

Previsão climática no Nordeste de Goiás: Aplicações de Big Data e Machine Learning

Climate Forecast in the Northeast of Goiás: Big Data and Machine Learning Applications

Gean Vitor Macedo de Sousa¹

Juliana Alves Vieira²

Sandir Rodrigues Campos³

<https://doi.org/10.5281/zenodo.17468577>

Resumo: Este trabalho propõe o desenvolvimento de um modelo preditivo que combina técnicas de Big Data e Machine Learning para análise climática na região do Nordeste Goiano. O estudo tem como foco principal a previsão de eventos climáticos extremos, como chuvas intensas e secas severas, com o objetivo de apoiar a tomada de decisões estratégicas nos setores agrícola e de gestão de recursos hídricos. A metodologia adotada envolve a coleta de dados multivariados oriundos de diversas fontes, incluindo satélites da NASA e ESA, estações meteorológicas do INMET e redes de sensores ambientais. O processamento desses dados será conduzido com o uso de frameworks robustos, como Apache Hadoop e Spark, permitindo a análise de grandes volumes de informação em tempo real. A etapa de modelagem utilizará algoritmos avançados de aprendizado de máquina, como Random Forest, que é um método baseado em múltiplas árvores de decisão o qual melhora a precisão ao reduzir o risco de overfitting, e XGBoost, uma técnica de gradient boosting altamente eficiente que combina diversos modelos fracos para formar um modelo robusto e preciso. Também serão aplicadas redes neurais para identificar padrões complexos em variáveis climáticas críticas, como precipitação, temperatura, umidade relativa do ar e concentração de gases do efeito estufa. Os resultados obtidos terão aplicação prática no planejamento agrícola regional, na otimização de sistemas de irrigação e na redução de riscos relacionados a eventos climáticos adversos, contribuindo de forma significativa para o desenvolvimento sustentável da região e para a segurança alimentar, gerando benefícios econômicos e ambientais.

Palavras-chave: *Machine Learning. Big Data. Python. Análise de dados.*

Abstract: This work proposes the development of a predictive model that combines Big Data and Machine Learning techniques for climate analysis in the Northeast region of Goiás, Brazil. The main focus of the study is the forecasting of extreme weather events, such as heavy rainfall and severe droughts, aiming to support strategic decision-making in the agricultural and water resource management sectors. The adopted methodology involves the collection of multivariate data from various sources, including NASA and ESA satellites, INMET meteorological stations, and environmental sensor networks. The processing of this data will be conducted using robust frameworks such as Apache Hadoop and Spark, enabling real-time analysis of large volumes of information. The modeling phase will employ advanced machine learning algorithms, such as Random Forest, a method based on multiple decision trees that enhances accuracy by reducing the risk of overfitting, and XGBoost, a highly efficient gradient boosting technique that combines several weak models to build a strong and accurate predictive model. Neural networks will also be applied to identify complex patterns in critical climate variables such as precipitation, temperature, relative humidity, and greenhouse gas concentrations. The results obtained will have practical applications in regional agricultural planning, irrigation system optimization, and risk reduction related to adverse climate events, significantly contributing to the region's sustainable development and food security, while generating economic and environmental benefits.

Keywords: *Machine learning, big data, python, data analysis.*

¹ Aluno. Iesgo. ORCID 0009-0005-7729-5681. E-mail: geanv7820@gmail.com

² Especialista. Iesgo. ORCID 0009-0003-5902-4373. E-mail: julianaalves.alves8@gmail.com.

³ Mestre. Iesgo. ORCID 0009-0000-4086-5752. E-mail: sandirr@gmail.com.

Introdução

Este trabalho tem como objetivo identificar e analisar padrões climáticos, uma tarefa cuja relevância tem se intensificado progressivamente no contexto contemporâneo. A capacidade de prever eventos meteorológicos proporciona benefícios que se estendem desde decisões operacionais simples até estratégias mais complexas de planejamento. No âmbito cotidiano, essa capacidade influencia escolhas práticas relacionadas à mobilidade urbana e segurança individual, permitindo que cidadãos se preparem melhor para condições adversas, reduzindo riscos e promovendo maior qualidade de vida.

No setor agrícola, por sua vez, essa capacidade exerce impacto direto na gestão eficiente de recursos hídricos, especialmente em propriedades que utilizam sistemas de irrigação mecanizada, como os pivôs centrais — estruturas responsáveis pela distribuição de água em movimentos circulares sobre extensas áreas cultivadas. A antecipação de variações climáticas torna-se, portanto, um elemento essencial para a redução de riscos, a otimização de recursos e a promoção da sustentabilidade na atividade agrícola, contribuindo para maior produtividade, menor desperdício e adaptação às mudanças climáticas.

Na região Centro-Oeste do Brasil, especialmente no nordeste goiano, o clima é caracterizado por duas estações bem definidas: a chuvosa e a seca. A alternância entre períodos de abundância e escassez hídrica acarreta desafios significativos, sobretudo durante a estiagem, quando a demanda por irrigação se intensifica. Durante longos períodos de estiagem, torna-se imprescindível a adoção de estratégias que promovam o uso racional da água, assegurando o abastecimento necessário sem comprometer a sustentabilidade do recurso.

Para além do setor agrícola, a análise de dados climáticos desempenha um papel essencial na formulação de políticas públicas eficazes, voltadas à prevenção e à resolução de problemas urbanos recorrentes, como enchentes e inundações. Esses eventos são comuns em grandes centros urbanos, especialmente naqueles estabelecidos próximos a corpos d'água, onde a impermeabilização do solo e o crescimento desordenado agravam os impactos. Nesses contextos, a utilização de sistemas de previsão meteorológica baseados em inteligência artificial, como o Aardvark Weather (Vaughan, et al., 2024), que integra redes neurais profundas e dados em tempo real para gerar previsões hiperlocais, tem se mostrado uma ferramenta promissora. Sua aplicação pode proporcionar respostas mais ágeis e precisas, colaborando com a gestão eficiente dos recursos e a implementação de medidas preventivas contra desastres naturais.

