



UNIVERSIDADE FEDERAL DE MATO GROSSO
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA AMBIENTAL

ADRIANO CARVALHO NUNES E ARAÚJO

PREVISÃO DE DENGUE BASEADA EM
APRENDIZADO DE MÁQUINA COMO
ESTRATÉGIA DE ADAPTAÇÃO CLIMÁTICA NO
BRASIL

CUIABÁ – MT
2025

ADRIANO CARVALHO NUNES E ARAÚJO

**PREVISÃO DE DENGUE BASEADA EM
APRENDIZADO DE MÁQUINA COMO
ESTRATÉGIA DE ADAPTAÇÃO CLIMÁTICA NO
BRASIL**

Dissertação apresentada ao Programa de Pós-Graduação em Física Ambiental da Universidade Federal de Mato Grosso como parte dos requisitos para obtenção do título de Mestre em Física Ambiental.

Orientador(a): Prof^a. Dr^a. DANIELA DE OLIVEIRA MAIONCHI

CUIABÁ – MT

2025

Dados Internacionais de Catalogação na Fonte.

A663p Araújo, Adriano Carvalho Nunes e.
Previsão de dengue baseada em aprendizado de máquina como estratégia de adaptação climática no Brasil tropical [recurso eletrônico] / Adriano Carvalho Nunes e Araújo. -- Dados eletrônicos (1 arquivo : 64 f., il. color., pdf). -- 2025.

Orientador: Daniela de Oliveira Maionchi.
Dissertação (mestrado) - Universidade Federal de Mato Grosso, Instituto de Física, Programa de Pós-Graduação em Física Ambiental, Cuiabá, 2025.
Modo de acesso: World Wide Web: <https://ri.ufmt.br>.
Inclui bibliografia.

1. Dengue. 2. Previsão. 3. Aprendizado de Máquina. 4. Clima. 5. Mato Grosso. I. Maionchi, Daniela de Oliveira, *orientador*. II. Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

Permitida a reprodução parcial ou total, desde que citada a fonte.



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE MATO GROSSO
PRÓ-REITORIA DE ENSINO DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA AMBIENTAL

FOLHA DE APROVAÇÃO

TÍTULO: PREVISÃO DE DENGUE BASEADA EM APRENDIZADO DE MÁQUINA COMO ESTRATÉGIA DE ADAPTAÇÃO CLIMÁTICA NO BRASIL TROPICAL

AUTOR: MESTRANDO ADRIANO CARVALHO NUNES E ARAÚJO

Dissertação defendida e aprovada em **30 de outubro de 2025**.

COMPOSIÇÃO DA BANCA EXAMINADORA

1. PROFA. DRA. DANIELA DE OLIVEIRA MAIONCHI [Presidente Banca / **Orientadora**]

INSTITUIÇÃO: UNIVERSIDADE FEDERAL DE MATO GROSSO

2. PROF. DR. JUNIOR GONÇALVES DA SILVA [Examinador Interno]

INSTITUIÇÃO: UNIVERSIDADE FEDERAL DE MATO GROSSO

3. PROF. DR. ERNANDES SOBREIRA OLIVEIRA JUNIOR [Examinador Externo]

INSTITUIÇÃO: UNIVERSIDADE DO ESTADO DE MATO GROSSO

CUIABÁ, 30/10/2025.



Documento assinado eletronicamente por **MARCELO SACARDI BIUDES**, **Coordenador(a) de Programas de Pós-Graduação em Física Ambiental - IF/UFMT**, em 31/10/2025, às 18:39, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Junior Gonçalves da Silva, Usuário Externo**, em 31/10/2025, às 18:46, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ernandes Sobreira Oliveira Junior, Usuário Externo**, em 02/11/2025, às 10:56, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **DANIELA DE OLIVEIRA MAIONCHI, Docente da Universidade Federal de Mato Grosso**, em 03/11/2025, às 10:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufmt.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **8613016** e o código CRC **0A317A94**.

A Deus, meu amor e meus familiares, a eles eu dedico este trabalho.

AGRADECIMENTOS

A Deus, por ter sido minha luz e minha fortaleza em todos os momentos desta caminhada. Agradeço pela sabedoria concedida, pela serenidade nos dias difíceis e pela força que me sustentou quando as incertezas pareciam maiores que os sonhos. Sem Sua presença, nada disso seria possível.

À minha amada esposa, Rafaella, por todo o amor, paciência e compreensão. Obrigado por acreditar em mim mesmo quando eu duvidava, por dividir comigo cada noite de estudo, cada momento de ansiedade e cada pequena vitória. Sua presença foi meu maior incentivo e seu carinho, o combustível que me manteve firme até o fim.

À minha família, que é o meu alicerce mais sólido. Agradeço por cada palavra de apoio, por cada gesto de afeto e por me ensinarem, desde sempre, o valor do esforço, da fé e da honestidade. Tudo o que conquistei é reflexo do amor e da base que vocês me deram.

À minha orientadora, Daniela Maionchi, por sua paciência, sabedoria e orientação cuidadosa. Sua dedicação e comprometimento foram fundamentais para o desenvolvimento deste trabalho. Agradeço por cada conselho, por cada correção e por acreditar no meu potencial. Levo comigo não apenas o aprendizado científico, mas também o exemplo de profissionalismo e humanidade.

Aos amigos e colegas que fiz durante o mestrado, por tornarem essa jornada mais leve e enriquecedora. A convivência, as trocas de ideias, as risadas e o apoio mútuo foram essenciais para seguir em frente. Cada um de vocês deixou uma marca importante nesta trajetória.

Por fim, a todos que, de alguma forma, contribuíram para que este sonho se tornasse realidade — meu sincero e profundo agradecimento. Este trabalho é o resultado de muitas mãos, corações e inspirações que me acompanharam nessa caminhada.

Resumo

A dengue constitui um dos maiores desafios de saúde pública em regiões tropicais, cuja transmissão é fortemente influenciada por variáveis climáticas e ambientais. Nesta dissertação, desenvolveu-se uma abordagem preditiva integrando dados epidemiológicos e meteorológicos para estimar a incidência de dengue em municípios de Mato Grosso, Brasil, entre 2001 e 2024. Os casos notificados foram obtidos da Secretaria Estadual de Saúde (SES-MT) e associados a variáveis do reanalisador ERA5-Land, incluindo temperatura média, máxima, mínima e precipitação diária. As séries foram agregadas em escala semanal, normalizadas e suavizadas por média móvel de 14 semanas, além da criação da variável *Pressão de Transmissão* (Pt), que sintetiza o efeito combinado entre incidência e suscetibilidade populacional. Foram avaliados quatro algoritmos de aprendizado de máquina — *Random Forest Regressor* (RFR), *Gradient Boosting Regressor* (GBR), *Support Vector Regressor* (SVR) e *k-Nearest Neighbors* (KNN) —, comparados por métricas de desempenho (MAE, RMSE, R^2 , coeficiente de Willmott e correlação de Pearson). Os resultados mostraram desempenho consistente dos modelos RFR e GBR, com melhor generalização e menor sobreajuste. A análise de interpretabilidade via SHAP revelou a *Pressão de Transmissão* como variável mais influente, seguida da precipitação e da temperatura mínima, evidenciando a relevância conjunta dos fatores epidemiológicos e climáticos. A metodologia proposta demonstra potencial para aplicação operacional em sistemas de vigilância e apoio à tomada de decisão, oferecendo previsões interpretáveis e reprodutíveis para contextos tropicais vulneráveis às variações climáticas.

Palavras-chave: Dengue; Previsão; Aprendizado de Máquina; Clima; Mato Grosso.

Abstract

Dengue fever remains one of the major public health challenges in tropical regions, with transmission dynamics strongly influenced by climatic and environmental conditions. This dissertation developed a predictive approach integrating epidemiological and meteorological data to estimate dengue incidence across municipalities in Mato Grosso State, Brazil, from 2001 to 2024. Reported dengue cases were obtained from the State Health Department (SES-MT) and combined with meteorological variables from ERA5-Land reanalysis, including mean, maximum, and minimum temperatures and daily precipitation. The datasets were aggregated to a weekly scale, normalized, and smoothed using a 14-week moving average. Additionally, a derived variable termed *Transmission Pressure* (Pt) was proposed to represent the combined effect of incidence and population susceptibility. Four machine learning algorithms were evaluated — *Random Forest Regressor* (RFR), *Gradient Boosting Regressor* (GBR), *Support Vector Regressor* (SVR), and *k-Nearest Neighbors* (KNN) — and their performance was assessed using multiple metrics (MAE, RMSE, R^2 , Willmott's index, and Pearson's correlation). The results indicated superior performance and generalization capacity for the RFR and GBR models, with reduced overfitting compared to SVR and KNN. Model interpretability analysis using SHAP values identified *Transmission Pressure* as the most influential predictor, followed by precipitation and minimum temperature, highlighting the combined role of epidemiological and climatic factors in dengue transmission. The proposed methodology demonstrates strong potential for operational application in epidemiological surveillance systems, providing interpretable and reproducible forecasts for dengue outbreaks in tropical regions sensitive to climatic variability.

Keywords: Dengue; Forecasting; Machine Learning; Climate; Mato Grosso.

LISTA DE FIGURAS

3.1	Mapa com os município de Cuiabá, Rondonópolis, Barra do Garças, Cáceres, Juína, Porto Alegre do Norte, Sinop, Tangará da Serra, Diamantino, Alta Floresta, Juara, Peixoto de Azevedo, Água Boa, Pontes e Lacerda, Colíder, e São Félix do Araguaia.	21
3.2	Mapa das regionais de saúde do estado de Mato Grosso.	22
3.3	Mapa da quantidade de casos de dengue por município de acordo com a regional de saúde. Fonte: Autoral	23
3.4	Esquema do ciclo de transmissão do vírus da dengue Secretaria da Saúde Gov RS.	26
3.5	Processo de aprendizagem supervisionada.	28
3.6	Método de <i>random forest</i>	29
3.7	Método de <i>gradient boosting</i>	30
3.8	Método do <i>support vector regressor</i>	31
3.9	Método de <i>KNN</i>	31
3.10	Dispersão dos dados previstos pelos medidos para os modelos de Água Boa (KNN), Alta Floresta (GBR), Barra do Garças (GBR), Cáceres (KNN), Colíder (RFR), Cuiabá (RFR), Diamantino (RFR), Juara (SVR), Juína (GBF), Peixoto de Azevedo(RFR), Pontes e Lacerda (RFR), Porto Alegre do Norte (RFR), Rondonópolis (RFR), São Félix do Araguaia (RFR), Sinop (RFR), Tangará da Serra (RFR).	34
3.11	Incidência anual dos casos de dengue por 100.000 habitantes das cidades de Água Boa, Cáceres, Colíder e Porto Alegre do Norte.	35
3.12	Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Água Boa (KNN), Alta Floresta (GBR), Barra do Garças (GBR) e Cáceres (KNN).	39
3.13	Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Colíder (RFR), Cuiabá (RFR), Diamantino (RFR) e Juara (SVR).	40
3.14	Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Juína (GBR), Peixoto de Azevedo (RFR), Pontes e Lacerda (RFR) e Porto Alegre do Norte (RFR).	41
3.15	Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Rondonópolis (RFR), São Félix do Araguaia (RFR), Sinop (RFR) e Tangará da Serra (RFR).	42

3.16	Comparação entre as variáveis mais importantes para os modelos de Água Boa (KNN), Alta Floresta (GBR), Barra do Garças (GBR), Cáceres (KNN), Colíder (RFR), Cuiabá (RFR), Diamantino (RFR) e Juara (SVR).	44
3.17	Comparação entre as variáveis mais importantes para os modelos de Juína (GBR), Peixoto de Azevedo (RFR), Pontes e Lacerda (RFR), Porto Alegre do Norte (RFR), Rondonópolis (RFR), São Félix do Araguaia (RFR), Sinop (RFR) e Tangará da Serra (RFR).	45
3.18	Dispersão dos dados previstos pelos medidos para os modelos de Água Boa (KNN), Alta Floresta (RFR), Barra do Garças (GBR), Cáceres (GBR), Colíder (KNN), Cuiabá (RFR), Diamantino (RFR), Juara (SVR), Juína (SVR), Peixoto de Azevedo (RFR), Pontes e Lacerda (KNN), Porto Alegre do Norte (RFR), Rondonópolis (SVR), São Félix do Araguaia (RFR), Sinop (SVR) e Tangará da Serra (GBR).	47
3.19	Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Água Boa (KNN), Alta Floresta (RFR), Barra do Garças (GBR) e Cáceres (GBR).	51
3.20	Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Colíder (KNN), Cuiabá (RFR), Diamantino (RFR) e Juara (SVR).	52
3.21	Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Juína (SVR), Peixoto de Azevedo (RFR), Pontes e Lacerda (KNN) e Porto Alegre do Norte (RFR).	53
3.22	Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Rondonópolis (SVR), São Félix do Araguaia (KNN), Sinop (KNN) e Tangará da Serra (GBR).	54

LISTA DE TABELAS

3.1	Informações sobre as populações das localidades estudadas e período considerado. <i>Ano inicial</i> e <i>Ano final</i> referem-se ao intervalo dos dados anuais de dengue disponíveis para cada município.	24
3.2	Variáveis Utilizadas na Modelagem Preditiva.	27
3.3	Desempenho dos modelos nos conjuntos de treino e teste.	35
3.4	Desempenho dos modelos nos conjuntos de treino e teste.	47

SUMÁRIO

Agradecimentos	4
Resumo	5
Abstract	6
1 Introdução Geral	11
1.0.1 Objetivos	12
2 Revisão Bibliográfica	13
2.1 Aprendizado de Máquina em Saúde e Doenças Infecciosas	13
2.2 Modelagem de Dengue com ML: Panorama Global	14
2.3 Modelagem de Dengue com ML: Evidências no Brasil	15
2.4 Síntese e Lacunas	16
3 Capítulo 1 - Artigo Científico	17
3.1 Resumo	17
3.2 Introdução	18
3.3 Materiais e Métodos	21
3.3.1 Área de Estudo	21
3.3.2 Coleta e Pré-Processamento de Dados	23
3.3.3 Modelos de Machine Learning	28
3.3.4 Avaliação dos Modelos	32
3.4 Resultados e Discussão	33
3.4.1 Previsão Semanal dos Casos de Dengue	33
3.4.2 Seleção de Variáveis	46
3.5 Conclusões	55
4 Conclusão Geral	58
Referências	59

1. Introdução Geral

A dengue permanece como um dos maiores desafios de saúde pública nas regiões tropicais e subtropicais, com crescimento expressivo da incidência nas últimas décadas e forte relação com fatores climáticos e socioambientais. No Brasil, a sazonalidade marcada pelo regime de chuvas e temperaturas elevadas favorece o ciclo do vetor *Aedes aegypti*, intensificando a transmissão em períodos chuvosos e em contextos urbanos com infraestrutura precária. Em Mato Grosso, essa dinâmica é modulada pela heterogeneidade climática dos biomas Amazônia, Cerrado e Pantanal, e por alterações de uso e cobertura da terra que afetam precipitação, temperatura e umidade — elementos críticos ao ciclo do vetor Leung et al. [2023], Lima et al. [2022].

Projeções recentes indicam que choques climáticos extremos, como excesso de chuva, estiagens prolongadas e ondas de calor, podem expandir a área de risco e alterar a sazonalidade da dengue, ora criando criadouros por acúmulo de água, ora estimulando o armazenamento domiciliar em reservatórios improvisados. A temperatura, por sua vez, modula taxas vitais do vetor e o período de incubação extrínseco do vírus, com efeitos não lineares e limiares térmicos específicos. Esses mecanismos, combinados a fatores demográficos e urbanos, impõem a necessidade de ferramentas preditivas robustas e interpretáveis para apoiar ações de vigilância e resposta Ong et al. [2023], Leung et al. [2023], Chen and Moraga [2025a].

Nesse contexto, este estudo integra dados epidemiológicos provenientes da Secretaria Estadual de Saúde de Mato Grosso (SES-MT) e variáveis meteorológicas derivadas do reanalisador ERA5-Land, estimando a incidência com base em projeções populacionais e propondo a variável *Pressão de Transmissão* (Pt), que incorpora o atraso biológico do ciclo vetor-hospedeiro. São comparados algoritmos de aprendizado de máquina — *Random Forest Regressor* (RFR), *Gradient Boosting Regressor* (GBR), *Support Vector Regressor* (SVR) e *k-Nearest Neighbors* (KNN) — com pré-processamento temporal (agregação semanal, normalização e média móvel de 14 semanas) e avaliação por múltiplas métricas de desempenho (MAE, RMSE, R^2 , d de Willmott e r de Pearson). A interpretabilidade é abordada por meio de valores SHAP, permitindo identificar a relevância relativa das variáveis epidemiológicas e meteorológicas — especialmente a Pt, a precipitação e a

temperatura mínima.

Os resultados obtidos para os municípios e regionais de saúde de Mato Grosso sugerem desempenho preditivo consistente, com boa aderência às variações semanais e capacidade de capturar picos epidêmicos quando se combinam informações epidemiológicas e climáticas. A análise de importância confirma a *Pressão de Transmissão* como variável central, seguida, em geral, por precipitação e temperatura mínima, demonstrando que o uso isolado de apenas um grupo de variáveis — epidemiológicas ou climáticas — degrada a acurácia e desloca temporalmente os picos previstos. Essas evidências reforçam o potencial das abordagens integradas e interpretáveis para aprimorar o planejamento e a resposta em saúde pública, alinhando-se às tendências recentes da literatura internacional sobre previsão de arboviroses mediadas por fatores climáticos Leung et al. [2023], Chen and Moraga [2025a,b].

1.0.1 Objetivos

- Geral: Prever a incidência de dengue com uma semana de antecedência, utilizando modelos de inteligência artificial baseados em variáveis climáticas e epidemiológicas, como estratégia de adaptação frente aos impactos das mudanças climáticas sobre a saúde pública.
- Específicos:
 - Investigar a influência da pressão de transmissão defasada e de variáveis meteorológicas na dinâmica temporal da dengue.
 - Testar diferentes algoritmos de machine learning (Random Forest, Gradient Boosting, Support Vector Regression e K-Nearest Neighbors) na previsão semanal de casos de dengue.
 - Comparar o desempenho dos modelos considerando diferentes combinações de variáveis de entrada (apenas climáticas, apenas epidemiológicas e ambas).
 - Interpretar os resultados dos modelos por meio da técnica SHAP, identificando as variáveis com maior impacto na previsão da incidência.
 - Avaliar a aplicabilidade dos modelos como ferramentas de apoio à gestão pública em cenários de risco climático, contribuindo para ações de mitigação e prevenção.

2. Revisão Bibliográfica

2.1 Aprendizado de Máquina em Saúde e Doenças Infecciosas

O aprendizado de máquina (ML) tem se consolidado como uma ferramenta poderosa na área da saúde, capaz de extrair padrões complexos e realizar previsões a partir de grandes volumes de dados heterogêneos. Sua aplicação abrange desde a vigilância epidemiológica e a detecção precoce de surtos, até o apoio diagnóstico, prognóstico e gestão de sistemas de saúde. Esses algoritmos, ao aprenderem a partir de dados históricos, conseguem identificar associações não lineares entre variáveis biológicas, ambientais e comportamentais, o que torna o ML particularmente adequado para lidar com doenças infecciosas, caracterizadas por dinâmicas multifatoriais e variabilidade espaço-temporal significativa.

Revisões recentes indicam uma rápida expansão do uso de modelos baseados em aprendizado supervisionado e não supervisionado para tarefas médicas, incluindo o uso de redes neurais profundas (*deep learning*), métodos de florestas aleatórias (*random forests*), modelos de *boosting* e arquiteturas híbridas que combinam componentes determinísticos e probabilísticos. No contexto das doenças infecciosas, destaca-se o uso crescente de abordagens explicáveis (*Explainable Artificial Intelligence* – XAI), que permitem compreender a contribuição de cada variável para a predição, aumentando a confiabilidade e a interpretabilidade dos modelos Andrade Girón et al. [2025], Lima et al. [2022].

Na área específica das arboviroses — dengue, zika e chikungunya —, essas técnicas têm sido aplicadas tanto à previsão epidemiológica quanto à classificação clínica e à triagem de casos suspeitos. Revisões sistemáticas mostram que o uso de ML em dados clínicos e laboratoriais alcança resultados promissores para o diagnóstico diferencial de dengue, permitindo identificar padrões de evolução da doença e prever casos graves com base em poucos parâmetros Andrade Girón et al. [2025]. No âmbito da vigilância, os modelos de ML possibilitam antecipar surtos e quantificar o risco em diferentes escalas espaciais e temporais, integrando dados ambientais, meteorológicos e socioeconômicos Lima et al. [2022].

Dessa forma, o aprendizado de máquina emerge não apenas como um recurso

analítico, mas como um componente estratégico de sistemas de vigilância inteligente, capaz de apoiar a formulação de políticas públicas de prevenção e controle de doenças infecciosas, especialmente em cenários tropicais sensíveis às variações climáticas.

2.2 Modelagem de Dengue com ML: Panorama Global

Nas últimas décadas, o aprendizado de máquina tem sido amplamente explorado para prever a incidência e a dinâmica espaço-temporal da dengue em diferentes contextos climáticos e socioeconômicos. A literatura internacional revela uma diversidade de abordagens — desde modelos estatísticos tradicionais até arquiteturas complexas de *deep learning* — utilizadas para capturar as relações não lineares entre variáveis meteorológicas (temperatura, precipitação, umidade), fatores demográficos e indicadores entomológicos.

A revisão sistemática conduzida por Leung et al. [2023] fornece um panorama abrangente das metodologias empregadas globalmente em previsão de surtos de dengue, destacando tanto a evolução histórica quanto as lacunas metodológicas. O estudo aponta que, embora os modelos de ML apresentem desempenho superior aos métodos lineares clássicos, ainda há pouca padronização quanto às métricas de avaliação, à escolha de variáveis e à validação cruzada temporal. Essa heterogeneidade dificulta a comparação direta entre estudos e a transposição dos modelos para novos contextos.

Comparações entre diferentes algoritmos mostram resultados consistentes de desempenho elevado para modelos de árvores impulsionadas por gradiente (XGBoost, LightGBM), florestas aleatórias e redes neurais recorrentes (LSTM), especialmente quando combinados com variáveis climáticas defasadas e índices vetoriais Ong et al. [2023]. Novas arquiteturas, como modelos híbridos que integram camadas de LSTM, GRU e *Transformers*, têm alcançado resultados ainda mais robustos em horizontes de previsão de curto e médio prazos, demonstrando a importância de incorporar dependências temporais e contextuais Zhou et al. [2024].

Mais recentemente, abordagens multitarefas foram propostas para previsão simultânea de múltiplas arboviroses, permitindo que informações compartilhadas entre doenças vetoriais auxiliem na previsão de picos epidêmicos de dengue e zika Farias et al. [2025]. Esse tipo de modelagem integrada reforça a tendência global de adotar modelos generalistas, capazes de operar em diferentes contextos ecológicos e epidemiológicos.

