



Artigos

Estrabão
Vol. (5): 81 - 91
© Autores
DOI: 10.53455/re.v5i1.207



Recebido em: 18/11/2023
Publicado em: 24/01/2024

Modelo Preditivo para a proliferação do *Aedes aegypti* em Itajaí (Santa Catarina): Uma abordagem integrando fatores climáticos locais e globais

Predictive Model for the proliferation of *Aedes aegypti* in Itajaí (Santa Catarina): An approach integrating local and global climatic factors

Eduardo Augusto Werneck Ribeiro ^{1A}, Cleusa Matiola, Mario Francisco Leal de Quadro, Matheus Ferreira de Souza, Isabel Cristina Bohn, João Augusto Brancher Fuck, Raul Borges Guimarães e Thiago Pereira Alves

Resumo:

Contexto: Itajaí, uma cidade costeira em Santa Catarina, enfrenta desafios singulares relacionados à proliferação do *Aedes aegypti*, vetor de várias doenças. Este artigo apresenta um modelo preditivo desenvolvido para prever os focos de *Aedes aegypti* na região, considerando a interação complexa entre variáveis climáticas locais e fenômenos globais, como El Niño e La Niña. **Metodologia:** Utilizando um algoritmo Random Forest, o modelo é capaz de capturar relações não-lineares nos dados, oferecendo insights sobre a influência de fatores climáticos na atividade do mosquito. A escolha deste algoritmo se deve à sua robustez e capacidade de considerar a multiplicidade de fatores que influenciam a proliferação do mosquito. Ao agrupar dados por estação do ano, o modelo incorpora nuances sazonais, refletindo as variações climáticas de Itajaí. **Considerações:** A integração de padrões climáticos mais amplos destaca a interconexão entre fatores locais e globais. Este modelo oferece uma ferramenta valiosa para as autoridades de saúde de Itajaí, permitindo ações proativas e otimização de recursos no combate ao *Aedes aegypti*. Em suma, este estudo propõe uma abordagem inovadora e prática para a prevenção de doenças transmitidas por mosquitos, com potencial para impactar positivamente a saúde pública em regiões similares.

Palavras-Chave: Random Forest, Modelo Preditivo, Variabilidade Climática, *Aedes aegypti*

Abstract

Context: Itajai, a coastal city in Santa Catarina, faces unique challenges related to the proliferation of *Aedes aegypti*, a vector of several diseases. This article presents a predictive model developed to forecast *Aedes aegypti* hotspots in the region, considering the complex interaction between local climatic variables and global phenomena such as El Niño and La Niña. **Methodology:** Using a Random Forest algorithm, the model is capable of capturing non-linear relationships in the data, providing insights into the influence of climatic factors on mosquito activity. The choice of this algorithm is due to its robustness and ability to consider the multiplicity of factors influencing mosquito proliferation. By grouping data by season, the model incorporates seasonal nuances, reflecting climatic variations in Itajai. **Considerations:** The integration of broader climatic patterns highlights the interconnection between local and global factors. This model offers a valuable tool for Itajai's health authorities, enabling proactive actions and resource optimization in the fight against *Aedes aegypti*. In short, this study proposes an innovative and practical approach to the prevention of mosquito-borne diseases, with potential to positively impact public health in similar regions.

Keywords: Random Forest, Predictive Model, Climatic Variability, *Aedes aegypti*

1 - professor do Instituto Federal Catarinense

A - Contato principal: eduardo.ribeiro@ifc.edu.br

Introdução

O fenômeno da mudança climática tem um alcance global e pode causar grandes impactos nos modos pelos quais as doenças transmitidas por vetores se propagam, dentre elas a dengue (Alexander et al, 2022). Utilizar modelos preditivos é fundamental para combater efetivamente a propagação do vetor diante da mudança climática. No grupo dos vários vetores, o *Aedes aegypti* se sobressai. Nesses modelos são levadas em conta as variáveis climáticas tais como a temperatura, umidade e precipitação e o possível impacto delas nas reproduções dos vetores.

A ocorrência da variabilidade climática na região sul do Brasil é marcada por uma grande complexidade decorrente das características específicas do terreno, assim como dos impactos causados pela continentalidade nos diferentes tipos de clima presentes nessas áreas. Isso cria mais obstáculos para entender as diversas alterações climáticas nessa área, que já se destaca pelos extremos de temperatura e pela grande variabilidade nas precipitações ao longo do ano. Contudo, os diferentes modos climáticos na região sul precisam ser analisados com maior profundidade em futuras pesquisas (Guerra & Quadros, 2021).

Diante desse contexto, busca-se por tecnologias que possibilitem a compreensão não só das mudanças climáticas como também de outras variáveis sociais e econômicas; isso permite realizar uma análise mais completa e sofisticada. Técnicas de aprendizado de máquina oferecem muitas opções interessantes em termos dos algoritmos a serem utilizados. Atualmente, há um crescente interesse no uso dos algoritmos para o aprendizado visando o reconhecimento e compreensão dos diferentes padrões encontrados nas imagens. A combinação eficiente entre as técnicas avançadas do geoprocessamento e os algoritmos com o uso do aprendizado da máquina é que possibilitou o êxito alcançado por Gonçalves (2021) na melhor identificação das silhuetas dos vegetais. A acurácia da técnica foi evidenciada ao ser capaz de criar um modelo classificador altamente preciso diante dos grandes volumes de dados.

Foi criada uma motivação para implementar esses modelos em estudos de variação climática envolvendo o *Aedes aegypti*. A partir do projeto Índice de Positividade de Armadilhas (IPA) (Ribeiro et al., 2023) procuramos desenvolver uma metodologia que pudesse integrar sensores e dados de estação meteorológica, imagens de drones, data loggers, banco de dados históricos das atividades de armadilhas e usar algoritmos computacionais do tipo aprendizado de máquina, para entender a dinâmica populacional do mosquito diante das mudanças climáticas e promover estratégias eficazes de controle. O resultado desta interação foi a construção de um modelo preditivo.

