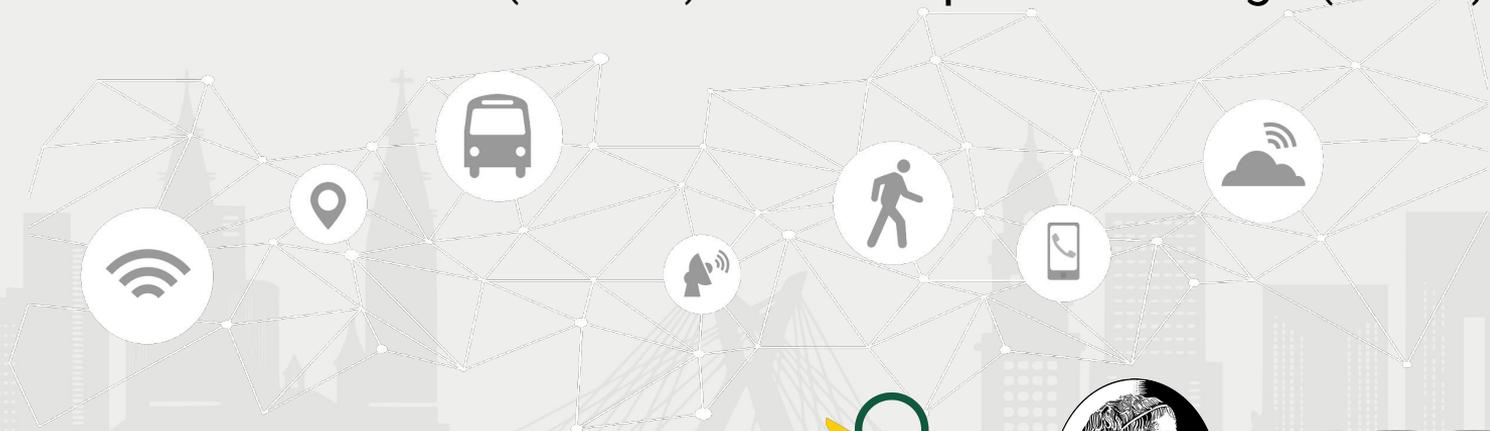


Visualizando e Prevendo a Saúde Urbana com Aprendizado de Máquina - Sífilis e Dengue

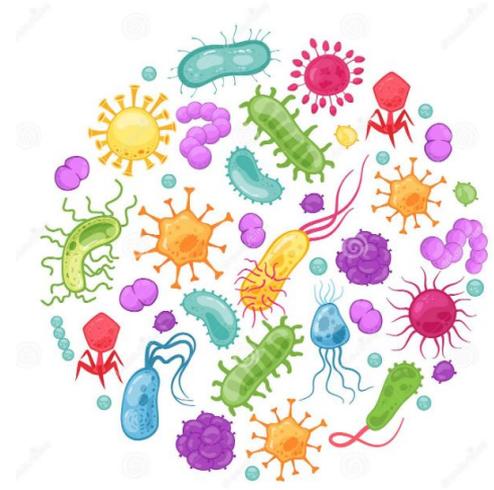
Alunos de Mestrado: Gabriely Pereira e Robson Aleixo (IME-USP)

Orientadores: Prof. Fabio Kon (IME-USP) e Prof. Raphael Camargo (UFABC)



Introdução

- A importância da epidemiologia
- Novas abordagens, aprendizado de máquina.
- Auxiliar as políticas públicas voltadas para saúde.
- Parceria com o Instituto de Estudos para Políticas de Saúde (IEPS) - FGV
- Trabalho em andamento



Doenças analisadas

- Sífilis

G1 CIÊNCIA E SAÚDE

Brasil tem 18 casos de sífilis por hora, diz Ministério da Saúde

Aumento é de 28,3%, comparado ao ano anterior. Ministro esteve no Rio neste sábado (7) para lançar Campanha de Prevenção às Infecções Sexualmente Transmissíveis.

