

ARTIGO DE PESQUISA/RESEARCH PAPER

Análise Comparativa de Modelos de Aprendizado de Máquina na Predição de Variáveis Climáticas para Aplicações Agrícolas

Comparative Analysis of Machine Learning Models for Climate Variable Prediction in Agricultural Applications

Francisco Ribeiro ✉ [Instituto Federal do Ceará (IFCE) – Campus Tianguá | francisco.ribeiro.silva07@aluno.ifce.edu.br]

Raquel Silveira 📧 [Instituto Federal do Ceará (IFCE) – Campus Tianguá | raquel_silveira@ifce.edu.br]

Clemilton Ferreira [Instituto Federal do Ceará (IFCE) – Campus Tianguá | clemil@ifce.edu.br]

✉ Instituto Federal do Ceará (IFCE), Campus Tianguá, CE-187, s/n, Tianguá, CE, 62320-000, Brasil.

Resumo. A previsão de variáveis climáticas é essencial para o planejamento agrícola, especialmente em regiões com grande variação na precipitação. Este trabalho avalia o desempenho de algoritmos de aprendizado de máquina na predição de temperatura, precipitação e umidade relativa do ar, utilizando dados climáticos mensais de Sobral-CE no período de 1973 a 2019. Foram analisados modelos como *Decision Tree*, *Random Forest*, *XGBoost*, KNN e MLP em cenários sem e com memória temporal. Os resultados indicam que a inclusão de memória temporal melhora o desempenho dos modelos, com destaque para os métodos de *ensemble*. A temperatura apresentou maior previsibilidade, enquanto a precipitação se mostrou mais desafiadora, devido a sua alta variabilidade. Como contribuição, o estudo apresenta uma análise comparativa integrada em uma série histórica extensa, evidenciando o potencial do aprendizado de máquina na previsão de variáveis climáticas, contribuindo para o planejamento agrícola regional.

Abstract. Climate variable forecasting is essential for agricultural planning, especially in regions with high rainfall variability. This study evaluates the performance of machine learning algorithms in predicting temperature, precipitation, and relative humidity using monthly data from Sobral-CE, Brazil, from 1973 to 2019. Models such as Decision Tree, Random Forest, and XGBoost were analyzed under two scenarios: with and without temporal lag. The results indicate that incorporating temporal memory improves model performance, with ensemble methods showing the best results. Temperature showed higher predictability, while precipitation proved to be more challenging. As a contribution, this study presents an integrated comparative analysis using a long-term historical series, highlighting the potential of machine learning to support regional agricultural planning.

Palavras-chave: Aprendizado de máquina, Previsão climática, Séries temporais, Agricultura

Keywords: Machine learning, Climate forecasting, Time series, Agriculture

Recebido/Received: 17 April 2026 • **Aceito/Accepted:** 07 July 2026 • **Publicado/Published:** 10 July 2026

1 Introdução

A variabilidade climática é um dos principais obstáculos para a sustentabilidade da produção agrícola, afetando especialmente as regiões semiáridas. Nessas regiões, a produtividade das culturas está profundamente ligada à distribuição irregular de fatores como a precipitação e a umidade relativa do ar, que determinam o êxito ou o fracasso das colheitas [Oliveira *et al.*, 2022; Lima *et al.*, 2025]. Nesse contexto, a previsão de variáveis climáticas torna-se uma ferramenta essencial para o planejamento agrícola, contribuindo para a tomada de decisão por parte de produtores e gestores [Born *et al.*, 2021; Garcia *et al.*, 2025].

Com o avanço das técnicas de aprendizado de máquina, diferentes abordagens têm sido propostas para modelagem e previsão de variáveis climáticas. Esses métodos apresentam vantagens em relação a técnicas estatísticas tradicionais, como os modelos autorregressivos de médias móveis (ARIMA) e a regressão linear múltipla, especialmente na capacidade de capturar relações não lineares e padrões complexos nos dados [Zhang *et al.*, 2025]. No entanto, a aplicação dessas técnicas ainda enfrenta desafios, como a alta variabilidade de certas variáveis, a presença de dados ausentes e a dependência

temporal inerente às variáveis climáticas [Chen *et al.*, 2023].

Embora modelos de aprendizado de máquina como o *Random Forest* e *XGBoost* tenham demonstrado eficácia em diversas aplicações de larga escala, a aplicação prática em séries temporais exige uma compreensão profunda de como esses algoritmos lidam com a complexidade dos dados [Sarker, 2021; Salman *et al.*, 2024].

Diante disso, este trabalho tem como objetivo analisar o desempenho de modelos de aprendizado de máquina na predição das variáveis climáticas temperatura, precipitação e umidade relativa do ar, utilizando dados históricos do município de Sobral-CE. Este estudo realiza uma análise comparativa de múltiplos algoritmos de aprendizagem de máquina, tais como *Decision Tree*, *Random Forest*, *eXtreme Gradient Boosting (XGBoost)*, *K-Nearest Neighbors (KNN)* e *Multilayer Perceptron (MLP)*, em duas configurações experimentais distintas (sem e com memória temporal), a partir de uma série histórica de longo prazo (46 anos) contribuindo para a compreensão do impacto da memória temporal na modelagem de variáveis climáticas em nível local.

2 Fundamentação Teórica

Para compreender a modelagem e a avaliação dos algoritmos propostos para o cenário local, faz-se necessário estabelecer os conceitos teóricos que sustentam esta pesquisa. Inicialmente, discute-se o papel das tecnologias digitais e dos sistemas preditivos aplicados à agricultura, destacando a importância da análise de dados climáticos para o planejamento agrícola. Em seguida, são apresentados conceitos relacionados à Inteligência Artificial e à Aprendizagem de Máquina, com ênfase em algoritmos de aprendizado supervisionado utilizados em tarefas de regressão. Por fim, são descritas as principais métricas de avaliação empregadas para analisar o desempenho dos modelos preditivos utilizados neste trabalho.

2.1 Tecnologias digitais e predição na agricultura

As tecnologias de predição aplicadas ao planejamento agrícola têm se tornado elementos centrais da chamada agricultura de precisão, permitindo diferentes formas de apoio à tomada de decisão no campo. Essas tecnologias utilizam sistemas computacionais, sensores, dados meteorológicos e algoritmos de inteligência artificial para prever o desempenho das culturas e indicar o melhor momento e local para o plantio. No contexto da agricultura moderna, essas ferramentas são empregadas como forma de aumentar a produtividade e reduzir riscos associados às variações climáticas e às condições do solo [Filippi et al., 2025].

Com o avanço da agricultura digital, diferentes tecnologias passaram a ser integradas com o objetivo de aprimorar a capacidade de previsão no planejamento agrícola. Entre essas tecnologias destacam-se as estações meteorológicas, que possibilitam a coleta contínua de dados ambientais. Essas informações podem ser processadas por modelos computacionais e combinadas com outras variáveis relevantes, como umidade do solo, fertilidade, temperatura e histórico climático, permitindo gerar recomendações mais precisas para o manejo agrícola. De acordo com Padhiary et al. [2025], a integração de inteligência artificial, Internet das Coisas (IoT) e computação em nuvem tem possibilitado o desenvolvimento de sistemas inteligentes capazes de otimizar o planejamento agrícola e melhorar a eficiência produtiva.

Além disso, o uso de técnicas de aprendizado de máquina tem ampliado significativamente a capacidade de previsão em sistemas agrícolas. Esses modelos preditivos podem auxiliar os produtores na escolha da cultura mais adequada, na definição do momento ideal de plantio e na adoção de práticas de manejo mais eficientes. Pesquisas recentes demonstram que algoritmos de aprendizado de máquina são capazes de prever o crescimento e o comportamento das culturas com elevada precisão, contribuindo para decisões agrícolas mais eficientes e sustentáveis [Mishra et al., 2024].

