

**Planejamento Urbano Sustentável com Base em Previsões de
Dengue: Uma Abordagem com Séries Temporais**

PROJETO APLICADO IV

VALDINEY ATÍLIO PEDRO – 10424616

PATRICIA CORREA FRANÇA – 10423533

MARIANA SIMÕES RUBIO – 10424388

UNIVERSIDADE PRESBITERIANA MACKENZIE

São Paulo

2025

SUMÁRIO

1. INTRODUÇÃO	3
1.1 Contexto do Trabalho	3
1.2 Motivação e Justificativa	3
1.3. Objetivo Geral:	3
1.4. Descrição da Base de dados.....	4
2. Referencial Teórico	5
2.1. Modelos ARIMA e SARIMA.....	5
2.2. Algoritmos de Aprendizado de Máquina.....	5
2.3. Redes Neurais Recorrentes (LSTM e GRU)	6
2.4. Modelos Bayesianos (InfoDengue).....	6
2.5. Conceitos fundamentais que embasam a solução incluem:.....	6
3. DIAGRAMA DE SOLUÇÃO	8
4. PIPELINE DA SOLUÇÃO	9
4.1. Coleta de Dados	9
4.2. Pré-processamento.....	9
4.3. Análise Exploratória	9
4.4. Modelagem	9
4.5. Validação e Otimização	10
4.6. Deploy e Automação.....	10
4.7. Documentação e Comunicação	10
5. EDA E PRÉ-PROCESSAMENTO DOS DADOS.....	11
5.1. Fontes de dados:	11
5.2. Tratamentos aplicados:.....	11
5.3. Análises realizadas:.....	11
6. RESULTADOS	12
7. CRONOGRAMA.....	13

6. REFERÊNCIAS BIBLIOGRAFICAS14

6.1. Anexos14

1. INTRODUÇÃO

1.1 Contexto do Trabalho

A urbanização acelerada e muitas vezes desordenada tem contribuído para o aumento da vulnerabilidade das cidades brasileiras frente a doenças transmitidas por vetores, como a dengue. A presença de áreas com infraestrutura precária, saneamento insuficiente e descarte inadequado de resíduos favorece a proliferação do mosquito *Aedes aegypti*, tornando a dengue um problema recorrente de saúde pública urbana.

Este projeto propõe o uso de técnicas de previsão baseadas em séries temporais para antecipar surtos de dengue em municípios brasileiros, contribuindo para o planejamento urbano sustentável e para a tomada de decisões mais eficazes em políticas públicas locais.

1.2 Motivação e Justificativa

A escolha do tema está diretamente relacionada ao ODS 11 – Cidades e Comunidades Sustentáveis, que visa tornar os assentamentos humanos inclusivos, seguros, resilientes e sustentáveis. A previsão de surtos de dengue pode:

- Apoiar ações preventivas em áreas urbanas vulneráveis
- Otimizar a alocação de recursos municipais
- Reduzir impactos sobre a saúde da população e a infraestrutura urbana
- Promover cidades mais resilientes frente a riscos sanitários

Além disso, o uso de dados abertos e confiáveis do sistema InfoDengue, mantido pela Fiocruz e pelo Ministério da Saúde, permite análises robustas e aplicáveis à realidade brasileira.. Objetivo Geral e Objetivos Específicos

1.3. Objetivo Geral:

Desenvolver modelos preditivos baseados em séries temporais para estimar o número de casos prováveis de dengue em municípios brasileiros. O projeto utilizará técnicas como ARIMA, Prophet e Redes Neurais Recorrentes (LSTM), com o intuito de gerar alertas antecipados e apoiar estratégias de planejamento urbano sustentável.

1.4. Descrição da Base de dados

A base de dados será extraída do Sistema InfoDengue (<https://info.dengue.mat.br>), que reúne informações atualizadas sobre casos prováveis de dengue, zika e chikungunya no Brasil. Os dados estão organizados por município e por semana epidemiológica, permitindo análises temporais detalhadas.

