias Juliano e o Kc é o coeficiente de cultivo da soja.

Tabela 2 – Coeficiente de cultivo (Kc) da soja.

Estádios	**Kc**
Estabelecimento –S-V2	0,58
Desenvolvimento – V2-R1	1,06
Floração – R1 – R3	1,39
Frutificação – R3 - R7	1,09
Maturação – R7- R8	0,55

Legenda: S-semeadura; V- vegetativo; R-reprodutivo, Kc da soja segundo Evangelista, 2017

O balanço hídrico mensal foi realizado pelo método de Thornthwaite e Mather (1955), com capacidade de água disponível (AWC) igual a 100 mm. Com a quantificação do BH, foram selecionadas Evapotranspiração de cultivo (mm) (CET), Evapotranspiração real (mm) (AET), Armazenamento (mm) (STO), Déficit (mm) (DEF) e Excedente (mm) (EXC) hídricos, além da temperatura média do ar, precipitação pluvial para compor o conjunto de variáveis independentes nos modelos de machine learning em cada estádio fenológico da planta.

A base de dados final foi composta por uma matriz de 36 colunas por 936 linhas (Figura 3). Cada uma das variáveis independentes T, P, CET, AET, STO, DEF and EXC foi estratificada em cinco subvariáveis correspondentes aos cinco meses do ciclo da soja determinados de acordo com o Calendário de Plantio e Colheita de Grãos no Brasil (Conab, 2019), os meses em que a semeadura (outubro a fevereiro) e a colheita (fevereiro a julho) de diversas culturas agrícolas é realizada ao longo do ano de acordo com a região do país. A soja de ciclo médio comumente adotada no MATOPIBA varia entre 126 e 137 dias.

Independent variables (Features)						Dependent variable
Mês	**NOV**	**DEZ**	**JAN**	**FEV**	**MAR**	
T	T_1	T_2	T_3	T_4	T_5	**Annual average yield**
P	P_1	P_2	P_3	P_4	P_5	
CET	CET_1	CET_2	CET_3	CET_4	CET_5	
AET	AET_1	AET_2	AET_3	AET_4	AET_5	
STO	STO_1	STO_2	STO_3	STO_4	STO_5	
DEF	DEF_1	DEF_2	DEF_3	DEF_4	DEF_5	
EXC	EXC_1	EXC_2	EXC_3	EXC_4	EXC_5	
Phases	**S-V6**	**R1-R2**	**R3-R5**	**R6-R8**	**R9**	

Figura 3 – Representação dos dados de entrada nos modelos. Legenda: T – Temperatura (°C); P – Precipitação (mm); CET – Evapotranspiração da cultura (mm); AET – Evapotranspiração real da cultura (mm); STO – Armazenamento (mm); DEF – Déficit hídrico (mm); EXC – Excesso hídrico (mm)

2.2.3 Modelos de machine learning

2.2.3.1 Random Forest – RF

O algoritmo Random Forest (RF) ajusta várias árvores de decisão a partir de várias sub amostras do conjunto de dados, e usa uma árvore média para melhorar a precisão preditiva e controlar o ajuste excessivo.

O parâmetro Profundidade máxima (MD) é que compreende a profundidade máxima profundidade da árvore na floresta foram ajustadas, utilizou-se MD = 10.

O n-estimadores (NE) representam o número de árvores na floresta, e entende-se que quanto maior o número de árvores que compõem a floresta, melhores serão os resultados, porém o tempo de execução do ajuste torna-se maior.Foi utilizado um NE = 5 para todas as simulações.

É importante ressaltar que o ajuste do parâmetro NE tende a um limite em que os resultados deixarão de ficar significativamente melhores.

2.2.3.2 RNA – Multilayer Perceptrom – RNA – MLP

O MLP é um algoritmo de aprendizado supervisionado. Ele aprende uma função de treinamento em um conjunto de dados para proporcionar uma ou mais saídas. Para se obter índices ótimos de previsão pela RNA, foram ajustadas diferentes hiperparâmetros.

Ajustou-se o hiperparâmetro “activation” que corresponde a função de ativação para ‘relu’: função de unidade linear retificada, retorna f (x) = max (0, x). O “solver” que é o solucionador de otimização de peso = ‘lbfgs’ que é um algoritmo de otimização de memória limitada Broyden–Fletcher–Goldfarb–Shanno – correspondendo a um otimizador na família de métodos quasi-Newton.

Determinou-se a taxa de aprendizado “constant”. E a taxa de aprendizado inicial para 0.001, além do parâmetro de penalidade “alpha” para 0.05. O hiperparametro “max_inter” que corresponde ao número de interações foi ajustando em 4000 e o “Random_state” igual a 0 (zero).

O hiperparâmetro “hidden_layer_sizes” corresponde ao número de neurônios em cada camada, foram realizados testes com três camadas ocultas, variando o número de neurônios em cada camada. Para obter o número de

neurônios foi utilizado um gerador de números aleatórios variando de 2 a 50, a fim de otimizar o tempo de processamento.

2.2.3.3 Suport vector machines - SVM

O SVM é um algoritmo avançado de aprendizado de máquina que funciona separando os vetores de suporte à distância máxima usando um hiperplano (Tehrany et al., 2015). Funciona com maior desempenho com número limitado de amostras. (Shaharum et al., 2018).

Vários núcleos estão disponíveis no SVM, Função de Base Radial (RBF), modelo Linear, e regressão polinomial foram escolhidos para regressão. Os parâmetros ajustados foram: “kernel”, que corresponde ao núcleo do algoritmo, o parâmetro “C” que é Parâmetro de regularização, o “gama”, o “degree” que corresponde ao grau da função polinomial do núcleo (“poli”). Todos os parâmetros foram ajustados para produzir o melhor resultado.

2.2.4 Avaliação do modelo

Nesta pesquisa utilizamos a validação cruzada como método de avaliação do desempenho do modelo. Usamos o pacote “Cross-validation” (CV) da biblioteca Scikit-Learn que usa uma abordagem denominada “k- fold CV”, onde o conjunto de dados é dividido em k conjuntos menores. A medida de desempenho relatada pela validação cruzada é então a média dos valores calculados no loop. (Pedregosa et al, 2011).

A CV utilizando o k-fold dividiu o conjunto de dados disponíveis aleatoriamente em 10 subconjuntos (k-fold=10). Para cada execução, um subconjunto foi utilizado para avaliação dos resultados classificados derivados dos outros 9, que são repetidos 10 vezes até que todos os 10 subconjuntos sejam usados uma vez como subconjunto de teste (Xu, 2019). Os modelos foram avaliados primeiramente quanto à sua precisão, representada pela raiz do erro quadrático médio (RMSE); a precisão representa uma medida do desempenho geral. Uma medida de acurácia também foi utilizada (R^2), que mede o grau de distância dos valores observados e estimados pelos modelos. Por fim, utilizamos uma medida de tendência, representada pelo erro médio de estimativa (EME),

que informa se estamos super ou subestimando os valores estimados (Equações 3 a 5):

$$EME = \frac{\sum_{i-1}^{N}(Yobs_i - \bar{Y}est_i)}{N} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Yobs_i - \bar{Y}est_i)^2}{N}} \quad (4)$$

$$R^2 = \frac{\sum_{i-1}^{N}(Yest - \hat{Y}obs)^2}{\sum_{i=1}^{N}(Yest - \bar{Y}obs)^2} \quad (5)$$

nas quais, $Yobs_i$ são os dados observados, $Yest_i$ os dados estimados e ou previstos e N é o número de observações.

Utilizamos o pacote Scikit-learn versão 0.23.2 para Python em todos os modelos de ML e validação cruzada.

2.3 Resultados e Discussão

O mapa térmico (Figura 4), revelou 2 grupos distintos de correlação com estádios fenológicos (eixo vertical) e dois agrupamentos climáticos (eixo horizontal). As variáveis climáticas foram claramente divididas em ARM, P e EXC, caracterizadas por oferta hídrica para o cultivo, sem perdas. Por outro lado, ETR, DEF, T e ETc formam outro conjunto de agregação, com perdas relativas de água, seja por evapotranspiração ou pelo déficit de água no solo.

No agrupamento, de acordo com os estágios fenólicos da planta, os resultados foram demonstrados em duas divisões: vegetativo (estabelecimento e desenvolvimento) e reprodutivo (frutificação e maturação) (Figura 4).

A partir do mapa térmico (Figura 4), observou-se que o grupo categorizado com oferta hídrica a planta, com exceção da fase de floração, apresenta correlação positiva em todos os demais estágios fenológicos da soja. Os grupos climáticos compostos por perdas hídricas, apresentam correlação negativas nos diferentes estágios fenológicos, com alta correlação nas fases de frutificação e maturação.

Santos et al. (2015), em estudos sobre o regime hídrico e produtividade de genótipos de soja, observaram que o excesso hídrico possui uma correlação

negativa com a produtividade, principalmente quando ocorre na fase que antecipa a floração. A causa dessa resposta é atribuída pelo alongamento das hastes principalmente na tentativa de aumentar a eficiência da absorção da radiação solar e causando crescimento excessivo das plantas. Essa relação leva a maior sensibilidade ao acamamento, que reduz a indução à floração e diminuindo o número de vagens por planta.

Em estudos objetivando avaliar a resposta da soja em diferentes suplementações de irrigação, Montoya et al. (2017) observaram que maiores ofertas hídricas proporcionaram aumento na matéria seca total e na produtividade de grãos, possibilitando a estabilidade da produtividade da cultura durante a fase reprodutiva da cultura.

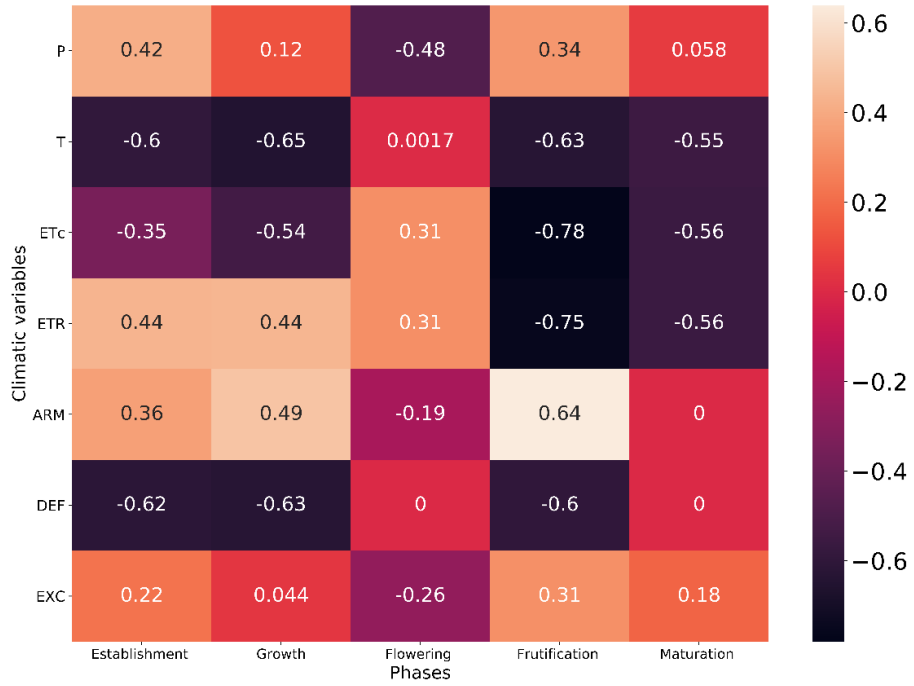

Figura 4 – Mapa térmico de correlação agrupado segundo clima e fases fenológicas da soja.

A capacidade preditiva dos algoritmos de ML avaliados apresentaram alto desempenho. O RF por meio de validação cruzada demonstrou maior acurácia e exatidão, com coeficiente de determinação R^2= 0,81 e RMSE de 176,93 kg ha^{-1} (Figura 5). Esses resultados mostram a robustez do modelo na previsão da produtividade da soja na região Sul do Maranhão. O menor desempenho foi observado pelo modelo RBF, com R^2 de 0,74 e RMSE de 213,58 kg ha^{-1}.

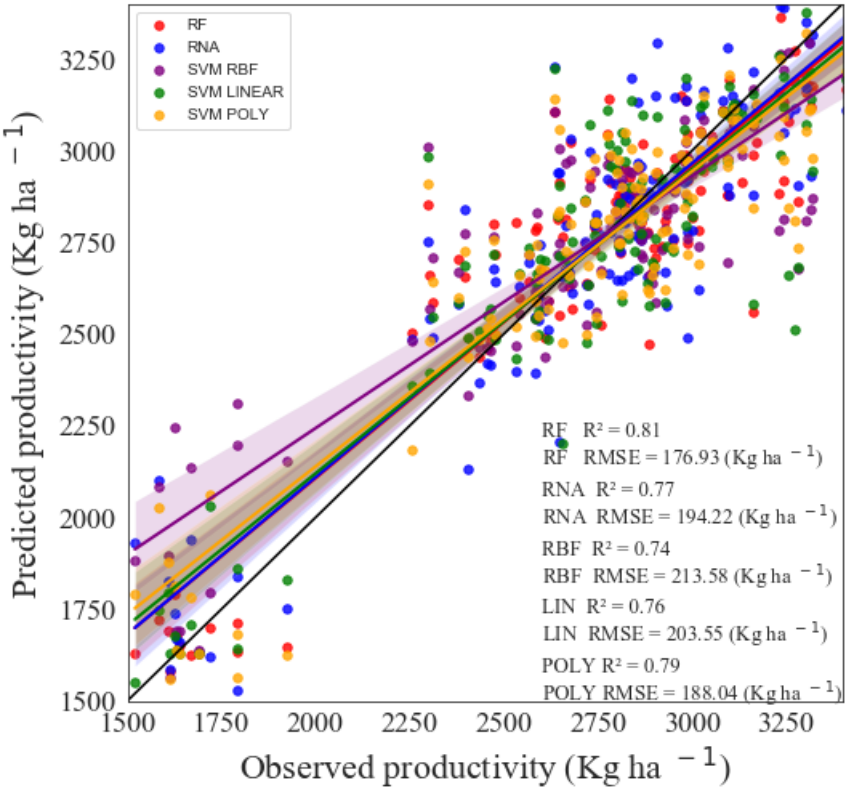

Figura 5 - Modelos de aprendizado de máquina, Floresta Aleatória - RF, Redes Neurais Artificiais - RNA, Máquinas de Vetores de Suporte - kernel -Função de base radial - SVM_RBF, Máquinas de Vetores de Suporte - kernel -Linear - SVM_LINEAR e Máquinas de Vetores de Suporte - kernel - Polinômio de terceira ordem - SVM_POLY, para previsão da produtividade da soja um mês antes da colheita.

