

Previsão Semanal de Casos de Dengue utilizando Séries Temporais e Algoritmos de Aprendizado de Máquina

Jurandy A. N. Junior¹, Giovanni Zanardo¹,
Fabiola G. Nakamura¹, Eduardo F. Nakamura¹

¹Instituto de Computação (IComp) – Universidade Federal do Amazonas (UFAM)
Av. Gen. Rodrigo Octávio, 6200, Coroado I, Setor Norte
69080-900 – Manaus – AM – Brasil

{jurandy.junior,gio.zanardo,fabiola,nakamura}@icomp.ufam.edu.br

Abstract. *Dengue represents an important public health challenge in Brazil, especially in tropical regions, where climatic factors favor vector proliferation and the recurrence of outbreaks. This paper presents partial results of a study on the weekly prediction of dengue cases using time series and machine learning algorithms, based on data from the InfoDengue platform between 2010 and 2025. Linear Regression, Random Forest, K-Nearest Neighbors, as well as the ensemble methods Stacking Regressor and Voting Regressor, were tested considering one-step and multi-step approaches. The results indicated better performance of the Stacking Regressor model in the one-step scenario, followed by Linear Regression. It is concluded that ensemble methods show promising results for dengue case prediction and may contribute to monitoring and prevention actions in public health.*

Keywords: *Time series; Machine learning; Dengue; Ensemble methods; Forecasting.*

Resumo. *A dengue representa um importante desafio à saúde pública no Brasil, especialmente em regiões tropicais, onde fatores climáticos favorecem a proliferação do vetor e a recorrência de surtos. Este trabalho apresenta resultados parciais de uma pesquisa sobre a previsão semanal de casos de dengue por meio de séries temporais e algoritmos de aprendizado de máquina, utilizando dados da plataforma InfoDengue entre 2010 e 2025. Foram testados os algoritmos Regressão Linear, Random Forest, K-Nearest Neighbors, além dos métodos ensemble Stacking Regressor e Voting Regressor, considerando abordagens one-step e multi-step. Os resultados indicaram melhor desempenho do modelo Stacking Regressor no cenário one-step, seguido pela Regressão Linear. Conclui-se que os métodos ensemble apresentam resultados promissores para a predição de casos de dengue e podem contribuir para ações de monitoramento e prevenção em saúde pública.*

Palavras-chave: *Séries temporais; Aprendizado de máquina; Dengue; Métodos ensemble; Previsão.*

1. Introdução

De acordo com a Organização Mundial da Saúde (OMS), em 2021, cerca de 1,65 bilhão de pessoas necessitavam de tratamento e cuidados relacionados às Doenças Tropicais Negligenciadas (DTNs), notificadas em 179 países e territórios [Nações Unidas 2023]. Entre

essas doenças, as arboviroses se destacam como importante desafio à saúde pública, especialmente em regiões tropicais e subtropicais, onde fatores climáticos e socioambientais favorecem a proliferação de vetores e a recorrência de surtos epidêmicos.

Nesse contexto, a dengue representa uma das arboviroses de maior impacto no Brasil, com elevada incidência e recorrentes sobrecargas aos sistemas de saúde, especialmente durante períodos sazonais de maior transmissão. Além dos impactos sanitários, a doença produz relevantes consequências sociais e econômicas, exigindo estratégias permanentes de vigilância, prevenção e controle.

Embora existam vacinas disponíveis e ações contínuas de combate ao vetor, o controle da dengue ainda enfrenta limitações estruturais, ambientais e operacionais [Dias et al. 2024]. Dessa forma, torna-se fundamental o uso de ferramentas analíticas capazes de antecipar o comportamento epidemiológico da doença, permitindo maior eficiência no planejamento de campanhas preventivas, na alocação de recursos públicos e na organização da rede assistencial.

A modelagem preditiva baseada em séries temporais associada a algoritmos de aprendizado de máquina apresenta-se como uma alternativa promissora para a previsão semanal de casos de dengue. Essas abordagens permitem identificar padrões históricos, sazonalidade e relações entre variáveis climáticas e epidemiológicas, contribuindo para previsões mais precisas e para o fortalecimento da tomada de decisão em saúde pública [Silva et al. 2020].

Este trabalho apresenta resultados parciais de uma pesquisa em andamento voltada à previsão semanal de casos de dengue, utilizando dados da plataforma InfoDengue [InfoDengue], com a aplicação de modelos de regressão e métodos ensemble, considerando abordagens one-step e multi-step, a fim de avaliar o desempenho preditivo desses algoritmos e sua aplicabilidade no monitoramento epidemiológico.

