

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
INSTITUTO DE CIÊNCIA E TECNOLOGIA
CÂMPUS DE SOROCABA

MAURICIO DE SOUZA E CASTRO

**UM ESTUDO COMPARATIVO DE MODELOS PREDITIVOS PARA PREVISÃO E
MONITORAMENTO DE DENGUE E INFLUENZA NO BRASIL**

Sorocaba
2026

MAURICIO DE SOUZA E CASTRO

**UM ESTUDO COMPARATIVO DE MODELOS PREDITIVOS PARA PREVISÃO E
MONITORAMENTO DE DENGUE E INFLUENZA NO BRASIL**

Trabalho de Conclusão de Curso apresentado à
Universidade Estadual Paulista (UNESP),
Instituto de Ciência e Tecnologia, Sorocaba,
como parte dos requisitos para obtenção do grau
de Bacharel em Engenharia de Controle e
Automação.

Orientador: Prof. Dr. Márcio Alexandre Marques

Sorocaba
2026

C355e Castro, Maurício de Souza e
Um estudo comparativo de modelos preditivos para previsão e
monitoramento de dengue e influenza no Brasil / Maurício de Souza e
Castro. -- Sorocaba, 2026
51 f. : il., tabs.

Trabalho de conclusão de curso (Bacharelado em Engenharia de
Controle e Automação) - Universidade Estadual Paulista (UNESP),
Instituto de Ciência e Tecnologia, Sorocaba
Orientador: Márcio Alexandre Marques

1. Aprendizagem de máquina. 2. Dengue. 3. Influenza. 4.
Epidemiologia. 5. Análise de séries temporais. I. Título.

MAURICIO DE SOUZA E CASTRO

**UM ESTUDO COMPARATIVO DE MODELOS PREDITIVOS PARA PREVISÃO E
MONITORAMENTO DE DENGUE E INFLUENZA NO BRASIL:**

Trabalho de Conclusão de Curso apresentado à Universidade Estadual Paulista (UNESP), Instituto de Ciência e Tecnologia, Sorocaba, para obtenção do título de Bacharel em Engenharia de Controle e Automação.

Data da defesa: 11/06/2026

Banca Examinadora:

Prof. Dr. Márcio Alexandre Marques

UNESP - Instituto de Ciência e Tecnologia – Campus de Sorocaba

Prof. Dr. Eduardo Verri Liberado

UNESP - Instituto de Ciência e Tecnologia – Campus de Sorocaba

Prof. Dr. Everson Martins

UNESP - Instituto de Ciência e Tecnologia – Campus de Sorocaba

RESUMO

A previsão de doenças infecciosas de transmissão vetorial e respiratória constitui desafio relevante para a saúde pública brasileira. Este trabalho propõe, implementa e compara três modelos preditivos - Prophet, Random Forest (RF) e Long Short-Term Memory (LSTM) - enriquecidos com regressores climáticos de temperatura e precipitação, para a previsão semanal de casos de influenza e dengue em todas as 27 Unidades Federativas do Brasil (UFs). Os dados de morbidade foram extraídos do Sistema de Informação de Agravos de Notificação (SINAN) e os climáticos do Instituto Nacional de Meteorologia (INMET), cobrindo o período de 2016 a 2024. A validação adota estratégia *walk-forward* com dois cenários independentes: treino até 2022 e teste em 2023; treino até 2023 e teste em 2024. Um painel interativo elaborado em Microsoft Power BI sintetiza os resultados para análise comparativa. Para influenza, o Prophet superou os demais no cenário 2023 (MASE médio: 0,37; 26 de 27 UFs melhores que o *naïve*), enquanto RF e LSTM lideraram em 2024 (MASE: 0,45 e 0,49). Para dengue, o RF foi superior em ambos os cenários (MASE: 0,45 em 2023; 3,73 em 2024). Todos os modelos falharam no surto histórico de dengue em 2024 ($MASE > 3,7$), evento fora da distribuição histórica sem paralelo nos dados de treinamento. Uma limitação estrutural crítica identificada é a alteração dos registros de Síndrome Respiratória Aguda Grave (SRAG) por COVID-19 entre 2020 e 2022, que compromete o treinamento dos modelos de influenza. Os resultados evidenciam que dois dos três modelos avaliados (Prophet e LSTM) apresentam desempenho insatisfatório especificamente para séries de dengue, com WAPE próximo a 100% mesmo em anos típicos. Para influenza, o Prophet demonstrou robustez superior no cenário 2023 e todos os modelos convergiram em 2024. O Random Forest mostrou-se o modelo com melhor desempenho em relação ao *naïve* para dengue em anos típicos (cenário 2023, $MASE = 0,45$), graças à combinação de atributos de defasagem explícitos e capacidade de modelar não linearidades clima-vetor. No cenário 2024, todos os modelos colapsaram diante do surto histórico, tornando a comparação de desempenho relativo irrelevante.

Palavras-chave: Prophet, Random Forest, LSTM, Dengue, Influenza, Regressores, climáticos, Previsão epidemiológica, MASE, Validação *walk-forward*, Power BI.

ABSTRACT

Forecasting vector-borne and respiratory infectious diseases is a relevant challenge for Brazilian public health. This study proposes, implements, and compares three predictive models - Prophet, Random Forest (RF), and Long Short-Term Memory (LSTM) - enriched with climate regressors of temperature and precipitation, for the weekly forecasting of influenza and dengue cases across all 27 Brazilian Federative Units (FUs). Morbidity data were extracted from the Notifiable Diseases Information System (SINAN) and climate data from the National Institute of Meteorology (INMET), covering the period from 2016 to 2024. Validation adopts a walk-forward strategy with two independent scenarios: training through 2022 and testing in 2023; training through 2023 and testing in 2024. An interactive dashboard developed in Microsoft Power BI synthesizes the results for comparative analysis. For influenza, Prophet outperformed the other models in the 2023 scenario (average MASE: 0.37; 26 of 27 FUs better than the naïve), while RF and LSTM led in 2024 (MASE: 0.45 and 0.49). For dengue, RF was superior in both scenarios (MASE: 0.45 in 2023; 3.73 in 2024). All models failed during the historic 2024 dengue outbreak ($\text{MASE} > 3.7$), an out-of-distribution event with no parallel in the training data. A critical structural limitation identified is the alteration of Severe Acute Respiratory Syndrome (SARS) records by COVID-19 between 2020 and 2022, which compromises influenza model training. Results show that two of the three evaluated models (Prophet and LSTM) perform unsatisfactorily specifically for dengue series, with WAPE near 100% even in typical years. For influenza, Prophet demonstrated superior robustness in the 2023 scenario and all models converged in 2024. Random Forest was the best model to outperform the naïve for dengue in typical years (2023 scenario, $\text{MASE} = 0.45$), owing to its explicit lag features and capacity to model climate-vector nonlinearities. In the 2024 scenario, all models collapsed before the historic outbreak, rendering relative performance comparisons irrelevant.

Keywords: Prophet. Random Forest. LSTM. Dengue. Influenza. Climate regressors. Epidemiological forecasting. MASE. Walk-forward validation. Power BI.

LISTA DE ABREVIATURAS E SIGLAS

ARIMA	Autoregressive Integrated Moving Average
DAX	Data Analysis Expressions
IBGE	Instituto Brasileiro de Geografia e Estatística
INMET	Instituto Nacional de Meteorologia
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MASE	Mean Absolute Scaled Error
RF	Random Forest
RMSE	Root Mean Squared Error
SINAN	Sistema de Informação de Agravos de Notificação
SRAG	Síndrome Respiratória Aguda Grave
UF	Unidade Federativa
WAPE	Weighted Absolute Percentage Error

SUMÁRIO

Sumário

1 INTRODUÇÃO	10
1.1 Objetivos.....	11
1.1.1 Objetivo geral	11
1.1.2 Objetivos específicos.....	11
2 REFERENCIAL TEÓRICO.....	12
2.1 Dengue e influenza no Brasil: epidemiologia e sazonalidade	12
2.2 Fundamentos de aprendizado de máquina.....	13
2.2.1 Aprendizado supervisionado e problema de regressão.....	13
2.2.2 Dilema viés-variância e sobreajuste	14
2.2.3 Árvores de decisão.....	14
2.2.4 Redes neurais artificiais e retropropagação	15
2.3 Prophet: <i>framework</i> de previsão de séries temporais	16
2.4 Random Forest para previsão de séries temporais.....	17
2.5 Redes Long Short-Term Memory (LSTM)	18
2.6 Regressores climáticos em modelagem epidemiológica	19
2.7 Métricas de avaliação de previsões	20
2.8 Validação <i>walk-forward</i>	21
2.9 Painéis interativos em Power BI para epidemiologia.....	22
3 METODOLOGIA.....	23
3.1 Fontes de dados	23
3.1.1 Dados epidemiológicos - SINAN	23
3.1.2 Dados climáticos - INMET	23
3.2 <i>Pipeline</i> de coleta e integração de dados	24
3.3 Engenharia de regressores	26
3.4 Configuração do modelo Prophet.....	26
3.5 Configuração do modelo Random Forest.....	26
3.6 Configuração do modelo LSTM.....	27
3.7 Estratégia de validação <i>walk-forward</i>	27

3.8 Painel interativo em Power BI.....	28
4 RESULTADOS E DISCUSSÃO	30
4.1 Visão geral comparativa	30
4.2 Resultados de influenza	31
4.2.1 Cenário influenza 2023.....	31
4.3 Resultados de dengue	33
4.3.1 Cenário dengue 2023	33
4.3.2 Cenário dengue 2024 - surto histórico.....	35
4.4 Impacto da pandemia de COVID-19 nos dados de treinamento	35
4.5 Análise de padrões geográficos e sazonalidade regional.....	36
4.6 Limitações do estudo e implicações metodológicas.....	41
4.7 Orientações práticas para sistemas de vigilância epidemiológica	41
5 CONCLUSÃO.....	43
6 TRABALHOS FUTUROS.....	45
7 USO DE INTELIGÊNCIA ARTIFICIAL.....	46
REFERÊNCIAS	47
APÊNDICE A	49

1 INTRODUÇÃO

O Brasil enfrenta, de forma recorrente, surtos de doenças infecciosas que impactam diretamente o sistema de saúde pública. A dengue e a influenza apresentam comportamento sazonal bem definido e são fortemente influenciadas por fatores climáticos como temperatura e precipitação. Em 2024, o país registrou o maior surto de dengue de sua história, com mais de 6 milhões de casos notificados, evidenciando a necessidade urgente de ferramentas capazes de antecipar picos epidemiológicos com antecedência suficiente para que gestores de saúde possam atuar preventivamente (MINISTÉRIO DA SAÚDE, 2024).

A previsão de séries temporais epidemiológicas constitui uma área de crescente interesse. Modelos estatísticos clássicos como ARIMA fornecem uma linha de base amplamente adotada, mas não incorporam naturalmente regressores externos nem capturam sazonalidades múltiplas sem extensões explícitas. Entre as alternativas modernas, o Prophet (TAYLOR; LETHAM, 2018) destaca-se pela robustez a dados faltantes e capacidade de modelar sazonalidades múltiplas com regressores externos. O Random Forest (RF) (BREIMAN, 2001) é um *ensemble* de árvores de decisão reconhecido por capturar relações não lineares entre variáveis climáticas e incidência de doenças. As redes Long Short-Term Memory - LSTM (HOCHREITER; SCHMIDHUBER, 1997) são redes neurais recorrentes capazes de aprender dependências temporais de longo prazo, particularmente relevantes em séries com memória sazonal.

Um fator complicador relevante é o impacto da pandemia de COVID-19 sobre os dados de vigilância epidemiológica. Entre 2020 e 2022, os registros de Síndrome Respiratória Aguda Grave (SRAG) no SINAN foram dominados por casos de COVID-19, gerando uma quebra estrutural severa na série histórica de influenza. Modelos treinados nesse período aprenderam padrões distorcidos, com reflexos diretos no desempenho preditivo, especialmente para redes neurais sensíveis à janela de dados mais recente.

A motivação para comparar abordagens algoritmicamente distintas reside na hipótese de que a adequação de cada modelo depende das características específicas da doença e do contexto temporal de treinamento. Estudos anteriores apresentados por Mussumeci e Coelho (2020) compararam LSTM e RF para dengue em escala nacional, porém sem incluir Prophet e sem adotar validação *walk-forward* em dois cenários independentes pós-pandêmicos.

Diante deste cenário, o presente trabalho tem como objetivo desenvolver, treinar e comparar os três modelos com regressores climáticos para a previsão semanal de casos de

influenza e dengue em todas as 27 unidades federativas do Brasil, utilizando validação *walk-forward* com cenários de teste em 2023 e 2024. Os resultados são apresentados em um painel interativo desenvolvido no Power BI para auxiliar a sua análise e visualização.

1.1 Objetivos

1.1.1 Objetivo geral

Desenvolver e comparar três modelos de previsão semanal de casos de influenza e dengue por unidade federativa - Prophet, Random Forest e LSTM - com regressores climáticos de temperatura e precipitação, avaliando o desempenho relativo de cada abordagem em dois cenários de validação *walk-forward* independentes (2023 e 2024).