Essas informações são fundamentais para embasar decisões estratégicas, planos de ação e políticas públicas. No setor agropecuário, viabilizam um planejamento mais eficiente das safras, com otimização no uso da água e redução de perdas. Também auxiliam na definição dos períodos ideais para plantio e colheita, aumentando a produtividade. No ambiente urbano, contribuem para a melhoria dos sistemas de drenagem e das ações de mitigação de enchentes. Dessa forma, tanto as áreas rurais quanto urbanas são beneficiadas, promovendo o desenvolvimento sustentável.

O Impacto do Big Data na Meteorologia Brasileira

O uso de Big Data na meteorologia tem crescido com a coleta de dados de satélites, sensores e mais de 700 estações meteorológicas operadas pelo INMET, que captam variáveis como precipitação, ventos e pressão atmosférica (Instituto Nacional de Meteorologia – INMET, 2025). Esses dados são integrados em modelos numéricos para gerar previsões mais precisas. O INPE também coleta informações por meio de satélites SCD e plataformas de coleta de dados distribuídas no país (INPE, 2025). Aliados a técnicas de Machine Learning, esses dados permitem prever eventos extremos como secas e chuvas intensas (Silva, et al., 2025), além de apoiar ações de mitigação e adaptação às mudanças climáticas (Oyarzabal, et al., 2023).

Machine Learning e Previsão Climática no Nordeste Goiano

No Nordeste Goiano, o clima semiárido e a escassez de dados históricos de precipitação apresentam desafios específicos para a modelagem climática. No entanto, o avanço das tecnologias de Big Data e Machine Learning tem permitido a criação de modelos mais precisos. Um estudo publicado na revista *Environmental Data Science* destaca que a aplicação de técnicas de Machine Learning, como UNet++ e ResNet, tem se mostrado eficaz na análise de grandes volumes de dados climáticos, como os pluviométricos. Isso contribui para o aprimoramento dos sistemas de organização do conhecimento e para uma melhor compreensão das dinâmicas climáticas da região. (Unal, et al., 2023)

Desafios e Oportunidades no Uso de Dados Climáticos no Semiárido

A escassez de estações meteorológicas e a dificuldade no acesso a dados atualizados constituem limitações significativas para a plena utilização das tecnologias de análise climática no semiárido. A integração de dados de sensoriamento remoto e a aplicação de técnicas avançadas de inteligência artificial, sendo essas estratégias capazes de permitir a geração de indicadores

climáticos mais precisos. Isso é especialmente relevante em regiões com baixa densidade de observação.

Iniciativas Locais e Projeções Futuras

Diversas iniciativas locais têm buscado integrar dados de diferentes fontes para aprimorar a análise e a previsão climática no Nordeste Goiano. Destaca-se o trabalho desenvolvido pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa) (Embrapa Cerrados), que realiza estudos sobre o impacto das mudanças climáticas na agricultura da região, utilizando ferramentas de modelagem agroclimática e análise de dados em larga escala. Por sua vez, o Instituto Nacional de Pesquisas Espaciais (INPE) mantém projetos de monitoramento climático e previsão de eventos extremos por meio de plataformas como o BDMEP (Banco de Dados Meteorológicos para Ensino e Pesquisa), que fornece acesso a dados históricos e em tempo real (INPE).

Big Data na Análise Climática

O conceito de Big Data é definido como uma abordagem capaz de processar dados que não podem ser manipulados, analisados ou catalogados por métodos convencionais devido à sua quantidade, complexidade e variedade. No campo da climatologia, o uso do Big Data oferece benefícios significativos. Uma das principais vantagens está no fato de possibilitar o processamento e a coleta de dados provenientes de múltiplas fontes e tipos, como satélites, estações meteorológicas e sensores remotos.

Tecnologias baseadas em Big Data permitem a detecção de padrões climáticos e a previsão de eventos extremos, como tempestades, furacões e secas. Isso possibilita a emissão de alertas antecipados para comunidades vulneráveis e o planejamento mais eficaz de ações emergenciais e agrícolas. Nesse contexto, o uso de Big Data na previsão de padrões climáticos é fundamental para compreender a dinâmica de regiões específicas. No Centro-Oeste do Brasil, por exemplo, a análise climática é complexa devido à geografia e topografia da região.

Essa complexidade se reflete na ocorrência de dois regimes sazonais distintos — chuvoso e de estiagem — cujas durações variam ao longo dos anos. Identificar esses padrões torna-se essencial para áreas como a agricultura, que é uma das principais atividades econômicas da região.

Importância da Coleta de Dados Climáticos

A coleta de dados climáticos é fundamental para a análise e previsão de fenômenos atmosféricos. Entretanto, essa atividade requer uma infraestrutura robusta e, em grande parte, de responsabilidade governamental. Fontes como o WorldView da NASA (NASA), o satélite Sentinel da Agência Espacial Europeia (ESA) (European Space Agency – ESA), dados abertos disponibilizados pelo INMET (Instituto Nacional de Meteorologia – INMET) e APIs públicas, como a Weather API (Weather API Documentation), desempenham um papel essencial nesse processo. As plataformas coletam e organizam dados provenientes de diversas regiões do mundo, reunindo informações cruciais, como a concentração de gases do efeito estufa, cobertura de nuvens, e temperatura da superfície terrestre, essas informações são insumos críticos para a modelagem climática preditiva.

Plataformas e Ferramentas de Big Data Aplicadas ao Clima

O Big Data enfrenta desafios significativos na análise de dados estruturados e não estruturados, além do enorme volume de informações envolvidas. Ferramentas como o Apache Hadoop se mostram eficazes para superar essas limitações. O Hadoop é um framework de código aberto projetado para manipular grandes volumes de dados de forma distribuída, dividindo-os em blocos e tratando-os simultaneamente em diversos nós de um cluster de servidores.

A arquitetura distribuída do Hadoop é especialmente vantajosa no tratamento de informações provenientes de satélites e outras fontes, bem como na manipulação de dados históricos e séries temporais. Essas funcionalidades do Hadoop permitem identificar padrões ao longo do tempo e prever eventos futuros, como mudanças sazonais e fenômenos climáticos extremos.