Os valores de tendência estimados pela EME, medindo o desvio entre o valor observado e o valor médio estimado, apresentam seus maiores erros para o modelo SVM_RBF, com 15,06 kg ha^{-1} (Figura 5). Portanto, observa-se que, em média, o SVM_RBF superestima os dados observados em campo. Porém, a superestimação é maior em valores com escalas menores, aproximadamente até 2.500 kg ha^{-1}, tendendo o modelo em escalas de produtividade maiores a subestimar os resultados.

Ao contrário do SVM_RBF, o modelo RF apresentou o melhor desempenho de tendência, com EME igual a 1,99 kg ha^{-1}. Estes resultados indicam que os valores observados estão muito próximos dos valores médios estimados, com o modelo, em média, tendendo a subestimar dos resultados

(Figura 5), mas com superestimações em valores menores de escala de produtividade e subestimações em valores de escala maior. Esses resultados com subestimação e superestimação em altas e baixas produtividades, respectivamente, podem ser explicados pelas limitações dos modelos ML em estimar anos com pico e declínio na produção.

Gopal e Bhargavi (2019), avaliaram a precisão dos algoritmos RNA, Support Vector Regression, K-Nearest Neighbour and RF para previsão de produtividade agrícola. Os resultados mostraram que o algoritmo de RF, assim como observados neste trabalho, atinge a maior precisão e acurácia na previsão produtiva das culturas.

A precisão e a acurácia são métricas importantes para determinar se o modelo de aprendizado de máquina está realmente entendendo as informações disponibilizadas. Esses algoritmos podem fornecer melhores recursos de generalização nos domínios espacial e temporal, o que é crítico para as previsões operacionais de produtividade de cultivos para grandes áreas de estudo. (Shao, 2015).

Os algoritmos de suporte de vetores de máquina também obtiveram resultados excelentes, as três configurações kernel apresentaram coeficiente de determinação R^2 próximos a 1, sendo SVM_RBF = 0,75, SVM_LINEAR = 0,89 e SVM_POLY = 0,89, quanto ao RMSE, SVM_RBF = 208,41 kg ha^{-1}, SVM_LINEAR = 137,27 kg ha^{-1} e SVM_POLY = 134,33 kg ha^{-1}.

Os resultados de RMSE de todos os modelos avaliados foram melhores que os encontrados por Sun et al., 2019, que utilizaram modelo Rede Neural Convolucional associado a Long Short-Term Memory (CNN-LSTM) para a previsão de produtividade de soja no final da temporada nos Estados Unidos e obtiveram RMSE de 329,53 kg ha^{-1}

A Figura 6 mostra a distribuição dos valores observados e dos valores previstos. A média dos modelos RF, RNA, SVM_Linear e SVM_Poly são próximas ao observado. O modelo RF é o que mais se assemelha aos dados reais. O primeiro e segundo quartis de ambos, observado e RF são semelhantes, esse resultado é explicado pelo RMSE de 203,55 kg ha^{-1} do modelo linear. Por apresentarem ótimo desempenho existe alta semelhança entre os valores de produtividade do cultivo com os valores previstos.

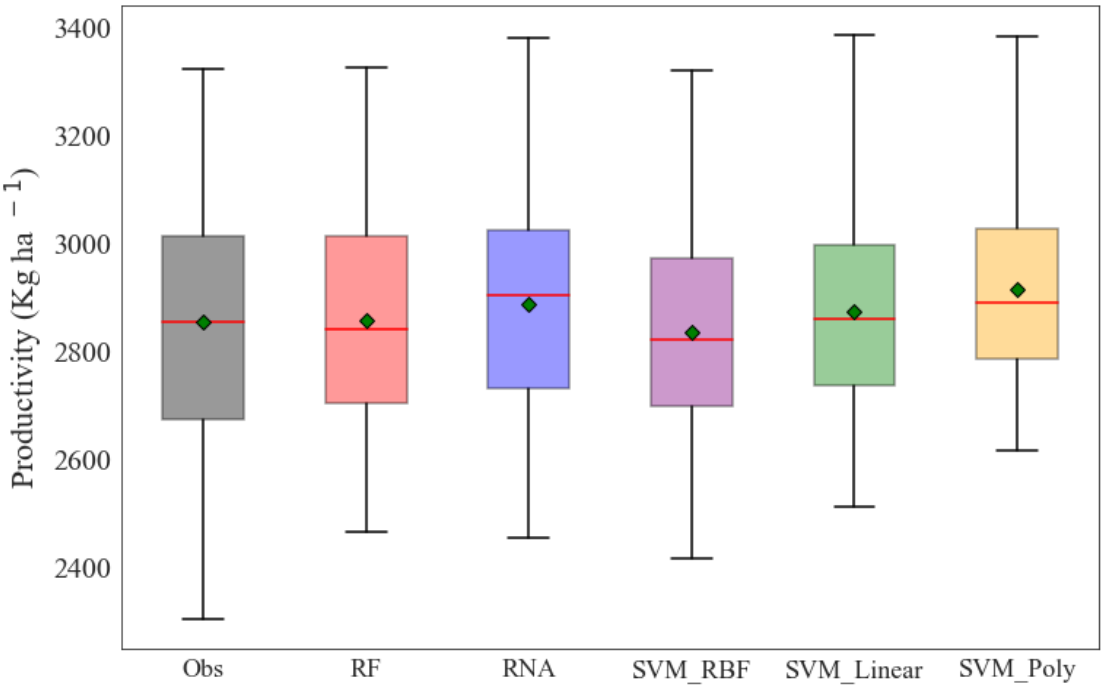

Figura 6 – Boxplot da produtividade observado e previsto por cada modelo de aprendizado de máquina. Legenda: losango = média, linha vermelha = mediana, caixa = 50% dos dados e traço = 99% dos dados.

A região sul do estado do Maranhão apresentou produtividade média de 2.730 kg ha^{-1}. Conforme a Figura 7.A, a produtividade observada é similar para a maioria dos municípios. Os modelos conseguiram capturar as pequenas variações entre os municípios. Sun et al., 2019, também observaram que há alta consistência entre a produtividade prevista pelo modelo CNN-LSTM e o resultado do USDA.

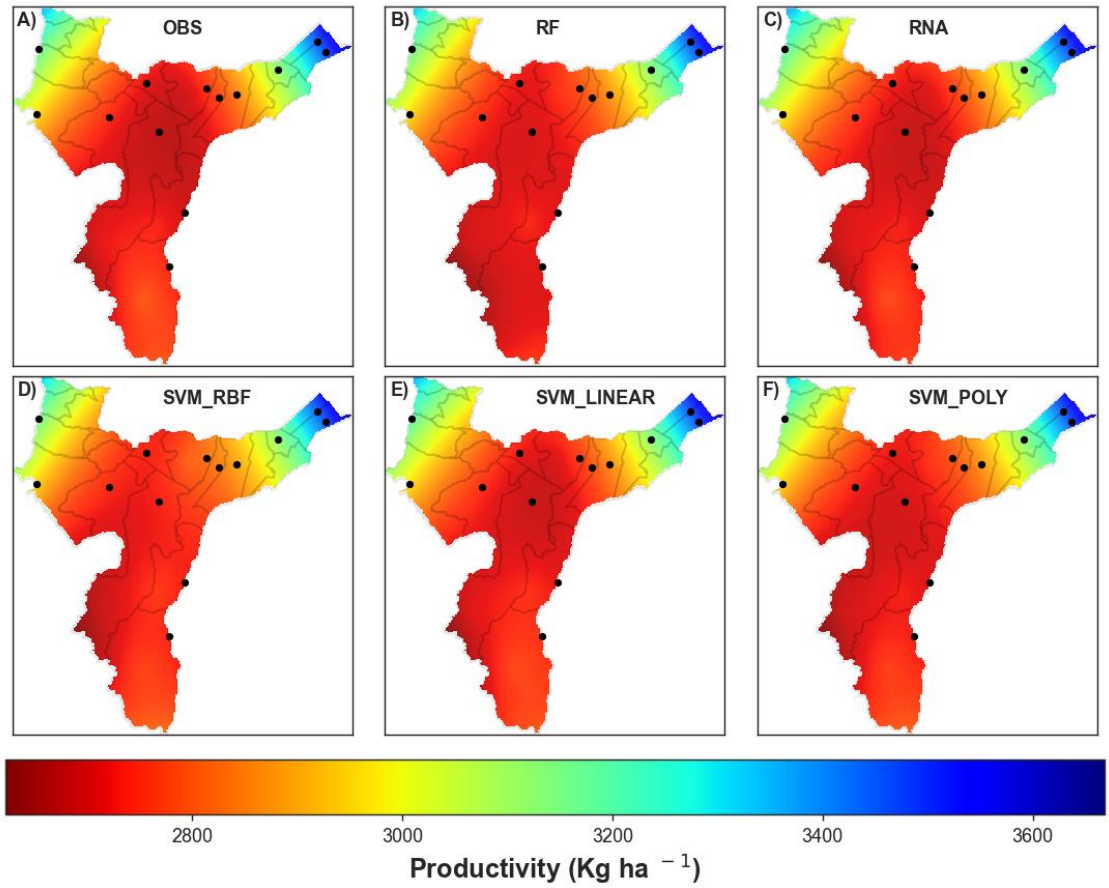

Figura 7 - Mapa da distribuição da produtividade de soja: A: produtividade observada, B: produtividade prevista por Random Forest, C: produtividade prevista por RNA, D: produtividade prevista por SVM_RBF, E: produtividade prevista por SVM_LINEAR, F: produtividade prevista por SVM_POLY.

Os mapas de diferença entre os valores previstos e observados promovem melhor entendimento entre as particularidades dos modelos quanto a previsão. A região Sul do local de estudo, apresentou subestimação em todos os modelos (Figura 8), esses resultados podem se dar pela menor concentração de pontos amostrais nesta região. Na parte norte, nordeste e noroeste da região sul do Maranhão é possível observar que o RF, RNA e SVM_LINEAR (Figuras 8.A, B, D), respectivamente superestimam a produtividade nestes locais. Já SVM_RBF (Figura 8.C) e SVM_POLY (Figura 8.E) com exceção da região norte, subestima a produtividade em comparação ao valor observado.

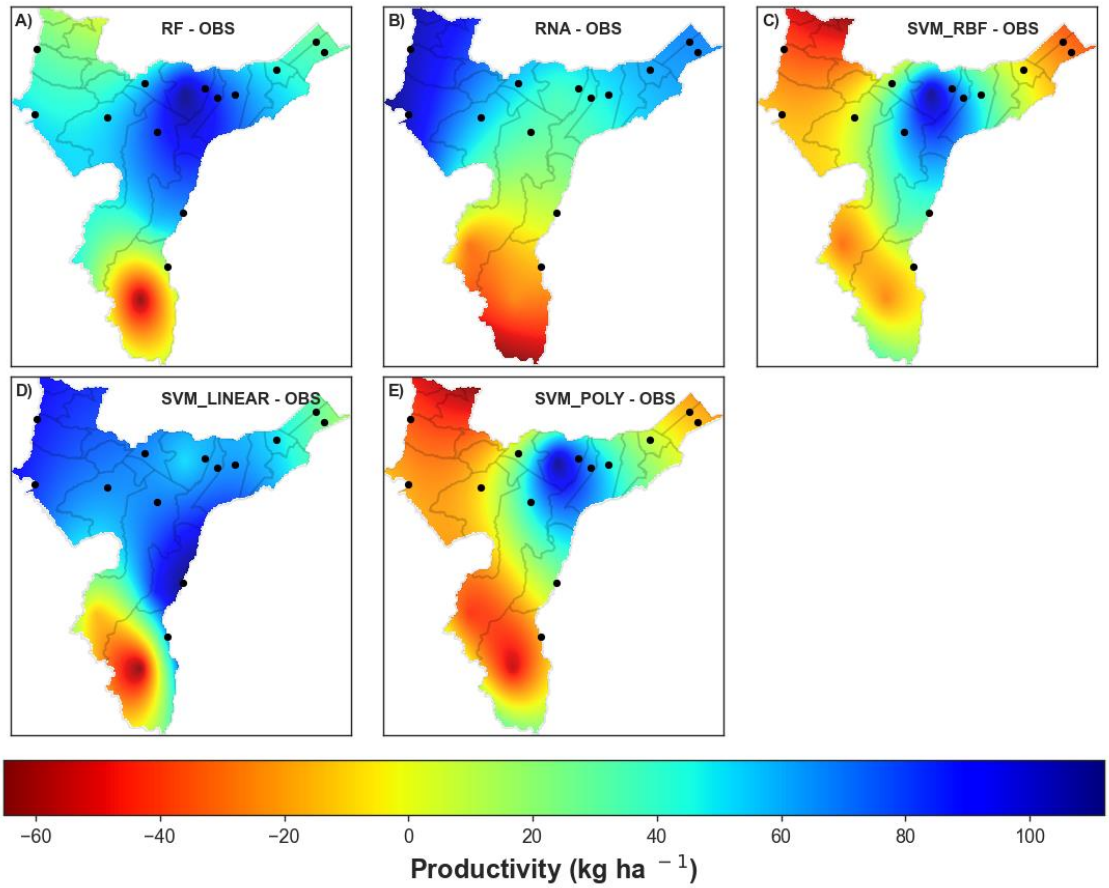

Figura 8 - Mapa da diferença entre a produtividade prevista e observada para cada modelo.

Como observado, o modelo SVM_RBF e SVM_POLY apresentaram variação de produção diferente dos demais modelos, com predominância de subestimações na parte noroeste e nordeste da região.

Maitiniyazi (2020) avaliou a fusão de dados multimodais usando sensores RGB, multiespectrais e térmicos para estimar a produtividade de grãos de soja, utilizando modelos de machine learning como “Deep Neural Network” (DNN), “Random Forest Regressor” (RF), “Support Vector Machines” (SVM) entre outros, obteve R^2= 0,72 e RMSE = 479 kg ha^{-1} para redes neurais profundas, R^2= 0,66 e RMSE= 526 kg ha^{-1} para o Random Forest e R^2= 0,67 e RMSE= 521 kg ha^{-1}, para o SVM.

2.4 Conclusão

A capacidade preditiva dos modelos de ML foi avaliada e comparada, produzindo alta precisão e acurácia na previsão de produtividade de soja no cerrado Brasileiro. O algoritmo RF obteve maior desempenho com R^2 = 0,81 e RMSE de 176,93 kg ha^{-1}. Por outro lado, o algoritmo SVM_RBF apresentou o menor desempenho com R^2 = 0,74 e RMSE de 213,58 kg ha^{-1}. Apesar disso, ambos os modelos são suficientemente precisos para prever a produtividade na região.