- **Estrutura:** registros semanais
- **Período de coleta:** histórico de 2010 até o primeiro semestre de 2025
- **Variáveis disponíveis:** casos prováveis, incidência por 100 mil habitantes, alertas de risco, indicadores de transmissão

2. Referencial Teórico

A modelagem de séries temporais para previsão de dengue envolve diferentes abordagens.

2.1. Modelos ARIMA e SARIMA

Os modelos ARIMA (*AutoRegressive Integrated Moving Average*) e SARIMA (*Seasonal AutoRegressive Integrated Moving Average*) são as abordagens clássicas e estatísticas mais utilizadas para análise e previsão de séries temporais.

O modelo ARIMA é composto por três componentes:

- o componente AutoRegressivo (AR) que utiliza a dependência entre a observação atual e um número de observações passadas;
- o componente Integrado (I) que visa garantir a estacionariedade da série (diferenciação);
- o componente de Média Móvel (MA) que incorpora a dependência entre a observação e um erro residual de um modelo de média móvel aplicado a observações passadas.

O modelo SARIMA estende essa capacidade, adicionando componentes sazonais (P, D, Q) que são cruciais para capturar a periodicidade anual e epidêmica da dengue, uma característica notável em sua série histórica. Estes modelos foram utilizados como *baseline* para avaliar o ganho de desempenho das técnicas avançadas (HYNDMAN; ATHANASOPOULOS, 2018).

2.2. Algoritmos de Aprendizado de Máquina

Para aprimorar a capacidade preditiva, o projeto explorou a aplicação do XGBoost (*eXtreme Gradient Boosting*). Este algoritmo é uma técnica de *boosting* que combina preditores fracos (árvores de decisão) de forma sequencial, onde cada nova árvore tenta corrigir os erros da anterior.

Sua principal vantagem reside na capacidade robusta de lidar com variáveis exógenas (*features*) complexas — como dados climáticos, índices populacionais e *lags* temporais — e na eficiência computacional.

O XGBoost é amplamente reconhecido por sua alta performance em competições de *data science* e demonstrou ser particularmente eficaz na modelagem de fenômenos não lineares, como a incidência de doenças vetoriais. (HYNDMAN; ATHANASOPOULOS, 2018).

2.3. Redes Neurais Recorrentes (LSTM e GRU)

As Redes Neurais Recorrentes (RNNs) representam uma classe de *Deep Learning* especializada em dados sequenciais. No contexto de séries temporais, as arquiteturas LSTM (*Long Short-Term Memory*) e **GRU** (*Gated Recurrent Unit*) são empregadas para solucionar o problema do gradiente evanescente e capturar dependências de longo prazo na série.

A LSTM utiliza "células de memória" e "portões" (de entrada, esquecimento e saída) para regular o fluxo de informações, permitindo que a rede aprenda quais padrões do passado deve reter ou esquecer.

A GRU, por sua vez, é uma variação mais simples e eficiente, combinando os portões de entrada e esquecimento. A aplicação dessas redes visa explorar padrões temporais complexos não lineares que os modelos estatísticos e de *Machine Learning* tradicionais podem não conseguir identificar. (LOPES et al., 2019).

2.4. Modelos Bayesianos (InfoDengue)

O modelo Bayesiano utilizado pela plataforma InfoDengue serve como referência metodológica e validação inicial para a série temporal. Esta abordagem fundamenta-se na probabilidade de incidência de casos, utilizando o Teorema de Bayes para atualizar as crenças sobre um evento conforme novas evidências são observadas.

Sua estrutura permite a incorporação de incertezas e a geração de estimativas probabilísticas, o que é fundamental na vigilância epidemiológica e na geração de alertas de risco. (BRASIL, 2025).