O Apache Spark, por outro lado, realiza o processamento em tempo real, destacando-se pela alta velocidade no tratamento de grandes conjuntos de dados. Diferentemente do Hadoop, o Spark executa as operações diretamente em memória, o que proporciona ganhos significativos de desempenho em situações que exigem respostas rápidas, como na análise contínua de dados de sensores e na detecção de padrões climáticos emergentes.

Devido a essas características, o Spark é ideal para lidar com dados climáticos em tempo real, como os gerados por sensores e satélites. A análise imediata é fundamental para prever eventos extremos com agilidade e precisão (Amazon Web Services).

Objetivos

Este artigo tem como objetivo principal avaliar a eficácia do uso integrado de Big Data e técnicas de Machine Learning pode contribuir na previsão de eventos climáticos, com ênfase na região do Nordeste Goiano.

A pesquisa busca medir o desempenho dessas tecnologias na análise de grandes volumes de dados climáticos. O objetivo é gerar informações relevantes que subsidiem estratégias eficazes em áreas sensíveis às variações do clima, como a agricultura e a gestão de recursos hídricos.

- a. Analisar grandes volumes de dados climáticos provenientes de fontes diversas.
- b. Verificar a viabilidade e a qualidade dos dados obtidos por meio de fontes abertas.
- c. Avaliar a aplicação desses dados em setores econômicos relevantes, como a agricultura.
- d. Desenvolver um modelo preditivo de Machine Learning voltado a eventos climáticos extremos.

Material

Os materiais utilizados neste estudo compreendem fontes de dados climáticos, ferramentas de Big Data e tecnologias complementares, descritos a seguir:

A. Fontes de Dados Climáticos:

- a. NASA WorldView Plataforma de sensoriamento remoto global (NASA, 2025).
- b. Sentinel Satellites (ESA): Satélites de observação que fornecem dados em tempo real (European Space Agency – ESA, 2025).
- c. INMET: Instituto Nacional de Meteorologia, órgão brasileiro responsável por séries históricas climáticas. (Instituto Nacional de Meteorologia – INMET, 2025).
- d. Weather API: API aberta que oferece informações meteorológicas globais (Weather API Documentation, s.d.).

Esses materiais permitiram a estruturação de um pipeline de análise que abrange desde a coleta automatizada de dados até a construção e validação de modelos preditivos.

b) Ferramentas e Plataformas de Big Data:

- a. Apache Hadoop: Framework para processamento distribuído de grandes volumes de dados (Apache Software Foundation, 2025).
- b. Apache Spark: Ferramenta para processamento em tempo real de dados em larga escala (Apache Software Foundation, 2025).

c) Ambiente de Desenvolvimento e Armazenamento:

- a. Python: Utilizada no desenvolvimento de scripts de processamento de dados (Python Software Foundation, 2025).
- b. Bibliotecas de Análise de Dados: Pandas, NumPy, Matplotlib: bibliotecas para análise numérica e visualização. (Pandas, 2025) (NumPy, 2025) (Matplotlib, 2025).
- c. Hadoop HDFS: sistema de arquivos distribuído para armazenamento de grandes volumes de dados. (Apache Software Foundation, 2025).

Metodologia

O processo metodológico deste estudo será dividido em três etapas principais: coleta de dados, processamento de dados e análise preditiva.

Coleta de Dados

A coleta de variáveis meteorológicas será realizada por meio de fontes externas, como as APIs da NASA, ESA, INMET e Weather API, que fornecem informações de sensores remotos e satélites. Dados históricos também serão coletados diretamente do INMET e de outras bases de dados climáticos. (NASA, 2025) (Instituto Nacional de Meteorologia – INMET, 2025) (Weather API Documentation, s.d.)

Entre essas bases destacam-se:

- ERA5: produzida pelo ECMWF, fornece reanálises climáticas globais com alta resolução temporal e espacial.
- CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data): oferece estimativas de precipitação quase globais, combinando dados de satélite com observações em superfície.

- MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2): fornece dados para reanálise da NASA com dados atmosféricos, terrestres e oceânicos.
- WorldClim: fornece variáveis bioclimáticas históricas e projetadas, utilizadas em modelagens ecológicas e estudos de impacto climático.

Os dados a serem coletados incluem:

- a) Temperatura da superfície terrestre
- b) Cobertura de nuvens
- c) Concentração de gases do efeito estufa
- d) Dados de Precipitação
- e) Umidade relativa do ar

Processamento de Dados

O processamento dos dados é dividido em duas etapas principais: tempo real e histórico. Cada etapa utiliza ferramentas específicas, conforme o volume de dados e a resposta exigida.

- 1) Hadoop: Trata-se de uma plataforma de software de código aberto amplamente utilizada para o armazenamento e o processamento distribuído de grandes volumes de dados. No contexto deste trabalho, o Hadoop foi empregado para lidar com extensos conjuntos de registros históricos meteorológicos. Ele organiza os dados em blocos distribuídos por diferentes nós por meio do HDFS (Hadoop Distributed File System) e permite o processamento paralelo por meio do modelo MapReduce. O MapReduce divide o processamento em duas etapas principais: Map, que realiza a leitura e transformação dos dados em pares chave-valor, e Reduce, que agrega os resultados de acordo com as chaves. Essa abordagem viabiliza a análise eficiente de séries temporais e a detecção de padrões sazonais complexos em grandes volumes de dados (Apache Software Foundation, 2025).
- 2) Python: Para o processamento, análise exploratória, limpeza e tratamento dos dados, a fim de filtrar e determinar quais são úteis e podem ser melhor aproveitados na análise e criação do modelo (Python Software Foundation, 2025).

Análise Preditiva

Para a análise preditiva, são utilizados diferentes algoritmos de aprendizado de máquina Machine Learning, tanto supervisionados quanto não supervisionados, todos treinados com dados históricos. Entre os métodos supervisionados, destacam-se a floresta aleatória Random Forest, o Extreme Gradient Boosting (XGBoost) e as máquinas de vetor de suporte ou Support Vector Machines (SVM), que são aplicados em tarefas de regressão e classificação, como a previsão de eventos climáticos — por exemplo, estiagens prolongadas, chuvas intensas e riscos de inundação.