Existem um conjunto de algoritmos que podem ser aplicados. Neste artigo abordaremos um modelo específico: o Random Forest. De acordo com Breiman (2001), o algoritmo Random Forest desempenha um papel fundamental nas técnicas de aprendizado de máquina. Essa abordagem utiliza um conjunto de árvores de decisão para melhorar a precisão da classificação e a capacidade de generalização dos modelos. Ao combinar várias árvores e seus votos, o algoritmo é capaz de superar as limitações individuais de cada árvore e fornecer resultados mais confiáveis e robustos.

Além disso, como um modelo preditivo, o Random Forest oferece uma base para o desenvolvimento e teste de diferentes métodos no campo do aprendizado de máquina. Através da randomização das árvores e da seleção aleatória de características, é possível explorar diversas abordagens para melhorar a precisão da classificação e a capacidade de generalização em diferentes conjuntos de dados. Essa diversidade de métodos no Random Forest permite a evolução contínua do aprendizado de máquina e a aplicação de sistemas que podem ajudar a sociedade a encontrar soluções para os problemas diários.

Ao explorar e testar diferentes métodos de Random Forest, entende-se que é possível melhorar a eficiência dos modelos preditivos e impulsionar o avanço das aplicações de machine learning em diversos setores, oferecendo soluções inovadoras para a sociedade, incluindo a área da saúde coletiva (Raman et al., 2021; Jia et al., 2017; Ochida et al., 2022).

Nesse contexto, a implementação do algoritmo Random Forest surge como uma ferramenta para desenvolver modelos preditivos desejado pelo projeto IPA (Ribeiro et al, 2023). Esse algoritmo permite não apenas a incorporação de variáveis climáticas essenciais, como a temperatura, mas também se adapta perfeitamente à natureza multifatorial das doenças transmitidas por vetores.

Ao combinar os dados meteorológicos com técnicas de aprendizado de máquina, espera-se desenvolver

um modelo que responda à questão central do estudo: até que ponto a combinação de dados meteorológicos e algoritmos de aprendizado de máquina, como o Random Forest, pode resultar em um modelo resolutivo para a reprodução do mosquito *Aedes aegypti*? Para isso, esta comunicação apresenta uma metodologia, detalhando como tais dados podem ser analisados e interpretados para informar estratégias de prevenção e controle da propagação do *Aedes aegypti* e de outros vetores que impactam a saúde pública (Ochida et al., 2022).

A combinação de dados meteorológicos coletados em diferentes escalas, como dados locais das armadilhas ovitrampas e dados globais da NOAA (2023) relacionados aos fenômenos El Niño e La Niña, oferece uma oportunidade de abordagem integrada para o monitoramento ambiental e epidemiológico. Essa abordagem também permite avançar em estudos mais detalhados sobre microclimas que influenciam diretamente o comportamento dos vetores de doenças e fornece um entendimento mais amplo dos padrões climáticos que afetam a distribuição e a prevalência desses vetores (Matiola, 2019, 2020).

Esta comunicação de pesquisa tem como objetivo apresentar a metodologia e os resultados parciais de uma proposta de modelagem que utilizou o algoritmo Random Forest, especificamente o método de “bagging”, a partir de dados climáticos no município de Itajaí - Santa Catarina. A utilização dessa abordagem combinada de dados meteorológicos e técnicas de aprendizado de máquina tem o potencial de melhorar a precisão da predição da eclosão de ovos do *Aedes aegypti* e fornecer informações para o desenvolvimento de estratégias eficazes de prevenção e controle do vetor da dengue e de outras doenças transmitidas.