A parte Sul da região de estudo, apresentou subestimação em todos os modelos preditivos, caracterizada principalmente pela menor densidade de pontos amostrais, causando incertezas nos valores interpolados de áreas não avaliada. Contudo, em regiões com maiores densidades de observações, como a parte norte, o algoritmo de interpolação tende a superestimar os resultados.

REFERÊNCIAS

Alambert, M. R., Umburanas, R. C., Schwerz, F., Reichardt, K., & Dourado-Neto, D. (2019). Stochastic estimation of potential and depleted productivity of soybean grain and oil. **International Journal of Plant Production**, 13, 103-116.

Alghamdi, M., Angelov, P., Gimenez, R., Rufino, M., & Soares, E. (2019, October). Self-organising and self-learning model for soybean yield prediction. In **2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)** (pp. 441-446). IEEE.

Alves GR, Teixeira IR, Melo FR, Souza RTG and Silva AG, (2018). Estimating soybean yields with artificial neural networks. **Acta Scientiarum. Agronomy**, 40, 35250.

Breiman, L. (2000). Randomizing outputs to increase prediction accuracy. Machine Learning, 40, 229-242.

Cai Y, Guan K, Lobell D, Potgieter AB, Wang S, Peng J et al., (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. **Agricultural and forest meteorology** 274:144–159.

Companhia Nacional De Abastecimento (CONAB). Grain planting and harvesting calendar in Brazil. (In Portuguese.) Brasília [Online]. 2019. p. 73. Disponível em: https://www.conab.gov.br/institucional/publicacoes/outraspublicacoes/item/7694-calendario-agricolaplantio-e-colheita.pdf [13 November 2019]

Companhia Nacional De Abastecimento (CONAB). Perspectives for agriculture, v.6, harvest 2018/2019. Brasília [online]. 2018. Disponível em: https://www.conab.gov.br/perspectivasparaaagropecuaria/item/download/22780ee707c6e6d44f06fe7b6a86ce6141652 [13 November 2019]

Evangelista, B. A., da Silva, F., Simon, J., & Campos, L. J. (2017). Climatic risk zoning for determining sowing dates for soybean crops in the MATOPIBA region. **Embrapa Pesca e Aquicultura-Boletim Pesqui e Desenvolv**.

Farias, J. R. B., Nepomuceno, A. L., & Neumaier, N. (2007). Ecophysiology of soy. **Embrapa Soy-Circular Technique (INFOTECA-E)**.

Han S, Tang Q, Xu D and Yang Z, (2019). Impacts of urbanization and agricultural development on observed changes in surface air temperature over mainland China from 1961 to 2006. **Theoretical and applied climatology** 135:1595–1607.

Instituto Brasileiro de Geografia e Estatistica (IBGE). Municipal agricultural production (PAM). [Online]. In: Sidra: system IBGE automatic recovery, Rio de Janeiro. 2019. Disponível em: https://sidra.ibge.gov.br/tabela/1612

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., & Fritschi, F. B. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. **Remote sensing of environment**, 237, 111599.

Maya Gopal PS and Bhargavi R, (2019). Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms. **Applied Artificial Intelligence** 33:621–642.

Mbangiwa, N. C., Savage, M. J., & Mabhaudhi, T. (2019). Modelling and measurement of water productivity and total evaporation in a dryland soybean crop. **Agricultural and forest meteorology** 266, 65-72.

Michelon GK, De Menezes PL, Júnior AC, Bazzi CL and Barbosa MM, (2017). Máquina De Vetores De Suporte Para Estimar A Produtividade Da Soja. **Revista Engenharia na Agricultura-REVENG**, 25(3), 240-248.

Montoya, F., García, C., Pintos, F., & Otero, A. (2017). Effects of irrigation regime on the growth and yield of irrigated soybean in temperate humid climatic conditions. **Agricultural Water Management**, 193, 30-45.

Moraes JR, Souza Rolim G, de Martorano LG, Oliveira Aparecido LE, de Oliveira MSP and Farias Neto JT, (2020). Agrometeorological models to forecast açaí (Euterpe oleracea Mart.) yield in the Eastern Amazon. **Journal of the Science of Food and Agriculture** 100:1558–1569.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al., (2011). Scikit-learn machine learning in Python. the Journal of machine Learning research 12:2825–2830.

Prela-Pantano A, Duarte AP, da Silva DF, Rolim GDS and Caser DV, (2011). Productivity of maize, and occurrence of enso precipitation in the region of the middle Paranapanema, SP, Brazil. **Rev Bras Milho e Sorgo** 10:146–157.

Sakamoto T, (2020). Incorporating environmental variables into a MODISbased crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm. **ISPRS Journal of Photogrammetry and Remote Sensing** 160:208–228.

Santos, E. D., Nepomuceno, A. L., Farias, J. R. B., MANDARINO, J., MERTZ-HENNING, L. M., de OLIVEIRA, M. C. N., & NEUMAIER, N. (2017). Regime hídrico e rendimento de genótipos de soja em condição de campo.

Schwalbert RA, Amado T, Corassa G, Pott LP, Prasad PVV and Ciampitti IA, (2020). Satellite-based soybean yield forecast: integrating machine learning and weather data for improving crop yield prediction in southern Brazil. **Agricultural and Forest Meteorolog** 284:107886.

Shaharum, N. S. N., Shafri, H. Z. M., Ghani, W. A. W. A. K., Samsatli, S., Yusuf, B., Al-Habshi, M. M. A., & Prince, H. M. (2018). Image classification for mapping oil palm distribution via support vector machine using scikit-learn module. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, 42, 133-137.

Shao, Y., Campbell, J. B., Taff, G. N., & Zheng, B. (2015). An analysis of cropland mask choice and ancillary data for annual corn yield forecasting using MODIS data. **International Journal of Applied Earth Observation and Geoinformation**, 38, 78-87.

Sharma, P., & Dupare, B. U. (2018). Varietal and yield of soybean in Madhya Pradesh. **Indian Farming**, 68(7).

Stackhouse PW, Perez R, Sengupta M, Knapp K, Mikovitz JC, Schlemmer J et al., An assessment of new satellite data products for the development of a long-term global solar resource at 10–100 km. **ASES Natl Sol Conf** 10:1–6 (2016).

Sun, J., Di, L., Sun, Z., Shen, Y., & Lai, Z. (2019). County-level soybean yield prediction using deep CNN-LSTM model. **Sensors**, 19(20), 4363.

Tehrany, M. S., Pradhan, B., Mansor, S., & Ahmad, N. (2015). Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. **Catena**, 125, 91-101.

Thornthwaite CW and Mather J, The Water Balance, (1955). Vol. 8. Drexel Institute of Technology: Laboratory of Climatology. Publications in Climatology, Centerton, NJ, p. 104.

Capítulo 3 – Modelos de machine learning para estimação do volume de eucalipto no Cerrado Brasileiro a partir de dados climáticos sazonais

Resumo – O eucalipto é uma espécie dominante nas florestas plantadas e é utilizado principalmente para produção de madeira, celulose e energia. Os Modelos de inteligência artificial, como Random Forest, são aplicados na estimativa da produtividade do eucalipto com base em variáveis ambientais e de manejo. A variabilidade climática durante o ciclo de uma floresta, além de eventos climáticos extremos, como secas, influencia significativamente na produtividade florestal. A compreensão da relação solo-planta-atmosfera é crucial para o cultivo de eucalipto. Esse estudo visa estimar o volume de madeira de eucalipto no cerrado brasileiro utilizando técnicas de machine learning e apenas dados climáticos como inputs dos modelos, abrangendo diferentes idades de crescimento em dois períodos do ano entre janeiro e junho e entre julho e dezembro. Os resultados mostram um ótimo desempenho dos modelos na estimativa do volume de madeira com base nos padrões meteorológicos da série histórica analisada. O modelo Random Forest caracterizou-se como o melhor modelo, apresentando as melhores métricas durante o treinamento e teste com, R^2= 0.93 e RMSE = 18.36 m^3ha^{-1} para o modelo janeiro-junho e R^2= 0.92 e RMSE = 19.52 m^3ha^{-1} para o modelo de julho-dezembro.

Palavras chaves: floresta digital, manejo florestal, inteligência artificial

3.1 Introdução

As florestas plantadas são cultivadas com a finalidade de produzir madeira, fibras e energia, (Seng Hua et al., 2022). Entre todas as espécies arbóreas plantadas, as espécies de Pinus (nativas e não nativas) são dominantes na maioria das regiões do mundo, enquanto as espécies não nativas de Eucalyptus são as mais comuns nos trópicos e subtrópicos (Messier et al., 2021).

O relatório Global Forest Resources Assessment (FRA) publicado pela Organização das Nações Unidas para a Alimentação e a Agricultura (FAO) em 2020, menciona que a área total de florestas é de 4,06 bilhões de hectares (ha), o que corresponde a 31% da área total do mundo (FAO, 2020). A FRA identificou duas grandes categorias de florestas, a saber, florestas em regeneração natural e florestas plantadas. De acordo com o relatório, as florestas naturais em regeneração cobrem cerca de 3,75 bilhões de hectares ou 93% da área florestal total. Enquanto isso, a área total de florestas plantadas globalmente é estimada em 294 milhões de hectares ou 7% da área florestal mundial (Seng Hua et al., 2022).

O eucalipto é tipicamente manejado em rotação curta para aumentar a economia com a produção de madeira, celulose, carvão vegetal e lenha (Zhou et al., 2018). É um cultivo muito adaptável, tolerando baixa fertilidade, solos ácidos e ricos em alumínio, estresse hídrico muitas vezes periódico, climas e tipos de solo diversos, e até mesmo danos causados por incêndios e insetos (Ramantswana et al., 2020).

O cultivo do eucalipto é uma das atividades florestais mais importantes referente à produção de madeira, papel e celulose e a modelagem para estimação da produtividade de forma precisa é essencial para otimizar tomadas de decisões quanto ao manejo florestal e a alocação de recursos (De Freitas et al., 2020).

Vários modelos têm sido desenvolvidos e aplicados para estimar a produtividade de eucalipto, porém a complexidade do ambiente florestal recomenda que a interação entre plantas e fatores ambientais seja analisada de forma conjunta (Billings, 1952).

Dentre as características complexas dentro de uma floresta, o crescimento é indiscutivelmente o mais estudado (de Oliveira Neto, 2022). Dessa forma, o entendimento dos fatores que impulsionam o da floresta através das práticas de silvicultura é de importância fundamental quando o objetivo é aumentar a produtividade florestal (Otto et al. 2013). No entanto, mesmo alcançando resultados significativos, ainda se conhece pouco sobre os fatores que regulam a produtividade florestal (Stape et al. 2004).

Para entender esses fatores, diversos pesquisadores têm se dedicado à elaboração de modelos que possam estimar a produtividade do eucalipto com base em variáveis climáticas, de solo e de manejo. Vieira et al. (2018) considerou o potencial das técnicas de inteligência artificial para serem utilizadas em estimativas florestais, propondo a aplicação de Redes Neurais Artificiais e de sistema adaptativo de inferência neurofuzzy para estimar o diâmetro a altura do peito (DBH) e a altura total da planta (H).

De Alcantara (2018) buscou desenvolver e testar modelos de redes neurais artificiais com o objetivo de estimar a produção e o desenvolvimento de estandes de eucalipto. Silva (2021) investigou a melhor alternativa para estimar a altura das árvores e a produção volumétrica de eucaliptos em sistemas agrosilvopastoris utilizando modelos de Redes Neurais Artificiais e Regressão.

Santana (2023) avaliou o desempenho de diferentes técnicas de machine learning como Regressão Linear Multipla, Random Forest e Suport Vector Machine para prever o volume de madeira de eucalipto. As variáveis de entrada foram DBH e a H. Todos os estudos alcançaram com sucesso seus objetivos durante a modelagem, demonstrando a eficácia do uso de variáveis biométricas.

A determinação das variáveis a serem utilizadas durante a modelagem é parte importante da investigação, pois a produção de eucalipto é influenciada por vários fatores, como características químicas, físicas e biológicas do solo, disponibilidade de água no solo, qualidade das plântulas, material genético, práticas de manejo, pragas e doenças e condições climáticas (Elli et al., 2020). Este último leva a uma variabilidade considerável na produção de eucalipto.

As variáveis climáticas têm forte correlação com o crescimento e desenvolvimento do eucalipto, apresentando potencial real para serem seus estimadores (Elli et al., 2020). A temperatura do ar, a disponibilidade hídrica e evapotranspiração influenciam diretamente na produtividade do eucalipto, uma

vez que afetam diretamente os processos fisiológicos, como fotossíntese, transpiração e crescimento (Campoe et al., 2016).

A ocorrência de eventos climáticos extremos, como secas e ondas de calor, pode afetar a produtividade do eucalipto. Eventos de seca, especialmente em áreas com baixa capacidade de armazenamento de água no solo reduz em até 50% a produtividade do eucalipto conforme avaliado por Christina et al. (2017).

Todo o processo de crescimento e produtividade são influenciados por uma infinidade de fatores que geram um grande volume de informações e a estratégia tradicional de avaliar e estimar o crescimento e a produtividade em floresta com base em dados climáticos, baseia-se na adoção de modelos de regressão (Silva et al., 2021).

Portanto, o objetivo desse estudo é estimar o volume de madeira de eucalipto por hectare no cerrado brasileiro a partir de dados climáticos, utilizando técnicas de machine learning para dois períodos do ano, entre janeiro e junho e entre julho e dezembro em diferentes idades de crescimento visando auxiliar o processo de inventário florestal e disponibilizar estimativas de produtividade de forma mais rápida e econômica para os tomadores de decisão em dois períodos com características climáticas distintas

3.2 Material e métodos

3. 2.1 Caracterização da área de estudo

A área de estudo corresponde à aproximadamente 400 mil ha de floresta de eucalipto plantado no Bioma Cerrado localizado no estado do Mato Grosso Sul, Brasil (Figura 1), com altitude variando entre 200 e 600 m.

O cerrado, normalmente se encontra associado a relevos que variam de suaves a levemente ondulados, com solos profundos, bem drenados e de baixa fertilidade. Entretanto, áreas campestres podem ocorrer sob solos mal drenados, onde o lençol freático aflora regularmente, o que impede o estabelecimento de espécies arbóreas (Silva, 2007).

