

Aprendizado de Máquina

Monitoramento e Análise Preditiva de
Dengue em Campinas

Contextualização

Cenário

1. Cidade de Campinas enfrenta surtos de epidemia de Dengue.
2. É algo recorrente. Acontece todos os anos, principalmente no verão.
3. Sobrecarrega o sistema de saúde público.

Problema

Dificuldade em alocar recursos preventivos devido a ausência de dados objetivos a respeito da sazonalidade (locais de maiores concentrações de casos) e dos grupos de riscos (idosos e crianças).

Objetivo

Mapear o comportamento temporal e demográfico da doença em 2025.

Explorar as bases de dados de dengue para responder perguntas-chave sobre tempo, espaço e perfil dos casos. A motivação é gerar evidências objetivas para mapear sazonalidade, localizar áreas prioritárias e perfis de risco, e apoiar a alocação de recursos e ações de controle do vetor, diante dos recentes casos de dengue em Campinas e do impacto na rede de saúde.

Metodologia

Fonte

Dados públicos do SINAN GOV – Arboviroses.

Recorte

Casos notificados em Campinas no ano de 2025.

Volume de dados

21.671 notificações processadas de 1.000.000 de dados registrados.

Tratamento dos dados

- Limpeza de duplicatas e valores nulos.
- Conversão de datas e idades.
- Categorização da gravidade (dengue clássica ou grave).

Análise Exploratória

O que aconteceu em 2025?

Pico sazonal

Alta concentração no verão e início do Outono

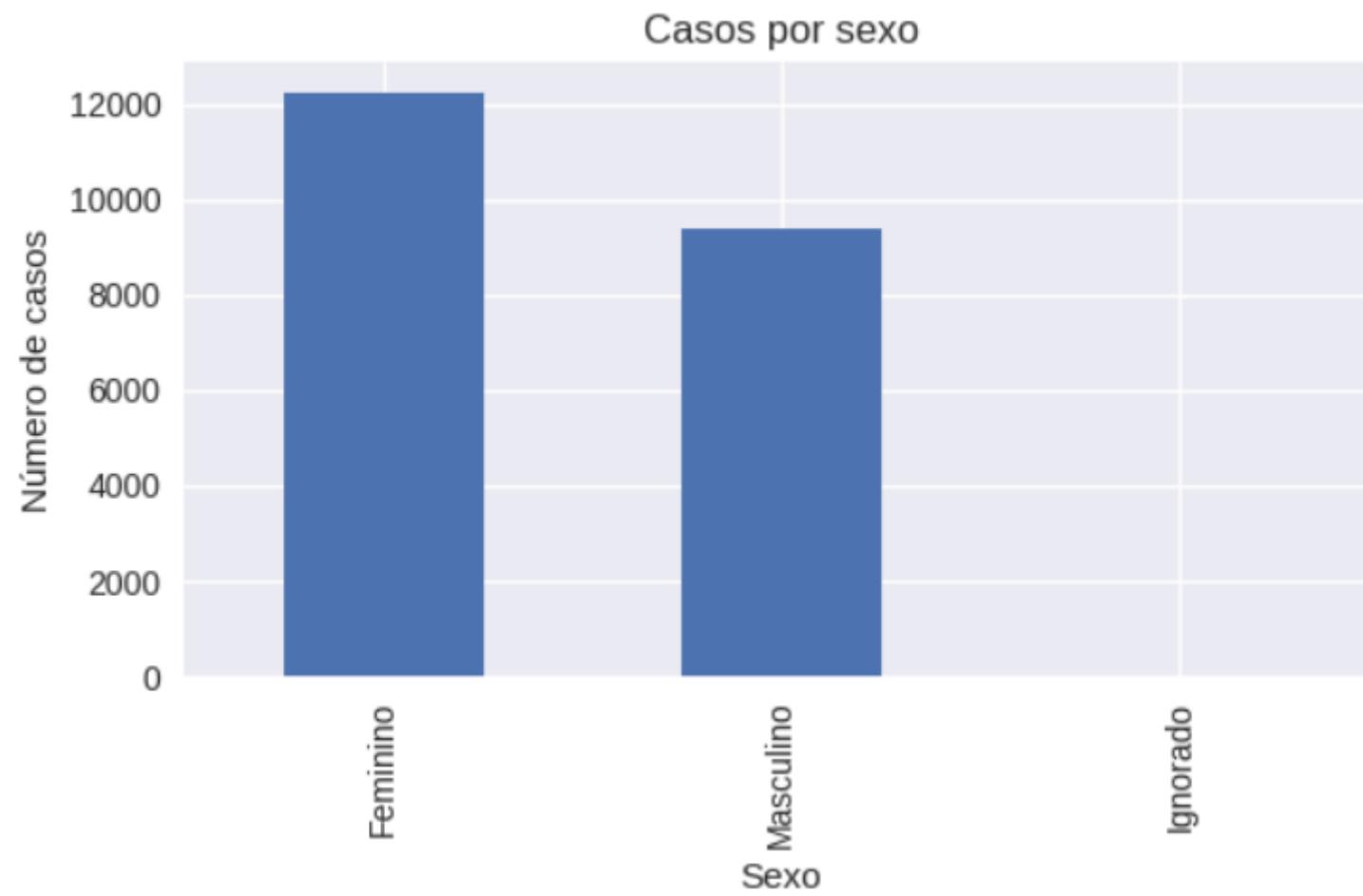
Meses

Março, Fevereiro e Abril concentraram **92%** dos casos do ano.