Em síntese, a literatura internacional converge para o reconhecimento de que o sucesso preditivo dos modelos de dengue depende da integração de variáveis meteorológicas, temporais e entomológicas, bem como da adoção de estratégias interpretáveis, como o uso de técnicas de explicabilidade (e.g., SHAP), que tornam os resultados mais transparentes

e aplicáveis à gestão em saúde pública.

2.3 Modelagem de Dengue com ML: Evidências no Brasil

No cenário brasileiro, observa-se um aumento expressivo no número de estudos aplicando aprendizado de máquina à previsão da dengue, refletindo tanto a disponibilidade crescente de dados climáticos e epidemiológicos quanto a necessidade de soluções preditivas para subsidiar políticas públicas. O país oferece um ambiente heterogêneo, com variações marcantes de clima, vegetação e densidade populacional, o que favorece o uso de ML para explorar as relações não lineares e regionais entre variáveis.

Um dos trabalhos mais abrangentes no contexto nacional foi conduzido por Chen and Moraga [2025a], que compararam métodos estatísticos e algoritmos de ML em séries temporais semanais de dengue no estado do Rio de Janeiro. Os autores mostraram que técnicas de aprendizado de máquina superaram modelos tradicionais (como ARIMA e SARIMA) na previsão de curto prazo, especialmente quando integradas a variáveis meteorológicas e defasagens de até oito semanas. O estudo também enfatiza a importância de considerar a autocorrelação temporal e a sazonalidade, aspectos frequentemente subestimados em modelos lineares.

Em escala nacional, Chen and Moraga [2025b] aplicaram redes neurais recorrentes do tipo LSTM, aprimoradas com análise de interpretabilidade via SHAP, para prever a incidência de dengue em múltiplos estados brasileiros. Os resultados demonstraram que a inclusão de efeitos espaciais e defasagens climáticas aumentou substancialmente a precisão do modelo, reforçando a necessidade de incluir variáveis ambientais com janelas temporais coerentes ao ciclo vetor-vírus.

De forma complementar, Sebastianelli et al. [2024] desenvolveram um modelo de *ensemble* reproduzível para previsão de surtos de dengue em escala nacional. O estudo destacou o potencial da combinação de múltiplos algoritmos — como *random forests*, regressão por gradiente e redes neurais — e a importância da reprodutibilidade científica por meio de *pipelines* documentados e publicamente disponíveis.

Outra contribuição relevante é o trabalho de Villela [2025], que utilizou a técnica de *path signatures* para identificar municípios brasileiros com maior probabilidade de apresentar incidência elevada de dengue. A abordagem, baseada em representações matemáticas de trajetórias temporais, mostrou-se eficaz para identificar padrões espaciais e temporais complexos, complementando métodos tradicionais de regressão.

Esses estudos demonstram que o Brasil tem se tornado um polo de inovação na aplicação de ML a problemas epidemiológicos, com abordagens que vão desde comparações

metodológicas locais até modelos nacionais interpretáveis e reprodutíveis. A convergência desses resultados evidencia o potencial das técnicas de aprendizado de máquina para apoiar estratégias de vigilância integrada e antecipar períodos críticos de transmissão da dengue no território brasileiro.

2.4 Síntese e Lacunas

A literatura revisada indica uma evolução significativa das abordagens baseadas em aprendizado de máquina para previsão de dengue, tanto em escala global quanto nacional. Três eixos principais emergem como determinantes do desempenho preditivo: (i) a combinação de dados epidemiológicos com variáveis meteorológicas e defasagens temporais que representem o ciclo de transmissão; (ii) a adoção de modelos com interpretabilidade, como SHAP, que permitem identificar os fatores mais influentes e garantir transparência na tomada de decisão; e (iii) a implementação de protocolos de validação temporal e espacial robustos, capazes de avaliar a generalização dos modelos em contextos regionais distintos.

Apesar desses avanços, ainda persistem desafios. Estudos apontam a falta de padronização na seleção de métricas e na comparação entre abordagens Leung et al. [2023], além de lacunas relacionadas à qualidade e completude dos dados epidemiológicos e climáticos. Mesmo entre os estudos brasileiros, nota-se uma heterogeneidade metodológica considerável, tanto na escolha dos algoritmos quanto nas estratégias de pré-processamento e defasagem temporal Chen and Moraga [2025a].

Nesse contexto, o presente trabalho busca contribuir ao propor uma abordagem integrada e interpretável que combina dados meteorológicos e epidemiológicos por meio da variável *Pressão de Transmissão*, incorporando defasagens temporais e interpretação via SHAP. Essa proposta se alinha às recomendações mais recentes da literatura, visando oferecer previsões mais precisas, explicáveis e aplicáveis à vigilância da dengue em Mato Grosso e em regiões tropicais similares.

3. Capítulo 1 - Artigo Científico

3.1 Resumo

A dengue é uma arbovirose tropical de alta incidência no Brasil, cuja dinâmica de transmissão está fortemente associada a variáveis climáticas. Frente ao avanço das mudanças climáticas e seus impactos sobre a distribuição do vetor *Aedes aegypti*, torna-se fundamental o desenvolvimento de ferramentas para auxiliar na vigilância epidemiológica. Este trabalho visa desenvolver modelos de previsão de casos de dengue utilizando técnicas de aprendizado de máquina, com foco em 16 municípios do estado de Mato Grosso, incluindo áreas de alta vulnerabilidade climática. Foram utilizados dados epidemiológicos da Secretaria de Estado de Saúde (SES-MT) e dados climáticos reanalisados do conjunto ERA5-Land. A incidência foi calculada com base em projeções populacionais, e a variável Pressão de Transmissão foi construída para considerar o atraso biológico do ciclo do vetor. Os modelos Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR), Support Vector Regressor (SVR) e K-Nearest Neighbors (KNN) foram testados em todos os municípios. Os resultados mostraram que, de forma geral, os modelos RFR e GBR apresentaram melhor desempenho, com erros médios absolutos (MAE) baixos e índices de Willmott e correlação de Pearson superiores a 0,90 na maioria dos municípios analisados. A análise de importância das variáveis, realizada com SHAP, evidenciou o papel central da Pressão de Transmissão e das variáveis meteorológicas – especialmente temperatura e precipitação – na predição dos casos. Os modelos baseados em inteligência artificial com integração de dados climáticos e indicadores epidemiológicos constitui uma abordagem promissora para fortalecer estratégias de adaptação às mudanças climáticas e controle da dengue em nível local.

Palavras-chave: arboviroses; aprendizado de máquina; previsão de epidemias; climatologia.

3.2 Introdução

A dengue é uma infecção viral transmitida aos humanos pela picada de fêmeas de mosquitos *Aedes aegypti* infectados Fiocruz, pertencente à família Flaviviridae, gênero Orthoflavivirus, e composta por quatro sorotipos distintos: DENV-1, DENV-2, DENV-3 e DENV-4 de Azevedo Assunção et al. [2025]. Nas últimas décadas, a incidência da doença cresceu de forma alarmante, passando de 505.430 casos registrados em 2000 para 5,2 milhões em 2019, com cerca de metade da população mundial atualmente em risco. Estima-se que ocorram entre 100 e 400 milhões de infecções anuais por dengue [WHO], o que evidencia a magnitude do problema de saúde pública. Globalmente, a doença é sub-notificada, e o Brasil lidera em número de casos reportados. Projeções indicam que, para cada caso de dengue notificado no país, podem existir até 12 casos reais não registrados Silva et al. [2016], reforçando a necessidade de estratégias eficazes de monitoramento e previsão de surtos.

A dengue tem sido um desafio persistente desde a década de 1980, com epidemias que atingem praticamente todas as regiões do Brasil Braga and Valle [2007]. Seu comportamento é fortemente sazonal, com aumento expressivo dos casos no início do ano, coincidindo com a estação chuvosa mais intensa de Azevedo Assunção et al. [2025]. A presença de água é essencial para o desenvolvimento das larvas, que passam por estágios aquáticos antes de se tornarem mosquitos adultos. Durante períodos de chuva, o acúmulo de água em recipientes, pneus e calhas cria ambientes ideais para a reprodução do vetor Benedum et al. [2018]. No estado de Mato Grosso, esse padrão é intensificado pelas condições climáticas típicas da região Centro-Oeste do Brasil, marcada por clima tropical, com estação chuvosa de outubro a abril e seca de maio a setembro. Essa dinâmica climática, aliada à expansão urbana desordenada e à degradação ambiental, cria um cenário propício para a proliferação do mosquito *Aedes aegypti* e para a manutenção do ciclo de transmissão da doença.

Os biomas presentes em Mato Grosso – Amazônia, Cerrado e Pantanal – encontram-se sob crescente pressão antrópica e ambiental, intensificada pelas mudanças climáticas e pelos eventos extremos. A região norte do estado apresenta uma elevada taxa de alteração da cobertura do solo, impulsionada pela expansão agropecuária e pelas queimadas ilegais, enquanto o Pantanal, no sul, enfrenta impactos severos decorrentes das queimadas e da redução de áreas alagáveis. O Cerrado, por sua vez, vem perdendo progressivamente sua vegetação nativa, comprometendo o equilíbrio hídrico regional e alterando os padrões de precipitação, que se tornam mais concentrados e irregulares. Segundo o PRODES/INPE (2024) Almeida et al. [2021], a área desmatada do Cerrado já alcança 236.700 km², e tais mudanças ambientais afetam diretamente a ecologia do *Aedes aegypti*, favorecendo a dis-

seminação da dengue ao modificar as condições de umidade e temperatura que influenciam o ciclo de vida do vetor.

As mudanças climáticas, caracterizadas por alterações nos padrões de chuva e temperatura, estão expandindo a área de transmissão da dengue para regiões antes consideradas inadequadas à sua propagação. Essa expansão geográfica da dengue é praticamente irreversível, com áreas de maior altitude e latitude antes consideradas barreiras naturais tornando-se zonas de alta incidência. Anomalias térmicas prolongadas, mesmo em regiões com histórico de baixa transmissão, estão impulsionando o aumento da dengue Ebi and Nealon [2016], Sousa et al. [2018].

Chuvas excessivas criam ambientes favoráveis à proliferação do vetor, com acúmulo de água em recipientes naturais e artificiais, especialmente em locais com saneamento precário. Por outro lado, períodos prolongados de estiagem induzem a população a armazenar água em reservatórios improvisados, aumentando o número de criadouros e, consequentemente, o risco de transmissão Lowe et al. [2021], Colón-González et al. [2013]. Embora chuvas intensas possam destruir criadouros e reduzir temporariamente a densidade larval, há consenso de que a precipitação acumulada mantém relação positiva com o aumento dos casos de dengue, indicando que tanto o excesso quanto a escassez de chuva desempenham papel determinante na dinâmica da doença.

A temperatura constitui outro fator climático decisivo na transmissão. O aumento térmico acelera o ciclo de vida do mosquito, intensifica a frequência de picadas e reduz o período de incubação do vírus, ampliando o potencial de transmissão. O desenvolvimento do vetor é mais eficiente entre 25 °C e 30 °C, enquanto temperaturas abaixo de 10 °C inviabilizam a eclosão e o crescimento das larvas. A faixa térmica ideal para a transmissão do vírus DENV situa-se entre 32 °C e 33 °C Liu et al. [2023], Barcellos et al. [2024a]. O calor também reduz o período de incubação extrínseco (EIP) do vírus dentro do mosquito, o que favorece a disseminação. Estudos indicam que o EIP, quando suscetível a temperaturas menores que 30°C, pode acelerar a transmissão da dengue Liu et al. [2023]. No entanto, temperaturas muito elevadas, acima de 36 °C, reduzem a sobrevivência dos mosquitos e podem provocar alterações estruturais irreversíveis no DENV-2 entre 31 °C e 35 °C, comprometendo sua infectividade Liu et al. [2023]. Assim, o regime térmico atua como um modulador da transmissão, com efeitos não lineares e fortemente dependentes da interação entre variáveis ambientais e biológicas.

Diante desse cenário, e considerando a crescente influência das mudanças climáticas sobre a incidência de doenças tropicais, diversas iniciativas globais têm buscado mitigar os impactos ambientais e promover o desenvolvimento sustentável. Entre elas, destacam-se os Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU), especialmente: ODS 3 – Saúde e Bem-Estar, com foco no objetivo 3.3, que visa

“acabar com as epidemias de AIDS, tuberculose, malária e doenças tropicais negligenciadas”, alinhando-se diretamente aos esforços de monitoramento e previsão da dengue; ODS 4 – Educação de Qualidade, em seu objetivo 4.7 (alínea b), que incentiva a promoção da educação científica e ambiental necessária para o desenvolvimento sustentável; ODS 13 – Ação Contra a Mudança Global do Clima, especificamente o objetivo 13.2, que orienta a integração das medidas de mudança climática nas políticas, estratégias e planejamentos nacionais; e ODS 16 – Paz, Justiça e Instituições Eficazes, com destaque para o objetivo 16.6, que propõe desenvolver instituições eficazes, responsáveis e transparentes, fundamentais para o fortalecimento dos sistemas de vigilância epidemiológica ODS. Nesse contexto, técnicas de aprendizado de máquina (Machine Learning – ML) têm emergido como ferramentas poderosas na modelagem e previsão de fenômenos ambientais complexos, sendo aplicadas em estudos sobre qualidade do ar, poluição atmosférica, desmatamento, monitoramento da biodiversidade e previsão de eventos climáticos extremos Ravindiran et al. [2023], Christin et al. [2019].

O presente estudo aplica técnicas de aprendizado de máquina para previsão semanal de dengue com base em dados históricos (2001–2024), visando identificar padrões e períodos críticos de transmissão, considerando variáveis climatológicas, projeções demográficas e a variável auxiliar Pressão de Transmissão, que incorpora o atraso biológico associado ao ciclo do vetor. As séries temporais foram processadas com janelas deslizantes e normalização, e diferentes técnicas foram adotadas, incluindo Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR), Support Vector Regressor (SVR) e K-Nearest Neighbors (KNN). Embora o foco seja o estado de Mato Grosso, a abordagem proposta apresenta potencial de generalização para outras regiões tropicais, oferecendo uma ferramenta robusta para o suporte à vigilância epidemiológica.

Em síntese, este trabalho reforça a importância de se abordar problemas de saúde pública como a dengue sob uma perspectiva interdisciplinar, integrando conhecimentos de epidemiologia, climatologia e ciência de dados. A análise aqui apresentada contribui para o avanço da modelagem preditiva aplicada à saúde ambiental e fornece subsídios técnicos e científicos para políticas públicas voltadas à mitigação dos efeitos das mudanças climáticas e à redução dos riscos associados à dengue em regiões tropicais. Ressalta-se, ainda, que a introdução da vacinação tende a modificar o cenário epidemiológico ao reduzir a suscetibilidade populacional, devendo ser considerada futuramente como variável interveniente nos modelos, uma vez que pode alterar padrões históricos de transmissão e influenciar a interpretação das previsões.

3.3 Materiais e Métodos

3.3.1 Área de Estudo

O presente trabalho tem como objeto de estudo os municípios do estado de Mato Grosso que apresentaram a maior incidência de casos de dengue. A classificação climática de Köppen aponta dois principais tipos no estado: tropical de savana (Aw) e tropical de monção (Am) Alvares et al. [2013]. A topografia do estado varia de planícies alagadas no Pantanal a planaltos no Cerrado, influenciando diretamente as condições climáticas. A Figura 3.1 destaca as localizações de estudo do presente trabalho.

Mapa de Localização dos Municípios de Mato Grosso

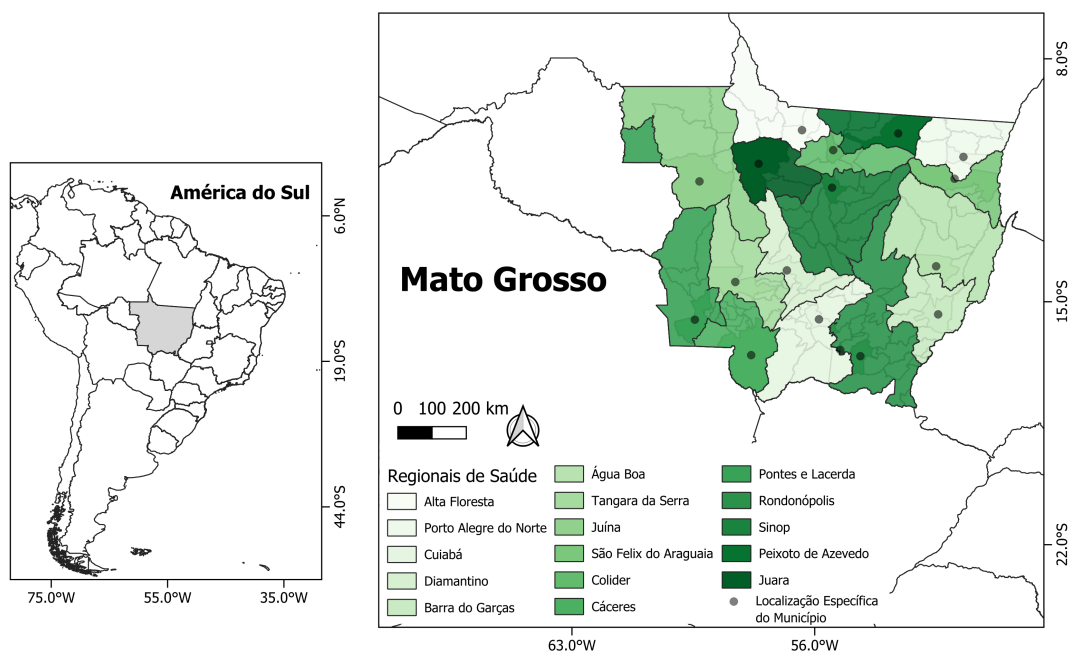


Figura 3.1: Mapa com os município de Cuiabá, Rondonópolis, Barra do Garças, Cáceres, Juína, Porto Alegre do Norte, Sinop, Tangará da Serra, Diamantino, Alta Floresta, Juara, Peixoto de Azevedo, Água Boa, Pontes e Lacerda, Colíder, e São Félix do Araguaia.

O estudo concentrou-se nos escritórios municipais das regionais de saúde de Mato Grosso [de Estado de Saúde MT] (Figura 3.2) por diversos motivos. Em primeiro lugar, essas regionais representam unidades administrativas e epidemiológicas fundamentais para a gestão da saúde pública, o que garante acesso a dados padronizados e de maior confiabilidade. Além disso, ao agrupar os municípios por regionais, é possível identificar variações locais na transmissão da dengue, influenciadas por fatores climáticos, sem perder a clareza

e a manejabilidade da análise. Essa escolha também dialoga diretamente com a estrutura organizacional adotada pela Secretaria Estadual de Saúde de Mato Grosso, possibilitando que os resultados tenham aplicação prática no fortalecimento da vigilância regional, na formulação de estratégias de intervenção e na alocação de recursos. A Figura 3.3, que apresenta a distribuição de casos de dengue por município dentro de cada regional, reforça a pertinência dessa abordagem, assegurando tanto o rigor científico quanto a relevância prática para o planejamento em saúde pública.

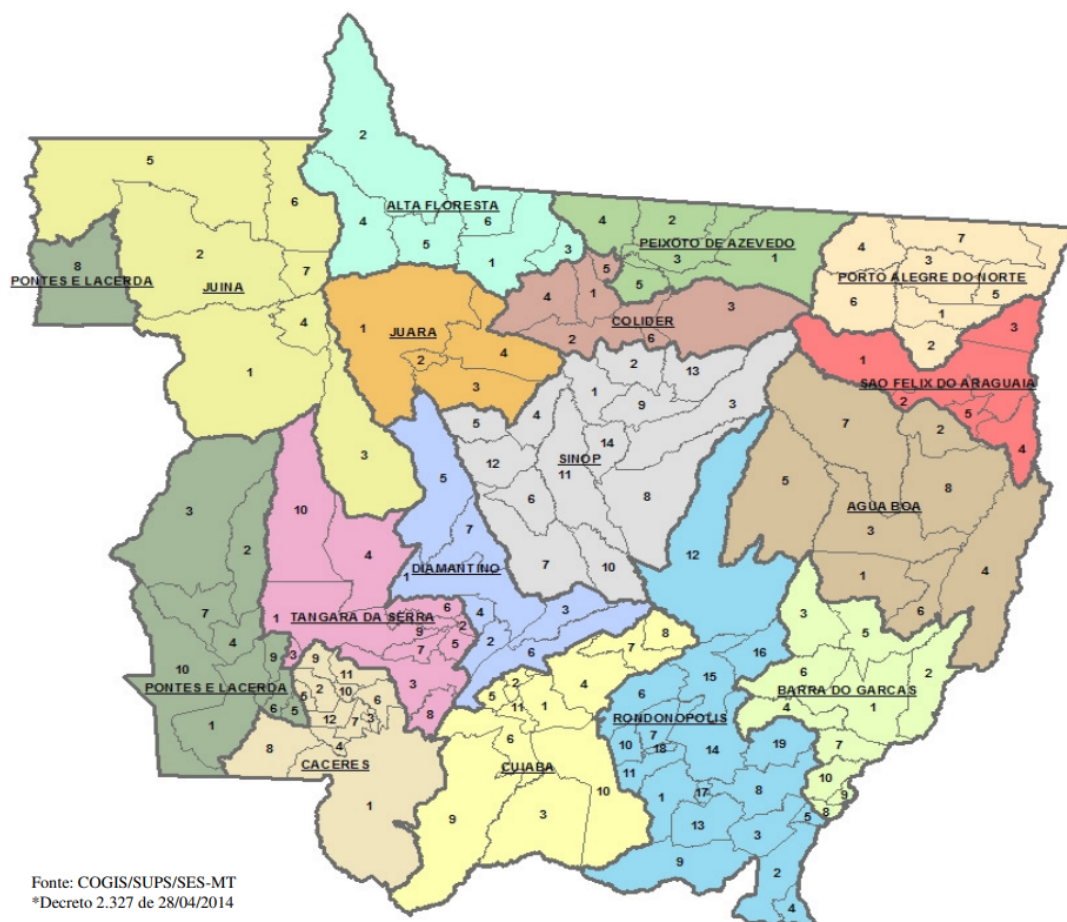


Figura 3.2: Mapa das regionais de saúde do estado de Mato Grosso.