Algoritmos não supervisionados, como o K-Means, também são utilizados para a identificação de padrões ocultos e agrupamentos naturais nos dados, como regiões com comportamento climático semelhante. A avaliação dos modelos supervisionados é realizada com base em métricas como raiz do erro quadrático médio ou Root Mean Square Error (RMSE) e precisão.

Os resultados são apresentados em gráficos e mapas interativos, que permitem a visualização espacial e temporal das previsões. Esses mapas incluem, por exemplo, visualizações geográficas em que diferentes regiões são coloridas de acordo com o risco climático previsto (como intensidade de chuva ou probabilidade de seca), com funcionalidades de zoom, seleção de períodos específicos e sobreposição de camadas de dados. Ferramentas como Folium, Plotly e GeoPandas são utilizadas para construir essas representações dinâmicas e interativas.

Resultados

Análise exploratória

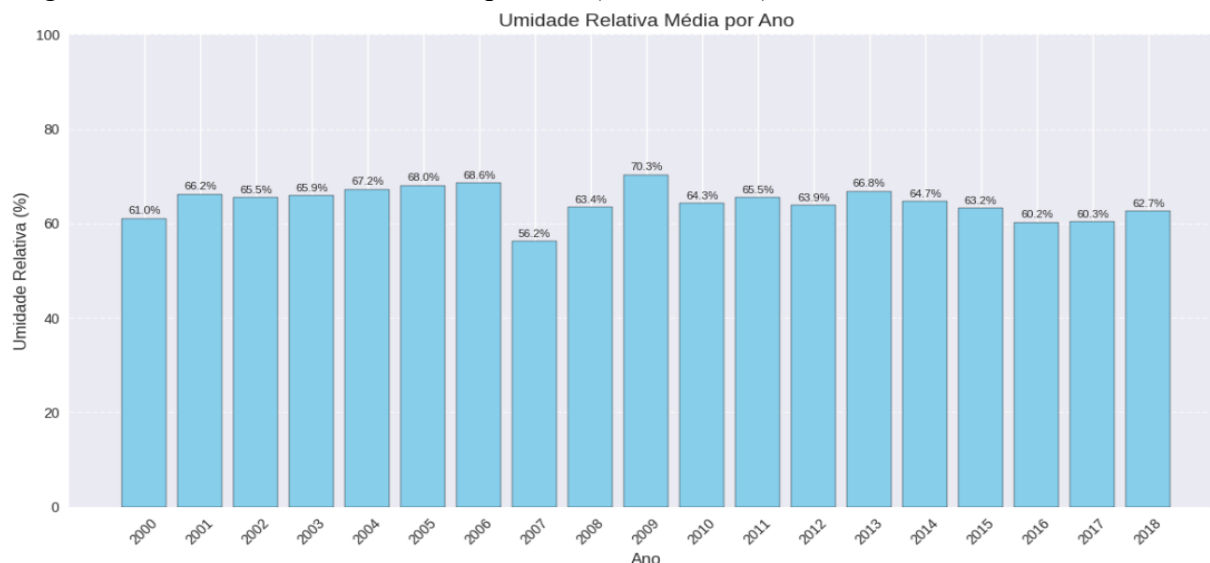
A análise exploratória é uma etapa fundamental para compreender a estrutura dos dados disponíveis. Entre seus principais objetivos estão a detecção de outliers e a avaliação do desbalanceamento na classe alvo (target), por serem cruciais à consistência estatística dos modelos preditivos.

A umidade relativa do ar é um dos principais elementos meteorológicos a serem analisados em estudos climatológicos. A Figura 1 apresenta a variação da umidade relativa média anual na região Nordeste de Goiás no período de 2000 a 2018. Verificou-se uma tendência de oscilação, com destaque para o ano de 2007, que apresentou o menor valor médio

de umidade (56,2%), e o ano de 2009, com o maior valor (70,3%). Observa-se ainda uma tendência de declínio nos valores médios a partir de 2014, o que pode indicar mudanças nas condições atmosféricas locais ou influência de eventos extremos, como secas prolongadas.

A Figura 1 apresenta a distribuição da umidade relativa média anual na região Nordeste de Goiás, entre os anos de 2000 e 2018, com variação entre 56,2% e 70,0%.

Figura 1 – Umidade Relativa Média por ano (2000 – 2018).

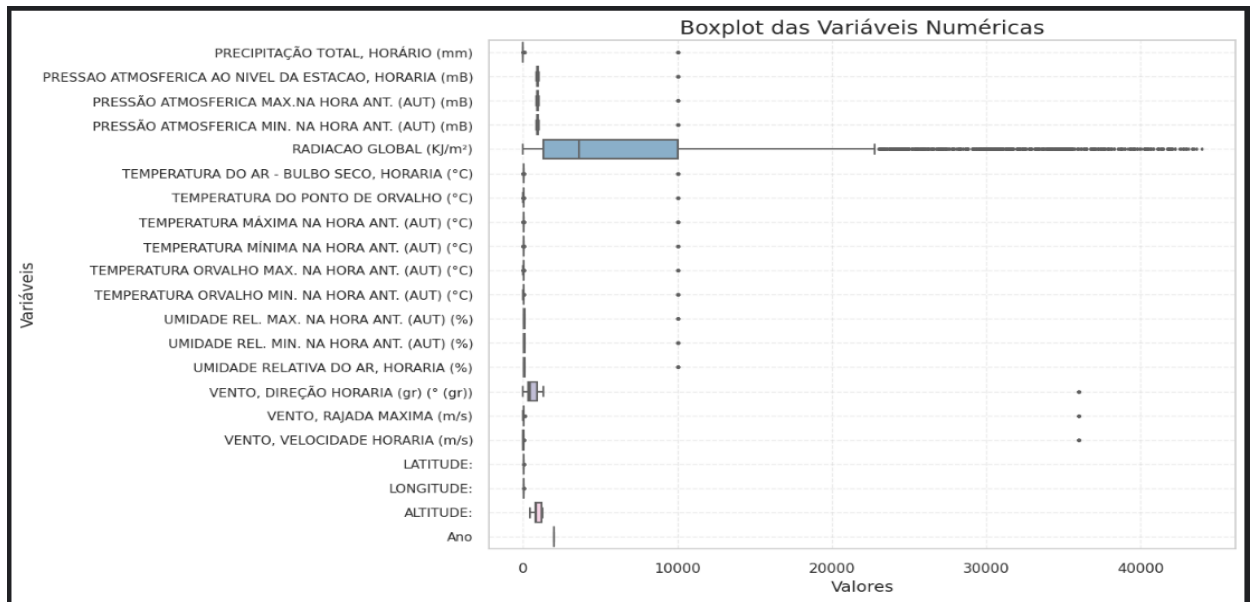


Fonte: Elaborado pelos autores (2025).

