

Predição de Temperatura em Castanhal: Análise de Tendências Climáticas Através de Aprendizado de Máquina.

¹Eric A. B. Mescouto, ¹Ferreira Junior, J. J. H.

¹Faculdade de Computação, Campus Universitário de Castanhal, Universidade Federal do Pará (UFPA).

Castanhal – PA – Brasil

{h.ericmescouto, josejailtonjunior}@gmail.com

Abstract. *The work presents a study on temperature prediction in Castanhal, in the state of Pará, using machine learning techniques applied to a temperature dataset. The historical database used in the study was obtained through the National Institute of Meteorology (INMET), covering the period from 2003 to 2024. Data processing and analysis were performed using Python and the pandas library, and the prediction models used were Random Forest Regressor and Linear Regression from the scikit-learn library. This study provides a contribution to the field of climate forecasting by proposing a computational methodology for predicting local temperatures.*

Resumo. *O trabalho apresenta um estudo sobre a predição de temperaturas em Castanhal, no estado do Pará, utilizando técnicas de aprendizado de máquina, aplicadas a um conjunto de dados de temperatura. A base de dados históricos, utilizada no estudo, foi obtida através do Instituto Nacional de Meteorologia (INMET), abrangendo o período de 2003 a 2024. O tratamento e análise dos dados foram feitos com Python e a biblioteca pandas, e os modelos de predição usados foram Random Forest Regressor e Regressão Linear, da biblioteca scikit-learn. Este estudo oferece uma contribuição para a área de previsão climática, propondo uma metodologia computacional para a predição de temperaturas locais.*

1. Introdução

As mudanças climáticas têm se tornado um dos maiores desafios globais da atualidade, sendo uma de suas principais consequências o aumento da frequência de eventos climáticos extremos, o que impacta desde ecossistemas até a qualidade de vida em todo o mundo. Em regiões como a Amazônia, as mudanças climáticas podem causar impactos profundos, incluindo o aumento da temperatura, a perda da biodiversidade, além de mudanças no regime de chuvas. A previsão de temperaturas é uma ferramenta essencial para compreender essas dinâmicas e auxiliar na formulação de estratégias de adaptação e mitigação dos impactos climáticos. Nesse contexto, as análises climáticas são fundamentais para oferecer subsídios científicos para promover um planejamento mais sustentável.

As previsões meteorológicas são tradicionalmente baseadas em modelos físicos complexos, que demandam elevado poder computacional. Contudo, o progresso das técnicas de aprendizado de máquina e a disponibilidade de dados históricos detalhados permitiram o desenvolvimento de modelos preditivos mais rápidos e adaptados.

Este estudo propõe a utilização de dois algoritmos de aprendizado de máquina, Regressão Linear e *Random Forest Regressor*, para prever temperaturas com base em dados históricos de 2003 a 2024. A escolha desses modelos se baseia em sua capacidade

de identificar tanto padrões lineares quanto não lineares nos dados, permitindo uma análise mais abrangente.

1.1. Objetivo Geral

Desenvolver um modelo capaz de avaliar e prever temperaturas baseado em técnicas de aprendizado de máquina, utilizando uma coleção histórica de dados meteorológicos do período de 2003 a 2024. O objetivo principal é criar um sistema de previsão de temperatura local que possa auxiliar no planejamento urbano, na saúde pública e nos estudos climáticos. Além de gerar previsões, o estudo busca compreender as tendências a longo prazo nas mudanças de temperatura, contribuindo para um melhor entendimento das dinâmicas climáticas e seus impactos.

Para alcançar esses objetivos, o modelo será fundamentado em um conjunto robusto de dados, garantindo a qualidade e a confiabilidade das informações utilizadas. Dessa forma, o projeto pretende oferecer uma ferramenta preditiva que não apenas auxilie na tomada de decisões estratégicas, mas também fortaleça estudos relacionados à sustentabilidade e à adaptação às mudanças climáticas.

1.2. Objetivos Específicos

- Realizar uma análise inicial dos dados históricos de temperatura do período de 2003 a 2024.
- Implementar modelos de aprendizado de máquina (Regressão Linear e *Random Forest Regressor*) para previsão de temperatura.
- Gerar gráficos a partir dos resultados obtidos para visualizar as previsões dos modelos além de representar os intervalos de confiança.
- Interpretar os resultados para identificar tendências futuras de temperatura.

2. Trabalhos Relacionados

Esta seção aborda trabalhos relacionados quanto a utilização de *Deep Learning* ou *Machine Learning* para previsão meteorológica.

Capuchinho (2021), em seu trabalho “Deep Learning para estimativa na escala mensal das temperaturas do ar para a região Centro-Oeste do Brasil”, utilizou Redes Neurais Artificiais (RNAs) para estimar temperaturas máximas, médias e mínimas do ar na escala mensal. O estudo analisou variáveis como altitude, latitude e umidade relativa do ar em diferentes cenários, obtendo coeficientes de determinação (R^2) superiores a 70% em diversos meses.