Dessa forma, o uso de tecnologias digitais e modelos preditivos representa um avanço significativo para a sustentabilidade e eficiência da agricultura moderna, contribuindo para minimizar riscos associados às mudanças climáticas e à variabilidade ambiental. Nesse contexto, a aplicação de técnicas de inteligência artificial e aprendizado de máquina torna-se uma estratégia promissora para a análise e previsão de variáveis climáticas, tema que é explorado nas seções

seguintes deste trabalho [Saha et al., 2025].

2.2 Inteligência Artificial

A Inteligência Artificial é definida como o ramo da ciência e da engenharia dedicado ao desenvolvimento de sistemas capazes de perceber seu ambiente, processar informações e tomar ações que maximizem suas chances de alcançar objetivos. Esses sistemas empregam métodos computacionais que permitem aprendizado, adaptação e tomada de decisão autônoma, reproduzindo aspectos do raciocínio humano [Sheikh et al., 2023].

2.2.1 Aprendizagem de Máquina

A Aprendizagem de Máquina (do inglês *Machine Learning*, ML) é um campo da inteligência artificial que estuda algoritmos e modelos capazes de aprender padrões e tomar decisões a partir de dados, sem depender de regras explicitamente programadas. Em vez de seguir regras estabelecidas por humanos, os modelos de aprendizado de máquina analisam os dados para encontrar padrões estatísticos e executar tarefas, como geração de resultados preditivos ou descritivos [Sarker, 2021].

Os métodos de ML são divididos em duas categorias principais: (i) aprendizado supervisionado e (ii) aprendizado não supervisionado. Cada categoria se difere conforme o objetivo do modelo [Pugliese et al., 2021].

No aprendizado supervisionado, o modelo é treinado usando pares de entrada e saída (previamente rotulada). Seu objetivo é aprender uma função que generalize para novos dados. Nesse contexto, duas tarefas são especialmente comuns: classificação e regressão. A classificação envolve prever categorias discretas — por exemplo, identificar se um e-mail é “spam” ou “não spam”, ou determinar a classe de um documento. Já a regressão, lida com a previsão de valores contínuos, como estimar o preço de um imóvel ou a probabilidade numérica de um evento ocorrer. Algoritmos como regressão linear, árvores de decisão, *Support Vector Machines* (SVM), *K-Nearest Neighbors* (KNN) e redes neurais são amplamente utilizados nessas tarefas [Sarker, 2021].

Enquanto que no aprendizado não supervisionado, os dados de entrada não possuem rótulos. A finalidade é identificar estrutura ou padrões inerentes, como agrupamentos (*clustering*), redução de dimensionalidade e detecção de anomalias. Algoritmos como *K-means*, *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN), *Principal Component Analysis* (PCA) e *Gaussian Mixture Model* (GMM) são exemplos de métodos representativos nesse cenário. Esse tipo de aprendizado é útil quando a rotulagem é cara ou impraticável [Woodman and Mangoni, 2023].

Este estudo enquadra-se no paradigma de aprendizado supervisionado, uma vez que os dados utilizados possuem rótulos associados. Como esses rótulos correspondem a valores contínuos, a tarefa é caracterizada como um problema de regressão. Dessa forma, as seções seguintes concentram-se nos fundamentos do aprendizado supervisionado aplicado à regressão, apresentando alguns dos principais algoritmos adequados a essa tarefa, bem como as métricas utilizadas para avaliar o desempenho dos modelos.

2.2.2 Algoritmos de Aprendizagem Supervisionada

Diante da caracterização do problema como uma tarefa de regressão no paradigma supervisionado, selecionou-se um conjunto de algoritmos amplamente consolidados na literatura, cujas características estruturais são descritas a seguir:

- **Decision Tree (DT):** Algoritmo de aprendizado supervisionado que organiza decisões em uma estrutura hierárquica de nós, segmentando os dados de acordo com os atributos que de forma mais eficiente separam as classes. Cada divisão visa diminuir a impureza do conjunto, gerando ramos que levam às folhas que representam as previsões. É um modelo não paramétrico, frequentemente empregado em tarefas de classificação e regressão [Silva and Silva Neto, 2023].
- **Random Forest (RF):** Algoritmo de aprendizado de máquina supervisionado que combina múltiplas árvores de decisão para melhorar a precisão e reduzir o sobreajuste. Conforme Salman *et al.* [2024], cada árvore é construída a partir de uma amostra aleatória dos dados e considera subconjuntos aleatórios de variáveis em cada divisão, permitindo que o modelo capture padrões complexos e identifique as variáveis mais relevantes.
- **XGBoost:** Algoritmo de aprendizado de máquina supervisionado baseado em *boosting* com gradiente extremo, que combina múltiplos modelos fracos em uma sequência para melhorar a capacidade preditiva geral. Essa estratégia de *boosting* possibilita que o algoritmo elabore árvores de decisão de maneira sequencial, em que cada árvore corrige os erros das anteriores, resultando em um modelo final mais robusto e preciso [Moore and Bell, 2022].
- **K-Nearest Neighbors (KNN):** Método de aprendizado supervisionado, não paramétrico e baseado em instâncias, comumente utilizado em tarefas de classificação e regressão. Ele é considerado um algoritmo de aprendizado preguiçoso, pois não exige uma fase de treinamento explícita. Em vez disso, realiza previsões analisando os dados de treinamento no momento da consulta, estimando a classe ou valor de uma nova instância com base nas instâncias vizinhas já conhecidas [Halder *et al.*, 2024].
- **Multilayer Perceptron (MLP):** Rede Neural Artificial (RNA) do tipo *feedforward*, composta por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. O MLP utiliza funções de ativação não lineares e o algoritmo de retropropagação do erro para ajustar seus pesos sinápticos, possibilitando a modelagem de relações complexas entre variáveis. Essa estrutura o torna adequado para tarefas de classificação e regressão [Bikku, 2020].

2.2.3 Métricas de Avaliação em Tarefas de Regressão

Para avaliar o desempenho dos modelos de aprendizagem de máquina, são empregadas métricas específicas, conforme a tarefa. Considerando que este estudo se concentra em uma tarefa de regressão, utilizam-se indicadores apropriados para esse tipo de problema, os quais quantificam o erro entre os valores previstos e os observados. Essas métricas permitem

avaliar tanto a precisão quanto a qualidade global das previsões. A seguir, são apresentadas as principais métricas adotadas nesta análise.

- **Coefficiente de determinação (R^2):** Mede a proporção da variabilidade da variável dependente que é explicada pelo modelo de regressão. Essa métrica indica o quanto o modelo consegue representar a variação observada nos dados, variando entre 0 e 1. Assim, valores mais altos de R^2 refletem melhor capacidade explicativa do modelo [Ferreira and Batista, 2021].
- **Erro quadrático médio (MSE):** Consiste na média do erro das previsões ao quadrado. Trata-se de uma avaliação que representa a diferença entre o valor estimado pelo modelo e o valor observado, cujo resultado é elevado ao quadrado. Este procedimento é repetido para todas as observações, somando-se os resultados, e, por fim, divide-se pela quantidade de observações previstas. Quanto menor esse valor, melhor é o modelo [Gimenes, 2023].
- **Erro médio absoluto (MAE):** Calcula a média das diferenças absolutas entre os valores previstos e os valores reais. Em termos simples, isso mostra a média das diferenças entre as previsões do modelo e os valores reais [Klehm *et al.*, 2025].
- **Symmetric Mean Absolute Percentage Error (sMAPE):** É uma métrica percentual que representa o erro relativo de maneira simétrica, ao normalizar a diferença entre os valores previstos e observados pelos valores absolutos de ambos [Klehm *et al.*, 2025].

3 Trabalhos Relacionados

A previsão de variáveis climáticas, tais como temperatura, umidade e precipitação, é fundamental para a agricultura, fornecendo informações essenciais para o planejamento do plantio, manejo da irrigação e tomada de decisões estratégicas. A literatura recente demonstra um crescente interesse na aplicação de modelos de aprendizado de máquina como ferramentas de apoio à tomada de decisão agrícola diante da variabilidade climática.