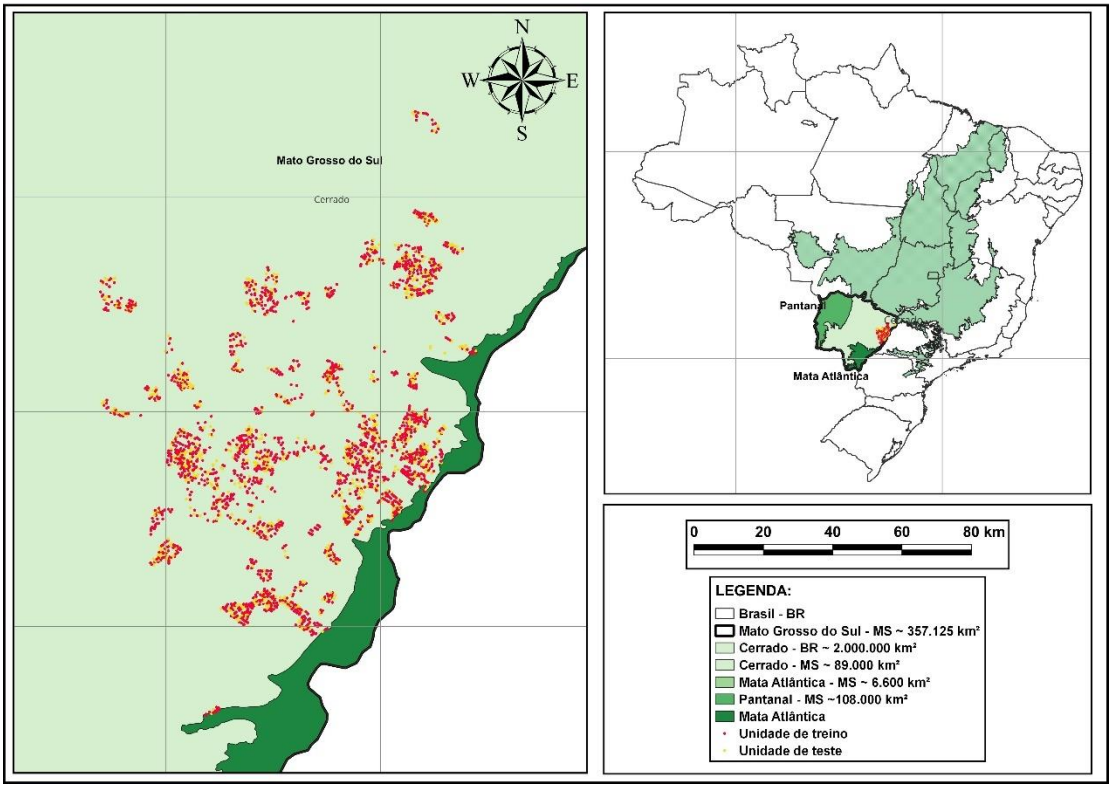

Figura. 1 – Mapa de localização das áreas de estudo e descrição do bioma Cerrado. As unidades de produção foram divididas aleatoriamente em um conjunto de treino (em vermelho – 70% do total de dados) e um conjunto de teste (em amarelo – 30% do total de dados)

3.2.2. Dados meteorológicos

A variabilidade climática, incluindo os dados das temperaturas do ar, dados da precipitação e a radiação solar global, desempenha um papel fundamental na determinação da produtividade das plantas, pois afeta diretamente o crescimento e o desenvolvimento das culturas e a disponibilidade de recursos hídricos (Thornton et al., 2014).

Para modelar o potencial de variação climática tanto espacial quanto temporalmente, a área de estudo selecionada dispõe de 17 estações meteorológicas próprias da empresa que em conjunto com as 21 estações do Instituto Nacional de Meteorologia (INMET), somam 38 estações, fornecendo dados climáticos em escala diária.

Nesse estudo, utilizamos dados diários de precipitação (P, mm dia^{-1}), temperatura máxima e mínima do ar (T, °C), umidade relativa (UR %) e radiação

solar no topo da atmosfera (Qo, MJ m^{-2} dia^{1}) entre janeiro de 2010 e dezembro de 2022, totalizando uma série histórica de 12 anos. Esses dados foram tabulados de forma mensal e em seguida foram espacializados usando interpolação por ponderação pelo inverso da distância (IDW). A evapotranspiração potencial foi calculada pelo método de Thornthwaite e Mather (1948).

Devido à sua localização geográfica e à influência de diferentes sistemas atmosféricos o clima é caracterizado como tropical chuvoso com inverno seco (Aw) de acordo com a classificação de Köppen, com temperatura média anual de 23,5 °C e precipitação média anual de cerca de 1400 mm, dividida em estação chuvosa e estação seca distintas (Alvares et al., 2013). O regime de chuvas no Mato Grosso do Sul é caracterizado pela estação chuvosa, que ocorre durante outubro a abril (figura 2). A estação seca se estende de maio a setembro, com redução das precipitações e aumento da frequência de dias sem chuva.

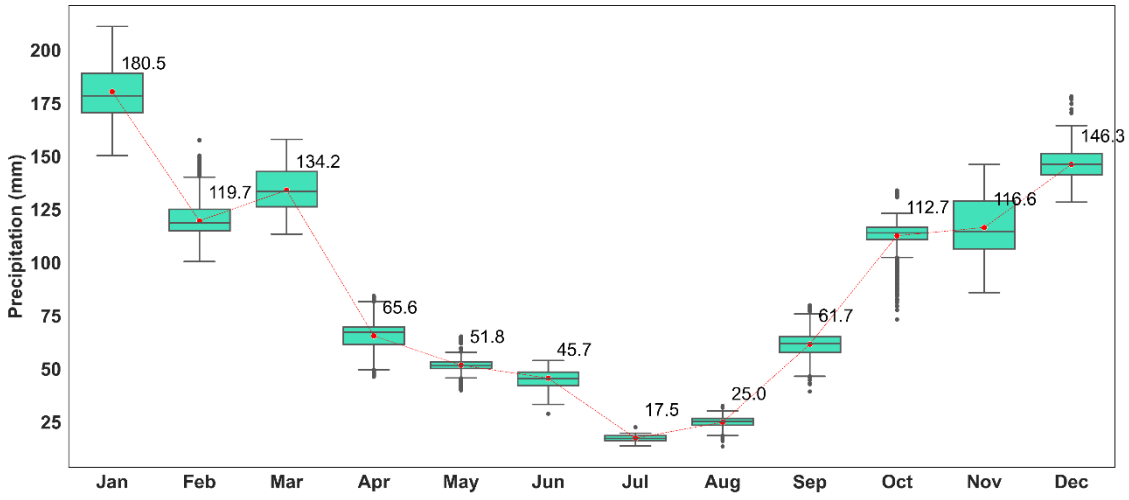

Figura. 2 – Distribuição dos valores de precipitação mensal para série histórica de 12 anos. O verão que ocorre entre dezembro e fevereiro é considerado chuvoso enquanto o inverno ocorrendo entre junho e agosto é caracterizado pelos baixos valores de precipitação na região do Cerrado – Mato Grosso do Sul, Brasil.

A amplitude da temperatura média da série histórica (figura 3) é entre 26 °C e 27 °C durante o período chuvoso e com valores variando entre 20 °C e 23 °C no período seco.

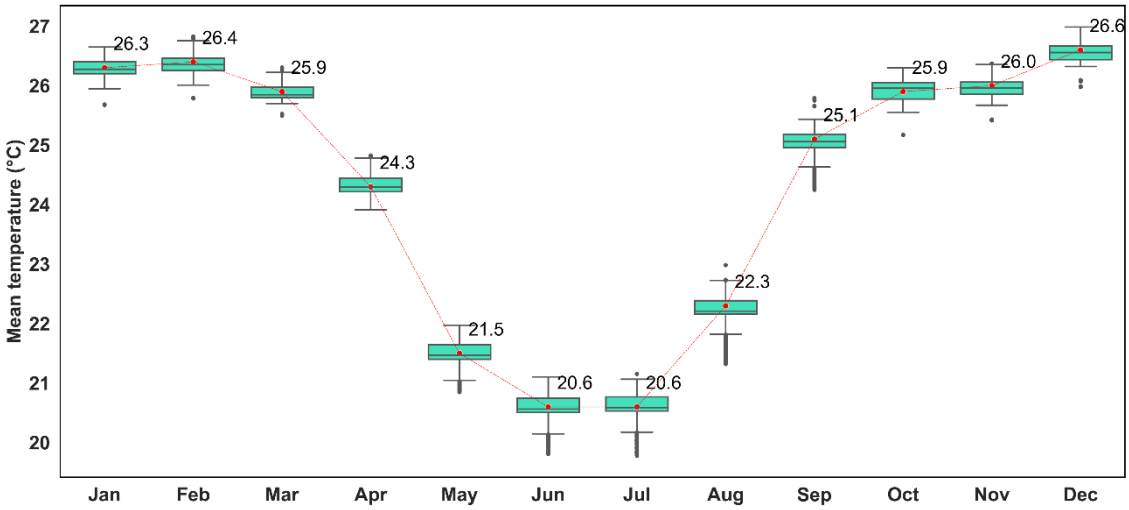

Figura. 3 – Temperatura média do ar mensal da área de estudo para o período de 12 anos com temperaturas na região do Cerrado – Mato Grosso do Sul, Brasil.

Foi calculado o balanço hídrico mensal pelo método de Thornthwaite e Mather (1955). A capacidade de água disponível (AWC) foi obtida das fontes de dados da Embrapa com valores entre 80 e 140 mm de acordo com cada coordenada central das unidades de produção (Embrapa, 2022). Em seguida foram selecionados os seguintes componentes do balanço hídrico para compor o conjunto de dados: Evapotranspiração potencial (mm) (PET) (figura 4), Evapotranspiração real (mm) (AET), Armazenamento de água do solo (mm) (STO), Déficit (mm) (DEF) e Excedente (mm) hídricos (EXC).

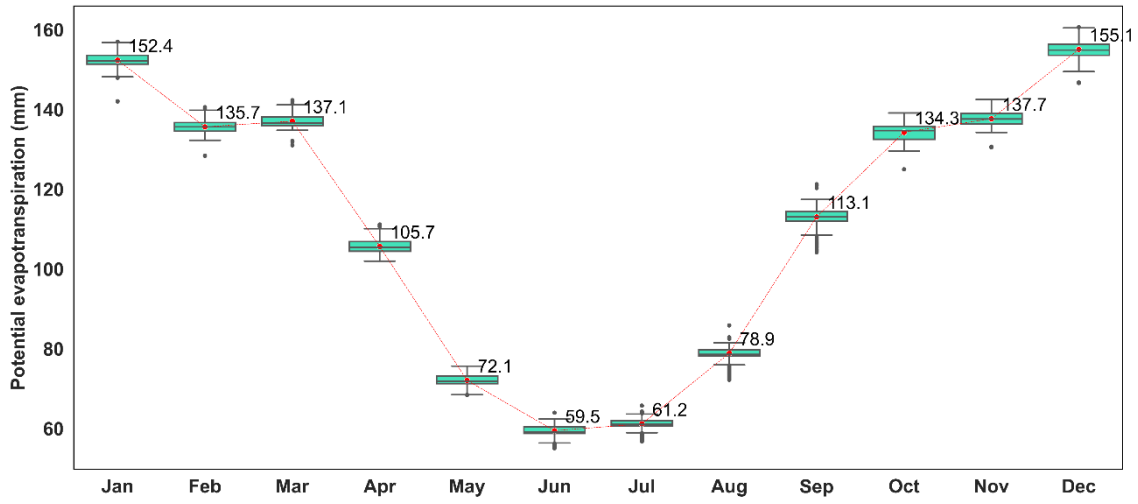

Figura. 4 – Evapotranspiração potencial mensal na região do Cerrado – Mato Grosso do Sul, Brasil.

A distribuição dos dados evapotranspiração potencial (PET) acompanha a distribuição da precipitação, com valores médios superiores a 150 mm nos meses mais chuvosos. Os valores mais elevados da PET ocorrem consequentemente nos meses em que a oferta de água é maior, além disso fatores ambientais podem impactar no potencial de variação da evapotranspiração (Mokhtar et al., 2020).

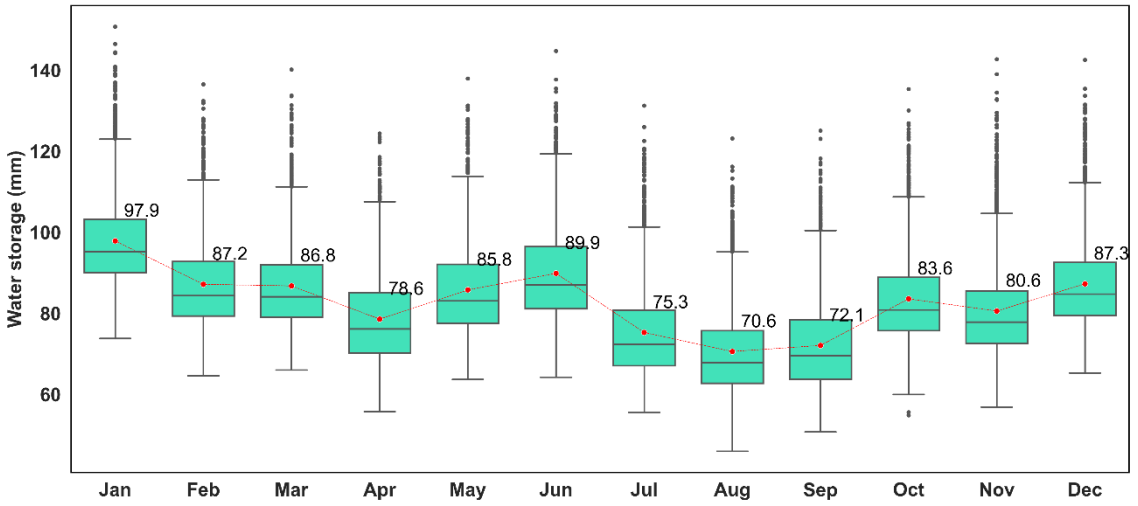

Figura. 5 – Distribuição dos valores de armazenamento mensal de água no solo para série histórica de 12 anos na região do Cerrado – Mato Grosso do Sul, Brasil.

A distribuição do armazenamento de água no solo (Figura 5) janeiro a dezembro mostra a diminuição no armazenamento médio de água no solo de janeiro (97,9 mm) a setembro (70,6 mm), seguida de um aumento entre outubro e dezembro, acompanhando a distribuição da precipitação.