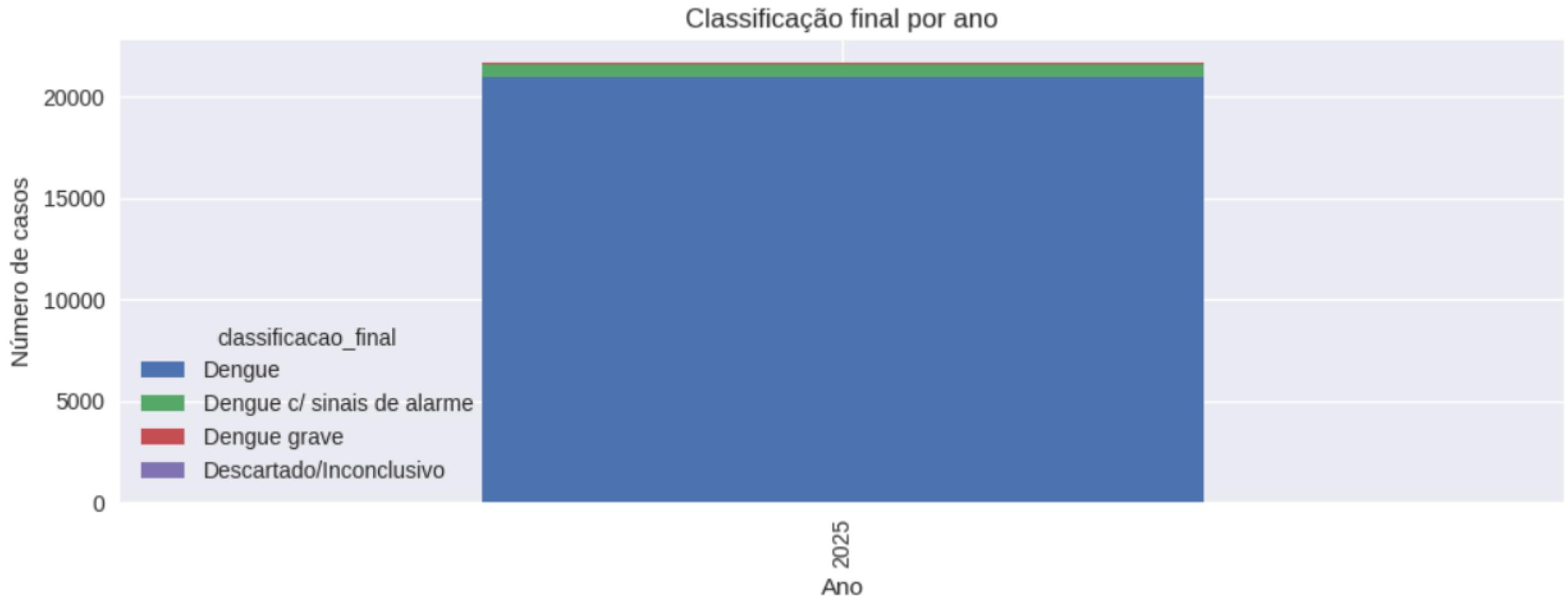
Tempo de Resposta

A mediana de atraso entre o início dos sintomas e a notificação foi de **3 dias**.

Análise Exploratória - Gráficos de Exemplos



Análise Exploratória - Gráficos de Exemplos



Perfil de Gravidade e Óbitos

Classificação dos casos

1. A maioria dos casos são leves
2. Os casos de sinais grave representam **3,1%** do total
3. Total de **13 óbitos**
4. Taxa de Letalidade (CFR): 0.59 óbitos a cada 1.000 casos
5. Os meses de março e abril representam as maiores taxas de letalidade

Abordagem e modelagem

Definição do Alvo [Target]

A variável alvo foi a severidade dos casos de dengue, no qual os casos graves foram classificados como (1) e os casos clássicos foram classificados como (0)

Algoritmos escolhidos

- Regressão Logística
- Random Forest (Floresta Aleatória)
- XGBoost (Extreme Gradient Boosting)

Divisão dos dados

Após todas as transformações e preparações, a base de dados foi dividida em conjuntos de treino (X, Y) e teste (X, Y). Uma proporção de **80% dos dados foi alocada para treino e 20% para teste.**

Resultados dos Modelos

Modelo ganhador

Random Forest e XGBoost, foram os modelos de melhor desempenho. O Random Forest destacou-se com o maior *Recall* e *ROC-AUC*, sendo excelente na identificação de casos severos. O XGBoost obteve o maior *F1-Score*, equilibrando precisão e recall. Ambos são recomendados

Variáveis mais importantes

- estacao_Verão
- HOSPITALIZ
- LEUCOPENIA
- CS_ESCOL_N
- CS_FLXRET

Conclusões e impacto

Diagnóstico

2025 foi um ano de alta concentração no primeiro quadrimestre exigindo mobilização rápida em Fevereiro e Março

Eficácia do Modelo

Random Forest: Alcançou uma Acurácia de **97.6%**, Precisão de **97.3%**, Recall de **97.6%** e F1-Score de **97.1%**. Seu ROC-AUC foi de **0.937**, indicando uma excelente capacidade de discriminação entre as classes de severidade.

XGBoost: Apresentou uma Acurácia de **97.3%**, Precisão de **97.3%**, Recall de **97.3%** e o melhor F1-Score de **97.3%**. Seu ROC-AUC foi de **0.926**.

Aplicações dos resultados na prática

- Suporte à triagem hospitalar
- Planejamento de estoque de insumos antes dos meses de pico
- Melhoria na visualização das informações

Obrigado!