Borella *et al.* [2022] realizaram previsões climáticas utilizando RNAs como apoio no processo de tomada de decisão para o plantio de alguns tipos de produtos agrícolas, tais como cenoura, maçã, uva e tomate. Foram usados os dados de precipitação total, velocidade média do vento, velocidade máxima média do vento, umidade, número de dias com precipitação, evaporação do piche, temperatura mínima, média e máxima. No final do estudo é apresentado um cronograma indicando a época propícia para plantar cada uma das culturas examinadas.

Rosa *et al.* [2022] aplicaram RNAs para estimar a produtividade de grãos de aveia considerando diferentes doses de nitrogênio e condições de cultivo. O experimento foi conduzido entre 2011 e 2016, com duas culturas de aveia branca ('Barbarasul' e 'Brisasul'), nas sequências soja/aveia e milho/aveia. As variáveis de entrada incluíram biomassa, dados meteorológicos e manejo nitrogenado, e a saída foi a produtividade de grãos, modelada no *Matlab* por meio do *toolbox Neural Network*. As RNAs mostraram alta precisão na previsão da produtividade, demonstrando ser um recurso eficiente

para estimar o rendimento da aveia em diferentes condições de manejo e clima.

Abraham *et al.* [2020] utilizaram RNAs do tipo *Nonlinear Autoregressive Exogenous* (NARX) para prever a área colhida, a produtividade e a produção de soja no Brasil, com dados de 1961 a 2016. O modelo contou com 10 neurônios, 6 *delays* e treinamento pelo algoritmo de *Levenberg–Marquardt* no *Matlab*, sendo comparado a modelos clássicos de séries temporais (linear, exponencial, polinomial). As RNAs apresentaram alto desempenho para área colhida e produção ($R > 0,99$), enquanto para produtividade o modelo linear foi mais eficaz. O estudo conclui que a RNA é uma abordagem confiável para previsão agrícola em séries temporais históricas.

Mesquita *et al.* [2022] realizaram uma análise climática para a previsão de precipitação na região de Sobral-CE, utilizando dados meteorológicos no intervalo de 2003 a 2020 obtidos por uma estação localizada na cidade, com o objetivo de auxiliar pequenos agricultores no gerenciamento de recursos hídricos. Os autores utilizaram diferentes algoritmos de aprendizado de máquina, tais como MLP, *Decision Tree* (DT) e KNN para a predição da precipitação. Ao final, observou-se que o MLP e o *Decision Tree* obtiveram altos índices de acertos, sugerindo a capacidade desses modelos no apoio à agricultura em regiões que podem sofrer com a estiagem.

Uma pesquisa conduzida no Paraná utilizou-se RNAs para prever a temperatura e a precipitação, avaliando os resultados em comparação com métodos tradicionais de persistência, que é uma técnica simples que assume que uma variável atmosférica tende a manter seu comportamento ao longo do tempo, apresentando dependência positiva entre valores sucessivos [Milléo *et al.*, 2020]. Os autores ressaltaram que a aplicação de médias móveis para filtrar os dados aumenta a exatidão das previsões, evidenciando que RNAs podem ser uma ferramenta eficaz para modelagem climática, particularmente em situações com variabilidade temporal considerável.

A **Tabela 1** apresenta uma comparação entre os trabalhos citados nesta seção e a abordagem proposta, apresentando o objetivo, variáveis climáticas analisadas, os algoritmos, a base de dados e o principal resultado encontrado. Diferentemente dos trabalhos relacionados, que em geral se apoiam em séries temporais curtas, dados diários e abordagens focadas isoladamente em precipitação ou temperatura, este estudo propõe uma modelagem de regressão baseada em dados mensais e em uma série histórica extensa, incorporando ainda a umidade relativa como variável climática adicional. Ademais, a realização de previsões conjuntas de precipitação, temperatura e umidade, aliada à análise comparativa de múltiplos algoritmos de aprendizagem de máquina, constitui o principal diferencial desta pesquisa. Essa abordagem integrada amplia a compreensão dos padrões climáticos locais e oferece uma ferramenta mais robusta de apoio ao planejamento da agricultura familiar, preenchendo lacunas relevantes identificadas na literatura.

4 Metodologia

A metodologia deste trabalho foi estruturada em cinco etapas principais, com o objetivo de organizar de forma sistemática o processo de predição climática para aplicações agrícolas: (i) coleta de dados, (ii) pré-processamento de dados climá-

ticos, (iii) desenvolvimento de modelos de aprendizado de máquina para predição da temperatura, precipitação e umidade, (iv) definição dos cenários experimentais e (v) avaliação do desempenho dos modelos.

A sequência detalhada dessas etapas e a integração do fluxo de informações estão ilustradas na **Figura 1**. Essa estrutura permite descrever de forma clara tanto os procedimentos de tratamento dos dados, quanto a construção e validação dos algoritmos. Cada etapa é detalhada a seguir.

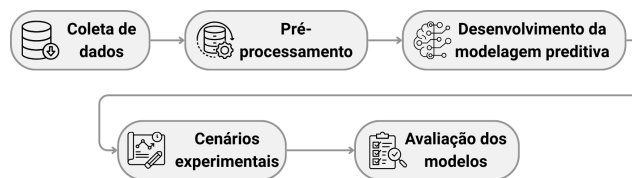


Figura 1. Fluxograma sequencial das etapas da metodologia proposta.

4.1 Coleta de dados

Os dados utilizados neste trabalho foram coletados a partir dos dados meteorológicos obtidos junto ao Instituto Nacional de Meteorologia (INMET) ¹.

Foi realizada a coleta de dados meteorológicos do município de Sobral, Ceará, no período de 31 de janeiro de 1973 a 31 de dezembro de 2019, correspondente ao intervalo máximo de registros disponíveis na estação meteorológica selecionada. Foram adquiridos dados de 46 anos, com valores mensais para as seguintes variáveis: evaporação do piche, nebulosidade, número de dias com precipitação pluvial, precipitação total, temperatura máxima média, temperatura média compensada, temperatura mínima média, umidade relativa do ar, velocidade máxima do vento e velocidade média do vento.

Portanto, o *dataset* utilizado neste trabalho possui 552 registros, com valores mensais para cada uma das variáveis analisadas.

4.2 Pré-processamento

O pré-processamento dos dados climáticos constitui uma etapa essencial para assegurar a qualidade das informações utilizadas no treinamento dos algoritmos.

Inicialmente, foi realizada a verificação do preenchimento dos valores das variáveis meteorológicas. A análise revelou a presença de valores ausentes em todas as variáveis, distribuídos ao longo de diferentes períodos. O tratamento desses dados faltantes considerou a natureza temporal, a sazonalidade e o comportamento específico de cada variável, com o objetivo de preservar padrões relevantes para a previsão de temperatura, precipitação e umidade. Assim, adotou-se a interpolação sazonal para o preenchimento dos valores ausentes, estratégia que permite manter as características temporais e sazonais das séries climáticas [Alejo-Sanchez *et al.*, 2025]. Portanto, realizou-se a imputação de valores ausentes por meio de interpolação sazonal, aplicada tanto à variável alvo quanto às variáveis predictoras, visando preservar os padrões temporais presentes nas séries de dados.

¹ <https://bdmep.inmet.gov.br/>

Tabela 1. Comparação entre os trabalhos da literatura e a abordagem proposta.