Por Carlos Brito, G1 Rio

08/02/2020 12h52 - Atualizado há 4 meses

[f](#) [t](#) [w](#) [in](#)

<https://g1.globo.com/ciencia-e-saude/noticia/2020/02/08/brasil-tem-18-casos-de-sifilis-por-hora-diz-ministerio-da-saude.ghtml>

- Dengue

Ministério da **Saúde**

Sistemas | Ouvidoria | Comunicação e Imprensa | Contatos | Assessoria de Imprensa

VOCÊ ESTÁ AQUI: PÁGINA INICIAL > ÚLTIMAS NOTÍCIAS > AGÊNCIA SAÚDE > CRESCE EM 264% O NÚMERO DE CASOS DE DENGUE NO PAÍS

Cresce em 264% o número de casos de dengue no país

Publicado: Segunda, 25 de Março de 2019, 12h06
Última atualização em Terça, 26 de Março de 2019, 17h24

[Tweeter](#) [Curtir 25 mil](#)

Os óbitos pela doença também aumentaram 67%, entre 30 de dezembro e 16 de março de 2019, em comparação ao mesmo período de 2018, sendo a maior concentração no estado de São Paulo

<https://www.saude.gov.br/noticias/agencia-saude/45314-cresce-em-264-o-numero-d-e-casos-de-dengue-no-pais>

Conteúdo

- Bases de Dados de Saúde
- Visualização de dados
- Aprendizado de Máquina



Conteúdo

- **Bases de Dados de Saúde**
- Visualização de dados
- Aprendizado de Máquina



Bases de dados do SUS

- **SINAN** - Sistema de Informação de Agravos de Notificação:
 - SINAN Sífilis adquirida
 - SINAN Sífilis gestacional
 - SINAN Sífilis congênita

- **SINAN** - Sistema de Informação de Agravos de Notificação:
 - SINAN Dengue

- **SIM** - Sistema de Informação sobre Mortalidade
- **SINASC** - Sistema de Informações sobre Nascidos Vivos

As bases obtidas são referentes ao período de 2000 a 2017 para a cidade do Rio de Janeiro.