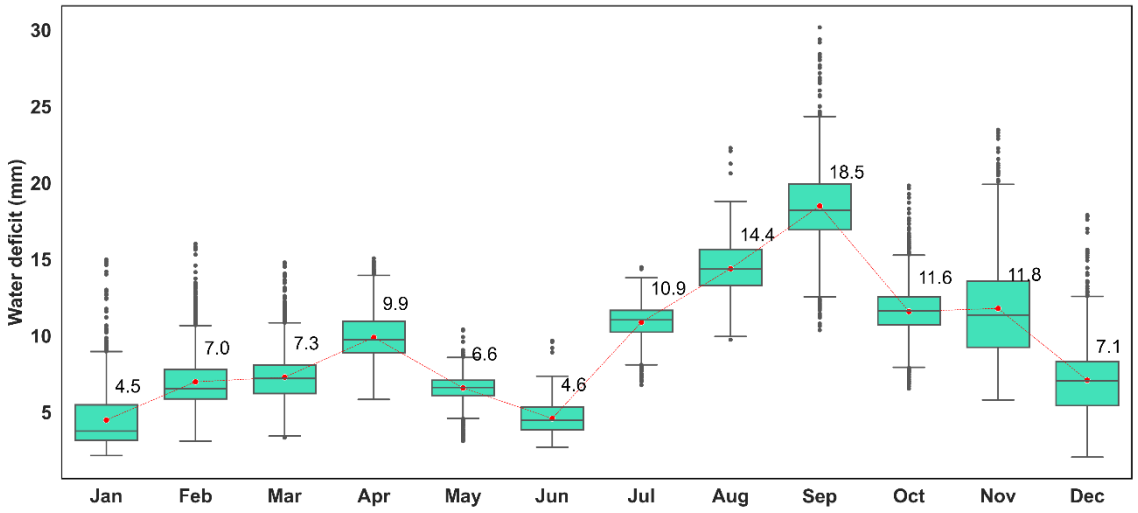

Figura. 6 – Distribuição dos valores mensais de deficiência de água no solo para série histórica de 12 anos na região do Cerrado – Mato Grosso do Sul, Brasil.

Ocorre o aumento do excedente de água no solo (EXC) entre outubro e março, período de maiores valores de precipitação (figura 7). O menor valor do excedente de água no solo ocorre em agosto com sua menor média (1,6 mm) influenciado pelos baixos valores de precipitação nesse período.

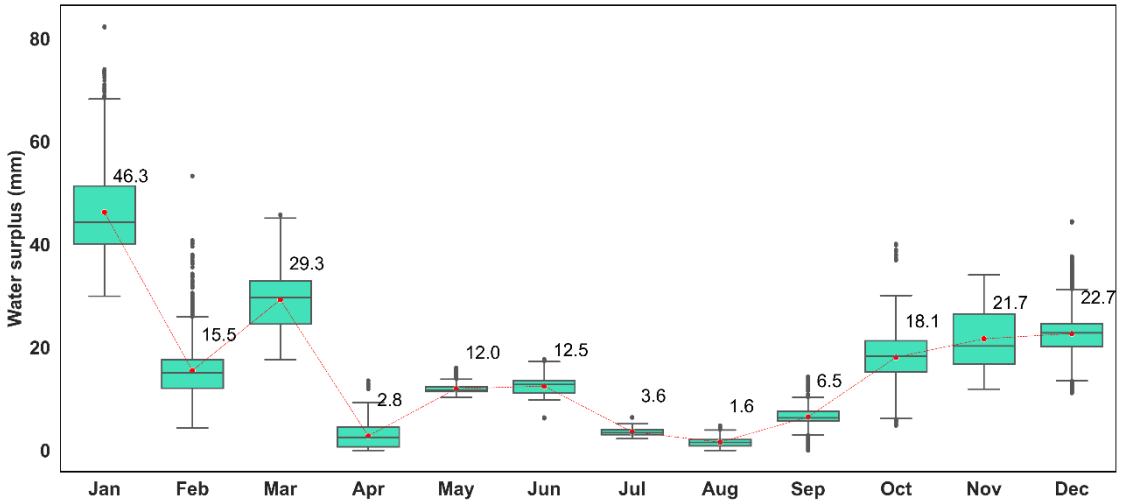

Figura. 7 – Distribuição dos valores de excedente de água no solo para série histórica de 12 anos na região do Cerrado – Mato Grosso do Sul, Brasil.

3.2.3 Inventario florestal.

O banco de dados amostral do inventario florestal utilizado nesse estudo apresenta parcelas de 400 m^2, com plantas em idades de 2 a 7 anos, com medições entre 2010 e 2022.

Uma pré análise para garantir a consistência dos dados foi realizada. A variável alvo desse estudo é a produtividade também conhecida como volume total com casca ($m^3\ ha^{-1}$) (figura 8).

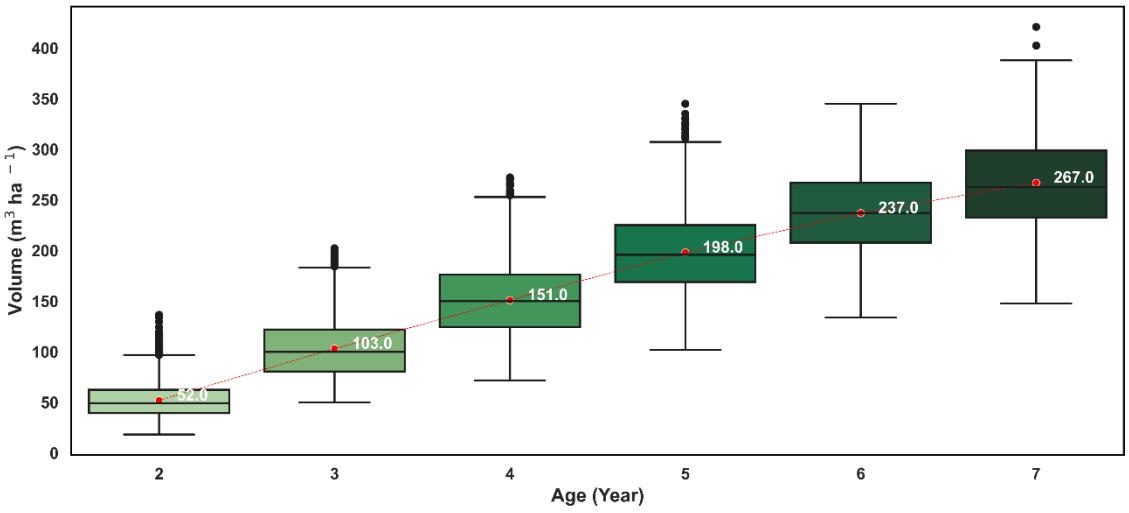

Figura. 8 – Curva de avaliação do volume total com casca ($m^3\ ha^{-1}$) ao longo do crescimento do eucalipto com medições a partir dos 2 anos de idade até aos 7 anos.

O volume total com casca é avaliado a partir do segundo ano, apresentando tendência de crescimento até aos 7 anos com incrementos anuais. Cada boxplot é composto por um número médio de amostras de 3.672 por idade. Com exceção da idade de 7 anos que apresenta número reduzido de amostras devido a demanda da indústria, a colheita ocorre na idade de 5 anos.

A análise da consistência dos dados florestais foi estendida para as variáveis diâmetro altura do peito (figura 9) e altura dominante (figura 10) para entendimento geral da curva de crescimento do eucalipto.

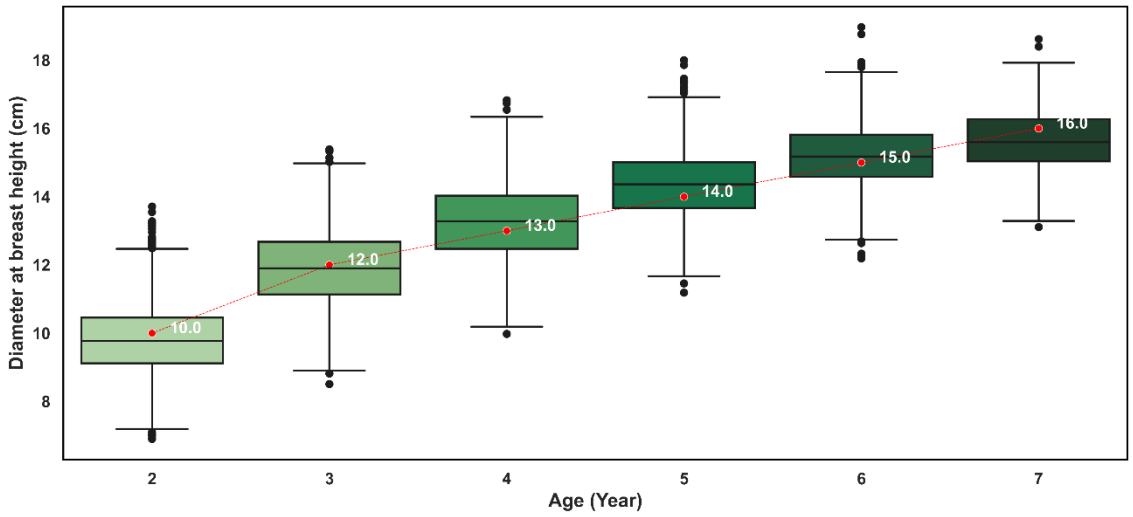

Figura. 9 – Curva de avaliação do diâmetro a altura do peito (cm) ao longo do crescimento do eucalipto com medições a partir dos 2 anos de idade até aos 7 anos.

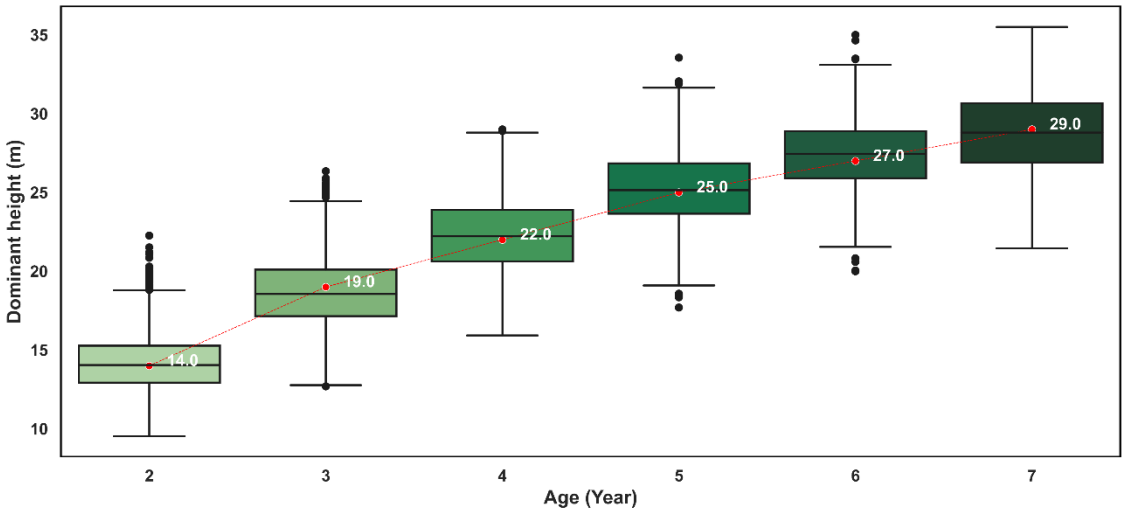

Figura. 10 – Curva de avaliação da altura dominante (m) ao longo do crescimento do eucalipto com medições a partir dos 2 anos de idade até aos 7 anos.

De modo geral as variáveis climáticas apresentam forte correlação com o volume (Figura 11). A precipitação influencia significativamente a produtividade da madeira devido ao aumento das chuvas ao longo dos períodos climáticos (Stape et al., 2010). A relação inversa pode ser interpretada para a deficiência Hídrica (DEF) sendo que o acúmulo de biomassa diminuiu significativamente no pico da estação seca (Rowland et al., 2014) período de maior déficit hídrico

ocorrendo entre julho e setembro meses em que o grau de correlação negativa entre DEF e VTCC (r = -0,49) é mais forte.

A correlação entre STO e VTCC (r = 0,089) no período de janeiro a junho é considerada fraca. No período de julho e dezembro o STO e VTCC (r = 0,49) apresentam forte correlação positiva. Essas interações entre crescimento de árvores e sazonalidade meteorológica é importante para usos práticos, tais como zoneamento de espécies visando a otimização da produção de madeira em árvores plantadas comercialmente, modelagem da produção e desenvolvimento de espécies específicas para distintas regiões climáticas, e para estimar potenciais impactos das mudanças climáticas no crescimento de diferentes espécies arbóreas (Janowiak et al., 2014; Campoe et al.,2016).

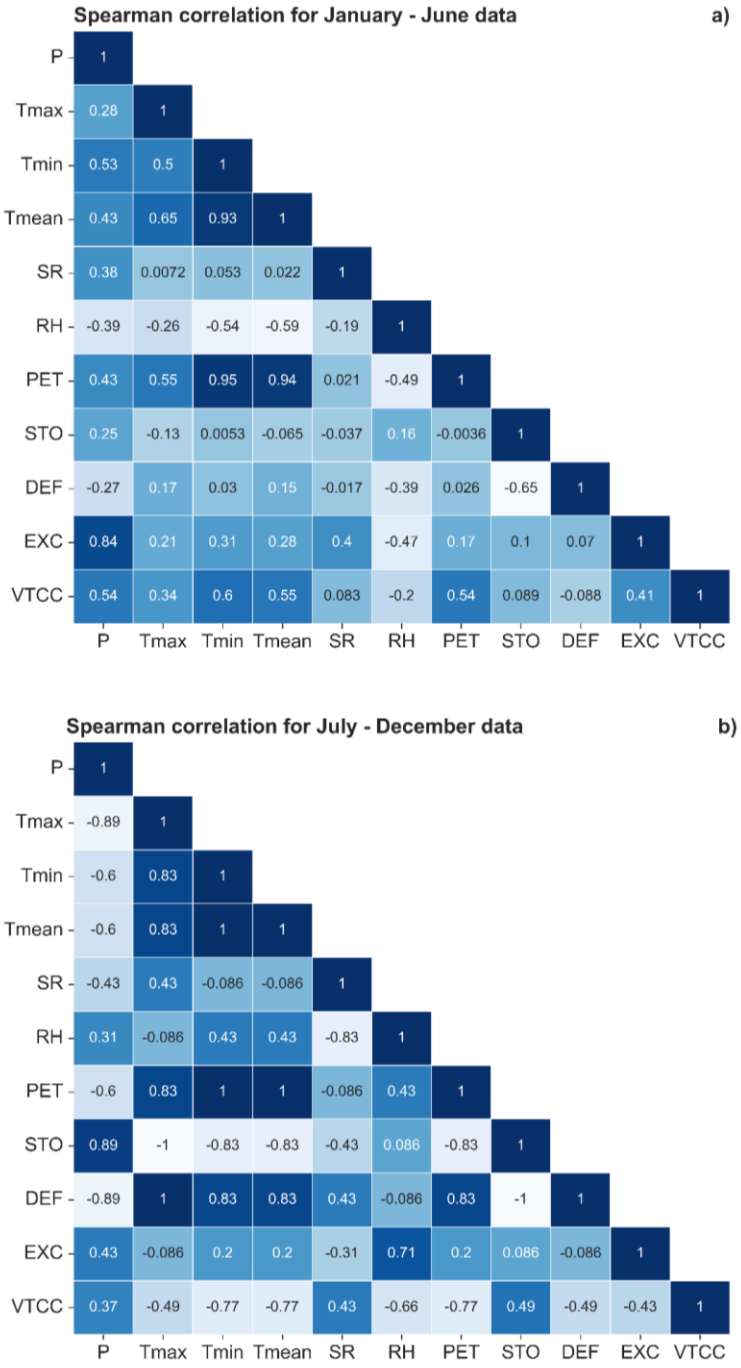

Figura. 11 – Correlação de Spearman das variáveis analisadas com a produtividade; a) período de janeiro a junho; b) período de julho a dezembro; VTCC – volume total com casca, P – precipitação, Tmax – temperatura máxima do ar, Tmin – temperatura mínima do ar, Tmean – temperatura média do ar, SR – radiação solar global, RH – umidade relativa, PET – Evapotranspiração potencial, STO – armazenamento de água no solo, DEF – Deficiência Hídrica, EXC – excedente hídrico.