Materiais e Método

Localização do estudo

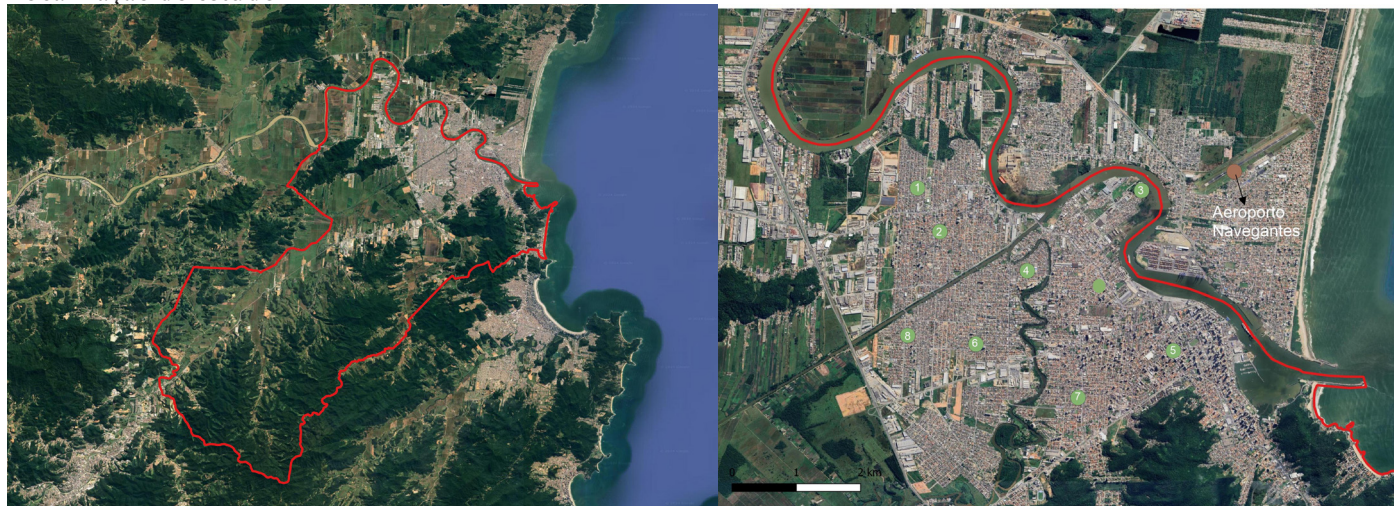


Figura 1 Município de Itajaí e os pontos de coleta

Coleta de Dados

Os dados utilizados neste estudo foram obtidos a partir de duas fontes principais. A primeira foi a Diretoria de Vigilância Epidemiológica de Santa Catarina (DIVE), que forneceu informações sobre a localização das armadilhas de mosquitos *Aedes aegypti* monitoradas em Itajaí. A segunda fonte foram os sensores termohidrográficos (KR420 Datalogger de temperatura e umidade) instalados nas proximidades das armadilhas para monitorar as condições ambientais locais. Foram escolhidos 9 pontos que fazem parte da malha de pontos de monitoramento, com base na maior atividade das armadilhas nos últimos 10 anos e que pudessem representar os diferentes padrões de uso do solo do município (residencial, comercial e industrial). Foram também utilizados os dados históricos da estação meteorológica do aeroporto de Navegantes, de 2014 a 2022 (Inmet), e os registros de focos do mosquito *Aedes aegypti* da Secretaria de Estado da Saúde de Itajaí (DIVE-SC) para treinar o modelo.

O desenvolvimento deste modelo foi realizado no Jupyter Notebook, uma aplicação web que facilita a execução interativa de código Python. O Jupyter Notebook é uma plataforma interativa para criar e

compartilhar documentos com códigos, equações, visualizações e texto. É muito usado por cientistas de dados e programadores Python para compartilhar documentos com códigos em tempo real e recursos multimídia. Pode ser executado em qualquer servidor com acesso SSH ou HTTP, seja em um computador pessoal ou na nuvem. Ele é compatível com a maioria das linguagens de programação, mas se destaca com Python. O Jupyter Notebook é muito utilizado em projetos de ciência de dados e machine learning, além de facilitar o uso eficiente da biblioteca scikit-learn (Jupyter, s.d.)

Pré-processamento de Dados e Etapas Preparatórias

Antes da análise, os dados foram processados usando a biblioteca pandas em Python. Foram realizadas as seguintes etapas de pré-processamento:

- Conversão de todos os números para o tipo de dado float64.
- Carregamento dos dados específicos para análise.
- Verificação para garantir que todos os dados eram numéricos.
- Substituição de valores NaN por zero.
- Preenchimento de valores ausentes com a média móvel da estação do ano correspondente.

Seleção de Características e Modelo Estatístico Adotado

Para criar o modelo preditivo, foi utilizada a técnica Random Forest “bagging”, uma abordagem de ensemble learning. A decisão por essa técnica se deu por sua capacidade de lidar com grandes conjuntos de dados, sua habilidade em lidar com variáveis não linearmente dependentes e sua robustez em relação a outliers.

Conforme a metodologia de Matiola (2020), foi organizado o banco de dados em diversas faixas e parâmetros. Os dados foram categorizados em faixas de temperatura específicas:

- Dias registrados com temperaturas menores ou iguais a 18 °C (C18)
- Temperaturas iguais a 19 e menores que 21°C (C19)
- Temperaturas iguais a 21 e menores que 23°C (C21)
- Temperaturas iguais a 23 e menores que 25°C (C23)
- Temperaturas iguais a 25 e menores que 30°C (C25)
- Temperaturas iguais ou maiores que 30°C (C30).

Além disso, outras variáveis foram consideradas na análise, incluindo:

- Temperatura média máxima no mês (TMAX)
- Média da precipitação (PREC)
- Dias registrados com precipitação menor do que 1 mm (Pm1)
- Dias registrados com precipitação igual a 1 e menor que 5 mm (P1)
- Dias registrados com precipitação igual a 5 e menor que 10 mm (P5)
- Dias registrados com precipitação igual ou maior que 10mm (P10)
- Focos registrados de *Aedes* (*Aedes*).

Essa organização detalhada do banco de dados permitiu a criação de condições adequadas para a construção do modelo. Para identificar as variáveis mais significativas para incluir no modelo preditivo, realizamos uma análise de correlação. Foi utilizada uma matriz de correlação, que é uma representação tabular dos coeficientes de correlação entre as variáveis do estudo. As correlações próximas de 1 ou -1 indicam uma forte relação linear, enquanto as correlações próximas de 0 indicam uma falta de relação. Selecionamos as variáveis com as maiores correlações para inclusão no modelo, assegurando que o mesmo pudesse captar os padrões mais significativos relacionados à presença do *Aedes aegypti*.

Treinamento e Validação do Modelo

O método Random Forest Bagging, inspirado em Breiman (2001), foi implementado utilizando a biblioteca Python. A implementação do algoritmo foi realizada com o auxílio da biblioteca de machine learning scikit-learn, que oferece uma variedade de ferramentas e algoritmos para a construção de modelos preditivos.

O scikit-learn possui uma classe chamada *RandomForestClassifier*, que permite a criação e treinamento de uma floresta aleatória utilizando o método de Bagging. Essa classe oferece diversos parâmetros ajustáveis, como o número de árvores na floresta e a profundidade máxima das árvores, que podem ser configurados de acordo com as necessidades do projeto.

Após a criação da floresta aleatória, o modelo é treinado utilizando os dados de treinamento disponíveis. Durante o treinamento, cada árvore da floresta é construída a partir de um subconjunto aleatório dos dados de treinamento, o que permite que o modelo aprenda padrões mais gerais e reduza a sensibilidade a ruídos e variações aleatórias nos dados.

Uma vez treinado, o modelo pode ser utilizado para fazer previsões em novos dados. Durante a fase de classificação, cada árvore da floresta contribui com um voto igual para a classe mais popular em um determinado dado de entrada. A classe final é determinada pelo voto majoritário das árvores.

