



Serviço Público Federal  
Ministério da Educação  
Fundação Universidade Federal de Mato Grosso do Sul



**ESCOLA SUPERIOR DE ADMINISTRAÇÃO E NEGÓCIOS – ESAN PROGRAMA  
DE PÓS-GRADUAÇÃO *STRICTO SENSU* EM ADMINISTRAÇÃO**

FRANCIANE RODRIGUES

**CLASSIFICAÇÃO DA ESTIAGEM NA SOJICULTURA COM MACHINE  
LEARNING: UMA ABORDAGEM APLICADA À REGIÃO DE  
DOURADOS/MS**

CAMPO GRANDE - MS  
2025

FRANCIANE RODRIGUES

**CLASSIFICAÇÃO DA ESTIAGEM NA SOJICULTURA COM MACHINE  
LEARNING: UMA ABORDAGEM APLICADA À REGIÃO DE  
DOURADOS/MS**

Tese apresentada como requisito parcial à obtenção do grau de Doutora em Administração. Curso de Pós-Graduação *stricto sensu* em Administração, nível de Doutorado da Escola de Administração e Negócios da Universidade Federal de Mato Grosso do Sul.

Orientador: Prof. Dr. Matheus Wemerson  
Gomes Pereira

CAMPO GRANDE - MS  
2025

*Dedico este trabalho ao meu esposo, Julio Alberto Peres Ferencz Júnior.*

## AGRADECIMENTOS

A Deus.

Ao meu esposo, Julio, pelo amor, companheirismo, parceria e por me incentivar a nunca desistir deste doutorado. Obrigada por tudo e por tanto. Amo-te!

Aos meus pais, Hélio e Tiana, ao meu irmão, Hélio Júnior, e à minha cunhada, Priscila, pelo amor incondicional e pelo suporte inestimável.

Ao meu orientador, Matheus, por sua escuta generosa, paciência inabalável e por estar sempre presente, mesmo nos meus silêncios. Obrigada por confiar no meu trabalho e por ser, antes de tudo, um orientador e professor verdadeiramente humano. Tive muita sorte!

Aos velhos amigos Marcos, Karen, Anna Rachel, Rafael, Ivanildo, Thiago, Alexandre e Cleison, pela amizade de tantos anos e por estarem ao meu lado em diferentes etapas deste percurso.

Aos novos amigos, Sandra Alice, por todo o carinho e cuidado para comigo, e Mário Alexandre, pela amizade, suporte, incentivo para terminar logo essa tese e parceria nesta fase final.

Aos amigos da SEMADESC: Ricardo Senna, Carlos Henrique e Carlos Eduardo (*in memoriam*) pelo apoio, incentivo desde o início do doutorado e pelas ricas contribuições.

À Inovvati Tecnologia, pelo apoio e por abrir espaço no mercado de trabalho a perfis acadêmicos como o meu. E, em especial, ao Wallace, pela generosidade em compartilhar ideias e conselhos profissionais ao longo da jornada e ao Daniel por toda sua transparência e amizade para comigo. Ambos são exemplos de liderança para mim.

À Datatech, especialmente a Renan e Laís, por todo o apoio, pela torcida de sempre e por acreditarem em mim desde o começo. Obrigada por abrirem espaço para quem vem da academia e por acolherem, com tanto carinho, minha ausência nos momentos decisivos do doutorado.

Aos professores e colegas da Escola Superior de Administração e Negócios (ESAN) pela oportunidade de desenvolver esta pesquisa em um ambiente acadêmico acolhedor e transformador.

*"Se eu vi mais longe, foi por estar sobre ombros de gigantes."  
Isaac Newton.*

## RESUMO

RODRIGUES, F. **Classificação da Estiagem na Sojicultura com Machine Learning: uma abordagem aplicada à região de Dourados/MS.** 2025. 90 f. Tese (Doutorado em Administração) – Escola de Administração e Negócios, Universidade Federal de Mato Grosso do Sul, Campo Grande, MS, 2025.

Este estudo teve como objetivo classificar a ocorrência de estiagens na cultura da soja em Dourados/MS, de modo a aprimorar o monitoramento climático e fornecer subsídios aos stakeholders do setor agrícola, favorecendo a tomada de decisão e a mitigação dos riscos associados à variabilidade pluviométrica. A pesquisa foi norteada pela seguinte questão: “como o desenvolvimento de um modelo de machine learning, baseado em um indicador específico de estiagem para a sojicultura, pode apoiar a classificação, o monitoramento contínuo e a mitigação dos impactos da estiagem em Dourados/MS?”. Os resultados demonstraram a eficiência do método proposto em identificar estiagens com base na precipitação acumulada e permitindo um acompanhamento sistemático das condições climáticas e sua relação com a produtividade agrícola. Uma das principais contribuições foi a formulação do conceito de “estiagem da cultura da soja”, que integra os critérios de estiagem e veranico, estabelecendo um limiar crítico de 20,6 mm em janelas móveis de cinco dias. Este indicador, inédito na literatura, mostrou-se sensível às exigências fenológicas da cultura e mais preciso para fins de monitoramento. A análise da precipitação evidenciou forte correlação com a produtividade, como verificado na comparação entre as safras 2021/2022, marcada por escassez pluviométrica e perdas significativas, e 2022/2023, caracterizada por melhor desempenho diante de chuvas regulares. A partir desse conceito, foram aplicados algoritmos de machine learning para classificar estiagens e avaliar seu potencial como ferramenta de apoio à decisão. O Random Forest apresentou acurácia elevada (96,24% no treinamento e 85,79% no teste), mas com número expressivo de falsos negativos (490), limitando sua utilidade em cenários críticos. O Extra Trees reduziu falsos negativos (112), mas gerou excesso de falsos positivos (640), comprometendo a confiabilidade. O XGB Classifier destacou-se como o modelo mais promissor, alcançando acurácia de 96,90% no treinamento e 87,36% no teste, com equilíbrio entre recall (0,87) e precisão (0,94) para a classe “com estiagem”, evidenciando maior robustez no trade-off entre segurança e confiabilidade. A relevância prática do modelo foi confirmada pela análise contrafactual, que demonstrou que, na ausência da estiagem de 2021/2022, a produtividade média poderia ter se aproximado de 54,43 sc/ha, contra as 24,08 sc/ha registradas, o que representou perdas de 56%. Além disso, a simulação econômica da aplicação seletiva de irrigação indicou um retorno sobre o investimento (ROI) de aproximadamente 470%, reforçando o potencial do modelo como suporte a estratégias de manejo hídrico, mesmo em solos de alta retenção, como os Latossolos Vermelhos da região. Diante desses resultados, o modelo mostra-se aplicável a sistemas de alerta climático e plataformas de monitoramento agrícola, com potencial de adoção por produtores, cooperativas e instituições públicas. Sua utilização pode orientar decisões como replantio, irrigação seletiva e ajustes no manejo durante o ciclo produtivo. Recomenda-se, para trabalhos futuros, a adaptação do método a outras regiões e culturas dependentes da chuva, com calibração de limiares conforme as condições edafoclimáticas locais. Ainda, sugere-se a inclusão de variáveis agronômicas e ambientais — como tipo de solo, práticas de manejo e características das cultivares — de modo a ampliar a precisão e a contextualização das previsões.

**Palavras-chave:** Inteligência artificial, agricultura, indicadores agroclimáticos.

## ABSTRACT

RODRIGUES, F. **Drought Classification in Soybean Cultivation Using Machine Learning: An Applied Approach to the Region of Dourados, Brazil.** 2025. 90 pages. Doctoral Tesys (Ph.D. in Administration) – School of Administration and Business, Federal University of Mato Grosso do Sul, Campo Grande, MS, 2025.

This study aimed to classify the occurrence of drought events in soybean cultivation in Dourados/MS, with the purpose of improving climate monitoring and providing support to agricultural stakeholders, thereby fostering decision-making and mitigating risks associated with rainfall variability. The research was guided by the following question: “How can the development of a machine learning model, based on a specific drought indicator for soybean cultivation, support classification, continuous monitoring, and the mitigation of drought impacts in Dourados/MS?” The results confirmed the efficiency of the proposed method in identifying droughts based on accumulated precipitation, enabling systematic monitoring of climate conditions and their relationship with agricultural productivity. One of the main contributions of this work was the formulation of the concept of “soybean crop drought,” which integrates the criteria of drought and veranico (short dry spells), establishing a critical threshold of 20.6 mm in five-day moving windows. This indicator, unprecedented in the literature, proved sensitive to the crop’s phenological requirements and more precise for monitoring purposes. Precipitation analysis revealed a strong correlation with productivity, as demonstrated by the comparison between the 2021/2022 season, marked by severe water scarcity and significant yield losses, and the 2022/2023 season, which showed better results due to more regular rainfall distribution. Based on this concept, different machine learning algorithms were applied to classify droughts and evaluate their potential as decision-support tools. The Random Forest Classifier achieved high accuracy (96.24% in training and 85.79% in testing) but generated a high number of false negatives (490), limiting its usefulness in critical scenarios. The Extra Trees Classifier reduced false negatives (112) but produced an excessive number of false positives (640), compromising reliability. The XGB Classifier stood out as the most promising model, achieving 96.90% accuracy in training and 87.36% in testing, with balanced performance for both classes, including a recall of 0.87 and precision of 0.94 for the “with drought” class. This demonstrated greater robustness in managing the trade-off between security (minimizing false negatives) and reliability. The practical relevance of the model was further validated by a counterfactual analysis, which estimated that, in the absence of the 2021/2022 drought, average yield could have approached 54.43 bags per hectare, compared with the observed 24.08 bags per hectare, representing a 56% loss. Additionally, an economic simulation of selective irrigation indicated a return on investment (ROI) of approximately 470%, highlighting the model’s potential to support water management strategies, even in high-retention soils such as the region’s Red Latosols. Given these results, the model proves applicable to climate alert systems and agricultural monitoring platforms, with adoption potential by producers, cooperatives, and public institutions. Its use can guide timely decisions such as replanting, selective irrigation, and management adjustments throughout the crop cycle. For future research, the adaptation of this method to other regions and rain-dependent crops is recommended, with calibration of thresholds according to local edaphoclimatic conditions. Furthermore, the integration of agronomic and environmental variables—such as soil type, management practices, and cultivar characteristics—could enhance the model’s accuracy and contextualization, strengthening its role as a data-driven decision-support tool.

**Keywords:** Artificial intelligence, agriculture, agroclimatic indicators.

## LISTA DE FIGURAS

Figura 1 - Representação gráfica do crescimento econômico no longo prazo.....	24
Figura 2 - Mapa regionalizado dos diferentes ciclos anuais de precipitação na América do Sul .....	27
Figura 3 - Esquema ilustrativo da localização dos sistemas climáticos na América do Sul durante o período chuvoso na baixa e na alta troposfera .....	28
Figura 4 - Estádios Fenológicos da Cultura da Soja .....	35
Figura 5 - Localização do município de Dourados (MS). .....	41
Figura 6 - Estruturação em Excel dos dados meteorológicos e da classificação da estiagem ..	45
Figura 7 - Fluxo metodológico para o modelo de classificação de estiagem da sojicultura.....	50
Figura 8 - Comparação da Produtividade da Soja em Dourados e Mato Grosso do Sul por Safra.....	52
Figura 9 - Janelas móveis de precipitação acumulada em Dourados/MS no ano safra 2021/2022 .....	55
Figura 10 - Precipitação por decêndio em Dourados/MS no ano safra 2021/2022 .....	56
Figura 11 - Quantidade de janelas de estiagens no ano de 2021/2022 em Dourados/MS. Acumulado de 5 dias à esquerda e 10 dias à direita .....	60
Figura 12 - Matriz de correlação entre as variáveis.....	62
Figura 13 - Matriz de confusão da aplicação do algoritmo Random Forest Classifier .....	66
Figura 14 - Matriz de confusão da aplicação do algoritmo Extra Trees Classifier .....	68
Figura 15 - Matriz de confusão da aplicação do algoritmo XGB Classifier .....	70
Figura 16 - Quantitativo real de janelas com estiagem à esquerda e sem estiagem à direita em Dourados/MS no acumulado de 5 dias de precipitação .....	71
Figura 17 – Comparação da produtividade da soja em Dourados/MS: cenário real vs. contrafactual.....	76

## LISTA DE QUADROS

Quadro 1 - Resumo comparativo de técnicas para estudos relacionados a seca e estiagens.....	33
Quadro 2 - Estádios de desenvolvimento da soja.....	36
Quadro 3 – Síntese dos métodos, tipos e técnicas de pesquisa.....	40
Quadro 4 - Matriz de confusão aplicada à classificação de estiagem na sojicultura.....	46
Quadro 5 – Comparativo entre algoritmos .....	48
Quadro 6 – Matriz de amarração .....	51
Quadro 7 - Resumo dos resultados do monitoramento diário de estiagem combinada.....	58

## LISTA DE TABELAS

Tabela 1 - Estatísticas descritivas das precipitações das janelas móveis.....	59
Tabela 2 - Performance de diferentes modelos de classificação com base em várias métricas.....	64
Tabela 3 - Relatório de resultados da aplicação do algoritmo <i>Random Forest Classifier</i> .....	66
Tabela 4 - Relatório de resultados da aplicação do algoritmo <i>Extra Trees Classifier</i> .....	68
Tabela 5 - Relatório de resultados da aplicação do algoritmo <i>XGB Classifier</i> .....	69
Tabela 6 - Comparativo dos modelos de aprendizado de máquina para classificação de estiagem na soja em Dourados/MS.....	72
Tabela 7 - Simulação econômica da aplicação seletiva de irrigação baseada na classificação de estiagem.....	77

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	<b>13</b>
1.1 PROBLEMÁTICA .....	15
1.2 JUSTIFICATIVA .....	18
1.3 OBJETIVOS.....	19
1.4 ORGANIZAÇÃO DA TESE.....	20
<b>2. FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>21</b>
2.1. EVOLUÇÃO DOS ESTUDOS DE CRESCIMENTO ECONÔMICO E SUA CONEXÃO COM A TECNOLOGIA E O CLIMA .....	21
2.2 DINÂMICA ATMOSFÉRICA E A VARIABILIDADE PLUVIOMÉTRICA .....	27
2.3 A TECNOLOGIA NO MONITORAMENTO DE SECAS E ESTIAGENS .....	30
2.4 CARACTERÍSTICAS E EXIGÊNCIAS CLIMÁTICAS DA SOJA.....	34
2.5 INTEGRAÇÃO DOS CONCEITOS TEÓRICOS.....	37
<b>3. PROCEDIMENTOS DE ANÁLISE</b> .....	<b>39</b>
3.1 CLASSIFICAÇÃO DA PESQUISA .....	39
3.2 ÁREA DE ESTUDO .....	40
3.3 BASE DE DADOS E HIPÓTESE DE PESQUISA .....	42
3.4 PROCEDIMENTOS METODOLÓGICOS .....	43
3.4.1 Coleta e Tratamento dos Dados.....	43
3.4.2 Definição do Indicador de Estiagem.....	43
3.4.3 Seleção e Engenharia de Variáveis.....	44
3.4.4 Formulação Matemática.....	44
3.4.5 Avaliação dos Modelos.....	45
3.4.6 Validação dos Modelos.....	47
3.4.7 Descritivo dos Algoritmos.....	47
3.4.8 Análise Contrafactual da Produtividade da Soja.....	49
3.4.9 Ferramentas Utilizadas.....	49
3.4.10 Fluxo dos procedimentos .....	49
3.5 MATRIZ DE AMARRAÇÃO.....	51
<b>4. RESULTADOS E DISCUSSÕES</b> .....	<b>52</b>
4.1 COMPARATIVO DE PRODUTIVIDADE DA SOJA.....	52
4.2 ANÁLISE DA PRECIPITAÇÃO NO ANO-SAFRA 2021/2022 .....	55
4.3 DETERMINAÇÃO DA ESTIAGEM DA CULTURA DA SOJA.....	59

4.4 MACHINE MACHINE LEARNING PARA CLASSIFICAÇÃO DA ESTIAGEM DA CULTURA DA SOJA .....	61
4.5 VALIDAÇÃO DO MODELO DE MACHINE LEARNING PARA ESTIAGEM DA CULTURA DE SOJA .....	70
4.6 DEFINIÇÃO DO MODELO DE CLASSIFICAÇÃO DA ESTIAGEM DA CULTURA DA SOJA.....	72
4.7 ANÁLISE CONTRAFACTUAL: IMPACTO DA ESTIAGEM NA SAFRA 2021/2022.....	75
<b>5. CONSIDERAÇÕES FINAIS .....</b>	<b>79</b>
REFERÊNCIAS .....	82

## 1. INTRODUÇÃO

A soja consolidou-se como um dos principais produtos agrícolas brasileiros, exercendo papel central no comércio exterior, na geração de emprego e renda e no desenvolvimento regional. Nas últimas cinco décadas, a produção nacional multiplicou-se em aproximadamente 262 vezes, evidenciando a relevância estratégica da cultura para a economia do país (EMBRAPA, 2021). Tal crescimento impulsionou a formação de um complexo agroindustrial que integra produção, industrialização e comercialização, promovendo sinergias entre diferentes cadeias produtivas e consolidando o Brasil como um dos maiores produtores e exportadores mundiais de soja (EMBRAPA, 2010).

Estudos da Embrapa Soja demonstram que, além do impacto econômico direto, a cultura exerce influência socioeconômica significativa em âmbito regional, atraindo mão de obra especializada, fomentando a inovação tecnológica e gerando transformações estruturais no espaço agrícola nacional (EMBRAPA, 2023).

Ademais, a expressiva participação do grão e de seus derivados no comércio exterior brasileiro confirma sua importância estratégica: aproximadamente 38% da produção de soja em grão, 63% do farelo e 44% do óleo destinam-se ao mercado externo, o que evidencia o papel do setor como vetor de dinamização econômica e de inserção internacional do agronegócio brasileiro (EMBRAPA, 2021b).

A diversificação das exportações tem sido um fator determinante nesse avanço, refletindo-se no desempenho do agronegócio brasileiro, que registrou US\$ 140 bilhões em exportações entre janeiro e outubro de 2024, um crescimento de 0,7% em relação ao mesmo período do ano anterior (MAPA, 2024). Nesse contexto, o complexo da soja se destacou, respondendo por US\$ 50,3 bilhões, representando 35,9% do total exportado pelo setor agropecuário brasileiro (MAPA, 2024). Além disso, o Brasil consolidou sua posição como maior exportador global do grão, alcançando 147,38 milhões de toneladas exportadas em 2024, evidenciando a relevância do setor na economia nacional e internacional (CONAB, 2024).

Esses resultados econômicos, especialmente os relacionados à sojicultura, estão diretamente relacionados aos avanços tecnológicos em cultivares, às inovações nas técnicas de manejo e à otimização dos fatores de produção, que incluem aspectos genéticos e ambientais (Santos *et al.*, 2017). O componente ambiental, por sua vez, é influenciado por condições edáficas e climáticas, que têm um papel crucial no desempenho da cultura da soja (Santos *et al.*, 2017). Sob a ótica climática, cerca de 80% da variabilidade da produção agrícola mundial

é atribuída às condições climáticas e meteorológicas ao longo do ciclo produtivo das culturas, influenciando diretamente os processos metabólicos das plantas e as atividades agrícolas (Farias *et al.*, 2009).

No Brasil, as variações na distribuição de chuvas e os eventos de estiagem são desafios constantes para a sojicultura, impactando tanto no desenvolvimento das culturas quanto ampliando a incerteza na produção (Arsego *et al.*, 2018). Essa imprevisibilidade climática pode ocasionar perdas parciais ou totais da produção, redução da qualidade do grão (Baioco *et al.*, 2023) e aumento da volatilidade no setor agrícola (Canal Rural, 2025).

Apesar do reconhecimento dos impactos da estiagem sobre a produtividade, ainda são escassas as ferramentas acessíveis que permitam ao produtor acompanhar, de forma contínua e em tempo oportuno, se a lavoura está sendo afetada por um déficit pluviométrico capaz de comprometer o desenvolvimento da cultura (Gonçalves *et al.*, 2019). Essa limitação tecnológica compromete a capacidade de resposta em campo e dificulta decisões estratégicas de manejo, especialmente durante os estágios mais sensíveis do ciclo da soja (APROSOJA- MS, 2022; Vicente e Martins, 2006; Mavi e Tupper, 2004).

Desse modo, a variabilidade climática e seus impactos na produção agrícola são temas amplamente estudados na economia agrícola e na administração rural. Ferreira *et al.* (2006) destacam que, ao longo da História Econômica, diversas teorias e modelos teóricos foram formulados para explicar e compreender os determinantes do crescimento de um produto em diferentes regiões.

No contexto agrícola, o progresso tecnológico possibilita a adoção de estratégias avançadas de monitoramento e gestão agrícola, como o uso de sistemas de integração de dados de múltiplas fontes (Golombieski e Vaz, 2024), previsão agrometeorológica e sensoriamento remoto (Richetti, 2018), modelagem climática (BRASIL, 2016) e técnicas com uso de inteligência artificial aplicadas à agricultura (Bandeira, 2022; Richetti, 2018).

Essas inovações permitem um acompanhamento mais preciso das condições climáticas e do solo, facilitando a tomada de decisões estratégicas para mitigar os efeitos adversos ao clima. No entanto, mesmo com elevados investimentos em tecnologias de manejo, como irrigação controlada, cultivares adaptadas à seca e práticas conservacionistas, a produção de soja ainda enfrenta desafios significativos diante de eventos climáticos extremos (Arsego *et al.*, 2018) com a ausência de uma proposta de monitoramento contínuo e de fácil acesso. A introdução de novas tecnologias digitais e analíticas, como modelos baseados em machine learning e plataformas integradas de dados climáticos, pode aprimorar a gestão dos riscos durante a safra.

## 1.1 PROBLEMÁTICA

A cultura da soja foi escolhida como objeto deste estudo, devido a sua importância histórica e econômica. Desde sua introdução, na década de 1960, passou a desempenhar um papel central no agronegócio local, especialmente na região da Grande Dourados, situada ao sul do estado de Mato Grosso do Sul (Michels *et al.*, 2004).

A expansão da cultura foi favorecida por condições climáticas adequadas e pela oferta de terras produtivas a custos acessíveis, fatores que estimularam a industrialização agrícola, impulsionaram a urbanização e criaram postos de trabalho (Michels *et al.*, 2004). O crescimento da produção é evidenciado pelos dados das safras entre 2012/2013 e 2023/2024, período em que a produção aumentou 104,2%, a área plantada cresceu 100,7% e a produtividade registrou um acréscimo de 1,8% (Casa Rural, 2024).

Silva *et al.* (1986) destacam em seu estudo que as condições climáticas exercem influência direta sobre o desempenho agrícola, podendo tanto favorecer o aumento do volume e da qualidade das safras em situações adequadas quanto comprometer a oferta de produtos quando adversas, com impactos significativos na economia. Especificamente, nos anos de 2021/2022 e 2022/2023, a influência das condições climáticas tem se mostrado determinante para os resultados das safras em Mato Grosso do Sul, evidenciando a vulnerabilidade da produção agrícola frente às variações do clima.

No ano-safra 2021/2022, a região sul, responsável por 62,4% da área plantada do estado e considerada a principal produtora de grãos, enfrentou uma das piores safras em uma década (APROSOJA-MS, 2022). A estiagem comprometeu significativamente a produção, resultando em uma produtividade média ponderada reduzida em Mato Grosso do Sul. Municípios tradicionalmente fortes na produção, como Maracaju, Sidrolândia, Ponta Porã, Dourados e Rio Brillhante, apresentaram produtividades abaixo de 46,78 sacas por hectare, afetando negativamente a média estadual, já que juntos representam 34% do peso na ponderação (APROSOJA-MS, 2022). No total, 30 dos 77 municípios superaram a média estadual, mas as condições climáticas adversas, tais como a estiagem e elevadas temperaturas, resultaram em uma queda de aproximadamente 38,49% na produtividade em relação à safra anterior, o que levou a uma redução de 34,68% no volume total de grãos (APROSOJA-MS, 2022).

Em contraste, o ano-safra 2022/2023 apresentou recuperação significativa. A regularidade das chuvas em momentos decisivos do ciclo fenológico contribuiu para um novo recorde de produção no estado (APROSOJA-MS, 2023). As principais regiões produtoras, sul e centro, responsáveis por 85,7% da produção estadual, mantiveram bons níveis de produtividade. Municípios como Maracaju, Ponta Porã, Sidrolândia, Dourados e Rio Brillhante registraram produtividades entre 54,43 e 70,44 sacas por hectare (APROSOJA-MS, 2023). O desempenho positivo também se estendeu ao norte do estado, com municípios como Alcinópolis, São Gabriel do Oeste, Chapadão do Sul e Costa Rica apresentando resultados expressivos, consolidando a recuperação da produção em Mato Grosso do Sul (APROSOJA-MS, 2023).

No ponto de vista econômico, essa relação entre clima e produção agrícola é evidenciada por Arias *et al.* (2015), que apontam perdas anuais médias de R\$ 11 bilhões — aproximadamente 1% do PIB agrícola brasileiro — atribuídas a fatores climáticos. Gonçalves *et al.* (2019) complementam essa análise ao registrar que determinadas perdas agrícolas ocasionadas por condições climáticas extremas resultaram em prejuízos expressivos para o tesouro nacional, corroborando com a vulnerabilidade do setor agrícola frente às variações climáticas. Em um estudo voltado para Mato Grosso do Sul, Casonato (2013) aponta que as quedas no Produto Interno Bruto (PIB) do estado nos anos de 2001 e 2002 foram diretamente influenciadas por condições climáticas adversas relacionadas à falta de chuvas que impactaram o setor agrícola.

Nesse contexto, Vicente e Martins (2006) enfatizam a importância de compreender os efeitos do clima sobre a produção agrícola, destacando que o isolamento desses impactos em relação a outros fatores determinantes da produtividade é fundamental para gerar estimativas mais precisas sobre o volume de produção. Segundo os autores, esse conhecimento pode orientar decisões estratégicas e embasar políticas públicas voltadas à mitigação dos efeitos econômicos provocados por fenômenos climáticos.

Complementando essa análise, Mavi e Tupper (2004) ressaltam que a informação climática pode ser uma componente estratégica nas ações de planejamento dos cultivos. De acordo com os autores, o acesso a dados climáticos auxilia nas decisões relacionadas às práticas agrícolas, contribuindo para a redução dos riscos associados às variações climáticas. Essa integração entre informação climática e tomada de decisão fortalece a resiliência dos sistemas agrícolas, preparando-os melhor

para enfrentar condições adversas que possam comprometer a produção (Mavi e Tuper, 2004).

Os estudos e relatórios de acompanhamento de safra reconhecem a influência das condições climáticas na produção agrícola e os impactos da estiagem sobre as lavouras de soja, mas não avançam em estratégias específicas que auxiliem o produtor a mitigar esses riscos. Além disso, não fornecem soluções tecnológicas que apoiem a tomada de decisão durante o ciclo de produção, o que limita a capacidade do produtor de identificar o momento em que a estiagem começa a comprometer a cultura. A ausência de informações em tempo oportuno dificulta decisões estratégicas, como o replantio ou a adoção de práticas corretivas, ampliando os riscos envolvidos na gestão da lavoura.