Trabalho	Objetivo	Variáveis analisadas	Algoritmos utilizados	Base de dados	Principal resultado
Borella et al. (2022)	Apoio à decisão de plantio agrícola	Temperatura, precipitação, umidade, vento e outras variáveis climáticas	RNA	Dados climáticos agrícolas	Definição de cronograma de plantio para diferentes culturas
Rosa et al. (2022)	Predição da produtividade de aveia	Biomassa, dados meteorológicos e nitrogênio	RNA	Dados experimentais (2011–2016)	Alta precisão na estimativa da produtividade
Abraham et al. (2020)	Predição da produção de soja	Área colhida, produtividade e produção	RNA NARX	Série histórica (1961–2016)	Elevado desempenho para área colhida e produção
Mesquita et al. (2022)	Previsão de precipitação em Sobral-CE	Precipitação	MLP, DT e KNN	Dados meteorológicos (2003–2020)	Bons resultados para MLP e DT
Milléo et al. (2020)	Previsão de temperatura e precipitação	Temperatura e precipitação	RNA	Dados meteorológicos do Paraná	RNAs apresentaram bom desempenho após aplicação de médias móveis
Este trabalho	Predição de variáveis climáticas para aplicações agrícolas	Temperatura, precipitação e umidade relativa do ar	DT, RF, XG-Boost, KNN e MLP	Série histórica mensal de Sobral-CE (1973–2019)	Comparação de múltiplos algoritmos com e sem memória temporal

4.3 Desenvolvimento da modelagem preditiva

Nesta etapa, foram desenvolvidos modelos de aprendizado de máquina voltados à predição de variáveis climáticas relevantes para aplicações agrícolas: temperatura média, precipitação e umidade relativa do ar. Para cada variável, foram desenvolvidos modelos específicos a partir de cinco algoritmos distintos, com o objetivo de explorar diferentes abordagens de aprendizado. A seguir, cada um dos modelos utilizados é descrito:

- **Decision Tree:** A configuração do modelo incluiu a limitação da profundidade máxima da árvore `max_depth = 5`, assim como restrições quanto ao número mínimo de amostras para divisão de nós `min_samples_split = 4` e para formação de folhas `min_samples_leaf = 2`, com o objetivo de reduzir o risco de sobreajuste.
- **Random Forest:** A configuração adotada utilizou `n_estimators = 200`, definindo o número de árvores na floresta, além de `random_state = 42` para reprodutibilidade dos resultados e `n_jobs = -1` para paralelização do processamento.
- **XGBoost:** A configuração adotada para este modelo incluiu `n_estimators = 600` e `learning_rate = 0.03`, controlando, respectivamente, o número de iterações do modelo e a taxa de aprendizado. A complexidade das árvores foi limitada por `max_depth = 4` e `min_child_weight = 3`, enquanto os parâmetros `subsample = 0.8` e `colsample_bytree = 0.7` foram utilizados para introduzir aleatoriedade no processo de treinamento, contribuindo para a redução de sobreajuste. Adicionalmente, foram empregados `gamma = 0.1` e `reg_lambda = 2.0` como mecanismos de regularização, e definida a função objetivo `objective = 'reg:squarederror'`, com avaliação baseada na mé-

trica `eval_metric = 'rmse'`. Assim como outros métodos baseados em árvores, não foi necessária a aplicação de normalização dos dados.

- **KNN:** Considerando que o algoritmo se baseia em medidas de distância entre amostras, foi realizada a normalização das variáveis de entrada por meio de padronização, a fim de evitar a influência desproporcional de variáveis com maior magnitude. O modelo foi configurado com `n_neighbors = 5`, definindo o número de vizinhos utilizados na etapa de predição.
- **MLP:** Devido à sua sensibilidade à escala dos dados, foi realizada a normalização das variáveis de entrada por meio de padronização. A configuração do modelo incluiu uma rede neural com duas camadas ocultas (`hidden_layer_sizes = (50, 25)`), função de ativação ReLU (`activation = 'relu'`) e otimização pelo algoritmo Adam (`solver = 'adam'`), com limite máximo de iterações definido como `max_iter = 1000`.

4.4 Cenários experimentais

Com o objetivo de avaliar o desempenho dos modelos em diferentes cenários de entrada, foram definidos cenários experimentais distintos aplicados a cada algoritmo descrito na seção anterior.

Para isso, foram avaliados dois cenários na construção dos modelos de predição das variáveis climáticas: (i) um cenário sem memória temporal, no qual as variáveis de entrada correspondem apenas aos valores observados no mês corrente; e (ii) um cenário com memória temporal, no qual foram incorporados, como preditores adicionais, os valores das variáveis referentes ao mês imediatamente anterior. A adoção de uma única defasagem temporal foi definida com base em análises preliminares, que indicaram que a inclusão de defasagens adicionais não produziu ganhos significativos no desempenho dos modelos.

4.5 Avaliação dos modelos

Para a avaliação dos modelos em ambos os cenários experimentais, utilizou-se a validação cruzada *k-fold* ($k = 5$) com os dados embaralhados. Destaca-se que, no cenário com memória temporal, o lag de 1 mês foi adicionado como um atributo de entrada antes da divisão dos dados. Como o objetivo do trabalho é comparar a capacidade dos algoritmos de aprender o comportamento das variáveis (e não prever o futuro em tempo real), o embaralhamento garante que os grupos de teste e treino sejam parecidos, sem destruir o histórico que foi guardado dentro de cada registro. As métricas de desempenho consideradas incluem MSE, MAE, sMAPE e R^2 , buscando capturar diferentes aspectos do erro e da capacidade preditiva dos algoritmos. O MSE e o MAE foram utilizados para quantificar o erro absoluto das previsões, enquanto o sMAPE permitiu uma análise relativa dos erros. O R^2 foi empregado para avaliar a capacidade explicativa dos modelos em relação à variabilidade dos dados observados.

5 Resultados e Discussão

Após a definição dos cenários experimentais e dos procedimentos de avaliação, são analisados os resultados obtidos pelos modelos de aprendizagem de máquina na predição das variáveis climáticas estudadas. Inicialmente, realiza-se uma análise descritiva das séries temporais de temperatura, precipitação e umidade relativa do ar, destacando suas principais características estatísticas e comportamentais. Em seguida, são apresentados e discutidos os desempenhos dos algoritmos avaliados, com base nas métricas MAE, sMAPE, MSE e coeficiente de determinação (R^2), permitindo a comparação entre as abordagens e a identificação dos modelos mais adequados para cada variável.

5.1 Análise dos Dados Climáticos

A base de dados utilizada apresenta limitações importantes, com registros disponíveis apenas até o ano de 2019. Ademais, as séries temporais apresentam lacunas no preenchimento dos registros, com percentuais de valores ausentes variando entre 12,41% e 38,3% nas variáveis analisadas, conforme apresentado na **Tabela 2**.

Tabela 2. Percentual de valores ausentes por variável

Variável	Valores ausentes (%)
Temperatura média	31,21
Umidade relativa do ar	30,67
Precipitação total	12,41
Velocidade média do vento	38,30
Velocidade máxima do vento	35,99
Número de dias com precipitação	35,82
Evaporação do piche	25,53
Nebulosidade	13,83
Temperatura mínima média	13,65
Temperatura máxima média	13,48

A **Figura 2** apresenta a série temporal da temperatura média antes do preenchimento dos dados ausentes, evidenciando lacunas ao longo do período analisado. A **Figura 3** mostra a série após o preenchimento. No período estudado, a temperatura variou entre 21,2°C e 29,2°C. Isso indica que,

dentro de todo o período estudado, a temperatura alcançou uma variação de 8°C. Nota-se que essa variável apresenta baixa amplitude de variação, com mudanças graduais entre os meses e ausência de oscilações abruptas. Além disso, observa-se um padrão sazonal definido e uma evolução relativamente suave ao longo do tempo. Essas características tendem a favorecer o desempenho de modelos de aprendizagem de máquina, uma vez que o processo de modelagem envolve o ajuste de relações mais estáveis e contínuas entre as variáveis.

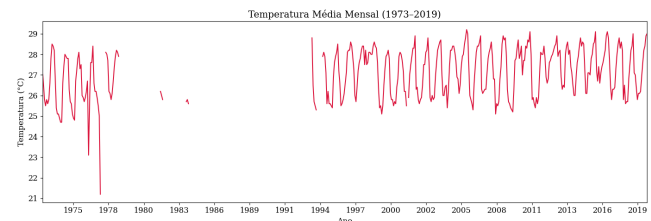


Figura 2. Temperatura média – antes da interpolação sazonal.