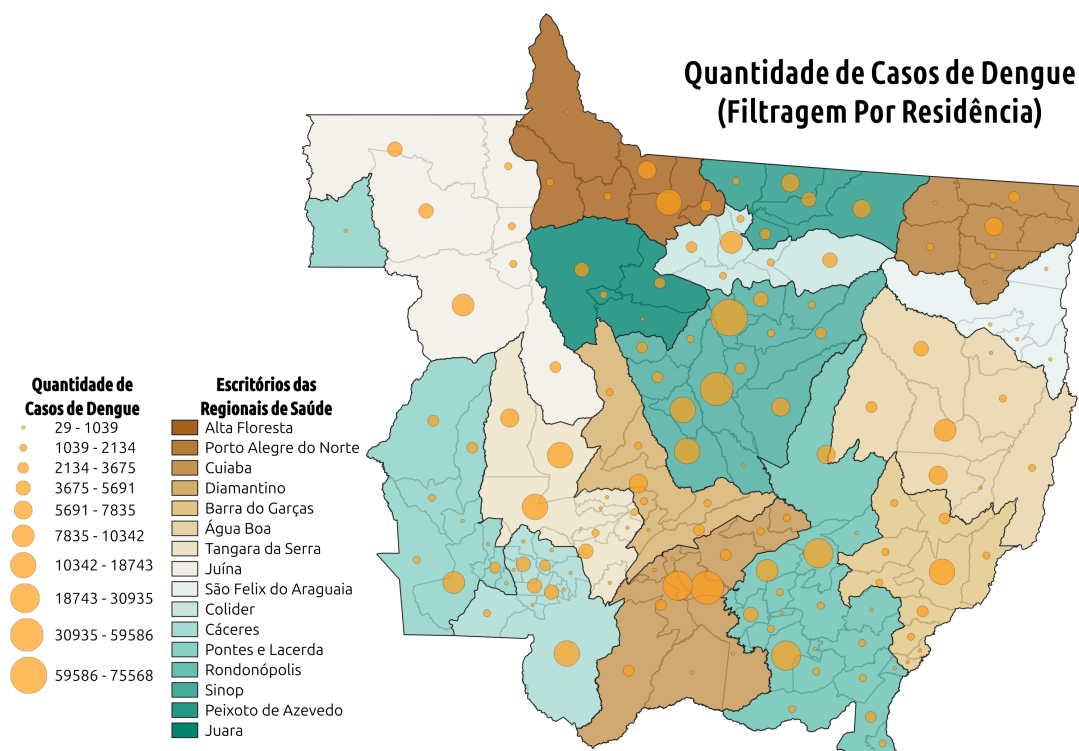


Figura 3.3: Mapa da quantidade de casos de dengue por município de acordo com a regional de saúde. Fonte: Autoral

3.3.2 Coleta e Pré-Processamento de Dados

Os dados foram coletados de fontes oficiais de saúde e clima:

- **Epidemiológicos:** Casos de dengue de 2001 a 2024 para Cuiabá e Várzea Grande obtidos da Secretaria Estadual de Saúde de Mato Grosso (SES-MT), considerando o endereço de residência dos pacientes de Saúde de Mato Grosso [2025]. O período principal de análise foi 2001–2024, quando os dados apresentam maior consistência sazonal e menor subnotificação.
- **Climáticos:** Variáveis diárias de temperatura média, mínima, máxima e precipitação foram obtidas do ERA5-Land via Google Earth Engine, com resolução aproximada de 11 km Copernicus Climate Change Service (C3S) [2019]. Os dados foram validados para todo o estado, mostrando boa concordância com outros produtos de reanálise Lotufo [2023].

Considerando que a densidade populacional é um fator que influencia na incidência da dengue, a evolução da população dos municípios avaliados foi estimada com base nos censos de 2010 e 2022 do IBGE. A opção por utilizar exclusivamente esses dois levantamentos se deve ao fato de que o Censo constitui a única fonte oficial com cobertura

completa e homogênea para todos os municípios brasileiros, garantindo comparabilidade entre localidades. Embora existam estimativas intercensitárias, elas apresentam incertezas maiores e, em alguns casos, discrepâncias em relação aos valores censitários consolidados. Dessa forma, o uso apenas dos dados de 2010 e 2022 minimiza vieses demográficos, assegura consistência metodológica entre os municípios e permite uma interpolação linear mais robusta para estimar a população anual utilizada no cálculo da incidência de dengue. Para isso, foi utilizado um modelo de regressão linear simples, ajustado pela equação:

$$P(t) = mt + b, \quad (3.1)$$

onde $P(t)$ é a população no ano t , m é a taxa de crescimento populacional e b o coeficiente linear ajustado.

Na Tabela 3.1 são mostradas as populações nos anos dos censos, bem como os parâmetros m e b obtidos para cada município avaliado.

Tabela 3.1: Informações sobre as populações das localidades estudadas e período considerado. *Ano inicial* e *Ano final* referem-se ao intervalo dos dados anuais de dengue disponíveis para cada município.

Município	2010	2022	m	b	Ano inicial	Ano final
Água Boa	20.844	29.219	697.92	-1381968.50	2001	2024
Alta Floresta	49.164	58.613	787.42	-1533543.50	2001	2024
Barra do Garças	56.560	69.210	1054.17	-2062315.00	2001	2024
Cáceres	87.942	89.681	144.92	-203340.50	2001	2024
Colider	30.864	31.370	42.17	-53891.00	2001	2024
Cuiabá	551.350	650.877	8293.92	-16119422.50	2001	2024
Diamantino	20.341	21.941	133.33	-247659.00	2001	2024
Juara	32.769	34.906	178.08	-325178.50	2001	2024
Juína	39.255	45.869	551.17	-1068590.00	2001	2024
Peixoto de Azevedo	30.762	32.714	162.67	-296198.00	2001	2024
Pontes e Lacerda	41.386	52.018	886.00	-1739474.00	2001	2024
Porto Alegre do Norte	10.748	12.127	114.92	-220234.50	2001	2024
Rondonópolis	195.476	244.911	4119.58	-8084886.50	2001	2024
São Felix do Araguaia	10.531	13.621	257.50	-507044.00	2001	2024
Sinop	111.676	196.312	7053.00	-14064854.00	2001	2024
Tangara da Serra	83.431	106.434	1916.92	-3769571.50	2001	2024

Para quantificar a propagação de dengue, a incidência foi calculada com base no número de casos registrados e na população estimada para cada ano no período analisado. A incidência (I) representa o número de casos por 100.000 habitantes e foi obtida pela equação,

$$\text{Incidência} = \frac{\text{Número de Casos}}{\text{População Estimada}} \times 100,000. \quad (3.2)$$

Com o objetivo de modelar a dinâmica de transmissão da dengue, adotou-se uma versão simplificada do modelo epidemiológico SIR (Suscetível–Infetado–Recuperado). No entanto, neste trabalho, o compartimento de recuperados foi desconsiderado, resultando em um modelo com apenas dois grupos: infectados e suscetíveis. Essa simplificação foi necessária devido à ausência de dados sobre o período infeccioso e à complexidade de estimar a imunidade populacional frente à circulação de múltiplos sorotipos do vírus da dengue.

Dessa forma, considera-se que um indivíduo está em uma de duas condições: ou está infectado, contribuindo para a incidência de casos, ou está suscetível à infecção. Ambos os cálculos foram feitos considerando uma população de referência de 100.000 habitantes.

A proporção de suscetíveis (Sh) foi então estimada como o complementar da incidência:

$$Sh(t) = \left(1 - \frac{\text{Incidência}(t)}{100.000}\right) \times 100.000. \quad (3.3)$$

Essa abordagem simplificada permite calcular diretamente a pressão de transmissão,

$$Pt(t) = I(t)Sh(t), \quad (3.4)$$

a partir da incidência observada, sem a necessidade de modelar o total da população ou o tempo de recuperação. Essa nova variável, no contexto do modelo SIR, representa a interação potencial entre indivíduos suscetíveis e infectados que pode levar a novas infecções. Pode ser interpretada como um indicador da intensidade potencial de transmissão, ou seja, quanto maior o número de infectados e suscetíveis, maior a probabilidade de novos casos, servindo como uma aproximação para a força da epidemia, mesmo sem modelar diretamente a população de mosquitos.

Para representar os efeitos do tempo no processo de transmissão da dengue, foi criada uma defasagem de 3 semanas na variável de pressão de transmissão. Isso significa que, ao analisar uma semana, estamos considerando a pressão de transmissão que ocorreu três semanas antes. Essa escolha de 3 semanas foi baseada no tempo médio necessário para que todo o ciclo de transmissão da dengue ocorra: desde que uma pessoa é infectada e começa a apresentar sintomas, passando pela transmissão do vírus ao mosquito, até que esse mosquito se torne capaz de infectar outra pessoa. Todo esse processo leva, em média, cerca de 22 dias, o que justifica o uso de uma defasagem de aproximadamente 21 dias (ou 3 semanas).

Na prática, a variável $Pt(t)$ representa a intensidade de interação entre pessoas in-

fectadas e suscetíveis três semanas antes da semana analisada. Isso é importante porque a dengue tem uma transmissão com atraso temporal: os casos registrados hoje foram, em parte, consequência de interações que ocorreram semanas atrás. Ou seja, pessoas infectadas há três semanas transmitiram o vírus aos mosquitos naquela época. Após o chamado período de incubação extrínseco (dentro do mosquito), esses mosquitos se tornaram capazes de infectar outras pessoas. As novas pessoas infectadas, por sua vez, passaram por um período de incubação intrínseco (dentro do corpo humano) antes de apresentarem sintomas e serem contabilizadas nos casos da semana atual. Assim, a defasagem de 3 semanas permite capturar esse efeito retardado da transmissão da dengue. A Figura 3.4 mostra o esquema da transmissão conforme descrito.

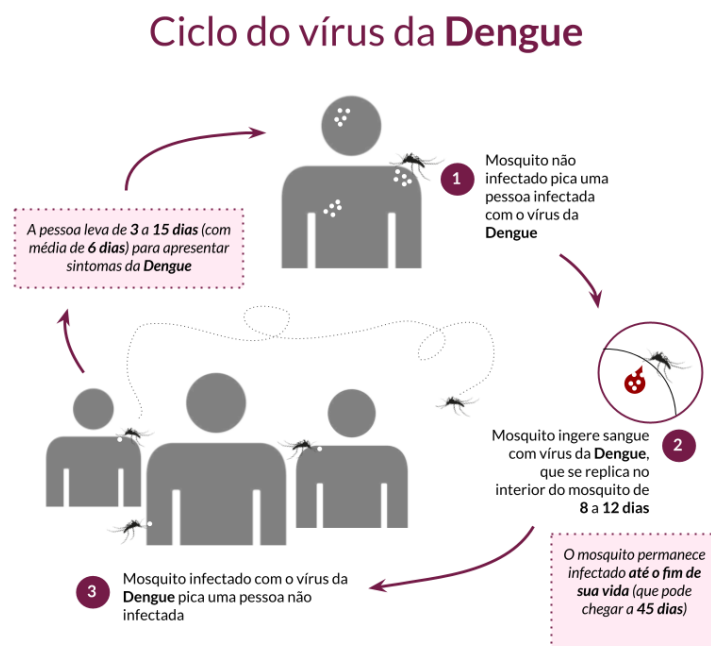


Figura 3.4: Esquema do ciclo de transmissão do vírus da dengue Secretaria da Saúde Gov RS.

Para reduzir variações semanais e suavizar a série temporal de incidência de dengue, foi aplicada uma média móvel de 14 semanas. Esse período foi escolhido com base no ciclo epidemiológico típico da doença em regiões tropicais, que inclui os períodos de incubação do vírus no mosquito e no humano, bem como variações sazonais relacionadas a fatores climáticos. A janela de 14 semanas permite capturar padrões sazonais e picos epidêmicos de forma eficiente, sem comprometer a resolução temporal necessária para análise preditiva.

Estudos anteriores indicam que a ocorrência de casos de dengue pode estar correlacionada com fatores epidemiológicos e climáticos com defasagens de várias semanas. Por exemplo, Shang et al. [2010] observaram que a incidência de casos locais de dengue estava significativamente correlacionada com casos importados e variáveis meteorológicas

com defasagens de 2 a 14 semanas [Shang et al., 2010]. Esses achados justificam a utilização de uma média móvel de 14 semanas, permitindo capturar a dinâmica do vetor e da transmissão da doença.

Para a criação do modelo, foram utilizadas as variáveis listadas na Tabela 3.2.

Tabela 3.2: Variáveis Utilizadas na Modelagem Preditiva.

Variável	Tipo	Abreviação	Unidade
Incidência	Dependente	I	-
Temperatura média diária	Independente	T_mean	°C
Temperatura mínima diária	Independente	T_min	°C
Temperatura máxima diária	Independente	T_max	°C
Precipitação diária	Independente	ppt	mm
Pressão de transmissão	Independente	Pt	-

O modelo foi treinado para considerar uma semana de dados anteriores como entrada para prever os casos de dengue na semana seguinte. Essa abordagem permite capturar a evolução temporal da epidemia e atualizar as previsões conforme novos dados são adicionados.

Como as variáveis independentes e dependentes dos conjuntos de dados estavam em escalas diferentes, foi aplicada a seguinte técnica de normalização:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (3.5)$$

a fim de garantir que todos os dados estivessem na mesma escala, facilitando o processo de treinamento dos modelos de aprendizado de máquina. Nessa equação, X representa o valor original, $X_{\min}$ o menor valor observado, $X_{\max}$ o maior valor observado, e X_{scaled} o valor normalizado.

Essa técnica de normalização garante que todas as variáveis de entrada contribuam de forma equivalente durante o processo de aprendizado. Sem essa normalização, variáveis com valores de magnitude maior poderiam dominar o processo de aprendizagem, ocultando a relevância de variáveis com escalas numéricas menores.

Resumindo, o pré-processamento incluiu as seguintes etapas: preenchimento de dias faltantes com valores nulos para o número de casos; estimativa populacional; agrupamento dos casos de dengue por semana; criação da variável Pressão de Transmissão; normalização das variáveis para ajustar as escalas; e aplicação da média móvel.

3.3.3 Modelos de Machine Learning

A área de aprendizado de máquina, um subcampo da Inteligência Artificial, permite que computadores aprendam por meio de algoritmos. Estes são alimentados com grandes volumes de dados, possibilitando a identificação de padrões e correlações entre variáveis, o que permite realizar previsões ou tomar decisões de forma autônoma Helm et al. [2020].

Este estudo tem como objetivo criar um modelo de predição tendo a incidência de casos de dengue como variável de saída, utilizando variáveis ambientais e pressão de transmissão como entradas. Para alcançar esse resultado, foram implementados algoritmos de regressão supervisionada, onde a variável de saída a ser prevista é contínua, ou seja, o objetivo é estimar um valor numérico Devulapalli [2015]. Esse método consiste em fornecer ao modelo um conjunto de dados previamente conhecidos, nos quais tanto as características de entrada quanto as saídas desejadas são informadas Vicentinia and Oliveira-Filho [2021]. Esse processo iterativo permite que o modelo aprenda a relação entre as variáveis e melhore sua capacidade de prever dados desconhecidos, como ilustrado na Figura 3.5. Por fim, a qualidade do modelo é avaliada utilizando dados que ele nunca viu antes, quantificando seu desempenho por meio de métricas estatísticas.

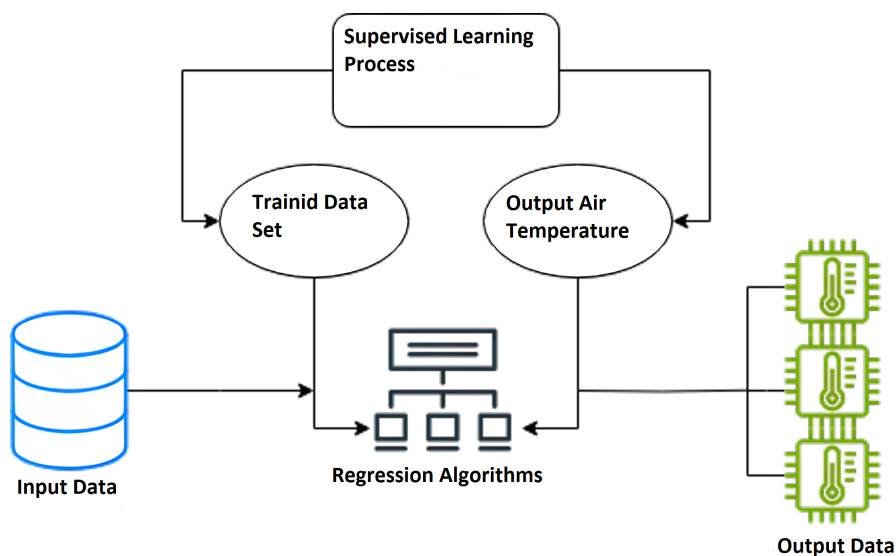


Figura 3.5: Processo de aprendizagem supervisionada.

Para a modelagem preditiva, foram empregados os seguintes algoritmos de *Machine Learning*:

1. O *RandomForestRegressor* (RFR) é um meta-estimador que ajusta diversas "Árvores aleatórias" em subamostras aleatórias do conjunto de dados e combina suas previsões por média para melhorar a acurácia e controlar o sobreajuste. As decisões de cada árvore se baseiam em valores provenientes de um vetor amostrado

aleatoriamente. Esse vetor é escolhido de forma independente e a partir da mesma distribuição para todas as árvores da floresta, como ilustrado na Figura 3.6. À medida que o número de árvores na floresta aumenta, o erro de generalização converge quase certamente para um limite. Esse erro, em uma floresta baseada em árvores, depende da força individual de cada árvore e da correlação entre elas Friedman [2000]. Suas principais vantagens incluem a significativa redução do risco de sobreajuste em comparação a árvores isoladas e a boa capacidade de lidar com alta dimensionalidade, com relevâncias de variáveis estimadas automaticamente.

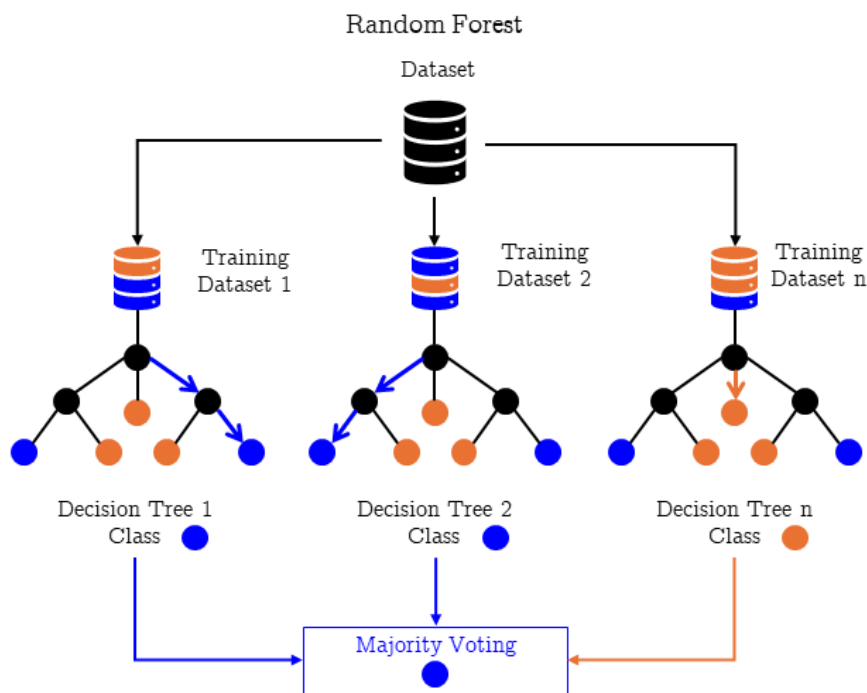


Figura 3.6: Método de *random forest*.

2. O *GradientBoostingRegressor* (GBR), em vez de se concentrar exclusivamente no espaço de parâmetros, explora a otimização numérica no espaço das funções. Essa abordagem estabelece uma conexão única entre expansões aditivas em etapas e a minimização por descida do gradiente mais acentuado. Assim, ele constrói um modelo aditivo em etapas, ajustando em cada estágio uma árvore de regressão ao gradiente negativo da função de perda especificada, como mostrado no Figura 3.7. Ele suporta diversas funções de perda, tornando-o flexível para diferentes perfis de erro. A cada nova árvore, o método corrige os resíduos remanescentes do conjunto anterior,

melhorando progressivamente a precisão. Suas vantagens residem na flexibilidade para otimização de várias funções de perda e na robustez ao sobreajuste, mesmo para grandes números de estimadores.

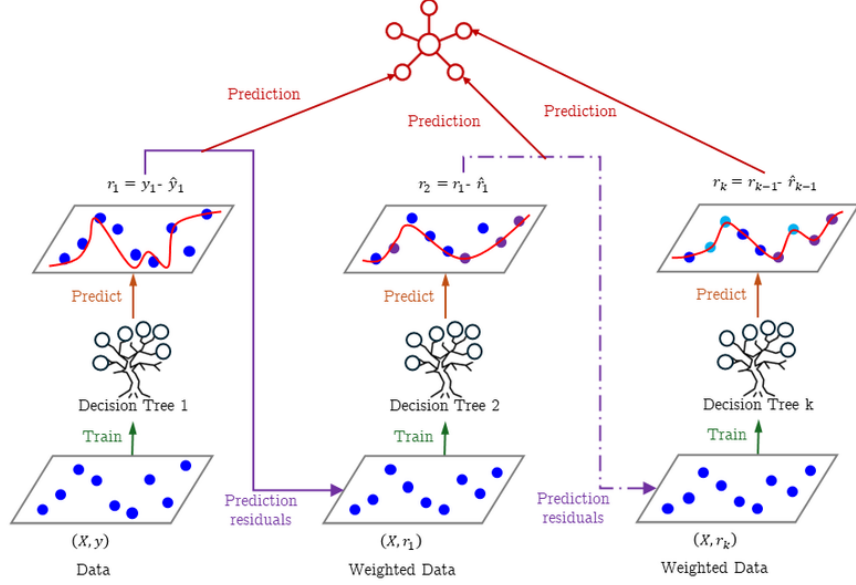


Figura 3.7: Método de *gradient boosting*.

3. O *Support Vector Regressor* (SVR) tem como objetivo encontrar um hiperplano de separação com a maior margem possível entre as classes, o que as torna robustas e eficazes em diversas aplicações. As SVRs conseguem lidar com dados que não são linearmente separáveis por meio do uso de funções kernel, como ilustrado na Figura 3.8. O principal propósito das funções kernel é permitir encontrar um hiperplano em um espaço de maior dimensão onde os pontos de dados possam ser linearmente separáveis, mesmo que não o sejam no espaço original das variáveis. Em outras palavras, os kernels permitem que as SVMs capturem relações complexas e fronteiras de decisão entre os dados que vão além da separação linear simples Boser et al. [1996]. A complexidade de treinamento cresce mais que quadraticamente com o número de amostras, o que pode impactar a performance em conjuntos muito grandes. É eficaz em espaços de alta dimensão e utiliza apenas os vetores de suporte na predição.