Diante desse cenário, a utilização de técnicas computacionais modernas, como *machine learning*, para classificar e prever eventos de estiagem, com base em dados de estações meteorológicas, apresenta-se como uma solução viável e eficiente para o monitoramento contínuo de períodos críticos que podem impactar a cultura da soja. Diferente de métodos mais complexos e onerosos, como o uso de imagens de satélite, que demandam altos custos operacionais e infraestrutura avançada (Mahmood *et al.*, 2019; Hu *et al.*, 2019), a proposta baseada em dados de precipitação local oferece uma alternativa de menor custo, simplicidade de implementação e fácil acesso.

Levantamentos indicam que 72% dos 447 produtores consultados em Mato Grosso do Sul afirmaram não utilizar softwares agrícolas (APROSOJA-MS, 2023), evidenciando a necessidade de soluções práticas, intuitivas e de mais simples compreensão. Considerando que muitos produtores já realizam medições básicas de precipitação por necessidade de acompanhamento, esta proposta se adapta à realidade do campo e oferece um método acessível para o monitoramento da estiagem.

Deste modo, este trabalho parte da hipótese de que a definição de um indicador de estiagem específico para a cultura da soja, combinando os conceitos de estiagem e veranico na estação chuvosa, permite uma caracterização mais precisa desse fenômeno. A aplicação de um modelo de *machine learning* para a classificação dessa estiagem possibilita o monitoramento contínuo da precipitação, identificando rapidamente períodos críticos que possam comprometer o desenvolvimento da cultura.

Sendo assim, a pergunta que norteia este trabalho é: como a construção de um modelo de *machine learning*, fundamentado em um indicador específico de estiagem para a sojicultura, pode contribuir para a classificação da estiagem, o monitoramento

climático contínuo e a redução dos riscos produtivos na região de Dourados/MS?

Responder a essa pergunta é necessário diante do papel da estiagem na redução da produtividade da soja, com efeitos diretos sobre os produtores e sobre a atividade agrícola. A ausência de métodos acessíveis e sistemáticos para identificar e acompanhar, em tempo adequado, os períodos de déficit pluviométrico dificulta a tomada de decisão no campo e amplia as incertezas no planejamento das safras.

Nesse contexto, compreender de que forma um modelo baseado em *machine learning*, fundamentado em um indicador específico de estiagem, pode contribuir para a classificação e o monitoramento contínuo é relevante para o desenvolvimento de ferramentas que auxiliem a adaptação da produção às variações climáticas. A aplicação da inteligência artificial nesse processo permite acompanhar o comportamento da precipitação, apoiar decisões relacionadas ao manejo da lavoura e promover maior controle sobre os impactos da variabilidade climática. Essa abordagem está alinhada às exigências do setor agrícola quanto ao uso de dados e tecnologias digitais.

## 1.2 JUSTIFICATIVA

O diferencial deste trabalho está na integração de diferentes áreas do conhecimento, superando a abordagem isolada comum em estudos acadêmicos. A partir do Modelo de Solow, o estudo demonstra como o progresso tecnológico, viabilizado por técnicas de *machine learning*, pode otimizar a eficiência produtiva por meio de um método de monitoramento contínuo. Simultaneamente, a agrometeorologia analisa os impactos das condições climáticas, como a estiagem, sobre a soja, fornecendo subsídios para interpretar como as variações pluviométricas afetam diretamente o desenvolvimento da cultura.

Nesse contexto, a tese apresenta inovações em três dimensões complementares. Do ponto de vista teórico, propõe-se uma abordagem interdisciplinar ao integrar a teoria do crescimento econômico com os fundamentos da gestão agrícola orientada por dados. A revisão de literatura sobre o crescimento econômico relacionado com tecnologia e clima é utilizada como base conceitual para refletir sobre o papel do progresso tecnológico na adaptação da produção agrícola às condições climáticas.

Em termos metodológicos, a pesquisa propõe um novo critério para a identificação de estiagem na sojicultura, construído a partir da combinação dos conceitos de estiagem e veranico, aplicados sobre janelas móveis de precipitação. Esse critério é operacionalizado por meio de técnicas de classificação automatizada, permitindo o monitoramento contínuo de déficits pluviométricos ao longo do ciclo produtivo.

A inovação empírica se materializa na aplicação do modelo a dados reais de precipitação do município de Dourados/MS, permitindo a validação do conceito proposto em diferentes safras agrícolas. A construção de um indicador específico para a soja e a utilização de dados locais reforçam a aplicabilidade da proposta, com possibilidade de adaptação para outros contextos produtivos.

Por fim, essa convergência permite uma análise integrada, com potencial de aplicação prática no apoio à tomada de decisão no campo e na redução dos riscos produtivos e econômicos associados à estiagem.

### 1.3 OBJETIVOS

O objetivo geral é classificar a ocorrência de estiagens na cultura da soja no município de Dourados (MS), com o intuito de aprimorar o monitoramento climático e oferecer subsídios aos stakeholders do setor agrícola, contribuindo para a tomada de decisão e a redução dos riscos associados à variabilidade pluviométrica.

Objetivos Específicos:

1. Definir um indicador de estiagem específico para a cultura da soja em Dourados/MS, combinando os conceitos de estiagem e veranico e considerando limites críticos de precipitação na estação chuvosa.
2. Examinar a distribuição da precipitação ao longo do ciclo produtivo da soja, identificando padrões históricos de estiagem e suas possíveis implicações na produtividade.
3. Desenvolver um modelo de machine learning para a classificação contínua da estiagem, com base no indicador proposto e utilizando variáveis climáticas selecionadas.

## 1.4 ORGANIZAÇÃO DA TESE

Esta tese está estruturada em cinco seções principais. A Seção 1, Introdução, apresenta o tema do estudo, a delimitação da problemática, os objetivos, a justificativa e a organização do trabalho. A Seção 2, Referencial Teórico, discute a Teoria do Crescimento Econômico e sua relação com a tecnologia e o clima, a variabilidade pluviométrica, o uso de tecnologias no monitoramento de estiagens, as exigências climáticas da soja e a integração desses conceitos na proposta do estudo.

A Seção 3, Procedimentos Metodológicos, descreve a natureza da pesquisa, a área de estudo, as bases de dados utilizadas, a hipótese formulada e os métodos de análise adotados, incluindo as técnicas estatísticas e computacionais aplicadas. A Seção 4, Resultados e Discussões, apresenta a análise da produtividade da soja, o comportamento da precipitação durante a safra 2021/2022, a definição do indicador de estiagem, a aplicação e validação dos modelos de *machine learning* e a análise contrafactual.

Por fim, a Seção 5, Considerações Finais, sintetiza as principais contribuições do estudo e aponta possibilidades de aplicação e desdobramentos futuros do modelo desenvolvido.

## 2. FUNDAMENTAÇÃO TEÓRICA

A variabilidade climática tem imposto desafios crescentes à produtividade agrícola, exigindo referenciais teóricos e metodológicos capazes de compreender e mitigar seus efeitos. Esta seção apresenta os fundamentos que orientam este estudo: a evolução dos estudos de crescimento econômico e sua conexão com a tecnologia e o clima; os sistemas atmosféricos que determinam a variabilidade pluviométrica em Mato Grosso do Sul; as exigências climáticas da soja e seus períodos críticos de vulnerabilidade hídrica; e, por fim, o papel das tecnologias no monitoramento de secas e estiagens. Esses elementos formam a base conceitual para o modelo de classificação de estiagem desenvolvido nos capítulos seguintes.

### 2.1 EVOLUÇÃO DOS ESTUDOS DE CRESCIMENTO ECONÔMICO E SUA CONEXÃO COM A TECNOLOGIA E O CLIMA

A relação entre crescimento econômico e tecnologia constitui um eixo central do pensamento econômico desde a formação da economia política clássica. Desde os primeiros estudos, a preocupação com os determinantes da riqueza das nações esteve intimamente associada à capacidade de elevar a produtividade e de responder aos limites impostos pela disponibilidade de recursos.

Neste sentido, Adam Smith destacou a divisão do trabalho como um mecanismo fundamental para elevar a produtividade, ao demonstrar que a especialização permitia ganhos crescentes de eficiência e ampliava a capacidade produtiva das nações (Smith, 1776).

Thomas Malthus, em uma perspectiva distinta, chamou atenção para os limites populacionais, defendendo que a taxa de crescimento da população tenderia a superar a capacidade de produção de alimentos, impondo barreiras estruturais à expansão da renda *per capita* (Malthus, 1798).

David Ricardo, por sua vez, formulou o princípio dos rendimentos decrescentes da terra e destacou o papel das vantagens comparativas no comércio internacional, argumentando que o crescimento econômico não poderia ser compreendido de forma isolada, mas sim como resultado da interação entre as limitações impostas pelos recursos naturais e a inserção das nações em um sistema de trocas globais (Ricardo, 1817).

No século XIX, Karl Marx reinterpretou a questão do crescimento a partir de uma abordagem crítica, argumentando que a tecnologia e o desenvolvimento das forças produtivas estavam indissociavelmente ligados à dinâmica da acumulação capitalista e às contradições

entre capital e trabalho (Marx, 1867). Para o autor, o avanço técnico não era apenas um fator de eficiência, mas um instrumento de ampliação das desigualdades sociais, de intensificação da exploração e de transformação estrutural das relações de produção.

Na mesma época, mas em um registro teórico distinto, Joseph Schumpeter deslocou a discussão ao enfatizar a inovação como força endógena e disruptiva no processo capitalista. Seu conceito de “destruição criativa” (Schumpeter, 1911; 1942) ressaltava que o capitalismo se desenvolve por meio de ciclos de inovação nos quais novas tecnologias substituem processos anteriores, remodelando continuamente a estrutura produtiva e gerando fases de expansão e crise.

Esses autores, ainda que partindo de premissas distintas, revelaram uma intuição central: o crescimento econômico não pode ser compreendido apenas pela ótica quantitativa da acumulação de capital e trabalho, mas exige considerar a qualidade das transformações tecnológicas e seus impactos sobre a produtividade, a distribuição e a sustentabilidade do sistema econômico.

A formalização matemática do crescimento econômico ocorreu a partir da primeira metade do século XX, com as contribuições de Roy Harrod (1939) e Evsey Domar (1946). Ambos formalizaram modelos de crescimento baseados na relação entre poupança, investimento e produto. Esses modelos, conhecidos como *Harrod-Domar*, explicavam a dinâmica do crescimento econômico a partir da taxa de poupança e da produtividade do capital, mas apresentavam problemas de instabilidade, pois pequenas variações nas taxas de investimento ou crescimento populacional poderiam gerar trajetórias explosivas ou recessivas de renda, comprometendo a explicação do crescimento sustentado (Harrod, 1939; Domar, 1946).

Finalmente, Moses Abramovitz (1956) contribuiu com uma análise empírica sobre o crescimento econômico norte-americano entre 1870 e 1946, demonstrando que grande parte da expansão do produto *per capita* não poderia ser explicada apenas pelo aumento do capital e do trabalho. Esse componente residual, chamado pelo autor de “medida de nossa ignorância”, indicava que fatores pouco compreendidos, especialmente o progresso técnico, exerciam papel central no crescimento econômico (Abramovitz, 1956). Essa lacuna impulsionou novos estudos voltados a identificar os elementos responsáveis por esse crescimento residual. Foi nesse contexto que Robert M. Solow desenvolveu seu modelo neoclássico de crescimento econômico, oferecendo uma explicação estruturada para o fenômeno observado por Abramovitz (1956).

Solow (1956) introduziu o conceito de progresso técnico exógeno como elemento-

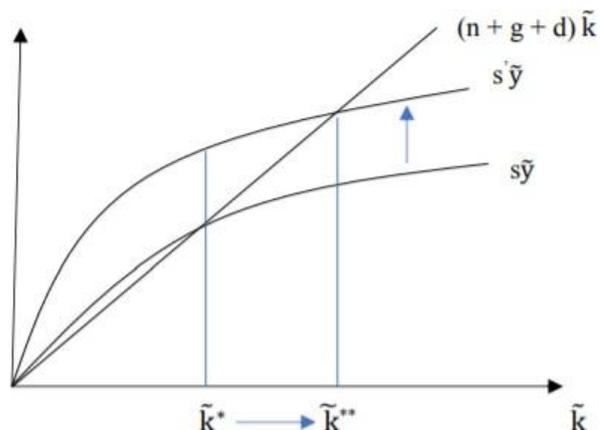
chave para explicar o aumento da produtividade não capturado pelo simples acúmulo de capital e trabalho. Seu modelo trouxe uma nova perspectiva ao debate econômico ao deslocar o foco dos fatores de demanda, predominantes nos modelos keynesianos, como os de Harrod (1939) e Domar (1946), para os fatores de oferta, enfatizando o papel da tecnologia como motor do crescimento sustentado (Solow, 1956).

Outra abordagem relevante para a análise da produtividade introduzida por Solow foi o uso de uma função de produção explícita dependente do tempo, apresentada em seus trabalhos de 1956 e 1957. Nessa perspectiva, a produtividade atua como um fator multiplicador na função de produção tradicional e é estimada como um resíduo, conhecido como Produtividade Total dos Fatores (PTF) (Solow, 1957). O Modelo de Solow trouxe contribuições fundamentais aos estudos sobre crescimento econômico (Carvalho, 2001), ao explicar a variação do produto per capita com base em fatores como poupança, acumulação de capital, crescimento demográfico e, principalmente, o progresso técnico exógeno.

Solow (1956) propôs um modelo teórico simplificado para analisar o crescimento econômico de longo prazo, partindo de algumas premissas fundamentais. Considera-se que os países produzem e consomem um único bem, representado em unidades do Produto Interno Bruto (PIB), sem comércio internacional e com tecnologia tratada como um fator exógeno ao sistema econômico (Solow, 1956). Esse modelo visa explicar o crescimento sustentado da economia ao longo do tempo, considerando a interação entre capital, trabalho e progresso técnico.

A Figura 1 está ilustrado a proposta do Modelo de Solow de forma gráfica. O avanço técnico provoca um deslocamento ascendente da curva de investimento bruto  $s_y$ , elevando os níveis de capital e produto *per capita*. Esse aumento impulsiona a economia a atingir um novo estado estacionário, caracterizado pela interseção entre a curva de investimento  $s_y$  e a reta, que representa o investimento necessário para manter o capital por trabalhador constante. Nesse novo equilíbrio, a economia opera em um patamar superior de produção *per capita*, com crescimento contínuo impulsionado pelo progresso técnico.

Figura 1 - Representação gráfica do crescimento econômico no longo prazo.



Fonte: Adaptado de Jones e Vollrath (2014).

Solow (1956) conclui que a inclusão do progresso técnico no modelo básico permite compensar os efeitos dos retornos decrescentes do capital, garantindo um crescimento sustentado ao longo do tempo e que sem essa adição, o modelo indicaria uma estagnação econômica após o alcance do estado estacionário. Assim, o avanço tecnológico é essencial para assegurar o crescimento econômico contínuo.

No ano seguinte, Solow (1957) analisou empiricamente o crescimento econômico e decompôs o crescimento do produto em suas fontes principais: capital, trabalho e progresso técnico. Neste artigo o autor demonstrou que a maior parte do crescimento econômico não poderia ser explicada apenas pelo aumento de capital e trabalho. O componente residual, frequentemente interpretado como progresso técnico, destacou o papel essencial da tecnologia como motor do crescimento econômico sustentado (Solow, 1957).

Apesar de sua simplicidade e da suposição de exogeneidade do progresso técnico, o Modelo de Solow forneceu uma visão geral sobre a dinâmica do crescimento econômico e a importância da inovação, bem como foi base para inúmeras pesquisas e debates, pavimentando o caminho para o desenvolvimento de teorias mais complexas que buscariam endogeneizar o progresso tecnológico. Nessa nova literatura buscou responder à questão deixada em aberto: de onde vem a tecnologia?

Com isso, a partir da década de 1980, a crescente evidência de que a inovação era um processo endógeno, surgiram os modelos de crescimento endógeno. Essa nova vertente da teoria econômica buscou explicar a inovação e o avanço tecnológico como resultados de decisões econômicas e investimentos deliberados dentro do próprio sistema econômico, com destaque para as contribuições de Paul Romer e Robert Lucas Jr.

Paul Romer, em seus trabalhos de 1986 e 1990, propôs que o conhecimento deveria ser tratado como um bem econômico que permite a existência de retornos crescentes e de efeitos de transbordamento que beneficiam toda a economia (Romer, 1986; Romer, 1990). Nesse modelo, a inovação surge como resultado de decisões deliberadas de investimento em pesquisa e desenvolvimento (P&D), incentivadas por instituições como o sistema de patentes e por políticas de apoio à inovação.

Robert Lucas, em paralelo, desenvolveu uma abordagem centrada no papel do capital humano como determinante do crescimento. Para o autor, a acumulação de conhecimento e habilidades gera externalidades positivas que aumentam a produtividade de toda a sociedade, explicando as diferenças persistentes entre países no longo prazo (Lucas, 1988; Lucas, 1993). Enquanto Paul Romer enfatiza o papel da inovação tecnológica e da P&D, Robert Lucas destaca a importância da educação, do aprendizado contínuo e da formação de competências.

Essa virada teórica representou uma mudança de paradigma de forma que a tecnologia deixou de ser considerada um fator externo e passou a ser explicada internamente, como consequência de decisões econômicas, sociais e políticas. Além disso, abriu caminho para a discussão sobre o papel das instituições, dos incentivos e da governança como elementos centrais para sustentar trajetórias de crescimento.

A partir do final do século XX e início do século XXI, a crescente evidência científica sobre as mudanças climáticas e a degradação ambiental impôs um novo desafio à teoria econômica. Os modelos de crescimento existentes, mesmo os endógenos, não incorporavam explicitamente os recursos naturais, os limites ambientais nem os efeitos das mudanças climáticas sobre a atividade produtiva, o que reduz sua capacidade explicativa em contextos onde essas variáveis têm papel relevante (Smulders, 1995; Stern, 1997). Essa lacuna levou ao surgimento e à consolidação da economia do clima e da economia ecológica, que buscam integrar explicitamente as variáveis ambientais e climáticas na análise do crescimento econômico.

Nesses modelos, a tecnologia assume um papel dual tanto como um vetor de impactos ambientais negativos (através de emissões e uso intensivo de recursos) quanto como a principal ferramenta para a mitigação desses impactos e para a adaptação a um cenário de mudanças climáticas. Autores como William Nordhaus, Kenneth Arrow e Partha Dasgupta foram pioneiros nessa nova fronteira da área de pesquisa.

Nordhaus (1994) desenvolveu os Modelos de Avaliação Integrada (Integrated Assessment Models – IAMs), que conectam equações econômicas a variáveis climáticas para estimar os custos do aquecimento global e definir trajetórias ótimas de mitigação.

Posteriormente, Nordhaus e Boyer (2000) expandiram esse arcabouço ao evidenciar que a política climática deve ser compreendida, em essência, como uma decisão econômica de natureza intertemporal, na qual se equaciona o dilema entre a busca por crescimento imediato e a preservação das condições ambientais para o futuro. Os autores aprofundaram a análise ao demonstrar que diferentes trajetórias de emissões de gases de efeito estufa resultam em custos econômicos distintos, o que torna a escolha das estratégias de mitigação e adaptação um elemento central na definição de políticas de desenvolvimento sustentável.

Paralelamente, a teoria do capital natural e a economia ecológica enfatizam a interdependência entre a economia e os ecossistemas naturais. Autores como Kenneth Arrow e Partha Dasgupta têm sido fundamentais nessa discussão.

Arrow *et al.* (1995) introduziram o conceito de capital natural, defendendo que os recursos ambientais devem ser considerados como ativos sujeitos à depreciação, da mesma forma que o capital físico e que ignorar esse aspecto leva a uma superestimação da riqueza de uma nação. Já Dasgupta (2021) reforçou essa visão ao propor que a prosperidade deve ser medida a partir da soma do capital físico, humano e natural, incluindo biodiversidade e serviços ecossistêmicos. Em ambos os trabalhos, a tecnologia é vista como uma ferramenta essencial para aumentar a eficiência no uso dos recursos, desenvolver substitutos para recursos escassos e restaurar ecossistemas degradados.

A inovação tecnológica é, portanto, um componente chave para a transição de uma economia mais circular e sustentável, que valorize e preserve o capital natural (Arrow *et al.*, 1995; Dasgupta, 2021). Esses modelos mais recentes expandem a compreensão da relação entre economia e tecnologia, incorporando a dimensão ambiental como um fator crítico para o crescimento de longo prazo.

Assim, a tecnologia tanto pode intensificar a exploração de recursos e aumentar as pressões ambientais, como também pode oferecer soluções para a economia como na eficiência no uso dos insumos e a restauração de ecossistemas degradados. Logo, tecnologias limpas, energias renováveis e sistemas digitais de monitoramento tornam-se centrais não apenas para a mitigação das mudanças climáticas, mas também para a redefinição do próprio conceito de crescimento econômico.

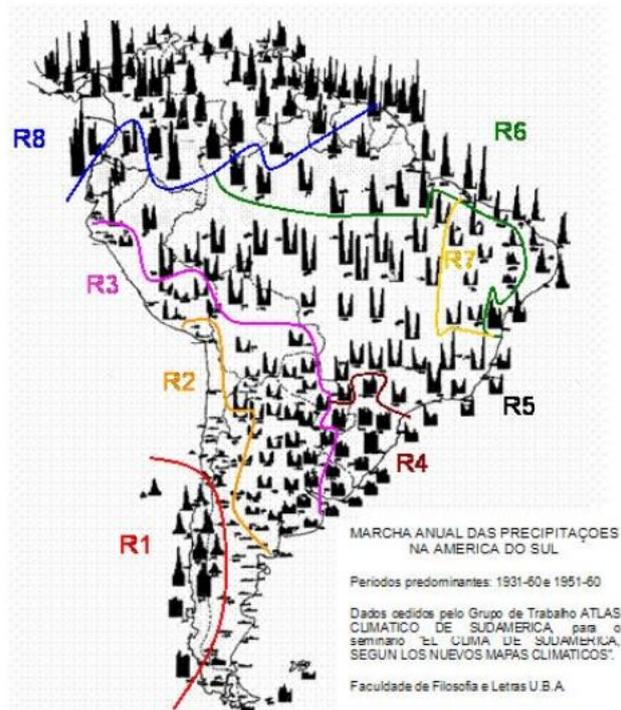
## 2.2 DINÂMICA ATMOSFÉRICA E A VARIABILIDADE PLUVIOMÉTRICA

Nesta seção é abordado o conceito de clima, os sistemas atmosféricos que influenciam o verão em Mato Grosso do Sul e os efeitos da mudança nos padrões climáticos. Além disso, é enfatizado que a estiagem como principal fator de queda na produtividade agrícola e apresenta suas definições e classificações.

O clima é convencionalmente caracterizado como uma descrição estatística das médias aritméticas das variáveis atmosféricas de superfície — incluindo temperatura, precipitação, radiação solar, vento e umidade — calculadas ao longo de um período padrão de trinta anos, refletindo o comportamento atmosférico médio esperado em determinada localidade (WMO, 2020). Em Mato Grosso do Sul, o clima durante o verão é influenciado por um regime de chuvas bem estabelecido, condicionado por sistemas atmosféricos predominantes no Hemisfério Sul.

Segundo Reboita *et al.* (2010), a América do Sul foi segmentada em oito setores climáticos devido a sua grande extensão territorial, sendo Mato Grosso do Sul incluído no setor R5, conforme figura 2.

Figura 2 - Mapa regionalizado dos diferentes ciclos anuais de precipitação na América do Sul.



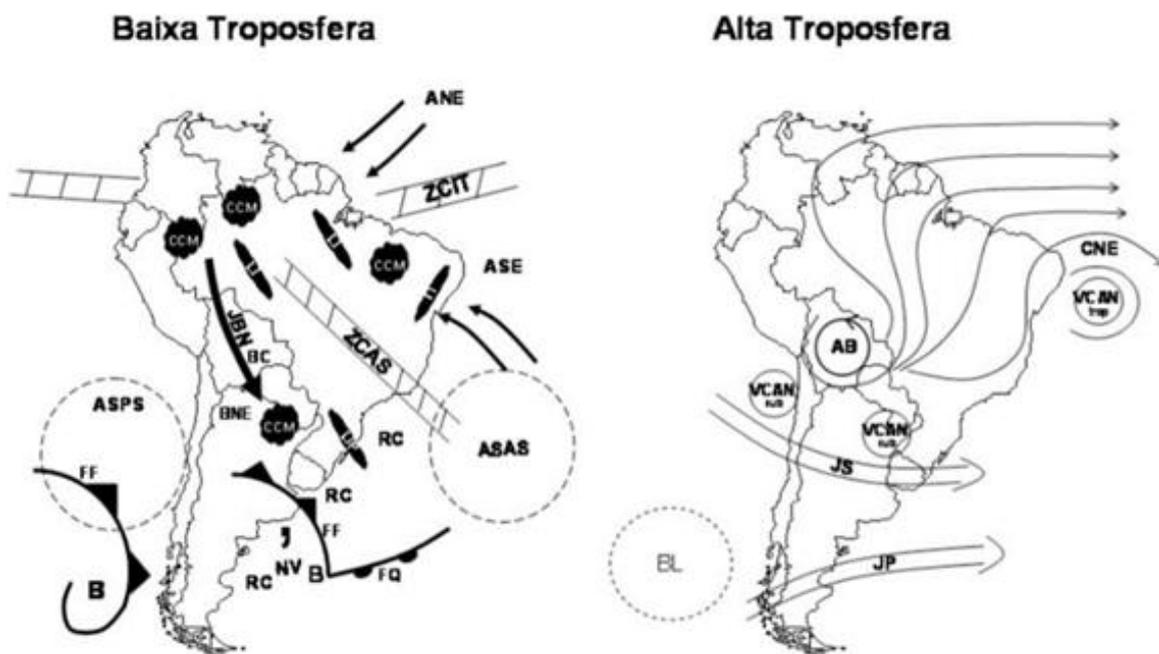
Fonte: Reboita *et al.* (2010)

O setor R5 abrange desde o noroeste até o sudeste brasileiro, incluindo o Equador e o norte do Peru, apresentando características típicas de um sistema de monção, com volumes pluviométricos elevados durante o verão e precipitações reduzidas no inverno. O aumento da radiação solar no Hemisfério Sul, principalmente durante o verão, intensifica a convecção atmosférica sobre o continente (Carvalho, 2011), criando condições favoráveis à formação de chuvas frequentes em Mato Grosso do Sul.

A combinação de diversos sistemas atmosféricos no verão exerce uma forte influência sobre o regime de chuvas do estado. Entre esses sistemas destacam-se os Vórtices Ciclônicos em Altos Níveis (VCAN), a Alta da Bolívia (AB) e os Jatos Subtropical e Polar (JS e JP), que se alinham devido ao solstício de verão (Reboita *et al.*, 2010). A atividade convectiva intensa na região amazônica, potencializada pelo aquecimento abaixo dos 500 hPa, facilita a formação da Alta da Bolívia — um anticiclone que favorece a transferência de calor da superfície para a atmosfera, resultando em instabilidades atmosféricas sobre as regiões Centro-Oeste e Sudeste do Brasil (Reboita *et al.*, 2010).

A posição geográfica estratégica de Mato Grosso do Sul, localizada no cruzamento de importantes sistemas atmosféricos sul-americanos (Zavattini, 2009), contribui para a definição da estação chuvosa, que geralmente se estende de setembro a março. Na Figura 3 é ilustrado esses sistemas climáticos descritos por Reboita *et al.* (2010).