1.1.2 Objetivos específicos

- a) Coletar, tratar e integrar dados epidemiológicos do SINAN e climáticos do INMET (2016-2024);
- b) Elaborar regressores climáticos com defasagem temporal adequada para cada doença;
- c) Treinar modelos Prophet, Random Forest e LSTM independentes por UF, doença e cenário;
- d) Validar os modelos via *walk-forward* com cenários de teste em 2023 e 2024;
- e) Avaliar e comparar o desempenho preditivo com RMSE, MAE, MASE e WAPE;
- f) Analisar o impacto da pandemia de COVID-19 como quebra estrutural nos dados de influenza;
- g) Desenvolver painel interativo no Power BI para visualização comparativa dos resultados.

2 REFERENCIAL TEÓRICO

2.1 Dengue e influenza no Brasil: epidemiologia e sazonalidade

A dengue é uma arbovirose transmitida pelo mosquito *Aedes aegypti*, endêmica em regiões tropicais e subtropicais (STANAWAY et al., 2016). No Brasil, a doença apresenta caráter hiperendêmico, com surtos anuais de magnitude variável. A transmissão é fortemente associada a fatores climáticos: temperaturas entre 25 °C e 30 °C aceleram o ciclo de vida do vetor, enquanto precipitações abundantes favorecem criadouros em recipientes com água parada (FOCKS et al., 1993; TAUILL, 2002). O pico de incidência geralmente segue a exposição climática com defasagem de duas a quatro semanas, correspondente ao tempo de desenvolvimento larval e ao período de incubação extrínseca no vetor (HUI et al., 2012).

Em 2024, o Brasil registrou mais de 6 milhões de casos notificados de dengue - o maior surto de sua história. Uma hipótese epidemiológica relevante para explicar essa magnitude é a chamada dívida imunológica: a redução da circulação viral durante a pandemia de COVID-19, entre 2020 e 2022, gerou uma população suscetível ampliada, especialmente para o sorotipo DENV-3, que havia circulado com menor frequência nos anos anteriores. A reintrodução deste sorotipo em uma população com menor imunidade de grupo, combinada a condições climáticas favoráveis, produziu a explosão epidêmica observada.

A influenza é uma infecção respiratória aguda causada pelos vírus Influenza A e B. No hemisfério sul, a circulação sazonal concentra-se tipicamente entre os meses de abril e agosto, associada ao inverno e ao início do outono (BRASIL, 2022). A transmissão do vírus influenza é favorecida por temperaturas baixas e baixa umidade relativa do ar, que aumentam a sobrevivência do vírus em aerossóis e superfícies (SHAMAN; KOHN, 2009).

A pandemia de COVID-19 representou uma ruptura sem precedentes na vigilância da influenza: entre 2020 e 2022, os dados de SRAG no SINAN passaram a refletir predominantemente nos casos de COVID-19, com supressão quase total da circulação sazonal da influenza em 2020 e retorno progressivo a partir de 2022-2023.

2.2 Fundamentos de aprendizado de máquina

Aprendizado de máquina (*Machine Learning*) é um ramo da inteligência artificial que estuda algoritmos capazes de aprender padrões a partir de dados e generalizar esse aprendizado para novas observações. Diferente da programação tradicional, em que regras são explicitadas pelo programador, nos algoritmos de aprendizado de máquina as regras emergem indutivamente do conjunto de dados de treinamento (BISHOP, 2006). Esta seção apresenta os fundamentos teóricos que embasam os modelos Prophet, Random Forest (RF) e Long Short-Term Memory (LSTM) utilizados neste trabalho.

2.2.1 Aprendizado supervisionado e problema de regressão

O aprendizado supervisionado é o paradigma mais amplamente utilizado em aprendizado de máquina. Dado um conjunto de treinamento com n exemplos rotulados $\{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_n, \mathbf{y}_n)\}$, onde $\mathbf{x}_i \in \mathbb{R}^p$ é o vetor de atributos e $\mathbf{y}_i \in \mathbb{R}$ é o rótulo alvo, o objetivo é aprender uma função $\mathbf{f}: \mathbb{R}^p \rightarrow \mathbb{R}$ que minimize o risco esperado sobre a distribuição conjunta $\mathbf{P}(\mathbf{X}, \mathbf{Y})$ (HASTIE; TIBSHIRANI; FRIEDMAN, 2009). Como $\mathbf{P}(\mathbf{X}, \mathbf{Y})$ é desconhecida, minimiza-se o risco empírico sobre os dados disponíveis, conforme a Equação (1):

$$L(\theta) = (1/n) \sum \ell(f_{\theta}(x_i), y_i), \quad i=1..n \quad (1)$$

Na Equação (1), θ representa os parâmetros aprendíveis do modelo e $\ell(\cdot, \cdot)$ é a função de perda. Para problemas de regressão - em que $\mathbf{y}_i$ é contínuo, como o número de casos de dengue em uma semana - a função de perda mais comum é o erro quadrático $\ell(\hat{\mathbf{y}}, \mathbf{y}) = (\mathbf{y} - \hat{\mathbf{y}})^2$, cuja minimização equivale a estimar a média condicional $\mathbf{E}[\mathbf{Y} | \mathbf{X} = \mathbf{x}]$.

Para séries temporais, o vetor de atributos $\mathbf{x}_t$ é tipicamente construído a partir de valores defasados da própria série e de variáveis exógenas, transformando o problema de previsão em um problema de regressão supervisionada padrão.

2.2.2 Dilema viés-variância e sobreajuste

Um resultado central da teoria do aprendizado estatístico é a decomposição do erro quadrático médio de generalização em três componentes, conforme a Equação (2) (GEMAN; BIENENSTOCK; DOURSAT, 1992):

$$E[(y - \hat{f}(x))^2] = [Bias(\hat{f}(x))]^2 + Var(\hat{f}(x)) + \sigma^2 \quad (2)$$

Na Equação (2), o viés (*Bias*) mede o erro sistemático causado por suposições incorretas sobre a forma funcional de $\mathbf{f}$; a variância (**Var**) mede a sensibilidade do modelo a flutuações no conjunto de treinamento; e σ^2 é o ruído irreduzível dos dados.

O dilema viés-variância expressa o compromisso fundamental do aprendizado de máquina: modelos muito simples (como regressão linear) apresentam alto viés, mas baixa variância - subajustam os dados; modelos muito complexos (como árvores de decisão profundas) apresentam baixo viés, mas alta variância - memorizam o ruído do treinamento e generalizam mal (sobreajuste).

Técnicas de regularização introduzem restrições ou penalidades para reduzir a variância ao custo de um pequeno aumento de viés. Exemplos incluem: o parâmetro de regularização **L2** (ridge) e **L1** (lasso) em modelos lineares; a limitação de profundidade máxima em árvores de decisão; a exclusão aleatória de neurônios em redes neurais; e a parada antecipada do treinamento por gradiente descendente, que interrompe as iterações quando o erro de validação começa a aumentar, evitando que o modelo memorize o ruído do conjunto de treinamento.

2.2.3 Árvores de decisão

Uma árvore de decisão para regressão particiona recursivamente o espaço de atributos em regiões disjuntas $\mathbf{R}_1, \dots, \mathbf{R}_J$ e atribui a cada região uma previsão constante $\hat{\mathbf{c}}_j$ (a média dos valores alvo na região). O particionamento binário é realizado escolhendo, a cada nó, a variável $\mathbf{j}$ e o limiar $\mathbf{s}$ que minimizam a soma dos erros quadráticos das partições resultantes, conforme a Equação (3):

$$\min_{\{j,s\}} [\sum (y_i - \hat{c}_1)^2 + \sum (y_i - \hat{c}_2)^2] \quad (3)$$

Na Equação (3), $\hat{c}_1$ e $\hat{c}_2$ são as médias de y nos conjuntos $\{i: x_{ij} \leq s\}$ e $\{i: x_{ij} > s\}$, respectivamente. Árvores mais profundas particionam o espaço em regiões cada vez menores, reduzindo o viés ao custo de aumentar a variância. Uma árvore completamente crescida - que particiona até que cada folha contenha uma única observação - tem viés zero no treinamento, mas variância máxima e péssima generalização. Esse comportamento extremo motivou o desenvolvimento de métodos de combinação de estimadores que unem múltiplas árvores para reduzir a variância.

2.2.4 Redes neurais artificiais e retropropagação

Uma rede neural artificial multicamada (Perceptron Multicamadas (MLP)) é composta por camadas de unidades (neurônios) interconectadas. Cada neurônio computa uma combinação linear de suas entradas seguida de uma função de ativação não linear, conforme a Equação (4):

$$a_j^{(l)} = \sigma(\sum_k w_{jk}^{(l)} \cdot a_k^{(l-1)} + b_j^{(l)}) \quad (4)$$

Na Equação (4), $a_j^{(l)}$ é a ativação do j -ésimo neurônio na camada l , $w_{jk}^{(l)}$ é o peso da conexão entre o k -ésimo neurônio da camada $l-1$ e o j -ésimo da camada l , $b_j^{(l)}$ é o viés, e σ é a função de ativação - tipicamente a função sigmoide $\sigma(z) = 1/(1+e^{-z})$ ou a função **ReLU** $\sigma(z) = \max(0, z)$, que combina eficiência computacional com boa capacidade de representação de não linearidades.

O treinamento é realizado pelo algoritmo de retropropagação (*backpropagation*), que computa o gradiente da função de perda em relação a cada peso da rede pela regra da cadeia do cálculo diferencial.

O gradiente é então utilizado pelo otimizador - tipicamente Adam (KINGMA; BA, 2015), uma variante adaptativa do gradiente descendente estocástico - para atualizar os pesos iterativamente.

A profundidade das redes neurais (número de camadas) é o que confere a elas a capacidade de aprender representações hierárquicas dos dados, aprendendo atributos elementares nas camadas iniciais e representações progressivamente mais abstratas nas camadas profundas.

Para séries temporais, as redes neurais recorrentes (RNNs) estendem a arquitetura MLP com conexões de retroalimentação que passam informação de um passo temporal para o seguinte, permitindo modelar dependências temporais. A limitação das RNNs simples - o desaparecimento do gradiente em sequências longas - é superada pelas redes LSTM, descritas na Seção 2.5 deste trabalho.

2.3 Prophet: *framework* de previsão de séries temporais

O Prophet é um *framework* de previsão de código aberto desenvolvido por Taylor e Letham (2018). O modelo decompõe a série temporal em componentes aditivos, conforme a Equação (5):

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \quad (5)$$

onde $\mathbf{g(t)}$ representa o componente de tendência - que pode ser linear ou logístico, $\mathbf{s(t)}$ representa a sazonalidade modelada por séries de Fourier, $\mathbf{h(t)}$ captura efeitos de feriados e eventos especiais, e $\mathbf{\varepsilon(t)}$ é o termo de erro, assumido como normalmente distribuído. A sazonalidade anual é representada por séries de Fourier de ordem $\mathbf{N}$, conforme a Equação (6):

$$s(t) = \sum [a_n \cos(2\pi nt/P) + b_n \sin(2\pi nt/P)], \quad n=1..N \quad (6)$$

onde $\mathbf{P}$ é o período (365,25 dias para sazonalidade anual). O número de termos $\mathbf{N}$ controla a flexibilidade do ajuste: valores maiores capturam sazonalidades mais complexas a custo de maior risco de sobreajuste. Pontos de inflexão são pontos de mudança automática na tendência, detectados por um prior de esparsidade (prior de Laplace com escala changepoint_prior_scale). Regressores externos são incorporados via `add_regressor()`, permitindo incluir variáveis climáticas como covariáveis aditivas no componente $\mathbf{s(t)}$.

Uma vantagem importante do Prophet é sua robustez a dados faltantes e à presença de *outliers* sazonais, o que o torna particularmente adequado para séries históricas com quebras estruturais.

2.4 Random Forest para previsão de séries temporais

O Random Forest (RF) é um método de combinação de estimadores proposto por Breiman (2001), composto por B árvores de decisão treinadas sobre amostras com reposição do conjunto de treinamento. A previsão final é a média das previsões individuais, conforme a Equação (7):

$$\hat{y} = (1/B) \sum T_b(x), \quad b=1..B \quad (7)$$

Na Equação (7), $T_b(x)$ é a previsão da **b-ésima** árvore para a entrada x . A seleção aleatória de subconjuntos de m variáveis em cada divisão (com $m \approx \sqrt{p}$, onde p é o total de atributos) garante diversidade entre as árvores e reduz a correlação entre elas, diminuindo a variância do combinador sem aumentar o viés.