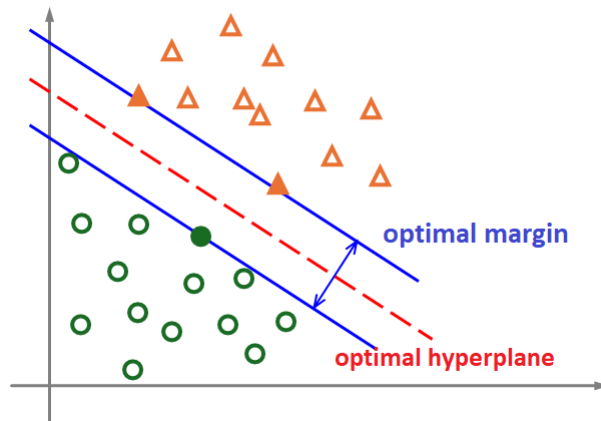


Figura 3.8: Método do *support vector regressor*.

4. O *K-Nearest Neighbors Regressor* (KNN) prevê novos valores calculando a média (ou interpolação ponderada) dos alvos dos k vizinhos mais próximos no espaço de características, como pode ser visto na Figura 3.9. Ele armazena todo o conjunto de treinamento e não exige fase de ajuste de modelo explícito. Esse método é não paramétrico e flexível, sem pressuposições sobre a distribuição dos dados. Suas vantagens incluem simplicidade, interpretabilidade e adaptação local a variações nos dados.

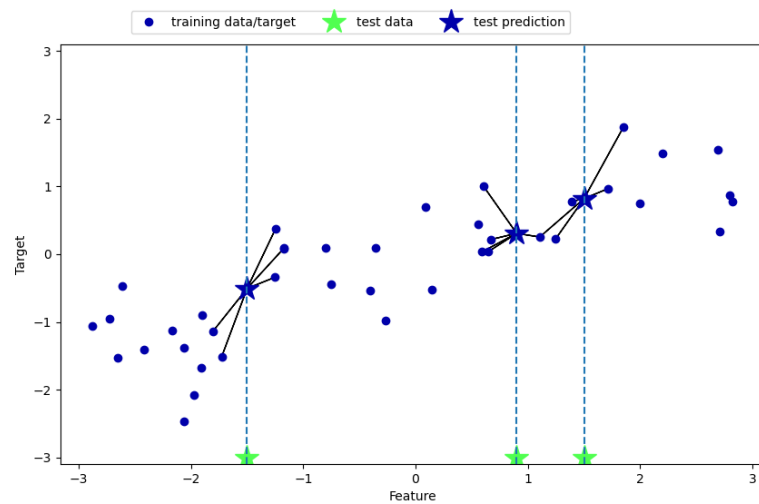


Figura 3.9: Método de *KNN*.

O processo de predição em aprendizado de máquina ocorre em duas etapas distintas. Na fase de treinamento, o algoritmo recebe os dados de entrada juntamente com as saídas esperadas. Já na fase de teste, apenas os dados de entrada são fornecidos, e cabe ao algoritmo gerar as saídas com base nos padrões aprendidos durante cada iteração de tempo (época) do treinamento. Em ambas as etapas, o desempenho pode ser avaliado por

meio da análise de métricas, com o objetivo de verificar a precisão do algoritmo e evitar problemas como sobreajuste (*overfitting*) e subajuste (*underfitting*).

Os dados foram divididos cronologicamente em 80% para o treinamento do modelo e 20% para o teste, respeitando a sequência temporal dos dados. Essa abordagem evita o vazamento de informações e segue o conceito de séries temporais, no qual observações passadas são utilizadas para prever eventos futuros. O conjunto de teste é fundamental para avaliar a qualidade do modelo por meio de métricas de desempenho.

3.3.4 Avaliação dos Modelos

Para mensurar a performance dos modelos preditivos, foram utilizadas as seguintes métricas,

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (3.6)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (3.7)$$

$$d = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (|y_i - \bar{y}| + |\hat{y}_i - \bar{y}|)^2}, \quad (3.8)$$

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}}, \quad (3.9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (3.10)$$

onde y_i é o valor observado no i -ésimo ponto de dados; $\hat{y}_i$ é o valor predito pelo modelo para o i -ésimo ponto de dados; $\bar{y}$ é a média dos valores observados y_i ; n é o número total de observações. A Raiz do Erro Quadrático Médio (RMSE) calcula a raiz quadrada do erro quadrático médio, penalizando erros maiores, sendo sensível a grandes desvios. O Erro Absoluto Médio (MAE) mede, em média, a diferença absoluta entre os valores previstos e os observados, com valores menores indicando maior precisão. O Índice de Concordância de Willmott é uma métrica de acordo que avalia a correspondência entre os valores previstos e os observados, onde valores próximos de 1 indicam alta concordância Willmott et al. [1985]. O índice de correlação de Pearson (r) mede a intensidade e a relação linear entre duas variáveis, assumindo valores entre -1 e 1. Valores próximos de 1 indicam forte correlação positiva, enquanto valores próximos de -1 indicam forte correlação negativa. O Coeficiente de Determinação (R^2) indica a proporção da variância dos dados que é explicada pelo modelo, com valores mais próximos de 1 indicando melhor desempenho. Essas métricas permitiram comparar a eficácia dos modelos, identificando

aquele com melhor capacidade de previsão e maior robustez.

A contribuição de cada variável foi avaliada pelo método SHAP (*SHapley Additive exPlanations*), que se baseia na teoria de jogos Mage.ai [2025] Ponce-Bobadilla et al. [2024]. A essência dos valores de Shapley é medir a contribuição de cada "jogador" para o resultado final de uma "coalizão" (um grupo de jogadores trabalhando juntos), garantindo que a soma das contribuições individuais seja igual ao resultado total. Ao aplicar os valores de *Shapley* à explicação de modelos de Machine Learning, as *features* do modelo são tratadas como "jogadores" e a previsão do modelo para uma observação específica é o "resultado" ou "pagamento" do jogo ?. A equação principal para o valor SHAP de uma *feature* j em uma instância x , para um modelo f com n *features* $N = \{1, 2, \dots, n\}$, é dada por:

$$f(x) = \phi_0 + \sum_{j=1}^n \phi_j, \quad \phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|! (n - |S| - 1)!}{n!} [V(S \cup \{j\}) - V(S)], \quad (3.11)$$

onde ϕ_j representa a contribuição da variável j para a previsão $f(x)$ da instância x e $\phi_0 = \mathbb{E}[f(x)]$ representa o valor médio do modelo sem nenhuma *feature*, $S \subseteq N \setminus \{j\}$ e $V(S) = \mathbb{E}[f(x) \mid x_S]$.

3.4 Resultados e Discussão

3.4.1 Previsão Semanal dos Casos de Dengue

A previsão semanal dos casos de dengue foi realizada para os 16 escritórios regionais de saúde do estado de Mato Grosso, pois desempenham papel fundamental na gestão da saúde pública. Para cada município, foram testados diferentes algoritmos de aprendizado de máquina, e os modelos com melhor desempenho no conjunto de teste foram selecionados para análise gráfica e interpretação dos resultados. A Figura 3.10 apresenta os gráficos de dispersão entre os valores previstos e observados para cada município, destacando os melhores modelos obtidos com base no conjunto de teste.

Os modelos de melhor desempenho foram os seguintes: KNN para Água Boa e Cáceres; GBR para Alta Floresta, Barra do Garças e Juína; RFR para Colíder, Cuiabá, Diamantino, Peixoto de Azevedo, Pontes e Lacerda, Porto Alegre do Norte, Rondonópolis, São Félix do Araguaia, Sinop e Tangará da Serra; SVR para Juara.

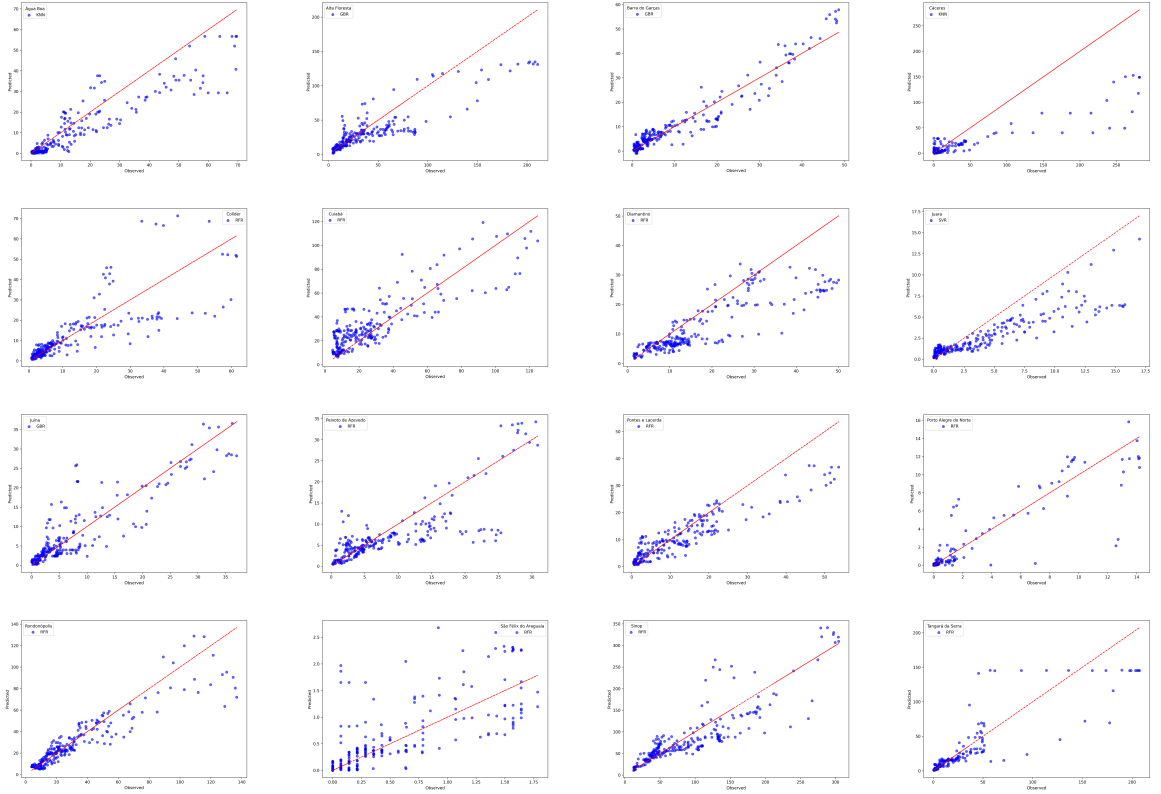


Figura 3.10: Dispersão dos dados previstos pelos medidos para os modelos de Água Boa (KNN), Alta Floresta (GBR), Barra do Garças (GBR), Cáceres (KNN), Colíder (RFR), Cuiabá (RFR), Diamantino (RFR), Juara (SVR), Juína (GBF), Peixoto de Azevedo (RFR), Pontes e Lacerda (RFR), Porto Alegre do Norte (RFR), Rondonópolis (RFR), São Félix do Araguaia (RFR), Sinop (RFR), Tangará da Serra (RFR).

De forma geral, os modelos testados demonstraram bom desempenho preditivo, com resultados satisfatórios tanto no conjunto de treino quanto no conjunto de teste. Isso é evidenciado pela concentração significativa dos pontos ao longo da linha de 45° nos gráficos de dispersão da Figura 3.10, o que indica que as previsões realizadas pelos modelos estiveram, em grande parte, próximas dos valores observados. Esse padrão reflete a capacidade dos algoritmos em capturar as variações semanais nos casos de dengue, mesmo diante de uma série temporal com flutuações bruscas associadas à natureza epidêmica da doença.

Algumas cidades como Água Boa, Cáceres, Colíder e Porto Alegre do Norte chegaram a apresentar um resultado aquém do ideal, apresentando nas métricas de erro valores muito altos e forte overfitting. Uma das possíveis razões é a baixa incidência de dengue dessas regiões, visto que o Sistema Único de Saúde (SUS) classifica a transmissão de dengue em 3 faixas com base em uma população de 100.000 habitantes: baixa (menos de 100 casos), média (de 100 a 300 casos) e alta (mais de 300 casos) Barcellos et al. [2024b]. Como visualizado na Figuras 3.11, essa oscilação entre anos de baixa e alta incidência

dificulta a detecção de padrões temporais consistentes pelos algoritmos de aprendizado de máquina, refletindo a natureza esporádica da transmissão em localidades de menor população.

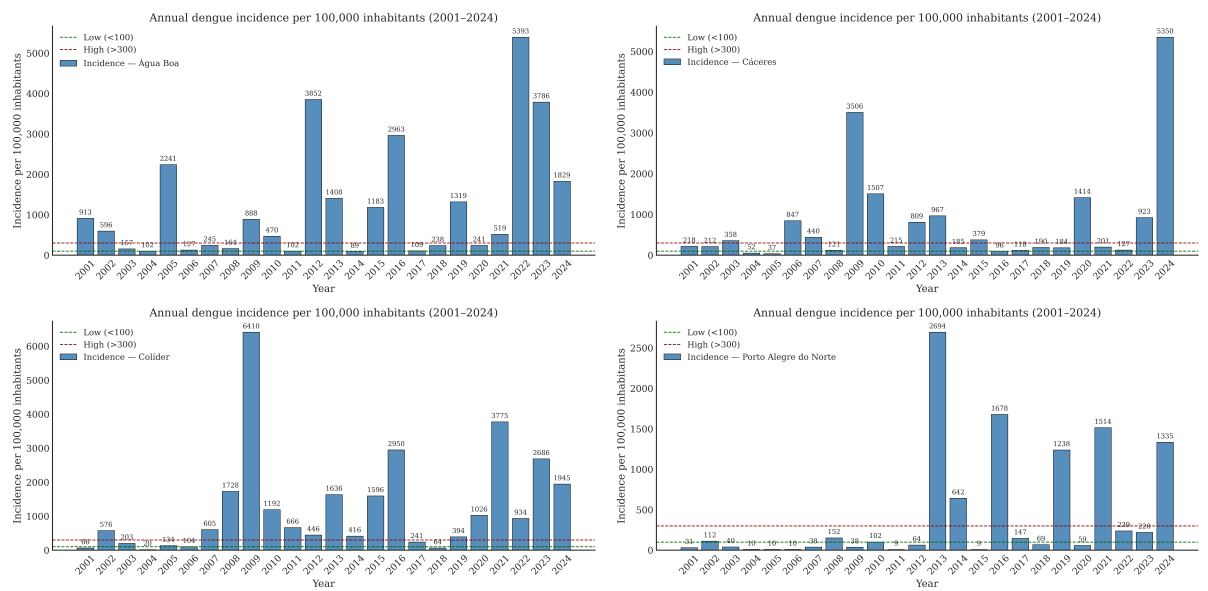


Figura 3.11: Incidência anual dos casos de dengue por 100.000 habitantes das cidades de Água Boa, Cáceres, Colíder e Porto Alegre do Norte.

A Tabela 3.3 sintetiza os resultados quantitativos do desempenho dos modelos em cada município, apresentando as principais métricas utilizadas na avaliação: Erro Quadrático Médio (MSE), Erro Absoluto Médio (MAE), Raiz do Erro Quadrático Médio (RMSE), Coeficiente de Willmott (d), Correlação de Pearson (r) e o Coeficiente de Determinação (R^2). Cada métrica oferece uma perspectiva complementar sobre o desempenho dos modelos, permitindo uma avaliação mais abrangente. Os valores destacados em negrito indicam os melhores resultados obtidos por município no conjunto de treino e teste, servindo como critério objetivo para a escolha dos modelos destacados nos gráficos de dispersão. Essa abordagem multifacetada de avaliação é essencial, visto que nenhuma métrica isolada é suficiente para capturar toda a complexidade do fenômeno preditivo.

Tabela 3.3: Desempenho dos modelos nos conjuntos de treino e teste.

Local	Métricas	Treino				Teste			
		RFR	GBR	SVR	KNN	RFR	GBR	SVR	KNN
Cuiabá	MSE	260.410	188.599	5753.459	613.316	209.703	320.059	393.884	934.459
	RMSE	16.137	13.733	75.852	24.765	14.481	17.890	19.847	30.569
	MAE	5.814	7.657	30.477	11.776	10.834	12.476	13.714	22.186
	Willmott	0.993	0.995	0.681	0.984	0.912	0.862	0.724	0.617
	Pearson	0.988	0.991	0.886	0.969	0.861	0.779	0.745	0.364

(continua na próxima página)

Local	Métricas	Treino				Teste			
		RFR	GBR	SVR	KNN	RFR	GBR	SVR	KNN
	R^2	0.974	0.982	0.436	0.940	0.709	0.556	0.453	-0.298
Rondonópolis	MSE	11.327	19.802	513.660	44.290	138.232	144.653	385.649	225.910
	RMSE	3.366	4.450	22.664	6.655	11.757	12.027	19.638	15.030
	MAE	1.700	2.743	9.711	3.734	6.901	7.160	12.816	10.556
	Willmott	0.998	0.996	0.808	0.991	0.949	0.947	0.812	0.916
	Pearson	0.996	0.992	0.910	0.983	0.924	0.920	0.981	0.921
	R^2	0.991	0.984	0.591	0.965	0.836	0.829	0.543	0.732
Barra do Garças	MSE	11.992	12.230	591.012	29.456	24.371	12.361	73.757	40.506
	RMSE	3.463	3.497	24.311	5.427	4.937	3.516	8.588	6.364
	MAE	1.430	2.061	9.748	2.845	3.014	2.438	6.354	4.395
	Willmott	0.996	0.996	0.536	0.991	0.967	0.981	0.753	0.929
	Pearson	0.993	0.993	0.839	0.983	0.952	0.963	0.876	0.868
	R^2	0.986	0.985	0.293	0.965	0.845	0.921	0.531	0.743
Cáceres	MSE	7.907	9.210	449.106	33.406	1635.455	1709.813	3443.190	1508.292
	RMSE	2.812	3.035	21.192	5.780	40.441	41.350	58.679	38.837
	MAE	1.194	1.634	8.150	2.784	14.924	17.085	21.453	15.373
	Willmott	0.996	0.995	0.400	0.982	0.766	0.751	0.296	0.790
	Pearson	0.993	0.991	0.766	0.969	0.833	0.801	0.891	0.919
	R^2	0.985	0.982	0.145	0.936	0.533	0.511	0.016	0.569
Juína	MSE	3.647	4.479	109.654	10.290	22.163	15.802	35.510	39.191
	RMSE	1.910	2.116	10.472	3.208	4.708	3.975	5.959	6.260
	MAE	0.870	1.307	4.583	1.660	2.729	2.514	3.666	3.857
	Willmott	0.995	0.994	0.673	0.985	0.927	0.945	0.792	0.871
	Pearson	0.991	0.988	0.848	0.973	0.864	0.896	0.914	0.764
	R^2	0.980	0.976	0.412	0.945	0.719	0.799	0.549	0.503
Porto Alegre do Norte	MSE	0.222	0.113	6.886	0.853	2.115	2.153	12.523	5.211
	RMSE	0.471	0.337	2.624	0.924	1.454	1.467	3.539	2.283
	MAE	0.134	0.138	0.759	0.264	0.571	0.601	1.487	0.900
	Willmott	0.992	0.996	0.298	0.969	0.954	0.955	0.376	0.909
	Pearson	0.986	0.993	0.873	0.942	0.912	0.914	0.845	0.848
	R^2	0.971	0.985	0.093	0.888	0.829	0.826	-0.012	0.579
Sinop	MSE	73.358	85.288	2890.070	196.220	1198.785	1342.848	2880.023	2273.483
	RMSE	8.565	9.235	53.759	14.008	34.623	36.645	53.666	47.681
	MAE	3.258	5.297	25.731	6.712	23.722	24.789	37.832	34.923
	Willmott	0.997	0.997	0.779	0.992	0.931	0.925	0.736	0.853
	Pearson	0.995	0.994	0.931	0.985	0.872	0.862	0.848	0.767
	R^2	0.989	0.987	0.556	0.970	0.737	0.706	0.369	0.502
Tangará da Serra	MSE	5.756	8.375	204.249	15.860	428.486	461.852	1519.493	848.473
	RMSE	2.399	2.894	14.292	3.982	20.700	21.491	38.981	29.129
	MAE	1.072	1.666	6.790	2.093	9.072	8.884	17.108	12.256
	Willmott	0.995	0.993	0.610	0.986	0.928	0.921	0.505	0.809
	Pearson	0.991	0.987	0.866	0.975	0.881	0.872	0.871	0.776
	R^2	0.982	0.973	0.346	0.949	0.770	0.753	0.186	0.545
Diamantino	MSE	0.444	0.611	12.384	1.249	71.790	89.677	229.667	79.809
	RMSE	0.666	0.782	3.519	1.118	8.473	9.470	15.155	8.934
	MAE	0.286	0.464	1.705	0.606	5.688	6.232	11.489	7.192
	Willmott	0.995	0.993	0.712	0.985	0.848	0.799	0.562	0.860
	Pearson	0.991	0.987	0.883	0.974	0.866	0.819	0.926	0.893
	R^2	0.981	0.974	0.467	0.946	0.564	0.455	-0.395	0.515

(continua na próxima página)

Local	Métricas	Treino				Teste			
		RFR	GBR	SVR	KNN	RFR	GBR	SVR	KNN
Alta Floresta	MSE	4.467	4.852	138.489	19.321	193.551	167.959	420.349	220.065
	RMSE	2.113	2.203	11.768	4.396	13.912	12.960	20.502	14.835
	MAE	0.947	1.360	4.740	1.991	8.565	7.923	11.588	9.529
	Willmott	0.996	0.996	0.740	0.981	0.874	0.900	0.670	0.885
	Pearson	0.992	0.991	0.866	0.966	0.880	0.892	0.879	0.854
	R^2	0.984	0.983	0.503	0.931	0.680	0.722	0.305	0.636
Juara	MSE	0.692	0.776	18.547	2.169	8.256	8.300	7.228	9.560
	RMSE	0.832	0.881	4.307	1.473	2.873	2.881	2.688	3.092
	MAE	0.387	0.539	1.929	0.747	1.760	1.709	1.741	2.067
	Willmott	0.997	0.996	0.851	0.989	0.844	0.846	0.834	0.817
	Pearson	0.994	0.993	0.912	0.981	0.741	0.741	0.904	0.727
	R^2	0.987	0.986	0.657	0.960	0.535	0.533	0.593	0.462
Peixoto de Azevedo	MSE	0.765	0.924	20.174	1.586	21.880	25.002	43.455	35.939
	RMSE	0.875	0.961	4.492	1.259	4.678	5.000	6.592	5.995
	MAE	0.403	0.602	1.894	0.753	2.913	2.991	4.419	3.868
	Willmott	0.995	0.995	0.770	0.991	0.891	0.881	0.699	0.854
	Pearson	0.992	0.990	0.899	0.982	0.815	0.792	0.914	0.768
	R^2	0.983	0.979	0.539	0.964	0.631	0.579	0.268	0.395
Água Boa	MSE	1.011	0.941	41.340	2.874	156.063	195.098	273.025	65.648
	RMSE	1.005	0.970	6.430	1.695	12.493	13.968	16.523	8.102
	MAE	0.417	0.566	2.407	0.738	6.564	7.667	9.691	5.137
	Willmott	0.996	0.996	0.639	0.988	0.771	0.707	0.565	0.932
	Pearson	0.993	0.993	0.901	0.979	0.776	0.678	0.939	0.924
	R^2	0.985	0.986	0.380	0.957	0.500	0.375	0.125	0.790
Pontes e Lacerda	MSE	0.959	1.469	24.146	2.768	31.946	36.657	57.973	44.004
	RMSE	0.979	1.212	4.914	1.664	5.652	6.055	7.614	6.634
	MAE	0.487	0.792	2.272	0.916	3.639	3.837	4.835	4.244
	Willmott	0.997	0.995	0.872	0.991	0.922	0.907	0.840	0.905
	Pearson	0.994	0.991	0.922	0.983	0.926	0.929	0.937	0.874
	R^2	0.988	0.981	0.692	0.965	0.782	0.750	0.604	0.700
Colíder	MSE	6.910	2.281	248.614	24.074	76.125	144.843	259.168	131.354
	RMSE	2.629	1.510	15.767	4.907	8.725	12.035	16.099	11.461
	MAE	0.834	0.872	5.300	1.824	4.954	6.049	10.002	7.309
	Willmott	0.993	0.998	0.270	0.974	0.886	0.807	0.472	0.723
	Pearson	0.988	0.996	0.765	0.955	0.796	0.680	0.909	0.695
	R^2	0.974	0.991	0.058	0.909	0.608	0.254	-0.335	0.323
São Félix do Araguaia	MSE	0.034	0.031	0.842	0.118	0.197	0.221	0.276	0.350
	RMSE	0.183	0.176	0.918	0.343	0.444	0.470	0.525	0.592
	MAE	0.074	0.097	0.363	0.174	0.280	0.295	0.356	0.414
	Willmott	0.992	0.992	0.466	0.969	0.849	0.844	0.547	0.647
	Pearson	0.985	0.986	0.840	0.943	0.747	0.760	0.780	0.430
	R^2	0.968	0.971	0.209	0.889	0.320	0.238	0.050	-0.207

Observa-se que, em praticamente todos os municípios, exceto São Félix do Araguaia, os modelos que se destacaram apresentaram valores quase perfeitos no conjunto de treino, com R^2 e Pearson próximos de 1. Isso sugere overfitting, confirmado pelo desempenho consideravelmente inferior no teste. Os modelos SVR e KNN, por outro lado,

obtiveram desempenho mais estável ao olhar para essas métricas em específico, mas apresentaram erros mais elevados em comparação ao RFR e o GBR entre treino e teste na maioria dos municípios.