O modelo Random Forest otimizado foi treinado usando um subconjunto dos dados (conjunto de treinamento) e validado usando um conjunto separado (conjunto de validação). Para o treinamento do modelo, foram utilizados dados históricos da estação meteorológica do aeroporto de Navegantes e registros de focos do mosquito *Aedes aegypti* da Secretaria de Estado da Saúde de Itajaí no período de 2014 a 2022. Assim, 80% dos dados foram utilizados para treinamento e 20% para teste.

Otimização de Hiperparâmetros

Para melhorar a performance do modelo, foi realizada uma otimização dos hiperparâmetros do Random Forest utilizando a técnica de pesquisa em grade (Grid Search). Foram testadas diversas combinações de parâmetros para encontrar a configuração que produziu os melhores resultados.

Seleção de Variáveis Baseada na Correlação

Antes de treinar o modelo, realizamos uma análise de correlação para identificar as faixas de temperatura que tinham a maior relação com a presença do mosquito *Aedes aegypti*. Por meio desta análise, selecionamos as faixas de temperatura que mostraram alta correlação com os focos do mosquito. Esta etapa ajudou a reduzir a dimensionalidade do modelo, tornando-o mais eficiente e evitando potenciais problemas de multicolinearidade.

Incorporação de Dados Climáticos Externos

Reconhecendo a influência significativa dos fenômenos El Niño e La Niña no clima e, conseqüentemente, na distribuição e atividade do *Aedes aegypti*, decidimos incorporar esses dados em nosso modelo. Para isso, recorreremos à NOAA (2023), para os dados atualizados sobre esses fenômenos. Estes dados foram integrados ao conjunto de dados principal para capturar a variação climática causada por El Niño e La Niña e entender melhor sua relação com a presença e atividade do mosquito.

Com o desenvolvimento deste modelo preditivo, passa-se a discutir os resultados sobre as atividades das armadilhas do mosquito *Aedes aegypti* no município de Itajaí (SC).

Avaliação Estatística do Modelo

Para avaliar a capacidade preditiva do nosso modelo, selecionamos o coeficiente de determinação (R^2) e a Raiz do Erro Quadrático Médio (RMSE) como nossas métricas principais. O R^2 é uma medida da quantidade de variação na variável dependente que é explicada pelas variáveis independentes no modelo. Um R^2 mais alto indica um melhor ajuste do modelo aos dados. O RMSE é uma medida da média dos erros quadráticos, proporcionando uma avaliação da diferença entre os valores previstos e os valores observados. Um RMSE

mais baixo indica previsões mais precisas. Essas métricas foram escolhidas por sua relevância em quantificar a performance do modelo em termos de ajuste e precisão.

Resultados

Na matriz gerada, observamos as seguintes correlações significativas:

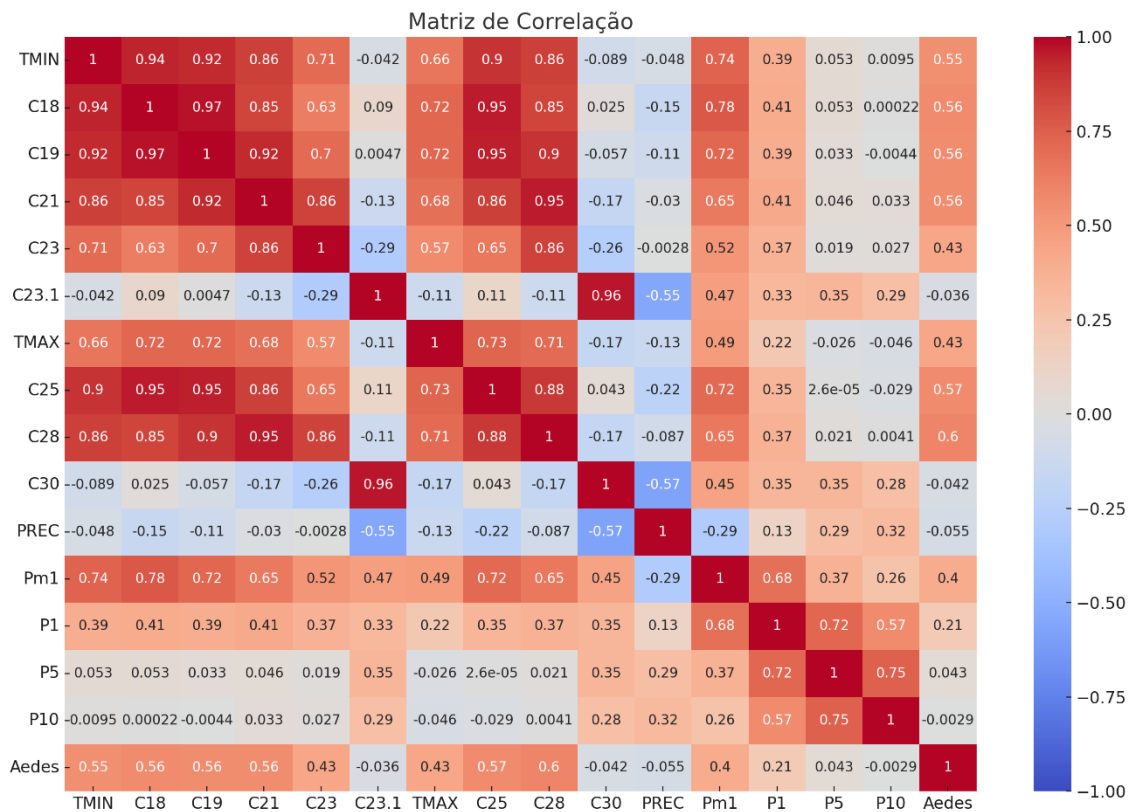


Figura 2 Matriz de Correlação.

As faixas de temperatura, representadas pelas variáveis TMIN, C18, C19, C21, C23, e C25, mostraram correlações fortes entre si, indicando uma relação linear positiva substancial.