Figura 3 - Esquema ilustrativo da localização dos sistemas climáticos na América do Sul durante o período chuvoso na baixa e na alta troposfera.



Fonte: Reboita et al. (2010).

Contudo, esse padrão climático tem sofrido alterações nas últimas décadas, afetando diretamente o regime de chuvas e gerando impactos no setor agrícola (APROSOJA-MS, 2023; APROSOJA-MS, 2022; Arias *et al.*, 2015; Casonato, 2013). A variabilidade climática, intensificada pelo processo de industrialização desde o século XVIII, provocou mudanças nos padrões atmosféricos tradicionais (Almeida, 2016).

O aumento nas emissões de gases de efeito estufa, impulsionado pelo uso excessivo de combustíveis fósseis, elevou as temperaturas médias globais e intensificou a frequência de eventos climáticos extremos (Almeida, 2016).

O Painel Intergovernamental sobre Mudanças Climáticas (IPCC, 2023) aponta que a agricultura responde por cerca de 22% das emissões globais de gases de efeito estufa, posicionando-se como um dos setores mais expostos aos impactos das mudanças climáticas. Essas transformações comprometem diretamente a produtividade agrícola e aumentam a frequência e severidade de eventos climáticos extremos.

Diante desse cenário, o IPCC (2023) destaca a urgência de implementar estratégias de adaptação para reduzir esses efeitos e fortalecer a resiliência do setor agrícola. No contexto de Mato Grosso do Sul, essas mudanças têm tornado o regime de chuvas mais irregular e imprevisível, impactando negativamente o desenvolvimento agrícola e elevando o risco de perdas produtivas (APROSOJA-MS, 2022).

Em destaque, a estiagem desponta como a principal causa da queda de produtividade nas lavouras de verão (Lennon, 2025; Rodrigues, 2025; Agência Estado, 2025). Neste sentido, o conceito de estiagem ainda não possui uma definição única e amplamente consensual nas áreas de climatologia e meteorologia, diferentemente do termo "seca", que apresenta um uso mais consolidado em diversas línguas (Freitas e Oliveira, 2018). No âmbito agrícola nacional, o Agritempo, Sistema de Monitoramento Agrometeorológico desenvolvido pela Embrapa Agricultura Digital em parceria com o Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura (Cepagri) da Unicamp, disponibiliza informações detalhadas sobre as condições meteorológicas e agrometeorológicas para diversos municípios brasileiros (Agritempo, 2025b). Integrado ao Zoneamento Agrícola de Risco Climático (Zarc), o sistema subsidia programas como o Garantia da Atividade Agropecuária (Proagro) e o Programa de Subvenção ao Prêmio do Seguro Rural (PSR), fornecendo boletins, gráficos e mapas para o planejamento agrícola (Agritempo, 2025b).

Seu uso é amplamente difundido entre produtores rurais, gestores públicos e pesquisadores, que acessam dados sobre estiagem agrícola, precipitação acumulada, necessidade de irrigação, manejo do solo, entre outros indicadores (Agritempo, 2025b).

Sendo assim, o Agritempo define estiagem como:

“Período sem ocorrência de chuvas, ou seja, com precipitação pluviométrica igual a zero, durante o qual pode haver rebaixamento de água de rios e cursos d’água.” (AGRITEMPO, 2025a).

Ainda segundo o Agritempo, a estiagem agrícola é definida como

“Período com precipitação pluviométrica inferior a 10 mm.” (AGRITEMPO, 2025a).

Por fim, o sistema classifica o veranico como sendo:

“Estiagem durante a estação chuvosa, com dias de calor intenso e insolação. Para ser considerado veranico, é necessário uma duração mínima de quatro dias.” (AGRITEMPO, 2025a)

Para complementar, o Instituto Nacional de Meteorologia não estabelece definições para estiagem, estiagem agrícola ou veranico, limitando-se a definir apenas o conceito de seca da seguinte forma:

“Clima excessivamente seco numa região específica. Deve ser suficientemente prolongado para que a falta de água cause sério desequilíbrio hidrológico” (INMET, 2025).

Essas definições destacam a complexidade dos fenômenos climáticos que influenciam diretamente o setor agrícola. A diferenciação entre seca, estiagem, estiagem agrícola e veranico permite compreender os distintos níveis de impacto sobre o solo, os recursos hídricos e o ciclo de desenvolvimento das culturas. No ambiente agrícola, os produtores, por meio da observação prática, já identificam a ocorrência do veranico e frequentemente o consideram em conjunto com os períodos de estiagem ou de seca ao avaliar os riscos potenciais às lavouras, reconhecendo sua influência direta sobre o rendimento e a qualidade das colheitas.

### 2.3 A TECNOLOGIA NO MONITORAMENTO DE SECAS E ESTIAGENS

O monitoramento de secas e estiagens se tornou cada vez mais relevante diante da crescente variabilidade climática e de seus impactos diretos sobre a agricultura. Embora índices tradicionais baseados em séries históricas de precipitação continuem sendo amplamente utilizados, sua eficácia depende da correta definição das escalas temporais e

espaciais de análise. Nesse contexto, apresenta-se a seguir um breve panorama dos estudos sobre seca e estiagens que utilizam tecnologias desde as tradicionais as mais sofisticadas usando inteligência artificial, acompanhado do detalhamento temporal e espacial desses trabalhos, conforme exposto no Quadro 1.

Blain e Brunini (2007) e Gonçalves *et al.* (2021) destacam que a seca, caracterizada por um déficit de precipitação em relação à média histórica, pode ser monitorada por meio de índices específicos que quantificam sua intensidade, frequência e duração. Esses índices, no entanto, apresentam resultados confiáveis apenas quando aplicados em escalas temporais apropriadas para cada tipo de seca. Ambos os estudos convergem na análise detalhada dos métodos e índices utilizados para o monitoramento de secas, enfatizando a importância da escolha adequada de escalas temporais e espaciais para garantir a eficácia dessas ferramentas.

Blain e Brunini (2007) enfatizam que o monitoramento da seca meteorológica deve ser feito em escalas mensais ou superiores devido à alta variabilidade das chuvas em períodos mais curtos, o que pode comprometer a precisão dos dados. Já a seca agrícola, diretamente relacionada à evapotranspiração real das culturas, exige avaliações em escalas menores, como o decêndio, permitindo identificar rapidamente deficiências hídricas que possam afetar o desenvolvimento das plantas. A escolha incorreta dessas escalas pode resultar em análises imprecisas, impactando negativamente decisões sobre manejo agrícola e políticas de mitigação.

Complementando essa perspectiva, Gonçalves *et al.* (2021) explora uma ampla gama de índices utilizados mundialmente no monitoramento de secas, como o Índice de Precipitação Padronizado (SPI) e o Índice de Umidade da Cultura (CMI), e discute suas aplicações em diferentes contextos.

O texto também aborda sistemas de monitoramento consolidados, como o Monitor de Secas dos Estados Unidos e o Monitor de Secas do Brasil, que integram dados climáticos, hidrológicos e de sensoriamento remoto para fornecer informações sobre a evolução e os impactos das secas. Essas plataformas permitem o planejamento de ações emergenciais e a mitigação dos efeitos adversos em setores como agricultura, abastecimento hídrico e energia. Ambos os estudos ressaltam a importância de sistemas de alerta precoce e que a utilização adequada dos índices é fundamental para aprimorar o monitoramento, prever eventos críticos e apoiar decisões estratégicas (Gonçalves *et al.*, 2021).

Para realizar esse monitoramento, alerta e previsão, novas técnicas foram desenvolvidas ao longo do tempo para mitigar riscos envolvendo déficits pluviométricos. Diversos estudos exploram diferentes metodologias baseadas em aprendizado de máquina

para o monitoramento e previsão de secas e estiagens, utilizando uma variedade de dados climáticos e hidrológicos.

Ganguli e Reddy (2014) utilizaram uma abordagem baseada em SVM (*Support Vector Machine*) para realizar previsões de secas regionais, integrando múltiplos *inputs* climáticos para gerar previsões mais precisas na Índia com dados mensais entre 1896 a 2005.

Feng *et al.* (2020) focaram na previsão de secas meteorológicas durante a estação de crescimento do trigo na Austrália, utilizando grandes padrões climáticos como entradas em modelos preditivos, evidenciando como teleconexões climáticas podem ser utilizadas para previsão de eventos locais.

Bouaziz *et al.* (2021) desenvolveram um modelo que combina dados de precipitação por sensoriamento remoto com o Índice Padronizado de Precipitação (SPI) para rastrear e prever secas em regiões áridas, destacando a capacidade do modelo em operar em áreas com escassez de dados meteorológicos tradicionais.

Carvalho *et al.* (2021) aplicaram técnicas de aprendizado de máquina no contexto brasileiro para detectar eventos de seca e visualizar padrões climáticos, proporcionando uma ferramenta útil para gestores de recursos hídricos. O trabalho propõe um modelo de classificação binária capaz de identificar anos secos e não secos com base em dados globais de temperatura da superfície terrestre e do mar e para isso usaram máquinas de vetores de suporte (SVM).

Dikshit *et al.* (2021) aprimoraram o modelo SPEI (Índice de Precipitação Evapotranspiração Padronizado) utilizando redes neurais LSTM (*Long Short-Term Memory*), capazes de capturar dependências temporais de longo prazo, melhorando significativamente a precisão das previsões de seca.

Wang *et al.* (2022) implementaram múltiplos modelos de *Extreme Learning Machine* (ELM) para previsão de secas hidrológicas em escalas mensais e sazonais, obtendo resultados robustos e confiáveis em diferentes bacias hidrográficas.

Elbeltagi *et al.* (2023) compararam três algoritmos distintos — *Random Forest* (RF), *Random Tree* (RT) e *Gaussian Process Regression* (GPR) — na previsão do SPI, concluindo que o RF apresentou o melhor desempenho em termos de acurácia e robustez.

Oliveira *et al.* (2023) desenvolveram um modelo baseado em redes neurais artificiais alimentado por dados de precipitação via satélite para fornecer alertas de estiagens e secas, demonstrando o potencial dessas tecnologias para aplicações práticas no monitoramento de recursos hídricos.

Hameed *et al.* (2023, 2024) exploraram abordagens de *deep learning* e híbridos,

comparando redes neurais profundas e máquinas de aprendizado extremo regularizadas (ELM) para prever secas com meses de antecedência, destacando que modelos híbridos podem superar redes profundas em determinadas condições climáticas tropicais.

Logo, esses estudos ilustram o avanço significativo no uso de técnicas de aprendizado de máquina para previsão e monitoramento de secas e estiagens, desde abordagens clássicas até métodos mais sofisticados, como *deep learning* e modelos híbridos. Eles demonstram como diferentes algoritmos podem ser aplicados para diversas escalas temporais e geográficas, utilizando uma ampla gama de variáveis climáticas, hidrológicas e de sensoriamento remoto, contribuindo para a melhoria da gestão de recursos hídricos e planejamento agrícola frente aos desafios impostos por eventos extremos de estiagem e seca. O Quadro 2 detalha as escalas temporais e os locais em que tais pesquisas foram aplicadas.

Quadro 1 - Resumo comparativo de técnicas para estudos relacionados a seca e estiagens

<b>Autores (Ano)</b>	<b>Técnica</b>	<b>Região Geográfica</b>	<b>Escala Temporal</b>	<b>Objetivo Principal</b>
Blain e Brunini (2007)	Índices meteorológicos (SPI, PDSI adap) e agrícola (IPER)	São Paulo	Mensal e decenal	Destacar a importância da escala temporal no monitoramento das secas
Gonçalves <i>et al.</i> (2021)	Revisão bibliográfica de índices de seca e metodologias de monitoramento	Mundial (com exemplos nos EUA, Europa e Brasil)	Não especificada (aplicável em múltiplas escalas, conforme os índices)	Revisar os principais índices e metodologias de monitoramento e previsão de secas, destacando sistemas de alerta precoce.
Ganguli e Reddy (2014)	Máquina de Vetores de Suporte (SVM)	Índia	Mensal (1896-2005)	Prever secas regionais integrando múltiplos <i>inputs</i> climáticos.
Feng <i>et al.</i> (2020)	Modelos preditivos (baseados em padrões climáticos)	Austrália (Cinturão do Trigo)	Sazonal (estação de crescimento, abril a novembro) 2001-2014	Prever a seca meteorológica durante a estação de crescimento do trigo.
Bouaziz <i>et al.</i> (2021)	Aprendizado de Máquina (sensoriamento remoto)	Norte da África (Regiões Áridas)	Mensal (1981-2019)	Rastrear e prever secas em áreas com poucos dados meteorológicos, usando dados de satélite.
Carvalho <i>et al.</i> (2021)	Máquina de Vetores de Suporte (SVM)	Brasil (Ceará)	Anual (1957-2018)	Classificar anos como 'secos' ou 'não secos' com base na temperatura da superfície do mar.

Dikshit <i>et al.</i> (2021)	Redes Neurais LSTM (Long Short-Term Memory)	Austrália	Mensal (SPEI-1 e SPEI-3, avaliados com previsão de até 1 mês) entre 1901-2018	Aprimorar a previsão do índice de seca SPEI, capturando dependências temporais de longo prazo.
Wang <i>et al.</i> (2022)	Extreme Learning Machine (ELM)	Bacias hidrográficas (não especificadas)	Mensal entre 1963-2017	Realizar a previsão de secas hidrológicas em diferentes escalas de tempo.
Elbeltagi <i>et al.</i> (2023)	Random Forest (RF), Random Tree (RT), Gaussian Process Regression (GPR)	Egito	Mensal (2000-2019)	Comparar algoritmos para prever o índice de seca SPI; Random Forest foi o mais preciso.
Oliveira <i>et al.</i> (2023)	Redes Neurais Artificiais (RNA)	Brasil (Região Sul)	Mensal (1996-2022)	Fornecer alertas de estiagens e secas usando dados de precipitação de satélite.
Hameed <i>et al.</i> (2023, 2024)	Deep Learning (Redes Neurais Profundas) e ELM Híbrido	Malásia (Região Tropical)	Multi-mensal (até 6 meses)	Comparar modelos de deep learning e híbridos para previsão de secas com meses de antecedência.

Fonte: Elaborado pela autora

## 2.4 CARACTERÍSTICAS E EXIGÊNCIAS CLIMÁTICAS DA SOJA

A soja (*Glycine max* (L.) Merrill), amplamente cultivada no Brasil para produção de grãos, é uma planta herbácea de porte baixo, com um sistema radicular bem desenvolvido composto por uma raiz principal e várias ramificações laterais, que facilitam a absorção de água e nutrientes do solo (Farias *et al.*, 2009).

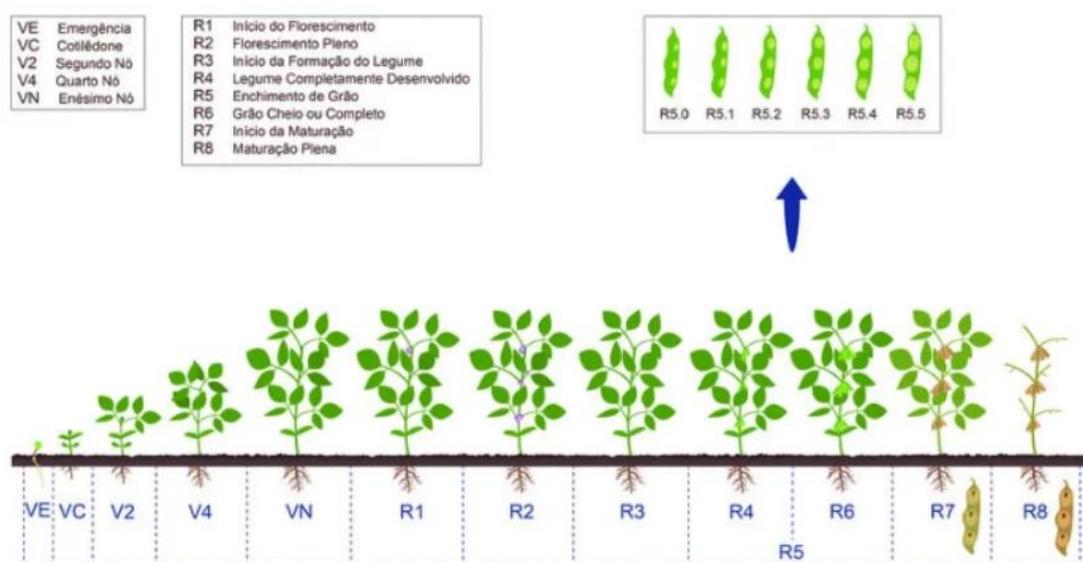
Suas folhas são majoritariamente trifolioladas, ou seja, formadas por três folíolos semelhantes a um trevo. As flores são pequenas e apresentam cores que variam entre branco, roxo ou tons intermediários, sendo autopolinizadas, o que significa que a fecundação ocorre na própria flor sem a necessidade de agentes externos (Farias *et al.*, 2009). Após esse processo, surgem as vagens, que abrigam de uma a cinco sementes arredondadas ou elípticas, geralmente de coloração amarela, com uma marca clara ou escura no ponto de ligação com a vagem (Farias *et al.*, 2009).

De acordo com Farias *et al.* (2007), entender a fenologia da soja, que descreve os estágios de desenvolvimento da planta ao longo de seu ciclo, é essencial para associar cada fase às necessidades agrônômicas específicas. O método mais utilizado mundialmente para

essa classificação foi desenvolvido por Fehr e Caviness (1977), devido a sua terminologia padronizada e objetiva. Esse sistema permite descrever com precisão o crescimento de uma única planta ou de uma lavoura inteira, independentemente da cultivar, facilitando o manejo e o acompanhamento do ciclo produtivo (Farias *et al.*, 2007).

O modelo de Fehr e Caviness (1977) divide o ciclo da soja em duas fases principais: a vegetativa (V), que corresponde ao crescimento das folhas e caules, e a reprodutiva (R), quando a planta floresce, forma vagens, produz grãos e atinge a maturação. Cada estágio é identificado por letras e números que indicam o desenvolvimento da planta desde a germinação até a colheita. Conforme estudo de Fordoński *et al.* (2023), o ciclo completo da cultura da soja, do plantio à colheita, variou entre 115 e 163 dias, ou seja, aproximadamente entre 3,8 e 5,5 meses. A Figura 4 é ilustrada os estádios vegetativos e reprodutivos, bem como os seus descritivos no Quadro 2.

Figura 4 - Estádios Fenológicos da Cultura da Soja



Fonte: Adaptado de Oliveira Junior *et al.* (2016).

Quadro 2 - Estádios de desenvolvimento da soja.

<b>Período</b>	<b>Estádio</b>	<b>Descrição</b>
<b>Vegetativo</b>	VE	Cotilédones acima da superfície do solo.
	VC	Cotilédones completamente abertos.
	V1	Folhas unifolioladas completamente desenvolvidas.
	V2	Primeira folha trifoliolada completamente desenvolvida.
	V3	Segunda folha trifoliolada completamente desenvolvida.
	Vn	Ante-enésima folha trifoliolada completamente desenvolvida.
<b>Reprodutivo</b>	R1	Início do florescimento – Uma flor aberta em qualquer nó do caule.
	R2	Florescimento pleno – Uma flor aberta, num dos 2 últimos nós do caule, com folha completamente desenvolvida.
	R3	Início da formação da vagem – Vagem com 5 mm de comprimento, num dos 4 últimos nós do caule, com folha completamente desenvolvida.
	R4	Vagem completamente desenvolvida – Vagem com 2 cm de comprimento, num dos 4 últimos nós do caule, com folha completamente desenvolvida.
	R5	Início do enchimento do grão – Grão com 3 mm de comprimento em vagem, num dos 4 últimos nós do caule, com folha completamente desenvolvida.
	R6	Grão cheio ou completo – Vagem contendo grãos verdes, preenchendo as cavidades da vagem em um dos 4 últimos nós do caule, com folha completamente desenvolvida.
	R7	Início da maturação – Uma vagem normal no caule, com coloração madura.
	R8	Maturação plena – 95% das vagens com coloração de madura.

Fonte: Adaptado de Farias *et al.* (2009).

Entre os fatores climáticos que influenciam o desenvolvimento da soja, a temperatura e a oferta de água são os mais relevantes (Farias *et al.*, 2007). A cultura apresenta seu melhor desempenho em regiões onde as temperaturas variam entre 20 °C e 30 °C, sendo 30 °C a temperatura ideal para seu crescimento. Além disso, a disponibilidade de água é essencial ao

longo do ciclo, com a planta podendo demandar até 8 mm diários (Farias *et al.*, 2009), tendo como fontes a precipitação, irrigação e armazenamento de água no solo.

A necessidade hídrica aumenta progressivamente com o crescimento da soja, atingindo o pico nas fases de floração e enchimento de grãos (R1-R6), períodos mais críticos em termos de suprimento de água (Farias *et al.*, 2009). Após essa etapa, a demanda por água diminui gradativamente à medida que a planta se aproxima da maturação. Para alcançar o máximo rendimento, é necessário que o volume total de água durante o ciclo tenha variação entre 450 e 800 mm (Farias *et al.*, 2007).

Além disso, Farias *et al.* (2009) afirmam que a escassez de chuvas durante o ciclo da soja pode comprometer seriamente o crescimento da planta e reduzir a produtividade. Em condições de estiagem, a soja apresenta crescimento limitado, folhas pequenas e murchas, além da queda de flores e abortamento de vagens, resultando em menor quantidade e qualidade de grãos. Ainda segundo os autores, por outro lado, o excesso de chuva também traz prejuízos, como o encharcamento do solo, que impede a respiração adequada das raízes e dificulta a absorção de nutrientes, deixando a planta mais vulnerável a doenças.

Por fim, além dos impactos no crescimento das plantas, o excesso de umidade no solo também interfere nas operações agrícolas, dificultando o uso de máquinas durante a colheita e, conseqüentemente, elevando as perdas de produção.

## 2.5 INTEGRAÇÃO DOS CONCEITOS TEÓRICOS

As mudanças climáticas têm provocado alterações significativas no regime de chuvas, impondo novos desafios à sojicultura. A irregularidade pluviométrica durante o verão, especialmente em fases fenológicas críticas, afeta diretamente a disponibilidade hídrica, resultando em perdas de produtividade e maior vulnerabilidade da cultura.

Os conceitos tradicionais de seca, estiagem, estiagem agrícola e veranico, embora úteis em um sentido amplo, mostram-se generalistas, pois não captam as necessidades específicas da soja em diferentes estádios de desenvolvimento. A combinação desses conceitos, adaptada à realidade fenológica da cultura, possibilita a criação de critérios mais precisos para identificar déficits hídricos que comprometem a produção.

Do ponto de vista teórico, a evolução dos estudos de crescimento econômico evidencia que a tecnologia desempenha papel determinante para sustentar a produtividade diante de restrições ambientais.

Desde as formulações clássicas, passando pela endogeneização da inovação e

chegando aos debates contemporâneos sobre clima e sustentabilidade, a literatura reforça que o progresso tecnológico não é apenas um motor de crescimento, mas também uma condição para compatibilizar atividade econômica e limites naturais. Transportada para a agricultura, essa perspectiva revela que sistemas tecnológicos de monitoramento, sensoriamento remoto e aprendizado de máquina funcionam como instrumentos de adaptação, permitindo antecipar riscos climáticos e otimizar a gestão hídrica.

Assim, a integração entre conceitos climáticos, exigências fenológicas da soja e tecnologias inovadoras estabelece um arcabouço robusto para enfrentar os desafios impostos pela variabilidade climática. Essa articulação fornece os fundamentos necessários para propor métodos de monitoramento e classificação de estiagens que ampliem a resiliência da sojicultura, auxiliando os produtores a ajustarem práticas agrícolas de forma mais eficiente e sustentável.

### 3. PROCEDIMENTOS DE ANÁLISE

#### 3.1 CLASSIFICAÇÃO DA PESQUISA

De acordo com Lakatos e Marconi (2003), o método científico é o conjunto de processos ou operações mentais aplicadas na investigação, considerando a linha de raciocínio adotada no processo de pesquisa para alcançar os objetivos propostos. Deste modo, os procedimentos para a realização deste projeto serão apresentados a seguir. O método de abordagem utilizado será o método hipotético-dedutivo de Karl Popper, que fornece as bases lógicas para o desenvolvimento deste trabalho. Segundo Gil (1999, p. 30):

“[...] quando os conhecimentos disponíveis sobre determinado assunto são insuficientes para a explicação de um fenômeno, surge o problema. Para tentar explicar as dificuldades expressas no problema, são formuladas conjecturas ou hipóteses. Das hipóteses formuladas, deduzem-se consequências que deverão ser testadas ou falseadas. Falsear significa tornar falsas as consequências deduzidas das hipóteses. Enquanto no método dedutivo se procura a todo custo confirmar a hipótese, no método hipotético-dedutivo, ao contrário, procuram-se evidências empíricas para derrubá-la” (Gil, 1999, p.30).

O método de procedimento, entendido como o conjunto de etapas concretas da investigação, será o método estatístico, que permite descrever e reduzir o fenômeno estudado em termos quantitativos (Lakatos e Marconi, 2003). Quanto à natureza, este trabalho é classificado como aplicado, pois visa gerar conhecimentos voltados para a aplicação prática, buscando solucionar um problema específico. No que se refere à forma de abordagem, o trabalho é de natureza quantitativa, com características exploratórias e descritivas. A pesquisa exploratória busca proporcionar maior familiaridade com o problema estudado, enquanto a abordagem descritiva visa caracterizar o fenômeno em estudo de forma sistemática. Os procedimentos técnicos incluem o levantamento bibliográfico e a análise experimental, fornecendo embasamento teórico e metodológico para o estudo. No que diz respeito à obtenção de dados, serão utilizados dados secundários de domínio público, os quais, segundo Mattar (1996, p. 134):

“[...] dados secundários são aqueles que já foram coletados, tabulados, ordenados e, às vezes, até analisados, com propósitos outros ao de atender às necessidades da pesquisa em andamento, e que estão catalogados à disposição dos interessados” (Mattar, 1996, p. 134).

No Quadro 3 é mostrado o resumo de toda a discussão referente ao referencial metodológico aplicado neste trabalho.

Quadro 3 – Síntese dos métodos, tipos e técnicas de pesquisa.

<b>Pesquisa</b>	<b>Classificação</b>	<b>Modalidade</b>
<b>Método</b>	De Abordagem (Lógicos)	Hipotético-Dedutivo
	De Procedimentos (Técnicos)	Estatístico
<b>Tipo</b>	Quanto à natureza	Aplicada
	Quanto a forma de abordagem	Quantitativa
	Quanto ao Objetivo Geral	Exploratório e Descritivo
	Quanto aos procedimentos Técnicos	Bibliográfica e experimental
<b>Técnica</b>	Quanto à obtenção de dados	Dados secundários de domínio público

Fonte: Elaborado pela autora.