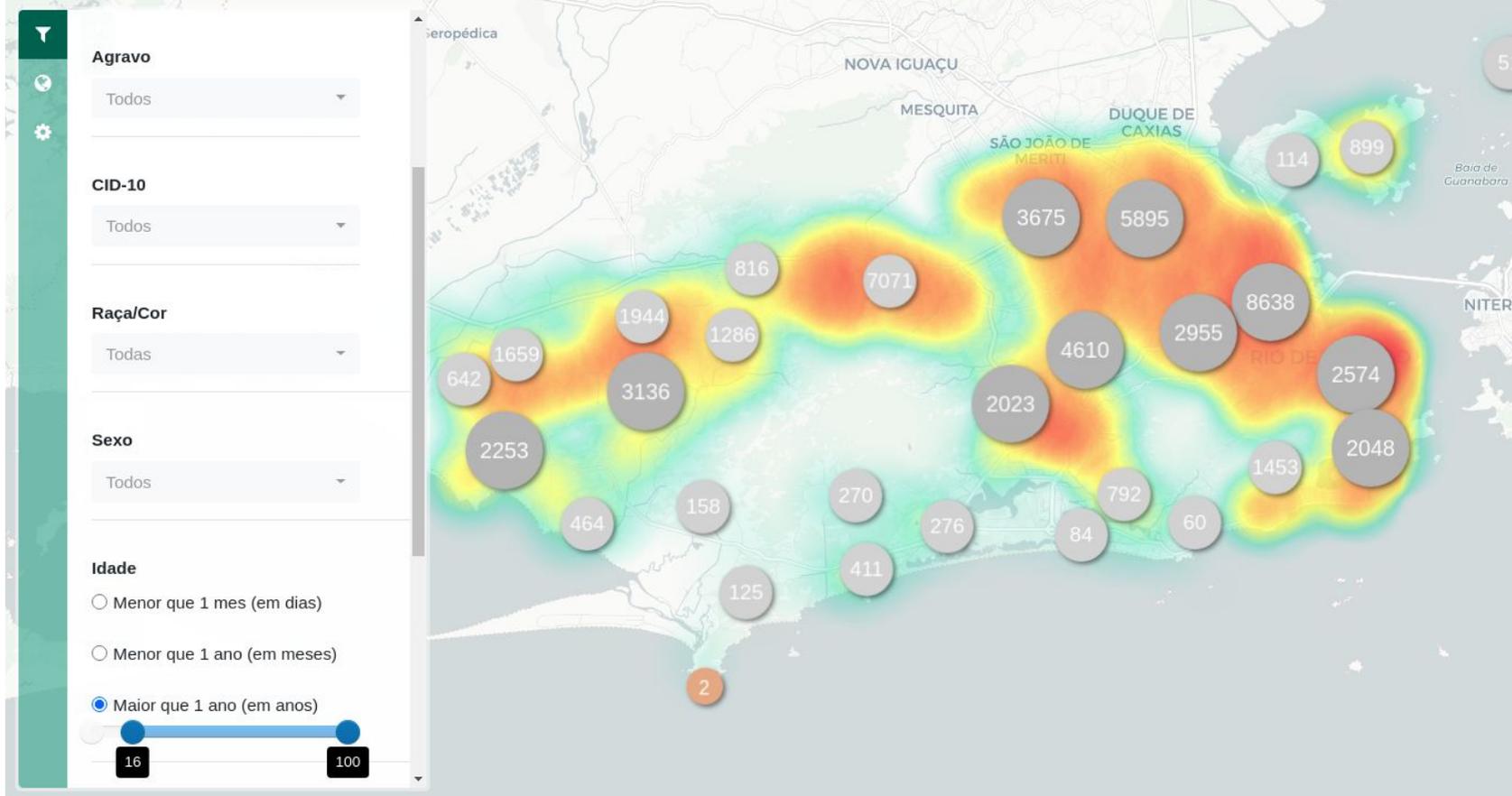
Conteúdo

- Bases de Dados de Saúde
- **Visualização de dados**
- Aprendizado de Máquina

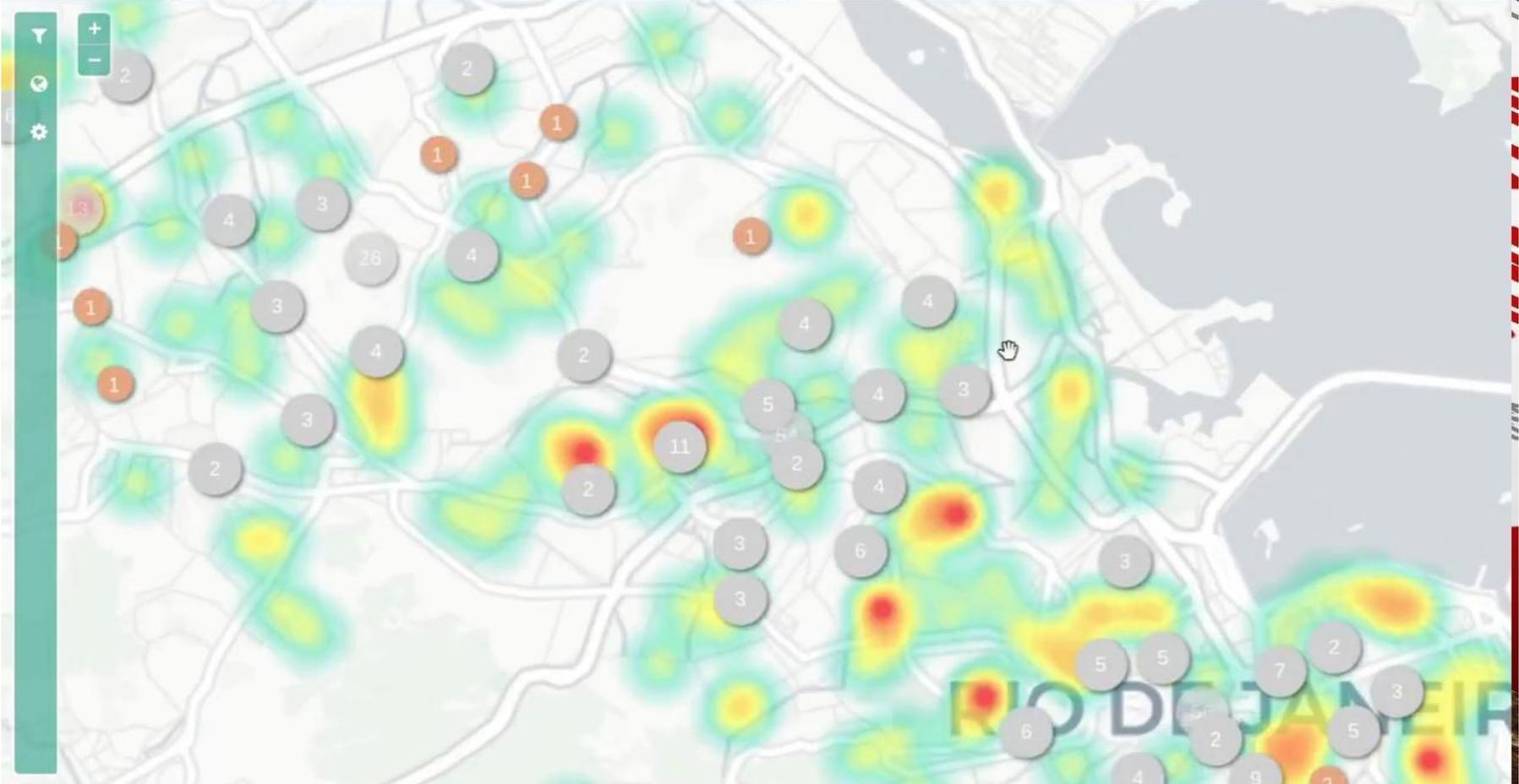


Visualização de dados

- Desenvolvimento de uma plataforma para **visualização georreferenciada** dos casos de Sífilis
- Objetivo de auxiliar **gestores de saúde** para criação de políticas públicas
- Os dados exibidos no mapa não expõe a nível de endereço, mas sim a nível de **setor censitário**



Casos de sífilis em maiores de 16 anos no Rio de Janeiro.