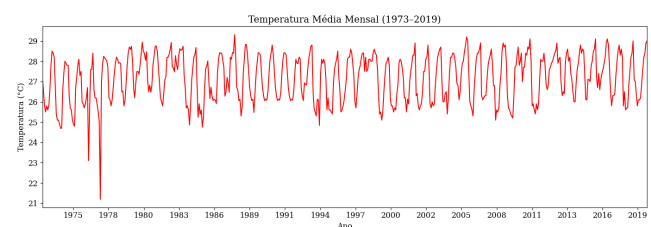


Figura 3. Temperatura média – após interpolação sazonal.

A **Figura 4** mostra o comportamento da precipitação antes do preenchimento dos dados ausentes, evidenciando falhas ao longo do intervalo analisado e elevada variabilidade temporal. A **Figura 5** indica a série após o preenchimento, observando-se a preservação do padrão irregular característico da variável. A precipitação apresentou amplitude de 517,0 mm, com valores mínimos de 0,0 mm e máximos de 517,0 mm, com predominância de meses com valores próximos de zero e ocorrência pontual de picos elevados, sem transições suaves entre os eventos. Todas essas particularidades tornam essa variável altamente instável, dificultando a aprendizagem do modelo.

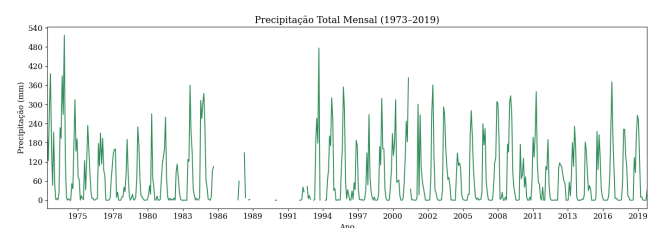


Figura 4. Precipitação – antes da interpolação sazonal.

O comportamento da umidade relativa antes do preenchimento dos dados ausentes pode ser observado na **Figura 6**. A **Figura 7** mostra a série após a imputação. A umidade relativa do ar mostrou uma variação moderada, superior à observada para a temperatura, porém significativamente inferior à da precipitação. Ao longo da série temporal analisada, a umidade apresenta registros mínimos de 51,7% e máximos de

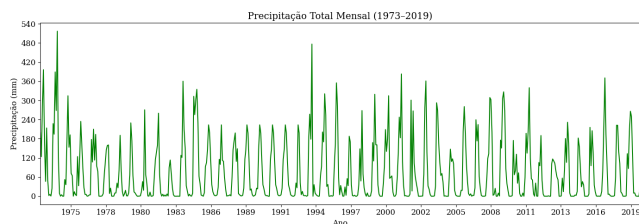


Figura 5. Precipitação – após interpolação sazonal.

95,0%, com amplitude observada de 43,3%. Observa-se que as mudanças ocorrem de forma relativamente gradual, sem oscilações abruptas e com distribuição contínua dos valores. Essas características tornam os dados parcialmente favoráveis ao aprendizado dos algoritmos, pois, embora apresentem um comportamento mais regular do que variáveis altamente instáveis (como a precipitação), ainda possuem variabilidade suficiente para exigir que os modelos capturem padrões sazonais e mudanças ao longo do tempo.

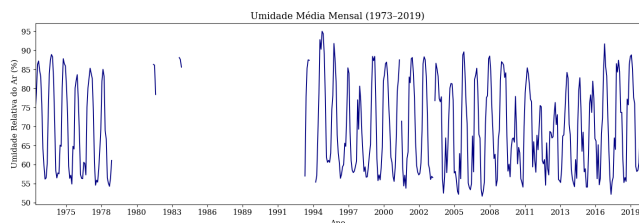


Figura 6. Umidade – antes da interpolação sazonal.

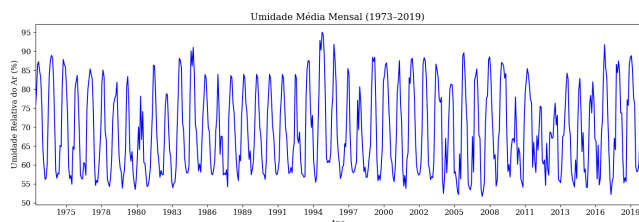


Figura 7. Umidade – após interpolação sazonal.

5.2 Algoritmos de Predição de Dados Climáticos

A compreensão da variabilidade e o preenchimento das lacunas dos dados climáticos dão suporte para a avaliação dos modelos de aprendizagem de máquina. Desse modo, o desempenho dos algoritmos foi comparado a partir de dois cenários experimentais: (i) análise sem memória temporal, e (ii) análise com memória temporal (*lag*). Para ambos os cenários, foram considerados modelos com diferentes algoritmos de aprendizagem de máquina, incluindo métodos baseados em árvores de decisão, técnicas de *ensemble*, algoritmos baseados em vizinhança e redes neurais artificiais. O objetivo é comparar o desempenho dessas técnicas na modelagem de predições de variáveis climáticas e verificar sua capacidade de capturar padrões presentes nos dados.

A Tabela 3 apresenta os resultados obtidos pelos modelos na predição das variáveis temperatura, precipitação e umidade relativa do ar. De forma geral, observa-se que o cenário com memória temporal apresenta refinou o desem-

penho dos algoritmos, embora a precisão continue variando conforme a natureza de cada variável alvo. Variáveis com forte sazonalidade e comportamento mais previsível apresentaram ajustes estatísticos superiores, enquanto a precipitação permaneceu como o desafio preditivo mais acentuado.

No caso da temperatura, os modelos no cenário com memória temporal apresentaram melhor desempenho. Os valores de MAE situaram-se entre 0,18 e 0,37, enquanto o sMAPE variou de 0,68 a 1,38. O MSE apresentou valores entre 0,12 e 0,35, e o R^2 oscilou entre 0,72 e 0,91, indicando um forte ajuste dos modelos à série térmica.

Para a precipitação, os resultados também são melhores para o cenário com memória temporal, apesar de confirmar a maior complexidade de modelagem, refletida em intervalos de erro mais amplos. Os valores de MAE variaram entre 21,57 e 26,00, enquanto o sMAPE situou-se entre 73,78 e 87,19. O MSE apresentou valores entre 1353,12 e 2018,57, e o R^2 variou de 0,78 a 0,85.

A variável umidade relativa do ar também exibiu o melhor comportamento médio entre os modelos no cenário com memória temporal. Os valores de MAE ficaram entre 2,03 e 2,74, com o sMAPE variando de 3,02 a 4,00. O MSE apresentou um intervalo entre 8,23 e 14,78, enquanto o R^2 atingiu patamares elevados, variando de 0,88 a 0,94.

Estes resultados reforçam que a predição de variáveis climáticas pode ser beneficiada pela inclusão de componentes autoregressivos. As subseções a seguir detalham o comportamento específico de cada algoritmo e identificam as técnicas que melhor performaram para cada variável individualmente.

5.2.1 Comportamento preditivo do modelo para a temperatura

Para a temperatura média, observaram-se os menores erros entre todas as grandezas analisadas, o que evidencia a estabilidade e a forte previsibilidade desta série. A baixa amplitude térmica e o padrão sazonal bem definido favoreceram uma modelagem altamente eficiente. No cenário com memória temporal, os algoritmos *XGBoost* e *Random Forest* apresentaram desempenhos equivalentes e superiores aos demais, com diferenças marginais de apenas 0,01 entre suas métricas. O *XGBoost* consolidou o melhor ajuste global com MSE de 0,12 e R^2 de 0,91, enquanto o *Random Forest* e o *XGBoost* destacaram-se com MAE de 0,18 e sMAPE de 0,68, conforme detalhado na Tabela 3.

O comportamento preditivo do *XGBoost* frente aos valores reais é ilustrado na Figura 8. Observa-se que o modelo replica com fidelidade quase integral a oscilação térmica mensal, sendo capaz de capturar tanto a tendência de longo prazo quanto as variações sazonais. Da mesma forma, o desempenho do modelo *Random Forest* é evidenciado na Figura 9.