As Figuras 3.12- 3.15 mostram, para os 16 municípios analisados, a evolução temporal dos casos de dengue previstos pelos modelos em comparação com os valores observados, abrangendo tanto o conjunto de treinamento quanto o conjunto de teste. Cada gráfico apresenta a série temporal semanal, permitindo visualizar o desempenho dos modelos ao longo de todo o período considerado, com destaque para a capacidade de adaptação às variações sazonais da doença.

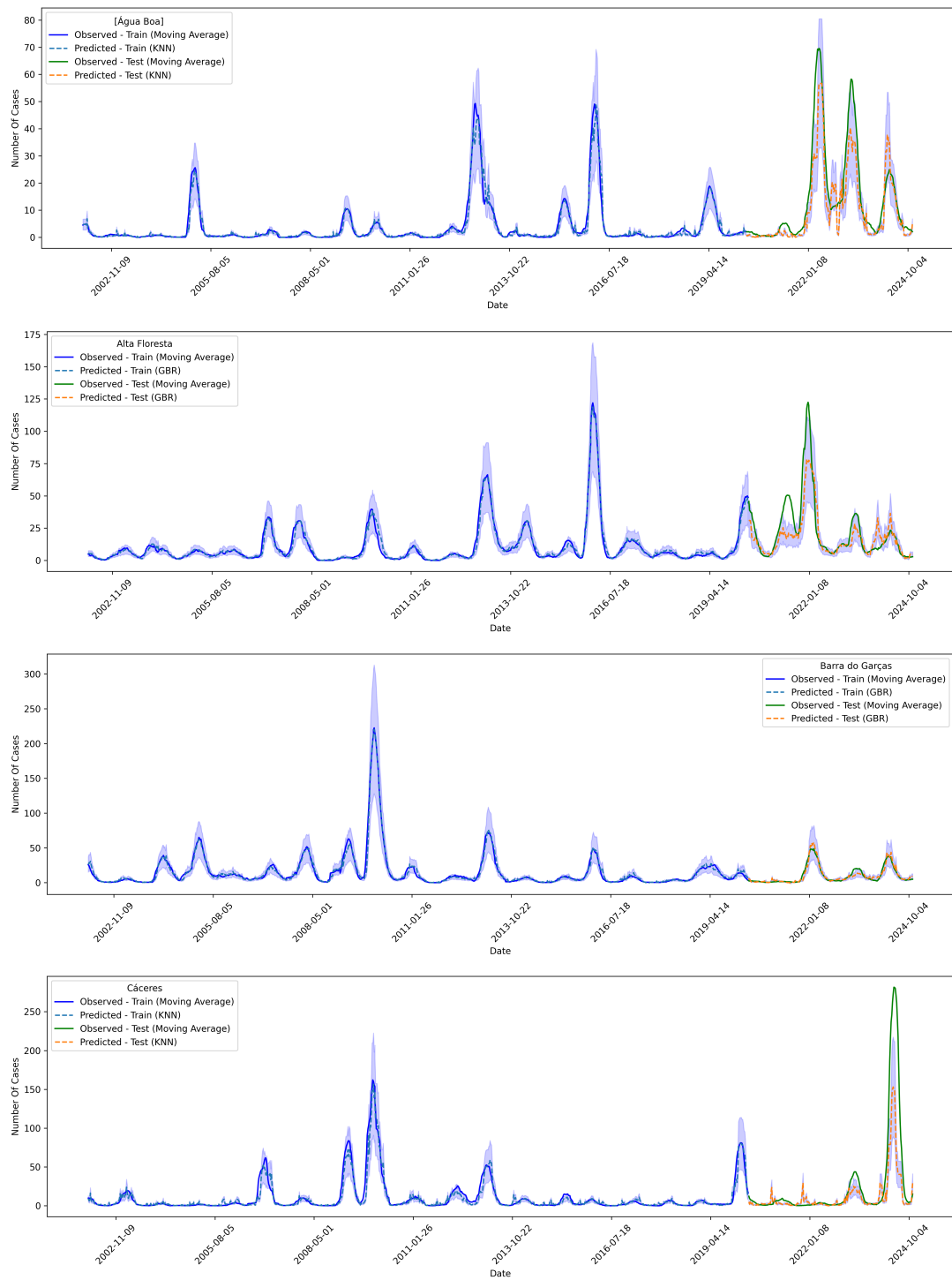


Figura 3.12: Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Água Boa (KNN), Alta Floresta (GBR), Barra do Garças (GBR) e Cáceres (KNN).

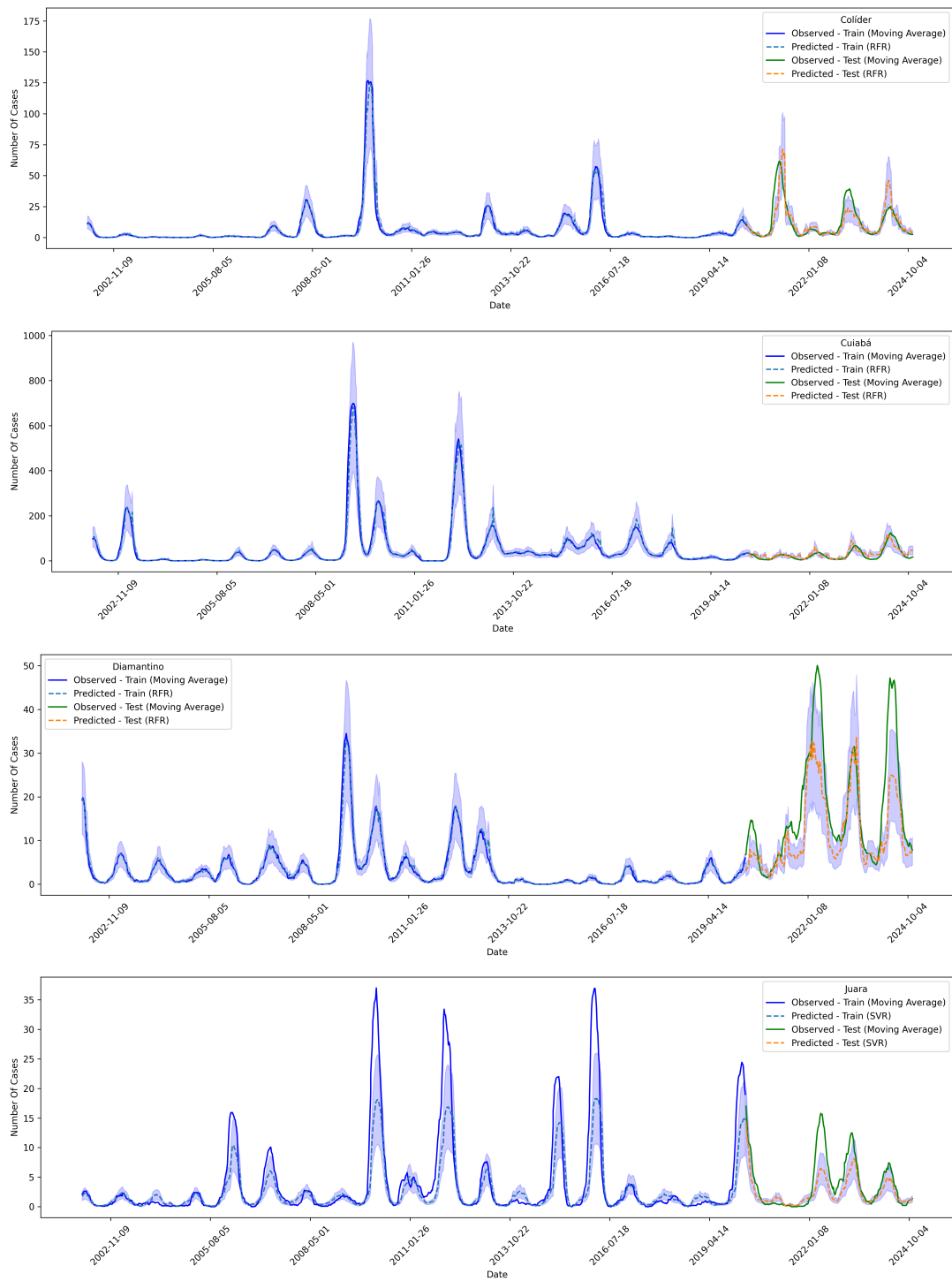


Figura 3.13: Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Colíder (RFR), Cuiabá (RFR), Diamantino (RFR) e Juara (SVR).

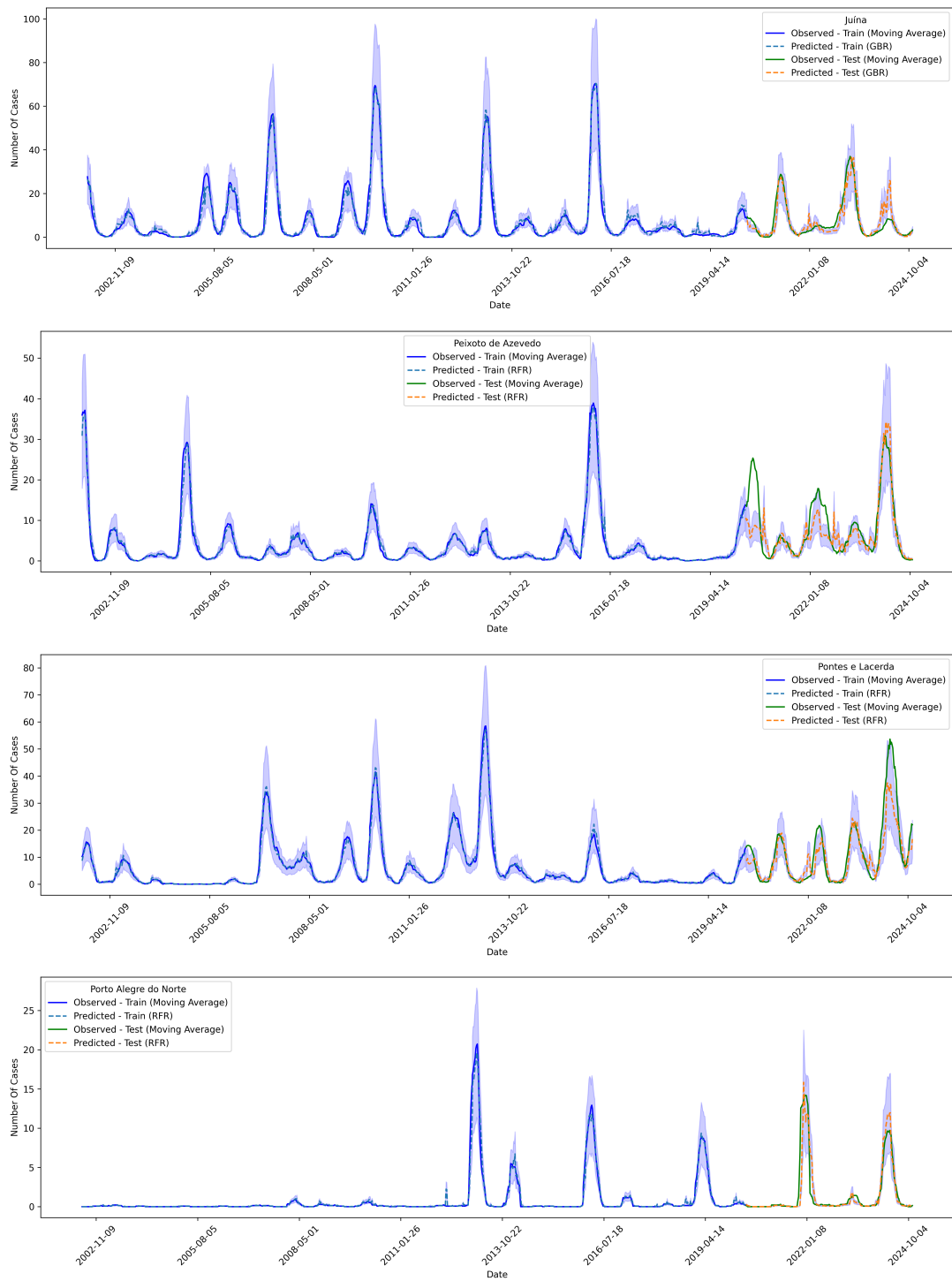


Figura 3.14: Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Juína (GBR), Peixoto de Azevedo (RFR), Pontes e Lacerda (RFR) e Porto Alegre do Norte (RFR).

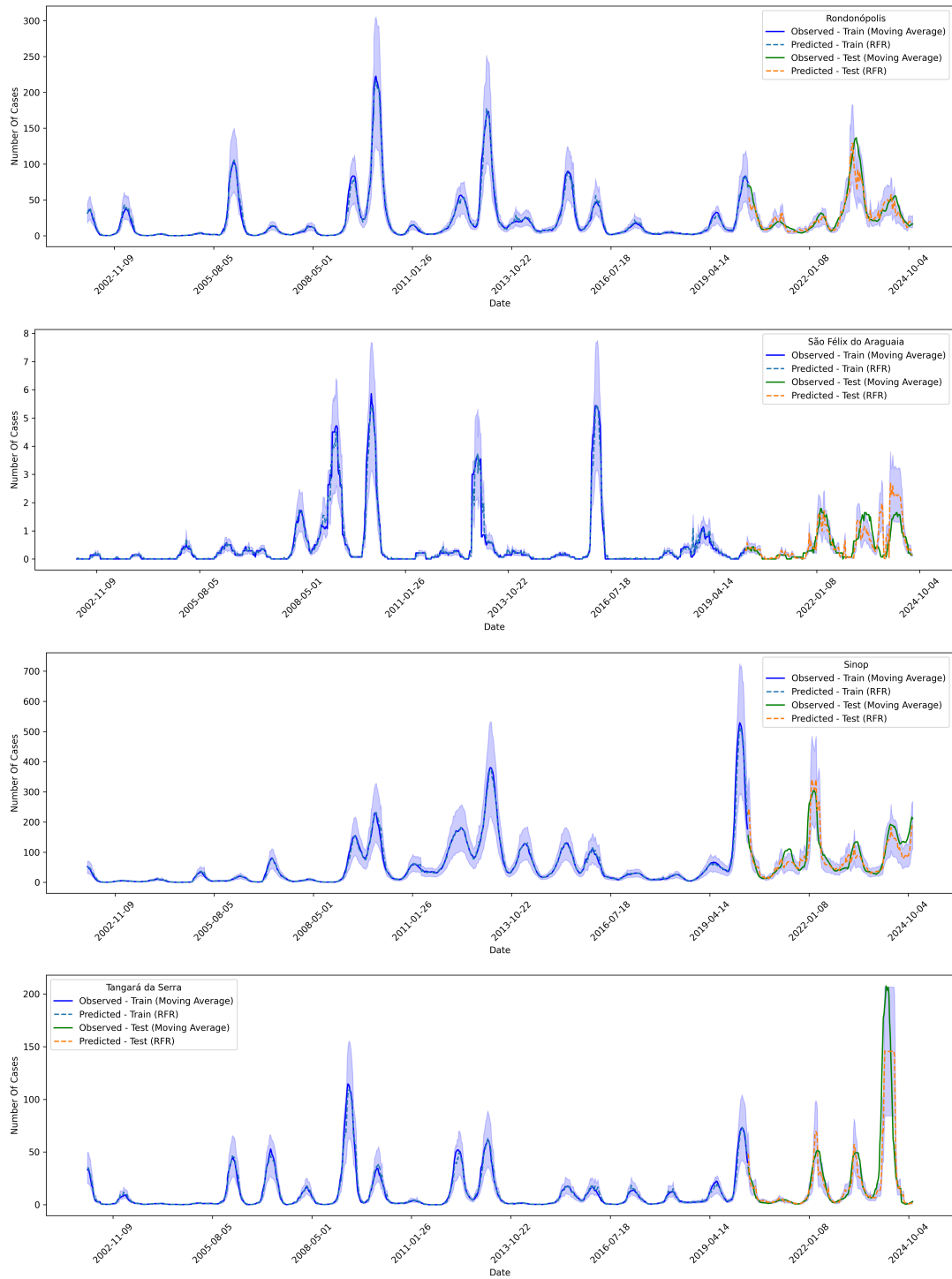


Figura 3.15: Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Rondonópolis (RFR), São Félix do Araguaia (RFR), Sinop (RFR) e Tangará da Serra (RFR).

De maneira geral, observa-se que os modelos conseguem acompanhar de forma satisfatória as flutuações nos dados reais, reproduzindo com boa fidelidade tanto os períodos de baixa incidência quanto os momentos de maior concentração de casos, associados a sur-

tos epidêmicos. A presença de picos acentuados nos gráficos, característica marcante da dinâmica da dengue em regiões tropicais como o estado de Mato Grosso, é capturada com razoável precisão, demonstrando que os modelos foram capazes de internalizar os padrões subjacentes de variação intersemanal.

Além disso, a sazonalidade típica da dengue, com aumentos regulares de casos nos meses mais quentes e chuvosos do ano, foi bem representada nas previsões, indicando que os modelos utilizaram efetivamente os dados climáticos e epidemiológicos como variáveis preditoras relevantes. Isso reforça a ideia de que a interação entre fatores ambientais, como temperatura e precipitação, e o comportamento epidemiológico foi bem assimilada pelos algoritmos de aprendizado de máquina empregados.

A qualidade dos ajustes nos conjuntos de teste é particularmente relevante, pois evidencia a capacidade de generalização dos modelos para dados não vistos, aspecto essencial em aplicações práticas de previsão em saúde pública. Os bons resultados visuais apresentados nessas figuras estão em consonância com os desempenhos quantitativos obtidos e resumidos na Tabela 3.3.

Por fim, as Figuras 3.16 e 3.17 apresentam os gráficos de importância das variáveis obtidos por meio da técnica SHAP (SHapley Additive exPlanations), aplicada ao melhor modelo de cada município. Em todos os casos, a variável Pressão de Transmissão (com defasagem) foi consistentemente a mais relevante, com forte impacto positivo nas previsões. As variáveis meteorológicas também tiveram importância significativa, especialmente a precipitação e a temperatura mínima. Por outro lado, a temperatura média e, em alguns municípios, a temperatura máxima, mostraram baixa contribuição, com valores de SHAP próximos de zero. Isso indica que tais variáveis podem ser descartadas em modelos futuros com pouco prejuízo ao desempenho.