Em contrapartida outros valores do conjunto de dados, apresentam outliers extremos em como mostra a Figura 2, o que é comum em dados de precipitação onde eventos raros de chuva intensa ocorrem, ou podem representar registros inconsistentes ou inválidos.

A figura 2 apresenta um Gráfico de Boxplot das variáveis numéricas meteorológicas registradas na região Nordeste de Goiás. A análise permite identificar possíveis outliers nos dados coletados entre os anos de 2000 e 2018, com destaque para valores extremos em pressão atmosférica e radiação global.

Figura 2 – Gráfico de Boxplot para análise de outliers.

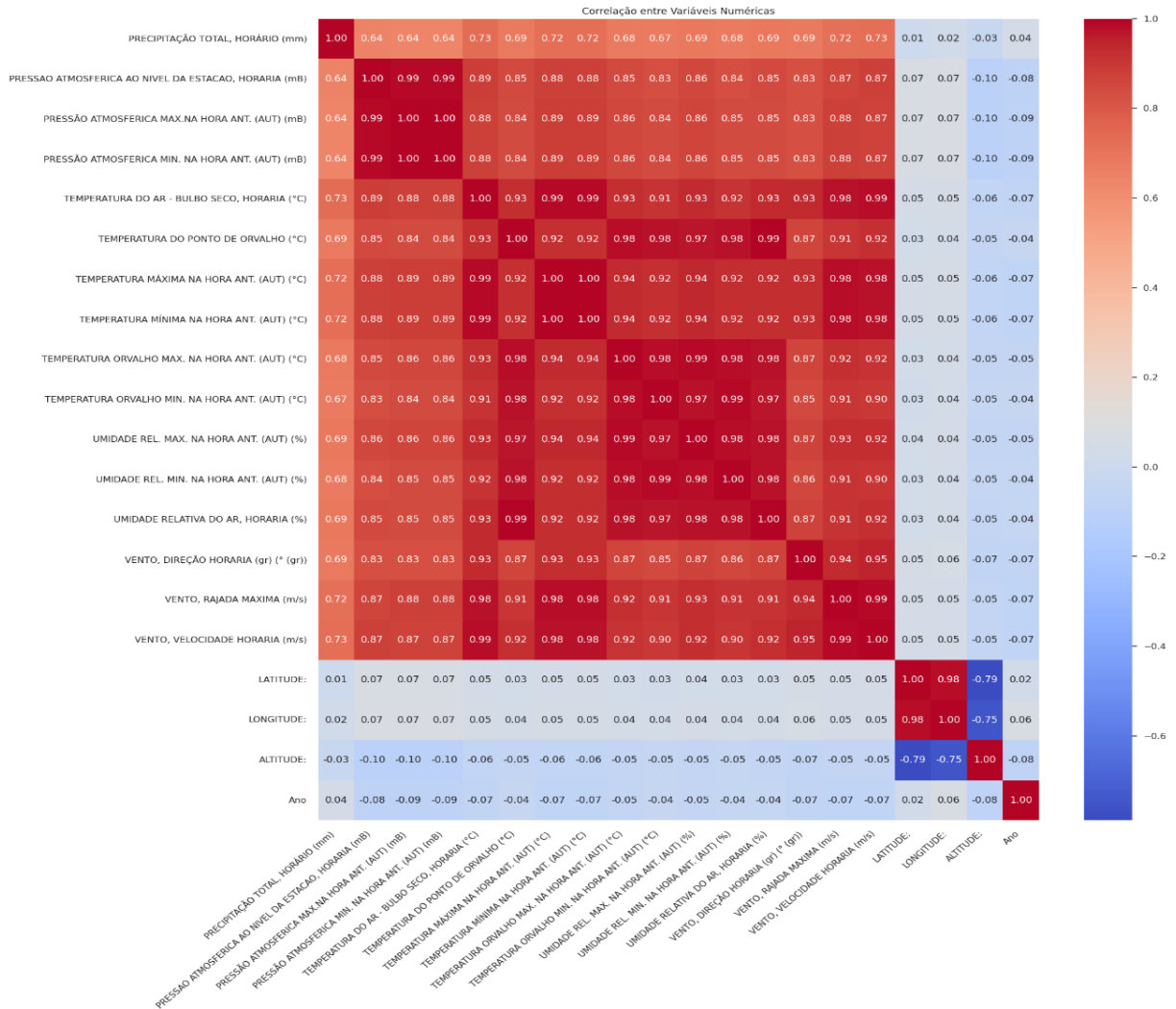


Fonte: Elaborado pelos autores (2025).

A etapa seguinte consiste na análise de correlação entre variáveis. A matriz de correlação apresentada revela os seguintes padrões:

A figura 3 apresenta um Mapa de correlação entre variáveis meteorológicas, com base no coeficiente de Pearson, mostrando a intensidade e a direção das relações entre os dados coletados na região.

Figura 3 – Gráfico de correlação de variáveis



Fonte: Elaborado pelos autores (2025).