Verifica-se uma elevada concordância entre os valores reais e os preditos pelos algoritmos *XGBoost* e *Random Forest*, com sobreposição quase completa das curvas ao longo da série temporal. Esse alinhamento indica que os modelos capturaram o comportamento sazonal da temperatura, acompanhando de forma consistente a oscilação mensal. As discrepâncias residuais são pequenas e concentram-se em pontos específicos da série, especialmente em eventos atípicos ou variações térmicas mais acentuadas que fogem ao padrão his-

Tabela 3. Resultados dos modelos para a predição de variáveis climáticas.

Cenário sem memória temporal						
Variável	Métrica	<i>Decision Tree</i>	RF	<i>XGBoost</i>	KNN	MLP
Temperatura	MAE	0,30	0,20	0,20	0,38	0,30
	sMAPE	1,12	0,75	0,76	1,42	1,13
	MSE	0,24	0,15	0,14	0,36	0,22
	R^2	0,81	0,88	0,89	0,71	0,83
Precipitação	MAE	27,42	22,28	22,31	23,33	22,65
	sMAPE	81,51	76,00	80,91	76,62	81,00
	MSE	2236,36	1420,48	1413,80	1666,06	1453,05
	R^2	0,75	0,84	0,84	0,82	0,84
Umidade relativa do ar	MAE	2,78	2,18	2,12	2,24	2,41
	sMAPE	4,07	3,21	3,12	3,30	3,54
	MSE	15,50	10,08	8,61	9,81	10,98
	R^2	0,88	0,92	0,93	0,92	0,91
Cenário com memória temporal						
Variável	Métrica	<i>Decision Tree</i>	RF	<i>XGBoost</i>	KNN	MLP
Temperatura	MAE	0,26	0,18	0,18	0,37	0,29
	sMAPE	0,96	0,68	0,68	1,38	1,08
	MSE	0,19	0,13	0,12	0,35	0,24
	R^2	0,85	0,90	0,91	0,72	0,81
Precipitação	MAE	26,00	21,73	21,57	22,67	22,59
	sMAPE	82,51	73,78	80,83	77,30	87,19
	MSE	2018,57	1409,22	1353,12	1521,32	1377,85
	R^2	0,78	0,85	0,85	0,83	0,85
Umidade relativa do ar	MAE	2,74	2,13	2,03	2,10	2,41
	sMAPE	4,00	3,14	3,02	3,09	3,54
	MSE	14,78	9,76	8,23	8,93	10,98
	R^2	0,88	0,92	0,94	0,93	0,91

tórico. Em suma, o ajuste visual confirma a robustez das métricas quantitativas, evidenciando que os modelos aprenderam satisfatoriamente o padrão da temperatura e entregou previsões estáveis, coerentes e com alta capacidade de generalização.

5.2.2 Comportamento preditivo do modelo para a precipitação

Para a variável precipitação, observam-se valores elevados de MSE, reflexo da sua alta variabilidade e distribuição assimétrica, caracterizada pela elevada frequência de dias sem chuva (0 mm) e pela ocorrência esporádica de eventos extremos. Embora a maior parte dos registros concentre-se entre 0 mm e 60 mm, a presença de picos que atingem aproximadamente 500 mm amplia significativamente a dispersão dos dados. Dado que o MSE penaliza erros de forma quadrática, os desvios associados a esses eventos extremos resultam em impactos expressivos na métrica de erro, tornando a modelagem um desafio acentuado.

Apesar da complexidade inerente à precipitação, os resultados na **Tabela 3** indicam um desempenho superior dos modelos baseados em *ensemble*. O *XGBoost* consolidou-se como o modelo de melhor precisão global, atingindo o menor

MSE (1353,12) e o maior R^2 (0,85), o que demonstra sua eficiência na minimização de desvios quadráticos. Paralelamente, o *Random Forest* destacou-se pela robustez na métrica sMAPE (73,78), sugerindo maior consistência na predição de valores relativos em comparação aos demais algoritmos.

Conforme evidenciado na **Figura 10**, o modelo *XGBoost* demonstra uma alta capacidade em replicar a dinâmica temporal da precipitação. Embora o algoritmo acompanhe com precisão a sazonalidade e a tendência geral da série, observa-se que, em episódios de precipitação extrema, o modelo ainda apresenta dificuldade em reproduzir a magnitude exata dos picos mais elevados. Percebe-se que, em eventos de alta intensidade, o *XGBoost* tende a suavizar valores extremos, resultando em subestimações pontuais em relação aos registros reais. Contudo, essa característica é menos acentuada do que nos modelos sem memória temporal, evidenciando que a inclusão da memória temporal permitiu ao algoritmo aprender com maior eficácia a estrutura dos dados, mitigando, ainda que não eliminando inteiramente, o desafio característico à previsão de eventos climáticos de grande magnitude.

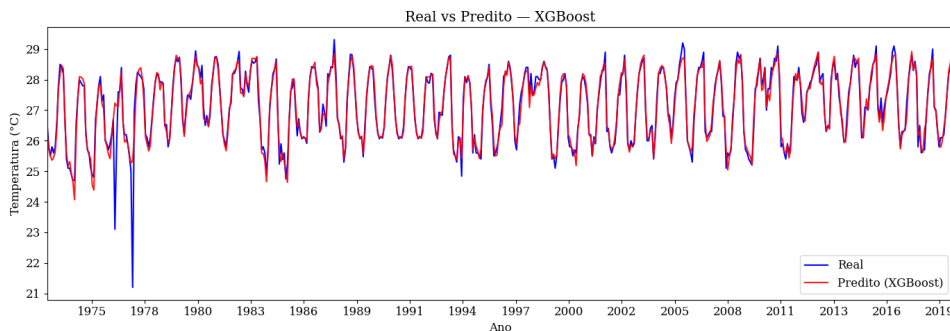


Figura 8. Valores observados e estimados pelo modelo *XGBoost* para a temperatura.

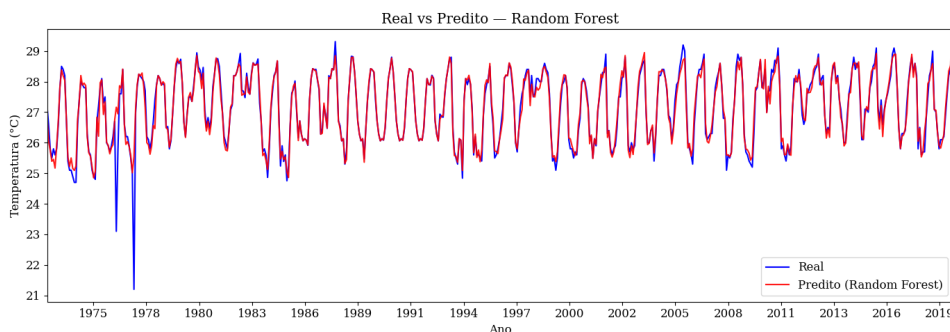


Figura 9. Valores observados e estimados pelo modelo *Random Forest* para a temperatura.

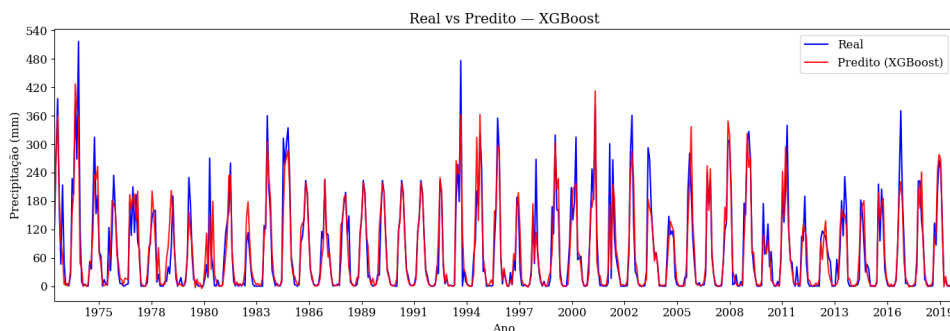


Figura 10. Valores observados e estimados pelo modelo *XGBoost* para a precipitação.