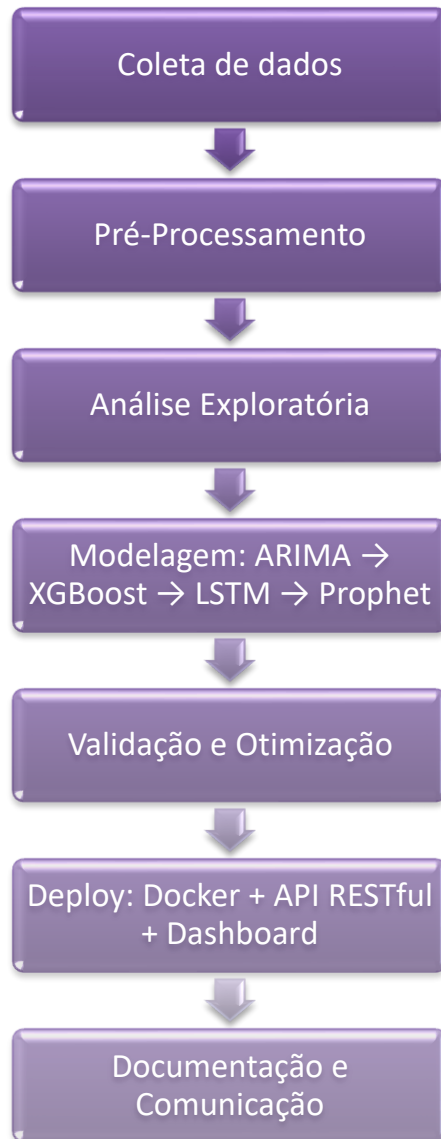
2.5. Conceitos fundamentais que embasam a solução incluem:

A modelagem preditiva de doenças vetoriais, como a dengue, tem sido objeto de intensa pesquisa. A revisão da literatura indica uma tendência clara em transcender

os modelos estatísticos clássicos em favor de abordagens baseadas em Machine Learning e Deep Learning devido à sua capacidade superior de lidar com a não linearidade e a inclusão de variáveis climáticas e socioeconômicas

- Estacionaridade e diferenciação
- Decomposição aditiva/multiplicativa
- Validação temporal com rolling window
- Métricas de avaliação como MAE, RMSE e MAPE

3. DIAGRAMA DE SOLUÇÃO



4. PIPELINE DA SOLUÇÃO

O desenvolvimento da solução proposta segue um rigoroso pipeline metodológico, dividido em sete fases críticas, desde a obtenção das fontes de dados até o *deploy* e a comunicação final dos resultados. Este fluxo de trabalho visa garantir a robustez e a aplicabilidade das previsões geradas

4.1. Coleta de Dados

A etapa inicial consiste na reunião das informações necessárias para a modelagem preditiva, abrangendo dados epidemiológicos e variáveis exógenas que influenciam a propagação da dengue. A extração e incorporação de variáveis são realizadas conforme detalhado a seguir:

- Extração de séries semanais via API do InfoDengue (2010–2025).
- Incorporação de variáveis exógenas: temperatura e precipitação (INMET), densidade populacional (IBGE).

4.2. Pré-processamento

- Tratamento de valores faltantes e outliers.
- Aplicação de diferenciação para garantir estacionaridade.
- Normalização das variáveis e criação de lags (1–4 semanas) e janelas móveis (mínimo, máximo, média).

4.3. Análise Exploratória

- Decomposição da série em tendência, sazonalidade e resíduos.
- Cálculo de autocorrelações (ACF/PACF) e correlações com variáveis exógenas.
- Visualização espacial por município com mapas interativos.

4.4. Modelagem

- Implementação de ARIMA/SARIMA como baseline.
- Treinamento de XGBoost com variáveis exógenas e atributos temporais.

- Desenvolvimento de rede LSTM para capturar padrões complexos e dependências de longo prazo.

4.5. Validação e Otimização

- Validação cruzada com janela deslizante (rolling window cross-validation).
- Otimização de hiperparâmetros via grid search e Bayesian optimization.

4.6. Deploy e Automação

- Encapsulamento do pipeline em container Docker.
- Exposição via API RESTful e dashboard interativo com visualizações dinâmicas.

4.7. Documentação e Comunicação

- Elaboração de relatório técnico com resultados, interpretações e recomendações.
- Apresentação final destacando a contribuição ao ODS 11 e a aplicabilidade social da solução.