Para previsão de séries temporais, o RF é adaptado ao paradigma supervisionado via “janelamento” temporal: a série é transformada em matriz de exemplos onde cada linha contém as defasagens da variável alvo e os regressores como atributos de entrada, e o alvo é o valor futuro a prever. A previsão multipasso é realizada em modo recursivo: o modelo prevê $t+1$, essa previsão é realimentada como *lag* para gerar $t+2$, e assim sucessivamente. Esse procedimento pode resultar em acúmulo de erro ao longo do horizonte de previsão, isto é, nas semanas mais distantes do ponto de corte, especialmente quando as previsões intermediárias divergem da realidade. O intervalo de confiança de 90% é obtido pelos percentis 5 e 95 das distribuições de previsões das B árvores individuais - explorando a floresta de quantis implícita - sem assumir normalidade dos resíduos.

2.5 Redes Long Short-Term Memory (LSTM)

As redes Long Short-Term Memory (LSTM), propostas por Hochreiter e Schmidhuber (1997), são uma arquitetura de rede neural recorrente projetada para modelar dependências temporais de longo prazo. A célula LSTM mantém um estado de memória $\mathbf{c}_t$ e um estado oculto $\mathbf{h}_t$ controlados por três portas — mecanismos de controle cujos parâmetros são ajustados automaticamente durante o treinamento via gradiente descendente: a porta de esquecimento ($\mathbf{f}_t$), a porta de entrada ($\mathbf{i}_t$) e a porta de saída ($\mathbf{o}_t$).

A porta de esquecimento $\mathbf{f}_t$, formalizada na Equação (8), determina quais informações do estado anterior são descartadas:

$$\mathbf{f}_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (8)$$

A porta de entrada $\mathbf{i}_t$ e o estado candidato $\tilde{\mathbf{c}}_t$, definidos na Equação (9), controlam novas informações a armazenar:

$$\mathbf{i}_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i); \quad \tilde{\mathbf{c}}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (9)$$

A Equação (10) atualiza o estado da célula combinando esquecimento e nova informação:

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \quad (10)$$

Por fim, a Equação (11) define a porta de saída $\mathbf{o}_t$ e o estado oculto $\mathbf{h}_t$ que completa o ciclo da célula:

$$\mathbf{o}_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o); \quad \mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad (11)$$

Nas Equações (8) a (11), σ é a função sigmoide, $\odot$ é o produto elemento a elemento (produto de Hadamard, isto é, a multiplicação termo a termo entre dois vetores ou matrizes de mesma dimensão), e $\mathbf{W}$ e $\mathbf{b}$ representam pesos e vieses aprendidos. O mecanismo de portas descrito nas Equações (8) a (11), permite que a rede controle seletivamente quais informações são mantidas ou descartadas ao longo do tempo, preservando gradientes ao longo de centenas de passos temporais e superando o problema do desaparecimento do gradiente que afeta redes recorrentes simples.

No presente trabalho, a arquitetura utilizada consiste em uma camada LSTM com 64 unidades, camada densa de 32 unidades com ativação ReLU e camada de saída com 1 unidade linear.

2.6 Regressores climáticos em modelagem epidemiológica

A relação entre fatores climáticos e incidência de doenças infecciosas é amplamente documentada. Para a dengue, temperatura e precipitação influenciam diretamente a proliferação e o ciclo de vida do *Aedes aegypti*: temperaturas entre 25 °C e 30 °C minimizam o período de incubação extrínseca do vírus no mosquito, enquanto eventos pluviométricos criam reservatórios temporários de água que ampliam os sítios de ovipostura (FOCKS et al., 1993).

Para a influenza, temperaturas baixas e baixa umidade absoluta do ar aumentam a estabilidade do vírus em aerossóis e em superfícies, favorecendo a transmissão (SHAMAN; KOHN, 2009).

A análise de correlação cruzada com defasagens de zero a oito semanas, conduzida neste trabalho para cada UF e doença, revelou que a correlação entre precipitação e incidência de dengue atinge pico para *lags* de duas a quatro semanas, fundamentando a janela de quatro semanas do regressor de temperatura. Dois regressores finais foram construídos:

- a) **temp_mm_4**: média móvel de quatro semanas da temperatura média semanal, capturando o efeito acumulado sobre a transmissão;
- b) **chuva_log1p**: transformação $\log(1 + \text{precipitação})$ da precipitação semanal acumulada, comprimindo a distribuição assimétrica da chuva e estabilizando a variância.

A transformação logarítmica da precipitação é justificada pela distribuição fortemente assimétrica à direita das séries pluviométricas brasileiras, onde eventos extremos de precipitação representam múltiplos desvios padrão da mediana.

O mesmo regressor é utilizado pelos três modelos, garantindo que as diferenças de desempenho observadas decorram das arquiteturas dos modelos e não da disponibilidade de informação.

2.7 Métricas de avaliação de previsões

A avaliação dos modelos utiliza quatro métricas complementares. O erro médio absoluto (MAE) mede o desvio médio na escala original, conforme a Equação (12):

$$MAE = (1/n) \sum |y_t - \hat{y}_t| \quad (12)$$

O RMSE (Root Mean Squared Error), Equação (13), penaliza erros grandes de forma quadrática, sendo mais sensível a *outliers* de previsão:

$$RMSE = \sqrt{[(1/n) \sum (y_t - \hat{y}_t)^2]} \quad (13)$$

O MASE (Mean Absolute Scaled Error), proposto por Hyndman e Koehler (2006) e formalizado na Equação (14), normaliza o MAE do modelo pelo MAE de um modelo *naïve* sazonal de referência (*modelo que, para cada semana t , prevê o valor observado na mesma semana do ano anterior: $\hat{y}_t = y_{t-52}$; ver Apêndice A*), tornando-o adimensional e comparável entre séries de escalas distintas:

$$MASE = MAE_modelo / MAE_naïve \quad (14)$$

onde MAE_*naïve* é calculado sobre o período de treinamento com o *naïve* sazonal de período 52 semanas ($\hat{y}_t = y_{t-52}$).

Um $MASE < 1$ indica que o modelo supera o *naïve sazonal*; $MASE = 1$ significa desempenho equivalente ao *naïve*. O MASE foi adotado como métrica principal deste trabalho por ser invariante à escala, fundamental para comparações entre os 27 estados com volumes de casos muito distintos.

O WAPE (Weighted Absolute Percentage Error) pondera os erros pelo volume de casos, sendo especialmente adequado para avaliar desempenho durante picos epidêmicos:

$$WAPE = (\sum |y_t - \hat{y}_t| / \sum y_t) \times 100\% \quad (15)$$

O WAPE atribui maior peso a semanas de alto volume (picos), sendo mais sensível à capacidade do modelo de prever a magnitude dos surtos, conforme a Equação (15).

Um $WAPE > 100\%$ indica que o modelo subestima ou superestima os casos em magnitude igual ou superior ao total do período, o que é particularmente revelador no contexto de surtos fora da distribuição histórica.

O MASE (Equação (14)) foi adotado como métrica principal por ser invariante à escala e comparável entre as 27 UFs.

2.8 Validação *walk-forward*

A validação progressiva *walk-forward* (também denominada validação com janela de expansão) simula fielmente o uso real do modelo: treina-se com todos os dados históricos disponíveis até um ponto de corte e prevê-se o período imediatamente seguinte, sem incorporar novos dados ao conjunto de treinamento durante o horizonte de previsão.

Essa estratégia evita o vazamento temporal de dados, problema presente na validação cruzada *k-fold* aplicada a séries temporais, onde amostras de validação podem preceder amostras de treinamento no tempo (HYNDMAN; ATHANASOPOULOS, 2021).

A escolha de dois cenários independentes - 2023 e 2024 - é motivada pela relevância epidemiológica de cada período: 2023 é o primeiro ano pós-pandêmico completo, com padrões de circulação viral em recuperação e dados de treinamento contaminados por COVID-19; 2024 é o ano do maior surto de dengue já registrado no Brasil, representando um teste extremo de generalização dos modelos. A comparação entre os dois cenários permite

avaliar a estabilidade do desempenho dos modelos ao longo do tempo, aspecto crítico para sistemas de vigilância epidemiológica em produção.

Como experimento complementar, foram avaliadas estratégias de validação *holdout* com múltiplos cenários históricos (pré-pandemia, pandemia, pós-pandemia); porém, os resultados obtidos não foram considerados representativos do desempenho real dos modelos, em razão de inconsistências nos dados de vigilância decorrentes das quebras estruturais do período analisado. Os resultados reportados neste trabalho referem-se exclusivamente à validação *walk-forward*.

2.9 Painéis interativos em Power BI para epidemiologia

A comunicação de resultados de modelos preditivos para gestores de saúde pública requer interfaces acessíveis que permitam análise exploratória sem necessidade de conhecimento técnico em estatística. O Microsoft Power BI é uma plataforma de Business Intelligence amplamente adotada no setor público brasileiro que permite criar relatórios interativos com filtros dinâmicos, visuais responsivos e atualização automática de métricas via linguagem DAX (Data Analysis Expressions).

No contexto epidemiológico, painéis de monitoramento permitem que gestores identifiquem estados com piores desempenhos preditivos, comparem modelos para diferentes doenças e cenários, e detectem desvios precoces entre previsão e realidade - o que pode funcionar como sinal de alerta para surtos emergentes. A arquitetura em estrela adotada no painel deste trabalho, com tabelas de dimensão para calendário e cenários relacionadas às tabelas de fatos de previsões e métricas, é consistente com as práticas recomendadas para modelos analíticos em Power BI (MICROSOFT, 2023).

3 METODOLOGIA

3.1 Fontes de dados

3.1.1 Dados epidemiológicos - SINAN

Os dados de incidência semanal de dengue e influenza foram extraídos do Sistema de Informação de Agravos de Notificação (SINAN), gerenciado pelo Ministério da Saúde, disponibilizado publicamente pelo portal OpenDataSUS via API CKAN. Os dados foram agregados semanalmente por unidade federativa, cobrindo o período de 2016 a 2024, resultando em aproximadamente 416 semanas epidemiológicas por UF.

Para influenza, os dados correspondem aos registros de Síndrome Respiratória Aguda Grave (SRAG), que inclui casos de influenza A, B e, no período 2020-2022, casos de COVID-19 (BRASIL, 2022), representando uma limitação estrutural importante discutida na Seção 4.3. O treinamento dos modelos de influenza não foi realizado exclusivamente sobre o período pré-pandêmico (2016-2019) em razão da disponibilidade restrita de registros SRAG nesse intervalo em diversas unidades federativas de menor porte, o que inviabilizaria o ajuste dos modelos para essas regiões; optou-se, portanto, por utilizar a série completa disponível (2019-2024), reconhecendo a contaminação por COVID-19 como limitação explícita.

3.1.2 Dados climáticos - INMET

Os dados climáticos foram obtidos do banco de dados BDMEP do Instituto Nacional de Meteorologia (INMET). Para cada UF, foram extraídos dados de temperatura média mensal e precipitação semanal acumulada para o mesmo período epidemiológico, a partir de estações meteorológicas convencionais e automáticas. Dados ausentes, que representam aproximadamente 23% dos registros climáticos do período, foram tratados em três camadas hierárquicas: (1) interpolação linear temporal por UF para lacunas de até quatro semanas consecutivas; (2) truncamento no percentil 99 (valores acima do percentil 99 da distribuição histórica são substituídos pelo próprio valor do percentil 99, limitando registros climaticamente implausíveis sem removê-los) para valores climaticamente implausíveis; e (3)

