

Série dos Seminários de Acompanhamento à Pesquisa

DEI
DEPARTAMENTO
DE ENGENHARIA
INDUSTRIAL

Número 19 | 09 2021

Efeitos das Medidas de Contenção no Número de Casos da COVID-19 nos Estados Brasileiros

Autor:

Paulo Henrique C. Simões



Série dos Seminários de Acompanhamento à Pesquisa



Número 19 | 04 2021

Efeitos das Medidas de Contenção no Número de Casos da COVID-19 nos Estados Brasileiros

Autor:

Paulo Henrique C. Simões

Orientadora: Paula Medina Maçaira
Coorientadora: Fernanda Araujo Baião



CRÉDITOS:

SISTEMA MAXWELL / LAMBDA
<https://www.maxwell.vrac.puc-rio.br/>

Organizadores: Fernanda Baião / Soraida Aguiar

Layout da Capa: Aline Magalhães dos Santos

Roteiro da apresentação

01

INTRODUÇÃO

02

BASE DE DADOS DE OXFORD

03

O MODELO XGBOOST

04

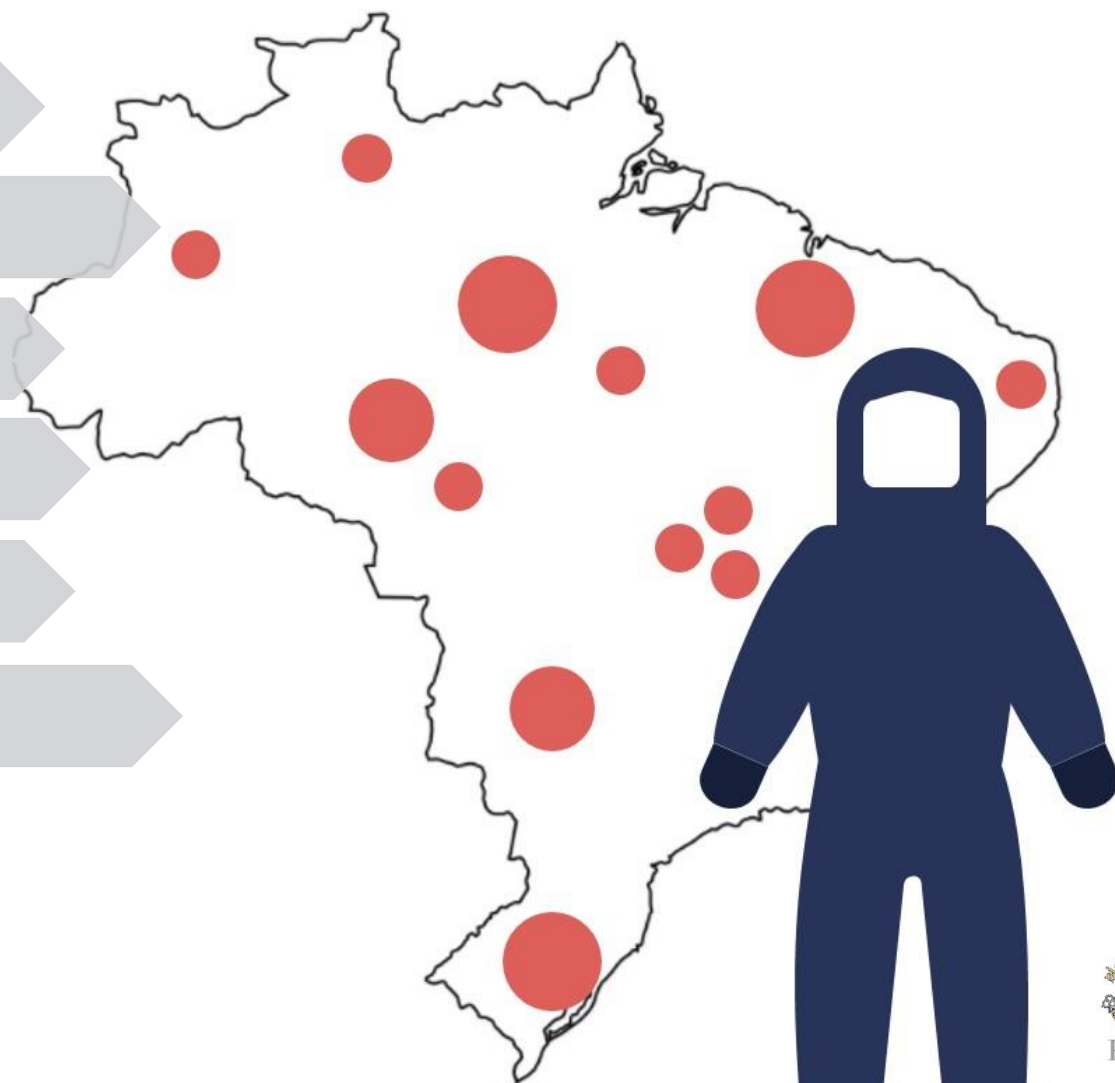
VISUAL TUNING XGBOOST WITH CARET