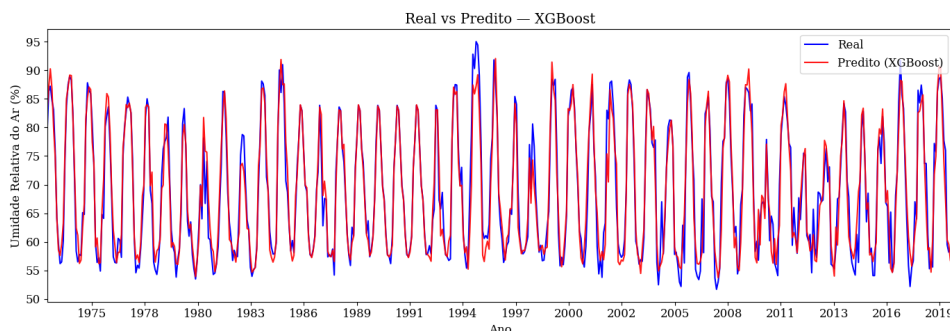


Figura 11. Valores observados e estimados pelo modelo *XGBoost* para a umidade relativa do ar.

5.2.3 Comportamento preditivo do modelo para a umidade relativa do ar

Para a umidade relativa do ar, os resultados indicam um desempenho superior e consistente, reflexo da variabilidade moderada e forte sazonalidade da série. O *XGBoost* destacou-se com os melhores indicadores em todas as métricas, atingindo um MAE de 2,03, sMAPE de 3,02, MSE de 8,23 e um coeficiente de determinação (R^2) de 0,94, conforme apresentado na **Tabela 3**.

Como evidenciado na **Figura 11**, observa-se uma elevada sobreposição entre os valores reais e as estimativas do modelo. O *XGBoost* foi capaz de replicar com precisão os ciclos sazonais e as variações de curto prazo, apresentando apenas discrepâncias mínimas em momentos de oscilações bruscas. Esse comportamento confirma a generalização do algoritmo para esta variável, demonstrando que a inclusão da memória temporal proporcionou um melhor ajuste à dinâmica da série real.

5.2.4 Comportamento preditivo geral dos modelos

A inclusão de memória temporal mostrou-se um fator determinante para o desempenho dos modelos preditivos. Variáveis climáticas apresentam forte dependência temporal, isto é, os valores observados em um determinado instante tendem a estar relacionados aos valores registrados em períodos anteriores. Ao incorporar memória temporal, os modelos passam a capturar padrões de autocorrelação, sazonalidade e persistência presentes nos fenômenos climáticos. Esse mecanismo amplia a capacidade de representação da dinâmica das séries, permitindo que os algoritmos aprendam não apenas relações entre variáveis, mas também a evolução temporal. Dessa forma, o cenário com memória temporal apresentou melhor generalização dos modelos.

De forma geral, nos cenários com série temporal, os resultados indicam que não houve um único algoritmo que dominasse de forma absoluta todas as métricas em todas as variáveis, embora os modelos baseados em *ensemble* tenham apresentado uma superioridade clara sobre os demais. Para a temperatura, o *XGBoost* e o *Random Forest* apresentaram desempenhos equivalentes, com diferenças marginais. Para a precipitação, o *XGBoost* consolidou-se como o modelo de melhor precisão global (menor MAE, MSE e maior R^2), enquanto o *Random Forest* manteve o mesmo valor de R^2 e uma ligeira vantagem em termos sMAPE. Já para a umidade relativa do ar, o *XGBoost* obteve os melhores resultados em todas as métricas, destacando-se pela alta capacidade de explicação da variância da série.

Esse comportamento é fundamentado na capacidade desses algoritmos em capturar relações não lineares e interações complexas entre as variáveis ao longo do tempo, pois combinam múltiplos estimadores para formar um preditor robusto, o que reduz a variância e melhora a generalização.

Além disso, os resultados evidenciam que a previsibilidade das variáveis climáticas está diretamente relacionada à sua variabilidade. Séries com forte componente sazonal e menor amplitude de ruído, como temperatura e umidade, permitiram ajustes quase integrais dos modelos. Em contrapartida, a precipitação impôs um desafio maior, resultando em erros quadráticos mais elevados devido à dificuldade dos modelos em atingir a magnitude total dos picos de chuva.

6 Conclusão

Este trabalho avaliou o desempenho de algoritmos de aprendizado de máquina na predição de variáveis climáticas: temperatura, precipitação e umidade relativa do ar, utilizando dados meteorológicos de uma série histórica mensal do município de Sobral-CE, obtidos junto ao INMET. Foram investigados diferentes algoritmos, incluindo *Decision Tree*, *Random Forest*, *XGBoost*, KNN e MLP. Os modelos foram avaliados em dois cenários experimentais: um sem a incorporação de memória temporal nas variáveis de entrada e outro com a inclusão de uma defasagem temporal correspondente ao mês anterior, permitindo analisar o impacto da informação temporal no desempenho das previsões.

Os resultados deste estudo evidenciam que a incorporação de dependência temporal é um fator determinante para o aprimoramento da capacidade preditiva de modelos de aprendizado de máquina aplicados a variáveis climáticas. De forma consistente, algoritmos baseados em *ensemble*, especialmente *XGBoost* e *Random Forest*, demonstraram maior robustez e desempenho superior, consolidando-se como abordagens mais adequadas para a modelagem de séries climáticas em contextos de elevada variabilidade.

A análise por variável indicou que séries mais estáveis e sazonais, como a temperatura, são mais previsíveis, alcançando MAE de 0,18 e R^2 de 0,91, enquanto a precipitação permanece como o principal desafio devido à sua alta variabilidade e ocorrência de eventos extremos, alcançando MAE de 21,57 e R^2 de 0,85. A umidade relativa do ar apresentou desempenho intermediário, com boa capacidade de modelagem, alcançando MAE de 2,03 e R^2 de 0,94.

Diante desses resultados, constata-se o potencial prático das modelagens preditivas desenvolvidas como ferramentas de suporte à tomada de decisão em aplicações agrícolas na região de Sobral-CE. A capacidade de antecipar o comportamento dessas variáveis fornece fundamentos essenciais para a definição estratégica das épocas de plantio, minimizando riscos de perda de safra por estresses hídricos ou térmicos. Além disso, as previsões viabilizam um manejo hídrico mais eficiente, auxiliando na otimização dos recursos na irrigação. Esse suporte tecnológico assume um papel ainda mais relevante no contexto da agricultura familiar local, historicamente vulnerável às incertezas climáticas, permitindo que pequenos produtores planejem suas atividades de cultivo de forma mais resiliente e baseada em dados.

Como principal contribuição, este estudo apresenta uma abordagem integrada para predição simultânea de múltiplas variáveis climáticas, utilizando uma base de dados de longo prazo e explorando diferentes algoritmos de aprendizado de máquina. Os resultados obtidos reforçam o potencial dessas técnicas como ferramentas de apoio ao planejamento agrícola, especialmente em contextos de agricultura familiar e regiões sujeitas à variabilidade climática.

Apesar dos avanços, limitações relacionadas à resolução temporal dos dados e à ausência de variáveis climáticas de larga escala indicam oportunidades de aprimoramento. Nesse sentido, estudos futuros devem explorar bases de dados mais granulares, integrar forçantes climáticas globais e avançar na conexão entre previsões meteorológicas e respostas agronômicas das culturas.

Declarações complementares

Agradecimentos

Os autores agradecem ao Instituto Federal do Ceará pelo suporte ao desenvolvimento deste trabalho.