2. Trabalhos Relacionados

Estudos recentes têm demonstrado a relevância da aplicação de modelos preditivos e técnicas de aprendizado de máquina na previsão de casos de dengue. Xiaoli Chen e Paula Moraga [Chen and Moraga 2025] analisaram tendências da dengue utilizando modelos estatísticos e algoritmos de machine learning combinados em métodos ensemble, com uma janela temporal de 312 semanas para capturar padrões de longo prazo. Os autores observaram que os modelos multi-step apresentaram melhor capacidade de previsão devido as janelas temporais mais longas e ao uso de ensembles: combinando regressores potentes em curto e médio prazo, como ARIMA, e modelos com alta capacidade de predição a longo prazo, como prophet.

De forma semelhante, Alessandro Sebastianelli et al. [Sebastianelli et al. 2024] propuseram um modelo ensemble composto por três algoritmos (CatBoost, SVM e LSTM) para a previsão semanal de casos de dengue em escala nacional. O estudo demonstrou resultados promissores na antecipação de surtos e destacou a possibilidade de replicação da metodologia em outros países, como o Peru, reforçando a relevância internacional desse tipo de abordagem preditiva.

Esses trabalhos evidenciam que a utilização de métodos ensemble podem ser mais precisos em comparação a modelos isolados, justificando sua adoção nesta pesquisa.

3. Metodologia

3.1. Base de Dados e Pré-processamento

Os dados utilizados neste estudo foram obtidos por meio da API pública da plataforma InfoDengue, contemplando registros semanais de casos confirmados de dengue no período de 2010 a 2025. Para esta etapa inicial da pesquisa, foi utilizada a série histórica do município do Rio de Janeiro.

As variáveis selecionadas foram: número semanal de casos confirmados (casos), temperatura média (tempmed) e umidade relativa média do ar (umidmed). A inclusão das variáveis climáticas justifica-se pela sua influência direta no ciclo de vida do vetor e na dinâmica de transmissão da doença.

O pré-processamento consistiu na organização cronológica crescente da base de dados, remoção de registros com valores ausentes e preparação das variáveis para construção das janelas temporais utilizadas nos modelos preditivos.

3.2. Modelagem

Para a modelagem preditiva, foram construídas variáveis defasadas (lags) a partir de janelas temporais deslizantes, permitindo capturar padrões de tendência, sazonalidade e atrasos entre fatores climáticos e a incidência da dengue. Nos modelos one-step, foi utilizada uma janela temporal de $T = 10$ semanas, com o objetivo de prever o número de casos na semana seguinte ($t + 1$), a partir das informações disponíveis até o tempo t .

Nos modelos multi-step, foram realizadas previsões de duas até doze semanas à frente ($t + 2$ até $t + 12$). Para essa abordagem, foi utilizada uma janela temporal de $T = 50$ semanas, buscando reduzir a influência de variações imediatas e priorizar a identificação de tendências mais amplas da série temporal. A divisão dos dados foi realizada em 80% para treinamento e 20% para teste, sem embaralhamento, preservando a ordem cronológica da série e evitando vazamento de informação temporal.

Foram avaliados cinco algoritmos, três regressores individuais: Regressão Linear, Random Forest Regressor e K-Nearest Neighbors Regressor ($K = 20$); e dois métodos ensemble: Stacking Regressor, utilizando Gradient Boosting Regressor como meta-modelo, e Voting Regressor. O objetivo foi comparar o desempenho dos modelos individuais com os métodos ensemble na previsão de casos semanais de dengue.

3.3. Métricas de Avaliação

O desempenho dos modelos foi avaliado por meio das métricas Mean Absolute Error (MAE) e Root Mean Squared Error (RMSE) [Hyndman and Athanasopoulos 2018]. Ademais, essas métricas serão comparadas ao RMSE e MAE do *Naive Approach*, definido como $\hat{y}(t + 1) = y(t)$.

4. Resultados Parciais

Os resultados obtidos demonstram diferenças significativas no desempenho dos modelos avaliados para a previsão semanal de casos de dengue. Embora as figuras apresentem a predição em toda a série temporal, com a comparação entre valores reais e preditos, apenas o conjunto de teste foi considerado para o cálculo das métricas de desempenho.

O *Naive Approach* resultou em um RMSE de 699.47 e um MAE de 292.52, logo, métodos com estatística maiores serem considerados ineficazes na previsão de casos de dengue.