Correlações Fortes ($|r| > 0.8$):

a. Variáveis de Temperatura e Umidade:

As temperaturas (bulbo seco, ponto de orvalho, máximas/mínimas) e umidade relativa apresentam correlações muito altas entre si (0.92 a 1.00), indicando que estão fortemente relacionadas. Por exemplo, temperatura do ar e temperatura máxima anterior têm $r=0.99$

b. Pressão Atmosférica:

As variáveis de pressão (horária, máxima e mínima) também mostram correlação quase perfeita entre si ($r=0.99$), sugerindo redundância (possibilidade de remover algumas para simplificar o modelo).

c. Vento:

Velocidade horária e rajada máxima têm $r=0.99$, enquanto direção do vento correlaciona-se bem com outras variáveis meteorológicas ($r \approx 0.83 - 0.97$ $r \approx 0.83 - 0.97$).

Correlações Moderadas ($0.5 < |r| < 0.8$):

Precipitação:

Correlaciona-se moderadamente com temperatura ($r \approx 0.69 - 0.73$ $r \approx 0.69 - 0.73$), umidade ($r \approx 0.68 - 0.69$ $r \approx 0.68 - 0.69$) e vento ($r \approx 0.72 - 0.73$ $r \approx 0.72 - 0.73$), o que é esperado, já que chuvas frequentemente acompanham mudanças nessas variáveis.

Pressão e Temperatura:

Pressão atmosférica tem $r \approx 0.85 - 0.89$ $r \approx 0.85 - 0.89$ com temperaturas, indicando que pressão tende a variar com a temperatura.

Correlações Fracas ou Nulas ($|r| < 0.3$):

Latitude, Longitude e Altitude:

Quase não se correlacionam com outras variáveis (valores próximos de zero ou negativos fracos, como $r = -0.10$ $r = -0.10$ entre altitude e pressão). Isso sugere que a localização geográfica tem pouca influência direta nas variáveis meteorológicas analisadas.

Treinando o modelo

Por se tratar de um modelo de previsão baseado em dados meteorológicos obtidos com frequência diária, o tratamento dos dados e o balanceamento das classes ainda estão em andamento. Esse tipo de *dataset* costuma apresentar uma grande quantidade de registros, o que demanda um processo rigoroso e criterioso de pré-processamento para garantir a qualidade das análises subsequentes.

Um dos principais desafios enfrentados nesta fase é o desbalanceamento da variável alvo, o que exige técnicas específicas para evitar que o modelo apresente viés para as classes majoritárias, comprometendo sua capacidade preditiva, especialmente em relação a eventos extremos, que são naturalmente mais raros.

Além disso, como evidenciado na análise exploratória e ilustrado no gráfico de boxplot, muitos atributos apresentam outliers significativos, o que é esperado em dados meteorológicos, particularmente em variáveis como precipitação, onde eventos atípicos (chuvas muito intensas ou secas severas) ocorrem de forma esporádica. Entretanto, também foi observada uma quantidade relevante de registros que destoam da realidade esperada, o que pode indicar erros de entrada manual, falhas nos sensores de coleta ou problemas de transmissão dos dados.

Próximos passos

As próximas etapas do desenvolvimento envolvem as seguintes ações:

- a. Realizar uma validação mais profunda dos dados coletados, identificando e tratando inconsistências, dados ausentes e outliers persistentes. Essa etapa também incluirá a verificação cruzada com outras fontes meteorológicas confiáveis, a fim de assegurar a integridade e a representatividade das variáveis analisadas.
- b. Após o tratamento dos dados, será realizada a seleção das variáveis mais relevantes para o modelo, com base em técnicas estatísticas e de aprendizado de máquina. Também serão desenvolvidas novas variáveis derivadas (engenharia de atributos), com o objetivo de melhorar o desempenho preditivo e reduzir redundâncias no conjunto de dados.
- c. Com os dados devidamente tratados e balanceados, os modelos de previsão serão treinados utilizando diferentes algoritmos (como Random Forest, XGBoost e Redes Neurais). Em seguida, os modelos serão avaliados com métricas apropriadas (acurácia, recall, F1-score, AUC, etc.) para determinar o desempenho em dados reais e sua capacidade de generalização.

Discussão

Os desafios enfrentados na análise dos dados e na construção do modelo preditivo evidenciam a importância estratégica da tecnologia e da ciência na previsão de eventos climáticos extremos. Esta seção discute os principais aspectos que ressaltam a relevância da análise preditiva, da ciência de dados e do uso de grandes volumes de informação no enfrentamento dessas questões, com ênfase na sustentabilidade, resiliência e tomada de decisões fundamentadas.

A importância da análise preditiva na previsão de eventos climáticos extremos

A análise preditiva é considerada indispensável para a identificação antecipada de eventos climáticos extremos, como longas estiagens que afetam o Centro-Oeste do Brasil (Elsayed, Wally, Alkabbany, Ali, & Farag, 2023). Essa região, marcada por forte atividade agrícola, depende da regularidade climática. Fenômenos como secas prolongadas afetam diretamente a produtividade das lavouras e, conseqüentemente, a economia local (Ikegwu, et al., 2024). Antecipar tais ocorrências por meio de modelos preditivos possibilita a implementação de estratégias de mitigação e preparação, tornando a análise climática fundamental para a sustentabilidade econômica e ambiental.

O papel da ciência de dados e do machine learning na construção de modelos preditivos

O avanço das tecnologias de ciência de dados e machine learning é diretamente responsável pelo desenvolvimento de modelos preditivos eficientes (Davenport & Diffebaugh, 2025). Esses recursos viabilizam algoritmos de alta precisão, capazes de detectar padrões climáticos complexos e antecipar variações que métodos tradicionais frequentemente não captam. Algoritmos como Random Forest destacam-se por lidar com grandes volumes de dados e identificar relações não-lineares, essenciais para a compreensão de fenômenos climáticos que impactam múltiplos setores sociais (Jain, et al., 2020).

A relevância do Big Data e a análise de grandes séries temporais

A eficácia desses modelos preditivos está diretamente associada à disponibilidade de dados em larga escala (Apache Software Foundation, 2025). O uso de ferramentas de processamento distribuído, como Hadoop e Apache Spark (Apache Software Foundation, 2025), viabiliza a manipulação de séries temporais extensas e complexas, aumentando a profundidade temporal da análise. Essa abordagem permite a detecção de variações sutis e a geração de previsões mais robustas e abrangentes, fornecendo subsídios valiosos para o planejamento estratégico de longo prazo em contextos sensíveis às mudanças climáticas.