Vieira (2021), em seu estudo “Modelo de estimativa para a previsão meteorológica para fins agrícolas utilizando Machine Learning”, desenvolveu um modelo de previsão decendial para elementos climáticos, como temperatura do ar e precipitação, utilizando Regressão Linear Múltipla (RLM). O modelo apresentou alta precisão, com coeficiente de determinação ajustado (R^2_{ajust}) superior a 0,96, destacando-se na previsão de precipitação para o manejo agrícola.

Embora os trabalhos mencionados explorem diferentes técnicas de aprendizado de máquina para estimativas climáticas, este estudo se diferencia pela aplicação de dois

modelos complementares (*Random Forest* e *Regressão Linear*) com o objetivo de prever temperaturas locais em Castanhal, Pará.

3. Fenômenos Climáticos

O impacto das mudanças climáticas no equilíbrio ambiental e nas atividades humanas tem se intensificado nas últimas décadas, tornando-se um tema central nas discussões globais. Fenômenos como o aumento da temperatura média global, derretimento das calotas polares, elevação do nível do mar e intensificação de eventos extremos, como secas e enchentes, são amplamente atribuídos ao aumento da concentração de gases de efeito estufa na atmosfera, resultado de atividades humanas, como a queima de combustíveis fósseis e o desmatamento. De acordo com o Relatório Síntese do Sexto Relatório de Avaliação do IPCC (2023), a temperatura média global já aumentou cerca de 1,1°C no período de 2011-2020 em comparação com os níveis pré-industriais (1850-1900). Cada incremento adicional no aquecimento global intensifica os riscos climáticos, afetando ecossistemas, biodiversidade e comunidades humanas de forma cada vez mais severa.

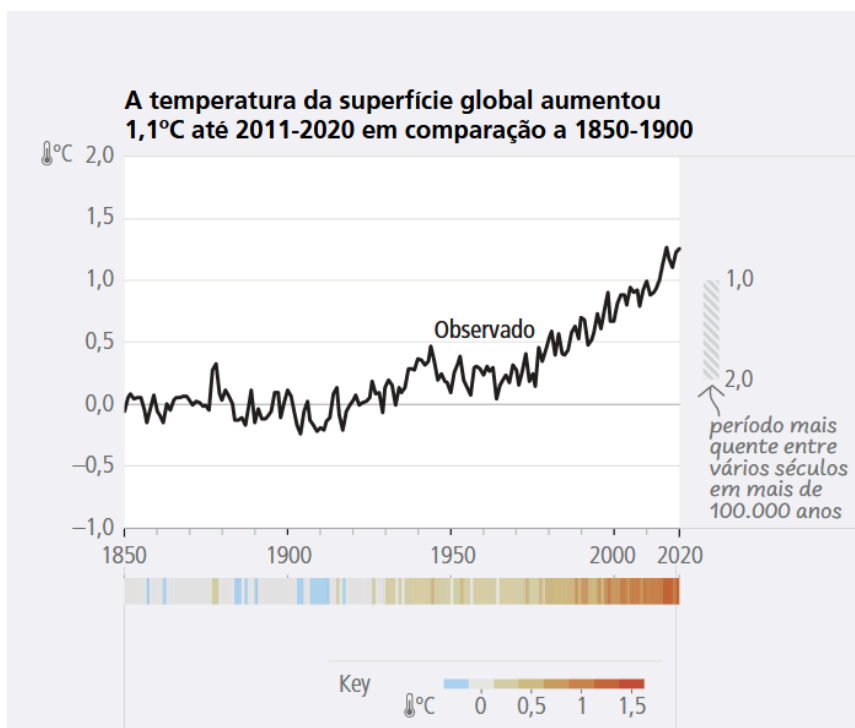


Figura 1. Evolução da temperatura da superfície global.

Fonte: Relatório Síntese do Sexto Relatório de Avaliação do IPCC (2023)

No Brasil, os efeitos das mudanças climáticas são evidentes em diferentes regiões. A Amazônia enfrenta taxas alarmantes de desmatamento, comprometendo seu papel como reguladora do clima global e como um dos maiores sumidouros de carbono do planeta. O semiárido nordestino sofre com secas prolongadas, impactando a agricultura de subsistência, a segurança hídrica e a qualidade de vida de milhões de

peças. Em outras áreas, como o Sul e Sudeste, eventos extremos, como enchentes, deslizamentos de terra e tempestades severas, têm causado prejuízos econômicos significativos, além de perdas humanas e danos à infraestrutura.