Contribuições dos autores

Francisco Ribeiro foi responsável pela concepção do estudo, coleta e análise dos dados, desenvolvimento dos modelos e redação do manuscrito. Francisca Raquel de Vasconcelos Silveira contribuiu com a supervisão, revisão e validação do trabalho. Clemilton da Silva Ferreira contribuiu com a orientação, revisão técnica e validação dos resultados. Todos os autores leram e aprovaram a versão final do manuscrito.

Conflitos de interesse

Os autores declaram que não têm nenhum conflito de interesses.

Disponibilidade de dados e materiais

Os conjuntos de dados (e/ou softwares) gerados e/ou analisados durante o estudo atual serão feitos mediante solicitação.

Referências

- Abraham, E. R., Mendes dos Reis, J. G., Vendrametto, O., Oliveira Costa Neto, P. L. d., Carlo Toloi, R., Souza, A. E. d., and Oliveira Moraes, M. d. (2020). Time series prediction with artificial neural networks: An analysis using Brazilian soybean production. *Agriculture*, 10(10):475. DOI: 10.3390/agriculture10100475.
- Alejo-Sanchez, L. E., Márquez-Grajales, A., Salas-Martínez, F., Franco-Arcega, A., López-Morales, V., Acevedo-Sandoval, O. A., González-Ramírez, C. A., and Villegas-Vega, R. (2025). Missing data imputation of climate time series: A review. *MethodsX*, page 103455. DOI: 10.1016/j.mex.2025.103455.
- Bikku, T. (2020). Multi-layered deep learning perceptron approach for health risk prediction. *Journal of Big Data*, 7(1):50. DOI: 10.1186/s40537-020-00316-7.
- Borella, L. d. C., Borella, M. R. d. C., and Corso, L. L. (2022). Climate analysis using neural networks as supporting to the agriculture. *Gestão & Produção*, 29:e06. DOI: 10.1590/1806-9649-2022v29e06.
- Born, L., Prager, S., Ramirez-Villegas, J., and Imbach, P. (2021). A global meta-analysis of climate services and decision-making in agriculture. *Climate Services*, 22:100231. DOI: 10.1016/j.cliser.2021.100231.
- Chen, L., Han, B., Wang, X., Zhao, J., Yang, W., and Yang, Z. (2023). Machine learning methods in weather and climate applications: A survey. *Applied Sciences*, 13(21):12019. DOI: 10.3390/app132112019.
- Ferreira, V. P. and Batista, J. L. F. (2021). Utilização de variáveis qualitativas de localização em equações volumétricas ea seleção de modelos. *Ciência Florestal*, 31(4):1968–1990. DOI: 10.5902/1980509840926.
- Filippi, P., Han, S. Y., and Bishop, T. F. (2025). On crop yield modelling, predicting, and forecasting and addressing the common issues in published studies. *Precision Agriculture*, 26(1):8. DOI: 10.1007/s11119-024-10212-2.
- Garcia, D., Silva, N., Rolim, J., Ferreira, A., Santos, J. A., Caimera, M. d. R., and Paredes, P. (2025). Prediction of crops cycle with seasonal forecasts to support decision-making. *Agronomy*, 15(6):1291. DOI: 10.3390/agronomy15061291.
- Giromes, L. E. (2023). Modelos não lineares aplicados às curvas de casos acumulados e óbitos da COVID-19 a nível Brasil e São Paulo. Disponível em: <http://hdl.handle.net/11449/239409>.
- Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., and Khraisat, A. (2024). Enhancing k-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, 11(1):113. DOI: 10.1186/s40537-024-00973-y.
- Klehm, R. O., Parreira, W. D., Dazzi, R. L. S., Fernandes, A. M. d. R., García, D. C., and González, G. V. (2025). A hybrid prediction model using statistical forecasters and deep neural networks. *Applied Sciences*, 15(23):12393. DOI: 10.3390/app152312393.
- Lima, R. C., Holand, F. S. R., Pimenta, V., and Alagão, M. L. (2025). Mudanças climáticas e adaptação produtiva na agricultura familiar do semiárido nordestino. Technical report, Cebrap. Disponível em: https://inclusaoprodutivarural.cebrap.org.br/website/wp-content/uploads/2025/11/RelatorioFinal_estudo1_MudancasClimaticas_2025.pdf.
- Mesquita, I. M. d., Carneiro, F. T., Alves, S. F., Albuquerque, L. T., and Aragão, F. A. P. (2022). Análise de algoritmos de aprendizagem de máquina para previsão de precipitações para utilização na agricultura familiar. In *Escola Regional de Computação do Ceará, Maranhão e Piauí (ERCEMAPI)*, pages 179–187. SBC. DOI: 10.5753/erce-mapi.2022.226472.
- Milléo, C., Almeida, R. C. d., et al. (2020). Emprego de redes neurais artificiais na previsão climática de temperatura e precipitação no estado do paraná. Disponível em: <https://hdl.handle.net/1884/69952>.
- Mishra, R., Shirsath, V., and Kadam, N. (2024). Machine learning for precision agriculture: Predictive analysis of crop growth frequencies. *International Journal of Intelligent Systems and Applications in Engineering*, 12(21s):986–997. Disponível em: <https://ijisae.org/index.php/IJISAE/article/view/5497>.
- Moore, A. and Bell, M. (2022). XGBoost, a novel explainable AI technique, in the prediction of myocardial infarction: A UK Biobank cohort study. *Clinical Medicine Insights: Cardiology*, 16:11795468221133611. DOI: 10.1177/11795468221133611.
- Oliveira, S. F., Prado, R. B., and Monteiro, J. M. G. (2022). Impactos das mudanças climáticas na produção agrícola e medidas de adaptação sob a percepção de atores e produtores rurais de Nova Friburgo, RJ. *Interações (Campo Grande)*, 23(4):1179–1201. DOI: 10.20435/inter.v23i4.3548.
- Padhiary, M., Kumar, A., and Sethi, L. N. (2025). Emerging technologies for smart and sustainable precision agriculture. *Discover Robotics*, 1(1):6. DOI: 10.1007/s44430-025-00006-0.
- Pugliese, R., Regondi, S., and Marini, R. (2021). Machine learning-based approach: Global trends, research directions, and regulatory standpoints. *Data Science and Management*, 4:19–29. DOI: 10.1016/j.dsm.2021.12.002.

- Rosa, J. A. d., Mantai, R. D., Peter, C. L., Basso, N. C. F., Jung, J. S., Bandeira, W. J. A., Carvalho, I. R., and da Silva, J. A. G. G. (2022). Redes neurais artificiais na previsibilidade da produtividade de grãos da aveia pelo uso do nitrogênio e condições do ciclo de desenvolvimento da cultura. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, 9(1). DOI: 10.5540/03.2022.009.01.0296.
- Saha, S., Kucher, O. D., Utkina, A. O., and Rebouh, N. Y. (2025). Precision agriculture for improving crop yield predictions: a literature review. *Frontiers in Agronomy*, 7:1566201. DOI: 10.3389/fagro.2025.1566201.
- Salman, H. A., Kalakech, A., and Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, 2024:69–79. DOI: 10.58496/BJML/2024/007.
- Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, 2(3):1–21. DOI: 10.1007/s42979-021-00592-x.
- Sheikh, H., Prins, C., and Schrijvers, E. (2023). Artificial intelligence: definition and background. In *Mission AI: The new system technology*, pages 15–41. Springer. DOI: 10.1007/978-3-031-21448-6_2.
- Silva, R. and Silva Neto, D. R. d. (2023). Inteligência artificial e previsão de óbito por covid-19 no brasil: uma análise comparativa entre os algoritmos logistic regression, decision tree e random forest. *Saúde Em Debate*, 46:118–129. DOI: 10.1590/0103-11042022E809.
- Woodman, R. J. and Mangoni, A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: present and future. *Aging Clinical and Experimental Research*, 35(11):2363–2397. DOI: 10.1007/s40520-023-02552-2.
- Zhang, H., Liu, Y., Zhang, C., and Li, N. (2025). Machine learning methods for weather forecasting: A survey. *Atmosphere*, 16(1):82. DOI: 10.3390/atmos16010082.