4.1. LinearRegressor

O modelo de Regressão Linear, no cenário one-step, obteve RMSE de 541,14 e MAE de 243,26, demonstrando estatísticas melhores que os demais modelos isolados, apesar de ser o algoritmo de implementação mais simples dentre os explorados neste artigo.

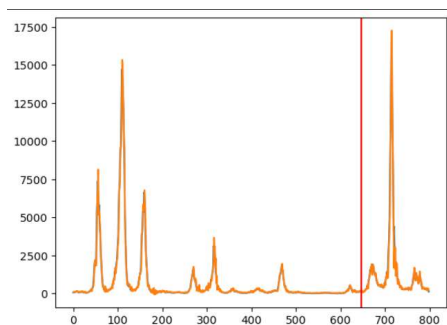


Figura 1. Resultados Linear Regressor One-Step – Número de Casos x Semana

4.2. KNearestNeighborsRegressor

O modelo K-Nearest Neighbors apresentou desempenho inferior em comparação aos demais métodos avaliados, com RMSE de 1266,12 e MAE de 471,92. Os resultados indicam a menor capacidade de generalização para previsão dos casos semanais dentre os métodos abordados, sendo um método ineficaz quando isolado.

4.3. RandomForestRegressor

O Random Forest Regressor apresentou desempenho intermediário, com RMSE de 736,32 e MAE de 284,73. Embora superior ao K-Nearest Neighbors, o modelo não superou o *Naive Approach*.

4.4. StackingRegressor

O modelo Stacking Regressor apresentou o melhor desempenho entre todos os algoritmos avaliados no cenário one-step, com RMSE de 474,60 e MAE de 224,75. Esses resultados evidenciam a eficácia da combinação entre diferentes regressores, permitindo melhor capacidade preditiva em relação aos modelos individuais. No cenário multi-step, o RMSE variou entre 710 e 2944, indicando um aumento progressivo do erro conforme o horizonte de previsão se ampliava, comportamento esperado, pois a informação providenciada ao modelo se torna cada vez mais distante do horizonte a ser predito.

4.5. VotingRegressor

O modelo *Voting Regressor* apresentou RMSE 717.99 de e MAE de 272.28 no *one-step*, representando uma eficácia ambígua, com o MAE menor e RMSE maior que o *Naive Approach*. Já no *multi-step* o valor de RMSE variou entre 1030 e 2421.

De forma geral, os resultados parciais indicam que os métodos *ensemble*, especialmente o Stacking Regressor, apresentam maior potencial para a previsão semanal de casos de dengue, reforçando sua aplicabilidade como ferramenta de apoio ao monitoramento epidemiológico e à tomada de decisão em saúde pública.

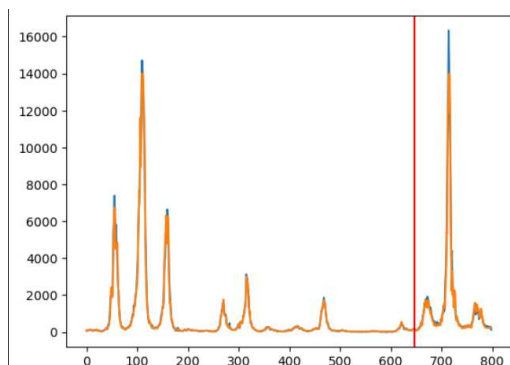


Figura 2. Resultados Stack One-Step – Número de Casos x Semana

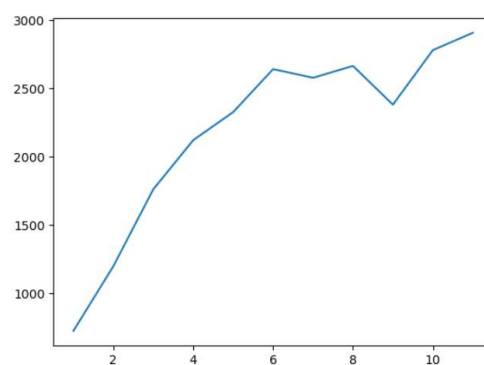


Figura 3. RMSE Stack Multi-Step x Número de Semanas à Frente

Figura 4. Resultados do modelo Stacking Regressor.

5. Conclusões Parciais

Este trabalho apresentou resultados parciais de uma pesquisa voltada à previsão semanal de casos de dengue por meio de séries temporais e algoritmos de aprendizado de máquina, utilizando dados da plataforma InfoDengue no período de 2010 a 2025. Foram avaliados cinco modelos preditivos, incluindo Regressão Linear, Random Forest, K-Nearest Neighbors e os métodos ensemble Stacking Regressor e Voting Regressor, considerando abordagens one-step e multi-step.