A variável *Aedes*, que representa os focos do mosquito, apresentou correlações variadas com as faixas de temperatura e precipitação, sugerindo diferentes graus de influência nas atividades do mosquito. Notavelmente, as temperaturas extremas, como representadas pelas variáveis C30 e PREC, mostraram uma correlação negativa com outras variáveis, indicando uma tendência oposta.

A variável P1 (dias registrados com precipitação igual a 1 e menor do que 5 mm) é considerada a mais importante, seguida de C25 (dias registrados com temperaturas entre 25 e 30°C) e TMIN (temperatura média mínima registrada no mês).

As variáveis C19 (dias registrados com temperaturas entre 19 e 21°C) e P5 (dias registrados com precipitação entre 5 e 10 mm) são consideradas menos importantes.

Esses resultados nos dão uma ideia das variáveis que o modelo considera mais relevantes. No entanto, isso não significa necessariamente que as variáveis com menor importância são inúteis. Em contextos complexos, como o ciclo de vida do *Aedes aegypti*, muitos fatores podem interagir de maneira não linear e complexa. Com base nisso, as variáveis C18, C19, TMIN, C25 e C21 parecem ser as mais fortemente correlacionadas positivamente com os focos de *Aedes aegypti*.

Para as figuras 2 e 3, as barras azuis representam os valores reais dos focos de *Aedes* e as barras vermelhas representam o erro absoluto entre as previsões.

A altura total de cada barra combinada (azul + vermelho) representa a previsão do modelo. Assim, pode-se comparar diretamente os valores reais (apenas a parte azul) com as previsões (a barra completa).

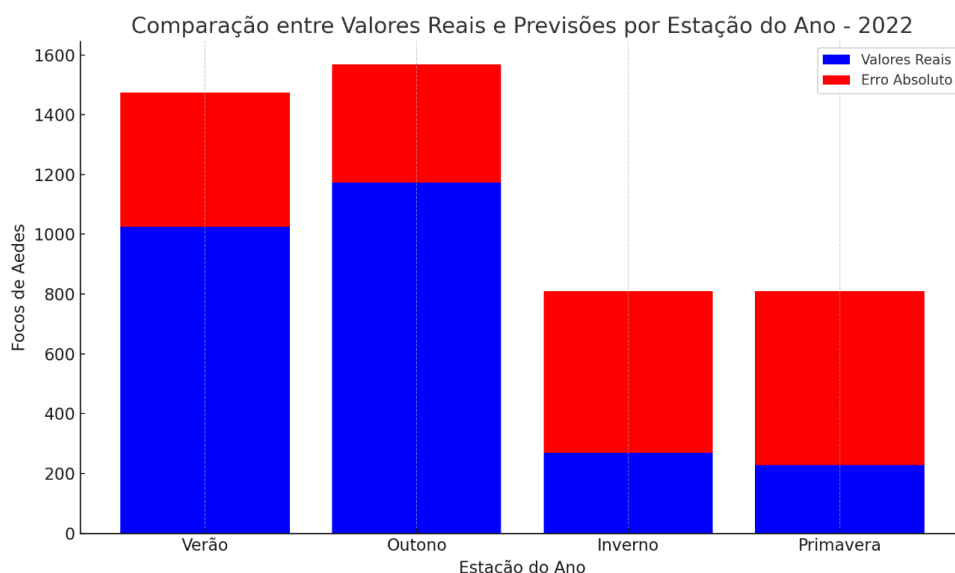


Figura 3 Avaliação do modelo

A Figura 3 ilustra a contagem total de focos do *Aedes aegypti* observados por estação, e a segmentação pela barra colorida indica a magnitude do erro absoluto da previsão do modelo.

Neste sentido, o desempenho do Modelo por Estação do Ano pode ser observado pelas estações do ano.

Verão: Durante o verão, observamos que o modelo teve um desempenho robusto, com os erros de previsão ocupando uma fração mínima das barras totais. Isso sugere que o modelo é bem calibrado para as condições típicas da estação, que historicamente mostram uma alta atividade do mosquito.

Outono e Primavera: No outono e na primavera, o modelo apresentou uma variação maior nos erros. Ainda que as previsões tenham sido em geral precisas, identificamos um ligeiro aumento nos erros absolutos, o que pode refletir na transição das condições climáticas e nos padrões de comportamento do mosquito.

Inverno: A precisão do modelo diminuiu durante o inverno, uma época do ano em que a atividade do mosquito é tradicionalmente mais baixa. Os erros absolutos nessas barras são relativamente maiores, destacando uma área para aprimoramento do modelo, possivelmente através da incorporação de variáveis específicas do inverno ou ajustes nos hiperparâmetros do modelo.

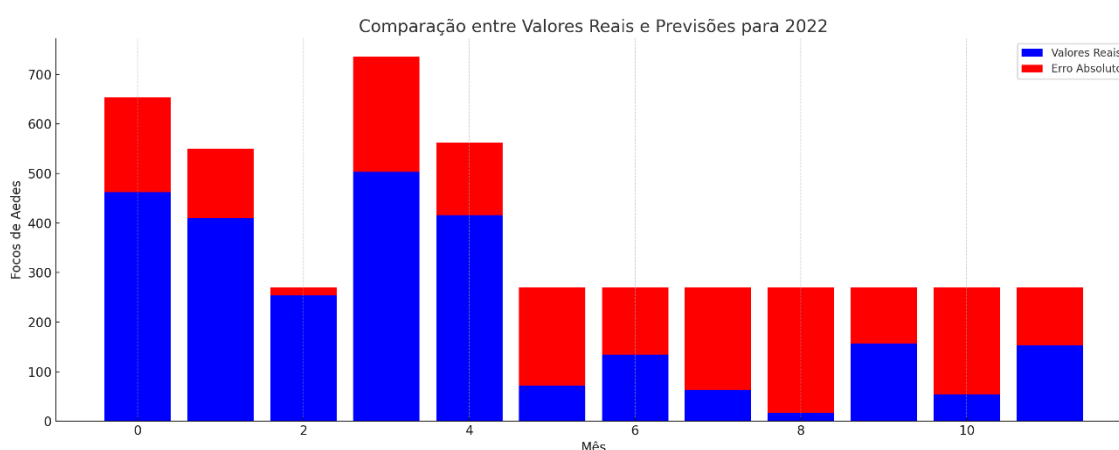


Figura 4 Avaliação do desempenho do modelo

O desempenho do modelo preditivo em uma base mensal através da Figura 4, ilustra as previsões de focos do *Aedes aegypti* comparadas com os valores observados. As barras azuis representam a quantidade real de focos registrados por mês, enquanto as barras vermelhas indicam o erro absoluto das previsões do modelo.