Conteúdo

- Bases de Dados de Saúde
- Visualização de dados
- **Aprendizado de Máquina**



Aprendizado de Máquina

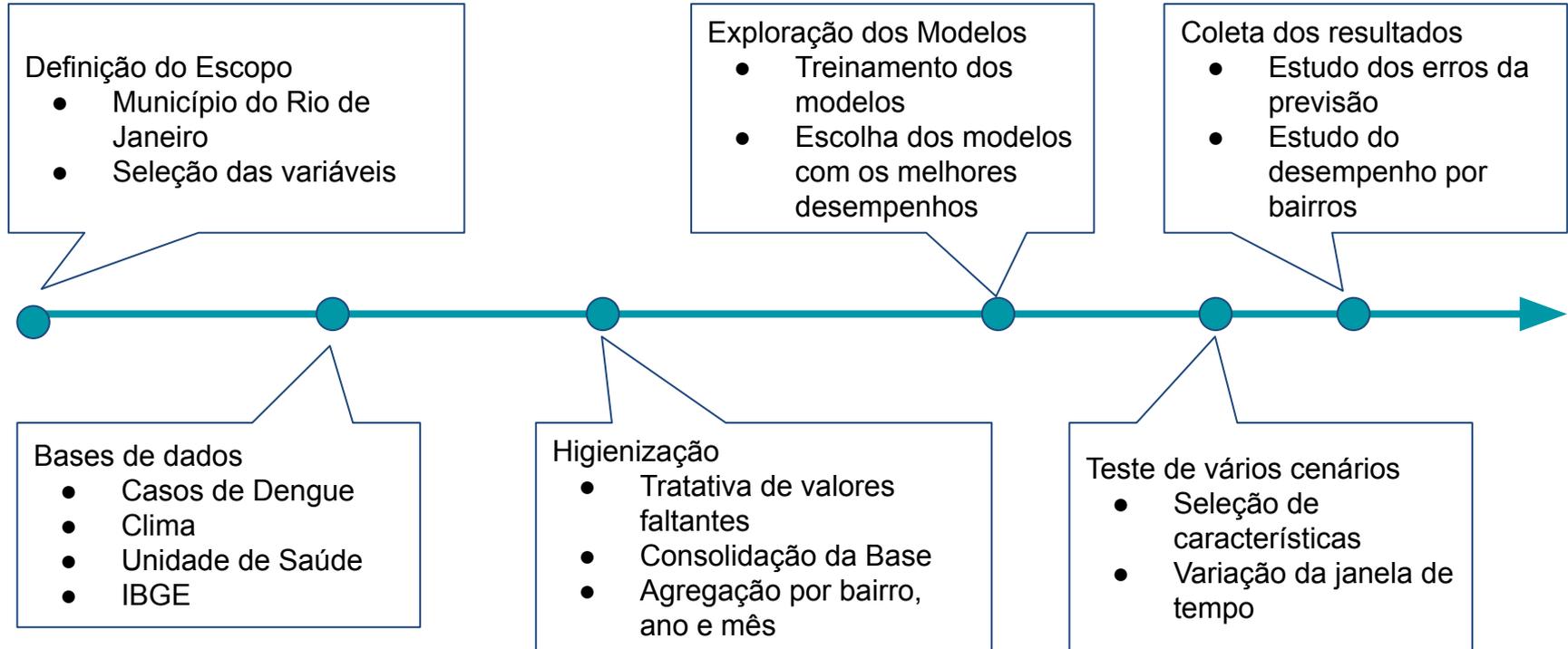
- Modelo de Regressão - Dengue

Utiliza dados sócio-demográfico-ambientais do passado para treinar um modelo para prever o número de casos futuros de dengue em cada bairro da cidade ao longo do tempo.

- Modelo de Classificação - Sífilis congênita

Utiliza dados sócio-demográficos do passado para treinar um modelo para prever se um determinado bebê adquirirá sífilis, por contágio da mãe.

O que foi feito até o momento - Dengue



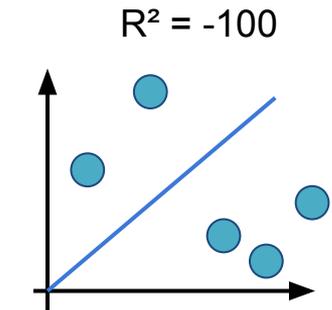
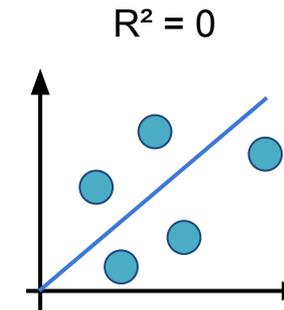
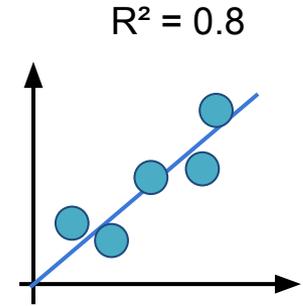
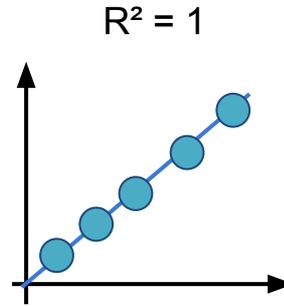
Informações utilizadas no modelo - Dengue