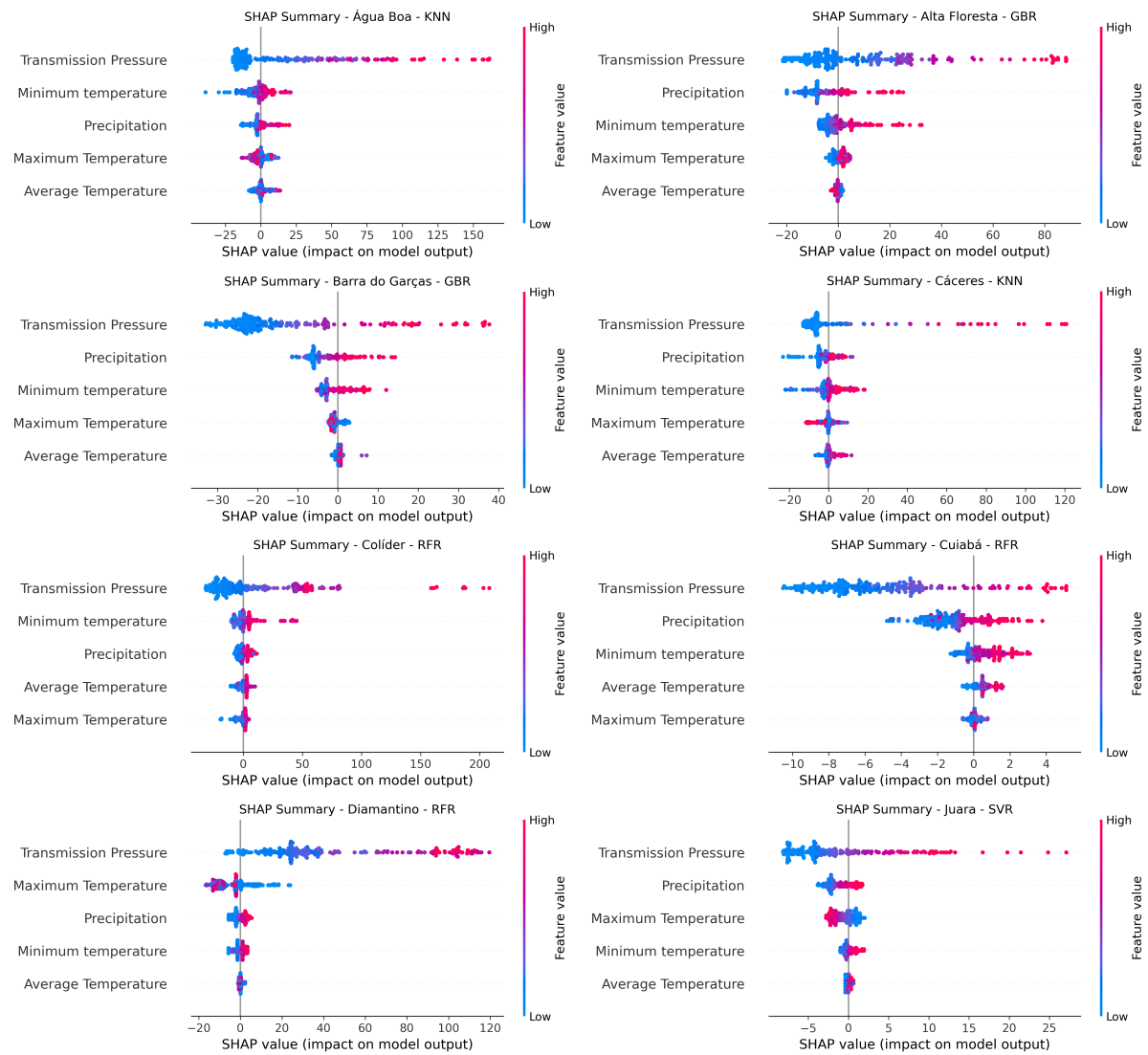


Figura 3.16: Comparação entre as variáveis mais importantes para os modelos de Água Boa (KNN), Alta Floresta (GBR), Barra do Garças (GBR), Cáceres (KNN), Colíder (RFR), Cuiabá (RFR), Diamantino (RFR) e Juara (SVR).

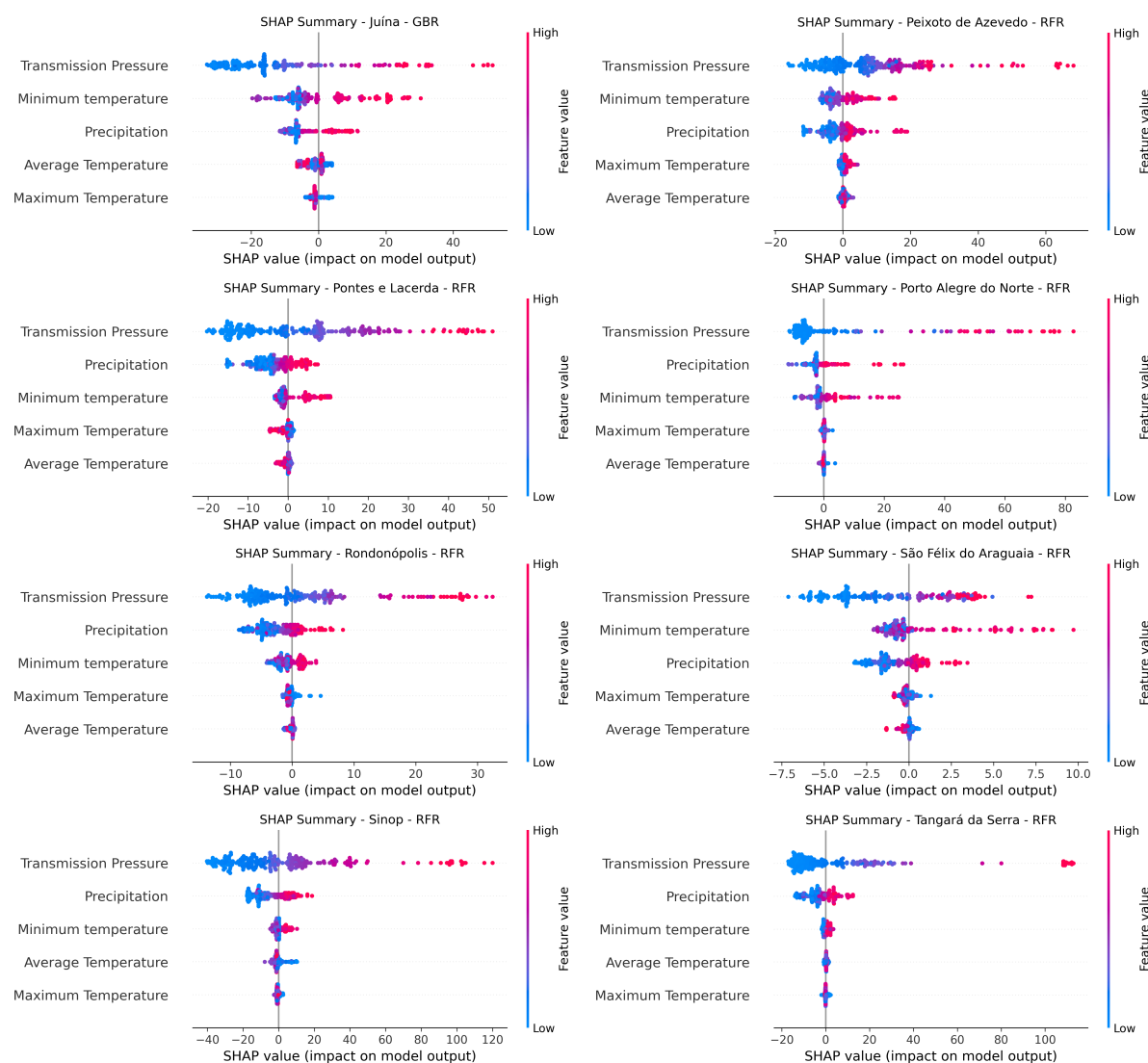


Figura 3.17: Comparação entre as variáveis mais importantes para os modelos de Juína (GBR), Peixoto de Azevedo (RFR), Pontes e Lacerda (RFR), Porto Alegre do Norte (RFR), Rondonópolis (RFR), São Félix do Araguaia (RFR), Sinop (RFR) e Tangará da Serra (RFR).

Em todos os municípios analisados, observou-se que, quando os modelos de Machine Learning utilizam apenas a variável de pressão de transmissão, as previsões se mostram imprecisas, com pontos de dispersão esparsos em relação à linha de 45° e tendência a prever os picos com atraso; por outro lado, ao utilizar exclusivamente variáveis meteorológicas, os modelos também não apresentam bom desempenho, antecipando indevidamente os picos de incidência. Esses comportamentos contrastam com os resultados obtidos quando ambas fontes de informação são combinadas.

Esses resultados demonstram que a combinação de dados epidemiológicos com variáveis meteorológicas melhora substancialmente a acurácia preditiva dos modelos. Além

disso, técnicas de interpretação como o SHAP oferecem informações importantes sobre o funcionamento interno dos modelos, permitindo maior confiabilidade e transparência nas aplicações voltadas à saúde pública.

É importante destacar que, em algumas localidades, a precipitação não se apresentou como a variável de maior relevância, ficando atrás da temperatura mínima. No entanto, manteve-se consistentemente entre os fatores mais importantes para a transmissão da dengue, corroborando os resultados de Almeida et al. [2022], que identificaram associação estatisticamente significativa entre a incidência da doença e a pluviosidade no estado de Mato Grosso ($R^2 = 0,0019$; $p < 0,05$). Em função dessa evidência, os modelos baseados nas variáveis consideradas mais relevantes foram desenvolvidos a partir da combinação de Pressão de Transmissão e Precipitação.

3.4.2 Seleção de Variáveis

Nesta etapa, foram utilizados os resultados da análise SHAP para selecionar, para cada localidade, as variáveis mais relevantes na previsão dos casos de dengue. A variável de maior importância foi, de forma consistente, a Pressão de Transmissão, seguida pela Precipitação, que se destacou como a segunda mais importante na maioria dos modelos. Essa escolha visa avaliar se o uso apenas das variáveis mais relevantes é suficiente para manter (ou até melhorar) o desempenho preditivo dos modelos, reduzindo a complexidade do modelo sem perda de qualidade.

Os resultados mostram que a adição da Precipitação corrige o deslocamento temporal observado ao utilizar exclusivamente a Pressão de Transmissão, permitindo que os modelos capturem com maior precisão os picos de incidência e a sazonalidade característica da doença. Isso é evidenciado tanto nos gráficos de dispersão (Figura 3.18) quanto nas métricas de desempenho (Tabela 3.4).

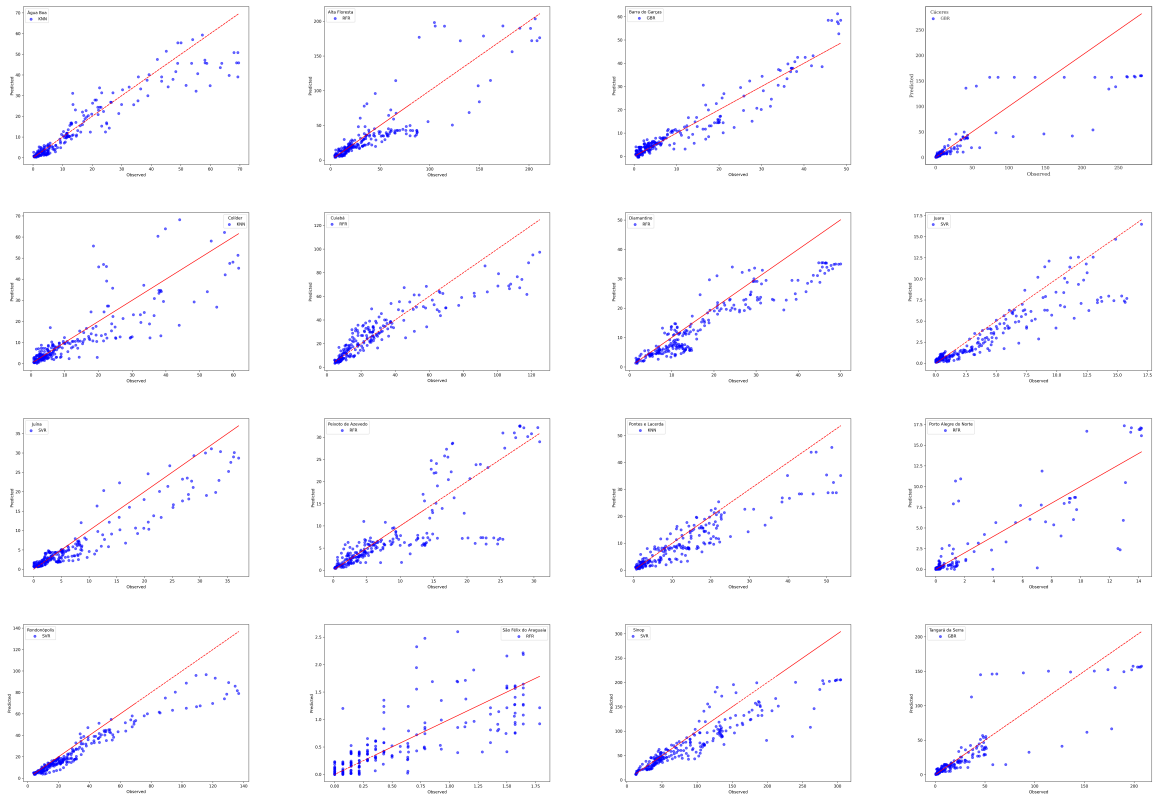


Figura 3.18: Dispersão dos dados previstos pelos medidos para os modelos de Água Boa (KNN), Alta Floresta (RFR), Barra do Garças (GBR), Cáceres (GBR), Colíder (KNN), Cuiabá (RFR), Diamantino (RFR), Juara (SVR), Juína (SVR), Peixoto de Azevedo (RFR), Pontes e Lacerda (KNN), Porto Alegre do Norte (RFR), Rondonópolis (SVR), São Félix do Araguaia (RFR), Sinop (SVR) e Tangará da Serra (GBR).

Os modelos reduzidos de melhor desempenho foram os seguintes: KNN para Água Boa, Colíder e Pontes e Lacerda; GBR para Barra do Garças, Cáceres e Tangará da Serra; RFR para Alta Floresta, Cuiabá, Diamantino, Peixoto de Azevedo, Porto Alegre do Norte e São Félix do Araguaia; SVR para Juara, Juína, Rondonópolis e Sinop.

Tabela 3.4: Desempenho dos modelos nos conjuntos de treino e teste.

Local	Métricas	Treino				Teste			
		RFR	GBR	SVR	KNN	RFR	GBR	SVR	KNN
Cuiabá	MSE	371.145	266.825	4445.734	884.862	129.450	128.565	163.499	160.313
	RMSE	19.265	16.335	66.676	29.747	11.378	11.339	12.787	12.661
	MAE	7.912	8.757	22.771	12.917	6.887	6.403	7.986	8.114
	Willmott	0.990	0.993	0.792	0.976	0.940	0.942	0.919	0.925
	Pearson	0.982	0.987	0.850	0.956	0.955	0.922	0.920	0.908
	R^2	0.964	0.974	0.564	0.913	0.821	0.820	0.773	0.777

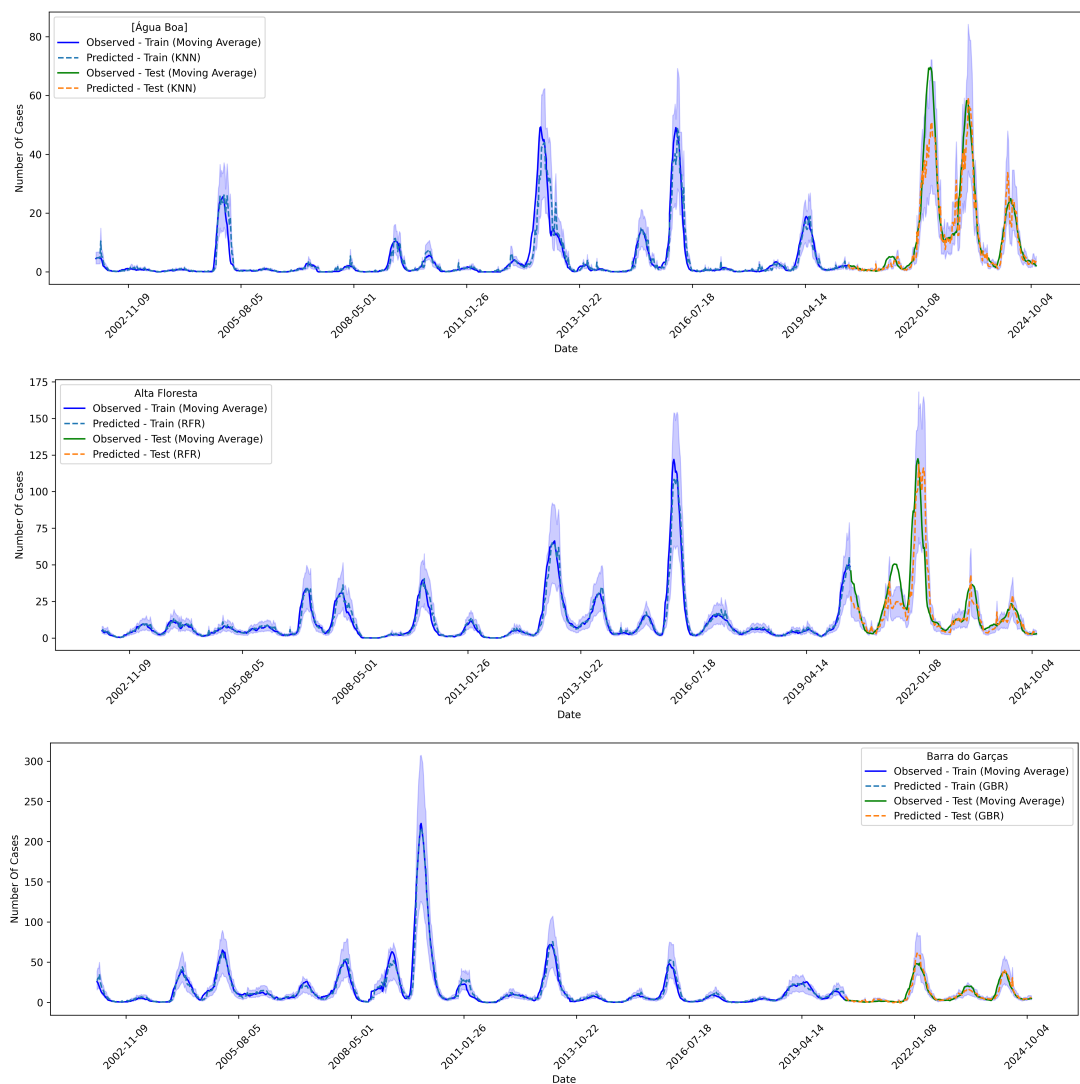
(continua na próxima página)

Local	Métricas	Treino				Teste			
		RFR	GBR	SVR	KNN	RFR	GBR	SVR	KNN
Rondonópolis	MSE	22.941	33.596	391.601	60.956	192.651	169.647	160.924	161.393
	RMSE	4.790	5.796	19.789	7.807	13.880	13.025	12.686	12.704
	MAE	2.471	3.421	7.615	4.119	8.229	7.378	7.684	7.155
	Willmott	0.995	0.993	0.871	0.987	0.930	0.939	0.936	0.941
	Pearson	0.991	0.987	0.901	0.976	0.883	0.900	0.967	0.915
	R^2	0.982	0.973	0.689	0.952	0.772	0.799	0.809	0.809
Barra do Gargas	MSE	19.257	19.735	565.300	41.039	21.321	13.492	22.324	28.352
	RMSE	4.388	4.442	23.776	6.406	4.618	3.673	4.725	5.325
	MAE	1.980	2.521	7.360	3.120	2.670	2.412	3.184	2.856
	Willmott	0.994	0.994	0.589	0.987	0.968	0.979	0.954	0.959
	Pearson	0.989	0.988	0.682	0.976	0.943	0.960	0.946	0.928
	R^2	0.977	0.976	0.324	0.951	0.865	0.914	0.858	0.820
Cáceres	MSE	28.385	27.592	263.639	56.924	998.157	924.156	3262.587	1135.760
	RMSE	5.328	5.253	16.237	7.545	31.594	30.400	57.119	33.701
	MAE	2.232	2.547	5.664	3.352	11.328	11.031	19.303	12.173
	Willmott	0.985	0.986	0.741	0.969	0.890	0.902	0.338	0.855
	Pearson	0.973	0.974	0.868	0.945	0.883	0.887	0.721	0.967
	R^2	0.946	0.947	0.498	0.892	0.715	0.736	0.068	0.675
Juína	MSE	10.580	13.234	78.537	25.959	18.242	14.664	12.374	13.477
	RMSE	3.253	3.638	8.862	5.095	4.271	3.829	3.518	3.671
	MAE	1.482	1.917	3.763	2.343	2.482	2.272	2.231	2.390
	Willmott	0.985	0.980	0.804	0.961	0.954	0.961	0.951	0.963
	Pearson	0.972	0.965	0.870	0.928	0.938	0.945	0.945	0.940
	R^2	0.943	0.929	0.579	0.861	0.769	0.814	0.843	0.829
Porto Alegre do Norte	MSE	0.384	0.238	4.833	1.202	3.459	4.704	8.016	4.087
	RMSE	0.619	0.488	2.198	1.096	1.860	2.169	2.831	2.022
	MAE	0.181	0.189	0.609	0.319	0.742	0.894	1.208	0.808
	Willmott	0.986	0.992	0.607	0.956	0.932	0.915	0.652	0.920
	Pearson	0.976	0.985	0.866	0.918	0.874	0.853	0.824	0.854
	R^2	0.949	0.969	0.363	0.842	0.720	0.620	0.352	0.670
Sinop	MSE	164.671	194.385	2715.792	397.409	1231.717	1365.207	1193.416	1170.063
	RMSE	12.832	13.942	52.113	19.935	35.096	36.949	34.546	34.206
	MAE	5.216	6.835	19.843	8.309	23.425	23.754	22.654	23.058
	Willmott	0.993	0.992	0.805	0.984	0.931	0.926	0.912	0.932
	Pearson	0.988	0.985	0.867	0.970	0.875	0.866	0.916	0.881
	R^2	0.975	0.970	0.583	0.939	0.730	0.701	0.738	0.744
Tangará da Serra	MSE	9.536	11.902	107.854	23.948	433.207	409.934	1408.436	422.781
	RMSE	3.088	3.450	10.385	4.894	20.814	20.247	37.529	20.562
	MAE	1.532	1.936	4.313	2.491	9.228	8.602	13.906	8.925
	Willmott	0.992	0.990	0.854	0.979	0.927	0.933	0.534	0.929
	Pearson	0.985	0.981	0.887	0.962	0.882	0.886	0.738	0.885
	R^2	0.969	0.962	0.655	0.923	0.768	0.780	0.246	0.774

(continua na próxima página)