Conferências internacionais, como a Conferência das Partes (COP), organizada pela Convenção-Quadro das Nações Unidas sobre Mudança do Clima (UNFCCC), têm sido fundamentais na definição de metas para reduzir as emissões de gases de efeito estufa e fortalecer a cooperação global. A realização da COP30 em 2025, na cidade de Belém-PA, destaca a relevância do Brasil no cenário climático, especialmente devido à sua posição estratégica na preservação das florestas tropicais. Além disso, o evento ressalta a necessidade de ações concretas para proteger os biomas brasileiros, que desempenham um papel crucial na regulação climática global e no enfrentamento das mudanças climáticas.

A análise de dados climáticos é essencial para compreender as tendências do aquecimento global e prever cenários futuros. Informações recentes do IPCC apontam que, mesmo com os compromissos atuais, é provável que o aquecimento global ultrapasse 1,5°C nas próximas décadas, caso ações mais robustas e imediatas não sejam implementadas. Essas análises podem orientar políticas públicas e estratégias voltadas à mitigação dos impactos climáticos, como a transição para fontes de energia renovável, o reflorestamento e a implementação de práticas agrícolas sustentáveis. Promover um desenvolvimento sustentável e resiliente ao clima é fundamental para garantir melhores condições para as próximas gerações, especialmente em países como o Brasil, onde os desafios climáticos estão diretamente ligados à conservação ambiental e à justiça social.

4. Metodologia

A metodologia adotada neste trabalho é composta por duas etapas principais: Coleta de Dados e o desenvolvimento de Modelos Matemáticos utilizados no estudo. Essas etapas foram estruturadas de forma a garantir a precisão e a confiabilidade dos resultados, além de possibilitar uma análise robusta e fundamentada no contexto do estudo.

Na etapa de Coleta de Dados, foram definidos os critérios metodológicos e a seleção das fontes de informações relevantes para o desenvolvimento da pesquisa. Os dados meteorológicos utilizados no estudo foram obtidos do Instituto Nacional de Meteorologia (INMET), que mantém registros climáticos desde o ano de 2000. Contudo, nos anos iniciais, o número de cidades com registros disponíveis é limitado. Para a cidade de Castanhal, os primeiros registros datam de 24 de janeiro de 2003. A seleção das fontes foi realizada com base em critérios de confiabilidade e representatividade, assegurando que os dados coletados fossem consistentes e adequados aos objetivos da pesquisa. Além disso, foram aplicadas técnicas de pré-processamento para tratar inconsistências e preparar os dados para as etapas subsequentes de análise.

O tratamento dos dados e o desenvolvimento dos modelos de previsão foram realizados utilizando a linguagem de programação Python, amplamente reconhecida por sua versatilidade e eficiência em análises computacionais e científicas. Para realizar o tratamento de dados e a verificação das médias de temperaturas dos anos anteriores, foi empregada a biblioteca pandas, devido à sua robustez no gerenciamento e manipulação

de dados. Após a obtenção das médias anuais, foram aplicados algoritmos de Regressão Linear e *Random Forest Regressor*, disponíveis na biblioteca *scikit-learn*, com o objetivo de realizar a predição das temperaturas. A escolha desses modelos foi fundamentada em sua capacidade de fornecer resultados precisos e interpretáveis, sendo sua validação realizada com base nos dados históricos coletados.

4.1. Coleta de Dados

A coleta de dados para este estudo teve como objetivo reunir informações meteorológicas confiáveis e representativas para a análise e modelagem proposta. Os dados foram obtidos do Instituto Nacional de Meteorologia (INMET), uma fonte reconhecida pela qualidade e abrangência de seus registros climáticos, que disponibiliza informações desde o ano 2000. Apesar disso, algumas cidades, como Castanhal, possuem registros incompletos nos anos iniciais, com dados disponíveis apenas a partir de 24 de janeiro de 2003.

As variáveis coletadas incluíram diversas informações meteorológicas, como a Precipitação Total, a Pressão Atmosférica e a Radiação Global. Contudo, o foco deste estudo restringiu-se às temperaturas máximas, utilizadas para calcular as médias anuais das temperaturas máximas com base nos valores diários. Além disso, foi realizada uma filtragem de horário, considerando apenas as temperaturas registradas entre 08:00 e 18:00, de modo a garantir maior consistência e relevância para a predição. A escolha das temperaturas máximas justifica-se pela sua relevância direta ao objetivo do trabalho, que é prever as temperaturas máximas e suas médias nos próximos anos.

Antes da análise, foi realizado um processo de tratamento dos dados, incluindo a identificação e remoção de valores ausentes ou inconsistentes, além da padronização do formato das informações. Essa etapa foi crucial para garantir a qualidade dos dados e evitar distorções nos resultados, assegurando que os modelos preditivos fossem baseados em informações confiáveis.