05

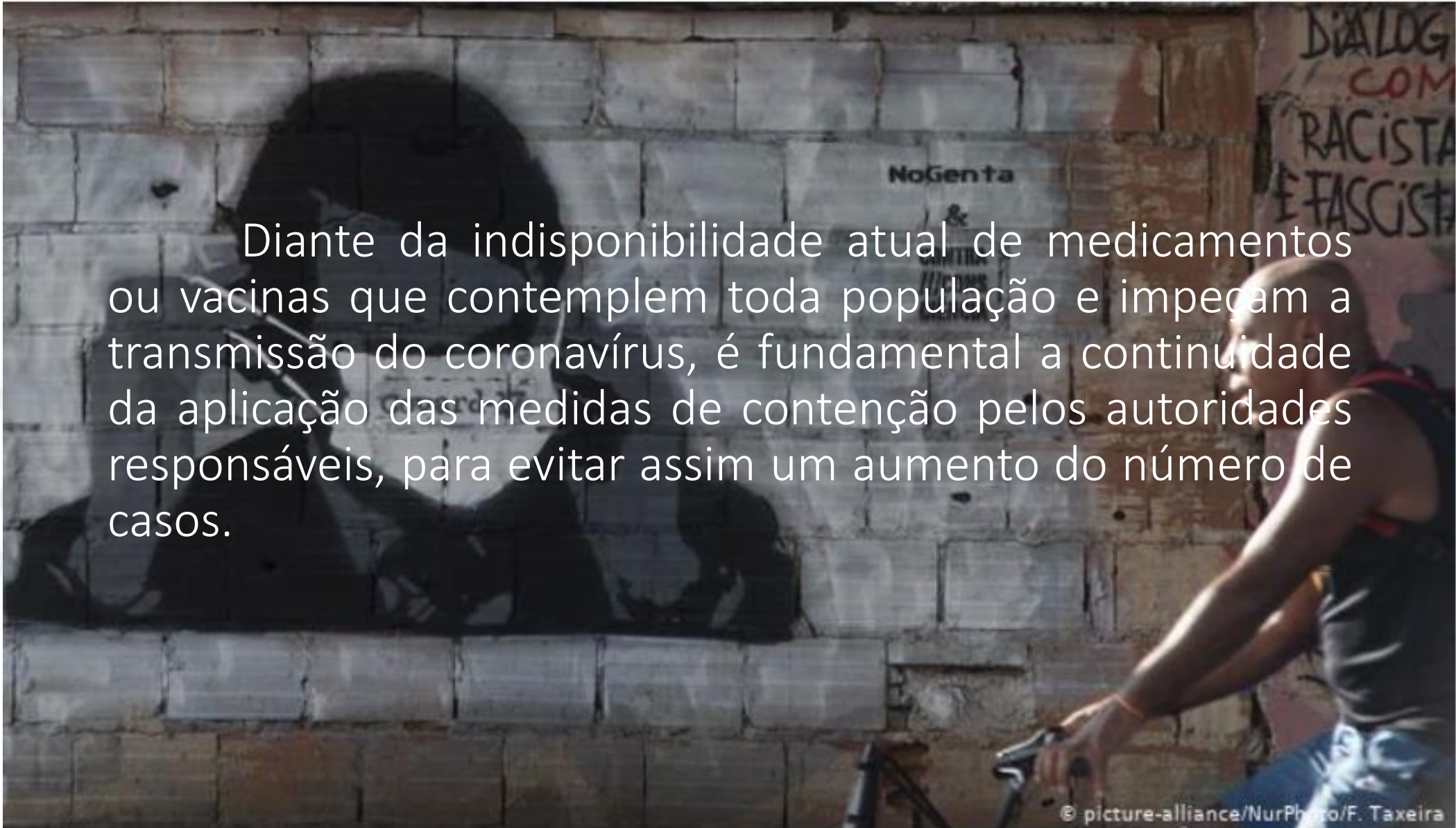
SHAPLEY VALUE

06

RESULTADOS PRELIMINARES



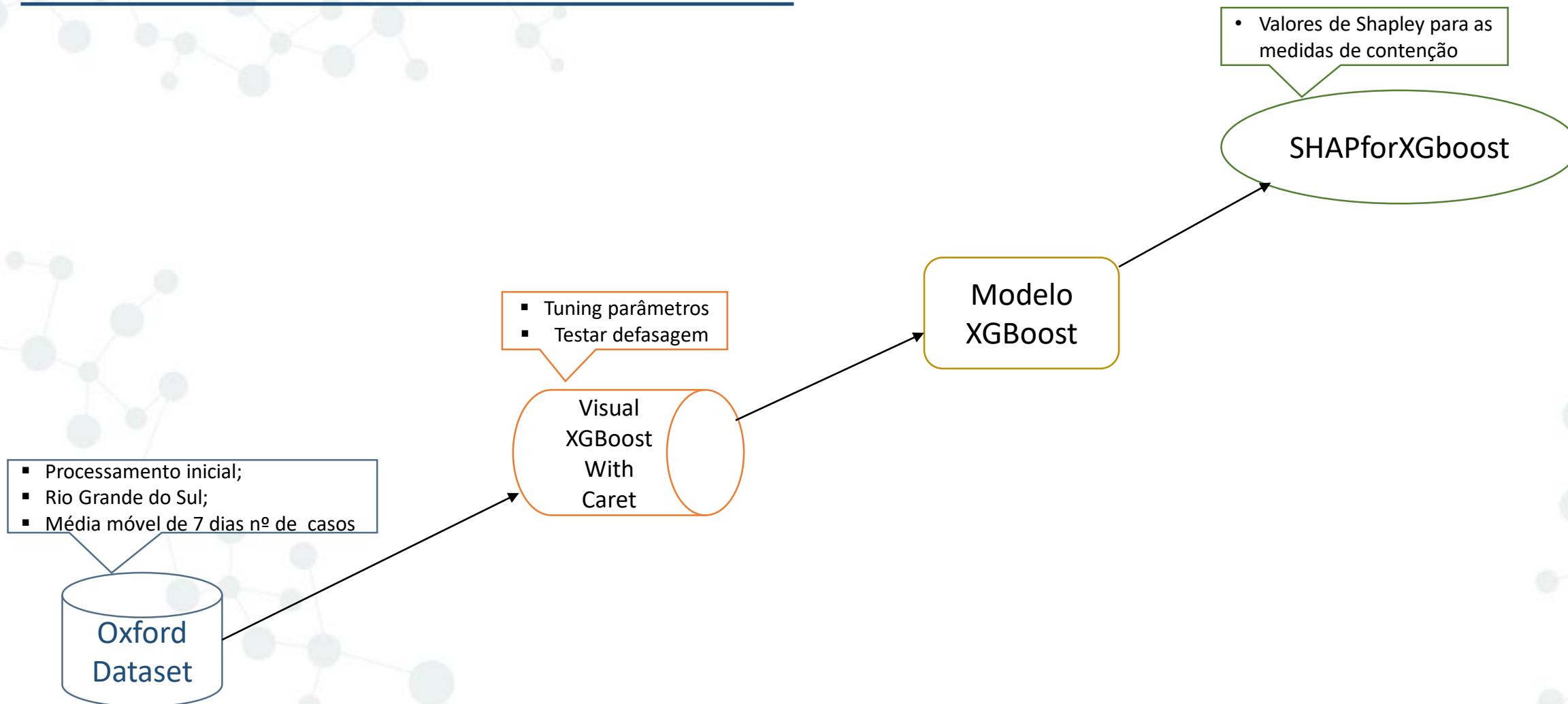
Diante da indisponibilidade atual de medicamentos ou vacinas que contemplem toda população e impeçam a transmissão do coronavírus, é fundamental a continuidade da aplicação das medidas de contenção pelas autoridades responsáveis, para evitar assim um aumento do número de casos.