Os modelos avaliados neste estudo foram a Regressão Linear Múltipla Stepwise (RLM), que usa critérios para selecionar variáveis explicativas automaticamente, como teste F, R^2 ajustado, critério de informação de Akaike (AIC) e critério de informação bayesiano (BIC) (LIU et al., 2021).

O Random Forest (RF) que é um algoritmo que ajusta várias árvores de decisão a partir de várias sub amostras do conjunto de dados, e usa uma árvore média para melhorar a precisão da estimação e controlar o ajuste excessivo (Breiman, 2001).

Suport vector machine (SVM) que é um algoritmo avançado de aprendizado de máquina que funciona separando os vetores de suporte à distância máxima usando um hiperplano (Tehrany et al., 2015). Funciona bem mesmo com o número limitado de amostras (Shaharum et al., 2018). Vários núcleos estão disponíveis no SVM, Função de Base Radial (RBF), modelo Linear, e regressão polinomial foram escolhidos para regressão.

O XGBoost desenvolve modelos por meio de um processo de treinamento aditivo. Esse algoritmo é baseado no aumento de árvores de decisão, que usa uma expressão eficiente de segunda ordem. Esse modelo é generalizável e evita o “overfitting” e “underfitting” das estimações (Chen & Guestrin, 2016; De Souza Diniz et al., 2023).

3.2.4 Regressão Linear Múltipla com seleção de variáveis por stepwise backward

A estruturação dos dados de entrada nos modelos ocorreu com a divisão do banco de dados em dois conjuntos um entre janeiro a junho e o outro conjunto entre julho a dezembro. E a decisão pela divisão do conjunto de dados em dois períodos, (Figura 11) foi realizada levando em consideração o critério da sazonalidade da região a fim de entender melhor o impacto da variação sazonal na estratégia de crescimento das florestas (Rowland et al., 2014). Portanto separamos as variáveis climáticas mensais em 1,2,3,4,5 e 6 corresponde à contagem de meses que formam o primeiro período de janeiro a junho. O mesmo esquema é aplicado no período de julho a dezembro, em que 1 passa a ser julho, 2 – agosto, 3 – setembro, 4 – outubro, 5 – novembro e 6 – dezembro.

Climatological features						Target
Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	
Jul	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**	
P_1	P_2	P_3	P_4	P_5	P_6	**Volume (m^3 ha^{-1})**
T_{max1}	T_{max2}	T_{max3}	T_{max4}	T_{max5}	T_{max6}	
T_{min1}	T_{min2}	T_{min3}	T_{min4}	T_{min5}	T_{min6}	
T_{mean1}	T_{mean2}	T_{mean3}	T_{mean4}	T_{mean5}	T_{mean6}	
RH_1	RH_2	RH_3	RH_4	RH_5	RH_6	
Qo_1	Qo_2	Qo_3	Qo_4	Qo_5	Qo_6	
ETP_1	ETP_2	ETP_3	ETP_4	ETP_5	ETP_6	
STO_1	STO_2	STO_3	STO_4	STO_5	STO_6	
DEF_1	DEF_2	DEF_3	DEF_4	DEF_5	DEF_6	
EXC_1	EXC_2	EXC_3	EXC_4	EXC_5	EXC_6	

Figura. 12 – Esquema de entrada de dados para os modelos sazonais de Random forest – RF, Regressão Linear Multipla – MLR, Suporte Vector Machine – SVM e XGBoost Regressor – XGB em que, P – precipitação (mm), Tmax – temperatura máxima do ar (°C), Tmin – temperatura mínima do ar (°C), Tmean – temperatura média do ar(°C), RH – umidade relativa, Qo – radiação no topo da atmosfera (), PET – evapotranspiração potencial (mm), STO – armazenamento de água no solo mm), DEF deficiência hídrica (mm), EXC excedente hídrico (mm).

Nesse trabalho usamos a significância estatística (valor–p < 0,05) para seleção de variáveis e regressão stepwise backward (Figura 13). A cada rodada do algoritmo, o valor-p de cada variável era testado para verificar se atendia a condição ‘p-value < 0,05’, caso, a condição fosse atendida, a variável era excluída e o modelo passava ser ajustado novamente em um procedimento stepwise até que restasse apenas as variáveis que apresentavam valor-p menor que 0,05.

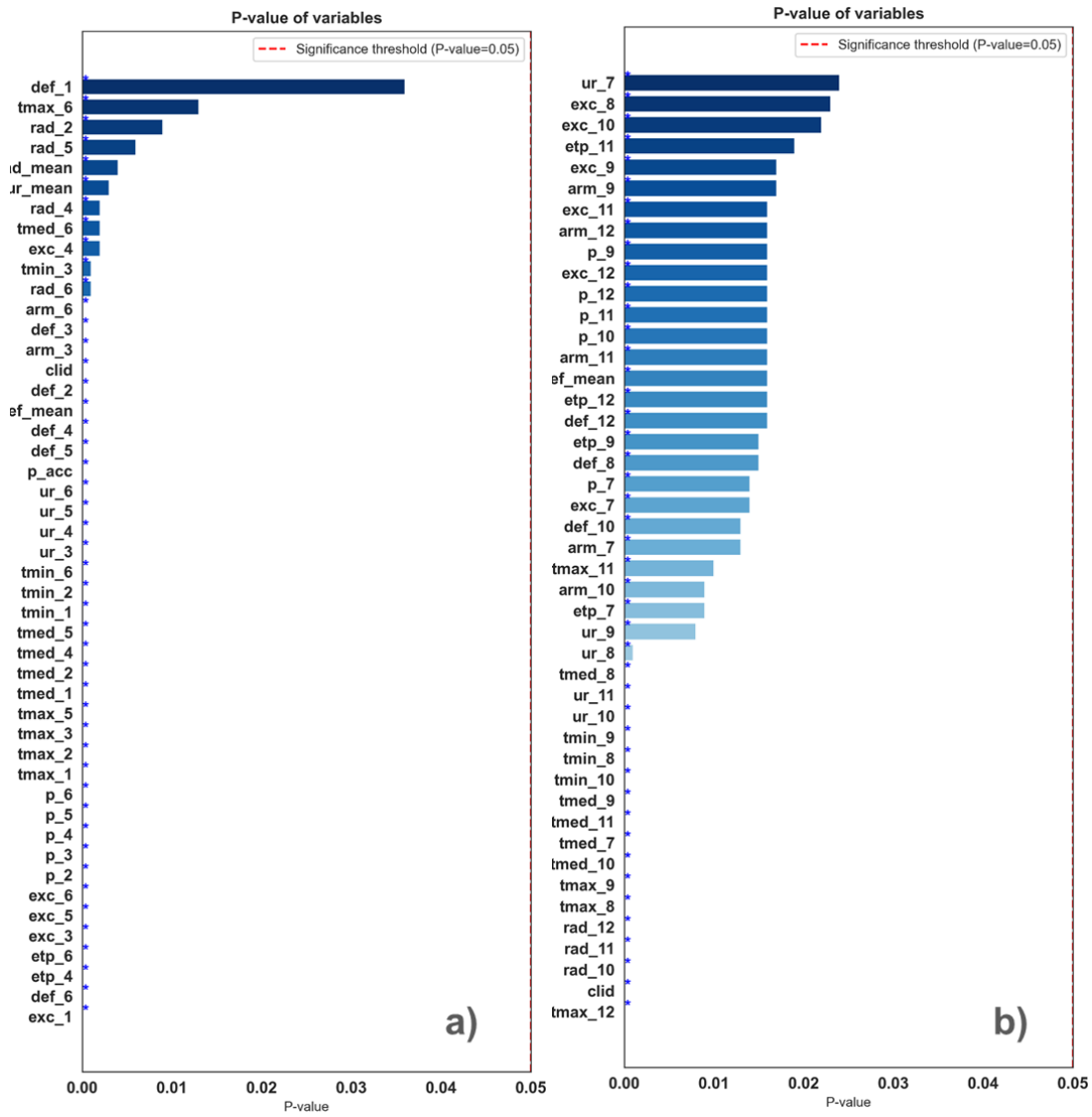

Figura. 13 – Seleção de variáveis via stepwise considerando ‘p-value’ menor que 0,05 como condição para determinação das variáveis a serem utilizadas para treinamento dos modelos; a) variáveis selecionadas via stepwise para o período de janeiro a junho; b) variáveis selecionadas via stepwise para o período de julho a dezembro.

A RLM com stepwise nos permitiu reduzir a quantidade de variáveis do banco com 71 variáveis independentes para 47 variáveis, mantendo o coeficiente de determinação R^2 = 0,91 e o RMSE em 21,15 m^3ha^{-1} para o teste inicial antes da execução do stepwise e R^2 = 0,91 e o RMSE em 21,11 m^3ha^{-1} para o período de janeiro a junho após a seleção das variáveis via regressão stepwise (Figura 14).

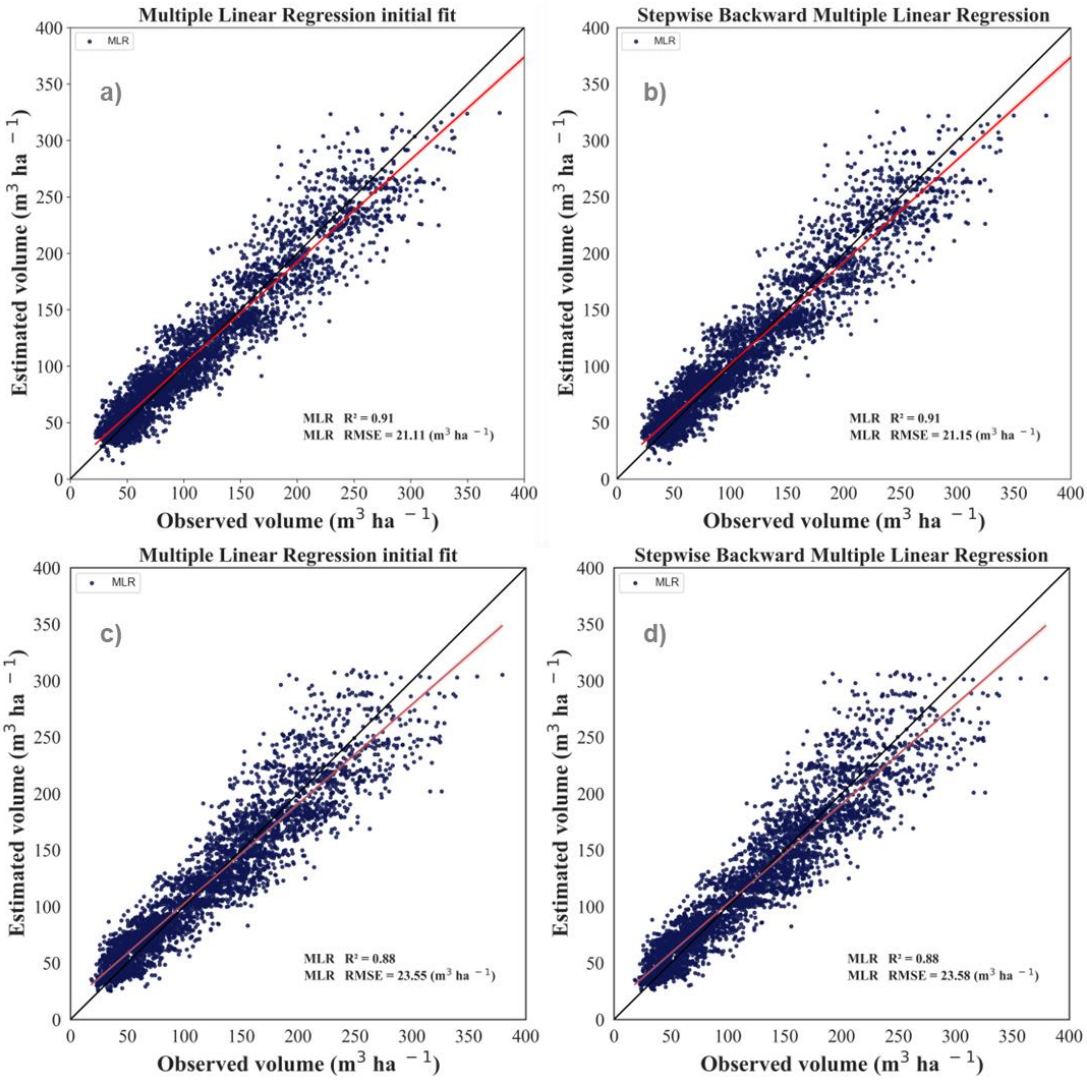

Figura.14 – Regressão Linear Multipla com Stepwise backward para seleção da variáveis; a) Ajuste inicial do modelo de regressão com 71 variáveis independentes para o período de janeiro a junho; b) Resultado do modelo de regressão após stepwise backward, com 47 variáveis para o período de janeiro a junho; c) Ajuste inicial do modelo de regressão com 71 variáveis independentes para o período de julho a dezembro; d) Resultado do modelo de regressão após stepwise backward, com 45 variáveis para o período de julho a dezembro.

Para o período de julho a dezembro, a RLM utilizando stepwise backward nos permitiu reduzir a quantidade de variáveis independentes para 45 variáveis independentes, com o coeficiente de determinação R^2 = 0,88 e o RMSE em 23,55 m^3ha^{-1} em comparação o modelo RLM sem aplicação de stepwise obteve R^2 = 0,88 e o RMSE em 23,58 m^3ha^{-1}.