A escolha dos dados do INMET foi motivada pela confiabilidade da fonte e pela abrangência temporal de seus registros, permitindo a construção de uma base sólida para o estudo. Esses dados sustentam tanto o tratamento estatístico quanto a modelagem preditiva, contribuindo diretamente para os resultados apresentados.

4.2. Modelos Matemáticos Utilizados

Os modelos matemáticos utilizados neste estudo foram desenvolvidos com o objetivo de realizar a predição das médias de temperaturas máximas para os anos subsequentes, com base nos dados históricos coletados. Para isso, foi empregada a linguagem de programação Python, reconhecida por sua versatilidade e ampla adoção em análises científicas e computacionais.

O processo de modelagem iniciou-se com o tratamento dos dados, realizado por meio da biblioteca *pandas*, que permitiu a manipulação eficiente dos dados meteorológicos e o cálculo das médias anuais de temperatura máxima. Após o pré-processamento, foram aplicados dois algoritmos de aprendizado de máquina: Regressão Linear e *Random Forest Regressor*, ambos disponíveis na biblioteca *scikit-learn*.

A escolha da Regressão Linear baseou-se na simplicidade e interpretabilidade do modelo, que permite identificar tendências lineares nos dados históricos. Já o algoritmo *Random Forest Regressor* foi selecionado por sua capacidade de lidar com relações não-lineares e sua robustez frente a dados complexos, sendo amplamente utilizado em problemas de predição devido à sua alta precisão. Ambos os modelos foram validados utilizando os dados históricos de temperaturas máximas, garantindo que as predições fossem baseadas em padrões consistentes observados ao longo do período analisado.

Dessa forma, a combinação de técnicas estatísticas e algoritmos de aprendizado de máquina possibilitou a construção de modelos confiáveis para a predição das médias de temperaturas máximas, atendendo aos objetivos propostos no estudo.

5. Resultados

Os resultados apresentados refletem a análise de dados climáticos históricos e projeções futuras, permitindo uma compreensão mais ampla das temperaturas ao longo do tempo. Com base em métodos estatísticos e modelos preditivos, busca-se identificar padrões, tendências e possíveis influências de fenômenos naturais e antrópicos sobre as temperaturas médias e máximas.

Além das análises gráficas realizadas, foi avaliada a confiabilidade dos modelos através da métrica Erro Absoluto Médio (MAE), que mede a diferença média entre os valores previstos e os reais. Para o modelo *Random Forest Regressor*, os resultados obtidos foram: Temperatura Média com MAE de 0,12 no conjunto de treino e 1,16 no conjunto de teste, e Temperatura Máxima com MAE de 0,24 no treino e 1,86 no teste. Esses valores indicam que o modelo apresenta boa precisão durante o treinamento, mas uma maior variação nos dados de teste, especialmente para as previsões de temperatura máxima.

5.1. Análise Descritiva

O gráfico da figura 2 apresenta as temperaturas médias anuais no período de 2003 a 2024. Os valores médios foram calculados anualmente e variam entre 27,70°C e 29,46°C, demonstrando uma certa variação ao longo dos anos. Pode-se observar um leve aumento nas temperaturas médias em alguns períodos, como em 2010 e 2015, possivelmente influenciado por fatores como o aquecimento global, que intensifica as temperaturas devido ao aumento de gases de efeito estufa, e fenômenos naturais que impactam diretamente o clima em escala global. Além disso, aspectos locais, como a urbanização e a redução da cobertura vegetal, podem ter contribuído para essas variações pontuais, ressaltando a interação entre causas naturais e atividades humanas na dinâmica climática.