Integração de múltiplas fontes de dados e modelos adaptáveis às variações climáticas

A integração de diversas fontes de dados — registros históricos, imagens de satélite (European Space Agency – ESA, 2025), sensores ambientais e informações atmosféricas globais — enriquece a base de conhecimento dos modelos preditivos (NASA, 2025). Essa abordagem melhora a qualidade das previsões ao considerar múltiplos fatores que influenciam

o clima, incluindo fenômenos naturais e mudanças globais. Modelos adaptáveis, que ajustam suas previsões com base em novas informações e tendências sazonais, são fundamentais para lidar com as rápidas e inesperadas mudanças climáticas, aumentando a confiabilidade das análises (Yang, et al., 2023).

Aplicações práticas na formulação de políticas públicas e gestão de riscos

As previsões meteorológicas geradas por essas tecnologias fundamentam ações concretas de órgãos públicos (Biesbroek, Badloe, & Athanasiadis, 2020). Diante do aumento da frequência e intensidade de eventos extremos no Brasil, como inundações e ondas de calor, o uso eficiente desses dados possibilita a formulação de políticas públicas eficazes, orientadas à prevenção de desastres e à proteção de populações vulneráveis. Estados e municípios podem elaborar planos de contingência mais eficientes, como a evacuação preventiva de áreas de risco, instalação de abrigos temporários, readequação de sistemas de drenagem urbana e campanhas de conscientização sobre cuidados durante eventos extremos. Além disso, podem aplicar incentivos para práticas agrícolas resilientes, ajustar políticas de uso da água e desenvolver estratégias de resposta rápida em setores essenciais como saúde, transporte e energia, fortalecendo a resiliência das infraestruturas e promovendo a segurança ambiental e social.

Impactos positivos na agricultura e na segurança alimentar

O setor agrícola é amplamente beneficiado pela previsão climática baseada em dados (Nogueira de Sousa, Ximenes de Brito, & Nogueira de Sousa Ximenes, 2025). A antecipação de períodos de seca ou variações extremas de temperatura permite que produtores rurais adotem práticas mais sustentáveis, como manejo racional da irrigação, escolha de cultivares resistentes e planejamento adequado de plantios e colheitas. Essas ações reduzem perdas econômicas, fortalecem a segurança alimentar e promovem a sustentabilidade do campo.

Monitoramento em tempo real e respostas rápidas a eventos extremos

A integração de modelos preditivos com sistemas de monitoramento em tempo real, por meio de sensores ambientais e satélites meteorológicos (Instituto Nacional de Meteorologia – INMET, 2025), representa um avanço significativo na agilidade e precisão das respostas a eventos meteorológicos severos. Dados continuamente atualizados permitem ajustes dinâmicos nas previsões, possibilitando a emissão antecipada de alertas para setores estratégicos como

defesa civil, energia e transportes. Essas ações preventivas contribuem para minimizar os impactos causados por fenômenos extremos (Hall & Sample, 2025).

Conclusão: tecnologia e ciência a favor da sustentabilidade e da resiliência

A convergência entre inteligência artificial, Big Data e conhecimento científico contribui para uma nova abordagem (Amazon Web Services, 2025). A aplicação dessas tecnologias na análise preditiva permite a formulação de estratégias inteligentes que favorecem o planejamento sustentável em setores públicos e privados. O investimento em modelos preditivos e sistemas integrados de monitoramento prepara governos, empresas e produtores para enfrentar as incertezas climáticas, tornando comunidades mais resilientes e promovendo um futuro seguro e sustentável (Python Software Foundation, 2025).

Conclusão

Diante da intensificação dos eventos climáticos extremos, torna-se urgente transcender abordagens tradicionais de monitoramento e previsão, incorporando tecnologias emergentes capazes de lidar com a complexidade e a escala dos dados ambientais. A utilização de Machine Learning associada a ferramentas de Big Data não apenas representa um avanço técnico, mas também um novo paradigma na forma como se compreende e se responde às mudanças climáticas. Mais do que prever, essas tecnologias oferecem suporte à tomada de decisão baseada em evidências, contribuindo para a formulação de políticas públicas resilientes e sustentáveis. Contudo, o sucesso dessa abordagem depende da qualidade dos dados, da infraestrutura tecnológica disponível e da capacidade de integração entre ciência, gestão pública e sociedade.

Este estudo evidenciou a importância do uso de modelos preditivos, como Regressão Linear (Linear Regression) e Floresta aleatória (Random Forest), na análise de padrões climáticos (Davenport & Diffebaugh, 2025), reforçando a capacidade preditiva na geração de previsões robustas. Previsões climáticas precisas podem beneficiar setores estratégicos como agricultura, infraestrutura e gestão de riscos, subsidiando decisões fundamentadas e formulação de políticas públicas eficazes (Biesbroek, Badloe, & Athanasiadis, 2020).

Embora os modelos estejam em fase de validação, os resultados parciais indicam caminhos promissores para o desenvolvimento de soluções eficazes na mitigação e previsão de eventos climáticos extremos (Nogueira de Sousa, Ximenes de Brito, & Nogueira de Sousa Ximenes, 2025). A integração entre tecnologia e ciência é essencial para antecipar desafios,

minimizar consequências negativas e facilitar a adaptação às novas condições ambientais emergentes

Para trabalhos futuros, recomenda-se explorar o uso de redes neurais profundas e modelos baseados em aprendizado contínuo, especialmente com dados em tempo real, para aprimorar ainda mais a precisão e a capacidade de resposta dos sistemas preditivos (Yang, et al., 2023). Além disso, investir no fortalecimento da infraestrutura de coleta e processamento de dados, assim como na ampliação da colaboração interdisciplinar entre pesquisadores, gestores públicos e comunidades, será fundamental para garantir o impacto prático e sustentável dessas tecnologias.

Essas inovações tecnológicas, aliadas ao avanço da pesquisa científica, serão determinantes para proteger comunidades, ecossistemas e setores produtivos, contribuindo diretamente para a promoção de políticas adaptativas de longo prazo.