O modelo demonstrou uma tendência consistente em suas previsões ao longo dos meses, com erros que variam proporcionalmente aos valores reais, indicando uma boa calibração do modelo em relação às flutuações temporais.

Precisão das Previsões: Em meses onde o erro absoluto é menor, como no mês 6 (julho), o modelo exibiu uma precisão notável na previsão da quantidade de focos de *Aedes aegypti*. Este desempenho robusto em determinados meses reforça a utilidade do modelo para ações de controle e prevenção.

Áreas de Melhoria: Nos meses onde o erro absoluto é mais significativo, como no mês 0 (janeiro) e mês 2 (março), identificamos oportunidades para refinar o modelo. Esses erros podem ser atribuídos a variáveis não consideradas pelo modelo ou a eventos atípicos que afetam o comportamento do mosquito.

A análise detalhada do desempenho mensal fornece insights valiosos sobre a aplicabilidade do modelo em diferentes períodos do ano. Por exemplo, a maior precisão do modelo durante os meses de pico de atividade do mosquito pode ser aproveitada para alocar recursos de maneira eficiente. Por outro lado, os meses com maior erro podem beneficiar-se de uma investigação mais aprofundada para entender as discrepâncias e aprimorar o modelo.

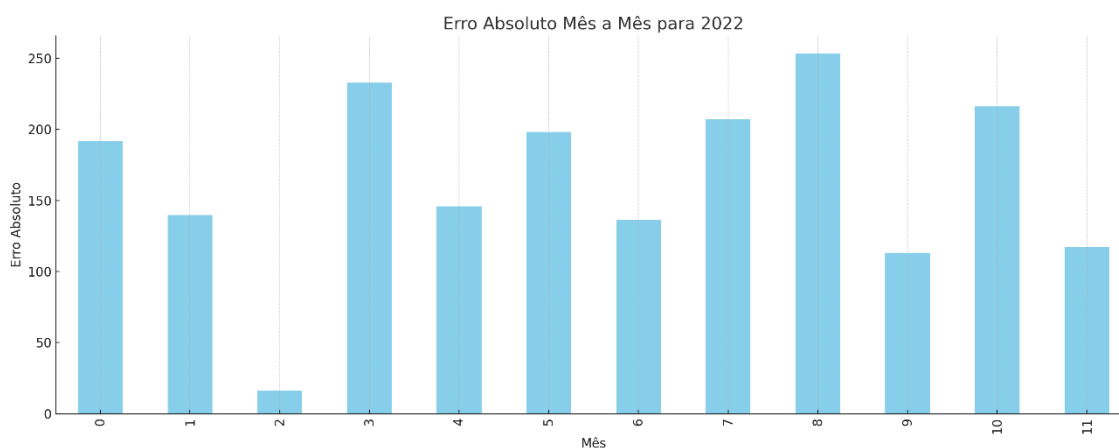


Figura 5 Erro Absoluto mês a Mês para 2022

Acompanhando nossa avaliação do modelo preditivo, a Figura 5 apresenta o gráfico dos erros absolutos para cada mês de 2022. Os erros absolutos são indicativos da diferença entre as previsões do modelo e os valores reais de focos de *Aedes aegypti*. Este gráfico é crucial para identificar os períodos de maior e menor acurácia do modelo.

Sobre a variação mensal dos erros, pode-se entender que:

Menor Erro: O mês que apresentou o menor erro absoluto foi o mês 3 (o mês com a barra mais curta), sugerindo que o modelo teve uma performance excepcionalmente boa na previsão da quantidade de focos de *Aedes aegypti* nesse período. A baixa discrepância entre os valores previstos e reais nesse mês pode ser atribuída a um conjunto de dados bem representativo ou a condições climáticas estáveis que são bem capturadas pelo modelo.

Maior Erro: Em contraste, o mês com o maior erro absoluto foi o mês 8 (setembro). Esse resultado destaca a necessidade de investigação adicional para compreender as causas desse desvio significativo. Fatores como eventos climáticos atípicos, mudanças no ecossistema local ou a introdução de novas práticas de controle de vetores podem ter influenciado essas previsões.

Desta forma, a análise dos erros absolutos revela que, enquanto o modelo é capaz de prever a tendência da atividade do mosquito, sua precisão varia ao longo do ano. Isso ressalta a complexidade dos padrões de comportamento do *Aedes aegypti* e a influência de múltiplos fatores ambientais e biológicos que podem afetar sua proliferação.