- Estações Meteorológicas
 - Temperatura
 - Precipitação
- Estabelecimentos de Saúde - SUS
 - Quantidade de estabelecimentos de saúde em cada bairro
 - Quantidade de serviços pré-natal e neonatal em cada bairro
 - Quantidade de serviços do programa da família e atendimento básico em cada bairro
- IBGE - Censo 2010
 - IDH
 - Taxa de alfabetização de adultos
 - Taxa de frequência escolar
 - Renda per capita
- Série histórica de casos de dengue em cada bairro da cidade agrupando da seguinte forma (meses):
 - t-1 até t-3
 - t-1 até t-6
 - t-1 até t-9
 - t-1 até t-12
- Informações sobre a ocorrência de dengue nos bairros vizinhos.
 - Valores máximo e mínimo

Como avaliar se as previsões estão corretas

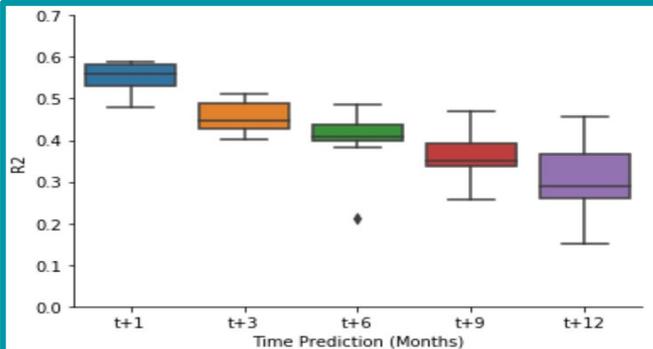
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad R^2 = (-\infty, 1)$$

- Floresta Aleatória: $R^2 = 0.48 \pm 0.11$
- Árvores Extras: $R^2 = 0.42 \pm 0.12$
- Regressão Linear: $R^2 = -3.3 \pm 3.38$
- Regressão Linear Múltipla: $R^2 = 0.35 \pm 0.09$

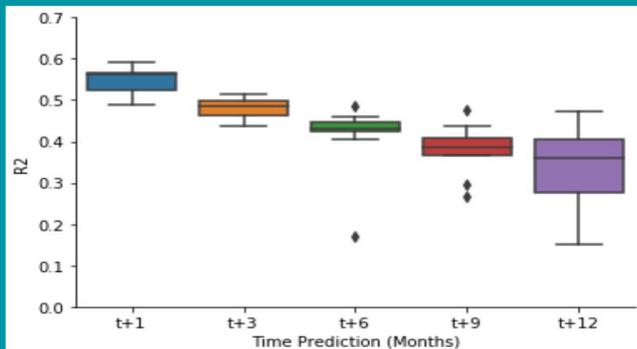


Resultados Preliminares - Dengue

Janela de 3 meses (t-1 até t-3)



Janela de 6 meses (t-1 até t-6)