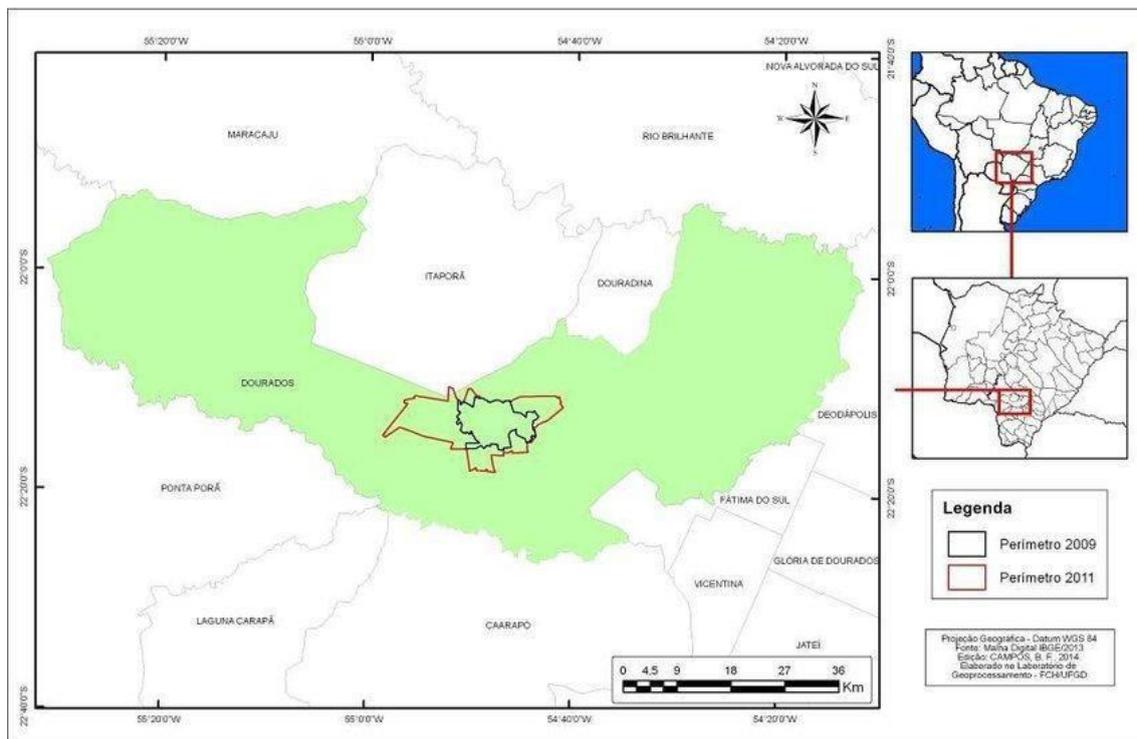
### 3.2 ÁREA DE ESTUDO

Dourados é um município brasileiro localizado no estado de Mato Grosso do Sul, a aproximadamente 225 km da capital, Campo Grande (IBGE, 2025). Com uma população estimada de 243.367 habitantes em 2022, é a segunda cidade mais populosa do estado (IBGE, 2024). Fundada em 20 de dezembro de 1935, Dourados destaca-se como um importante centro agropecuário e de serviços na região sul do estado (IBGE, 2025).

A economia local é fortemente baseada na agricultura, com destaque para o cultivo de soja, milho e cana-de-açúcar, além da pecuária (IBGE, 2025). A cidade também abriga diversas instituições de ensino superior, como a Universidade Federal da Grande Dourados (UFGD) e a Universidade Estadual de Mato Grosso do Sul (UEMS), consolidando-se como um pólo educacional no estado (IBGE, 2025).

Geograficamente, Dourados está situada na região Centro-Oeste do Brasil, próxima à Serra de Maracaju e inserida na bacia do Rio Paraná (IBGE, 2025). A vegetação original é composta por campos limpos, cerrados e matas tropicais (IBGE, 2025). Na Figura 5 é mostrada a localização do município em questão.

Figura 5 - Localização do município de Dourados (MS).



Fonte: Oliveira e Costa (2018).

O clima de Dourados é classificado como Cwa, segundo a classificação climática de Köppen, caracterizado por verões quentes e úmidos e invernos secos com temperaturas moderadas (Alvares *et al.*, 2013).

A temperatura média anual é de 23,1 °C, com máximas médias de 31,9 °C nos meses de verão e mínimas médias de 10,7 °C nos meses de inverno. Eventos extremos incluem registros de temperaturas acima de 40 °C e mínimas negativas, com a menor temperatura registrada de -1,9 °C (Fietz *et al.*, 2024).

A precipitação média anual é de 1.375 milímetros, com maior concentração durante o verão, especialmente entre dezembro e fevereiro. O inverno apresenta menores índices pluviométricos, sendo julho o mês menos chuvoso, com média de 42,3 milímetros. Períodos prolongados sem chuva são comuns no inverno, com eventos de até 92 dias consecutivos sem precipitação registrados (Fietz *et al.*, 2024).

### 3.3 BASE DE DADOS E HIPÓTESE DE PESQUISA

Os dados utilizados neste estudo foram extraídos de diferentes fontes, abrangendo tanto informações meteorológicas quanto índices de produtividade agrícola. Os registros de produtividade da soja, expressos em sacas por hectare, foram obtidos nos relatórios anuais da APROSOJA-MS, cobrindo o período de 2016 a 2024, tanto para o estado de Mato Grosso do Sul quanto para o município de Dourados.

A base de dados empregada e grande parte do trabalho refere-se à estação meteorológica da Embrapa Dourados (MS), situada nas coordenadas latitude 22° 16' 31" S, longitude 54° 49' 06" W a uma altitude de 408 metros (EMBRAPA, 2025). Essa estação foi selecionada por apresentar uma série histórica completa para o período de análise, com dados diários contínuos e ininterruptos.

As estações meteorológicas de Dourados (MS), operadas pela Embrapa Agropecuária Oeste, compreendem unidades convencionais e automáticas ao longo da sua história. A estação convencional iniciou suas atividades em junho de 1979, realizando três leituras diárias dos principais elementos meteorológicos, incluindo temperatura, umidade do ar, velocidade do vento, insolação e precipitação. Em maio de 1999, passou a operar parcialmente, com apenas uma leitura diária de temperatura e chuva, sendo desativada em dezembro de 2000 (Fietz *et al.*, 2024).

A estação automática começou a operar em janeiro de 2001, seguida pela instalação de uma segunda unidade em outubro de 2002. Essas estações realizam medições contínuas de temperatura e umidade do ar, direção e velocidade do vento, precipitação e radiação solar global. A partir de junho de 2013, os dados passaram a integrar o sistema Guia Clima, responsável pelo monitoramento agrometeorológico regional e emissão de alertas em tempo real (Fietz *et al.*, 2024).

As medições são realizadas por instrumentos como termômetros, anemômetros, piranômetros e pluviômetros a cada hora. As estações também registram eventos extremos, como geadas, ondas de calor, períodos sem precipitação e ventos fortes, fornecendo informações para o planejamento agrícola e o monitoramento climático regional (Fietz *et al.*, 2024). Ressalta-se que os dados do Instituto Nacional de Meteorologia não foram utilizados devido à ausência de informações sobre temperatura e precipitação. A seleção das variáveis utilizadas neste estudo baseou-se em pesquisas anteriores e na disponibilidade de dados em escala municipal e temporal da estação meteorológica.

Foram consideradas as variáveis de data, temperatura média, temperatura mínima, temperatura máxima e precipitação, com registros diários no período de 1º de junho de 1979 a 31 de dezembro de 2023.

Por fim, este trabalho parte da hipótese de que a definição de um indicador de estiagem específico para a cultura da soja, integrando os conceitos de estiagem e veranico durante a estação chuvosa, permite uma caracterização mais precisa desse fenômeno e viabiliza seu monitoramento contínuo por meio de técnicas de machine learning.

### 3.4 PROCEDIMENTOS METODOLÓGICOS

Foi desenvolvido um modelo de machine learning supervisionado com o objetivo de classificar a ocorrência de estiagem na cultura da soja na região de Dourados/MS, com base em dados históricos de precipitação diária. O processo metodológico seguiu as etapas de construção da base de dados, definição da variável-alvo, escolha do algoritmo de classificação, treinamento e avaliação do modelo.

#### 3.4.1 Coleta e Tratamento dos Dados

Foram utilizados dados diários de precipitação e temperatura (máxima, mínima e média) do período de 1979 a 2023, provenientes da estação meteorológica local. O tratamento incluiu a remoção de outliers, preenchimento de valores ausentes e normalização das variáveis, assegurando a consistência e a qualidade da base de dados. Os dados foram divididos em conjuntos de treinamento (70%) e teste (30%), assegurando que as safras de 2021/2022 e 2022/2023 fossem reservadas exclusivamente para validação externa.

#### 3.4.2 Definição do Indicador de Estiagem

A estiagem foi caracterizada combinando conceitos de estiagem e veranico, considerando a precipitação acumulada em janelas móveis de 5 e 10 dias. O limiar crítico foi estabelecido em 20,6 mm para a janela de 5 dias, com base em estatísticas descritivas da série histórica.

### 3.4.3 Seleção e Engenharia de Variáveis

A matriz de correlação indicou que as variáveis mais relevantes para a classificação da estiagem foram temperatura média (°C), precipitação diária (mm), ano, mês, dia do ano e a razão entre os acumulados de precipitação de 5 e 10 dias. As demais variáveis foram descartadas com o objetivo de reduzir a complexidade e o risco de sobreajuste no modelo.

### 3.4.4 Formulação Matemática

A variável resposta  $Y$  é binária, representando a ocorrência de estiagem (1) ou sua ausência (0) em uma janela móvel de cinco dias consecutivos. Com base na análise estatística da distribuição de precipitação diária e nas necessidades hídricas da cultura da soja, definiu-se que:

$$Y = \begin{cases} 1, & \text{se } \sum_{i=t-4}^t P_i < 20,6 \text{ mm} \\ 0, & \text{caso contrário} \end{cases} \quad (1)$$

Onde  $P_i$  representa a precipitação registrada no dia  $i$ . As variáveis explicativas  $X$  utilizadas no modelo foram derivadas a partir da série temporal de dados meteorológicos, com foco em características diretamente relacionadas à dinâmica climática da região. Foram consideradas a temperatura média diária e a precipitação acumulada (chuva), ambas extraídas de registros históricos.

Além dessas variáveis físicas, foram incluídas variáveis temporais como o ano, o mês e o dia do ano, com o intuito de captar padrões sazonais e efeitos recorrentes ao longo do ciclo produtivo. Também foi incorporada a taxa de variação do acumulado de precipitação entre janelas móveis de cinco e dez dias consecutivos, permitindo identificar mudanças abruptas no regime de chuvas.

Essas variáveis, combinadas, fornecem ao modelo informações suficientes para discriminar padrões associados à ocorrência de estiagem de forma supervisionada. Assim, o objetivo do modelo é encontrar uma função  $f$  tal que  $Y = f(X)$  minimize a taxa de erro na classificação da estiagem.

Para operacionalizar essa formulação matemática, todas as variáveis foram estruturadas em uma planilha no Excel, onde se organizaram os dados meteorológicos originais (temperatura média, máxima, mínima e precipitação diária), as variáveis derivadas (acumulados de cinco e dez dias e suas taxas de variação) e as classificações

binárias da estiagem (0 = ausência; 1 = ocorrência). A classificação foi implementada segundo diferentes metodologias de referência e pela proposta deste estudo, o que possibilitou a comparação sistemática entre critérios. A planilha também contou com recursos de coloração condicional, permitindo destacar visualmente os períodos críticos e facilitar a análise exploratória dos resultados. O exemplo dessa operacionalização é mostrada na Figura 6.

Figura 6 - Estruturação em Excel dos dados meteorológicos e da classificação da estiagem.

data	Tmedia	Tmax	Tmin	chuva	acumulado5dias	acumulado10dias	MetodologiaAGR	classdiário (literatura)	class5dias (média)	class10dias (média)	classAgr (Literatura)
8/22/2021	27.2	35.7	17.5	0	#REF!	#REF!					
8/23/2021	27.9	36.8	19.1	0	#REF!	#REF!					
8/24/2021	29.2	36.3	22.0	0	#REF!	#REF!					
8/25/2021	28.0	35.7	17.6	0	#REF!	#REF!					
8/26/2021	15.2	17.7	13.1	2,5	2,5	#REF!					
8/27/2021	14.3	15.9	13.3	40,1	42,6	#REF!					
8/28/2021	17,8	22,2	14,4	2,2	44,8	#REF!					
8/29/2021	21,7	28,7	15,5	1,1	45,9	#REF!					
8/30/2021	21,5	29,3	15,9	0	45,9	#REF!					
8/31/2021	23,2	31,1	16,8	0	43,4	45,9	45,9				
9/1/2021	24,4	32,4	17,1	0	3,3	45,9		1	1	0	
9/2/2021	26,1	34,5	18,7	0	1,1	45,9		1	1	0	
9/3/2021	26,5	36,5	16,3	0	0	45,9		1	1	0	
9/4/2021	28,1	37,3	17,9	0	0	45,9		1	1	0	
9/5/2021	24,4	30,3	20,7	3,1	3,1	46,5		1	1	0	
9/6/2021	28,0	36,2	19,5	0	3,1	6,4		1	1	1	
9/7/2021	30,4	38,6	22,0	0	3,1	4,2		1	1	1	
9/8/2021	27,9	35,4	18,3	0	3,1	3,1		1	1	1	
9/9/2021	18,2	22,1	14,9	2,5	5,6	5,6		1	1	1	
9/10/2021	17,5	22,1	13,8	0,3	2,8	5,9	5,9	1	1	1	1
9/11/2021	21,8	29,7	14,8	1,9	4,7	7,8		1	1	1	
9/12/2021	27,2	35,5	19,5	0	4,7	7,8		1	1	1	
9/13/2021	29,8	36,8	22,8	0	4,7	7,8		1	1	1	
9/14/2021	27,4	36,7	20,0	1,6	3,8	9,4		1	1	1	
9/15/2021	23,4	30,8	17,3	0	3,5	6,3		1	1	1	
9/16/2021	24,1	33,7	18,3	39,5	41,1	45,8		0	0	0	
9/17/2021	24,7	32,3	18,8	0,1	41,2	45,9		1	0	0	
9/18/2021	28,2	36,9	20,6	0	41,2	45,9		1	0	0	
9/19/2021	30,2	38,7	22,4	0	39,6	43,4		1	0	1	
9/20/2021	30,5	39,2	21,5	0	39,6	43,1	43,1	1	0	1	1

Fonte: Elaborado pela autora.

### 3.4.5 Avaliação dos Modelos

Foi utilizado o Lazy Predict (2022), uma biblioteca que permite testar rapidamente diversos algoritmos de aprendizado de máquina, avaliando 27 classificadores. Os três modelos com melhor desempenho – *Random Forest Classifier*, *Extra Trees Classifier* e *XGB Classifier* – foram selecionados para análise mais detalhada. O desempenho dos modelos foi avaliado por meio de métricas como acurácia, precisão, recall e F1-score, além da análise da matriz de confusão.

A acurácia mede a proporção de previsões corretas em relação ao total de amostras, sendo útil como uma métrica geral, especialmente em conjuntos de dados equilibrados. A precisão, por sua vez, indica a proporção de verdadeiros positivos entre todas as previsões positivas, enquanto o recall (ou sensibilidade) mede a proporção de verdadeiros positivos identificados entre todos os casos realmente positivos. A especificidade, por outro lado, expressa a proporção de verdadeiros negativos corretamente identificados entre todos os casos realmente negativos, sendo fundamental para avaliar a capacidade do modelo em reconhecer a classe “sem estiagem” neste estudo. Já o F1-score é a média harmônica entre precisão e recall,

sendo uma métrica particularmente relevante em contextos com desequilíbrio entre as classes (Geron, 2019). As fórmulas utilizadas para o cálculo dessas métricas foram:

$$\text{Acurácia} = (\text{VP} + \text{VN}) / (\text{VP} + \text{VN} + \text{FP} + \text{FN}) \quad (2)$$

$$\text{Precisão} = \text{VP} / (\text{VP} + \text{FP}) \quad (3)$$

$$\text{Recall (sensibilidade)} = \text{VP} / (\text{VP} + \text{FN}) \quad (4)$$

$$\text{Especificidade} = \text{VN} / (\text{VN} + \text{FP}) \quad (5)$$

$$\text{F1-score} = 2 \times (\text{Precisão} \times \text{Recall}) / (\text{Precisão} + \text{Recall}) \quad (6)$$

Onde VP = verdadeiros positivos, VN = verdadeiros negativos, FP = falsos positivos e FN = falsos negativos. Além dessas métricas, a matriz de confusão foi utilizada para verificar, de forma detalhada, os acertos e erros cometidos pelo modelo, identificando a distribuição das classificações verdadeiras e previstas (Geron, 2019). Essa abordagem possibilita compreender não apenas a performance global do classificador, mas também sua capacidade de identificar corretamente os casos de estiagem e evitar falsos positivos, que poderiam comprometer a tomada de decisão em campo.

No contexto deste trabalho, os verdadeiros positivos (VP) correspondem a períodos de estiagem corretamente identificados, enquanto os falsos negativos (FN) representam estiagens reais não detectadas, o erro mais crítico do ponto de vista da segurança do produtor. Os falsos positivos (FP) indicam alertas de estiagem emitidos sem que ela tenha ocorrido, podendo reduzir a confiança no sistema, e os verdadeiros negativos (VN) correspondem a dias sem estiagem corretamente classificados. Dessa forma, a matriz de confusão, apresentada no Quadro 4, constitui a base para interpretar a natureza dos erros cometidos e avaliar se o modelo atende ao critério de segurança estabelecido para o monitoramento da sojicultura.

Quadro 4 - Matriz de confusão aplicada à classificação de estiagem na sojicultura.

	<b>Sem estiagem</b>	VN	FP
<b>Real</b>	<b>Com estiagem</b>	FN	VP
		<b>Sem estiagem</b>	<b>Com estiagem</b>
		<b>Classificado pelo modelo (previsto)</b>	

Fonte: Elaborado pela autora

### 3.4.6 Validação dos Modelos

A validação foi realizada utilizando dados das safras 2021/2022 e 2022/2023, que não foram incluídos no treinamento. O objetivo foi avaliar a capacidade dos modelos de identificar estiagens em diferentes condições meteorológicas. As estimativas foram comparadas com a ocorrência real das estiagens, verificando a proximidade dos resultados previstos em relação aos valores observados.

Optou-se por uma divisão simples entre treino (70%) e teste (30%), respeitando a ordem cronológica da série temporal. Essa estratégia visa evitar vazamento de informação entre os conjuntos e garantir que o modelo seja avaliado com base em dados efetivamente posteriores aos utilizados no treinamento. Não foi utilizada validação cruzada do tipo k-fold, justamente para preservar a estrutura sequencial dos dados meteorológicos, o que é essencial em estudos baseados em séries temporais.

### 3.4.7 Descritivo dos Algoritmos

O presente estudo utilizou os algoritmos Random Forest Classifier, Extra Trees Classifier e XGB Classifier, amplamente reconhecidos por sua robustez, interpretabilidade e bom desempenho em problemas envolvendo variáveis numéricas e categóricas, além da capacidade de capturar relações não lineares entre os dados (Geron, 2019).

O Random Forest Classifier é um algoritmo de aprendizado em conjunto (ensemble) que utiliza o método de bagging (bootstrap aggregating) para construir múltiplas árvores de decisão de forma independente, com amostragens aleatórias dos dados e dos atributos. A predição final é realizada por meio da votação majoritária entre as árvores da floresta. Esse modelo tende a reduzir o overfitting observado em árvores individuais, promovendo uma melhor generalização (Breiman, 2001).

O Extra Tree Classifier (Extremely Randomized Trees) compartilha princípios semelhantes ao Random Forest, mas introduz um maior nível de aleatoriedade durante a construção das árvores. Enquanto o Random Forest escolhe os melhores pontos de divisão com base em critérios como Gini ou entropia, o Extra Trees seleciona aleatoriamente os pontos de corte para cada atributo, tornando as árvores ainda mais diversas. Isso tende a reduzir o viés do modelo, ao custo de um leve aumento da variância, o que pode ser benéfico em conjuntos de dados com muitos ruídos (Geurts et al., 2006).

Já o XGB Classifier é uma implementação otimizada do algoritmo Gradient Boosting, desenvolvido na biblioteca Boost (eXtreme Gradient Boosting). Ao contrário dos dois anteriores, que constroem árvores em paralelo, o XGB Classifier constroi árvores

sequencialmente, de modo que cada nova árvore corrige os erros cometidos pelas anteriores.

O algoritmo busca minimizar uma função de perda utilizando técnicas de regularização, como L1 (Lasso) e L2 (Ridge), que atuam penalizando a complexidade do modelo. Além disso, o algoritmo possui uma estrutura eficiente que permite a paralelização do processo de treinamento reduzindo o tempo de execução, controle de overfitting e excelente desempenho preditivo, mesmo em conjuntos de dados desbalanceados (Chen e Guestrin, 2016). No Quadro 5 é apresentado um comparativo entre as principais características dos algoritmos adotados.

Quadro 5 – Comparativo entre algoritmos.

Aspecto	<i>Random Forest</i>	<i>Extra Trees</i>	<i>XGBoost</i>
<b>Funcionamento</b>	Várias árvores com votação final.	Várias árvores com divisões aleatórias.	Árvores sequenciais que corrigem erros.
<b>Velocidade</b>	Rápido.	Muito rápido.	Moderada, mas otimizada.
<b>Precisão</b>	Boa.	Boa, com mais variação.	Muito alta.
<b>Facilidade de uso</b>	Fácil.	Muito fácil.	Exige mais ajustes.
<b>Robustez a ruídos</b>	Boa.	Alta.	Alta, com bom ajuste.
<b>Diferencial</b>	Equilíbrio entre desempenho e simplicidade.	Alta aleatoriedade e rapidez.	Alta precisão com controle de overfitting.

Fonte: Elaborado pela autora com base em Breiman (2001), Geurts *et al.* (2006), Chen e Guestrin (2016).

Para lidar com o desbalanceamento da variável-alvo (estiagem), foi aplicada a técnica SMOTE (*Synthetic Minority Over-sampling Technique*), que gera exemplos sintéticos da classe minoritária com base em vizinhos próximos (Chawla *et al.*, 2002).

Os hiperparâmetros de cada algoritmo foram ajustados manualmente por meio de testes empíricos controlados, com variações sucessivas nas configurações de parâmetros como profundidade máxima, número de árvores e taxa de aprendizado (no caso do *XGB Classifier*), até alcançar os melhores resultados observados nas métricas de avaliação. Essa abordagem, embora mais simples do que métodos automatizados como *grid search* ou *random search*, é recomendada em contextos exploratórios ou com limitação de recursos computacionais (Geron, 2019). O modelo final selecionado foi aquele que apresentou melhor desempenho nos dados de validação externa.

#### 3.4.8 Análise Contrafactual da Produtividade da Soja

Para estimar o impacto potencial da estiagem na produtividade da soja, foi realizada uma análise contrafactual com base na comparação entre a safra afetada (2021/2022) e um cenário hipotético de condições climáticas regulares.

Como referência, utilizou-se a produtividade média registrada em Dourados/MS na safra 2022/2023, reconhecida por sua distribuição regular de chuvas. A diferença entre a produtividade real da safra 2021/2022 e a produtividade esperada no cenário contrafactual foi utilizada para estimar a perda em sacas por hectare.

Esse valor serviu de base para a simulação econômica da aplicação seletiva de irrigação. Foram consideradas, ainda, estimativas de custo de produção irrigada e o preço médio da saca de soja vigente no período, a fim de estimar o retorno financeiro da intervenção proposta.

#### 3.4.9 Ferramentas Utilizadas

As ferramentas utilizadas neste estudo incluíram o Python, empregado no processamento dos dados, na construção e validação dos modelos de machine learning. O Microsoft Excel foi utilizado para a organização e tabulação dos dados meteorológicos e estatísticos, facilitando a manipulação inicial das informações. Por fim, o software Origin foi adotado para a criação de gráficos para visualização das séries temporais e das janelas móveis de precipitação, permitindo representar visualmente os padrões e variações observados ao longo do período analisado.

#### 3.4.10 Fluxo dos procedimentos

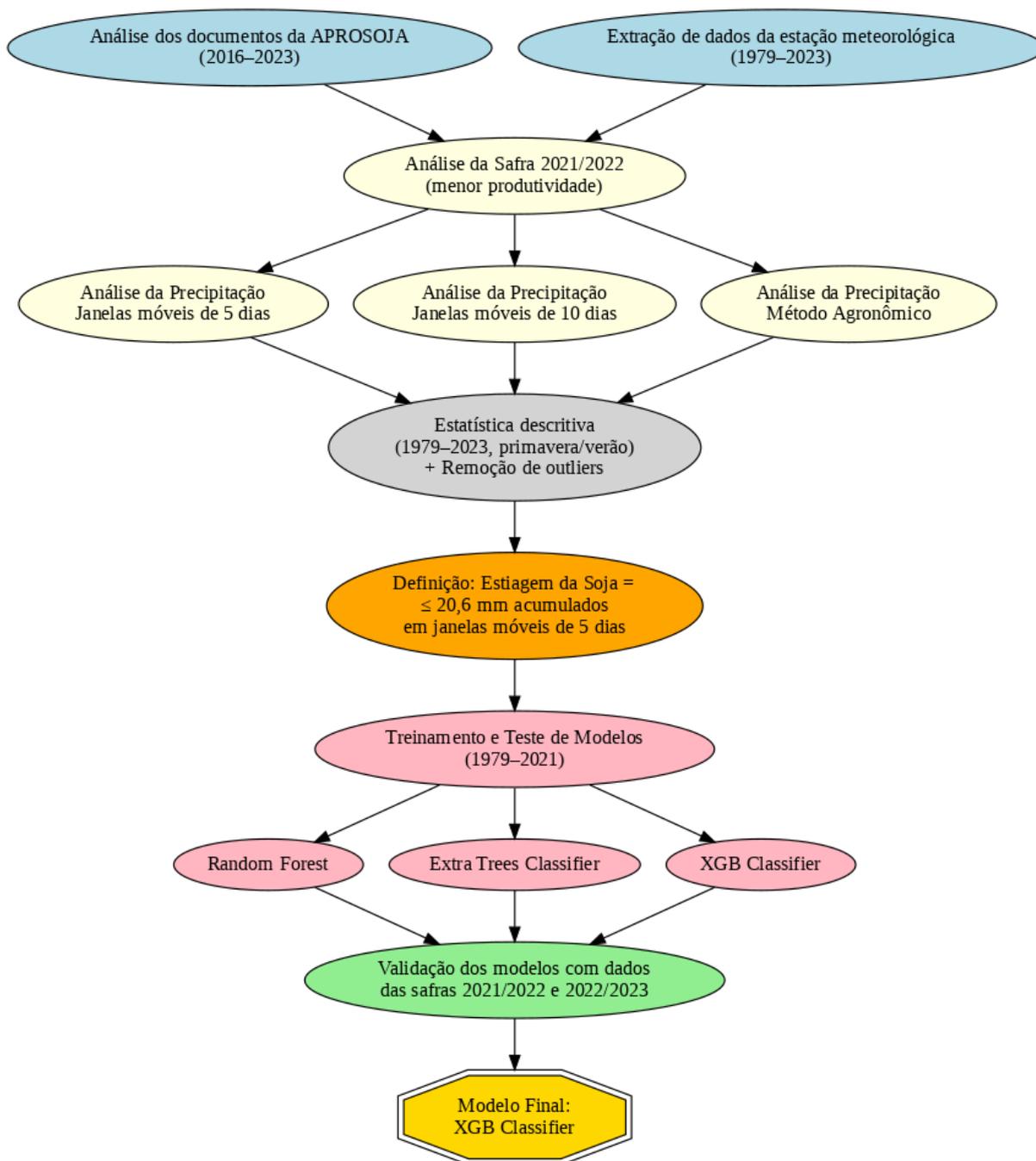
O diagrama apresenta as etapas principais do desenvolvimento da pesquisa. Inicialmente, foi realizada a análise dos documentos da APROSOJA (2016–2023) e a extração dos dados meteorológicos históricos (1979–2023). Em seguida, destacou-se a safra de 2021, identificada como a de menor produtividade, sobre a qual foram conduzidos estudos de comportamento das chuvas em janelas móveis de 5 e 10 dias, além do método agrônomo por decêndio.