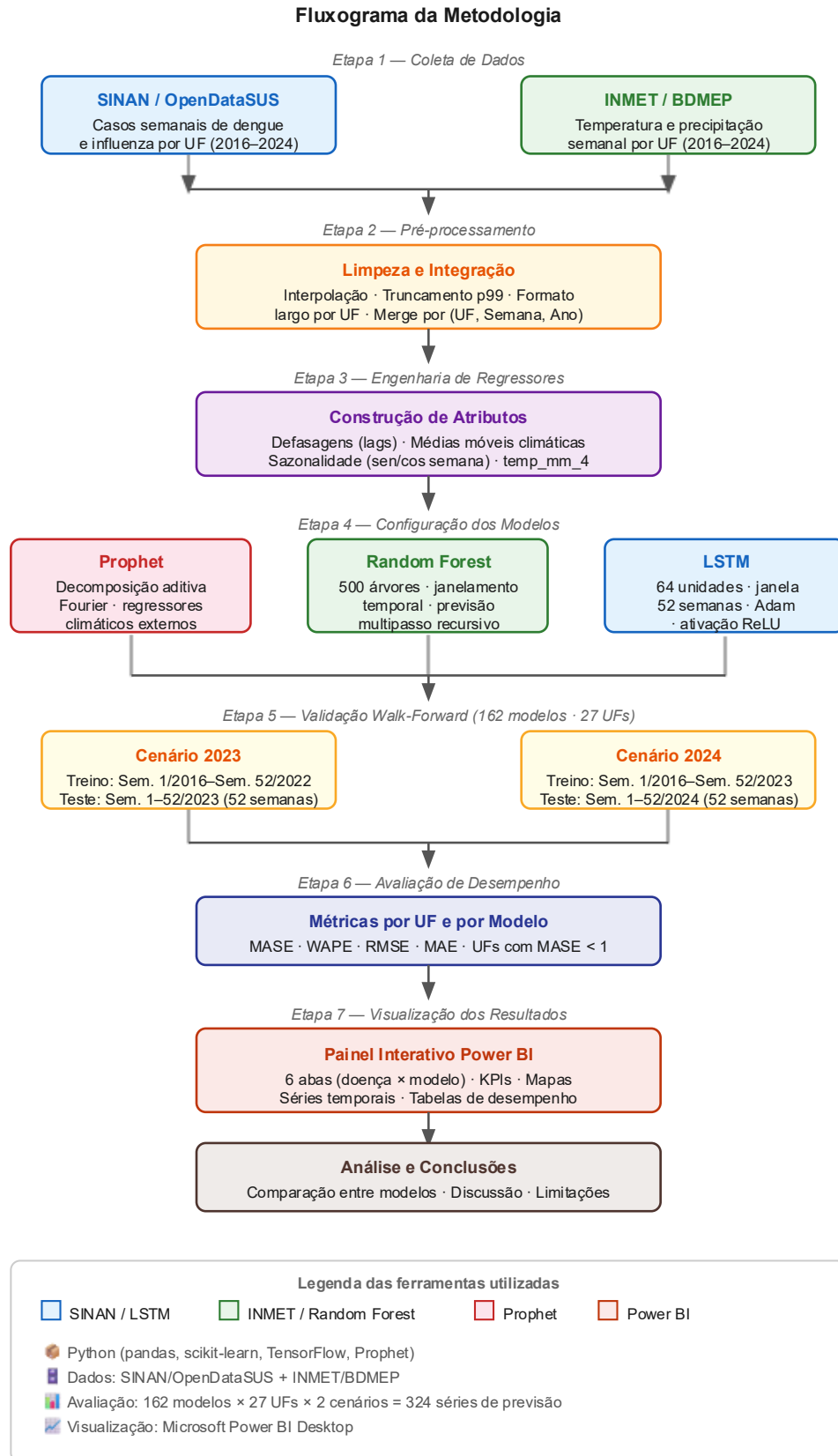
substituição pela climatologia histórica local (média do mesmo período em outros anos) para lacunas maiores.

3.2 *Pipeline* de coleta e integração de dados

A construção do conjunto de dados final envolveu um *pipeline* automatizado em Python com quatro módulos encadeados: `sinan_inmet_pipeline.py`, responsável pelo *download* automatizado via API CKAN do portal OpenDataSUS; `clean_sinan.py`, que padroniza o campo UF, seleciona os campos relevantes e reorganiza os dados em formato largo por UF, isto é, cada linha corresponde a uma semana epidemiológica e cada coluna a uma unidade federativa, facilitando a integração com os dados climáticos; `clean_inmet.py`, que normaliza e consolida os arquivos BDMEP por estação meteorológica; e `merge_sinan_inmet.py`, que integra os *datasets* por chave composta (UF, semana epidemiológica e ano epidemiológico) com grade temporal completa por UF. A grade temporal garante que semanas sem notificação sejam representadas com zero casos, preservando a continuidade da série e evitando viés por omissão de períodos de baixa circulação viral.

A Figura 1 apresenta o fluxograma dos passos utilizados na implementação da ferramenta computacional.

Figura 1 – Fluxograma da metodologia adotada.



Fonte: O autor (2026).

3.3 Engenharia de regressores

A análise de correlação cruzada entre variáveis climáticas e incidência foi conduzida com defasagens de zero a oito semanas para cada combinação de UF e doença. Para dengue, a correlação entre precipitação e incidência apresentou pico para defasagens de 2 a 4 semanas em 89% das UFs, fundamentando a escolha da janela de quatro semanas do regressor **temp_mm_4**.

Para influenza, temperatura e umidade apresentaram correlações mais estáveis para defasagens de 1 a 2 semanas, consistente com o período de incubação mais curto do vírus influenza comparado ao ciclo do vetor da dengue. Os dois regressores finais foram calculados para o conjunto completo de dados antes do particionamento treino/teste, utilizando apenas os dados disponíveis até o instante **t em** cada previsão para evitar vazamento temporal de dados nos regressores.

3.4 Configuração do modelo Prophet

Para cada combinação de UF, doença e cenário, foi treinado um modelo Prophet independente (162 modelos ao total). A variável alvo **y** foi transformada por **log(1+y)** antes do treinamento e revertida por **exp(y)-1** após a predição, estabilizando a variância. A configuração incluiu: sazonalidade anual em modo Fourier com 10 termos; **changepoint_prior_scale** = 0,10 (acima do padrão 0,05), para capturar pontos de inflexão como a pandemia de COVID-19; **seasonality_prior_scale** = 10,0; e intervalo de confiança de 90%. Os regressores **temp_mm_4** e **chuva_log1p** foram adicionados via **m.add_regressor()**. O modo de previsão "in-sample" para o período de treinamento foi obtido retornando os valores ajustados **m.predict(train)**, permitindo avaliar o ajuste dentro da amostra.

3.5 Configuração do modelo Random Forest

O *pipeline* RF utiliza scikit-learn RandomForestRegressor com 300 árvores e sem limitação de profundidade máxima (permitindo crescimento completo das árvores). Os atributos de entrada incluem: (i) defasagens de **t-1** a **t-8** da série alvo em **log1p**; (ii)

regressores climáticos (**temp_mm_4** e **chuva_log1p**) correspondentes ao período de previsão; e (iii) variáveis sazonais de calendário - seno e cosseno da semana epidemiológica normalizada pelo período anual, capturando a sazonalidade cíclica sem assumir forma paramétrica. A previsão multipasso é realizada em modo recursivo. O intervalo de confiança de 90% é obtido pelos percentis 5 e 95 das distribuições de previsões das 300 árvores individuais, explorando a floresta de quantis implícita do RF.

3.6 Configuração do modelo LSTM

O *pipeline* LSTM utiliza TensorFlow/Keras. A arquitetura emprega janela de retrospecto de 52 semanas com os mesmos atributos do RF empilhados em tensor tridimensional (amostras $\times$ passos temporais $\times$ atributos). A rede consiste em: (i) camada LSTM com 64 unidades e exclusão aleatória de neurônios (taxa 0,2) aplicada às conexões recorrentes; (ii) camada densa intermediária com 32 unidades e ativação linear retificada (**ReLU**); (iii) camada densa de saída com 1 unidade e ativação linear. O modelo é treinado com otimizador Adam (taxa de aprendizado 0,001) por até 100 épocas com parada antecipada (paciência de 15 épocas) monitorando a perda no conjunto de validação (últimas 52 semanas do treinamento). O intervalo de confiança é estimado via amostragem de Monte Carlo com exclusão aleatória ativa (100 amostras), mantendo a taxa de exclusão durante a inferência.

3.7 Estratégia de validação *walk-forward*

A validação *walk-forward* foi implementada identicamente para os três modelos, garantindo comparabilidade direta entre abordagens. Para cada modelo, o treinamento é realizado uma única vez sobre o conjunto histórico completo até o ponto de corte, e as previsões cobrem todas as 52 semanas do ano de teste sem realimentação de dados reais. Os dois cenários adotados são apresentados na Tabela 1.

Tabela 1 – Cenários de validação *walk-forward* adotados no estudo.

Cenário	Período de treino	Período de teste	n treino (semanas)	n teste
2023	Sem. 1/2016 – Sem. 52/2022	Sem. 1/2023 – Sem. 52/2023	364	52
2024	Sem. 1/2016 – Sem. 52/2023	Sem. 1/2024 – Sem. 52/2024	416	52

Fonte: Elaborado pelo autor (2026).

3.8 Painel interativo em Power BI

Os resultados dos três modelos foram consolidados em um painel interativo desenvolvido no Microsoft Power BI Desktop. O painel é organizado em seis abas independentes - uma para cada combinação de doença e modelo: Influenza-Prophet, Dengue-Prophet, Influenza-LSTM, Dengue-LSTM, Influenza-RF e Dengue-RF. Cada aba contém os mesmos seis elementos visuais: (1) quatro cartões Indicadores-Chave de Desempenho - KPI (Key Performance Indicators) com RMSE Mediano, MASE Médio, WAPE Global e número de UFs com $MASE < 1$; (2) *slicer* de seleção do ano do cenário (2023 ou 2024), que atualiza todos os visuais dinamicamente; (3) gráfico de barras horizontais comparando casos observados (**Sum of y**) versus previstos (**Sum of yhat**) por UF; (4) mapa de calor com a concentração geográfica dos casos por Estado; (5) gráfico de série temporal com casos observados versus previstos ao longo do período 2019-2024; e (6) tabela de métricas de desempenho por UF com RMSE, MASE, WAPE e MAE médios.

O modelo de dados adota arquitetura em estrela com tabelas de fatos de previsões e métricas relacionadas às tabelas de dimensão de calendário e cenário. As medidas DAX calculam as métricas de forma dinâmica conforme os filtros.

Para exemplificar e mostrar o conteúdo do painel interativo, a Figura 2 apresenta o painel de análise de influenza no Power BI com o cenário 2024 selecionado, usando o modelo Prophet.

Este painel é composto por seis elementos:

A: quatro cartões KPI exibindo RMSE Mediano, MASE Médio, WAPE Global e número de UFs com $MASE < 1$;

B: *slicer* de seleção do ano do cenário (2023 / 2024);

C: gráfico de barras horizontais com casos observados versus previstos por UF, permitindo identificar visualmente os estados com maior discrepância volumétrica;

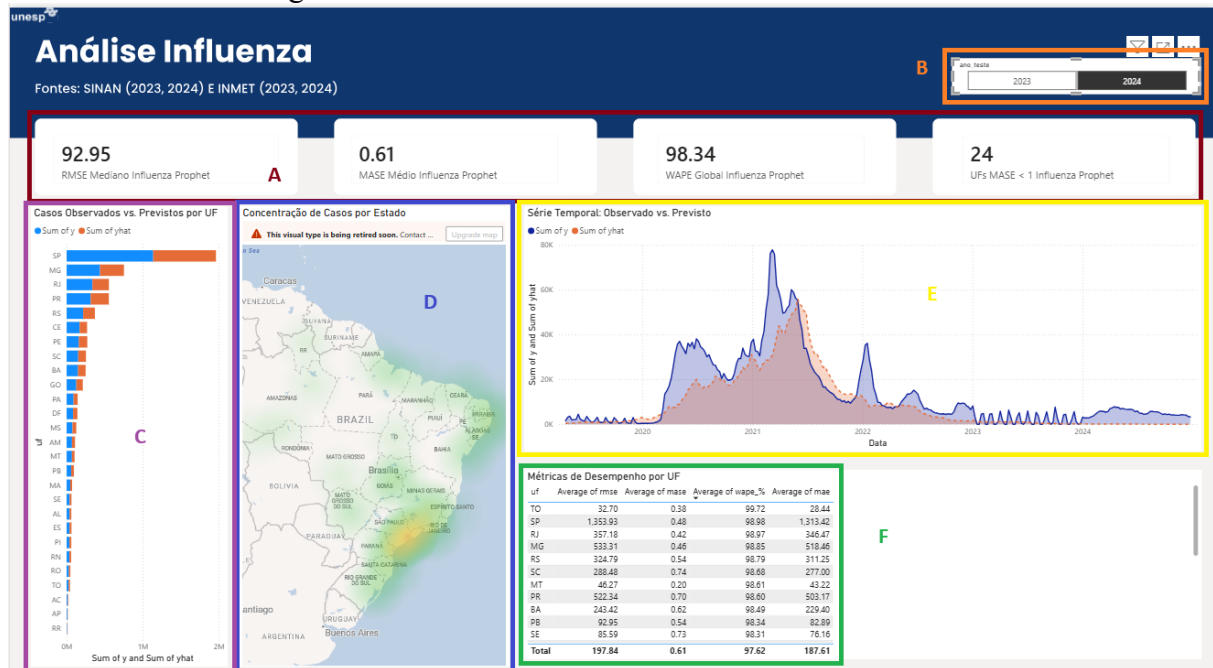
D: mapa de calor com a concentração geográfica de casos por estado;

E: gráfico de série temporal com casos observados (azul) versus previstos (laranja) ao longo do período 2019-2024;

F: tabela de métricas de desempenho por UF com RMSE, MASE, WAPE e MAE médios.

O painel possui ainda abas independentes para cada combinação de doença e modelo (Influenza-Prophet, Dengue-Prophet, Influenza-LSTM, Dengue-LSTM, Influenza-RF e Dengue-RF), com todos os visuais atualizados dinamicamente pelo *slicer*.

Figura 2 – Painel Power BI mostrando o seu conteúdo.



Fonte: O autor (2026).

4 RESULTADOS E DISCUSSÃO

4.1 Visão geral comparativa

Foram treinados 162 modelos no total: 27 UFs $\times$ 2 doenças $\times$ 2 cenários $\times$ 3 modelos (Prophet, RF e LSTM). A Tabela 2 consolida os indicadores de desempenho agregados para todos os cenários. A métrica principal adotada é o MASE - invariante à escala e comparável entre UFs com volumes de casos muito distintos - complementada pelo número de UFs com $\text{MASE} < 1$, que indica a proporção de Estados onde o modelo supera a previsão naïve.