Local	Métricas	Treino				Teste			
		RFR	GBR	SVR	KNN	RFR	GBR	SVR	KNN
Diamantino	MSE	0.444	0.611	12.384	1.249	71.790	89.677	229.667	79.809
	RMSE	0.666	0.782	3.519	1.118	8.473	9.470	15.155	8.934
	MAE	0.286	0.464	1.705	0.606	5.688	6.232	11.489	7.192
	Willmott	0.995	0.993	0.712	0.985	0.848	0.799	0.562	0.860
	Pearson	0.991	0.987	0.883	0.974	0.866	0.819	0.926	0.893
	R^2	0.981	0.974	0.467	0.946	0.564	0.455	-0.395	0.515
Alta Floresta	MSE	11.190	11.463	135.110	29.511	152.758	163.365	430.375	201.928
	RMSE	3.345	3.386	11.624	5.432	12.360	12.781	20.745	14.210
	MAE	1.599	1.931	4.105	2.591	7.179	6.906	10.410	8.362
	Willmott	0.989	0.989	0.762	0.971	0.933	0.933	0.624	0.904
	Pearson	0.981	0.980	0.803	0.946	0.879	0.879	0.804	0.839
	R^2	0.960	0.959	0.515	0.894	0.748	0.730	0.289	0.666
Juara	MSE	1.671	2.051	12.525	3.992	6.280	4.881	4.133	6.567
	RMSE	1.293	1.432	3.539	1.998	2.506	2.209	2.033	2.563
	MAE	0.617	0.790	1.572	0.977	1.416	1.224	1.208	1.459
	Willmott	0.992	0.990	0.916	0.980	0.892	0.917	0.924	0.883
	Pearson	0.985	0.981	0.914	0.963	0.821	0.863	0.916	0.814
	R^2	0.969	0.962	0.769	0.926	0.647	0.725	0.767	0.630
Peixoto de Azevedo	MSE	1.298	1.831	15.981	2.470	23.240	25.628	26.893	26.459
	RMSE	1.139	1.353	3.998	1.572	4.821	5.062	5.186	5.144
	MAE	0.584	0.787	1.605	0.898	2.832	2.864	3.212	3.090
	Willmott	0.992	0.989	0.837	0.985	0.899	0.888	0.821	0.889
	Pearson	0.986	0.979	0.898	0.971	0.818	0.798	0.906	0.804
	R^2	0.970	0.958	0.635	0.944	0.609	0.568	0.547	0.554
Água Boa	MSE	2.442	2.135	33.668	6.432	58.934	87.564	239.576	37.506
	RMSE	1.563	1.461	5.802	2.536	7.677	9.358	15.478	6.124
	MAE	0.623	0.733	1.923	1.046	3.789	4.559	7.730	3.351
	Willmott	0.990	0.992	0.737	0.974	0.937	0.903	0.599	0.964
	Pearson	0.983	0.984	0.850	0.951	0.923	0.874	0.804	0.948
	R^2	0.963	0.968	0.495	0.903	0.811	0.719	0.232	0.880
Pontes e Lacerda	MSE	2.013	2.630	21.037	4.774	34.348	36.881	46.443	31.972
	RMSE	1.419	1.622	4.587	2.185	5.861	6.073	6.815	5.654
	MAE	0.717	0.945	1.898	1.193	3.553	3.661	3.901	3.434
	Willmott	0.993	0.991	0.898	0.984	0.919	0.910	0.881	0.927
	Pearson	0.987	0.983	0.898	0.969	0.928	0.926	0.911	0.931
	R^2	0.974	0.966	0.732	0.939	0.765	0.748	0.683	0.782
Colíder	MSE	16.306	6.146	209.614	36.293	69.816	120.996	142.703	59.585
	RMSE	4.038	2.479	14.478	6.024	8.356	11.000	11.946	7.719
	MAE	1.295	1.256	4.196	1.981	4.122	4.616	7.143	4.463
	Willmott	0.983	0.994	0.429	0.961	0.912	0.862	0.656	0.914
	Pearson	0.970	0.989	0.725	0.929	0.841	0.776	0.922	0.846
	R^2	0.938	0.977	0.206	0.862	0.640	0.377	0.265	0.693
São Félix do Araguaia	MSE	0.062	0.055	0.355	0.145	0.137	0.155	0.118	0.107
	RMSE	0.249	0.235	0.596	0.381	0.370	0.393	0.343	0.328
	MAE	0.102	0.119	0.241	0.169	0.231	0.237	0.233	0.219
	Willmott	0.984	0.986	0.861	0.961	0.875	0.852	0.848	0.873
	Pearson	0.971	0.974	0.889	0.930	0.774	0.737	0.918	0.829
	R^2	0.942	0.948	0.666	0.864	0.529	0.467	0.594	0.630

Em praticamente todos os cenários, observou-se melhora no desempenho dos modelos quando treinados com apenas duas variáveis. Em Barra do Garças, o desempenho manteve-se semelhante, mas com redução do overfitting. Em Colíder, apesar de uma pequena melhora, a baixa incidência de casos (Figura 3.11) limitou a capacidade de generalização do modelo. Situação semelhante ocorreu em Peixoto de Azevedo e em Porto Alegre do Norte — neste último, o modelo com todas as variáveis apresentou desempenho superior ao obtido com a seleção das duas mais importantes. Em Sinop, também não houve mudanças significativas no desempenho, embora tenha sido observada redução do overfitting.



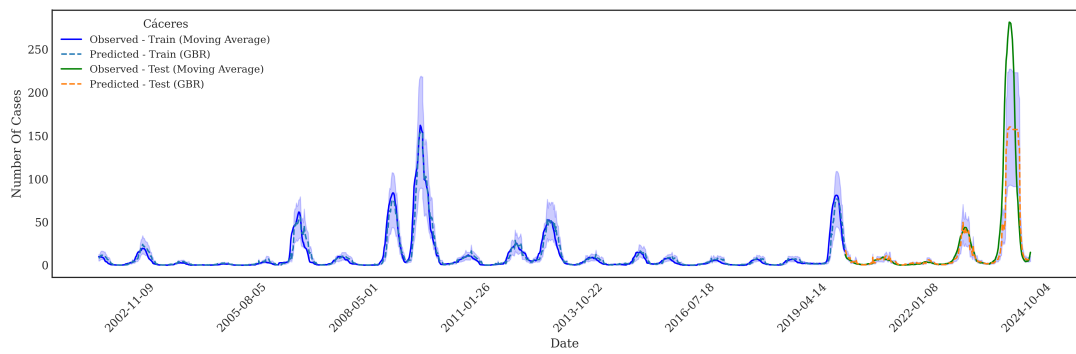
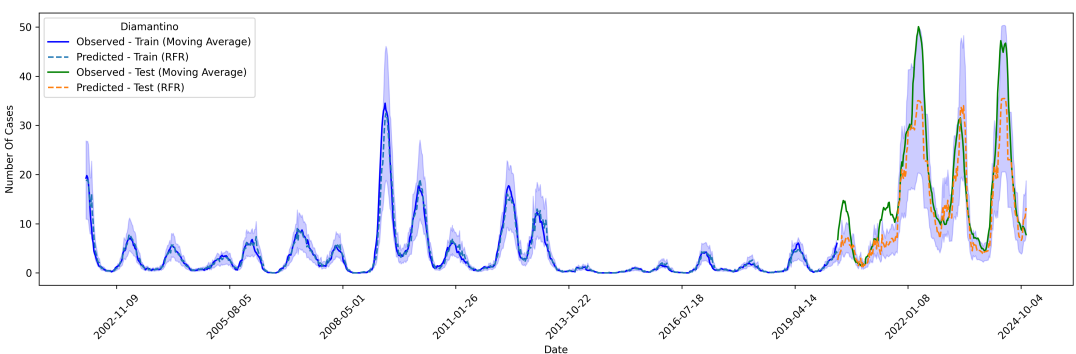
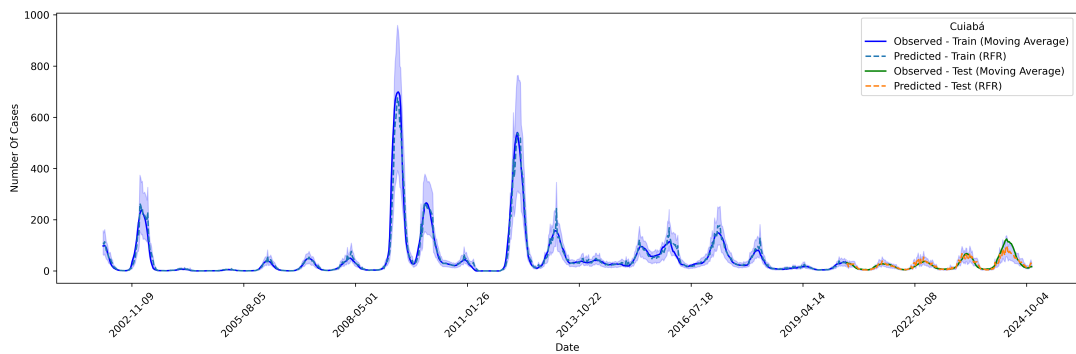
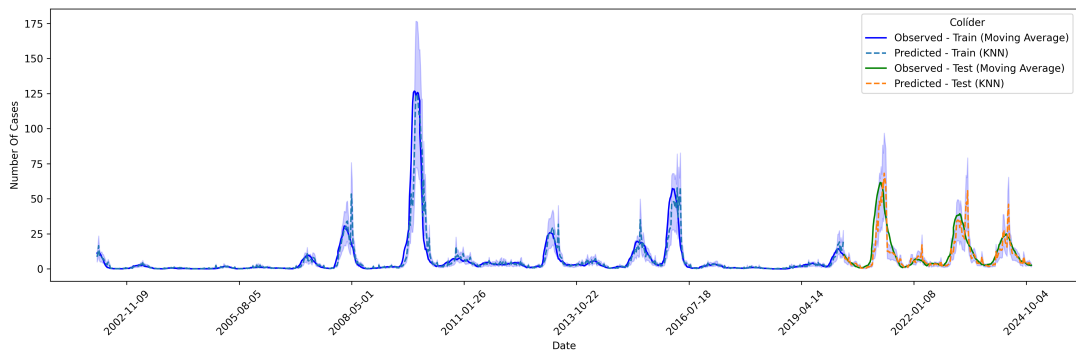


Figura 3.19: Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Água Boa (KNN), Alta Floresta (RFR), Barra do Garças (GBR) e Cáceres (GBR).



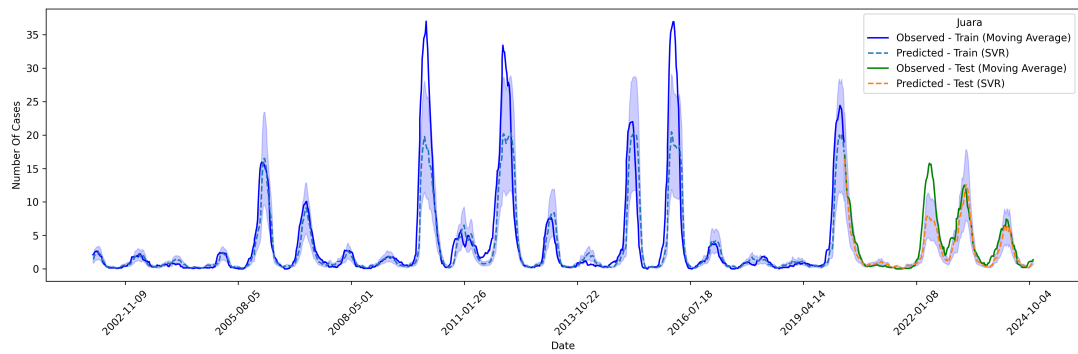
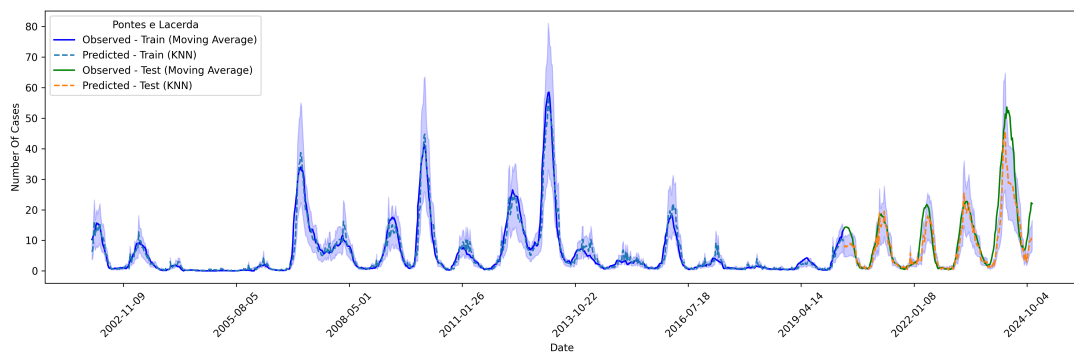
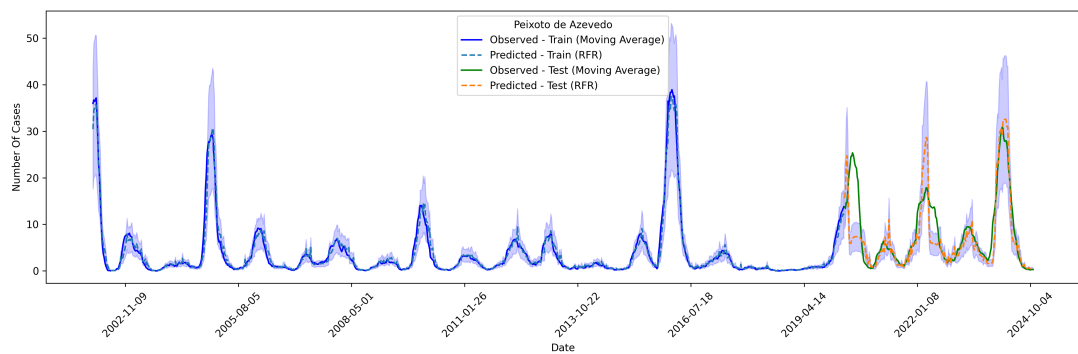
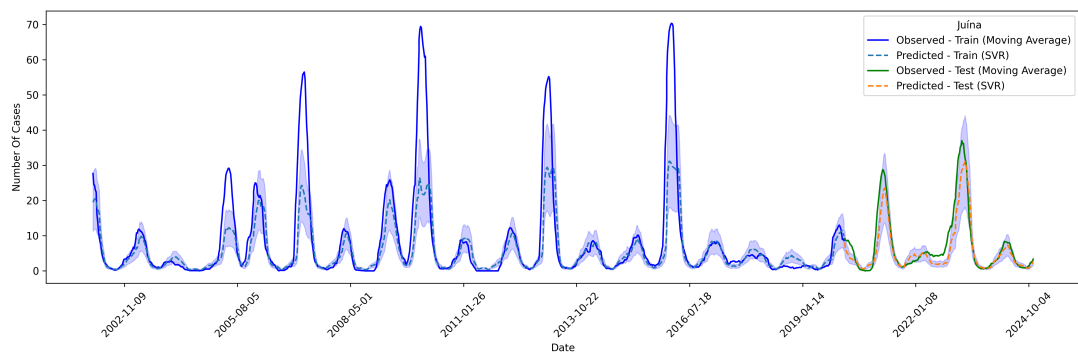


Figura 3.20: Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Colíder (KNN), Cuiabá (RFR), Diamantino (RFR) e Juara (SVR).



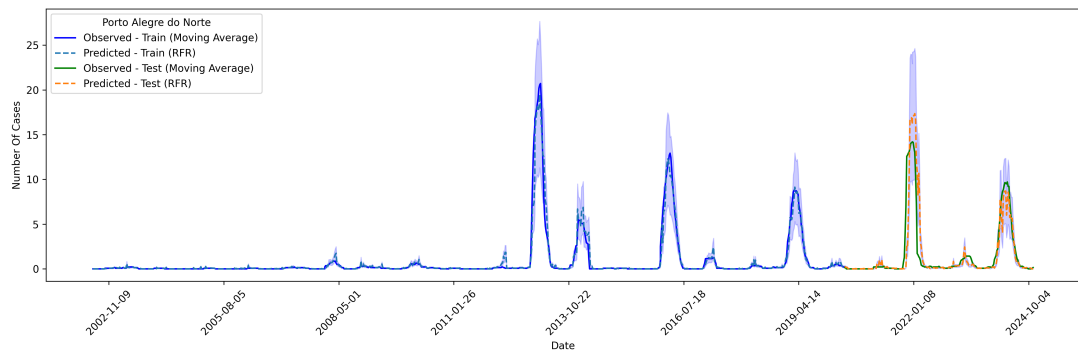
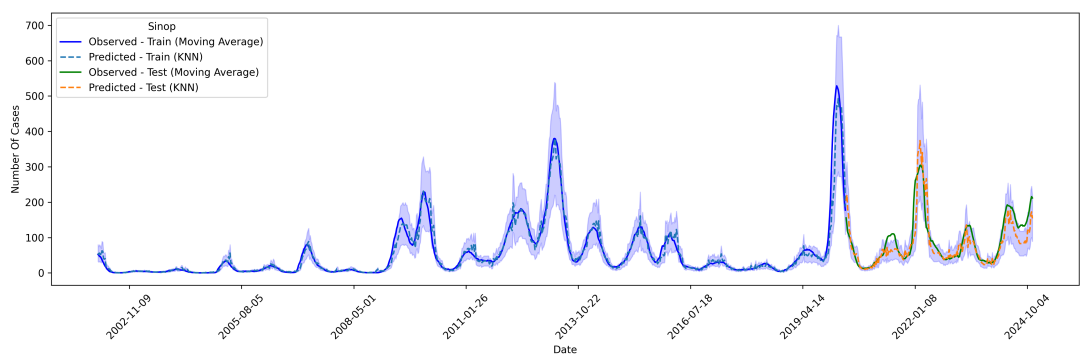
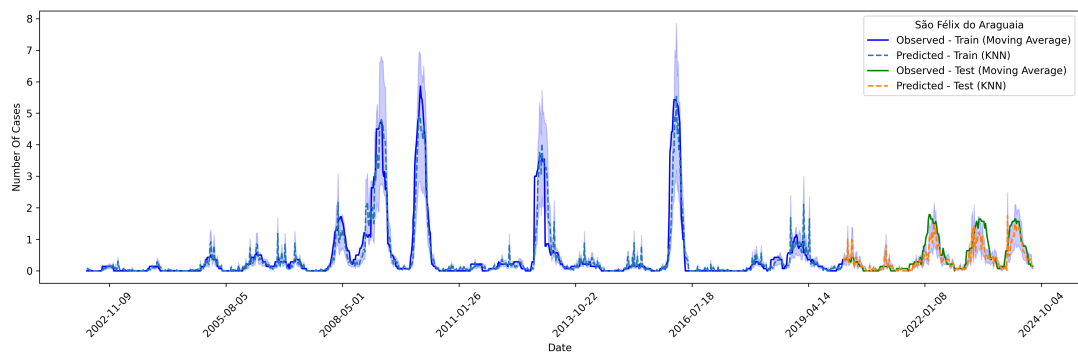
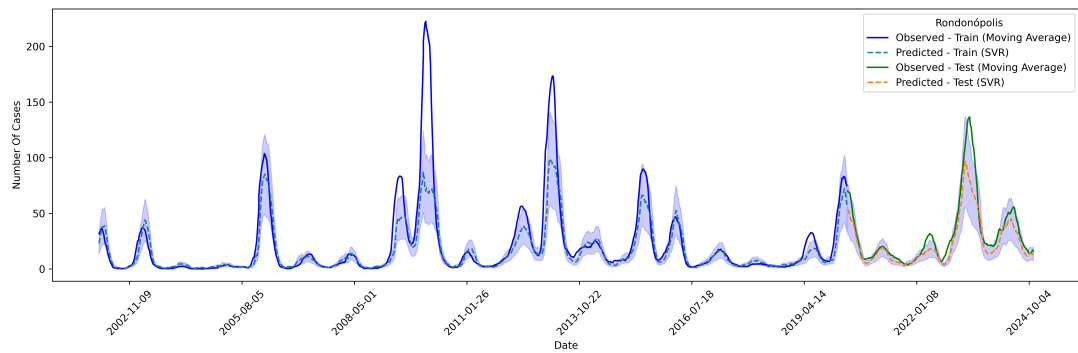


Figura 3.21: Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Juína (SVR), Peixoto de Azevedo (RFR), Pontes e Lacerda (KNN) e Porto Alegre do Norte (RFR).



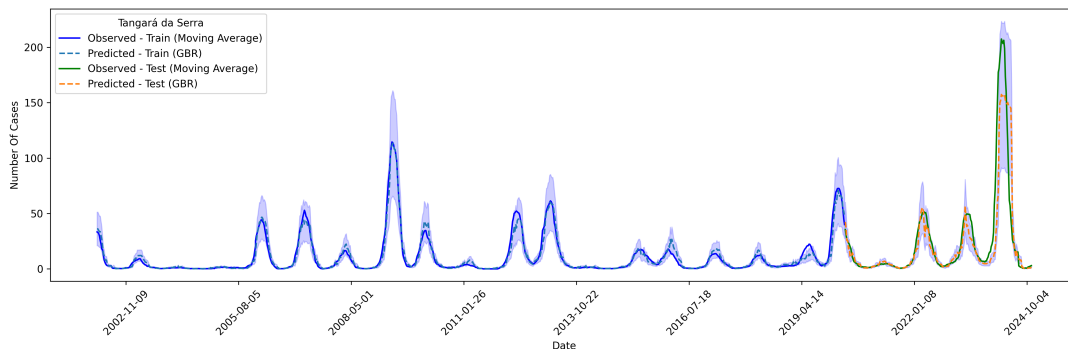


Figura 3.22: Comparação entre os dados reais e os previstos (treino + teste) para os modelos de Rondonópolis (SVR), São Félix do Araguaia (KNN), Sinop (KNN) e Tangará da Serra (GBR).

Esses resultados reforçam a centralidade da Precipitação como uma variável meteorológica crítica na previsão da incidência de dengue, evidenciando sua influência direta na dinâmica populacional do vetor *Aedes aegypti* e, por consequência, na propagação da doença. Quando combinada com a variável Pressão de Transmissão — que representa o acúmulo de casos anteriores e, portanto, a presença contínua do vírus na população —, a capacidade preditiva dos modelos é significativamente ampliada. Esse achado destaca a importância de integrar indicadores climáticos e epidemiológicos na construção de modelos preditivos, de modo a capturar com maior precisão os padrões temporais e espaciais da doença.

Além disso, os resultados indicam que, para alguns algoritmos de aprendizado de máquina, a utilização de um conjunto reduzido de variáveis selecionadas com base em critérios estatísticos ou de importância relativa pode ser mais vantajosa do que o uso irrestrito de todos os dados disponíveis. Essa estratégia de seleção de atributos permite a construção de modelos mais eficientes, menos suscetíveis ao overfitting e com maior interpretabilidade, favorecendo sua incorporação em sistemas práticos de apoio à decisão. O equilíbrio entre complexidade e desempenho é um fator-chave na adoção de tecnologias de previsão em contextos reais, especialmente quando os recursos computacionais ou operacionais são limitados.

Com base nos resultados obtidos, este estudo corrobora a relevância e a viabilidade do uso de modelos baseados em inteligência artificial para a previsão antecipada de casos de dengue em nível municipal. A possibilidade de estimar com até uma semana de antecedência os comportamentos epidêmicos da doença, incluindo os picos sazonais de incidência, representa uma ferramenta estratégica para as equipes de vigilância em saúde. Essa capacidade de antecipação pode contribuir para ações preventivas mais eficazes, como o direcionamento geográfico de campanhas de controle vetorial, mobilização

comunitária, alocação de recursos e intensificação do monitoramento em períodos críticos.

Essas contribuições tornam-se ainda mais significativas no contexto das mudanças climáticas globais, que têm alterado os padrões ambientais e ampliado a distribuição geográfica de vetores transmissores de arboviroses. Assim, o presente trabalho não apenas responde a uma demanda científica e tecnológica, mas também se insere em um esforço mais amplo de enfrentamento dos desafios sanitários emergentes, promovendo resiliência adaptativa em sistemas públicos de saúde. Essa abordagem dialoga diretamente com os Objetivos de Desenvolvimento Sustentável (ODS) das Nações Unidas, especialmente o ODS 3 (Saúde e Bem-Estar), ao fortalecer mecanismos de prevenção e resposta a doenças infecciosas, e o ODS 13 (Ação Contra a Mudança Global do Clima), ao utilizar ferramentas baseadas em ciência e tecnologia para reduzir os impactos de eventos climáticos sobre a saúde humana.