Posteriormente, aplicou-se a estatística descritiva da estiagem da cultura da soja, seguida da preparação da base de dados com as informações meteorológicas, que foi dividida em conjuntos de treinamento e teste (1979–2021) e validação (safras 2021/2022 e 2022/2023). Na sequência, foram selecionados os três modelos de melhor desempenho e avaliados com base em suas métricas de classificação e escolhido de melhor

desempenho.

A Figura 7 apresenta o fluxograma metodológico que organiza, de forma sequencial e estruturada, as etapas de coleta de dados, análises de precipitação, definição da regra de negócio, modelagem e validação, resultando no modelo final de classificação de estiagem na sojicultura.

Figura 7 - Fluxo metodológico para o modelo de classificação de estiagem da sojicultura.



Fonte: Elaborado pela autora.

### 3.5 MATRIZ DE AMARRAÇÃO

A Matriz de Amarração, proposta por Mazzon (2018), é uma ferramenta que verifica a adequação entre o modelo adotado, os objetivos, as questões ou hipóteses formuladas e o tratamento dos dados. Essa matriz organiza a configuração da pesquisa por meio de uma formulação matricial. No Quadro 6, apresenta-se a Matriz de Amarração que fundamenta este estudo.

Quadro 6 – Matriz de amarração.

Problema de Pesquisa	Como a construção de um modelo de machine learning, fundamentado em um indicador específico de estiagem para a sojicultura, pode contribuir para a classificação da estiagem, o monitoramento climático contínuo e a redução dos riscos produtivos na região de Dourados/MS?
Objetivo Geral	Classificar a ocorrência de estiagens na cultura da soja no município de Dourados (MS), com o intuito de aprimorar o monitoramento climático e oferecer subsídios aos stakeholders do setor agrícola, contribuindo para a tomada de decisão e a redução dos riscos associados à variabilidade pluviométrica.
Hipótese	A definição de um indicador de estiagem específico para a cultura da soja, integrando os conceitos de estiagem e veranico durante a estação chuvosa, permite uma caracterização mais precisa desse fenômeno e viabiliza seu monitoramento contínuo por meio de técnicas de machine learning.
Objetivos Específicos	<ol style="list-style-type: none"> <li>1. Definir um indicador de estiagem específico para a cultura da soja em Dourados/MS, combinando os conceitos de estiagem e veranico e considerando limites críticos de precipitação na estação chuvosa.</li> <li>2. Examinar a distribuição da precipitação ao longo do ciclo produtivo, identificando padrões históricos de estiagem.</li> <li>3. Desenvolver um modelo de machine learning para a classificação contínua da estiagem, com base no indicador proposto e em variáveis climáticas selecionadas.</li> </ol>
Variáveis	<ul style="list-style-type: none"> <li>- Precipitação diária (mm).</li> <li>- Acumulado de precipitação (5 e 10 dias).</li> <li>- Índice de variação da precipitação (razão entre acumulados).</li> <li>- Ano, mês e dia do ano.</li> <li>- Temperatura média, máxima e mínima.</li> <li>- Classificação binária de estiagem (0 = não; 1 = sim).</li> </ul>
Técnica de Análise	<ul style="list-style-type: none"> <li>- Estatística descritiva e exploratória</li> <li>- Construção de indicador climático com base em janelas móveis</li> <li>- <i>Machine learning</i> para classificação: <i>Random Forest Classifier</i>, <i>Extra Tree Classifier</i> e <i>XGB Classifier</i>.</li> <li>- Validação com dados de safras contrastantes.</li> <li>- Matriz de confusão, análise de desempenho (acurácia, precisão, <i>recall</i> e <i>F1-score</i>).</li> <li>- Análise contrafactual.</li> </ul>

Fonte: Elaborado pela autora.

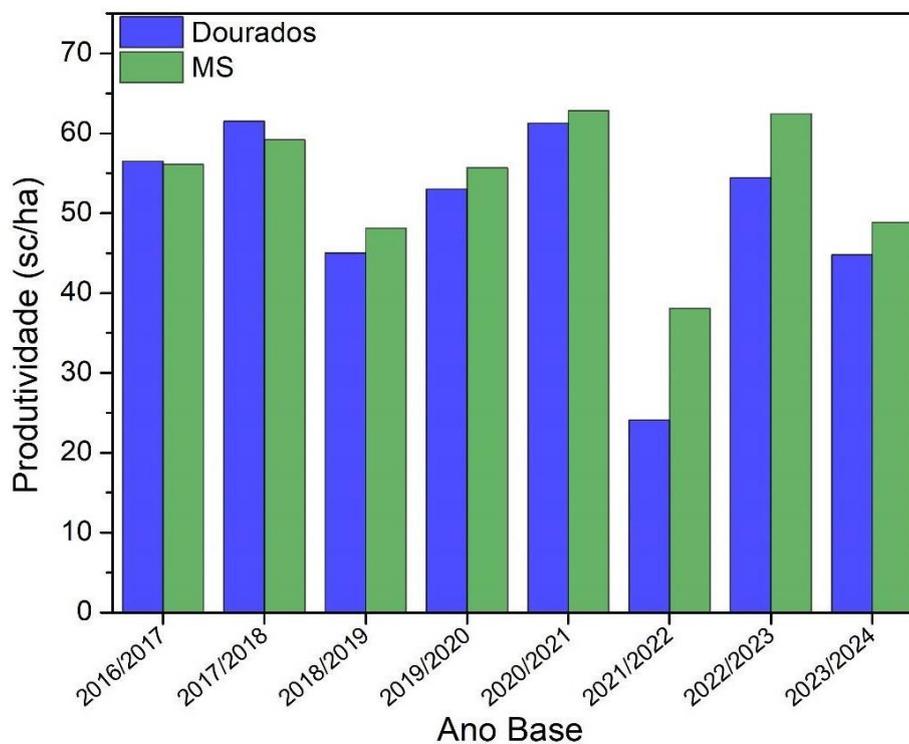
## 4. RESULTADOS E DISCUSSÕES

Esta seção apresenta os resultados da pesquisa, abrangendo a análise da produtividade da soja em Dourados/MS, a caracterização da distribuição da precipitação, a definição de um indicador específico de estiagem e a aplicação de modelos de *machine learning* para sua classificação. Os resultados incluem a avaliação comparativa entre metodologias de monitoramento, o desempenho dos algoritmos para classificação da estiagem, a validação com safras contrastantes e a análise contrafactual.

### 4.1 COMPARATIVO DE PRODUTIVIDADE DA SOJA

A avaliação da produtividade da soja em Dourados e no estado de Mato Grosso do Sul, entre as safras de 2016/2017 e 2023/2024, revela oscilações ao longo dos anos, conforme apresentado na Figura 8.

Figura 8 - Comparação da Produtividade da Soja em Dourados e Mato Grosso do Sul por Safra.



Fonte: Elaborado pela autora, com dados da APROSOJA-MS referentes aos ciclos de safra de 2016/2017 a 2023/2024.

No ciclo de 2016/2017, Dourados obteve uma produtividade de 56,5 sc/ha, levemente superior à média estadual de 56,1 sc/ha, resultando em uma diferença relativa positiva de 1% (APROSOJA, 2017). No ano seguinte, 2017/2018, o município manteve a produtividade acima da média estadual, atingindo 61,51 sc/ha em comparação aos 59,17 sc/ha registrados em Mato Grosso do Sul, o que corresponde a uma diferença positiva de 4% (APROSOJA, 2018).

A partir da safra 2018/2019, houve uma redução na produtividade em Dourados em relação à média estadual, com variações percentuais negativas entre 3% e 7% (APROSOJA, 2019; APROSOJA,2020). Em 2020/2021, observou-se uma recuperação, com o município registrando 61,26 sc/ha, embora ainda abaixo da média estadual (APROSOJA, 2021). Conforme apontado pela APROSOJA-MS (2018), as condições climáticas influenciaram negativamente o desenvolvimento da cultura durante esse período, devido à ocorrência de duas a três estiagens em momentos críticos do ciclo, afetando fases essenciais como floração, formação de canivettino e enchimento de grãos. Apesar desses fatores, produtores que adotaram técnicas de manejo mais avançadas conseguiram mitigar parte dos efeitos negativos, o que impactou diretamente os resultados da safra (APROSOJA-MS, 2018).

O maior recuo na produtividade ocorreu em 2021/2022, quando Dourados alcançou apenas 24,08 sc/ha, enquanto Mato Grosso do Sul registrou 38,05 sc/ha, apresentando uma diferença negativa de 58%. De acordo com a APROSOJA-MS (2022), essa safra foi uma das menos produtivas dos últimos dez anos na região sul do estado, principal área produtora de grãos, que teve sua produção comprometida pela estiagem. Fietz *et al.* (2022) destacam que o fenômeno climático *La Niña* contribuiu para a escassez e má distribuição das chuvas, associadas a temperaturas elevadas, impactando significativamente as lavouras da Grande Dourados.

Em 2022/2023, a produtividade em Dourados apresentou recuperação, atingindo 54,43 sc/ha, ainda inferior à média estadual de 62,44 sc/ha, o que resultou em uma diferença negativa de 15%. Essa melhora foi atribuída ao aumento das precipitações em períodos críticos do ciclo fenológico da soja, proporcionando maior regularidade nas condições das lavouras e reduzindo os efeitos adversos do clima (APROSOJA-MS, 2023).

Por fim, durante a safra de 2023/2024, Mato Grosso do Sul registrou uma das menores produtividades médias ponderadas dos últimos dez anos. O resultado foi

afetado por fatores climáticos, incluindo temperaturas elevadas e déficit hídrico, que comprometeram grandes áreas do estado (APROSOJA-MS, 2024).

As regiões centro e sul, que juntas correspondem a 84,4% da área cultivada com soja, foram as mais impactadas, ocasionando uma redução no potencial produtivo das lavouras (APROSOJA-MS, 2024). O mesmo relatório aponta que as temperaturas médias durante o ciclo agrícola apresentaram elevação em relação aos anos anteriores, provocando estresse térmico nas plantas e afetando o desenvolvimento fenológico da cultura.

Além disso, o volume de chuvas esteve entre 40% e 60% abaixo do esperado, agravando o estresse hídrico nas lavouras. Essas condições foram associadas à forte atuação do fenômeno *El Niño*, que intensificou o aumento das temperaturas e reduziu a precipitação em Mato Grosso do Sul, impactando diretamente o desempenho produtivo estadual (APROSOJA-MS, 2024).

Os relatórios analisados não apresentam a profundidade técnica para discutir sobre questões de *El Niño* e *La Niña*. Neste contexto, destaca-se o trabalho de Cai *et al.* (2020), o qual evidencia que as mudanças projetadas na circulação atmosférica e nas condições oceânicas, especialmente relacionadas às fases do *El Niño* Oscilação Sul (ENOS), podem ter implicações diretas para o Centro-Oeste do Brasil. Durante episódios de *El Niño*, a região tende a registrar secas ou precipitação abaixo da média, condição que pode ser agravada em cenários de aquecimento global devido à maior frequência e intensidade desses eventos extremos. Por outro lado, episódios de *La Niña* costumam favorecer o aumento das chuvas, embora a magnitude e a distribuição espacial desse efeito dependam da intensidade do fenômeno e das interações com o sistema climático global.

Além disso, os autores ressaltam que o aquecimento global intensifica a variabilidade do ENOS, aumentando a probabilidade de ocorrência de eventos extremos. Essa maior instabilidade pode gerar impactos mais severos sobre a disponibilidade hídrica, a agricultura e os processos de evapotranspiração no Centro-Oeste brasileiro, uma vez que os regimes de precipitação passam a ser modulados por alterações na circulação atmosférica e nos padrões de temperatura da superfície oceânica (Cai *et al.*, 2020).

#### 4.2 ANÁLISE DA PRECIPITAÇÃO NO ANO-SAFRA 2021/2022

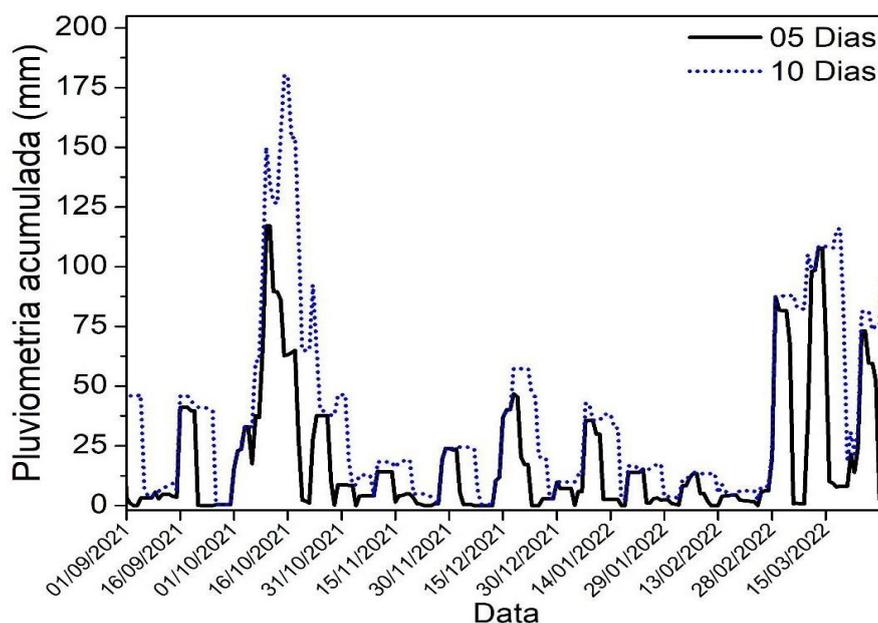
Como mencionado na seção anterior, a safra 2021/2022 foi caracterizada por estiagens e distribuição irregular das chuvas, impactando diretamente a produtividade da soja na região de Dourados.

Na Figura 9 é exibida a análise em janelas móveis da precipitação acumulada de 5 e 10 dias, proposta inovadora do presente trabalho, o que permite visualizar os períodos de maior e menor volume de chuvas ao longo da safra. Os resultados indicam que houve picos de precipitação concentrados entre outubro e novembro de 2021, seguidos por períodos prolongados de baixa precipitação.

No intervalo entre dezembro de 2021 e fevereiro de 2022, período de estágio fenológico de R1-R6 considerado crítico, observou-se uma redução expressiva nos acumulados de chuva, com sucessivos períodos de estiagem, onde os valores acumulados em 10 dias permaneceram baixos. O final do ciclo, em março de 2022, apresentou uma recuperação parcial da precipitação, porém tardiamente para impactar significativamente o desenvolvimento da cultura.

Essa distribuição indica que a soja foi exposta a períodos de chuva insuficiente em momentos críticos do seu ciclo, como floração e enchimento de grãos, o que contribuiu para a redução da produtividade observada na safra.

Figura 9 - Janelas móveis de precipitação acumulada em Dourados/MS no ano safra 2021/2022.



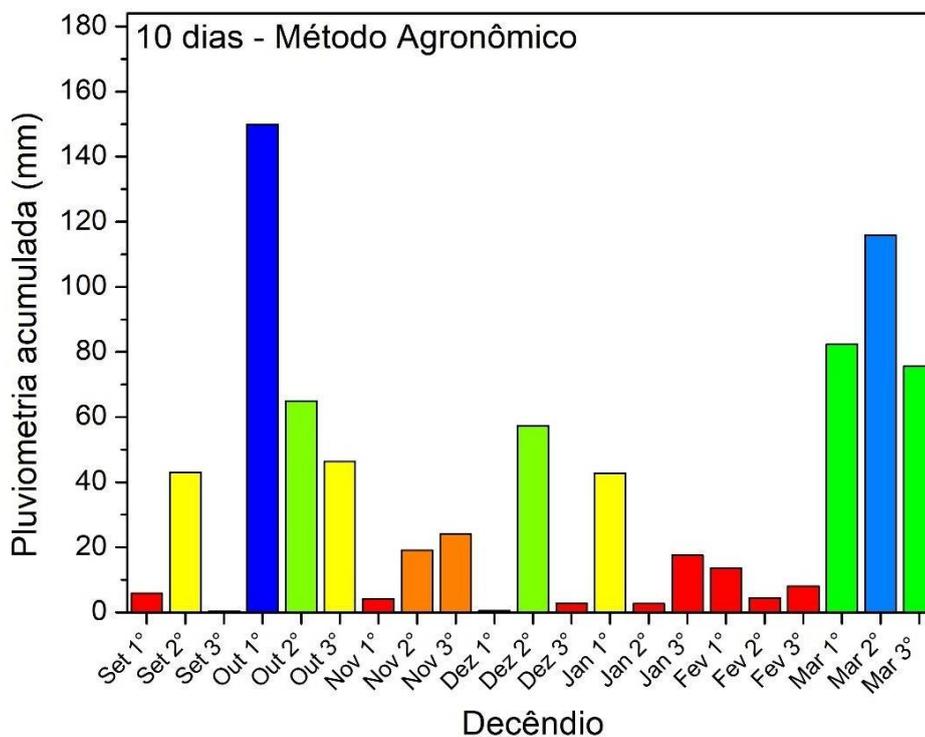
Fonte: Elaborado pela autora.

O segundo método, baseado na precipitação acumulada por decêndio (método agrônômico), exibido na Figura 10, permitiu uma visão mais segmentada da distribuição da chuva ao longo da safra.

Os resultados indicaram que no primeiro e segundo decêndios de outubro houve altos acumulados de chuva, garantindo umidade inicial para o estabelecimento da cultura. No entanto, a partir de novembro, os valores acumulados começaram a diminuir, sendo observada uma redução acentuada em dezembro e janeiro, com decêndios apresentando acumulados muito baixos, caracterizando períodos críticos de estiagem. Nos meses de fevereiro e março, houve certa recuperação da precipitação, mas com irregularidade, onde alguns decêndios ainda apresentaram acumulados reduzidos.

Esses padrões reforçam que a irregularidade das chuvas e os períodos de estiagem foram determinantes para a queda na produtividade da safra 2021/2022. O acúmulo de chuvas nos estágios iniciais da cultura foi insuficiente para compensar o déficit pluviométrico observado nas fases mais críticas, resultando em redução expressiva do rendimento.

Figura 10 - Precipitação por decêndio em Dourados/MS no ano safra 2021/2022.



Fonte: Elaborado pela autora.

A definição da melhor estratégia de monitoramento diário da precipitação com foco na estiagem combinada exige uma análise comparativa entre métodos que permitam identificar períodos de precipitação reduzida de forma eficiente. Ambas apresentam vantagens e limitações que influenciam a precisão da detecção de estiagens e a aplicabilidade dos dados para modelagem.

O método de janelas móveis de 5 e 10 dias calcula a precipitação acumulada dentro de um período deslizando, atualizando-se diariamente conforme novos registros de chuva são incorporados e os mais antigos são descartados. Essa abordagem apresenta alta resolução temporal, permitindo um monitoramento contínuo das variações pluviométricas ao longo do tempo. Além disso, é altamente sensível a mudanças no tempo, possibilitando a rápida identificação de estiagens curtas durante o ciclo, sem a necessidade de esperar intervalos fixos para avaliar a ocorrência de possíveis déficits hídricos.

A flexibilidade desse método permite testar diferentes limiares para classificar a estiagem, ajustando-se conforme a necessidade do modelo. Outro fator relevante é sua compatibilidade com modelos de previsão, pois a atualização diária possibilita alimentar algoritmos de aprendizado de máquina com dados dinâmicos e em tempo real. No entanto, essa abordagem também apresenta algumas limitações. A variação diária da precipitação pode gerar oscilações na classificação, exigindo filtros ou técnicas de suavização para evitar instabilidades no monitoramento. Além disso, o processamento contínuo das janelas móveis pode demandar maior capacidade computacional, uma vez que requer cálculos diários da precipitação acumulada.

Por outro lado, a metodologia baseada no acumulado por decêndio segmenta a precipitação em períodos fixos de 10 dias dentro de cada mês, totalizando três decêndios mensais. Essa abordagem é amplamente utilizada em estudos agrônômicos e climatológicos por sua facilidade de implementação e interpretação. A agregação dos dados em intervalos fixos reduz a influência de eventos extremos pontuais, suavizando oscilações que poderiam ocorrer em séries temporais diárias. Além disso, a análise por decêndio permite comparações diretas entre diferentes safras e facilita a detecção de padrões climáticos ao longo dos anos.

Entretanto, essa abordagem apresenta algumas limitações para o monitoramento diário de estiagem. Como a atualização dos dados ocorre apenas a cada 10 dias, há um atraso na identificação de estiagens curtas, o que pode comprometer a detecção de períodos críticos de baixa precipitação que ocorrem dentro de um único

decêndio. Além disso, a rigidez temporal desse método pode resultar na perda de informações sobre eventos secos relevantes que não coincidam com o início ou fim de um decêndio. Outra desvantagem é sua menor eficiência para modelos de previsão baseados em séries temporais diárias, uma vez que a baixa granularidade dos dados dificulta a adaptação do método para sistemas de monitoramento dinâmico.

Logo, a comparação entre esses métodos evidencia que a abordagem de janelas móveis de 5 e 10 dias é a mais adequada para um monitoramento diário de estiagem combinada. A janela de 10 dias suaviza a série e pode representar melhor o acúmulo hídrico relevante para a planta, o que não é o objetivo central do trabalho. Por outro lado, a janela de 5 dias capta variações mais rápidas nas questões de estiagem e mitigação de risco, sendo a janela escolhida para realizar as análises subsequentes. Abaixo, no Quadro 7 é mostrado a síntese dos resultados:

Quadro 7 - Resumo dos resultados do monitoramento diário de estiagem combinada.

<b>Critério</b>	<b>Janelas Móveis (5 a 10 dias)</b>	<b>Acumulado por Decêndio</b>
<b>Resolução temporal</b>	Alta (resolução diária).	Baixa (atualização a cada 10 dias).
<b>Sensibilidade às Variações</b>	Alta (detecta estiagens curtas).	Baixa (pode suavizar eventos críticos).
<b>Facilidade de Interpretação</b>	Média (requer cálculos diários).	Alta (acúmulo fixo facilita análise).
<b>Compatibilidade com modelos de previsão</b>	Alta (permite atualizações diárias).	Baixa (menor granularidade dos dados).
<b>Identificação de estiagens curtas</b>	Boa (janela de 5 dias); Ruim (janela de 10 dias).	Ruim (pode não detectar períodos secos e curtos).
<b>Suavização das oscilações extremas</b>	Baixa (variação diária pode ser alta).	Alto (valores fixos evitam oscilações diárias).
<b>Complexidade computacional</b>	Alta (requer processamento contínuo).	Baixo (fácil cálculo e agregação dos dados).

Fonte: Elaborado pela autora.

### 4.3 DETERMINAÇÃO DA ESTIAGEM DA CULTURA DA SOJA

É importante destacar que esses resultados foram obtidos após a remoção de *outliers*, garantindo que eventos extremos de precipitação não influenciassem a definição do limiar de estiagem. Além disso, a análise foi realizada exclusivamente no período de safra, compreendido entre setembro e março, abrangendo o histórico de precipitação de 1979 a 2023. Esse critério permitiu uma avaliação mais precisa das condições de precipitação dentro do intervalo fenológico da soja, excluindo períodos que não interferem diretamente no desenvolvimento da cultura. Com base nesses resultados, foi definida a estiagem da cultura da soja em Dourados, um conceito que combina as definições de estiagem e veranico, estabelecendo um limiar crítico de precipitação para caracterizar a ausência de chuvas durante o período produtivo da cultura.

A partir da estatística descritiva da precipitação acumulada, verificou-se que o valor médio do acumulado em 5 dias, após a remoção de *outliers*, foi de 20,6 mm. Esse valor foi utilizado como limiar para classificação da estiagem, ou seja, períodos em que a precipitação acumulada em 5 dias foi inferior a 20,6 mm foram considerados estiagens, enquanto valores acima desse limite foram classificados como não estiagem. Na Tabela 1 é exibida as estatísticas descritivas das janelas móveis.

Tabela 1 - Estatísticas descritivas das precipitações das janelas móveis.

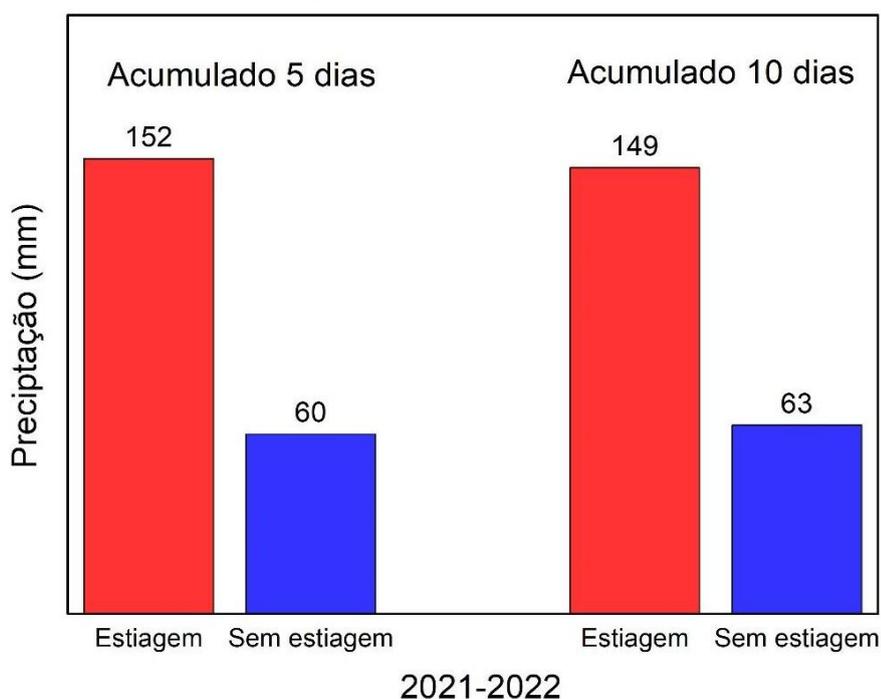
<b>Estatística Descritiva</b>	<b>Janela Móvel (5 dias acumulados em milímetros)</b>	<b>Janela Móvel (10 dias acumulados em milímetros)</b>
Contagem	9128,0	9200,0
Média	20,6	43,9
Desvio-padrão	22,2	35,6
Mínimo	0,0	0,0
25%	1,3	14,6
50%	13,4	37,1
75%	35,5	65,2
Máximo	89,2	150,4

Fonte: Elaborado pela autora.

A análise da safra 2021/2022 mostrou que, utilizando esse método, foram identificadas 152 janelas de estiagem para o acumulado de 5 dias e 149 para o acumulado de 10 dias, enquanto os períodos sem estiagem foram 60 e 63, respectivamente.