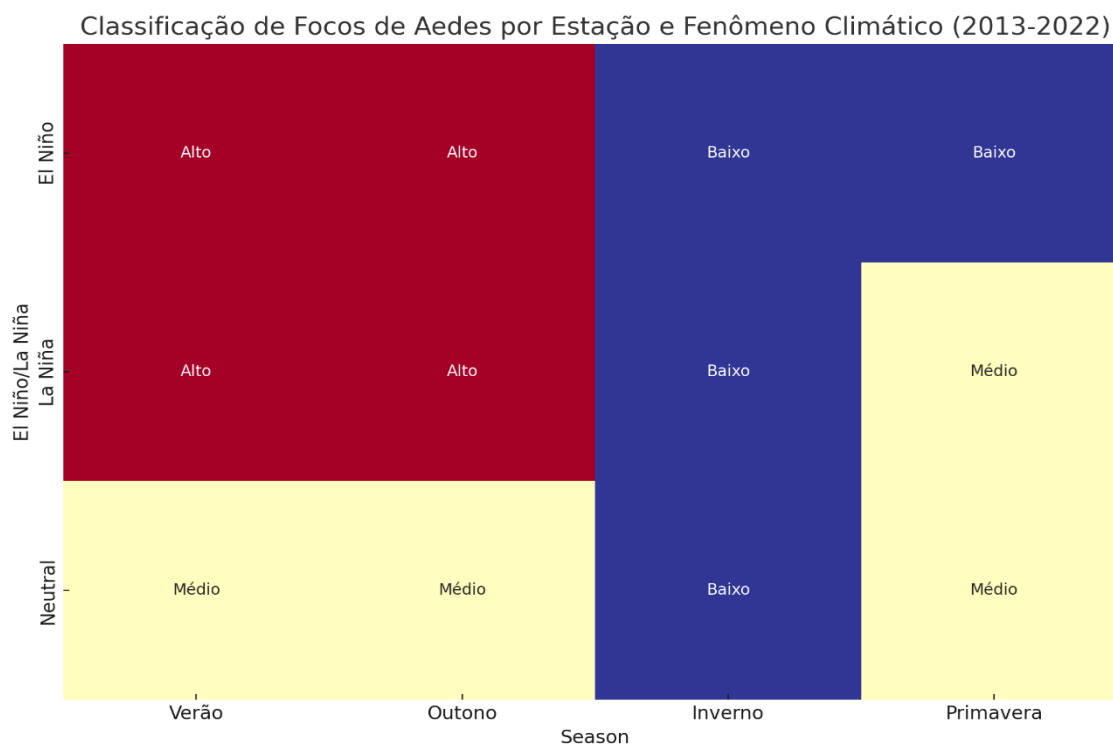


Figura 6 - Modelo preditivo de focos de *Aedes* por estação para Itajaí ajustado ao fenômeno El Niño e La Niña

A Figura 6 destaca a relação entre as estações do ano, fenômenos climáticos El Niño e La Niña, e a atividade de focos de *Aedes aegypti* no município de Itajaí. A matriz utiliza um esquema de cores intuitivo para denotar a intensidade dos focos do mosquito: vermelho para alto, azul para baixo e amarelo para médio.

El Niño/La Niña: Durante os eventos El Niño e La Niña, observamos uma predominância de focos altos no verão e outono (quadrantes vermelhos), o que sugere que as condições climáticas extremas associadas a esses fenômenos favorecem a proliferação do mosquito.

Períodos Neutros: Em condições neutras, sem a influência de El Niño ou La Niña, a atividade do mosquito é classificada como média em todas as estações, exceto na primavera, onde se mantém média (quadrantes amarelos).

Pode-se observar que a variação sazonal é representada pelos quadrantes sendo que os azuis correspondentes ao inverno e à primavera, sob condições de El Niño e La Niña, indicam uma atividade mais baixa do *Aedes aegypti*, o que está em consonância com o comportamento sazonal típico do mosquito.

Discussão

Os resultados da pesquisa desenvolvida demonstram uma relação clara entre as estações do ano, os fenômenos climáticos El Niño e La Niña, e a atividade de focos de *Aedes aegypti* no município de Itajaí. Observamos uma predominância de focos altos de *Aedes aegypti* durante os eventos de El Niño e La Niña no verão e outono, sugerindo que as condições climáticas extremas associadas a esses fenômenos favorecem a proliferação do mosquito, em consonância com o estudo de Jia et al (2017) sobre o *Aedes albopictus*.

Essas descobertas dialogam com outros estudos de destaque na literatura científica, como o de Santos et al (2022), que se debruçou sobre a distribuição atual e futura do mosquito *Aedes aegypti* na Ilha da Madeira. Assim como nosso estudo, este trabalho revela que um clima mais quente pode aumentar a presença do mosquito, especialmente em regiões que atualmente não são consideradas adequadas devido a temperaturas mais baixas ou outras condições climáticas. Isto sugere que o aquecimento do clima poderia potencialmente ampliar as áreas geográficas suscetíveis à dengue e outras doenças transmitidas por este vetor.

Os resultados do nosso modelo, que obteve um coeficiente de determinação (R^2) de 0,58 e um RMSE

de 303,7 no conjunto de teste, reforçam a ideia de que o modelo pode prever os focos com um erro que não ultrapassa o número máximo de focos observados, indicando sua aplicabilidade prática. No entanto, a magnitude do RMSE sugere que há espaço para melhoria, destacando a necessidade de explorar variáveis adicionais ou técnicas de modelagem mais sofisticadas para aprimorar as previsões.

Como esclarecido por Chicco et al. (2021), o coeficiente de determinação (R^2) é uma métrica estatística que mostra o quanto a variação de uma variável dependente é explicada pela variação de uma ou mais variáveis independentes em um modelo de regressão. Os autores também enfatizam que não há um valor “ideal” para o R^2 , pois sua interpretação depende do contexto do modelo e dos dados. Em geral, um R^2 mais alto indica que o modelo é mais eficaz em explicar a variação da variável dependente, mas um R^2 menor não significa necessariamente que o modelo seja insuficiente. O RMSE (Root Mean Square Error) serve para indicar o quão bem o modelo se ajusta aos dados. Um RMSE menor indica que o modelo é mais preciso na previsão dos valores da variável dependente, conforme Chicco et al (2021).

Exemplificando tal necessidade, o estudo de Raman et al. (2021) na Tailândia, ao investigar modelos preditivos para a dengue, mostrou a relevância dos fatores comportamentais e de conhecimento na modelagem da distribuição do vetor da dengue. Segundo os autores, ao agregar o conhecimento, atitudes e práticas (KAP) nos modelos, estes se mostraram mais influentes na previsão da prevalência do *Aedes aegypti* do que os fatores socioeconômicos e de paisagem.