- O objetivo do trabalho é identificar dentre as medidas de contenção à pandemia adotadas no Brasil (Rio Grande do Sul) quais foram mais eficientes no combate ao avanço da covid-19.

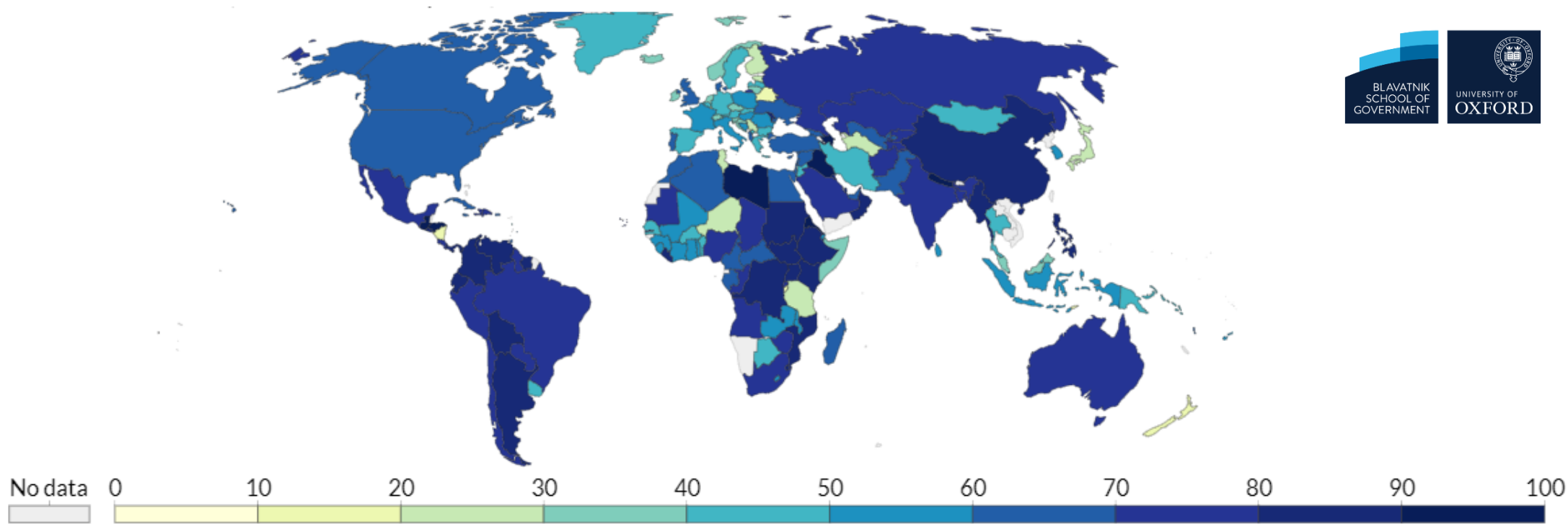


FLUXOGRAMA DO TRABALHO



DATASET

O Oxford COVID-19 Government Response Tracker (OxCGRT) fornece um sistema de medidas capaz de compreender como as respostas governamentais têm evoluído ao longo do tempo, registrando diariamente a política em vigor para cada um dos seus indicadores



Indicadores do Dataset

Contenção e fechamento	
Código	Nome
C1	Fechamento de escolas
C2	Fechamento de locais de trabalho
C3	Cancelamento de eventos públicos
C4	Restrições a aglomerações
C5	Restrições ao transporte público
C6	Confinamento domiciliar
C7	Restrições de movimentação interna
C8	Restrições a viagens internacionais

Indicadores do Dataset

Sistemas de saúde	
Indicador	Nome
H1	Campanha pública de informação
H2	Política de testagem
H3	Rastreamento de contatos
H4	Investimentos emergenciais em saúde
H5	Investimento em vacinas contra a Covid-19
H6	Proteção facial

Resposta econômica	
Indicador	Nome
E1	Suporte à renda
E2	Alívio contratual/de débitos para domicílios
E3	Medidas fiscais
E4	Concessão de auxílio internacional
Diverso	
M1	Outras respostas

Medidas de Contenção do Dataset que compõem o índice de restrição

Contenção e fechamento + H1

Indicador	Nome	Valor	Codificação
C1	Fechamento de escolas	0, 1, 2, 3	0 - No measures / 1 - recommend closing / 2 - Require closing (only some levels or categories, eg just high school, or just public schools) / 3 - Require closing all levels
C2	Fechamento de locais de trabalho	0, 1, 2, 3	0 - No measures / 1 - recommend closing (or work from home) / 2 - require closing (or work from home) for some sectors or categories of workers / 3 - require closing (or work from home) all-but-essential workplaces