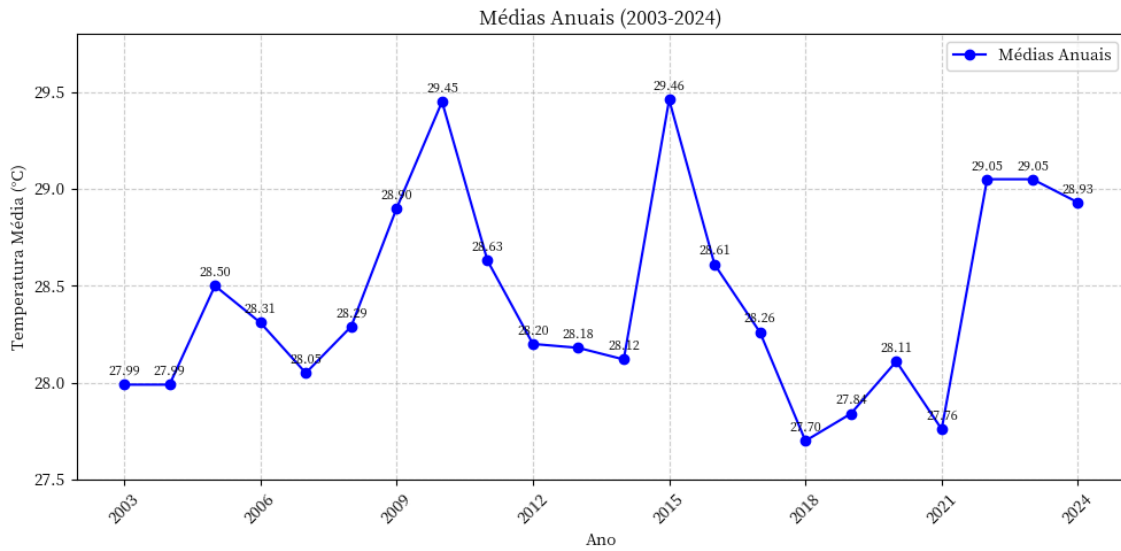


Figura 2. Gráfico das temperaturas médias anuais no período de 2003 a 2024.

O gráfico da figura 3 apresenta as temperaturas máximas anuais entre 2003 e 2024, destacando um pico significativo em 2015, quando foi registrada uma das maiores temperaturas do período. Esse aumento pode ser atribuído ao fenômeno climático global conhecido como "*super El Niño*", que, segundo o Instituto Nacional de Meteorologia (INMET), provocou um aquecimento anormal das águas do Oceano Pacífico. Esse fenômeno impactou diretamente o clima de diversas regiões, incluindo o Brasil, reduzindo a ocorrência de chuvas e intensificando o calor.

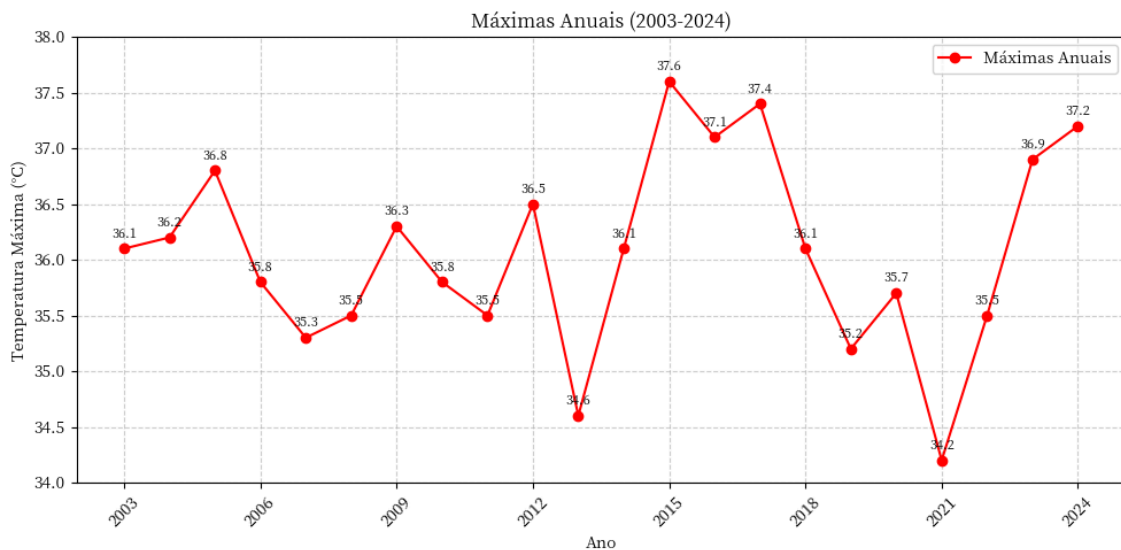


Figura 3. Gráfico das temperaturas máximas anuais no período de 2003 a 2024.

5.2. Predições do Modelo de Regressão Linear

O gráfico da figura 4 apresenta a predição das temperaturas médias, obtidas pelo modelo de Regressão Linear, no período de 2025 a 2040, com os valores exibidos em barras e um intervalo de confiança de $\pm 5\%$ representado pelas linhas de erro. Observa-se um leve aumento gradual das temperaturas ao longo dos anos, variando de 28,56°C em 2025

para 28,75°C em 2040. Essa tendência sugere um aquecimento contínuo no período analisado.