Referências

AMAZON WEB SERVICES. **AWS - Serviços de computação em nuvem**. 2025. Disponível em: <https://aws.amazon.com/pt/>. Acesso em: 10 abr. 2025.

APACHE SOFTWARE FOUNDATION. **Apache Hadoop**. 2025. Disponível em: <https://hadoop.apache.org/>. Acesso em: 10 abr. 2025.

APACHE SOFTWARE FOUNDATION. **Apache Spark**. 2025. Disponível em: <https://spark.apache.org/>. Acesso em: 10 abr. 2025.

BIESBROEK, R.; BADLOE, S.; ATHANASIADIS, I. N. Machine learning for research on climate change adaptation policy integration: an exploratory UK case study. **Springer Nature**, v. 20, n. 3, p. 13, 2020. DOI: 10.1007/s10113-020-01677-8.

DAVENPORT, F. V.; DIFFENBAUGH, N. S. Using Machine Learning to Analyze Physical Causes of Climate Change: A Case Study of U.S. Midwest Extreme Precipitation. **Advancing Earth and Space Sciences - AGU**, v. 52, n. 3, p. e2021GL093787, 2025. DOI: 10.1029/2021GL093787.

ELSAYED, A. et al. Leveraging machine learning to enhance climate models: a review. **Arxiv.org**, 15 nov. 2023. Disponível em: <https://arxiv.org/abs/2311.09413>. Acesso em: 15 nov. 2023.

EMBRAPA CERRADOS. **Portal Embrapa**. 2025. Disponível em: <https://www.embrapa.br/cerrados>. Acesso em: 10 abr. 2025.

EUROPEAN SPACE AGENCY – ESA. **Explore Copernicus satellite missions - Sentinel Online**. 2025. Disponível em: <https://sentinels.copernicus.eu/>. Acesso em: 10 abr. 2025.

HALL, R.; SAMPLE, I. AI-driven weather prediction breakthrough reported. **The Guardian**, 20 mar. 2025. Disponível em: <https://www.theguardian.com/technology/2025/mar/20/ai-aardvark-weather-prediction-forecasting-artificial-intelligence>. Acesso em: 20 mar. 2025.

IKEGWU, C. A. et al. Recently emerging trends in big data analytic methods for modeling and combating climate change effects. **Springer Open**, v. 7, n. 1, 2024. DOI: 10.1186/s42162-024-00307-5.

INSTITUTO NACIONAL DE METEOROLOGIA – INMET. **Portal INMET**. 2025. Disponível em: <https://portal.inmet.gov.br/>. Acesso em: 10 abr. 2025.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS – INPE. **Instituto Nacional de Pesquisas Espaciais (INPE)**. 2025. Disponível em: <https://www.gov.br/inpe/pt-br>. Acesso em: 12 abr. 2025.

JAIN, P. et al. A review of machine learning applications in wildfire science and management. **Arxiv**, v. 301, 19 ago. 2020. DOI: 10.1016/j.jenvman.2021.113949.

MATPLOTLIB. **Matplotlib: Visualization with Python**. 2025. Disponível em: <https://matplotlib.org/>. Acesso em: 25 abr. 2025.

NASA. **NASA Worldview**. 2025. Disponível em: <https://worldview.earthdata.nasa.gov/>. Acesso em: 10 abr. 2025.

NOGUEIRA DE SOUSA, R.; XIMENES DE BRITO, R.; NOGUEIRA DE SOUSA XIMENES, J. Estudo comparativo a partir da aplicação de técnicas de aprendizagem profunda baseadas em dados pluviométricos coletados por estação meteorológica automática. **Revista de Sistemas e Computação - RSC**, v. 15, n. 2, 2025. DOI: 10.12345/rsc.2025.7593.

NUMPY. **NumPy reference**. 2025. Disponível em: <https://numpy.org/doc/stable/reference/>. Acesso em: 15 abr. 2025.

OYARZABAL, R. S. et al. Pesquisadores(as) do Cemaden publicam estudo de revisão científica sobre previsão de secas e o uso da inteligência artificial. **Centro Nacional de Monitoramento e Alertas de Desastres Naturais - Cemaden/MCTI**, 2023. Disponível em: <https://www.gov.br/cemaden/pt-br/assuntos/noticias-cemaden/pesquisadores-as-do-cemaden-publicam-estudo-de-revisao-cientifica-sobre-previsao-de-secas-e-o-uso-da-inteligencia-artificial>. Acesso em: 2025.

PANDAS. **pandas.pydata.org**. 2025. Disponível em: <https://pandas.pydata.org/>. Acesso em: 15 maio 2025.

PYTHON SOFTWARE FOUNDATION. **Python.org**. 2025. Disponível em: <https://www.python.org/downloads/>. Acesso em: 10 abr. 2025.

SILVA, S. T. et al. Predicting temperatures in Brazilian states capitals via Machine Learning. **arXiv preprint arXiv:2505.11511**, 2 maio 2025. Disponível em: <https://arxiv.org/abs/2505.11511>. Acesso em: 2 maio 2025.

SIRMACEK, B.; VINUESA, R. Remote sensing and AI for building climate adaptation applications. **Science Direct**, 2024. Disponível em: https://library.wmo.int/doc_num.php?explnum_id=11344. Acesso em: 2024.

UNAL, A. et al. Climate model-driven seasonal forecasting approach with deep learning. **cambridge.org**, v. 2, 2023. DOI: 10.1017/eds.2023.10.

VAUGHAN, A. et al. Aardvark Weather: End-to-End Data-Driven Weather Forecasting. **Cornell University**, 2024. Disponível em: <https://arxiv.org/abs/2404.00411>. Acesso em: 2024.

WEATHER API DOCUMENTATION. **Weather API**. [s.d.]. Disponível em: <https://www.weatherapi.com/>.

YANG, J. et al. Assessing climate risks from satellite imagery with machine learning: A case study of flood risks in Jakarta. **Science Direct**, v. 215, p. 12, 2023. DOI: 10.1016/j.envres.2023.114432.