Portanto, a discussão sobre os resultados da nossa pesquisa, em diálogo com outros estudos existentes, reforça a necessidade de se considerar uma multiplicidade de fatores, incluindo climáticos, comportamentais e de conhecimento, na elaboração de estratégias efetivas para o controle do *Aedes aegypti*. A integração dessas variáveis em nosso modelo preditivo oferece uma ferramenta valiosa para a prevenção de doenças transmitidas por mosquitos, com potencial para impactar positivamente a saúde pública em Itajaí e em outras regiões com características semelhantes.

Considerações finais

A partir dos dados meteorológicos e da aplicação do algoritmo de aprendizado de máquina Random Forest, desenvolvemos um modelo com capacidade significativa para prever a reprodução do mosquito *Aedes aegypti* em Itajaí. A correlação entre as variáveis climáticas e a atividade do mosquito confirma o potencial desses modelos para orientar estratégias de controle de vetores.

Identificamos, contudo, limitações e oportunidades de melhoria. A precisão do modelo variou durante o ano e a influência de eventos climáticos extremos, como El Niño e La Niña, sugere a necessidade de incorporar variáveis adicionais ou ajustar o modelo para capturar a complexidade do comportamento do mosquito.

Acreditamos que a combinação de dados meteorológicos com algoritmos de aprendizado de máquina pode resultar em um modelo eficaz para a reprodução do *Aedes aegypti*. Porém, a eficácia desse modelo depende de uma compreensão completa das interações complexas entre o mosquito, seu ambiente e as condições climáticas. A pesquisa contínua, a experimentação e a integração de diferentes tipos de dados e abordagens de modelagem são essenciais para melhorar a precisão e utilidade desses modelos.

Destacamos que este modelo pode ser replicado para todo o estado de Santa Catarina. Isso não só confirmará a eficácia do modelo, mas também permitirá análises mais profundas em períodos do ano com condições climáticas menos constantes, considerando a ampla variabilidade do clima em Santa Catarina.

Agradecimentos

Esta pesquisa foi contemplada pelo Edital de chamada Pública FAPESC Nº 16/2020 PROGRAMA PESQUISA PARA O SUS: Gestão Compartilhada em Saúde - PPSU (2021TR000550)

Créditos

Eduardo Augusto Werneck Ribeiro – supervisão, manuscrito original

Cleusa Matiola – coleta e tratamento de dados, análise dos dados

Mario Francisco Leal de Quadro – desenho metodológico, análise dos dados e revisão final

Matheus Ferreira de Souza - coleta e tratamento de dados, análise dos dados

Isabel Cristina Bohn - coleta e tratamento de dados, análise dos dados e revisão final

João Augusto Brancher Fuck - análise dos dados e revisão final

Raul Borges Guimarães - análise dos dados e revisão final

Thiago Pereira Alves - análise dos dados e revisão final

Referências

Alexander J, Wilke ABB, Mantero A, Vasquez C, Petrie W, Kumar N, et al. (2022) Using machine learning to understand microgeographic determinants of the Zika vector, *Aedes aegypti*. PLoS ONE 17(12): e0265472. <https://doi.org/10.1371/journal.pone.0265472>

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ. Computer science, 7, e623. <https://doi.org/10.7717/peerj-cs.623>

Guerra, C., & Quadro, M. (2021). Estudo da variabilidade diurna da precipitação região sul do Brasil. Estrabão, 2(1), 205–209. <https://doi.org/10.53455/re.v2i.56>

Jupyter. (s.d.). Recuperado de <https://jupyter.org/>

Matiola, C. Avaliação da relação entre focos de *Aedes aegypti* com a distribuição espacial da temperatura no município de Itajaí/SC, por geoprocessamento. Metodologias e Aprendizado, 1, 30 – 35, 2019. <https://doi.org/10.21166/metapre.v1i0.645>

Matiola, C. et al. O uso de dados de temperatura e precipitação de MERRA2 para compreender a dinâmica ecológica do *A. aegypti* no município de Chapecó/SC - 2007 a 2017. Revista Brasileira de Geografia Física, v. 12, n. 4, p. 1385-1398, nov. 2019. <https://doi.org/10.26848/rbgf.v12.4.p1385-1398>

Ochida, N., Mangeas, M., Dupont-Rouzeyrol, M., Dutheil, C., Forfait, C., Peltier, A., ... & Menkès, C. E. (2022). Modeling present and future climate risk of dengue outbreak, a case study in new caledonia. Environmental Health, 21(1). <https://doi.org/10.1186/s12940-022-00829-z>

Rahman, M. S., Pientong, C., Zafar, S., Ekalaksananan, T., Paul, R. E., Haque, U., Rocklöv, J., & Overgaard, H. J. (2021). Mapping the spatial distribution of the dengue vector *Aedes aegypti* and predicting its abundance in northeastern Thailand using machine-learning approach. One Health, 13(7446), 100358. <https://doi.org/10.1016/j.onehlt.2021.100358>

Ribeiro, E., Matiola, C., Mattedi, M., Bohn, I., Fuck, J., Guimaraes, R., Quadro, M., & Alves, T. (2023). Desafios e Aprendizados na Aplicação do Projeto Índice de Positividade de Armadilhas (IPA): Um Estudo sobre o Controle do *Aedes aegypti* em Santa Catarina, Brasil. Metodologias E Aprendizado, 6, 740–743. <https://doi.org/10.21166/metapre.v6i.4174>

Santos, J. M., Capinha, C., Rocha, J., & Sousa, C. A. (2022). The current and future distribution of the yellow fever mosquito (*Aedes aegypti*) on Madeira Island. PLOS Neglected Tropical Diseases, 16(9), e0010715. <https://doi.org/10.1371/journal.pntd.0010715>