Isso significa que 71,7% das janelas móveis de 5 dias e 70,3% das janelas móveis de 10 dias foram classificadas como estiagem. Esses valores demonstram consistência entre as duas abordagens, reforçando a validade do critério adotado para a identificação da estiagem da cultura da soja em Dourados. O comparativo entre o quantitativo de estiagem verificadas no período nas janelas móveis de 5 e 10 dias está apresentado na Figura 11.

Figura 11 - Quantidade de janelas de estiagens no ano de 2021/2022 em Dourados/MS. Acumulado de 5 dias à esquerda e 10 dias à direita.



Fonte: Elaborado pela autora.

A similaridade entre os quantitativos obtidos nas janelas móveis de 5 e 10 dias sugere que a frequência de estiagens não é significativamente alterada pelo tamanho da janela de agregação da precipitação, o que reforça a robustez do método utilizado. Além disso, o fato da estiagem ter sido identificada em mais de 70% do período analisado confirma a severa irregularidade da distribuição de chuvas durante a safra 2021/2022, o que está alinhado com os impactos negativos na produtividade da soja.

Dessa forma, a estiagem da cultura da soja em Dourados proposta neste estudo se fundamenta na identificação quantitativa de períodos críticos de baixa precipitação, proporcionando um indicador objetivo para a caracterização de déficits pluviométricos durante a safra. A definição do limiar crítico de 20,6 mm em 5 dias representa um avanço metodológico na identificação da estiagem agrícola, permitindo seu uso para monitoramento climático e estudos de impacto na produtividade da cultura.

#### 4.4 MACHINE LEARNING PARA CLASSIFICAÇÃO DA ESTIAGEM DA CULTURA DA SOJA

A modelagem da estiagem da cultura da soja em Dourados foi realizada com base na classificação binária do fenômeno, considerando um limiar crítico de precipitação acumulada em cinco dias inferior a 20,6 mm. Esse valor foi definido a partir da estatística descritiva da série histórica de precipitação, após a remoção de *outliers* de precipitação.

Para garantir a integridade da análise e evitar vazamento de informações entre treino e teste, a base de dados foi segmentada da seguinte forma: o primeiro conjunto de dados abrange o período de 1º de junho de 1979 a 31 de agosto de 2021, sendo utilizado para o treinamento e teste dos modelos. O segundo conjunto, correspondente ao intervalo de 1º de setembro de 2021 a 31 de dezembro de 2023, foi reservado exclusivamente para a etapa de validação.

Dentro desse segundo conjunto, etapa de validação, foram definidos dois subconjuntos distintos: o primeiro, referente à safra 2021/2022, compreende o período de 1º de setembro de 2021 a 31 de março de 2022, caracterizado pela maior incidência de estiagem; o segundo, relativo à safra 2022/2023, engloba os dados meteorológicos de 1º de setembro de 2022 a 31 de março de 2023, correspondendo a uma safra favorável. A separação desses períodos possibilita a comparação entre anos com diferentes níveis de severidade climática.

Para a construção do modelo, foram utilizadas variáveis derivadas da precipitação e fatores temporais, como mês e dia do ano, além da razão entre os acumulados de 5 e 10 dias, permitindo captar variações na distribuição das chuvas. Para realizar esta análise, foi construída uma matriz de correlação, exibida na Figura 12.

Figura 12 - Matriz de correlação entre as variáveis.

<b>Data</b>	-												
<b>Temperatura Média</b>	0,11	-											
<b>Temperatura Máxima</b>	0,09	0,92	-										
<b>Temperatura Mínima</b>	0,04	0,88	0,69	-									
<b>Chuva (mm)</b>	0,00	-0,06	-0,16	0,10	-								
<b>Acumulado de 5 dias (mm)</b>	-0,01	-0,03	-0,15	0,12	0,48	-							
<b>Acumulado de 10 dias (mm)</b>	-0,01	0,06	-0,07	0,19	0,36	0,75	-						
<b>Ano</b>	1,00	0,11	0,09	0,04	0,00	0,00	-0,01	-					
<b>Mês</b>	0,01	-0,03	0,00	-0,10	-0,01	-0,02	-0,03	-0,02	-				
<b>Dia do ano</b>	0,01	-0,03	0,00	-0,10	-0,01	-0,02	-0,03	-0,02	-0,02	-			
<b>Varição acumulada de 5 e 10 dias</b>	-0,01	-0,11	-0,21	0,04	0,31	0,62	0,23	-0,01	-0,03	-0,03	-		
<b>Estiagem</b>	0,02	0,03	0,14	-0,11	-0,38	-0,78	-0,60	0,02	0,02	0,02	-0,62	-	
	<b>Data</b>	<b>Temp. Média</b>	<b>Temp. Máx.</b>	<b>Temp. Mín.</b>	<b>Chuva (mm)</b>	<b>Acumulado de 5 dias (mm)</b>	<b>Acumulado de 10 dias (mm)</b>	<b>Ano</b>	<b>Mês</b>	<b>Dia do ano</b>	<b>Varição acumulada de 5 e 10 dias</b>	<b>Estiagem</b>	

Fonte: Elaborado pela autora.

A matriz de correlação indicou que a variável estiagem possui uma relação negativa expressiva com os acumulados de precipitação, sendo de -0,78 para o acumulado de 5 dias e -0,60 para o acumulado de 10 dias, demonstrando a relevância desses atributos na classificação do fenômeno.

Além disso, a variável chuva apresentou correlação negativa de -0,38 com

estiagem, indicando que sua inclusão no modelo contribui para a identificação de padrões de estiagem. A razão entre os acumulados de 5 e 10 dias também apresentou correlação negativa moderada com estiagem (-0,62), reforçando sua importância na análise. Por outro lado, variáveis como temperatura máxima, temperatura mínima e temperatura média apresentaram correlações muito baixas com estiagem.

Diante desses resultados, a seleção de variáveis foi conduzida com base nos coeficientes de correlação, resultando na escolha das seguintes variáveis para compor o modelo de *machine learning* para classificação de estiagem da soja: precipitação diária (chuva), ano, mês, dia do ano, variação entre os acumulados de 5 e 10 dias e temperatura média, sendo esta última incluída por elevar os níveis de performance do modelo e ser uma variável relevante aos estudos de estiagem. A exclusão das demais variáveis buscou reduzir a complexidade do modelo, mantendo apenas os atributos com maior relevância para a classificação da estiagem.

A avaliação dos modelos de aprendizado de máquina voltados à classificação da estiagem na cultura da soja em Dourados/MS considerou um conjunto diversificado de algoritmos, cuja performance foi comparada, inicialmente, com base em cinco métricas: acurácia, acurácia balanceada, área sob a curva ROC (ROC AUC), *F1-score* e tempo de execução.

Essas métricas permitem uma análise abrangente da eficácia de cada modelo, especialmente em contextos de classes desbalanceadas, como é o caso da ocorrência de estiagem. Entre os algoritmos avaliados, destacaram-se o *Extra Trees Classifier*, *XGB Classifier* e *Random Forest Classifier*, que apresentaram os melhores resultados em termos de desempenho e estabilidade.

A acurácia indica o percentual de acertos totais do modelo, enquanto a acurácia balanceada corrige possíveis distorções causadas por desequilíbrios nas classes. O ROC AUC mede a capacidade do modelo de distinguir entre estiagem e não estiagem, e o *F1-score* reflete o equilíbrio entre precisão e sensibilidade. O tempo de execução, por sua vez, expressa a eficiência computacional.

A Tabela 2 sintetiza os resultados gerais obtidos, evidenciando o desempenho superior dos modelos baseados em árvores de decisão que foram analisados individualmente, que combinaram alta acurácia com tempos de execução compatíveis com aplicações práticas.

Tabela 2 - Performance de diferentes modelos de classificação com base em várias métricas.

<i>Model</i>	<i>Accuracy</i>	<i>Balanced Accuracy</i>	<i>ROC AUC</i>	<i>F1 Score</i>	<i>Time Taken (s)</i>
<i>ExtraTreesClassifier</i>	0,88	0,89	0,89	0,89	1,86
<i>XGBClassifier</i>	0,87	0,87	0,87	0,87	0,18
<i>RandomForestClassifier</i>	0,87	0,87	0,87	0,87	1,85
<i>BaggingClassifier</i>	0,86	0,86	0,86	0,86	0,64
<i>LGBMClassifier</i>	0,86	0,86	0,86	0,86	0,20
<i>ExtraTreeClassifier</i>	0,84	0,84	0,84	0,84	0,04
<i>SVC</i>	0,80	0,83	0,83	0,80	4,71
<i>NuSVC</i>	0,80	0,83	0,83	0,80	10,16
<i>DecisionTreeClassifier</i>	0,84	0,83	0,83	0,84	0,15
<i>KNeighborsClassifier</i>	0,82	0,83	0,83	0,82	0,74
<i>LabelSpreading</i>	0,82	0,82	0,82	0,82	13,63
<i>LabelPropagation</i>	0,82	0,82	0,82	0,82	18,44
<i>AdaBoostClassifier</i>	0,78	0,82	0,82	0,78	0,57
<i>SGDClassifier</i>	0,79	0,82	0,82	0,80	0,11
<i>PassiveAggressiveClassifier</i>	0,77	0,81	0,81	0,78	0,03
<i>LinearSVC</i>	0,80	0,80	0,80	0,80	0,04
<i>CalibratedClassifierCV</i>	0,80	0,80	0,80	0,80	0,22
<i>RidgeClassifier</i>	0,79	0,80	0,80	0,80	0,03
<i>RidgeClassifierCV</i>	0,79	0,80	0,80	0,80	0,10
<i>LinearDiscriminantAnalysis</i>	0,79	0,80	0,80	0,80	0,05
<i>BernoulliNB</i>	0,79	0,80	0,80	0,79	0,04
<i>LogisticRegression</i>	0,80	0,80	0,80	0,80	0,07
<i>NearestCentroid</i>	0,79	0,79	0,79	0,80	0,03
<i>QuadraticDiscriminantAnalysis</i>	0,80	0,77	0,77	0,80	0,03
<i>Perceptron</i>	0,78	0,77	0,77	0,78	0,04
<i>GaussianNB</i>	0,80	0,76	0,76	0,79	0,06
<i>DummyClassifier</i>	0,33	0,50	0,50	0,17	0,03

Fonte: Elaborado pela autora.

Os três modelos foram treinados e testados separadamente e com um conjunto de dados balanceado utilizando a técnica SMOTE, garantindo uma distribuição equitativa (Chawla *et al.*, 2011) entre as classes "estiagem" e "não estiagem". As

métricas analisadas foram acurácia, precisão, *recall* e *F1-score*, permitindo uma comparação objetiva do desempenho de cada modelo.

O resultado do algoritmo *Random Forest Classifier* é apresentado na Tabela 3, na qual se observa uma acurácia de 96,24% no treinamento e de 85,79% no teste, evidenciando que o desempenho do modelo foi reduzido quando aplicado a novos dados. A precisão para classificar a estiagem alcançou 94%, enquanto a sensibilidade (*recall*) foi de 84%.

Em termos práticos, isso significa que o modelo identificou corretamente 84% dos casos reais de estiagem, mas deixou de detectar aproximadamente 16%, configurando falsos negativos. Para a classe “sem estiagem”, a precisão foi de 74% e a sensibilidade (*recall*) de 89%, sugerindo maior dificuldade do algoritmo em prever corretamente os dias sem estiagem. A matriz de confusão ilustrada na Figura 13 confirma essa tendência, evidenciando um número expressivo de falsos negativos.

A análise das matrizes de confusão mostra que o modelo não atende plenamente ao critério de segurança esperado para o produtor. No conjunto de treinamento, foram registrados 488 falsos negativos — situações em que a estiagem ocorreu, mas o modelo classificou como “sem estiagem” — contra apenas 53 falsos positivos, que representam casos em que o modelo indicou estiagem inexistente. No conjunto de teste, os falsos negativos totalizaram 490, enquanto os falsos positivos foram 168.

Esses resultados demonstram que o algoritmo erra quase três vezes mais por não identificar a estiagem quando ela realmente ocorre do que por emitir alertas incorretos. Tal comportamento é crítico, pois compromete a confiabilidade do modelo em cenários decisivos para o manejo agrícola.

Os indicadores reforçam essa interpretação: no teste, a sensibilidade (*recall classe 1*) para a classe “com estiagem” foi de 0,84, o que equivale a uma taxa de falsos negativos de aproximadamente 16%, enquanto a especificidade (*recall classe 0*) da classe “sem estiagem” foi de 0,89, correspondente a uma taxa de falsos positivos em torno de 11%.

Embora a precisão para a classe “com estiagem” tenha sido elevada (0,94), o valor classificado negativo apresentou-se relativamente baixo ( $\approx 0,74$ ). Isso significa que, quando o modelo indica “sem estiagem”, há uma chance considerável de erro, o que representa maior risco ao produtor do que o recebimento de um alerta falso.

Diante desse cenário, e considerando que no contexto agrícola a segurança do

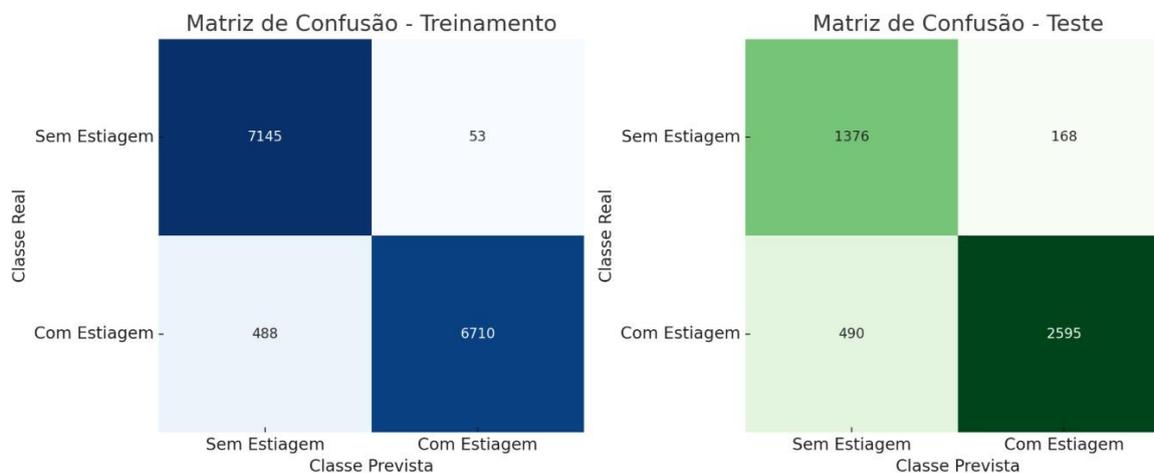
produtor deve ser priorizada, a configuração do classificador deve buscar a redução dos falsos negativos, ainda que isso resulte em um número maior de falsos positivos.

Tabela 3 - Relatório de resultados da aplicação do algoritmo Random Forest Classifier.

Métrica	Treinamento	Teste	Diferença (%)
Acurácia	96,24%	85,79%	↓ 10,45%
Precisão - Classe 0 (sem estiagem)	0,94	0,74	↓ 20 pontos
Precisão - Classe 1 (com estiagem)	0,99	0,94	↓ 5 pontos
Recall - Classe 0 (sem estiagem)	0,99	0,89	↓ 10 pontos
Recall - Classe 1 (com estiagem)	0,93	0,84	↓ 9 pontos
<i>FI-score</i> - Classe 0 (sem estiagem)	0,96	0,81	↓ 15 pontos
<i>FI-score</i> - Classe 1 (com estiagem)	0,96	0,89	↓ 7 pontos
Média <i>FI-score</i>	0,96	0,85	↓ 11 pontos

Fonte: Elaborado pela autora

Figura 13 - Matriz de confusão da aplicação do algoritmo *Random Forest Classifier*.



Fonte: Elaborado pela autora.

O resultado do algoritmo *Extra Trees Classifier* é apresentado na Tabela 4, no qual a acurácia atingiu 91,84% no treinamento e 83,75% no teste, revelando uma redução de aproximadamente 8 pontos percentuais quando o modelo foi aplicado a novos dados. A precisão para classificar estiagem foi de 96%, enquanto a sensibilidade (recall) foi de 79%. Isso significa que o modelo identificou corretamente 79% dos casos reais de estiagem, mas deixou de detectar cerca de 21%, configurando falsos negativos. Já para a classe “sem estiagem”, a precisão foi de 69% e a sensibilidade de 93%, indicando maior dificuldade do modelo em classificar corretamente os dias sem estiagem. A matriz de confusão ilustrada na Figura 14 confirma esse comportamento.

A análise detalhada das matrizes de confusão mostra que o modelo apresenta limitações quanto ao critério de segurança esperado para o produtor. No conjunto de treinamento, foram observados 124 falsos negativos — situações em que a estiagem ocorreu, mas o modelo classificou como “sem estiagem” — contra 1.051 falsos positivos, nos quais houve indicação de estiagem sem que ela tivesse ocorrido. Já no conjunto de teste, os falsos negativos chegaram a 112, enquanto os falsos positivos totalizaram 640. Esse resultado demonstra que, embora o algoritmo apresente desempenho equilibrado entre as métricas, ele tende a errar mais ao gerar falsos alarmes de estiagem do que ao deixar de detectar estiagens reais, comportamento distinto do observado no *Random Forest*.

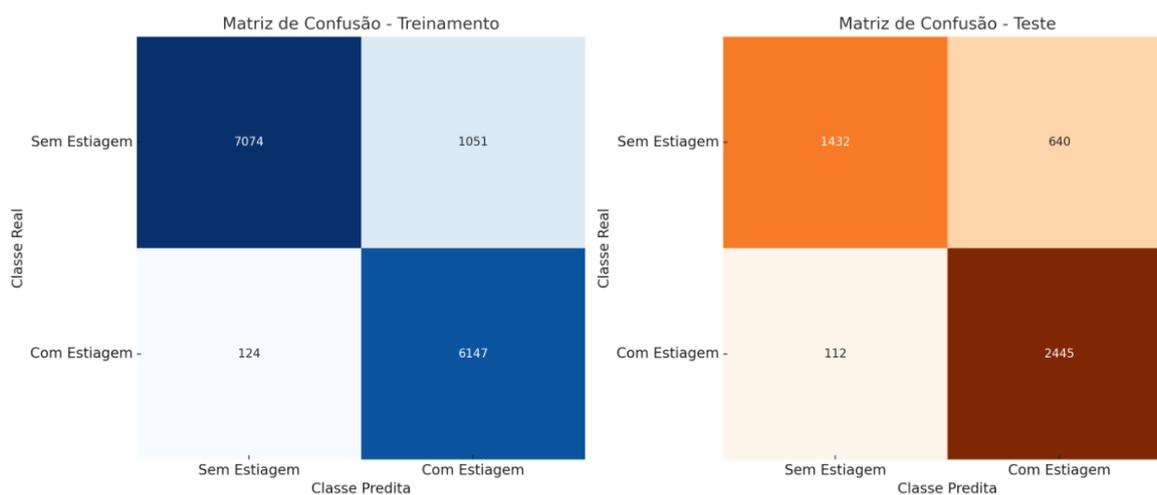
Os indicadores corroboram essa interpretação: no teste, a sensibilidade da classe “com estiagem” foi de 0,79, representando uma taxa de falsos negativos de aproximadamente 21%, enquanto a especificidade da classe “sem estiagem” foi de 0,93, correspondente a uma taxa de falsos positivos em torno de 7%. Embora a precisão da classe “com estiagem” tenha se mantido elevada (0,96), o desempenho mais baixo na classe “sem estiagem” (precisão de 0,69) comprometeu a capacidade do modelo em distinguir corretamente os dias sem estiagem.

Nesse sentido, ainda que o *Extra Trees Classifier* tenha apresentado resultados consistentes, seu comportamento de erro é menos alinhado ao objetivo de segurança do produtor do que o desejável. Isso ocorre porque, ao privilegiar a redução dos falsos negativos em relação ao *Random Forest*, o modelo passa a incorrer em maior número de falsos positivos, o que pode gerar excesso de alertas indevidos. Embora tais alertas não representem risco direto à produção, podem comprometer a confiança do usuário no sistema de monitoramento.

Tabela 4 - Relatório de resultados da aplicação do algoritmo *Extra Trees Classifier*.

Métrica	Treinamento	Teste	Diferença (%)
Acurácia	91,84%	83,75%	↓ 8,09%
Precisão - Classe 0 (sem estiagem)	0,87	0,69	↓ 18 pontos
Precisão - Classe 1 (com estiagem)	0,98	0,96	↓ 2 pontos
Recall - Classe 0 (sem estiagem)	0,98	0,93	↓ 5 pontos
Recall - Classe 1 (com estiagem)	0,85	0,79	↓ 6 pontos
F1-score - Classe 0 (sem estiagem)	0,92	0,79	↓ 13 pontos
F1-score - Classe 1 (com estiagem)	0,91	0,87	↓ 4 pontos
Média F1-score	0,92	0,83	↓ 9 pontos

Fonte: Elaborado pela autora.

Figura 14 - Matriz de confusão da aplicação do algoritmo *Extra Trees Classifier*.

Fonte: Elaborado pela autora

O resultado do algoritmo *XGB Classifier* é apresentado na Tabela 5, na qual se observa acurácia de 96,90% no treinamento e 87,36% no teste, evidenciando uma redução de aproximadamente 9,5 pontos percentuais quando o modelo foi aplicado a novos dados. A precisão para prever estiagem foi de 94%, enquanto a sensibilidade (*recall*) foi de 87%. Isso significa que o modelo identificou corretamente 87% dos casos reais de estiagem, mas deixou de detectar cerca de 13%, configurando falsos negativos. Para a classe “sem estiagem”, a precisão foi de 77% e a sensibilidade de 88%, demonstrando maior dificuldade na correta identificação dos dias sem estiagem. A matriz de confusão ilustrada na Figura 15 confirma essa tendência, revelando tanto a ocorrência de falsos negativos quanto de falsos positivos.

A análise detalhada das matrizes de confusão reforça essa interpretação. No conjunto de treinamento, foram observados 108 falsos negativos — situações em que a estiagem ocorreu, mas o modelo classificou como “sem estiagem” — contra 338 falsos positivos, em que o modelo sinalizou estiagem sem que ela tivesse ocorrido. Já no conjunto de teste, os falsos negativos chegaram a 182, enquanto os falsos positivos foram 403. Esses resultados indicam que, diferentemente do *Extra Trees*, o *XGB Classifier* apresenta um equilíbrio maior entre os dois tipos de erro, mas ainda assim mantém a tendência de falhar ao não identificar estiagens reais, o que compromete a confiabilidade do sistema em cenários críticos.

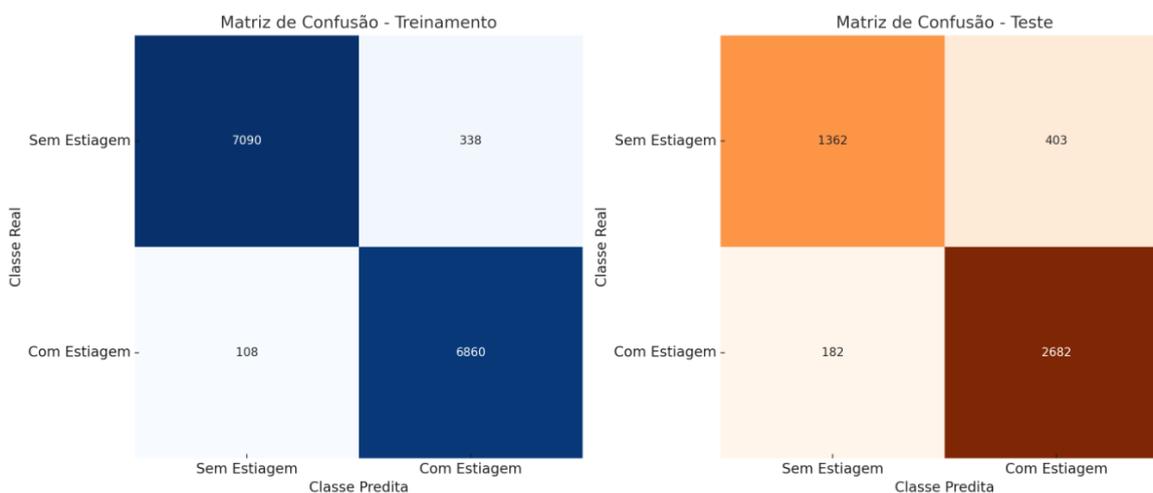
Os indicadores quantitativos corroboram essa análise: no teste, a sensibilidade para a classe “com estiagem” foi de 0,87, o que corresponde a uma taxa de falsos negativos em torno de 13%, enquanto a especificidade para a classe “sem estiagem” foi de 0,88, indicando uma taxa de falsos positivos próxima de 12%. Apesar da elevada precisão da classe “com estiagem” (0,94), o desempenho mais modesto na classe “sem estiagem” (precisão de 0,77) impacta a confiabilidade das previsões.

Mesmo assim, o *XGB Classifier* demonstrou desempenho superior às demais abordagens nas métricas globais, conciliando alta acurácia, boa precisão e elevada sensibilidade. Contudo, seu padrão de erro ainda exige atenção, especialmente no que diz respeito à redução dos falsos negativos, pois a falha em identificar uma estiagem real representa maior risco ao produtor do que a emissão de alertas indevidos.

Tabela 5 - Relatório de resultados da aplicação do algoritmo *XGB Classifier*.

Métrica	Treinamento	Teste	Diferença (%)
Acurácia	96,90%	87,36%	↓ 9,54%
Precisão - Classe 0 (sem estiagem)	0,95	0,77	↓ 18 pontos
Precisão - Classe 1 (com estiagem)	0,98	0,94	↓ 4 pontos
Recall - Classe 0 (sem estiagem)	0,98	0,88	↓ 10 pontos
Recall - Classe 1 (com estiagem)	0,95	0,87	↓ 8 pontos
<i>F1-score</i> - Classe 0 (sem estiagem)	0,97	0,82	↓ 15 pontos
<i>F1-score</i> - Classe 1 (com estiagem)	0,97	0,90	↓ 7 pontos
Média <i>F1-score</i>	0,97	0,86	↓ 11 pontos

Fonte: Elaborado pela autora.

Figura 15 - Matriz de confusão da aplicação do algoritmo *XGB Classifier*.

Fonte: Elaborado pela autora.

Em síntese, enquanto o *Random Forest* apresenta desempenho elevado, mas falha de forma mais crítica ao não identificar estiagens, e o *Extra Trees* reduz os falsos negativos ao custo de um excesso de falsos positivos, o *XGB Classifier* se consolida como a alternativa mais adequada.

Sua capacidade de manter elevado desempenho global, ao mesmo tempo em que diminui de forma relativa as falhas em estiagens reais, o torna mais alinhado ao objetivo central deste estudo: classificar a ocorrência de estiagens na cultura da soja em Dourados (MS), aprimorando o monitoramento climático e fornecendo subsídios mais confiáveis à tomada de decisão dos stakeholders do setor agrícola, mesmo que isso resulte em maior número de alertas preventivos.