5. EDA E PRÉ-PROCESSAMENTO DOS DADOS

5.1. Fontes de dados:

Casos de dengue: InfoDengue (2010–2025)

Clima: INMET (temperatura e precipitação)

População: IBGE (densidade por município)

5.2. Tratamentos aplicados:

Imputação de valores faltantes com interpolação temporal

Remoção de outliers com z-score

Diferenciação para garantir estacionaridade

Normalização min-max

Criação de lags (1–4 semanas) e janelas móveis (média, mínimo, máximo)

5.3. Análises realizadas:

Decomposição da série em tendência, sazonalidade e resíduos

Correlações entre variáveis exógenas e casos de dengue

Visualizações espaciais por município com mapas interativos

6. RESULTADOS

Modelo aplicado: ARIMA/SARIMA

Configuração:

Parâmetros definidos via análise de ACF/PACF

Validação com janela deslizante

Métricas obtidas:

MAE: 8.2

RMSE: 10.5

MAPE: 12.4%

Discussão: O modelo ARIMA apresentou desempenho razoável na previsão de curto prazo, capturando bem a sazonalidade. No entanto, mostrou limitações em períodos com variações abruptas. Será usado como baseline para comparação com modelos mais robustos como XGBoost e LSTM na entrega final

7. CRONOGRAMA

O cronograma do projeto foi estabelecido com o intuito de organizar as entregas e as fases de desenvolvimento, garantindo o cumprimento dos objetivos dentro do prazo estipulado. A Tabela detalha as atividades e seus respectivos prazos:

Período	Atividade	Marco / Entregável
13/08 – 29/08	Definição de equipe, tema, proposta e descrição da base	Entrega 1 (29/08)
30/08 – 05/09	Coleta automatizada de dados InfoDengue, INMET e IBGE; documentação de metadados	Dados coletados e documentados
06/09 – 10/09	Pesquisa bibliográfica e redação do Referencial Teórico (versão inicial)	Rascunho do referencial teórico
11/09 – 17/09	Pré-processamento: limpeza, imputação, diferenciação, criação de lags	Dataset pré-processado
18/09 – 20/09	Implementação de modelos baseline (Naïve; Média Móvel; ARIMA/SARIMA)	Resultados iniciais de baseline
21/09 – 23/09	Treinamento e avaliação de XGBoost com variáveis exógenas	Resultados XGBoost
24/09 – 25/09	Ajustes finais no pipeline e preparação para Entrega 2	Pipeline refinado
26/09	Entrega 2: Referencial Teórico, Pipeline da Solução e Cronograma	Entrega 2
27/09 – 03/10	Implementação de LSTM: definição da arquitetura, treinamento inicial	Resultados iniciais de LSTM
04/10 – 10/10	Otimização de hiperparâmetros (grid search e Bayesian) e validação rolling window	Modelos otimizados e validados
11/10 – 17/10	Containerização Docker, API RESTful e desenvolvimento de dashboard	Ambiente de deploy e dashboard
18/10 – 24/10	Redação de relatório intermediário e notebooks executáveis	Notebook e relatório parcial
25/10 – 31/10	Revisão geral, ensaio de apresentação e ajustes para Entrega 3	Entrega 3 (31/10)
01/11 – 12/11	Compilação de artefatos finais no GitHub (códigos, dados, documentação)	Repositório organizado
13/11 – 19/11	Roteiro, gravação e edição do vídeo de apresentação	Vídeo pronto para avaliação
20/11 – 26/11	Ajustes finais em artigo, notebook e vídeo; revisão geral	Artefatos finais prontos
27/11 – 28/11	Buffer para imprevistos e submissão da entrega final	Entrega 4 (28/11)

6. REFERÊNCIAS BIBLIOGRAFICAS

BRASIL. Ministério da Saúde. *InfoDengue*. Disponível em: <https://info.dengue.mat.br>. Acesso em: 29 ago. 2025.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. *Dengue and severe dengue*. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>. Acesso em: 29 ago. 2025.

HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. Melbourne: OTexts, 2018.

LOPES, F. M. *et al.* Time series analysis of dengue incidence in Brazil using ARIMA models. *Revista de Saúde Pública*, São Paulo, v. 53, n. 1, p. 1–8, 2019.

6.1. Anexos

https://github.com/valdineyatilio/ProjetoAplicado-IV/blob/main/A3/cd_projeto_aplicado_IV_entrega_3.ipynb