Ademais, os achados obtidos abrem espaço para o desenvolvimento de plataformas digitais inteligentes e acessíveis, baseadas em modelos previamente treinados ou ajustáveis conforme as particularidades de cada região. Tais sistemas poderiam permitir que profissionais da saúde ou gestores públicos inserissem dados epidemiológicos e climáticos atualizados, gerando previsões automáticas e interpretações claras sobre os fatores que influenciam a evolução da doença. Essa proposta é particularmente relevante em estados com grandes extensões territoriais e realidades climáticas diversas, como Mato Grosso, onde a descentralização da vigilância epidemiológica pode beneficiar-se enormemente de soluções tecnológicas adaptáveis e de fácil uso.

Portanto, os resultados aqui apresentados não apenas oferecem evidências técnicas sólidas sobre a aplicabilidade de modelos preditivos de dengue, como também contribuem para estratégias de inovação em saúde pública, promovendo a integração entre ciência de dados, vigilância epidemiológica e políticas de adaptação às mudanças climáticas. Como exemplo, pode-se citar a utilização desses modelos em sistemas de alerta precoce que classifiquem semanalmente os municípios por faixas de risco, orientando ações de controle vetorial, planejamento de leitos e campanhas de comunicação em saúde. A perspectiva de transformar esses modelos em ferramentas operacionais escaláveis representa um avanço importante na capacidade de resposta das autoridades de saúde frente aos desafios sanitários contemporâneos e futuros.

3.5 Conclusões

Com base nos resultados obtidos, evidencia-se a importância da integração entre variáveis meteorológicas – com destaque para a Precipitação – e a variável epidemiológica de Pressão de Transmissão na construção de modelos robustos para previsão da incidência

de dengue. Essa combinação se mostrou fundamental para captar, com maior precisão, os padrões sazonais e os picos epidêmicos da doença, superando o desempenho de modelos que utilizaram variáveis de forma isolada. A precipitação, por sua estreita relação com a dinâmica reprodutiva do vetor *Aedes aegypti*, funcionou como um sinal climático precursor, enquanto a Pressão de Transmissão refletiu o histórico local de circulação viral, conferindo aos modelos uma capacidade contextualizada de previsão.

Nos 16 municípios analisados – Água Boa, Alta Floresta, Barra do Garças, Cáceres, Colíder, Cuiabá, Diamantino, Juara, Juína, Peixoto de Azevedo, Pontes e Lacerda, Porto Alegre do Norte, Rondonópolis, São Félix do Araguaia, Sinop e Tangará da Serra –, os modelos de aprendizado de máquina demonstraram capacidade de adaptação aos diferentes perfis epidemiológicos e climáticos regionais, mesmo diante de grandes variações no número de casos notificados entre os anos de 2013 e 2024. Os algoritmos SVR, RFR, GBR e KNN apresentaram desempenhos variáveis conforme a localidade, o que destaca a necessidade de ajustes finos e personalizados por município, respeitando as peculiaridades locais. Apesar dessas variações, todos os modelos foram capazes de capturar os principais surtos epidêmicos e tendências sazonais da dengue, com bom desempenho tanto nos conjuntos de treino quanto nos de teste.

A alternância entre os modelos mais eficazes em cada município revela a complexidade da modelagem epidemiológica, influenciada por fatores como volume de dados, qualidade dos registros, padrão climático local e intensidade da circulação viral. Ainda assim, os modelos apresentaram estabilidade, coerência estatística e capacidade de generalização, sendo que, em diversos casos, a redução criteriosa do número de variáveis – com base em testes empíricos e indicadores de importância – resultou em modelos mais eficientes, menos sobreajustados e com maior potencial de interpretação, o que é crucial para a transparência e aceitação de ferramentas tecnológicas no contexto da saúde pública.

Do ponto de vista da gestão pública e da vigilância em saúde, os resultados reforçam o potencial dos modelos baseados em inteligência artificial como ferramentas preditivas estratégicas, capazes de antecipar a evolução da dengue com até uma semana de antecedência. Essa capacidade de previsão fornece uma janela temporal de atuação preventiva, permitindo o redirecionamento oportuno de recursos humanos e materiais, a intensificação do controle vetorial em áreas críticas e o planejamento de ações educativas e assistenciais com base em evidências. A adoção dessas ferramentas em esferas municipais pode representar um avanço significativo na eficiência operacional dos serviços de saúde, especialmente em estados com grandes dimensões territoriais e heterogeneidade climática, como Mato Grosso.

Além disso, a proposta metodológica e os resultados obtidos estão alinhados aos Objetivos de Desenvolvimento Sustentável (ODS), em especial ao ODS 3 (Saúde e Bem-

Estar), ao contribuir com soluções orientadas para a prevenção de doenças infecciosas, e ao ODS 13 (Ação Contra a Mudança Global do Clima), ao reconhecer a influência crescente das variabilidades climáticas na saúde ambiental e humana, e propor formas adaptativas de resposta. O uso de modelos preditivos com base em ciência de dados, aprendizado de máquina e integração de múltiplas fontes de informação representa um caminho promissor para o enfrentamento das arboviroses em cenários de mudança ambiental.

Por fim, este trabalho contribui não apenas para o avanço técnico na modelagem preditiva de doenças tropicais, mas também para o desenvolvimento de abordagens práticas, escaláveis e adaptáveis, com potencial de aplicação em sistemas inteligentes de apoio à decisão. A transformação desses modelos em plataformas acessíveis e automatizadas pode democratizar o uso da inteligência artificial no setor público, fortalecendo a resiliência dos sistemas de saúde frente aos desafios sanitários atuais e emergentes, e promovendo inovação com impacto social direto e mensurável. Como exemplo de aplicação prática, tais modelos poderiam integrar rotinas semanais da Secretaria Estadual de Saúde, gerando boletins automáticos de risco para cada município, alimentados por dados epidemiológicos e climáticos atualizados, e fornecendo alertas antecipados que orientem decisões operacionais — como intensificação de visitas dos agentes de endemias, distribuição de insumos, priorização de municípios para ações de bloqueio e planejamento sazonal de equipes.

4. Conclusão Geral

A presente dissertação teve como objetivo principal desenvolver e avaliar modelos preditivos de casos de dengue a partir da integração de dados epidemiológicos e variáveis meteorológicas em municípios de Mato Grosso, utilizando técnicas de aprendizado de máquina. A proposta partiu do reconhecimento de que a dengue é uma doença multifatorial, cuja dinâmica é fortemente modulada por fatores ambientais, climáticos e socioeconômicos, e que métodos tradicionais de previsão nem sempre capturam as relações não lineares e defasadas que caracterizam o ciclo vetor-vírus-hospedeiro.

A metodologia desenvolvida neste trabalho incorporou séries temporais semanais de casos notificados de dengue (2013–2024) obtidas da Secretaria Estadual de Saúde de Mato Grosso (SES-MT), associadas a variáveis meteorológicas derivadas do ERA5-Land, incluindo temperatura média, máxima, mínima e precipitação. As variáveis foram submetidas a um processo de padronização e suavização temporal, com a aplicação de uma média móvel de 14 semanas, de forma a reduzir ruídos e representar melhor as tendências epidemiológicas e climáticas. Além disso, foi proposta a variável *Pressão de Transmissão* (Pt), construída a partir do produto entre incidência e suscetibilidade, representando uma métrica dinâmica da força de infecção ao longo do tempo.

Os modelos de aprendizado de máquina testados — *Random Forest Regressor* (RFR), *Gradient Boosting Regressor* (GBR), *Support Vector Regressor* (SVR) e *k-Nearest Neighbors* (KNN) — foram avaliados a partir de múltiplas métricas de desempenho (MAE, MSE, RMSE, R^2 , coeficiente de Willmott e correlação de Pearson), tanto em conjuntos de treino quanto de teste. Os resultados evidenciaram que, de modo geral, os modelos RFR e GBR apresentaram as melhores performances, com maior capacidade de generalização e menor sobreajuste, enquanto o KNN e o SVR mostraram limitações em capturar picos epidêmicos.

A análise de interpretabilidade realizada via SHAP confirmou que a variável *Pressão de Transmissão* (Pt) foi o principal determinante na explicação das previsões, seguida pelas variáveis de precipitação e temperatura mínima. Essa hierarquia de importância reforça a hipótese de que as condições climáticas, em especial o regime de chuvas e as temperaturas mais amenas, influenciam diretamente a dinâmica de transmissão do vetor,

ao mesmo tempo em que a Pt sintetiza o efeito acumulado de fatores biológicos e populacionais. Os resultados mostraram ainda que o uso combinado de variáveis epidemiológicas e climáticas melhora substancialmente o desempenho dos modelos, em comparação com o uso isolado de apenas um desses grupos.

Do ponto de vista espacial, os municípios analisados apresentaram padrões distintos de resposta às variáveis meteorológicas, refletindo a heterogeneidade climática e ecológica do estado de Mato Grosso, que abrange porções dos biomas Amazônia, Cerrado e Pantanal. Essa diversidade ambiental destaca a importância de abordagens regionalizadas e adaptativas, capazes de incorporar particularidades locais à modelagem preditiva.

A integração de aprendizado de máquina com dados meteorológicos e epidemiológicos demonstrou ser uma estratégia eficaz para aprimorar o monitoramento e a previsão da dengue, oferecendo subsídios valiosos para a vigilância epidemiológica. Além de fornecer previsões com bom desempenho estatístico, os modelos interpretáveis permitem identificar condições favoráveis à transmissão antes do aumento efetivo dos casos, viabilizando ações preventivas mais direcionadas.

Como perspectivas futuras, recomenda-se a ampliação da base de dados para incluir variáveis socioeconômicas, de uso e cobertura da terra, e de mobilidade humana, bem como a incorporação de abordagens de aprendizado profundo e espacial (como CNNs e GNNs), que possam capturar relações entre municípios vizinhos. A construção de modelos híbridos, que combinem componentes mecanicistas e empíricos, também se apresenta como uma linha promissora para representar de forma mais completa o sistema de transmissão.

Em síntese, este trabalho contribui para o avanço da modelagem preditiva da dengue em regiões tropicais, ao propor uma estrutura metodológica reprodutível, interpretável e sensível às variáveis climáticas. Os resultados obtidos reforçam o potencial do aprendizado de máquina como ferramenta de apoio à tomada de decisão em saúde pública, especialmente em contextos de elevada vulnerabilidade ambiental e climática, como o estado de Mato Grosso.

Referências Bibliográficas

URL <https://brasil.un.org/pt-br/sdgs>.

Claudio Almeida, Luis Maurano, Dalton Valeriano, Gilberto Câmara, Lubia Vinhas, Alessandra Gomes, Antonio Monteiro, Arlesson Antonio Souza, Camilo Rennó, Daniel e Silva, Marcos Adami, Maria Escada, Marisa Motta, and Silvana Amaral. Methodology for forest monitoring used in prodes and deter projects, 01 2021.

Clayton Alcarde Alvares, José Luiz Stape, Paulo Cesar Sentelhas, JL de M Gonçalves, Gerd Sparovek, et al. Köppen’s climate classification map for brazil. *Meteorologische zeitschrift*, 22(6):711–728, 2013.

Daniel Cristobal Andrade Girón, William Joel Marín Rodriguez, Flor de María Lioo-Jordan, and Jose Luis Ausejo Sánchez. Machine learning and deep learning models for dengue diagnosis prediction: A systematic review. In *Informatics*, volume 12, page 15. MDPI, 2025.

Christovam Barcellos, Vanderlei Matos, Raquel Lana, and Rachel Lowe. Climate change, thermal anomalies, and the recent progression of dengue in brazil. *Scientific Reports*, 14, 03 2024a. doi: 10.1038/s41598-024-56044-y.

Christovam Barcellos, Vanderlei Matos, Raquel Martins Lana, and Rachel Lowe. Climate change, thermal anomalies, and the recent progression of dengue in brazil. *Scientific reports*, 14(1):5948, 2024b.

Corey M. Benedum, Osama M. E. Seidahmed, Elfatih A. B. Eltahir, and Natasha Markuzon. Statistical modeling of the effect of rainfall flushing on dengue transmission in singapore. *PLOS Neglected Tropical Diseases*, 12(12):1–18, 12 2018. doi: 10.1371/journal.pntd.0006935. URL <https://doi.org/10.1371/journal.pntd.0006935>.

Bernhard Boser, Isabelle Guyon, and Vladimir Vapnik. A training algorithm for optimal margin classifier. *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, 5, 08 1996. doi: 10.1145/130385.130401.

- Ima Aparecida Braga and Denise Valle. Aedes aegypti: histórico do controle no Brasil. *Epidemiologia e Serviços de Saúde*, 16:113 – 118, 06 2007. ISSN 1679-4974. doi: 10.5123/S1679-49742007000200006.
- Xiang Chen and Paula Moraga. Assessing dengue forecasting methods: a comparative study of statistical models and machine learning techniques in rio de janeiro, brazil. *Tropical medicine and health*, 53(1):52, 2025a.
- Xiang Chen and Paula Moraga. Forecasting dengue across brazil with lstm neural networks and shap-driven lagged climate and spatial effects. *BMC public health*, 25 (1):973, 2025b.
- Sylvain Christin, Éric Hervet, and Nicolas Lecomte. Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10):1632–1644, 2019.
- Felipe J. Colón-González, Carlo Fezzi, Iain R. Lake, and Paul R. Hunter. The effects of weather and climate change on dengue. *PLoS Neglected Tropical Diseases*, 7(11): e2503, November 2013. doi: 10.1371/journal.pntd.0002503. URL <https://doi.org/10.1371/journal.pntd.0002503>.
- Copernicus Climate Change Service (C3S). Era5 land monthly averaged data from 1950 to present, 2019. URL <https://doi.org/10.24381/cds.68d2bb30>. Acessado em: 27 fev. 2025.
- Tatiane Gomes de Almeida, Ernandes Sobreira Oliveira Júnior, and Claumir Cesar Muniz. Regionais de saúde e os casos de dengue no mato grosso: a chuva como principal fator para a proliferação do aedes aegypti. *Revista Ciência Geográfica*, 26(01):437–453, 2022.
- Luiz Felipe de Azevedo Assunção, Adrieny Álida Guilherme de Macêdo, Luiza Zielke da Silva, Letícia Lopes da Costa, João Victor de Melo Amaral, Zelda Maria dos Santos Miranda, Rodolpho Marcell Medeiros Costa de Melo, Beatriz Araujo da Costa, Christiane Maria Passos Marcos, Julianna Beatriz Arruda de Moraes, et al. Sazonalidade na incidência da dengue no brasil nos últimos 10 anos: Uma revisão epidemiológica. *Brazilian Journal of Implantology and Health Sciences*, 7(3):139–151, 2025.
- STI Secretaria de Estado de Saúde MT. Ses-mt - secretaria de estado de saúde de mato grosso. <https://www.saude.mt.gov.br/unidade/superintendencia-de-politicas-de-saude/296/regioes-de-saude-de-mato-grosso>.
- Secretaria Estadual de Saúde de Mato Grosso. Ses mt, 2 2025. URL <http://appweb3.saude.mt.gov.br/dw>. [Online; accessed 27. Feb. 2025].

- Krishna Devulapalli. Neural networks for classification and regression. *Biom Biostat Int J*, 2(6):183–185, 2015. doi: 10.15406/bbij.2015.02.00046. URL <https://doi.org/10.15406/bbij.2015.02.00046>.
- Kristie L. Ebi and Joshua Nealon. Dengue in a changing climate. *Environmental Research*, 151:115–123, 2016. ISSN 0013-9351. doi: <https://doi.org/10.1016/j.envres.2016.07.026>. URL <https://www.sciencedirect.com/science/article/pii/S0013935116303127>.
- Lucas Farias, Talita Silva, and Pedro Araújo. Multitask lstm for arboviral outbreak prediction using public health data, 05 2025.
- Fiocruz. URL <https://fiocruz.br/taxonomia-geral-05-doencas/dengue>.
- Jerome Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29, 11 2000. doi: 10.1214/aos/1013203451.
- J. Matthew Helm, Andrew M. Swiergosz, Heather S. Haeberle, Jaret M. Karnuta, Jonathan L. Schaffer, Viktor E. Krebs, Andrew I. Spitzer, and Prem N. Ramkumar. Machine learning and artificial intelligence: Definitions, applications, and future directions. *Current Reviews in Musculoskeletal Medicine*, 13(1):69–76, 2 2020. ISSN 1935-9748. doi: 10.1007/s12178-020-09600-8. URL <https://doi.org/10.1007/s12178-020-09600-8>.
- Xing Yu Leung, Rakibul M Islam, Mohammadmehdi Adhami, Dragan Ilic, Lara McDonald, Shanika Palawaththa, Basia Diug, Saif U Munshi, and Md Nazmul Karim. A systematic review of dengue outbreak prediction models: Current scenario and future directions. *PLoS neglected tropical diseases*, 17(2):e0010631, 2023.
- Clarisse Lins de Lima, Ana Clara Gomes da Silva, Giselle Machado Magalhães Moreno, Cecilia Cordeiro da Silva, Anwar Musah, Aisha Aldosery, Livia Dutra, Tercio Ambrizzi, Iuri VG Borges, Merve Tunali, et al. Temporal and spatiotemporal arboviruses forecasting by machine learning: a systematic review. *Frontiers in Public Health*, 10:900077, 2022.
- Zhuanzhuan Liu, Qingxin Zhang, Liya Li, Junjie He, Jinyang Guo, Zichen Wang, Yige Huang, Zimeng Xi, Fei Yuan, Yiji Li, and Tingting Li. The effect of temperature on dengue virus transmission by aedes mosquitoes. *Frontiers in Cellular and Infection Microbiology*, 13:1242173, 2023. doi: 10.3389/fcimb.2023.1242173. URL <https://www.frontiersin.org/articles/10.3389/fcimb.2023.1242173/full>. Acessado em: 16 de julho de 2024.
- Juliana Barbosa da Silva Lotufo. *Validação de elementos climáticos estimados por sensoriamento remoto e reanálises na região Centro-Oeste do Brasil*. Tese de doutorado,

Programa de Pós-Graduação em Física Ambiental, Instituto de Física, Universidade Federal de Mato Grosso, Cuiabá, Mato Grosso, Brasil, December 2023.

Rachel Lowe, Sophie A. Lee, Kathleen M. O'Reilly, Oliver J. Brady, Leonardo Bastos, Gabriel Carrasco-Escobar, Rafael de Castro Catão, Felipe J. Colón-González, Christovam Barcellos, Marília Sá Carvalho, Marta Blangiardo, Håvard Rue, and Antonio Gasparrini. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in brazil: a spatiotemporal modelling study. *The Lancet Planetary Health*, 5(4):e209–e219, 2021. doi: 10.1016/S2542-5196(20)30292-8. URL [https://doi.org/10.1016/S2542-5196\(20\)30292-8](https://doi.org/10.1016/S2542-5196(20)30292-8). Acessado em: 16 de julho de 2024.

Mage.ai. How to interpret and explain your machine learning models using shap values. <https://m.mage.ai/how-to-interpret-and-explain-your-machine-learning-models-using-shap-values-471c2635b78e>, 2025. Accessed: 2025-07-04.

Song Quan Ong, Pradeep Isawasan, Ahmad Mohiddin Mohd Ngesom, Hanipah Shahar, As' malia Md Lasim, and Gomesesh Nair. Predicting dengue transmission rates by comparing different machine learning models with vector indices and meteorological data. *Scientific reports*, 13(1):19129, 2023.

Ana Victoria Ponce-Bobadilla, Vanessa Schmitt, Corinna S Maier, Sven Mensing, and Sven Stodtmann. Practical guide to shap analysis: explaining supervised machine learning model predictions in drug development. *Clinical and Translational Science*, 17(11):e70056, 2024.

Gokulan Ravindiran, Gasim Hayder, Karthick Kanagarathinam, Avinash Alagumalai, and Christian Sonne. Air quality prediction by machine learning models: A predictive study on the indian coastal city of visakhapatnam. *Chemosphere*, 338:139518, 2023.

Alessandro Sebastianelli, Dario Spiller, Raquel Carmo, James Wheeler, Artur Nowakowski, Ludmilla Viana Jacobson, Dohyung Kim, Hanoch Barlevi, Zoraya El Raiss Cordero, Felipe J Colon-Gonzalez, et al. A reproducible ensemble machine learning approach to forecast dengue outbreaks. *Scientific reports*, 14(1):3807, 2024.

Secretaria da Saúde Gov RS. Dengue. <https://saude.rs.gov.br/dengue-cuidados>. Acessado em: 22 de julho de 2025.

Chuin-Shee Shang, Chi-Tai Fang, Chung-Ming Liu, Tzai-Hung Wen, Kun-Hsien Tsai, and Chwan-Chuen King. The role of imported cases and favorable meteorological conditions in the onset of dengue epidemics. *PLoS Neglected Tropical Diseases*, 4(8):e775, 2010.

- M. O. Silva, S. Moreno Rodrigues, et al. Accuracy of dengue reporting by national surveillance system, brazil. *Emerging Infectious Diseases*, 22(2):336–339, 2016. doi: 10.3201/eid2202.150495. URL <https://wwwnc.cdc.gov/eid/article/22/2/15-0495-t1>.
- Tatiane Cristina Moraes de Sousa, Flavia Amancio, Sandra de Sousa Hacon, and Christovam Barcellos. Doenças sensíveis ao clima no brasil e no mundo: revisão sistemática. *Revista Panamericana de Salud Pública*, 42:e85, 2018.
- Eduardo D. Vicentinia and Antonio G. Sampaio de Oliveira-Filho. Introdução às redes neurais para regressões não-lineares: Ajuste de superfícies de energia potencial. *Química Nova*, 44(2):229–234, 2021. doi: 10.21577/0100-4042.20170650. URL <http://dx.doi.org/10.21577/0100-4042.20170650>.
- Daniel AM Villela. Predicting high dengue incidence in municipalities of brazil using path signatures. *Scientific Reports*, 15(1):26733, 2025.
- WHO. World health organization. URL <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>.
- Cort Willmott, Steven Ackleson, Robert Davis, Johannes Feddema, Katherine Klink, David Legates, James O’Donnell, and Clinton Rowe. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research*, 09 1985. doi: 10.1029/JC090iC05p08995.
- Yanguo Zhou, Liu Yuan, Ran Bai, Chenyu Liu, and Yi Zhang. A hybrid model for dengue epidemic prediction. In *Proceedings of the 2024 5th International Symposium on Artificial Intelligence for Medicine Science*, pages 327–331, 2024.