3.2.5 Otimização de parâmetros dos modelos de machine learning.

Algoritmo de aprendizado de máquina (ML) tem sido amplamente utilizado em muitos domínios de aplicações, incluindo sistemas de recomendação, visão computacional, processamento de linguagem natural. Isso ocorre porque eles são genéricos e demonstram alto desempenho em problemas de análise de dados (Yang & Shami, 2020).

Em geral, a construção de um modelo eficaz é um processo complexo e demorado que envolve determinar o algoritmo apropriado e obter uma arquitetura de modelo ideal, ajustando seus parâmetros (Elshawi et al., 2019). Os parâmetros são os parâmetros que são usados para configurar um modelo de ML por exemplo, o parâmetro de penalidade, a função de ativação e os tipos de otimizador em uma rede neural e o tipo de kernel em uma máquina vetorial de suporte (Diaz et al., 2017).

O ajuste de parâmetros é considerado um componente chave da construção de um modelo ML eficaz, especialmente para modelos de ML baseados em árvores (Yang & Shami, 2020).

A escolha do melhor ajuste de parâmetros para o modelo se deu por meio do Algoritmo de busca exaustiva “GridSearchCV” da biblioteca para linguagem de programação Python scikit-learn, que testa todas as diferentes de topologias e parâmetros solicitados de forma ordenada em busca da melhor otimização, ou seja, determina o mínimo geral da função da soma de quadrados entre os dados observados e estimados (Pedregosa et al, 2011).

Para realizar a pesquisa exaustiva inicialmente determinou-se os parâmetros que seriam testados para cada um dos modelos (Tabela 1). A variáveis utilizadas nessa etapa foram as selecionadas via stepwise backward para cada período janeiro a junho e julho a dezembro separadamente.

Tabela 1 – Modelos avaliados e seus respectivos parâmetros com as faixas de valores testadas na pesquisa exaustiva (“GridSearchCV”), cada modelo apresentou um conjunto de parâmetros otimizado que foi selecionado de acordo a melhor métrica de avaliação (R^2 e RMSE).

MODELO	PARÂMETROS	FAIXA DE VALORES	VALORES OTIMIZADOS
Random Forest Regressor	n_estimator	Entre 2 e 45	29
	max_deph	Entre 3 e 40	18
Multiple Linear Regression	weights, intercept	-	-
Suport Vector Machine	C	Entre 0.1 e 10	0.1
	kernel	Linear e RBF	Linear
XGBoost	n_estimator	50, 100 e 200	100

Todos esses parâmetros foram combinados pelo algoritmo de pesquisa exaustiva. O resultado foi filtrado por meio de um loop afim de se separar as melhores configurações para obter os melhores ajustes.

3.2.6 Métricas de avaliação dos modelos

As métricas utilizadas para avaliar o desempenho e a seleção dos melhores modelos foram o coeficiente de determinação R^2 (equação 1) e o erro médio quadrático RMSE (equação 2).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Yobs_i - \bar{Y}est_i)^2}{n}} \quad (1)$$

$$R^2 = \frac{\sum_{i-1}^{N}(Yobs - \hat{Y}est)^2}{\sum_{i=1}^{N}(Yobs - \bar{Y}obs)^2} \quad (2)$$

nas quais, $Yobs_i$ são os dados observados e $Yest_i$ os dados estimados.

3.3 Resultados e Discussão

O processo de otimização dos parâmetros dos modelos de machine learning é fundamental para o desenvolvimento do modelo final de alto desempenho. Os ótimos resultados obtidos pelos modelos para estimativa (Tabela 2) do volume (produtividade do eucalipto) tendo as variáveis climáticas

e o balanço hídrico como preditores, mostram o potencial e a forte relação das condições e eventos climáticos como preditoras do volume de madeira de eucalipto.

Tabela 2 – Resultado das métricas dos modelos aplicados ao conjunto de dados de teste, para avaliação do desempenho sazonal.

MODELO SEMESTRAL JANEIRO - JUNHO				
MODELO	**TREINAMENTO**		**TESTE**	
	R^2	**RMSE**	**R^2**	**RMSE**
Random Forest Regressor	0.99	7.09	0.93	17.97
Multiple Linear Regression	0.91	21.15	0.90	21.38
Suport Vector Machine	0.89	22.18	0.90	21.80
XGBoost	0.96	13.43	0.93	18.20
MODELO SEMESTRAL JULHO - DEZEMBRO				
MODELO	**TREINAMENTO**		**TESTE**	
	R^2	**RMSE**	**R^2**	**RMSE**
Random Forest Regressor	0.99	7.95	0.92	19.29
Multiple Linear Regression	0.96	13.23	0.97	12.41
Suport Vector Machine	0.96	13.97	0.96	12.96
XGBoost	0.99	7.08	0.98	9.81

Os resultados dos testes dos modelos calibrados usando dados climáticos e de balanço hídrico apenas entre janeiro e junho de toda a série histórica (Figura 15), mostram o alto desempenho dos modelos testados, sendo possível observar um coeficiente de determinação de R^2 = 0.93 e RMSE = 17.97 m^3ha^{-1} para o modelo Random Forest, R^2 = 0.90 e RMSE = 21.38 m^3ha^{-1} para a Regressão Linear Multipla, R^2 = 0.90 e RMSE = 21.80 m^3ha^{-1} para o Suport Vector Machine e R^2 = 0.93 e RMSE = 18.20 m^3ha^{-1} para o modelo XGB.

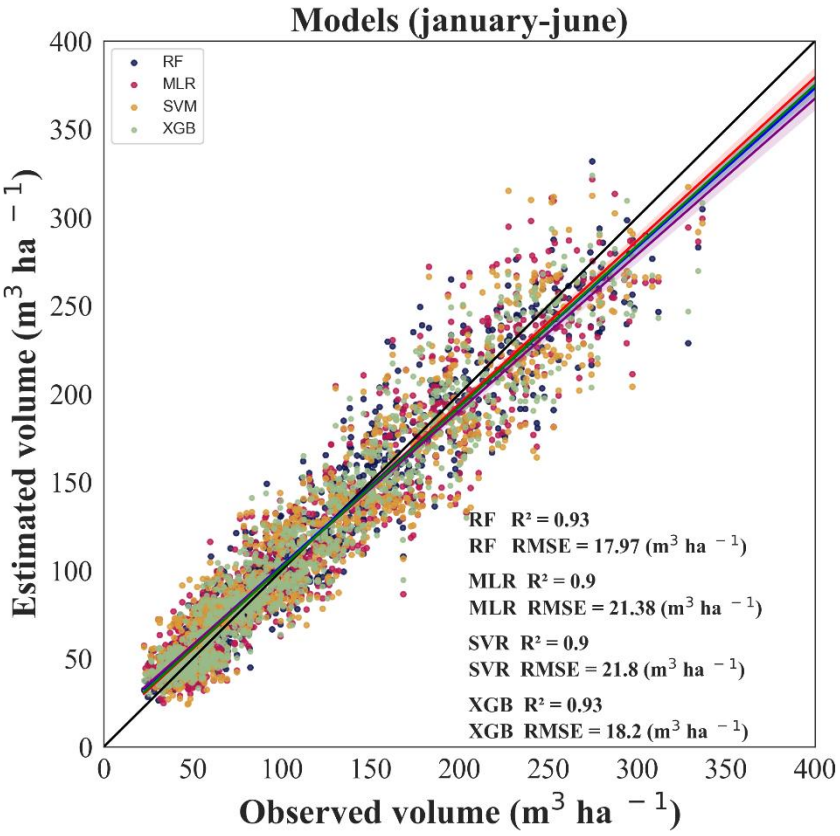

Figura. 15 – Teste dos modelos Random forest – RF, Regressão Linear Multipla – MLR, Suporte Vector Machine – SVM e XGBoost Regressor – XGB, com conjunto de dados climáticos entre janeiro e julho da série histórica de 12 anos, considerando florestas entre 2 e 7 anos de idade na região do Cerrado – Mato Grosso do Sul, Brasil.

Enquanto os testes dos modelos calibrados apenas entre julho e dezembro de toda a série histórica (Figura 16), mostraram também alto desempenho, porém diferindo razoavelmente entre as médias (Figuras 17 e 19) demonstrando a capacidade de compreensão dos padrões espaço-temporal das variáveis climáticas e a sua correlação com a produtividade. O coeficiente de determinação para os modelos de julho-dezembro foi de $R^2 = 0.92$ e RMSE = 19.29 m^3ha^{-1} para o modelo Random Forest, $R^2 = 0.97$ e RMSE = 12.41 m^3ha^{-1} para a Regressão Linear Multipla, $R^2 = 0.96$ e RMSE = 12.96 m^3ha^{-1} para o Suport Vector Machine e $R^2 = 0.98$ e RMSE = 9.81 m^3ha^{-1} para o modelo XGB.

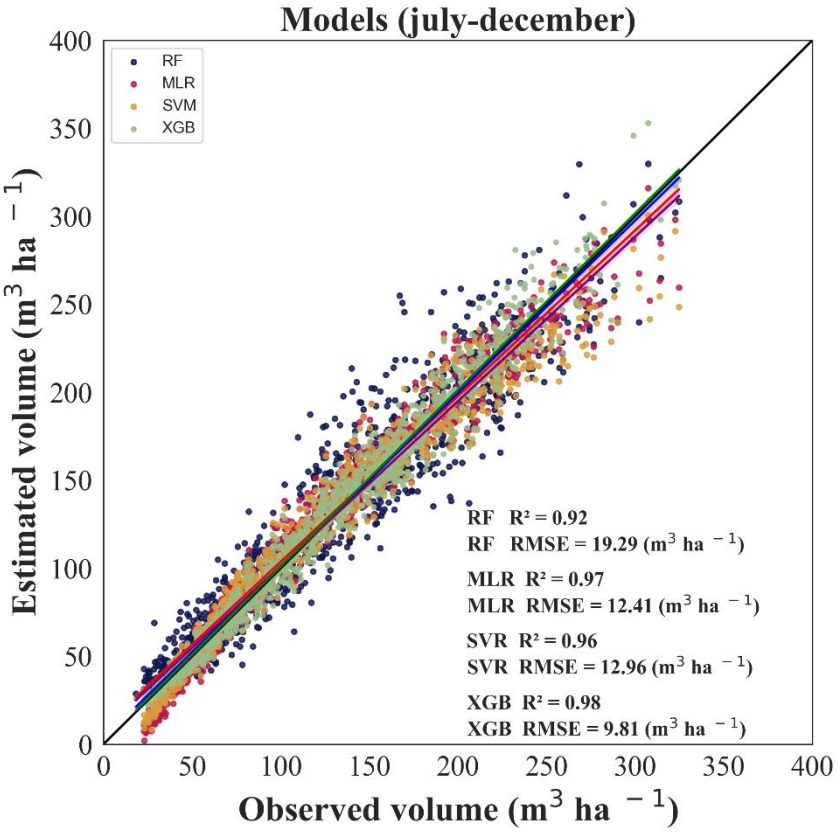

Figura. 16 – Teste dos modelos Random forest – RF, Regressão Linear Multipla – MLR, Suporte Vector Machine – SVM e XGBoost Regressor – XGB, com conjunto de dados climáticos entre janeiro e julho da série histórica de 12 anos, considerando florestas entre 2 e 7 anos de idade na região do Cerrado – Mato Grosso do Sul, Brasil.

Diversos estudos sobre a otimização e aplicação de modelos de machine learning para estimar o volume do eucalipto, fazem uso principalmente de variáveis biométricas como altura e diâmetro à altura do peito em associação com algumas variáveis climáticas, como a chuva e a temperatura o que nos remete aos modelos mecanísticos tradicionais que necessitam de variáveis biométricas para fazerem estimações.

Eli et al., (2019), utilizando uma abordagem multimodelo ensemble (modelo FAO, APSIM e 3-PG) melhorou o desempenho das estimativas de produtividade de eucalipto, com maior R^2 variando entre 0,85 e 0,89. Alvares et al, (2023), utilizando o modelo de machine learning Decision Trees aplicado ao zoneamento da produtividade do eucalipto e em múltiplos conjuntos de dados

operacionais de crescimento de árvores e fatores ambientais encontraram um alto coeficiente de determinação R^2= 0,91, um RMSE de 12,3 m3ha-1.

Durante a avaliação da modelagem de volume individual com casca utilizando diferentes técnicas de modelagem aplicadas a um banco de dados composto por variáveis biométricas, Cordeiro et al., 2022 verificou que tanto as redes neurais quanto as Support Vector Machine, obtiveram alta precisão com R2 = 0,99. É importante ressaltar que as variáveis biométricas (altura e DHP) comumente encontras em estudos de modelagem, apresentam elevada correlação com o volume.

A avaliação do incremento do volume ao longo das idades de crescimento é visível no gráfico médio para os modelos de janeiro a julho (Figura 17).

Os resultados médios dos modelos ao longo das idades 2, 3 e 4 anos são próximos quando se é possível observar a maior tendência de crescimento do eucalipto (Scolforo, 2019)

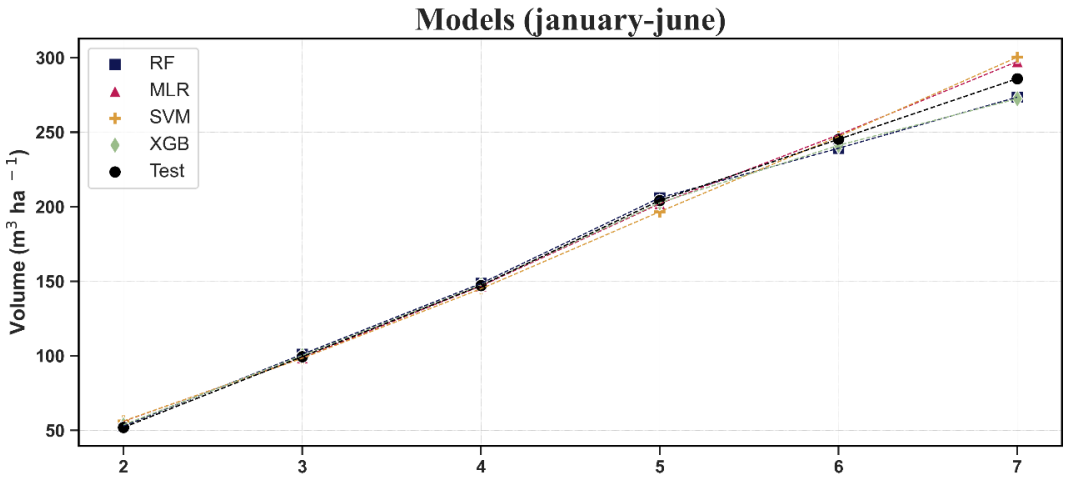

Figura. 17 – Volume estimado ao longo do crescimento do eucalipto, considerando os valores médios para cada idade do conjunto de teste dos modelos (janeiro-junho) Random forest – RF, Regressão Linear Multipla – MLR, Suporte Vector Machine – SVM e XGBoost Regressor – XGB, para a região do Cerrado – Mato Grosso do Sul, Brasil.