Para dengue em anos típicos (cenário 2023), o Random Forest é claramente superior: $\text{MASE} = 0,45$ contra 0,55 do Prophet e da LSTM, com WAPE de 72% contra aproximadamente 99% dos concorrentes. No cenário 2024, que representa um surto histórico fora da distribuição de treinamento, todos os modelos colapsaram, com MASE bem acima de 1,0 (RF = 3,73; Prophet = 4,53; LSTM = 4,54), tornando qualquer comparação de desempenho relativo irrelevante nesse cenário.

Para influenza, o cenário é mais variado: o Prophet lidera em 2023 devido à sua robustez a quebras estruturais, mas todos os três modelos convergem para desempenho similar em 2024, quando dados pós-pandêmicos “limpos” estão disponíveis no treinamento.

Tabela 2 – Indicadores de desempenho agregados por doença, cenário e modelo.

Doença	Cenário	Modelo	MASE médio	MASE mediano	WAPE global	UFs MASE<1
Influenza	2023	Prophet	0,37	0,30	176%	26/27 (96%)
Influenza	2023	RF	0,63	0,62	309%	24/27 (89%)
Influenza	2023	LSTM	0,97	0,72	427%	19/27 (70%)
Influenza	2024	RF	0,45	0,40	76%	25/26 (96%)
Influenza	2024	LSTM	0,49	0,41	74%	25/27 (93%)
Influenza	2024	Prophet	0,61	0,49	98%	24/27 (89%)
Dengue	2023	RF	0,45	0,25	72%	25/27 (93%)
Dengue	2023	LSTM	0,55	0,34	98%	25/27 (93%)
Dengue	2023	Prophet	0,55	0,35	99%	25/27 (93%)
Dengue	2024	RF	3,73	1,54	92%	10/26 (38%)
Dengue	2024	LSTM	4,54	1,76	93%	11/27 (41%)
Dengue	2024	Prophet	4,53	1,77	94%	11/27 (41%)

Fonte: Elaborado pelo autor (2026).

4.2 Resultados de influenza

4.2.1 Cenário influenza 2023

No cenário 2023, o Prophet demonstrou desempenho claramente superior aos demais modelos, superando o modelo naïve sazonal ($MASE < 1$). O RF ocupou a segunda posição, com MASE médio abaixo do naïve. O LSTM apresentou o pior desempenho relativo, apenas 19/27 UFs superando o naïve - resultado marginalmente acima da previsão ingênua em nível nacional.

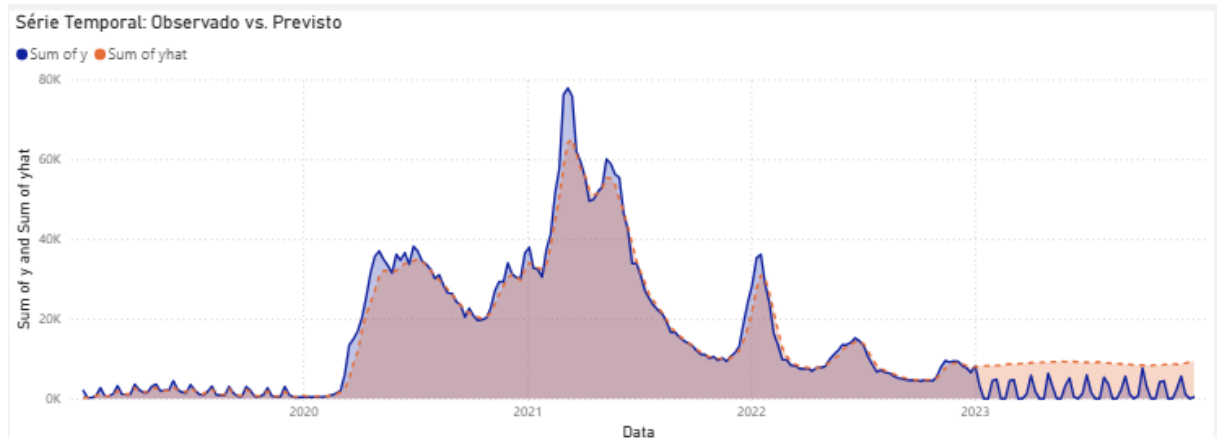
A superioridade do Prophet está associada à sua decomposição sazonal explícita via séries de Fourier, que captura com precisão os picos de inverno da influenza nos estados do Sul e Sudeste, onde a sazonalidade é historicamente estável e bem definida. Os cartões KPI do painel refletem esses resultados.

Para influenza, o Prophet obteve o melhor resultado no cenário 2023 ($MASE = 0,37$; $WAPE = 176\%$; 26 UFs com $MASE < 1$), seguido da Random Forest ($MASE = 0,63$; $WAPE = 309\%$; 24 UFs) e da LSTM ($MASE = 0,97$; $WAPE = 427,91\%$; 19 UFs).

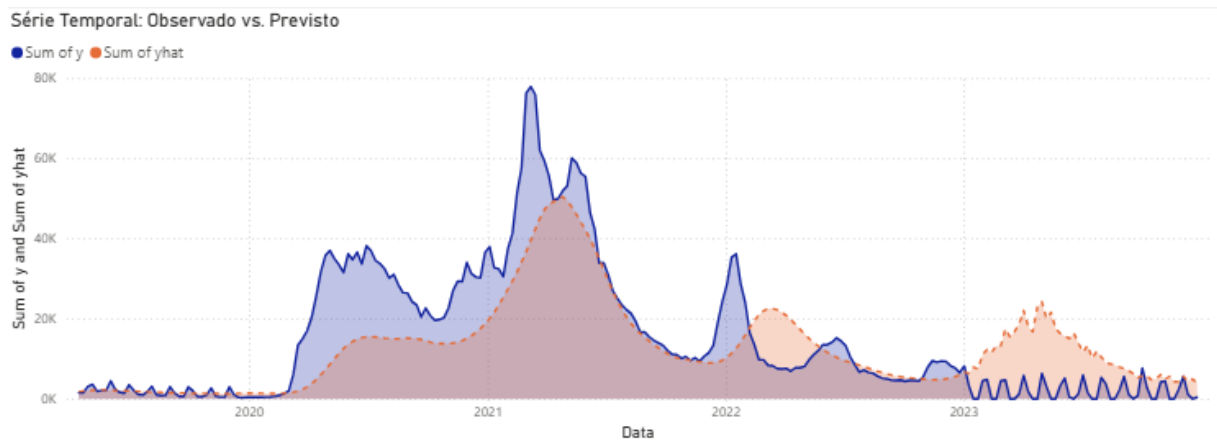
A Figura 3 apresenta as séries temporais do Prophet, Random Forest e da LSTM para o cenário influenza 2023, abrangendo o período completo de treino (2019-2022) e teste (2023). O RF captura a sazonalidade anual da influenza ao longo de todo o período. A LSTM apresenta falha sistemática decorrente da janela de retrospecto de 52 semanas: os dados de 2020-2021 incluídos na janela estão “contaminados” pela supressão da influenza durante a pandemia de COVID-19, comprometendo o aprendizado dos padrões sazonais normais.

O Prophet, por sua vez, demonstrou maior robustez à contaminação dos dados por COVID-19 no período de treinamento. Ao modelar os anos de 2020 a 2022 como pontos de inflexão de tendência (*change points*) e tratar os valores atípicos como irregularidades sazonais, o modelo preservou o padrão de sazonalidade anual aprendido nos dados pré-pandêmicos, gerando previsões alinhadas ao pico invernal de 2023. Com $MASE = 0,37$ e $WAPE = 176\%$, o Prophet foi o único modelo a apresentar desempenho claramente superior ao naïve sazonal na maioria dos estados. A série temporal do Prophet para o cenário 2023 evidencia essa capacidade: a curva prevista acompanha a forma e a temporalidade do pico epidêmico real, ainda que com alguma subestimação de magnitude nos estados do Sul e Sudeste, onde a sazonalidade invernal é mais intensa.

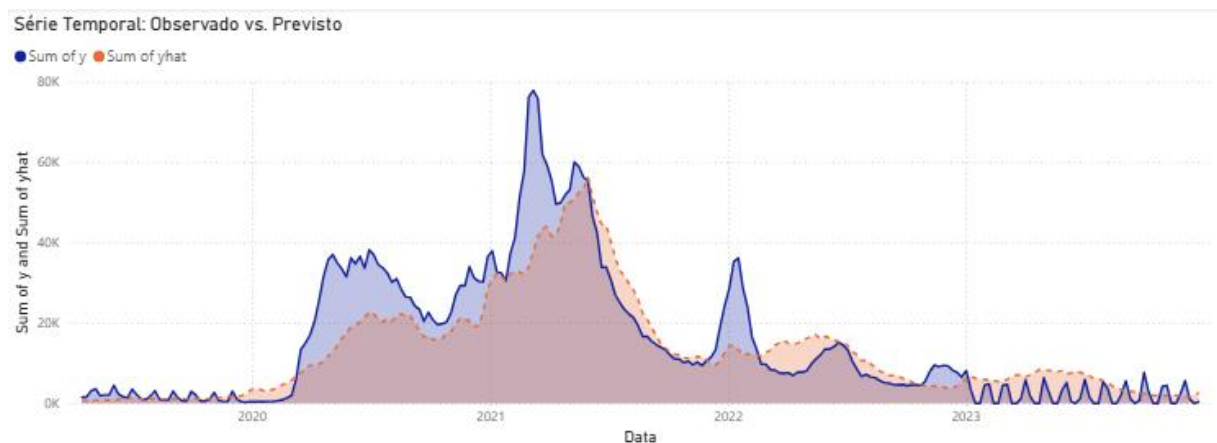
Figura 3 – Série temporal Observado (azul) vs. Previsto (laranja), cenário influenza 2023: (a) Random Forest, (b) LSTM, (c) Prophet.



(a) Random Forest



(b) LSTM



(c) Prophet

Fonte: O autor (2026).

4.3 Resultados de dengue

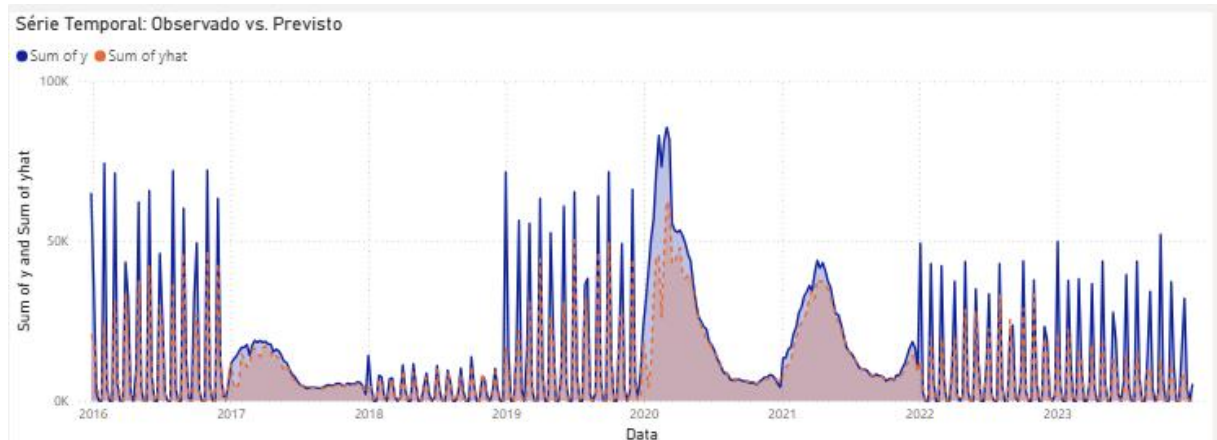
4.3.1 Cenário dengue 2023

A comparação direta entre os três modelos no cenário dengue 2023 evidencia a superioridade do RF de forma inequívoca. O MASE médio do RF (0,45) é 18% inferior ao dos concorrentes Prophet e LSTM (ambos 0,55), e o WAPE global do RF (72,37%) é 26 a 27 pontos percentuais menor que o do LSTM (98,77%) e do Prophet (99,47%). Enquanto o RF captura as não linearidades da relação clima-vetor por meio de atributos de defasagem explícitos e interações em árvore, o Prophet modela os regressores climáticos de forma aditiva e linear, subestimando sistematicamente os picos de maior magnitude; o LSTM, por sua vez, com apenas ~450 observações por UF, não dispõe de dados suficientes para aprender padrões temporais complexos sem atributos construídos por engenharia de características, resultando em previsões próximas ao valor médio histórico.