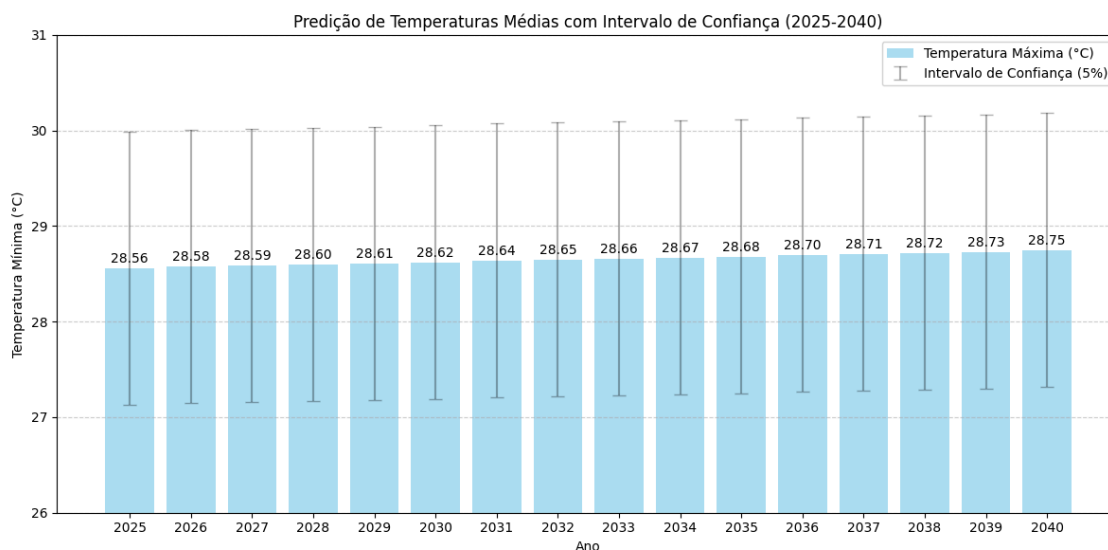


Figura 4. Gráfico da predição das temperaturas médias obtidas por meio do modelo de Regressão Linear.

O gráfico da figura 5 apresenta a predição das temperaturas máximas de 2025 a 2040. Observa-se uma tendência de leve aumento ao longo do período, com temperaturas variando de 36,16°C em 2025 para 36,28°C em 2040. Apesar do crescimento gradual, as variações são pequenas e sugerem uma relativa estabilidade no aumento das temperaturas máximas.

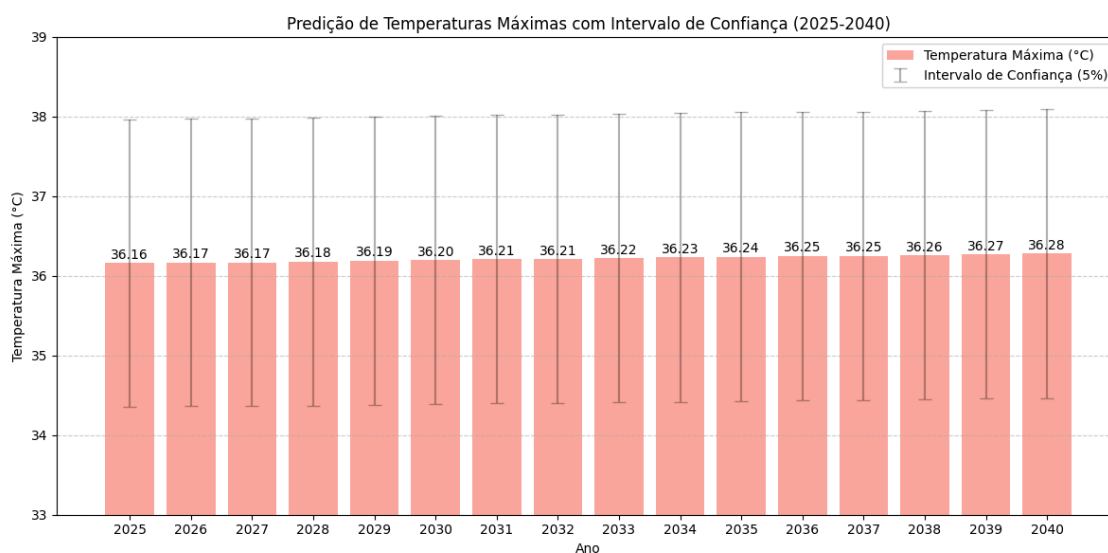


Figura 5. Gráfico da predição das temperaturas máximas obtidas por meio do modelo de Regressão Linear.

5.3. Predições do Modelo *Random Forest Regressor*

O gráfico da figura 6 apresenta a predição das temperaturas médias, obtidas pelo modelo *Random Forest Regressor*, no período de 2025 a 2040. Observa-se uma tendência de crescimento gradual ao longo do período, com a temperatura média passando de 28,38°C em 2025 para 28,50°C em 2040. Apesar de oscilações intermediárias, como o pico de 28,61°C em 2029, o comportamento geral indica um aumento sutil, mas consistente, nas temperaturas médias. Esse padrão sugere possíveis impactos de mudanças climáticas ou outros fatores de longo prazo que influenciam as condições térmicas da região analisada.