Na idade 6 e 7 nota-se diferença entre os modelos, vale ressaltar que normalmente há poucos dados de inventário florestal nessas idades o que limita os modelos de compreenderem melhor as relação clima-planta. Com isso o modelo Suporte Vector machine - SVM e a Regressão Linear Multipla – RLM

tendem a superestimar os resultados para essas idades com isso o SVM estima valores aos 6 anos de idade com uma diferença de 2 m^3ha^{-1} e de 14 m^3ha^{-1} em relação ao valor observado no conjunto de teste para as idades de 6 e 7 anos respectivamente (Figura 18). Enquanto o Random Forest e o XGB tendem a subestimar os resultados das estimações.

Age	Test	RF	(RF - Test)	MLR	(MLR -Test)	SVM	(SVM -Test)	XGB	(XGB -Test)
2	52	53	1	56	4	56	4	54	2
3	99	101	2	99	-1	98	-1	100	1
4	147	149	1	147	0	145	-2	147	0
5	204	206	2	202	-2	197	-8	203	-2
6	245	239	-6	248	3	247	2	241	-4
7	286	274	-12	297	11	300	14	272	-13

Figura. 18 – Valores médios de volume total com casca por idade utilizados no conjunto de teste dos modelos seguido dos resultados médios estimados com avaliação dos resíduos entre o estimado e observado para o modelo de janeiro-junho.

Já para os modelos testados com dados de julho a dezembro (Figura 19), observa-se maior variação, uma vez que as condições climáticas desse período são mais severas e isso associado à poucas informações na idade de 7 anos provoca a superestimação dos resultados (Figura 20).

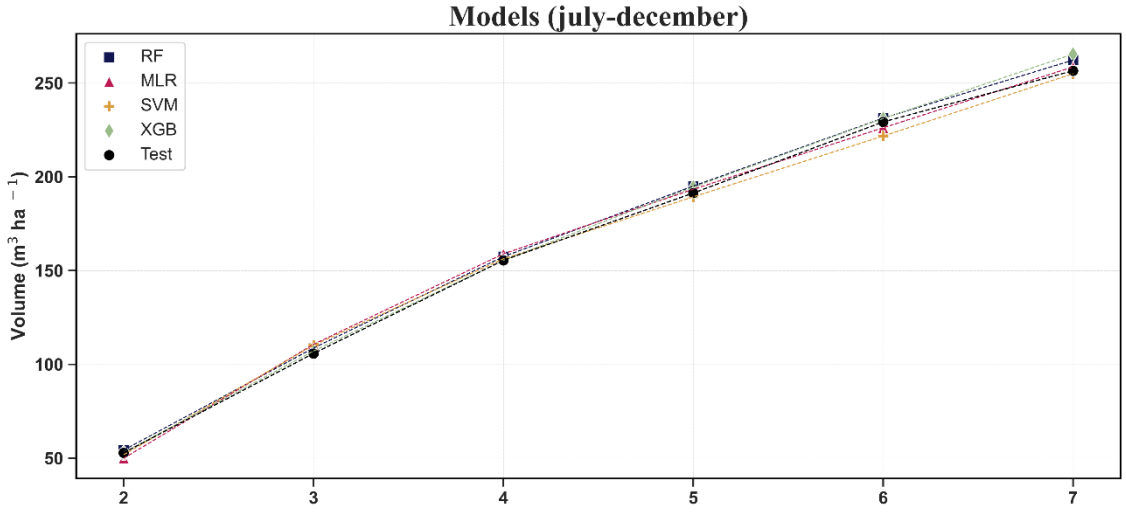

Figura. 19 – Comportamento da curva do volume estimado ao longo do crescimento do eucalipto, considerando os valores médios para cada idade do conjunto de teste dos modelos (julho-dezembro) Random forest – RF, Regressão Linear Multipla – MLR, Suporte Vector Machine – SVM e XGBoost Regressor – XGB, para a região do Cerrado – Mato Grosso do Sul, Brasil.

Os modelos Random Forest e XGboost tendem a superestimar os resultados em todas as idades, enquanto o modelo SVM subestima valores nas idades 2, 5, 6 e 7 anos, com diferenças de -1 m^3ha^{-1}, -2 m^3ha^{-1}, -8 m^3ha^{-1} e -2 m^3ha^{-1} respectivamente, em relação ao valor observado no conjunto de teste (Figura 18).

Age	Test	RF	(RF - Test)	MLR	(MLR -Test)	SVM	(SVM -Test)	XGB	(XGB -Test)
2	53	54	1	50	-3	52	-1	53	0
3	106	109	3	111	5	110	4	107	1
4	155	157	2	159	3	156	1	155	0
5	191	195	4	193	2	189	-2	194	3
6	229	231	2	226	-3	222	-8	231	2
7	257	262	6	259	2	255	-2	265	9

Figura. 20 – Valores médios de volume total com casca por idade utilizados no conjunto de teste dos modelos seguido dos resultados médios estimados com avaliação dos resíduos entre o estimado e observado para o modelo de julho-dezembro.

Compreender a relação solo planta atmosfera é fundamental para a modelagem. Entender como o componente clima se relaciona com o eucalipto é o primeiro passo para a criação de modelos de estimativa mais rápidos e precisos. A disponibilidade de dados climáticos histórico permite a avalição temporal de modelos de machine learning, elevando assim a confiabilidade para seu uso em estimativas do volume de eucalipto. A evolução no campo de Machine Learning aplicado ao setor florestal permite que modelos específicos, treinados, testados e recorrentemente validados, sejam utilizados para auxiliar em diversos processo como modelagem e avaliação dos impactos de eventos climáticos extremos.

3.4. Conclusões

Todos os modelos apresentam elevadas acurácias, porém o Random Forest caracterizou-se como o melhor modelo, apresentando as melhores métricas durante o treinamento e teste com, R^2= 0.93 e RMSE = 18.36 m^3ha^{-1} para o modelo janeiro-junho e R^2= 0.92 e RMSE = 19. 52 m^3ha^{-1} para o modelo de julho-dezembro.

Para o período de janeiro a junho, o modelo Suporte Vector machine - SVM e a Regressão Linear Multipla – RLM tendem a superestimar os resultados em 2 m^3ha^{-1} e de 14 m^3ha^{-1} para as idades de 6 e 7 anos respectivamente, enquanto os modelos Random Forest e o XGB subestimam os resultados nessas idades.

Para o período de julho a dezembro, os modelos Random Forest superestimam os resultados em todas as idades, enquanto o modelo SVM subestimam valores nas idades 2, 5, 6 e 7 anos, com diferenças de -1 m^3ha^{-1}, -2 m^3ha^{-1}, -8 m^3ha^{-1} e -2 m^3ha^{-1} respectivamente, em relação ao valor observado.

REFERÊNCIAS

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. D. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. **Meteorologische *zeitschrift***, *22*(6), 711-728.

Alvares, Clayton Alcarde et al. Decision-Tree Application to Predict and Spatialize the Wood Productivity Probabilities of Eucalyptus Plantations. **Forests**, v. 14, n. 7, p. 1334, 2023.

Billings, W. Dwight. The environmental complex in relation to plant growth and distribution. **The Quarterly Review of Biology**, v. 27, n. 3, p. 251-265, 1952.

Breiman, Leo. Random forests. **Machine learning**, v. 45, p. 5-32, 2001.

Campoe, Otávio C. et al. Meteorological seasonality affecting individual tree growth in forest plantations in Brazil. **Forest Ecology and Management**, v. 380, p. 149-160, 2016.

Chen, Tianqi; GUESTRIN, Carlos. Xgboost: A scalable tree boosting system. In: **Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining**. 2016. p. 785-794.

Christina, Mathias et al. Importance of deep water uptake in tropical eucalypt forest. **Functional Ecology**, v. 31, n. 2, p. 509-519, 2017.

Cordeiro, Márcio Assis et al. Volumetric estimates in eucalyptus stands using support vector machines and artificial neural networks. **Madera y bosques**, v. 28, n. 1, 2022.

de Alcântara, Aline Edwiges Mazon et al. Use of artificial neural networks to assess yield projection and average production of eucalyptus stands. **African Journal of Agricultural Research**, v. 13, n. 42, p. 2285-2297, 2018.

de Oliveira Neto, Ricardo Rodrigues et al. Estimation of Eucalyptus productivity using efficient artificial neural network. **European Journal of Forest Research**, v. 141, n. 1, p. 129-151, 2022.

de Souza Diniz, Juliana Maria Ferreira et al. Estimating stem volume of Eucalyptus sp. and Pinus sp. plantations in Brazil, using Sentinel-1B and ALOS-2/PALSAR-2 data. **Journal of Applied Remote Sensing**, v. 17, n. 1, p. 014513-014513, 2023.

Diaz, Gonzalo I. et al. An effective algorithm for hyperparameter optimization of neural networks. **IBM Journal of Research and Development**, v. 61, n. 4/5, p. 9: 1-9: 11, 2017.

Elli, Elvis Felipe et al. Intercomparison of structural features and performance of Eucalyptus simulation models and their ensemble for yield estimations. **Forest Ecology and Management**, v. 450, p. 117493, 2019.

Elli, Elvis Felipe; SENTELHAS, Paulo Cesar; BENDER, Fabiani Denise. Impacts and uncertainties of climate change projections on Eucalyptus plantations productivity across Brazil. **Forest Ecology and Management**, v. 474, p. 118365, 2020.

Elshawi, Radwa; MAHER, Mohamed; SAKR, Sherif. Automated machine learning: State-of-the-art and open challenges. **arXiv preprint arXiv:** 1906.02287, 2019.

FAO 2020 Global Forest Resources Assessment 2020 (Rome: FAO) (available at: https://doi.org/10.4060/ca8753en)

He, Haibo et al. Incremental learning from stream data. **IEEE Transactions on Neural Networks**, v. 22, n. 12, p. 1901-1914, 2011.

Li, Ying chang et al. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. **Scientific reports**, v. 10, n. 1, p. 9952, 2020.

Liu, Yingxia et al. Analysis of spatio-temporal variation of crop yield in China using stepwise multiple linear regression. **Field Crops Research**, v. 264, p. 108098, 2021.

Maulud, Dastan; ABDULAZEEZ, Adnan M. A review on linear regression comprehensive in machine learning. **Journal of Applied Science and Technology Trends**, v. 1, n. 4, p. 140-147, 2020.

Messier, C., Bauhus, J., Sousa-Silva, R., Auge, H., Baeten, L., Barsoum, N., ... & Zemp, D. C. (2022). For the sake of resilience and multifunctionality, let's diversify planted forests!. **Conservation Letters**, 15(1), e12829.

Otto, Marina Shinkai Gentil et al. Photosynthesis, stomatal conductance and productivity of Eucalyptus clones under different soil and climatica conditions. **Revista Árvore**, v. 37, p. 431-439, 2013.

Pedregosa, Fabian et al. Scikit-learn: Machine learning in Python. **the Journal of machine Learning research**, v. 12, p. 2825-2830, 2011.

Ramantswana, Muedanyi; GUERRA, Saulo Philipe Sebastião; ERSSON, Back Tomas. Advances in the mechanization of regenerating plantation forests: A review. **Current forestry reports**, v. 6, p. 143-158, 2020.

Rowland, Lucy et al. The sensitivity of wood production to seasonal and interannual variations in climate in a lowland Amazonian rainforest. **Oecologia**, v. 174, p. 295-306, 2014.

Santana, Dthenifer Cordeiro et al. Machine Learning Methods for Woody Volume Prediction in Eucalyptus. **Sustainability**, v. 15, n. 14, p. 10968, 2023.

Scolforo, Henrique Ferraco et al. Eucalyptus growth and yield system: Linking individual-tree and stand-level growth models in clonal Eucalypt plantations in Brazil. **Forest Ecology and Management**, v. 432, p. 1-16, 2019.

Seng Hua, Lee et al. Engineering wood products from Eucalyptus spp. **Advances in Materials Science and Engineering**, v. 2022, p. 1-14, 2022.

Shaharum, N. S. N. et al. Image classification for mapping oil palm distribution via support vector machine using scikit-learn module. **The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. 42, p. 133-137, 2018.

Silva, Jeferson Pereira Martins et al. Prognosis of forest production using machine learning techniques. **Information Processing in Agriculture**, 2021.

Silva, L. C. R. Dinâmica de transição e interações entre fitofisionomias florestais e formações vegetacionais abertas do bioma Cerrado. Brasília-DF. 2007. 168p. Dissertação (Mestrado em Ciências Florestais) – Departamento de Engenharia Florestal, Universidade de Brasília, Brasília, 2007.

Stape, Jose Luiz et al. The Brazil Eucalyptus Potential Productivity Project: Influence of water, nutrients and stand uniformity on wood production. **Forest Ecology and Management**, v. 259, n. 9, p. 1684-1694, 2010.

Stape, Jose Luiz et al. Water use, water limitation, and water use efficiency in a Eucalyptus plantation. **Revista Bosque**, v. 25, n. 2, p. 35-41, 2004.

Tehrany, Mahyat Shafapour et al. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. **Catena**, v. 125, p. 91-101, 2015.

Thornton, P. K., Ericksen, P. J., Herrero, M., & Challinor, A. J. (2014). Climate variability and vulnerability to climate change: a review. **Global change biology**, *20*(11), 3313-3328.

Van de Ven, Gido M.; TUYTELAARS, Tinne; TOLIAS, Andreas S. Three types of incremental learning. **Nature Machine Intelligence**, v. 4, n. 12, p. 1185-1197, 2022.

Vieira, Giovanni Correia et al. Prognoses of diameter and height of trees of eucalyptus using artificial intelligence. **Science of the Total Environment**, v. 619, p. 1473-1481, 2018.

Yang, Li; SHAMI, Abdallah. On hyperparameter optimization of machine learning algorithms: Theory and practice. **Neurocomputing**, v. 415, p. 295-316, 2020.

Zhou, Xiaoguo et al. Effects of understory management on trade-offs and synergies between biomass carbon stock, plant diversity and timber production in eucalyptus plantations. **Forest Ecology and Management**, v. 410, p. 164-173, 2018.