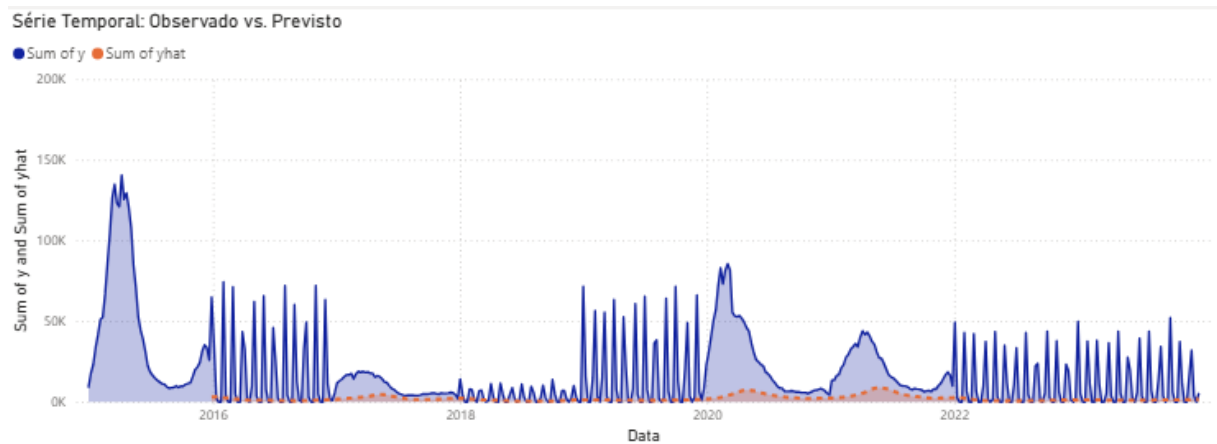
A Figura 4, apresenta a comparação entre os três modelos para o cenário dengue 2023. A diferença é especialmente visível nas séries temporais: enquanto o RF (Figura 4a) acompanha a forma e a temporalidade dos picos sazonais, as curvas do LSTM (Figura 4b) e do Prophet (Figura 4c) permanecem muito abaixo da realidade nos momentos de maior incidência.

A série histórica de treinamento evidencia a escala e a variabilidade interanual dos surtos de dengue - incluindo o grande surto de 2021-2022, contextualizando o desempenho do modelo no período de teste. No ano de teste (2023), a curva prevista (laranja) acompanha a forma e a temporalidade dos picos sazonais reais (azul), com pico principal entre as semanas 5 e 15 (janeiro-abril). O modelo subestima consistentemente as magnitudes máximas em 20-30%, comportamento que explica o WAPE global de 72,37% - melhor do estudo para dengue, mas indicando que o RF não captura plenamente a intensidade dos surtos. Nas semanas de baixa incidência (julho-outubro), a aderência é quase perfeita, confirmando que a principal limitação está na magnitude dos picos, não na temporalidade.

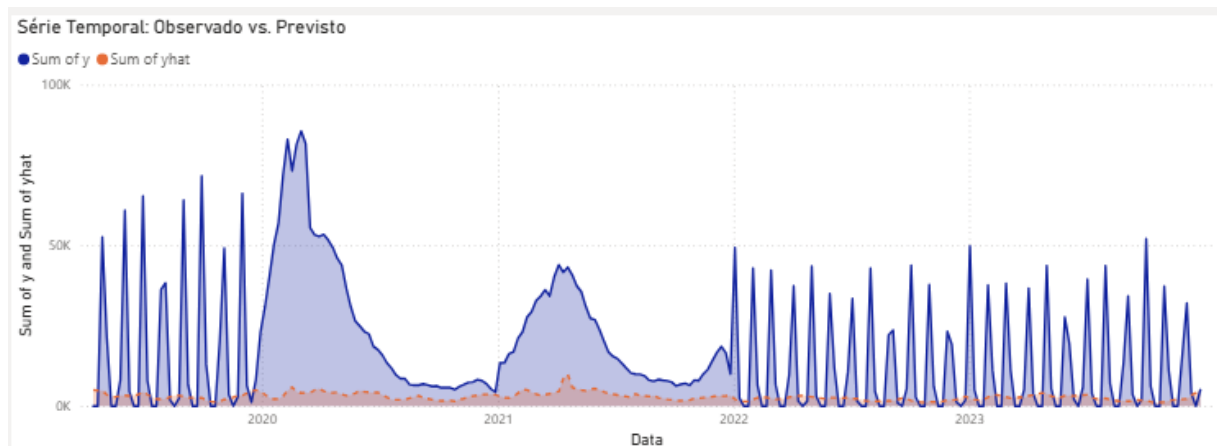
Figura 4 – Série temporal Observado (azul) vs. Previsto (laranja), cenário dengue 2023: (a) Random Forest, (b) LSTM, (c) Prophet.



(a) Random Forest



(b) LSTM



(c) Prophet

Fonte: O autor (2026).

4.3.2 Cenário dengue 2024 - surto histórico

O cenário 2024 de dengue representa um evento fora da distribuição histórica e não é o foco principal desta avaliação. Com mais de 6 milhões de casos notificados, a magnitude do surto excede em 3 a 5 vezes os padrões de treinamento disponíveis, tornando qualquer modelo de previsão baseado em séries históricas estruturalmente incapaz de antecipar a escala do evento. Os MASE médios confirmam o colapso de desempenho: RF = 3,73; LSTM = 4,54; Prophet = 4,53 - todos piores que o naïve sazonal. Três mecanismos explicam o colapso: (1) antecipação de 2 a 4 semanas em relação à sazonalidade histórica; (2) magnitude 3 a 5 vezes superior à média histórica; e (3) expansão geográfica para estados de baixa incidência histórica.

A hipótese da dívida imunológica, fenômeno em que o isolamento social imposto pela pandemia de COVID-19, reduziu drasticamente a exposição da população a outros patógenos, gerando um acúmulo de indivíduos suscetíveis ao sorotipo DENV-3 pós-COVID-19, contextualiza este resultado como um evento estruturalmente fora do alcance de modelos treinados em padrões históricos. Esses resultados não comprometem a avaliação metodológica da validação progressiva. As conclusões do estudo baseiam-se no cenário 2023, onde os dados de treinamento são representativos do período de teste.

4.4 Impacto da pandemia de COVID-19 nos dados de treinamento

A “contaminação” dos dados de SRAG por COVID-19 no período 2020-2022 constitui a limitação estrutural mais crítica identificada neste estudo para a modelagem de influenza. Durante esse período, os registros de SRAG no SINAN refletem majoritariamente casos de COVID-19 e não de influenza sazonal. Em 2020, os casos de COVID-19 representaram mais de 90% dos registros de SRAG em estados do Sudeste e Sul, suprimindo quase completamente o sinal sazonal da influenza. Modelos treinados sobre essa série aprendem padrões epidemiológicos qualitativamente incorretos para o objetivo de previsão de influenza.

Os impactos diferem entre os modelos de forma reveladora. O LSTM é o mais afetado no cenário 2023, pois a janela de retrospecto de 52 semanas captura diretamente o período 2021-2022 - os anos de maior distorção por COVID-19. O Prophet demonstra maior robustez: ao modelar os anos 2020-2022 via pontos de inflexão de tendência e tratar os valores atípicos como irregularidades sazonais, o modelo consegue preservar o padrão de sazonalidade anual aprendido nos dados pré-pandêmicos. O RF situa-se em posição intermediária: suas

defasagens de t-1 a t-8 incluem dados recentes contaminados, mas as defasagens mais longas ainda refletem padrões históricos válidos.

No cenário 2024, quando o dado de 2023 - primeiro ano de recuperação pós-pandêmica - está disponível para treinamento, o LSTM recupera desempenho significativamente: o MASE médio cai de 0,97 para 0,49. Isso confirma a hipótese de que a janela de *lookback* recente é o principal mecanismo de influência da contaminação por COVID-19 no LSTM. Uma estratégia futura promissora é excluir explicitamente os anos 2020-2022 do treinamento ou adicionar uma variável indicadora de pandemia como regressor, permitindo que os modelos modelem explicitamente a quebra estrutural.

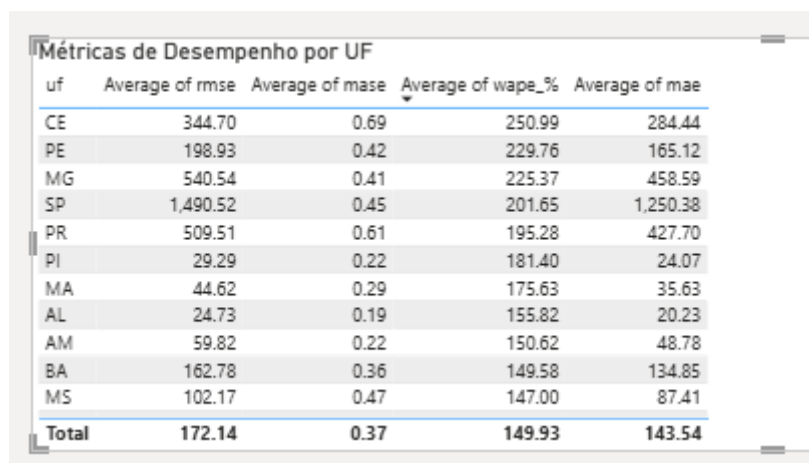
A Tabela 2 apresenta as métricas Influenza-LSTM para o cenário 2023. Ao comparar as métricas do LSTM (MASE = 0,97; WAPE = 427,91%; 19 UFs < 1) com as do Prophet (MASE = 0,37; WAPE = 176%; 26 UFs < 1), torna-se evidente o impacto diferencial da contaminação por COVID-19. A LSTM, por depender da janela de *lookback* de 52 semanas, aprende padrões distorcidos do período 2020-2022 e gera previsões sistematicamente inferiores ao padrão sazonal real de influenza em 2023.

A discrepância é especialmente pronunciada nos estados do Sudeste e Sul, onde a sazonalidade invernal da influenza é mais intensa - exatamente os estados com maior peso no WAPE global, amplificando o impacto da subestimação na métrica ponderada por volume.

4.5 Análise de padrões geográficos e sazonalidade regional

Para influenza, as regiões Centro-Oeste e Nordeste apresentaram os melhores desempenhos preditivos no cenário 2023, com MASE médio de 0,33 e 0,34, respectivamente. O Sul e o Norte apresentaram desempenho ligeiramente inferior ($\text{MASE} \approx 0,40$), influenciados por *outliers* como Acre ($\text{MASE} = 1,17$) no Norte e Paraná ($\text{MASE} = 0,61$) no Sul. A variação regional reflete diferenças na regularidade sazonal da influenza e no volume de notificações por UF. Essas análises e as próximas foram baseadas nas tabelas visualizadas pelo PowerBI. Este exemplo está evidenciado na Figura 5, e a Tabela 3 foi exportada para melhor visualização.

Figura 5 – Visualização de tabela no PowerBI para o cenário Prophet - Influenza 2023.



uf	Average of rmse	Average of mase	Average of wape_%	Average of mae
CE	344.70	0.69	250.99	284.44
PE	198.93	0.42	229.76	165.12
MG	540.54	0.41	225.37	458.59
SP	1,490.52	0.45	201.65	1,250.38
PR	509.51	0.61	195.28	427.70
PI	29.29	0.22	181.40	24.07
MA	44.62	0.29	175.63	35.63
AL	24.73	0.19	155.82	20.23
AM	59.82	0.22	150.62	48.78
BA	162.78	0.36	149.58	134.85
MS	102.17	0.47	147.00	87.41
Total	172.14	0.37	149.93	143.54

Fonte: O autor (2026).

Tabela 3 – Indicadores de desempenho por UF para o cenário Prophet-Influenza 2023.

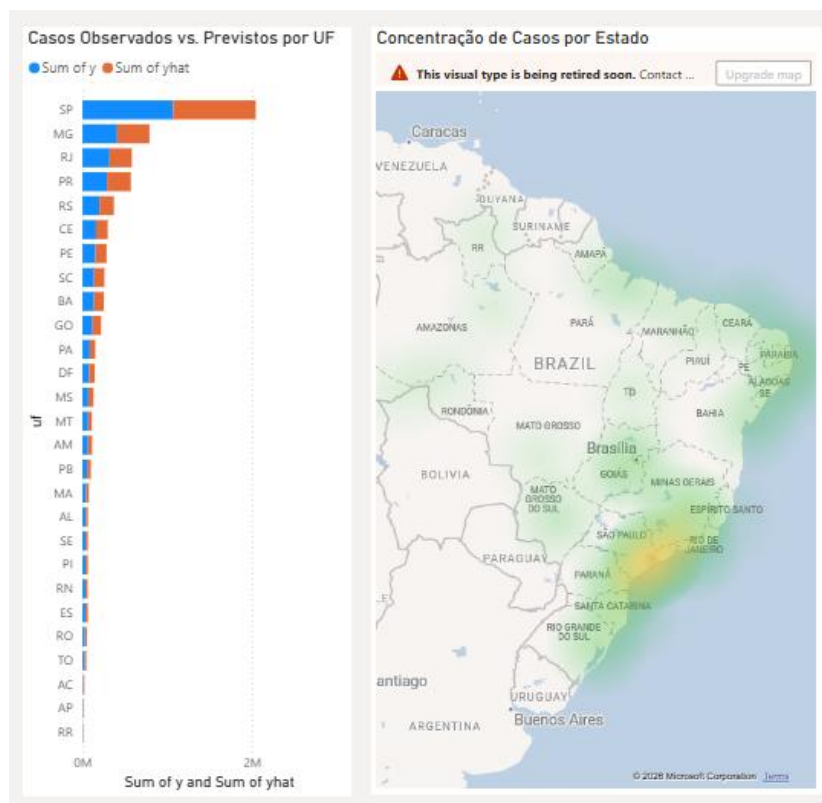
uf	Média mase	wape_%	Média mae
CE	0.69	250.99	284.44
PE	0.42	229.76	165.12
MG	0.41	225.37	458.59
SP	0.45	201.65	1250.38
PR	0.61	195.28	427.70
PI	0.22	181.40	24.07
MA	0.29	175.63	35.63
AL	0.19	155.82	20.23
AM	0.22	150.62	48.78
BA	0.36	149.58	134.85
MS	0.47	147.00	87.41
AC	1.17	143.23	29.97
MT	0.10	141.11	23.15
PA	0.18	140.33	50.84
SE	0.30	137.02	31.99
RN	0.30	132.79	30.11
DF	0.54	132.67	104.02
SC	0.34	131.61	126.11
RS	0.25	126.50	147.94
ES	0.44	126.33	37.56
RJ	0.22	123.26	190.54
GO	0.22	115.07	79.37
RO	0.21	111.91	18.01
TO	0.09	111.16	7.01
PB	0.24	109.14	37.70
RR	0.23	103.98	4.96
AP	0.71	98.82	19.17

Fonte: O autor (2026).