#### 4.5 VALIDAÇÃO DO MODELO DE MACHINE LEARNING PARA ESTIAGEM DA CULTURA DE SOJA

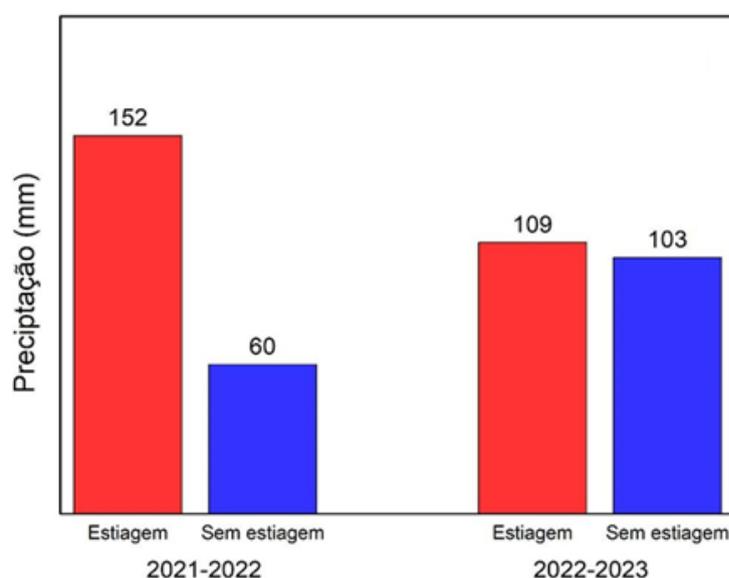
Para validar os modelos de classificação da estiagem na cultura da soja, foram utilizados dados correspondentes às safras de 2021/2022 (de 1º de setembro de 2021 a 31 de março de 2022) e 2022/2023 (de 1º de setembro de 2022 a 31 de março de 2023). Esse procedimento permitiu avaliar a capacidade dos algoritmos em identificar estiagens em um período não incluído no treinamento, garantindo a análise de sua eficácia em diferentes condições climáticas.

A escolha dessas safras foi motivada por seus padrões distintos de

precipitação: 2021/2022 foi marcada por maior ocorrência de estiagens, enquanto 2022/2023 apresentou uma distribuição mais regular das chuvas. Dessa forma, a validação possibilitou testar a robustez dos modelos em cenários contrastantes.

O gráfico apresentado na Figura 16 indica a quantidade real de janelas móveis de 5 dias com estiagem e sem estiagem ocorridas em cada safra, enquanto na Tabela 6 são comparados esses valores com as estimativas dos modelos.

Figura 16 - Quantitativo real de janelas com estiagem à esquerda e sem estiagem à direita em Dourados/MS no acumulado de 5 dias de precipitação.



Fonte: Elaborado pela autora.

Na safra 2021/2022, os modelos indicaram uma predominância de estiagens, o que corrobora com a análise empírica da distribuição da precipitação no período. O *XGB Classifier* apresentou a melhor estimativa, com 72% das janelas classificadas como estiagem, aproximando-se do valor real. O *Random Forest Classifier* identificou 67% das janelas como estiagem, enquanto o *Extra Tree Classifier* foi mais conservador, com 57%. Esses valores estão alinhados com a distribuição observada na série histórica, em que o acumulado de precipitação apresentou falhas prolongadas durante a fase crítica do ciclo da soja, afetando diretamente o desenvolvimento da cultura.

Para a safra 2022/2023, que apresentou um regime de chuvas mais regular, houve uma redução na frequência de estiagens classificadas pelos modelos. O *Random Forest Classifier* identificou estiagem em 56% das janelas móveis de 5 dias, o *Extra Trees Classifier* estimou 47%, e o *XGB Classifier* apontou 63%. A redução na

classificação de estiagens pelos modelos acompanha a menor incidência observada no período, evidenciando a maior regularidade pluviométrica da safra 2022/2023. Os resultados são mostrados na Tabela 6.

Tabela 6 - Comparativo dos modelos de aprendizado de máquina para classificação de estiagem na soja em Dourados/MS.

Safra	Estiagem Observada	Estiagem Estimada - RandomForest (%)	Estiagem Estimada - ExtraTrees (%)	Estiagem Estimada - XGBClassifier (%)
2021/2022	152 (71,7%)	141 (67%)	121 (57%)	153 (72%)
2022/2023	109 (51,4%)	118 (56%)	100 (47%)	133 (63%)

Fonte: Elaborado pela autora.

A análise comparativa entre as safras revela que os modelos foram mais precisos na identificação das estiagens em um ano com menor precipitação, enquanto em uma safra com chuvas mais distribuídas, houve uma leve superestimação da estiagem. Esse comportamento pode estar relacionado ao viés dos modelos, treinados com dados históricos em que a estiagem foi predominante e a falta de treinamento com novos dados, já que foram reservados para validação.

De qualquer forma, a abordagem baseada em janelas móveis de 5 dias mostrou-se eficaz para capturar estiagens em condições adversas, mas pode apresentar um maior número de falsos positivos em anos de precipitação mais equilibrada.

#### 4.6 DEFINIÇÃO DO MODELO DE CLASSIFICAÇÃO DA ESTIAGEM DA CULTURA DA SOJA

A adoção do algoritmo *XGB Classifier* como modelo principal neste estudo justifica-se por sua elevada capacidade de capturar relações não lineares, identificar padrões complexos e lidar com variáveis altamente correlacionadas — características comuns em séries temporais climáticas. Diferentemente dos modelos econométricos tradicionais, como a regressão logística, regressão linear e classificadores lineares, os algoritmos baseados em árvores, como o *XGBoost*, dispensam suposições paramétricas rígidas e se adaptam melhor à natureza dos dados ambientais.

Diversos estudos reforçam essa vantagem dos modelos de *machine learning*, especialmente em contextos agroclimáticos. Jeong et al. (2016) evidenciaram a

superioridade dos algoritmos baseados em árvores, como o *Random Forest*, na previsão de produtividade agrícola em escala regional e global. Ao comparar *Random Forest* com a regressão linear múltipla (MLR) aplicada às culturas de trigo, milho e batata, os autores constataram que o *Random Forest* apresentou desempenho consistentemente superior em todos os cenários avaliados. Os valores de RMSE obtidos com o *Random Forest* variaram entre 6% e 14% da produtividade média observada, enquanto os modelos MLR apresentaram erros significativamente maiores, entre 14% e 49%.

De forma complementar, Crane-Droesch (2018), por exemplo, propôs uma rede neural semiparamétrica para previsão de produtividade agrícola nos Estados Unidos, utilizando dados de clima, solo e rendimento do milho. O estudo comparou o desempenho de diferentes abordagens de modelagem para prever a produtividade agrícola, incluindo regressões lineares tradicionais (OLS), redes neurais totalmente não paramétricas e redes neurais semiparamétricas (SNN). Os resultados demonstraram que o modelo SNN apresentou o menor erro médio quadrático (251,5), superando tanto a regressão paramétrica (334,4) quanto a rede neural totalmente não paramétrica (638,6).

Complementarmente, o estudo de Ravi e Baranidharan (2020) reforça a superioridade do algoritmo *XGBoost* em relação a outros métodos de aprendizado de máquina na previsão da produtividade agrícola. Ao utilizar dados de precipitação, temperatura e área cultivada para prever a produção agrícola em quatro estados do sul da Índia, os autores compararam o desempenho de cinco algoritmos para execução do modelo: regressão linear, regressão do vetor de suporte, árvore de decisão, *Random Forest* e *XGBoost*. O *XGBoost* obteve o melhor desempenho em todas as métricas avaliadas, com coeficiente de determinação ( $R^2$ ) de 0,9391, superando inclusive o *Random Forest*, que apresentou  $R^2$  de 0,9314. Além do alto  $R^2$ , o *XGBoost* também apresentou os menores valores de erro médio quadrático (MSE) e erro absoluto médio (MAE), indicando maior precisão na previsão e menor desvio entre os valores estimados e observados. Esses resultados demonstram a robustez do *XGBoost* na modelagem de variáveis agrícolas, reforçando sua escolha como base metodológica do presente estudo.

Mais recentemente, um exemplo relevante que compara métodos econométricos tradicionais e algoritmos de *machine learning* no contexto agrícola é o estudo de Du *et al.* (2025). Os autores analisaram a adoção e o abandono de práticas

conservacionistas, como o uso de culturas de cobertura, entre produtores rurais no estado de Iowa (EUA). O objetivo foi avaliar a capacidade preditiva de diferentes abordagens para compreender o comportamento dos agricultores diante de decisões ambientais estratégicas. O estudo comparou diretamente a performance da regressão logística, amplamente empregada em análises econômicas, com a do algoritmo *Random Forest*, representativo dos métodos de *machine learning*. Os resultados revelaram que o *Random Forest* alcançou uma acurácia de 92,1%, superando os 85,1% obtidos pela regressão logística. Além disso, o modelo de árvore apresentou maior robustez na presença de *outliers* e melhor capacidade de capturar interações não lineares e heterogeneidade entre os produtores, elementos essenciais em contextos agrícolas marcados por variabilidade socioeconômica e ambiental.

No presente estudo, os resultados obtidos também corroboram essas evidências. Os algoritmos baseados em árvores — especialmente o *XGB Classifier*, seguido por *Random Forest* e *Extra Trees* — apresentaram as melhores performances nas tarefas de classificação de estiagem. O *XGB Classifier* obteve acurácia de 87,36% e *F1-score* de 0,90 para a classe “estiagem”, com o menor índice de falsos negativos entre todos os modelos testados. O *Random Forest Classifier* apresentou acurácia de 85,79% e *F1-score* de 0,89, enquanto o *Extra Trees Classifier* alcançou 83,75% de acurácia e *F1-score* de 0,87.

Por outro lado, os modelos econométricos e estatísticos lineares testados — já apresentados anteriormente na Tabela 2 que está na Seção 4.4 tais como: como regressão logística (*Logistic Regression*), classificador de crista (*Ridge Classifier*), classificador de gradiente estocástico (*Stochastic Gradient Descent Classifier – SGDClassifier*), máquina de vetores de suporte linear (*Linear Support Vector Classifier – LinearSVC*) e análise discriminante linear (*Linear Discriminant Analysis – LinearDiscriminantAnalysis*) — apresentaram desempenhos consideravelmente inferiores.

A acurácia desses algoritmos variou entre 78% e 80%, com *F1-scores* também próximos de 0,80, além dos valores de áreas sob a curva ROC (AUC) mais baixas, geralmente próximas de 0,80, quando comparadas às métricas superiores dos algoritmos baseados em árvores. Devido a esse desempenho inicial significativamente inferior e à baixa capacidade de modelar padrões não lineares e multivariados, esses algoritmos para o modelo de classificação de estiagem foram descartados para análises mais aprofundadas neste estudo, por não apresentarem resultados comparáveis em

termos de precisão, equilíbrio entre classes e robustez diante da complexidade dos dados climáticos envolvidos.

Dessa forma, a escolha do algoritmo *XGB Classifier* como modelo final não apenas se fundamenta em seu desempenho superior nos experimentos realizados, como também está em consonância com os avanços da literatura recente, consolidando-se como a alternativa mais robusta, precisa e adequada ao problema de classificação de estiagem na sojicultura na região de Dourados/MS.

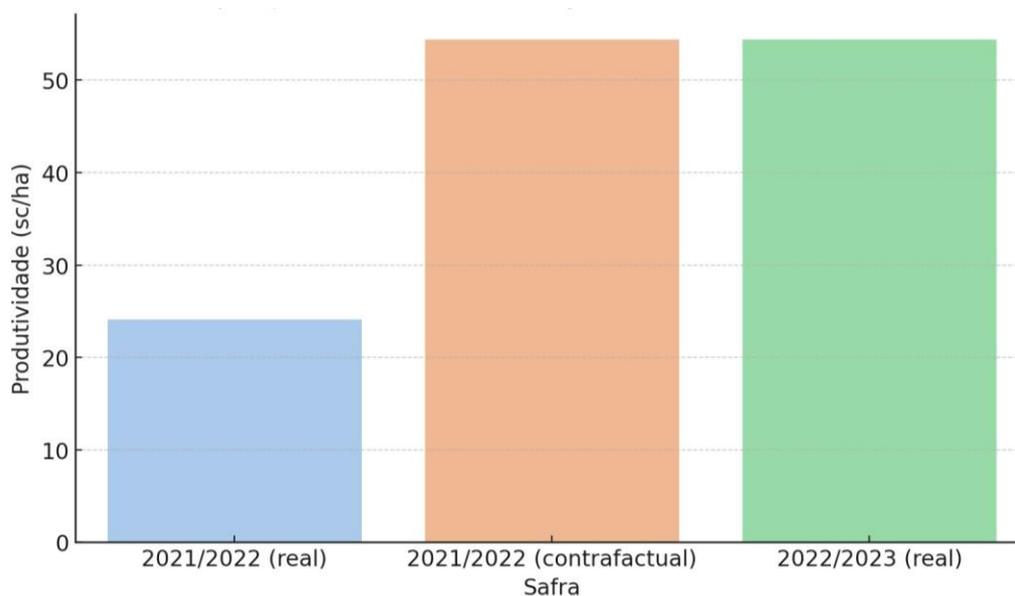
#### 4.7 ANÁLISE CONTRAFACTUAL: IMPACTO DA ESTIAGEM NA SAFRA 2021/2022

A safra de soja 2021/2022 em Dourados/MS foi marcada por um dos piores desempenhos produtivos da última década, registrando apenas 24,08 sacas por hectare (sc/ha), enquanto a média estadual foi de 38,05 sc/ha, segundo dados da APROSOJA-MS (2022). Essa expressiva queda de produtividade foi atribuída, principalmente, à estiagem severa associada ao fenômeno *La Niña*, que ocasionou precipitações entre 40% e 60% abaixo do esperado, além de prolongados períodos de estresse térmico (APROSOJA-MS, 2022).

Com o intuito de mensurar o impacto direto da estiagem sobre a produtividade, foi realizada uma análise contrafactual. Considerou-se um cenário hipotético no qual as condições climáticas da safra 2021/2022 tivessem se assemelhado àquelas observadas em 2022/2023, reconhecida por sua distribuição regular de chuvas. Para essa safra posterior, Dourados/MS alcançou produtividade média de 54,43 sc/ha (APROSOJA-MS, 2023), valor utilizado como parâmetro de comparação.

A simulação indicou que, na ausência da estiagem severa, a produtividade de 2021/2022 poderia ter se aproximado ao patamar observado em 2022/2023, sugerindo um prejuízo estimado de 30,35 sc/ha. Essa diferença representa uma redução de aproximadamente 56% no rendimento potencial da cultura, evidenciando os efeitos adversos da irregularidade pluviométrica sobre o desempenho da soja. A Figura 17 resume visualmente essa comparação entre o cenário real e o contrafactual.

Figura 17 – Comparação da produtividade da soja em Dourados/MS: cenário real vs. contrafactual.



Fonte: Elaborado pela autora com dados da APROSOJA-MS (2022; 2023).

Além da análise da produtividade, estimou-se o impacto financeiro da perda observada. Considerando o preço médio da saca de soja em Dourados no ano de 2022, de R\$ 177,40 (APROSOJA-MS, 2022), a perda de 30,35 sc/ha corresponde a um prejuízo de aproximadamente R\$ 5.384,09 por hectare.

Em uma área de 100 hectares, a perda financeira totalizaria cerca de R\$ 538.409,00. Em escala estadual, estimativas apontaram que a estiagem da safra 2021/2022 causou perdas superiores a R\$ 3 bilhões, com redução de mais de um milhão de toneladas na produção de soja em Mato Grosso do Sul (Campo Grande News, 2022).

Nesse cenário, o modelo de classificação de estiagem proposto neste estudo se apresenta como uma ferramenta estratégica para a tomada de decisão agrônômica. Ao identificar janelas críticas de precipitação, o modelo oferece subsídios técnicos para intervenções mais precisas, como a aplicação de irrigação suplementar em estádios fenológicos sensíveis da cultura, a exemplo do enchimento de grãos.

Para avaliar a viabilidade dessa estratégia, foi considerada a estimativa de custo da soja irrigada na região sul de Mato Grosso do Sul na safra 2015/2016, calculada em R\$ 3.108,32 por hectare (Richetti *et al.*, 2015). Embora não represente dados recentes, trata-se da referência mais atualizada e regionalmente específica disponível. A simulação considerou a aplicação seletiva da irrigação em apenas 40%

de uma área de 100 hectares, definidos como talhões críticos com base na interpretação do produtor a partir das indicações do modelo, resultando em um custo total estimado de R\$ 124.332,80.

Admitindo-se a recuperação de aproximadamente um terço da perda total estimada de 30,35 sc/ha, o que corresponde a cerca de 10 sacas por hectare irrigado, a receita bruta adicional alcançaria R\$ 709.600,00. Subtraído o custo da irrigação, o lucro líquido estimado seria de R\$ 585.267,20. O retorno sobre o investimento (ROI), calculado pela razão entre o lucro líquido e o valor investido, resulta em aproximadamente 470%. Esse resultado indica que, para cada real investido, haveria um retorno líquido de R\$ 4,70, evidenciando a alta rentabilidade da irrigação direcionada como medida mitigadora em cenários de estiagem, mesmo quando aplicada de forma seletiva.

Esses valores reforçam a viabilidade econômica da aplicação seletiva da irrigação com base no modelo de classificação proposto, destacando seu papel como ferramenta de apoio à tomada de decisão no campo. Logo, a simulação econômica está apresentada na Tabela 7.

Tabela 7 - Simulação econômica da aplicação seletiva de irrigação baseada na classificação de estiagem.

<b>Item</b>	<b>Valor estimado</b>
Área total da propriedade	100 hectares
Percentual da área irrigada	40% (40 hectares)
Custo médio por hectare irrigado <sup>1</sup>	R\$ 3.108,32
Custo total da irrigação	R\$ 124.332,80
Produtividade recuperada com irrigação	400 sacas
Valor da saca de soja <sup>2</sup>	R\$ 177,40
Receita adicional	R\$ 709.600,00
Lucro líquido estimado (receita adicional – custo de irrigação)	R\$ 585.267,20
<b>Retorno sobre o Investimento (ROI)</b>	<b>≈ 470%</b>

Fonte: Elaborado pela autora com base em Richetti *et al.* (2015)<sup>1</sup> e APROSOJA-MS (2022)<sup>2</sup>.

Já a quantidade ideal de irrigação a ser aplicada, contudo, não é fixa. Com base na abordagem de janelas móveis de cinco dias, recomenda-se que o produtor utilize o acumulado pluviométrico recente em comparação ao limiar de 20,6 mm como

parâmetro técnico. Sempre que a precipitação acumulada nesse intervalo estiver abaixo desse valor, especialmente em fases críticas do ciclo da soja, a irrigação suplementar deve ser considerada. A lâmina a ser aplicada pode ser ajustada conforme as previsões meteorológicas de curto prazo, o estágio fenológico da cultura e os registros recentes de precipitação, otimizando o uso da água e evitando desperdícios.

Além do monitoramento das chuvas, as características edafoclimáticas da região devem ser incorporadas à avaliação da necessidade de irrigação. Em Dourados/MS, predomina o Latossolo Vermelho distroférrico, um solo profundo, muito argiloso, com capacidade de água disponível (CAD) estimada em cerca de 83 milímetros por metro de profundidade (Amaral *et al.*, 2000). Embora essa capacidade de retenção hídrica seja elevada, sua efetividade depende da reposição contínua de umidade, o que está diretamente ligado à regularidade e ao volume das chuvas.

Quando o acumulado pluviométrico em janelas móveis de cinco dias permanece abaixo do limiar definido neste estudo, o balanço hídrico da planta tende a se desequilibrar, comprometendo o desenvolvimento fisiológico, mesmo em solos com boa capacidade de armazenamento. Sob condições de alta demanda evaporativa, o estoque hídrico do solo se torna insuficiente em pouco tempo, iniciando processos de estresse que impactam diretamente a produtividade.

Nesse contexto, o modelo de classificação de estiagem contribui não apenas para identificar situações críticas com base em dados observados, mas também para orientar a irrigação de forma mais eficiente. Quando integrado às previsões meteorológicas de curto prazo, viabiliza uma estratégia de manejo adaptativo, com potencial para mitigar perdas e otimizar o uso dos recursos hídricos, mesmo em condições edáficas favoráveis. Essa abordagem promove maior resiliência produtiva frente à variabilidade climática.

## 5. CONSIDERAÇÕES FINAIS

Este estudo teve como objetivo classificar a ocorrência de estiagens na cultura da soja no município de Dourados/MS, com o intuito de aprimorar o monitoramento climático e oferecer subsídios aos stakeholders do setor agrícola, contribuindo para a tomada de decisão e a redução dos riscos associados à variabilidade pluviométrica. A pergunta de pesquisa que orientou esta investigação é: “Como o desenvolvimento de um modelo de machine learning, baseado em um indicador específico de estiagem para a sojicultura, pode apoiar a classificação, o monitoramento contínuo e a mitigação dos impactos da estiagem na sojicultura em Dourados/MS?” foi respondida de forma positiva. Os resultados confirmam que o método proposto é eficiente para identificar estiagens com base na precipitação acumulada, permitindo um acompanhamento sistemático das condições climáticas e sua relação com a produtividade agrícola.

Uma das principais contribuições da pesquisa foi a criação do conceito de "estiagem da cultura da soja", que combina os critérios de estiagem e veranico e define um limiar crítico de 20,6 mm em janelas móveis de cinco dias. Essa abordagem, inédita na literatura, mostrou-se sensível às exigências fenológicas da cultura e mais precisa para fins de monitoramento climático. A análise da precipitação ao longo do ciclo produtivo revelou sua forte correlação com a produtividade, como demonstrado na comparação entre as safras 2021/2022, marcada por escassez pluviométrica, e 2022/2023, que apresentou melhor desempenho diante de uma distribuição mais regular das chuvas.

A partir desse conceito, foi possível aplicar diferentes algoritmos de *machine learning* para classificar as estiagens e avaliar seu potencial de uso como ferramenta de apoio à decisão. A análise comparativa evidenciou diferenças relevantes entre os três modelos testados, não apenas em termos de desempenho estatístico, mas sobretudo quanto ao critério de segurança do produtor, que prioriza a redução dos falsos negativos — isto é, a minimização do risco de não detectar estiagens reais.

O *Random Forest Classifier* apresentou boa acurácia geral (96,24% no treinamento e 85,79% no teste), com desempenho consistente na classe “com estiagem” (precisão de 0,94 e recall de 0,84). No entanto, suas matrizes de confusão revelaram 490 falsos negativos no teste, número quase três vezes maior do que os falsos positivos (168). Esse padrão de erro indica que o modelo tende a falhar em situações de estiagem real, justamente o cenário mais crítico para a tomada de decisão agrícola. Assim, apesar do bom desempenho numérico, seu alinhamento com a segurança do produtor é limitado.

O *Extra Trees Classifier*, por sua vez, obteve acurácia inferior em comparação ao *Random Forest* (91,84% no treinamento e 83,75% no teste) e apresentou maior desequilíbrio entre classes. Enquanto conseguiu reduzir o número de falsos negativos (112 no teste), aumentou expressivamente os falsos positivos (640). Ou seja, o modelo passou a emitir mais alertas de estiagem sem que esta estivesse presente, o que compromete a precisão da classe “sem estiagem” (0,69) e pode reduzir a confiança do produtor no sistema. Ainda que o excesso de falsos positivos seja menos prejudicial do que a ocorrência de falsos negativos, o modelo mostrou fragilidade em manter o equilíbrio entre segurança e confiabilidade.

O *XGB Classifier* destacou-se com o melhor desempenho global, alcançando 96,90% de acurácia no treinamento e 87,36% no teste, além de métricas equilibradas para ambas as classes. No teste, apresentou 182 falsos negativos e 403 falsos positivos, com recall de 0,87 para a classe “com estiagem” e precisão de 0,94. Esses resultados evidenciam um comportamento mais balanceado entre evitar falhas na detecção de estiagens e limitar a emissão de alertas indevidos, ainda que permaneça o desafio de reduzir os falsos negativos.

Em termos de alinhamento com o critério de segurança, o *XGB Classifier* mostrou-se o modelo mais promissor, por combinar alta performance numérica com maior robustez diante do *trade-off* entre segurança e confiabilidade.

Em síntese, enquanto o *Random Forest* apresenta desempenho elevado, mas erra de forma mais crítica ao não identificar estiagens, e o *Extra Trees* reduz os falsos negativos às custas de excessivos falsos positivos, o *XGB Classifier* se posiciona como a alternativa mais adequada. Sua capacidade de manter alto desempenho global, ao mesmo tempo em que reduz de forma relativa as falhas em estiagens reais, o torna mais alinhado ao objetivo central deste estudo.

A relevância prática do modelo foi evidenciada por uma análise contrafactual, a partir da qual se estimou que, na ausência da estiagem registrada em 2021/2022, a produtividade média poderia ter se aproximado do patamar de 54,43 sc/ha observado em 2022/2023. Considerando a produtividade real de 24,08 sc/ha na safra afetada, a perda potencial chegou a 30,35 sc/ha, representando uma redução de 56% no rendimento. Esse dado reforça o impacto da irregularidade pluviométrica sobre o desempenho da lavoura e a necessidade de ferramentas de apoio à decisão capazes de antecipar situações críticas.

Além da função de monitoramento climático, o modelo também demonstrou aplicabilidade prática no apoio a decisões agronômicas, com destaque para o manejo da irrigação. A simulação econômica da aplicação seletiva demonstrou que a irrigação em parte da área cultivada pode gerar um retorno financeiro altamente vantajoso, com recuperação

parcial da produtividade nas áreas afetadas e lucro líquido significativo em relação ao valor investido. A simulação indicou um retorno sobre o investimento (ROI) estimado em aproximadamente 470%, o que significa que, para cada real investido, haveria um retorno líquido de R\$ 4,70. Esse resultado reforça a relevância do modelo não apenas como ferramenta diagnóstica, mas como suporte técnico para estratégias de mitigação de perdas em contextos de instabilidade climática, mesmo em solos com alta capacidade de retenção de água, como os Latossolos Vermelhos da região.

Diante disso, observa-se que o modelo pode ser incorporado a sistemas de alerta climático, integrando-se a plataformas de monitoramento agrícola já existentes ou a painéis customizados de uso prático. Sua adoção por cooperativas, empresas do setor e instituições públicas se mostra viável, especialmente no contexto de políticas voltadas à adaptação climática no campo. A classificação contínua da condição pluviométrica favorece decisões oportunas, como o replantio, a irrigação e o ajuste do manejo ao longo do ciclo produtivo.

Com base nas classificações geradas, recomenda-se que os produtores utilizem o modelo como ferramenta complementar à gestão da lavoura, adotando ações como escalonamento do plantio, ajuste na adubação ou intervenções emergenciais diante da iminência de déficit hídrico. Para trabalhos futuros, sugere-se a adaptação deste método para outras regiões produtoras e culturas agrícolas dependentes da chuva, com a devida calibração dos limiares conforme os padrões locais. Este avanço permitirá o fortalecimento de estratégias de gestão baseadas em dados, com maior precisão e adaptabilidade aos diferentes contextos produtivos.

Por fim, embora o modelo tenha apresentado bom desempenho com base exclusivamente em variáveis meteorológicas, reconhece-se que a inclusão de fatores agronômicos e ambientais, como tipo de solo, práticas de manejo e características das cultivares, pode ampliar sua capacidade de classificação. Recomenda-se, assim, o desenvolvimento de modelos que integrem múltiplas fontes de dados, favorecendo uma tomada de decisão mais contextualizada, precisa e eficaz.