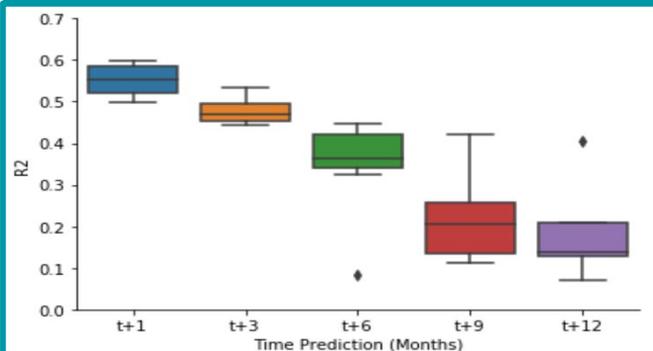


Base de Treino:
2014 - 2016

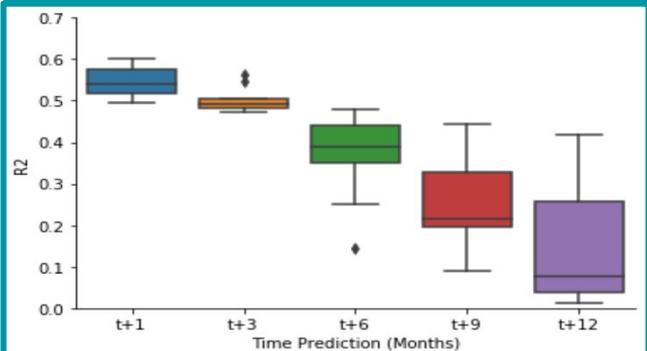
Base de Teste:
2017

Validação Cruzada
Estratégia de partição: 10

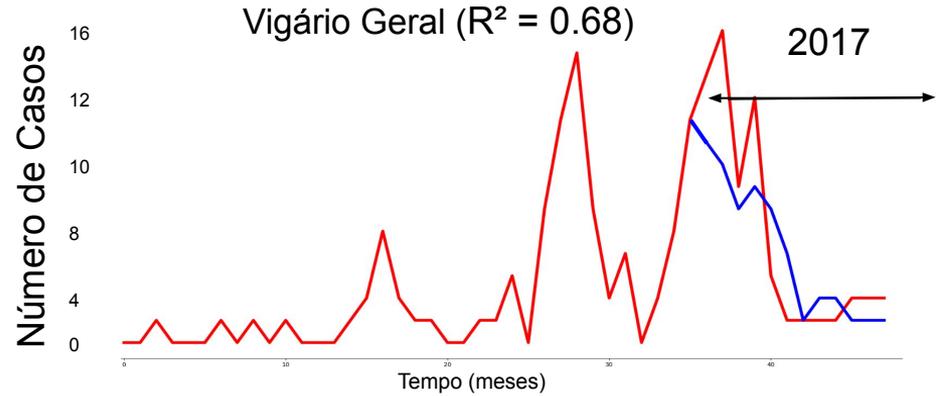
Janela de 9 meses (t-1 até t-9)



Janela de 12 meses (t-1 até t-12)



Resultados Preliminares - Dengue

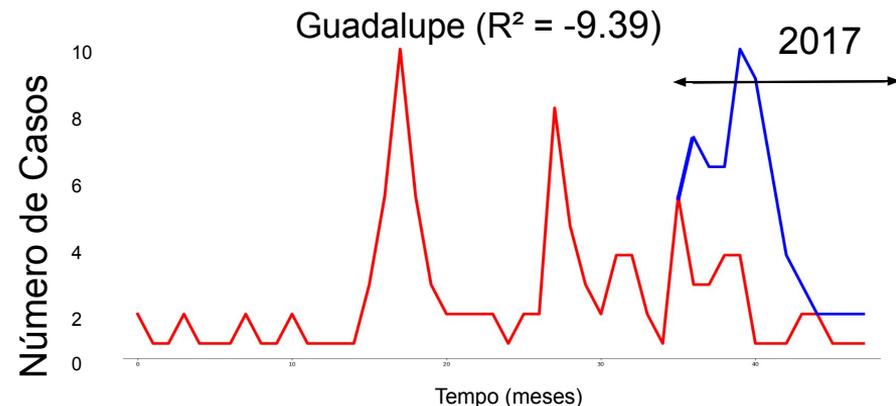
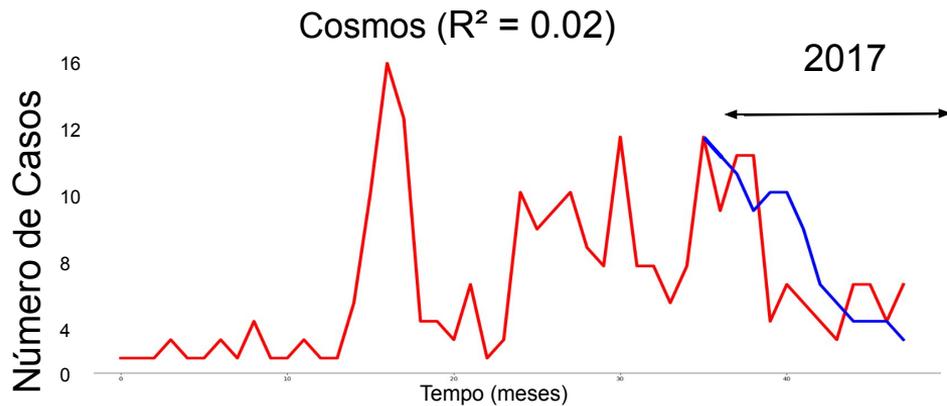