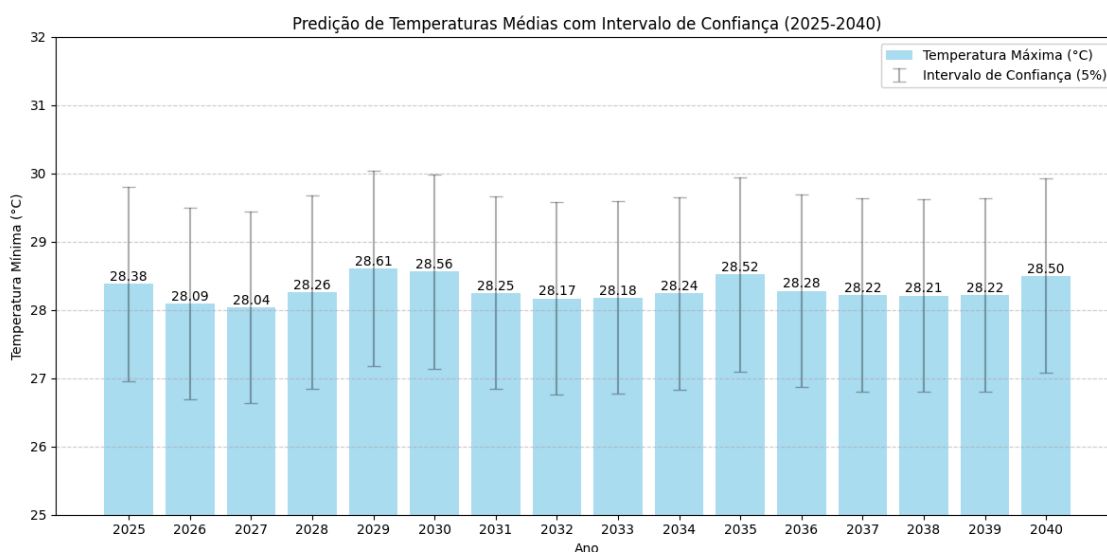


Figura 6. Gráfico da predição das temperaturas médias obtidas por meio do modelo *Random Forest Regressor*.

O gráfico da figura 7 apresenta a predição das temperaturas máximas de 2025 a 2040, obtidas pelo modelo *Random Forest Regressor*. Observa-se uma leve oscilação nos valores, com um pico inicial em 2025 (36,66°C) e uma estabilização em torno de 35,7°C nos anos seguintes. Ao comparar as predições obtidas com os dados históricos de 2003 a 2024, é identificado que a máxima prevista é superior aos menores registros dos anos anteriores, como 35,3°C em 2009 e 34,2°C em 2021. Isso indica uma possível tendência de elevação nas temperaturas mínimas ao longo do tempo, sugerindo um cenário de aquecimento gradual.

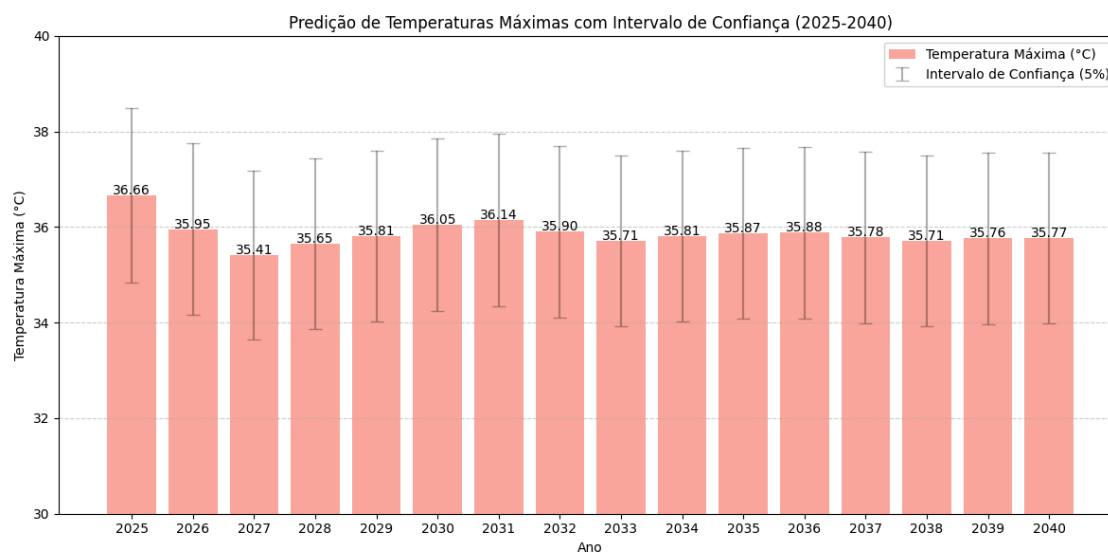


Figura 7. Gráfico da predição das temperaturas máximas obtidas por meio do modelo *Random Forest Regressor*.