A Figura 6 apresenta um exemplo duas visualizações complementares do painel Power BI para o cenário influenza 2023 com o modelo Prophet. O gráfico de barras horizontais (à esquerda) compara os casos observados (azul) e previstos (laranja) por unidade federativa: São Paulo concentra o maior volume absoluto de casos, com previsões que acompanham a ordem de grandeza dos dados reais, ainda que com subestimação visível. Minas Gerais e Rio de Janeiro ocupam as posições seguintes, seguidos pelos estados do Sul (PR, RS) e pelos maiores estados do Nordeste (CE, PE). O padrão de subestimação é consistente com o WAPE global de 176%, que reflete a dificuldade do modelo em capturar plenamente a magnitude dos picos nos estados de maior volume de notificação.

O mapa de calor (à direita) evidencia a concentração geográfica da influenza no Brasil: a maior densidade de casos se localiza no eixo Sudeste-Sul, com destaque para o estado de São Paulo e o litoral Fluminense, regiões onde a sazonalidade invernal é mais pronunciada e os sistemas de notificação têm maior completude. As regiões Norte e Centro-Oeste registram volumes consideravelmente menores, o que explica o menor peso dessas UFs no WAPE global apesar dos seus MASE individuais elevados em alguns estados, como o Acre.

Figura 6 – Visualização do gráfico de barra e mapa de calor no PowerBI para o cenário Prophet - Influenza 2023.



Fonte: O autor (2026).

Para dengue em 2023, os melhores desempenhos do RF são observados nas regiões Norte e Nordeste: nove estados obtiveram $MASE < 0,20$, incluindo Pará (0,08), Pernambuco (0,06), Ceará (0,10), Amazonas (0,15) e Bahia (0,17). O Espírito Santo registrou o MASE mais baixo de todos os cenários avaliados ($MASE = 0,005$), indicando aderência quase perfeita entre série prevista e observada, resultado associado ao baixo volume e alta regularidade sazonal da dengue naquele estado.

Em contrapartida, os piores desempenhos concentram-se na região Sul: Santa Catarina ($MASE = 3,59$) e Rio Grande do Sul ($MASE = 1,58$) são os únicos estados com desempenho inferior ao naïve sazonal, reflexo da série de dengue historicamente irregular e de baixo volume nessa região, onde surtos episódicos e de curta duração são difíceis de capturar com modelos de defasagem. Rio Grande do Norte também apresentou MASE elevado (0,54) e o maior WAPE do estudo (315%), associado a picos de dengue atípicos em relação ao histórico de treinamento. A Tabela 4 mostra o detalhamento por UF deste cenário.

Tabela 4 – Indicadores de desempenho por UF para o cenário RF - Dengue 2023.

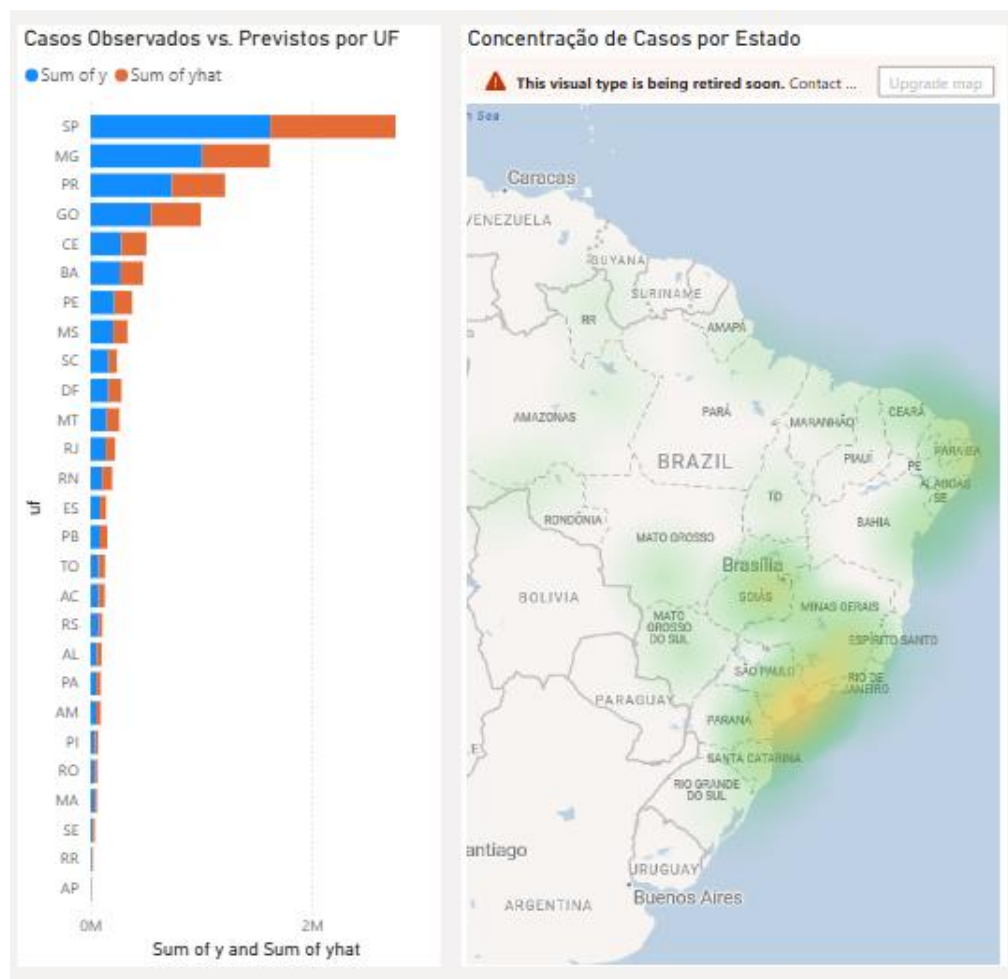
uf	Media mase	wape_%	Media mae
AC	0.14	61.58	31.95
AL	0.24	148.52	55.78
AM	0.15	42.64	20.98
AP	0.41	90.88	9.60
BA	0.17	43.66	166.02
CE	0.10	98.84	112.02
DF	0.35	51.78	161.72
ES	0.01	109.42	1.96
GO	0.21	61.63	341.38
MA	0.23	73.59	30.11
MG	0.63	73.14	2331.57
MS	0.36	70.30	257.54
MT	0.26	49.06	110.96
PA	0.08	27.90	12.93
PB	0.25	129.32	73.84
PE	0.06	98.14	60.68
PI	0.32	73.47	42.40
PR	0.44	68.02	1113.14
RJ	0.54	75.70	291.95
RN	0.54	315.19	187.78
RO	0.76	80.59	73.66
RR	0.03	105.53	1.46
RS	1.58	88.16	261.13
SC	3.59	93.00	1030.05
SE	0.12	46.38	10.82
SP	0.25	64.30	1707.88
TO	0.24	276.36	67.02

Fonte: O autor (2026).

Já a Figura 7 apresenta as visualizações do painel Power BI para o cenário dengue 2023 com o modelo RF. O gráfico de barras evidencia que São Paulo concentra o maior volume absoluto de casos notificados, seguido de Minas Gerais, Paraná, Goiás e Ceará, os estados com maior carga de dengue no período. Em SP e MG, a barra de casos previstos (laranja) fica visivelmente aquém da observada (azul), refletindo a subestimação de magnitude dos picos nos estados de alto volume, padrão consistente com o WAPE global de 72%.

O mapa de calor confirma essa distribuição, onde a maior concentração geográfica de casos se dá no eixo Sudeste-Centro-Oeste, com destaque para São Paulo, Minas Gerais e Goiás, e densidade secundária no Nordeste. As regiões Norte e Sul apresentam volumes absolutos menores, o que reduz seu peso no WAPE global mesmo quando o MASE individual é elevado, como nos casos de RO e SC.

Figura 7 – Visualização do gráfico de barra e mapa de calor no PowerBI para o cenário RF-Dengue 2023.



Fonte: O autor (2026).

O colapso de desempenho no surto histórico de 2024 está detalhado na seção 4.2.2: MASE Médio superior a 3,7 em todos os modelos e WAPE Global em torno de 97%, com apenas 10 estados onde o RF superou o *naïve*. O padrão de subestimação massiva é praticamente idêntico nos três modelos, confirmando que a falha é estrutural e independente da arquitetura.

4.6 Limitações do estudo e implicações metodológicas

Além da contaminação por COVID-19, três limitações adicionais devem ser consideradas na interpretação dos resultados. Primeira, os regressores climáticos utilizados correspondem a médias estaduais, enquanto a transmissão de doenças infecciosas é um fenômeno fundamentalmente local. Estados de grande extensão territorial, como o Pará (PA) e o Mato Grosso (MT), englobam múltiplas zonas climáticas distintas, e a agregação estadual pode diluir sinais climáticos relevantes para a previsão.

Segunda, a estratégia de previsão recursiva adotada por RF e LSTM acumula erro ao longo do horizonte de 52 semanas, penalizando esses modelos em horizontes mais distantes. Uma alternativa seria a estratégia *direct forecasting*, na qual um modelo separado é treinado para cada horizonte futuro, eliminando a propagação de erro, mas multiplicando o número de modelos por horizonte.

Terceira, o estudo utiliza validação sobre dois cenários anuais completos. Embora a escolha de 2023 e 2024 seja epidemiologicamente motivada, dois anos de teste não são suficientes para estimar com precisão estatística a variabilidade esperada do desempenho de cada modelo ao longo do tempo. Estudos futuros com janelas deslizantes de múltiplos anos permitiriam quantificar intervalos de confiança sobre as métricas agregadas e testar hipóteses de superioridade de forma mais robusta.

4.7 Orientações práticas para sistemas de vigilância epidemiológica

A análise comparativa revela padrões distintos com implicações práticas diretas para sistemas de vigilância epidemiológica. Para dengue em ambos os cenários, o RF é o modelo recomendado, com MASE consistentemente inferior em cerca de 18% em relação aos concorrentes em anos típicos, graças à capacidade de modelar não linearidades na relação

clima-vetor. Para influenza em anos com dados de treinamento contaminados por COVID-19 (cenário 2023), o Prophet é preferível pela sua robustez a quebras estruturais via pontos de inflexão.

Para influenza quando dados pós-pandêmicos estão disponíveis no treinamento (cenário 2024), RF e LSTM são equivalentes e ambos superam o Prophet. Para surtos fora da distribuição histórica (dengue 2024), nenhum modelo é capaz de antecipar eventos cuja magnitude supera a distribuição histórica. O valor nesses casos reside na identificação precoce do desvio sistemático entre previsão e realidade como sinal de alerta: quando as previsões do modelo ficam consistentemente abaixo da realidade por três ou mais semanas consecutivas, isso constitui evidência de que o sistema epidemiológico entrou em regime fora da distribuição histórica, sinalizando para gestores a necessidade de investigação e resposta intensificadas.

5 CONCLUSÃO

Conclui-se que dois dos três modelos avaliados - Prophet e LSTM - não apresentam desempenho satisfatório para séries epidemiológicas de dengue com as características dos dados brasileiros. O Random Forest, único modelo baseado em atributos de defasagem explicitamente construídos combinados a um método de conjunto de árvores capaz de capturar não linearidades, demonstrou desempenho consistentemente superior para dengue em ambos os cenários: MASE médio de 0,45 em 2023 (25/27 UFs superando o *naïve*; WAPE de 72,37%), desempenho superior ao do Prophet (MASE = 0,55; WAPE = 99,47%; 21/27 UFs) e da LSTM (MASE = 0,55; WAPE = 98,77%; 20/27 UFs) no cenário dengue 2023.