Base de Teste: 2017

Valores Previstos

Valores Reais

(previsão ruim)



Conclusões - Dengue

- Floresta Aleatória foi a técnica que obteve o melhor desempenho
- Janela de 6 meses foi a que mostrou melhor desempenho para previsões de longo prazo (12 meses)
- Existe diferença no valor do R^2 entre os bairros

Trabalhos em Andamento - Dengue

- Aplicação do modelo em outras doenças (p. ex.: Influenza)
- Considerar informações de bairros vizinhos
- Aplicação da rede neural *Long Short-Term Memory* (LSTM)
- Estender a análise para todo o estado do Rio de Janeiro
- Incluir novas variáveis
- Melhoria do modelo para diminuição do erro

O que foi feito até o momento - Sífilis

- Definição do escopo inicial:
 - Objetivo: Sífilis Congênita
 - Região: cidade do Rio de Janeiro
- Higienização das bases (SINASC, SINAN Gestante e SINAN Congênita)
- Seleção das variáveis a serem utilizadas para treinar o modelo
- Treinamento dos modelos
- Testes e análise dos primeiros resultados

Resultados Preliminares - Sífilis

Avaliação das Métricas

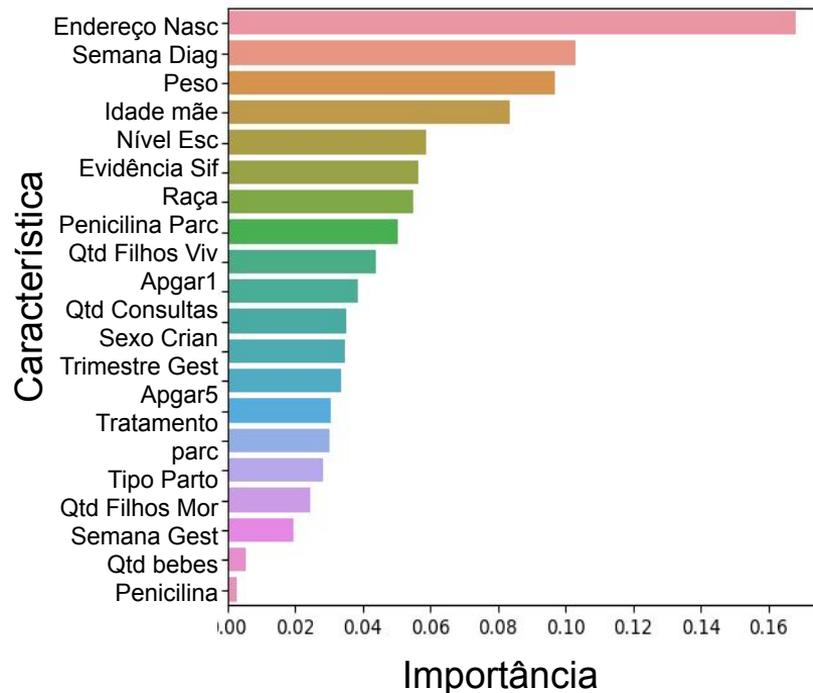
	R-True	R-False
P-True	606	131
P-False	230	537

- Acurácia: 0.78 ± 0.01
- Precisão: 0.81 ± 0.02
- Cobertura: 0.70 ± 0.02
- Métrica-F1: 0.75 ± 0.01

Trabalhos em andamento

- Seleção de características
- Ampliação para o estado do Rio de Janeiro
- Aplicação de variáveis das regiões

Análise das Características





INCT

InterSCity



IME-USP



UFABC



CAPES



interscity.org

raphael.camargo@ufabc.edu.br

gabriely.pereira@usp.br

robson.aleixo@usp.br