6. Considerações Finais

Este trabalho apresentou um estudo sobre a predição de temperaturas em Castanhal, no estado do Pará, utilizando técnicas de aprendizado de máquina, com base em dados históricos de 2003 a 2024 obtidos do Instituto Nacional de Meteorologia (INMET). Foram implementados dois modelos preditivos, *Random Forest Regressor* e Regressão Linear, que demonstraram boa capacidade de identificar padrões climáticos nos dados analisados. A avaliação dos resultados através da métrica MAE revelou que o modelo *Random Forest Regressor* apresentou maior precisão durante o treinamento, embora tenha exibido maior variação nos dados de teste, especialmente para a previsão de temperatura máxima.

Os resultados obtidos destacam a utilidade das técnicas de aprendizado de máquina para a análise e previsão de tendências climáticas locais, contribuindo para o planejamento urbano, saúde pública e estudos climáticos em regiões específicas, como Castanhal. Além disso, a abordagem proposta neste trabalho reforça a importância de utilizar modelos complementares para capturar tanto padrões lineares quanto não lineares nos dados, promovendo uma análise mais robusta e confiável.

Entre as limitações do estudo, destaca-se a necessidade de incorporar variáveis climáticas adicionais, como umidade relativa, precipitação e velocidade do vento, que poderiam enriquecer as previsões e ampliar a compreensão das dinâmicas climáticas locais. Para trabalhos futuros, sugere-se a aplicação de outros algoritmos de aprendizado de máquina, como redes neurais profundas e modelos de ensemble, além de uma análise comparativa entre diferentes técnicas para identificar as mais adequadas ao contexto climático da região. Também é recomendada a inclusão de dados de outras cidades amazônicas, visando expandir o escopo geográfico e oferecer subsídios para políticas públicas voltadas à adaptação às mudanças climáticas.

Por fim, este estudo contribui para o avanço das pesquisas em previsão climática na Amazônia, oferecendo uma metodologia prática e replicável para outras localidades. Com o crescente impacto das mudanças climáticas, trabalhos como este são essenciais para fortalecer estratégias de mitigação e adaptação, promovendo um desenvolvimento sustentável e resiliente ao clima.

7. Referências

- PYTHON. Python. Disponível em: <<https://www.python.org/>>. Acesso em: 7 fev. 2025.
- NUMPY. NumPy. Disponível em: <<https://numpy.org/>>. Acesso em: 7 fev. 2025.
- MATPLOTLIB. Matplotlib: Python Plotting — Matplotlib 3.1.1 Documentation. Disponível em: <<https://matplotlib.org/>>. Acesso em: 7 fev. 2025.
- PANDAS. Python Data Analysis Library. Disponível em: <<https://pandas.pydata.org/>>. Acesso em: 7 fev. 2025.
- SCIKIT-LEARN. scikit-learn: Machine Learning in Python. Disponível em: <<https://scikit-learn.org/stable/>>. Acesso em: 7 fev. 2025.
- INMET. Instituto Nacional de Meteorologia - INMET. Disponível em: <<https://portal.inmet.gov.br/>> Acesso em: 7 fev. 2025.
- CAPUCHINHO, F. F. Deep Learning para estimativa na escala mensal das temperaturas do ar para a região Centro-oeste do Brasil. 2021. 90 f. Dissertação (Mestrado em Engenharia Agrícola) - Câmpus Central - Sede: Anápolis - CET, Universidade Estadual de Goiás, Anápolis-GO. Disponível em: <<http://www.bdt.ueg.br/handle/tede/967>>. Acesso em: 21 set. 2025.
- VIEIRA, I. C. O. Modelo de estimativa para a previsão meteorológica para fins agrícolas utilizando machine learning. 2022 - Universidade Estadual Paulista (UNESP), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal-SP. Disponível em: <<http://hdl.handle.net/11449/237490>>. Acesso em: 21 set. 2025.
- GLOBO. Belém atinge maior temperatura dos últimos 100 anos com “super El Niño”. Disponível em: <<https://g1.globo.com/pa/para/noticia/2015/12/belem-atinge-maior-temperatura-dos-ultimos-100-anos-com-super-el-nino.html>>. Acesso em: 7 fev. 2025.
- MCTI. Situação Atual e Tendências PAINEL INTERGOVERNAMENTAL SOBRE MUDANÇA DO CLIMA MUDANÇA DO CLIMA 2023 Relatório Síntese Um Relatório do Painel Intergovernamental sobre Mudança do Clima. [s.l: s.n.]. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/sirene/publicacoes/relatorios-do-ipcc/arquivos/pdf/copy_of_IPCC_Longer_Report_2023_Portugues.pdf>. Acesso em: 7 fev. 2025.