Os resultados demonstraram um desempenho misto, com destaque para o Stacking Regressor no cenário one-step, seguido pela Regressão Linear, demonstrando que métodos ensemble podem apresentar maior capacidade preditiva em comparação a modelos individuais, mesmo quando compostos de modelos não muito eficazes quando isolados, evidenciando que cada regressor possui vantagens e desvantagens. Nos cenários multi-step, observou-se maior degradação dos resultados à medida que o horizonte de previsão aumentava, demonstrando a complexidade da antecipação de surtos epidemiológicos em longo prazo.

Com base nos resultados obtidos, o projeto terá como objetivo expandir os seguintes tópicos:

- **Filtros e Wavelets:** a implementação de filtros contribui na remoção de distúrbios da série temporal [Tamilselvi et al. 2025], gerando assim um modelo capaz de analisar tendências gerais e, consequentemente, produzir previsões mais precisas.
- **Expansão da Base de Dados:** de forma semelhante ao ponto anterior, uma base de dados maior pode ajudar na capacidade de generalização do modelo. Para isso, pode-se agrupar regiões brasileiras com características climáticas semelhantes, a fim de potencializar os regressores e possibilitar uma análise mais expansiva do país. Ademais, há a possibilidade de treinar o modelo em uma base de dados, e testá-la em outra, semelhante ao experimento realizado por [Sebastianelli et al. 2024].
- **Explorar Outros Modelos de Regressão:** como foi explorado em trabalhos semelhantes, há vários algoritmos comumente utilizados para previsões de séries temporais que não foram abordados neste projeto, como *Prophet* e *SARIMAX*.

Conforme esses modelos forem incluídos no escopo deste projeto, espera-se que a eficácia dos *ensembles* aumente, por serem capazes de analisar os pontos fortes e fracos de cada modelo [Aceña et al. 2022].

Referências

- Aceña, V., Martín de Diego, I., Fernández, R. R., and Moguerza, J. M. (2022). Minimally overfitted learners: A general framework for ensemble learning. *Knowledge-Based Systems*, 254:109669. Disponível em: <https://doi.org/10.1016/j.knosys.2022.109669>. Acesso em: 23 mar. 2026.
- Chen, X. and Moraga, P. (2025). Assessing dengue forecasting methods: a comparative study of statistical models and machine learning techniques in rio de janeiro, brazil. *Tropical Medicine and Health*. Disponível em: <https://doi.org/10.1186/s41182-025-00723-7>. Acesso em: 20 mar. 2026.
- Dias, R. I. R. et al. (2024). Impacto das medidas de prevenção e promoção da saúde na epidemiologia da dengue no brasil: uma revisão sistemática. *Brazilian Journal of Implantology and Health Sciences*, 6(3):1069–1078. Disponível em: <https://doi.org/10.36557/2674-8169.2024v6n3p1069-1078>. Acesso em: 20 mar. 2026.
- Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts, Melbourne, 2 edition. Disponível em: <https://otexts.com/fpp2/>. Acesso em: 13 abr. 2026.
- InfoDengue. Infodengue – serviços de api. Disponível em: <https://info.dengue.mat.br/services/api>. Acesso em: 15 out. 2025.
- Nações Unidas (2023). Mais de 1,6 bilhão precisam de cuidados contra doenças tropicais negligenciadas. ONU News. Disponível em: <https://news.un.org/pt/story/2023/01/1808967>. Acesso em: 1 abr. 2025.
- Sebastianelli, A., Spiller, D., Carmo, R., et al. (2024). A reproducible ensemble machine learning approach to forecast dengue outbreaks. *Scientific Reports*, 14:3807. Disponível em: <https://doi.org/10.1038/s41598-024-52796-9>. Acesso em: 20 mar. 2026.
- Silva, N. d. S. et al. (2020). Avaliação da relação entre a climatologia, as condições sanitárias (lixo) e a ocorrência de arboviroses (dengue e chikungunya) em quixadáce no período entre 2016 e 2019. *Revista Brasileira de Meteorologia*, 35:485–492. Disponível em: <https://doi.org/10.1590/0102-77863530014>. Acesso em: 23 mar. 2023.
- Tamilselvi, C., Paul, R. K., Yeasin, M., et al. (2025). Novel wavelet-lstm approach for time series prediction. *Neural Computing and Applications*, 37:10521–10530. Disponível em: <https://doi.org/10.1007/s00521-024-10561-z>. Acesso em: 20 mar. 2026.