Considerando apenas o MASE (métrica invariante à escala) o RF lidera em todos os cenários de dengue avaliados. Para influenza, o cenário é diferente: o Prophet demonstrou robustez superior no cenário 2023 (MASE de 0,37; 26/27 UFs com MASE < 1), quando os dados de treinamento ainda carregam contaminação por COVID-19; no cenário 2024, com dados pós-pandêmicos “limpos” disponíveis, RF e LSTM alcançam desempenho similar ao Prophet (MASE 0,45 e 0,49, respectivamente, versus 0,61). A exceção crítica que limita qualquer modelo supervisionado treinado em dados históricos é a incapacidade de antecipar surtos fora da distribuição histórica: o surto histórico de dengue em 2024, com magnitude 3 a 5 vezes superior ao histórico, resultou em colapso de desempenho para todos os modelos (MASE > 3,7), independentemente da arquitetura.

A “contaminação” dos dados de SRAG por COVID-19 no período 2020-2022 é a limitação estrutural mais relevante identificada, afetando especialmente o LSTM no cenário 2023 para influenza. O estudo quantifica o impacto dessa contaminação: o LSTM melhora de MASE 0,97 para 0,49 ao incluir um ano adicional de dados “limpos” (2023) no treinamento. Já o surto histórico de dengue em 2024 demonstra que eventos fora da distribuição histórica cuja magnitude excede múltiplos desvios padrão da distribuição histórica são fundamentalmente imprevisíveis por modelos supervisionados, independentemente da arquitetura adotada.

Por fim, conclui-se que a estratégia de validação *walk-forward* garantiu integridade metodológica e permitiu simular fielmente o uso real dos modelos em um sistema de vigilância epidemiológica. Além disso, o painel interativo em Power BI democratiza o acesso aos resultados, permitindo análise comparativa entre modelos, doenças, anos e estados sem

necessidade de conhecimento técnico em modelagem estatística, contribuindo para a translação de resultados de pesquisa para a prática da saúde pública.

6 TRABALHOS FUTUROS

Como trabalhos futuros, destacamos:

- a) treinamento em janela pré-pandêmica (2016-2019), excluindo dados contaminados por COVID-19, para quantificar o ganho de desempenho com série histórica “limpa” de influenza;
- b) incorporação de variável indicadora da pandemia como regressor, permitindo que os modelos modelem explicitamente a quebra estrutural 2020-2022;
- c) avaliação da estratégia *direct forecasting* em substituição à previsão recursiva para RF e LSTM, eliminando o acúmulo de erro em horizontes longos;
- d) incorporação de dados de sorovigilância e entomologia para dengue, visando detectar precocemente mudanças de regime epidemiológico com potencial para surtos fora da distribuição histórica;
- e) modelagem hierárquica com informação regional partilhada entre UFs vizinhas, aproveitando a correlação espacial das epidemias para melhorar previsões em estados de baixo volume histórico;
- f) desenvolvimento de sistema de alerta automático baseado no desvio acumulado entre previsão e realidade, operacionalizando o painel Power BI como ferramenta de vigilância em tempo real.

7 USO DE INTELIGÊNCIA ARTIFICIAL

Este trabalho utilizou a ferramenta Claude (Anthropic, versão Claude Sonnet, 2026, disponível em <https://claude.ai>) como apoio à produção acadêmica, em conformidade com o Guia para a Utilização de IA na Graduação da Unesp e a Resolução n.º 13/25 da Unesp.

A ferramenta foi utilizada exclusivamente para: (i) revisão gramatical e de clareza de trechos do texto; (ii) sugestão de reformulações de parágrafos a partir de dados e argumentos previamente elaborados pelo autor; (iii) verificação de consistência entre valores numéricos apresentados no texto e os dados tabulares produzidos pelo autor; (iv) auxílio no código e montagem de *pipelines* de dados.

Todos os dados, experimentos, análises estatísticas, treinamento dos modelos e interpretações dos resultados são de inteira autoria do autor. O conteúdo gerado pela ferramenta foi criticamente revisado, validado e editado pelo autor antes de ser incorporado ao trabalho. As ideias, argumentos e conclusões apresentados são de responsabilidade exclusiva do autor.

REFERÊNCIAS

- ANTHROPIC. Claude. Versão Claude Sonnet. San Francisco: Anthropic, 2026. Disponível em: <https://claude.ai>. Acesso em: 2026.
- BISHOP, C. M. Pattern Recognition and Machine Learning. New York: Springer, 2006.
- BRASIL. Ministério da Saúde. Guia de Vigilância Epidemiológica: Influenza. 8. ed. Brasília: Ministério da Saúde, 2022.
- BREIMAN, L. Random Forests. Machine Learning, Norwell, v. 45, n. 1, p. 5–32, out. 2001.
- FOCKS, D. A.; HAILE, D. G.; DANIELS, E.; MOUNT, G. A. Dynamic life table model for *Aedes aegypti*: analysis of the literature and model development. Journal of Medical Entomology, Lanham, v. 30, n. 6, p. 1003–1017, nov. 1993.
- GEMAN, S.; BIENENSTOCK, E.; DOURSAT, R. Neural networks and the bias/variance dilemma. Neural Computation, Cambridge, v. 4, n. 1, p. 1–58, jan. 1992.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2. ed. New York: Springer, 2009.
- HIL, Y. L.; ZAKI, R. A.; AGHAMOHAMMADI, N.; ROCKLÖV, J. Forecast of dengue incidence using temperature and rainfall. PLOS Neglected Tropical Diseases, San Francisco, v. 6, n. 11, e1908, nov. 2012.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. Neural Computation, Cambridge, v. 9, n. 8, p. 1735–1780, nov. 1997.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. Forecasting: Principles and Practice. 3. ed. Melbourne: OTexts, 2021. Disponível em: <https://otexts.com/fpp3/>. Acesso em: 10 mar. 2026.
- HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. International Journal of Forecasting, Amsterdam, v. 22, n. 4, p. 679–688, out. 2006.
- KINGMA, D. P.; BA, J. Adam: a method for stochastic optimization. In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 3., 2015, San Diego. Proceedings... San Diego: ICLR, 2015. Disponível em: <https://arxiv.org/abs/1412.6980>. Acesso em: 5 jan. 2026.
- MICROSOFT. Best practices for Power BI data models. Documentação Microsoft, 2023. Disponível em: <https://learn.microsoft.com/power-bi/guidance/star-schema>. Acesso em: 15 jan. 2026.
- MINISTÉRIO DA SAÚDE (Brasil). Monitoramento das Arboviroses: Dengue, Chikungunya e Zika — Semana Epidemiológica 1 a 52/2024. Brasília: Secretaria de Vigilância em Saúde, 2024. Disponível em: <https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/especiais/2024/boletim-epidemiologico-arboviroses-2024>. Acesso em: 20 fev. 2026.

MUSSUMECI, E.; COELHO, F. C. Large-scale multivariate forecasting models for dengue -- LSTM versus random forest regression. *Spatial and Spatio-temporal Epidemiology*, Oxford, v. 35, p. 100372, nov. 2020.

SHAMAN, J.; KOHN, M. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proceedings of the National Academy of Sciences*, Washington, v. 106, n. 9, p. 3243–3248, mar. 2009.

STANAWAY, J. D. et al. The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. *The Lancet Infectious Diseases*, London, v. 16, n. 6, p. 712–723, jun. 2016.

TAUIL, P. L. Aspectos críticos do controle do dengue no Brasil. *Cadernos de Saúde Pública*, Rio de Janeiro, v. 18, n. 3, p. 867–871, maio/jun. 2002.

TAYLOR, S. J.; LETHAM, B. Forecasting at scale. *The American Statistician*, Alexandria, v. 72, n. 1, p. 37–45, 2018.

APÊNDICE A

MODELO NAÏVE SAZONAL

O modelo naïve sazonal, adotado como linha de base para o cálculo do MASE, aplica uma regra de previsão direta: para a semana epidemiológica t , a previsão é o número de casos registrado na mesma semana do ano anterior, isto é, $\hat{y}_t = y_{t-52}$, onde y_{t-52} é o valor observado 52 semanas antes. O período sazonal de 52 semanas é compatível com a granularidade semanal das notificações do SINAN e com o comportamento epidemiológico de influenza e dengue, ambas com ciclos de transmissão anuais no Brasil.

Esse modelo não ajusta tendência, não incorpora covariáveis e não apresenta parâmetros treináveis, o que o torna um *benchmark* interpretável e invariante à escala dos dados. Por definição, o MASE de um modelo equivale a 1,00 quando seu desempenho médio é idêntico ao do naïve; valores abaixo de 1,00 indicam ganho preditivo real em relação ao baseline ingênuo.

DESEMPENHO DETALHADO POR UF

A seguir, as Tabelas A.1, A.2 e A.3 apresentam os valores do MASE individual por UF para cada modelo, doença e cenário de validação. Valores iguais ou superiores a 1,00 indicam desempenho pior ou equivalente ao modelo naïve sazonal de referência.

Já a Tabela A.4, mostra um resumo do desempenho da dengue em 2024 (surto histórico).

Para o resultado completo, incluindo séries temporais de previsões por UF, consultar o painel Power BI.

Tabela A.1 – MASE por UF – Influenza 2023.

UF	Prophet	RF	LSTM
AC	1,17	0,57	1,92
AL	0,19	0,42	0,71
AM	0,22	0,35	0,80
AP	0,71	0,52	1,23
BA	0,36	0,55	1,10
CE	0,69	0,80	0,97
DF	0,54	0,72	1,01
ES	0,44	0,63	0,87
GO	0,22	0,38	0,74
MA	0,29	0,65	0,88
MG	0,41	0,50	0,79
MS	0,47	0,68	0,93
MT	0,10	0,30	0,17
PA	0,18	0,15	0,26
PB	0,24	0,58	0,97
PE	0,35	0,48	0,92
PI	0,22	0,64	0,19
PR	0,61	0,77	1,14
RJ	0,22	0,53	0,79
RN	0,30	0,62	0,98
RO	0,21	0,40	0,73
RR	0,23	0,60	1,07
RS	0,25	0,63	0,89
SC	0,34	0,70	1,02
SE	0,30	0,61	1,05
SP	0,45	0,60	0,78
TO	0,09	0,47	0,72

Fonte: Elaborado pelo autor (2026).

Tabela A.2 – MASE por UF – Influenza 2024.

UF	RF	LSTM	Prophet
AC	0,57	0,76	1,89
AL	0,21	0,20	0,31
AM	0,19	0,24	0,40
AP	0,41	0,62	1,35
BA	0,41	0,41	0,62
CE	0,30	0,28	0,47
DF	0,54	0,55	0,81
ES	0,71	0,68	1,02
GO	0,23	0,24	0,41
MA	0,32	0,28	0,49
MG	0,33	0,32	0,46
MS	0,55	0,52	0,83
MT	0,13	0,13	0,20
PA	0,19	0,20	0,32
PB	0,36	0,34	0,54
PE	0,26	0,23	0,42
PI	0,14	0,13	0,25
PR	0,45	0,43	0,70
RJ	0,24	0,24	0,42
RN	0,31	0,32	0,53
RO	0,22	0,24	0,40
RR	0,65	0,57	0,98
RS	0,33	0,34	0,54
SC	0,45	0,48	0,74
SE	0,45	0,48	0,73
SP	0,27	0,26	0,48
TO	0,25	0,32	0,38

Fonte: Elaborado pelo autor (2026).

Tabela A.3 – MASE por UF – Dengue 2023.

UF	RF	LSTM	Prophet
AC	0,24	0,26	0,24
AL	0,17	0,11	0,17
AM	0,35	0,28	0,35
AP	0,44	0,19	0,44
BA	0,38	0,33	0,38
CE	0,11	0,08	0,11
DF	0,60	0,57	0,60
ES	0,01	0,00	0,01
GO	0,31	0,36	0,31
MA	0,30	0,25	0,30
MG	0,85	0,78	0,85
MS	0,53	0,55	0,53
MT	0,59	0,61	0,59
PA	0,26	0,21	0,26
PB	0,20	0,18	0,20
PE	0,08	0,06	0,08
PI	0,39	0,33	0,39
PR	0,63	0,68	0,63
RJ	0,70	0,65	0,70
RN	0,16	0,13	0,16
RO	0,93	0,88	0,93
RR	0,04	0,03	0,04
RS	1,88	1,72	1,88
SC	3,94	3,61	3,94
SE	0,29	0,24	0,29
SP	0,39	0,40	0,39
TO	0,12	0,10	0,12

Fonte: Elaborado pelo autor (2026).

Tabela A.4 – Resumo de desempenho – Dengue 2024 (surto histórico).

Indicador	RF	LSTM	Prophet
MASE médio	3,73	4,54	4,53
MASE mediano	1,54	1,76	1,77
WAPE global	92%	93%	94%
UFs MASE < 1	10/26	11/27	11/27

Fonte: Elaborado pelo autor (2026).