## REFERÊNCIAS

- ABRAMOVITZ, Moses. Resource and output trends in the United States since 1870. *American Economic Review*, v. 46, n. 2, p. 5-23, 1956.
- AGRITEMPO. Glossário. Sistema de Monitoramento Agrometeorológico. (2025a) Disponível em: <https://www.agritempo.gov.br/br/glossario/>. Acesso em: 23 fev. 2025.
- AGRITEMPO. Sobre nós. Sistema de Monitoramento Agrometeorológico. (2025b) Disponível em: <https://www.agritempo.gov.br/br/sobre-nos/>. Acesso em: 23 fev. 2025.
- AGÊNCIA ESTADO. Estiagem derruba qualidade das lavouras de soja em MS, diz relatório. UOL Notícias, 15 jan. 2025. Disponível em: <https://noticias.uol.com.br/ultimas-noticias/agencia-estado/2025/01/15/estiagem-derruba-qualidade-das-lavouras-de-soja-em-ms-diz-relatorio.htm>. Acesso em: 23 fev. 2025.
- ALMEIDA, H. A. de. Climatologia aplicada à geografia. Campina Grande: EDUEPB, 2016.
- ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P.C.; GONÇALVES, J. L de M.; SPAROVEK, G. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, vol. 22, n. 6, p. 711–728, 2013.
- AMARAL, J. A. M.; MOTCHI, E. P.; OLIVEIRA, H.; CARVALHO FILHO, A.; NAIME, U. J.; SANTOS, R. D. Levantamento semidetalhado dos solos do campo experimental de Dourados, da Embrapa Agropecuária Oeste, município de Dourados, MS. Dourados: Embrapa Agropecuária Oeste; Rio de Janeiro: Embrapa Solos, 2000. 68 p. (Embrapa Agropecuária Oeste. Documentos, 22; Embrapa Solos. Documentos, 15).
- ARIAS, D.; MENDES, P.; ABEL, P. (Coord.). Revisão rápida e integrada da gestão de riscos agropecuários no Brasil. Brasília, DF: Banco Mundial, 2015. 76 p. Disponível em: [https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1044224/1/Revisaorapidaei\\_ntegradadagestao.pdf](https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1044224/1/Revisaorapidaei_ntegradadagestao.pdf). Acesso em: 09 nov. 2020.
- ARROW, K.; BOLIN, B.; COSTANZA, R.; DASGUPTA, P.; FOLKE, C.; HOLLING, C. S.; JANSSON, B.-O.; LEVIN, S.; MÄLER, K.-G.; PERRINGS, C.; PIMENTEL, D. Economic growth, carrying capacity, and the environment. *Science*, v. 268, n. 5210, p. 520– 521, 1995.
- ARSEGO, D. A.; FERRAZ, S. E. T.; STRECK, N. A.; CARDOSO, A. de O.; ZANONI, A. J. Estudos do impacto de diferentes índices associados ao El Niño Oscilação Sul na produtividade de soja no Rio Grande do Sul. *Ciência e Natura*, v. 40, p. 82-87, 2018.
- ASSOCIAÇÃO DOS PRODUTORES DE SOJA DE MATO GROSSO DO SUL. Acompanhamento da Safra - Circular 205/2017: Produtividade Safra Soja 2016/2017. Campo Grande: APROSOJA-MS, 2017.
- ASSOCIAÇÃO DOS PRODUTORES DE SOJA DE MATO GROSSO DO SUL. Acompanhamento da Safra - Circular 253/2018: Produtividade Safra Soja 2017/2018. Campo Grande: APROSOJA-MS, 2018.

ASSOCIAÇÃO DOS PRODUTORES DE SOJA DE MATO GROSSO DO SUL. Acompanhamento da Safra - Circular 302/2019: Produtividade Safra Soja 2018/2019. Campo Grande: APROSOJA-MS, 2019.

ASSOCIAÇÃO DOS PRODUTORES DE SOJA DE MATO GROSSO DO SUL. Errata - Resultados da Safra - Circular 353/2020: Produtividade Safra Soja 2019/2020. Campo Grande: APROSOJA-MS, 2020.

ASSOCIAÇÃO DOS PRODUTORES DE SOJA DE MATO GROSSO DO SUL. Resultados da Safra - Circular 404/2021: Produtividade Safra Soja 2020/2021. Campo Grande: APROSOJA-MS, 2021.

ASSOCIAÇÃO DOS PRODUTORES DE SOJA DE MATO GROSSO DO SUL. Resultados da Safra - Circular 454/2022: Produtividade Safra Soja 2021/2022. Campo Grande: APROSOJA-MS, 2022.

ASSOCIAÇÃO DOS PRODUTORES DE SOJA DE MATO GROSSO DO SUL. Resultados da Safra - Circular 509/2023: Produtividade Safra Soja 2022/2023. Campo Grande: APROSOJA-MS, 2023.

ASSOCIAÇÃO DOS PRODUTORES DE SOJA DE MATO GROSSO DO SUL. Resultados da Safra - Circular 560/2024: Produtividade Safra Soja 2023/2024. Campo Grande: APROSOJA-MS, 2024.

BAIOCO, R.; ORBEN, J.; PEREIRA, J.; ESSER, J.; FERREIRA SANTOS, R. Impacto da estiagem sobre a safra de soja 2021/22 no Núcleo Regional de Cascavel. *International Journal of Environmental Resilience Research and Science*, v. 5, p. 1-11, 2023. DOI: 10.48075/ijerrs.v5i1.30572.

BANDEIRA, P. M. da C. Estimativa de produtividade da cultura da soja utilizando técnicas de inteligência artificial. 2022. 51p. Dissertação (Mestrado em Engenharia Agrícola). Universidade Federal de Viçosa, Viçosa - MG, 2022.

BLAIN, G. C.; BRUNINI, O. Análise da escala temporal de monitoramento das secas agrícolas e meteorológicas no estado de São Paulo. *Revista Brasileira de Meteorologia*, v. 22, n. 2, p. 255–261, 2007.

BOUAZIZ, M.; MEDHIOUB, E.; CSAPLOVISC, E. A machine learning model for drought tracking and forecasting using remote precipitation data and a standardized precipitation index from arid regions. *Journal of Arid Environments*, Academic Press, 2021.

BRASIL. Modelagem Climática e Vulnerabilidades Setoriais à Mudança do Clima no Brasil. Ministério do Meio Ambiente, Brasília, 2016. Disponível em: <https://repositorio.enap.gov.br/bitstream/1/8030/7/ModelagemClimticaeVulnerabilidadeSetoriaisMudanadoClimanoBrasilcompressed.pdf>. Acesso em: 22 fev. 2025.

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

CAI, W.; SANTOSO, A.; WANG, G.; WU, L.; COLLINS, M.; LENGAINNE, M.; POWER, S.; TIMMERMANN, A. ENSO response to greenhouse forcing. In: McPHADEN, Michael J.;

SANTOSO, Agus; CAI, Wenju (org.). *El Niño Southern Oscillation in a Changing Climate*. Hoboken: John Wiley & Sons, 2020. p. 289–307. DOI: <https://doi.org/10.1002/9781119548164.ch13>

CAMPO GRANDE NEWS. Seca causa prejuízo de quase R\$ 3 bilhões na safra de soja em MS. *Campo Grande News*, 7 fev. 2022. Disponível em: <https://www.campograndenews.com.br/lado-rural/seca-causa-prejuizo-de-quase-r-3-bilhoes-na-safra-de-soja-em-ms>. Acesso em: 21 jun. 2025.

CANAL RURAL. Impactos do clima e da política externa no mercado da soja. Canal Rural, 2024. Disponível em: <https://www.canalrural.com.br/agricultura/projeto-soja-brasil/impactos-do-clima-e-da-politica-externa-no-mercado-da-soja/>. Acesso em: 21 fev. 2025.

CARVALHO, P. G. M. de. As vertentes teóricas da produtividade. *Revista de Economia Contemporânea*, Rio de Janeiro, 5(2): 67-92, 2001.

CARVALHO, M. A. V. Variabilidade da largura e intensidade da Zona de Convergência Intertropical Atlântica: aspectos observacionais e de modelagem. 2011. 137f. Dissertação (Mestrado em Meteorologia) – Instituto Nacional de Pesquisas Físicas, São José dos Campos, 2011.

CASONATO, L. O papel do agronegócio no crescimento econômico sul-mato-grossense à luz do modelo de Solow. *Revista de Economia Agrícola*, São Paulo, v. 60, n. 1, p. 31-39, 2013.

CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, v. 16, p. 321-357, 2002. Disponível em: <https://doi.org/10.48550/arXiv.1106.1813>. Acesso em: 25 fev. 2025.

CHEN, T.; GUESTRIN, C. *XGBoost: A Scalable Tree Boosting System*. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016. Disponível em: <https://xgboost.readthedocs.io/en/stable/>. Acesso em: 25 fev. 2025.

CRANE-DROESCH, A. *Machine learning methods for crop yield prediction and climate change impact assessment in agriculture*. *Environmental Research Letters*, v. 13, n. 11, p. 114003, 2018.

COMPANHIA NACIONAL DE ABASTECIMENTO (CONAB). Safra - Série Histórica de Grãos. Brasília, 2024. Disponível em: <https://portaldeinformacoes.conab.gov.br/safra-serie-historica-graos.html>. Acesso em: 21 fev. 2025.

DASGUPTA, P. *The Economics of Biodiversity: The Dasgupta Review*. Londres: HM Treasury, 2021.

DIKSHIT, A.; PRADHAN, B.; HUETE, A. An improved SPEI drought forecasting approach using the long short-term memory neural network. *Journal of Environmental Management*, v. 283, 2021.

DOMAR, E. D. Capital Expansion, Rate of Growth and Employment. *Econometrics*, n°. 14, 1946.

DU, Zhushan; FENG, Hongli; ARBUCKLE, J. Exploring the complementarity between traditional econometric methods and machine learning – an application to adoption and disadoption of conservation practices. *Applied Economics*, [S.l.], publicado em 17 fev. 2025. DOI: <https://doi.org/10.1080/00036846.2025.2462792>. Acesso em: 21 jun. 2025.

ELBELTAGI, A.; PANDE, C. B.; KUMAR, M.; TOLCHE, A. D.; SINGH, S. K.; KUMAR, A.; VISHWAKARMA, D. K. Prediction of meteorological drought and standardized precipitation index based on the random forest (RF), random tree (RT), and Gaussian process regression (GPR) models. *Environmental Science and Pollution Research*, v. 30, n. 15, p. 43183–43202, 2023.

EMBRAPA. Guia Clima – Embrapa Agropecuária Oeste. Disponível em: <https://clima.cpao.embrapa.br/>. Acesso em: 25 fev. 2025.

EMBRAPA. Importância socioeconômica da soja. Agência de Informação Embrapa, 2021. Disponível em: <https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/soja/pre-producao/socioeconomia/importancia-socioeconomica-da-soja>. Acesso em: 9 set. 2025.

EMBRAPA. Evolução e perspectivas de desempenho econômico do complexo soja. Londrina: Embrapa Soja, 2010. (Documentos, 319). Disponível em: <https://www.infoteca.cnptia.embrapa.br/bitstream/doc/887037/1/Doc3192ED1.pdf>. Acesso em: 9 set. 2025.

EMBRAPA. Evolução da cultura da soja: Censos Agropecuários 2006 e 2017. In: CONGRESSO DA SOBER, 61., 2023. Anais [...]. Brasília: Embrapa, 2023. Disponível em: <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1156293/1/Evolucao-Congresso-SOBER-2023.pdf>. Acesso em: 9 set. 2025.

EMBRAPA. Mercado da soja e derivados. Agência de Informação Embrapa, 2021b. Disponível em: <https://www.embrapa.br/agencia-de-informacao-tecnologica/cultivos/soja/pre-producao/socioeconomia/mercado>. Acesso em: 9 set. 2025.

FARIAS, J. R. B.; NEPOMUCENO, A. L.; NEUMAIER, N. Ecofisiologia da soja. Londrina: Embrapa Soja, 2007. 8 p. (Embrapa Soja. Circular técnica, 48). Disponível em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/470308/ecofisiologia-da-soja>. Acesso em: 23 fev. 2025.

FARIAS, J.R.B.; NEUMAIER, N.; NEPOMUCENO, A.L. Soja. In: MONTEIRO, J.E.B.A. Agrometeorologia dos cultivos: o fator meteorológico na produção agrícola. Brasília, DF: INMET, 2009. cap.15, p.261-277.

FEHR, W. R.; CAVINESS, C. E. Stages of soybean development. Ames: Iowa State University. Depto. of Science and Technology, 1977. 11 p.

FENG, P.; WANG, B.; LUO, J.-J.; LIU, D. L.; WATERS, C.; JI, F.; RUAN, H.; XIAO, D.; SHI, L.; YU, Q. Using large-scale climate drivers to forecast meteorological drought

condition in growing season across the Australian wheatbelt. *Science of the Total Environment*, v. 724, 2020.

FERREIRA, M. de O.; RAMOS, L. M.; ROSA, A. L. T. Crescimento da agropecuária cearense: comparação entre as produtividades parciais e total. *Revista de Economia e Sociologia Rural*, vol. 44, n° 03, p. 503-524, 2006.

FIETZ, C. R., COMUNELLO, É., FLUMIGNAN, D. L. (2022). Em dezembro de 2021 as temperaturas foram elevadas e as chuvas escassas na região da Grande Dourados. Embrapa Agropecuária Oeste, *Boletim Agropecuário* 12. Disponível em: <https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1138828/1/BA-12-2021.pdf>

FIETZ, C. R.; FISCH, G.; COMUNELLO, E.; FLUMIGNAN, D. L. O clima de Dourados e Rio Brillhante, Mato Grosso do Sul. Dourados, MS: Embrapa Agropecuária Oeste, 2024. 49 p. (Documentos / Embrapa Agropecuária Oeste, n. 150).

FORDOŃSKI, G.; OKORSKI, A.; OLSZEWSKI, J.; DAŃBROWSKA, J.; PSZCZÓŁKOWSKA, A. The Effect of Sowing Date on the Growth and Yield of Soybeans Cultivated in North-Eastern Poland. *Agriculture*, v. 13, n. 12, 2199, 2023. DOI: 10.3390/agriculture13122199

FREITAS, M. J. C. C.; OLIVEIRA, F. H. de. Estiagem no Oeste Catarinense: diagnóstico e resiliência (relatório técnico-científico). Florianópolis: Editora Udesc, 2018.

GANGULI, P.; JANGA REDDY, M. Ensemble prediction of regional droughts using climate inputs and the SVM-copula approach. *Hydrological Processes*, v. 28, n. 19, p. 4989–5009, 2014.

GERON, A. *Mãos à obra: aprendizado de máquina com Scikit-Learn, Keras e TensorFlow*. 2. ed. Rio de Janeiro: Alta Books, 2019.

GEURTS, P.; ERNST, D.; WEHENKEL, L. Extremely randomized trees. *Machine Learning*, v. 63, p. 3–42, 2006.

GIL, A. C. *Métodos e técnicas de pesquisa social*. São Paulo: Atlas, 1999.

GONÇALVES, S. L.; FARIAS, J. R. B de.; SIBALDELLI, R. N. R. Eventos climáticos adversos e seus impactos para as culturas de soja, milho e trigo no Brasil. Londrina, PR: Embrapa Soja. n. 420, p. 48, 2019.

GONÇALVES, S. T. N.; VASCONCELOS JUNIOR, F. C.; SAKAMOTO, M. S.; SILVEIRA, C. S.; MARTINS, E. S. P. R. Análise dos índices e metodologias para o monitoramento de secas: uma revisão. *Revista Brasileira de Meteorologia*, v. 36, n. 3 (Suplemento), p. 495–511, 2021.

GOLOMBIESKI, E. E.; VAZ, M. S. M. G. Modelo de integração de dados climáticos e imagens no monitoramento agrícola. In: Congresso Internacional de Agricultura de Precisão. Anais [...], 2024. Disponível em: [https://www.ciaca-conf.org/wp-content/uploads/2024/12/CIACA2024\\_PT\\_F\\_034\\_Everz-Golombieski.pdf](https://www.ciaca-conf.org/wp-content/uploads/2024/12/CIACA2024_PT_F_034_Everz-Golombieski.pdf). Acesso em: 22 fev. 2025.

GUIA clima. Dourados: Embrapa Agropecuária Oeste, 2025. Disponível em: [www.cpaio.embrapa.br/clima/](http://www.cpaio.embrapa.br/clima/). Acesso em 26 fev. 2025.

HAMEED, M. M.; RAZALI, S. F. M.; MOHTAR, W. H. M. W.; RAHMAN, N. A.; YASEEN, Z. M. Machine learning models development for accurate multi-months ahead drought forecasting: Case study of the Great Lakes, North America. PLoS ONE, v. 18, n. 10, 2023.

HAMEED, M. M.; RAZALI, S. F. M.; MOHTAR, W. H. M. W.; ALSAYDALANI, M. O. A.; YASEEN, Z. M. Deep learning versus hybrid regularized extreme learning machine for multi-month drought forecasting: A comparative study and trend analysis in tropical region. Heliyon, v. 10, n. 1, 2024.

HARROD, R. An essay in dynamic theory. Economic Journal, v.49, n.193, p.14-33, 1939.

HU, J.; LIU, X.; YANG, M.; SONG, J. A Comparison and Strategy of Semantic Segmentation on Remote Sensing Images. *arXiv preprint* arXiv:1905.10231, 2019. Disponível em: <https://arxiv.org/abs/1905.10231>. Acesso em: 22 fev. 2025.

IBGE. Dourados (MS) – Panorama. Cidades. Disponível em: <https://cidades.ibge.gov.br/brasil/ms/dourados/panorama>. Acesso em: 12 set. 2025.

INSTITUTO NACIONAL DE METEOROLOGIA (INMET). Glossário. Disponível em: <https://portal.inmet.gov.br/glossario/glossario#S>. Acesso em: 25 fev. 2025.

IPCC. Climate Change 2023: Synthesis Report. Summary for Policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva: IPCC, 2023. Disponível em: [https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC\\_AR6\\_SYR\\_SPM.pdf](https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_SPM.pdf). Acesso em: 23 fev. 2025.

JEONG, J. H.; RESOP, J. P.; MUELLER, N. D.; FLEISHER, D. H.; YUN, K.; BUTLER, E. E.; TIMLIN, D. J.; SHIM, K. M.; GERBER, J. S.; REDDY, V. R.; KIM, S. H. Random forests for global and regional crop yield predictions. PLOS ONE, v. 11, n. 6, p. e0156571, 2016.

JONES, C. I.; VOLLRATH, D. Introdução à Teoria do Crescimento Econômico. 3º ed. Rio de Janeiro, RJ: Elsevier, 2014.

LAKATOS, E. M.; MARCONI, M. de A. Fundamentos de metodologia científica. 5. ed. São Paulo: Atlas, 2003.

LAZY PREDICT. LazyPredict Documentation. (2022) Disponível em: <https://lazypredict.readthedocs.io/en/latest/>. Acesso em: 25 fev. 2025.

LENNON, S. Estiagem compromete floração e plantio da soja. Agrolink, 24 jan. 2025. Disponível em: [https://www.agrolink.com.br/noticias/estiagem-compromete-floracao-e-plantio-da-soja\\_498775.html](https://www.agrolink.com.br/noticias/estiagem-compromete-floracao-e-plantio-da-soja_498775.html). Acesso em: 23 fev. 2025.

LUCAS JR., R. E. On the mechanics of economic development. *Journal of Monetary Economics*, v. 22, n. 1, p. 3-42, 1988.

LUCAS JR., R. E. Making a miracle. *Econometrica*, v. 61, n. 2, p. 251-272, 1993.

MAHMOOD, A.; RIZVI, S. A. R.; RIZVI, S. M. H. Apache Spark Accelerated Deep Learning Inference for Large Scale Satellite Image Analytics. arXiv preprint arXiv:1908.04383, 2019. Disponível em: <https://arxiv.org/abs/1908.04383>. Acesso em: 22 fev. 2025.

MALTHUS, T. An essay on the principle of population. London: J. Johnson, 1798.

MARX, K. O Capital: crítica da economia política. Livro I. Hamburgo: Otto Meissner Verlag, 1867.

MATTAR, F. N. Pesquisa de marketing: metodologia, planejamento. 3. ed. São Paulo: Atlas, 1996.

MAVI, H. S.; TUPPER, G. J. *Agrometeorology: Principles and Application of Climate Studies in Agriculture*. New York: Food Products Press, 2004. 364 p

MAZZON, J. A. Using the methodological association matrix in marketing studies [Uso da matriz de amarração metodológica em marketing]. *Revista Brasileira de Marketing – REMark*, v. 17, n. 5, p. 747-770, 2018.

MICHELS, I.; SILVESTRINI, R.; SARDINHA, V.; SPROESSER, R. Sojicultura. Campo Grande, MS: UFMS, 2004.

MINISTÉRIO DA AGRICULTURA, PECUÁRIA E ABASTECIMENTO (MAPA). MAPA fortalece exportações agrícolas com recordes históricos e diversificação de mercados neste ano. Brasília, 2024. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/noticias/mapa-fortalece-exportacoes-agricolas-com-recordes-historicos-e-diversificacao-de-mercados-neste-ano>. Acesso em: 21 fev. 2025.

NORDHAUS, W. D. *Managing the Global Commons: The Economics of Climate Change*. Cambridge, MA: MIT Press, 1994.

NORDHAUS, W. D.; BOYER, J. *Warming the World: Economic Models of Global Warming*. Cambridge, MA: MIT Press, 2000.

OLIVEIRA, G. G. de; RUIZ, L. F. C.; GUASSELLI, L. A.; TEIXEIRA, B.; KLABIN, G. Modelo baseado em redes neurais artificiais e dados de chuva por satélites para alerta de estiagens e secas. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 20., 2023, Florianópolis. Anais [...]. São José dos Campos: INPE, 2023. Disponível em: <https://proceedings.science/sbsr-2023/trabalhos/modelo-baseado-em-redes-neurais-artificiais-e-dados-de-chuva-por-satelites-para?lang=pt-br>. Acesso em: 7 set. 2025.

OLIVEIRA JUNIOR, A. de; CASTRO, C. de; PEREIRA, L. R.; DOMINGOS, C. da S. *Estádios fenológicos e marcha de absorção de nutrientes da soja*. Paiçandu: Fortgreen; Londrina: Embrapa Soja, 2016.

OLIVEIRA, L. F.; COSTA, M. H. Mapa de localização da cidade de Dourados, Mato Grosso do Sul. In: *Análise da Dinâmica Urbana e Uso do Solo em Dourados/MS*. 2018. Disponível em: [https://www.researchgate.net/figure/Figura-1-Mapa-de-localizacao-da-cidade-de-Dourados-Mato-Grosso-do-Sul\\_fig1\\_329926693](https://www.researchgate.net/figure/Figura-1-Mapa-de-localizacao-da-cidade-de-Dourados-Mato-Grosso-do-Sul_fig1_329926693). Acesso em: 26 fev. 2025.

RAVI, R.; BARANIDHARAN, B. Crop yield prediction using XGBoost algorithm. *International Journal of Recent Technology and Engineering (IJRTE)*, [S.l.], v. 8, n. 5, p. 3516–3520, jan. 2020. DOI: <https://doi.org/10.35940/ijrte.D9547.018520>. Acesso em: 21 jun. 2025.

REBOITA, M. S.; GAN, M. A.; DA ROCHA, R. P.; AMBRIZZI, T. Regime de precipitação na América do Sul: uma revisão bibliográfica. *Revista Brasileira de Meteorologia*, v. 25, n.2, p. 185–204, 2010.

RICARDO, D. *On the principles of political economy and taxation*. London: John Murray, 1817.

RICHETTI, A.; FLUMIGNAN, D. L.; ALMEIDA, A. C. S. Viabilidade econômica da soja irrigada na safra 2015/2016, na região sul de Mato Grosso do Sul. Dourados, MS: Embrapa Agropecuária Oeste, 2015. 8 p. (Embrapa Agropecuária Oeste. Comunicado Técnico, 224). ISSN 1679-0472.

RICHETTI, J. Modelos agrometeorológicos, espectrais e de inteligência artificial para estimação de produtividade de soja. 2018. 83p. Tese (Doutorado em Engenharia Agrícola). Universidade Estadual do Oeste do Paraná, Cascavel – PR, 2018.

RODRIGUES, A. B. Estiagem prejudica desenvolvimento das lavouras de verão no RS. Governo do Estado do Rio Grande do Sul, 24 jan. 2025. Disponível em: <https://www.agricultura.rs.gov.br/estiagem-prejudica-desenvolvimento-das-lavouras-de-verao-no-rs>. Acesso em: 23 fev. 2025.

ROMER, P. M. Increasing returns and long-run growth. *Journal of Political Economy*, v. 94, n. 5, p. 1002–1037, 1986.

ROMER, P. M. Endogenous technological change. *Journal of Political Economy*, v. 98, n. 5, p. S71-S102, 1990.

SANTOS, A. C.; OLIVEIRA, B. A. de.; GOMES, I.F.; GROFF, A. M. Fatores e técnicas de produção e sua influência na produtividade e qualidade da soja. In: *Encontro de Engenharia de Produção Agroindustrial*, 11., 2017, Campo Mourão. Anais [...]. Campo Mourão: EEPA, 2017.

SCHUMPETER, J. A. *The Theory of Economic Development*. Cambridge, MA: Harvard University Press, 1911.

SCHUMPETER, J. A. *Capitalism, Socialism and Democracy*. New York: Harper & Brothers, 1942.

SILVA, G. L. S. P.; VICENTE, J. R.; CASER, D. V. Variações do tempo e produtividade agrícola: um subsídio à previsão de safras no estado de São Paulo. Campinas: Fundação Cargill, 1986.

SMITH, A. An Inquiry into the Nature and Causes of the Wealth of Nations. London: W. Strahan and T. Cadell, 1776.

SMULDERS, S. Entropy, environment, and endogenous economic growth. *International Tax and Public Finance*, v. 2, n. 2, p. 319–340, 1995.

SOLOW, R. M. A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, v. 70, n. 1, p. 65–94, 1956.

SOLOW, R. M. Technical change and the aggregate production function. *The review of Economics and Statistics*, v. 39, n. 3, p. 312–320, 1957.

STERN, N. The Economics of Climate Change: The Stern Review. Londres: HM Treasury, 1997.

VICENTE, J. R.; MARTINS, R. Condições do tempo e produtividade total dos fatores na agricultura paulista, 1962-2002. *Informações Econômicas*, São Paulo, v.36, n.6, 2006.

WANG, G. C.; ZHANG, Q.; BAND, S. S.; DEGHANI, M.; CHAU, K. W.; THO, Q. T.; ZHU, S.; SAMADIANFARD, S.; MOSAVI, A. Monthly and seasonal hydrological drought forecasting using multiple extreme learning machine models. *Engineering Applications of Computational Fluid Mechanics*, v. 16, n. 1, p. 1364–1381, 2022.

WMO (World Meteorological Organization). What is climate? 2020. Disponível em: [https://www.wmo.int/pages/prog/wcp/ccl/faq/faq\\_doc\\_en.html](https://www.wmo.int/pages/prog/wcp/ccl/faq/faq_doc_en.html). Acesso em: 25 jun. 2020.

ZAVATTINI, J. A. As chuvas e as massas de ar no estado de Mato Grosso do Sul: estudos geográficos com vista à regionalização climática. 1. ed. São Paulo: Editora Cultura Acadêmica, 2009. 214p.