C3	Cancelamento de eventos públicos	0, 1, 2	0- No measures / 1 - Recommend cancelling / 2 - Require cancelling
C4	Restrições a aglomerações	0, 1, 2, 3, 4	0 - No restrictions / 1 - Restrictions on very large gatherings (the limit is above 1000 people) / 2 - Restrictions on gatherings between 101-1000 people / 3 - Restrictions on gatherings between 11-100 people 4 - Restrictions on gatherings of 10 people or less
C5	Restrições ao transporte público	0, 1, 2	0 - No measures / 1 - Recommend closing (or significantly reduce volume, route, means of transport available) / 2 - Require closing (or prohibit most citizens from using it)
C6	Confinamento domiciliar	0, 1, 2, 3	0 - No measures / 1 - recommend not leaving house / 2 - require not leaving house with exceptions for daily exercise, grocery shopping, and 'essential' trips / 3 - Require not leaving house with minimal exceptions (e.g. allowed to leave only once a week, or only one person can leave at a time)
C7	Restrições de movimentação interna	0, 1, 2	0 - No measures / 1 - Recommend not to travel between regions, cities / 2 – internal movement restrictions in place
C8	Restrições a viagens internacionais	0, 1, 2, 3, 4	0 - No measures / 1 – Screening / 2 - Quarantine arrivals from high-risk regions / 3 - Ban on arrivals from some regions 4 – Ban on all regions or total border closure
H1	Campanha pública de informação	0, 1, 2	0 -No COVID-19 public information campaign / 1 - public officials urging caution about COVID-19 / 2 - coordinated public information campaign (e.g. across traditional and social media)

A base de dados produz um índice linear que ilustram as respostas governamentais: O Stringency Index condensa as medidas de contenção para cada dia da pandemia.

Stringency Index

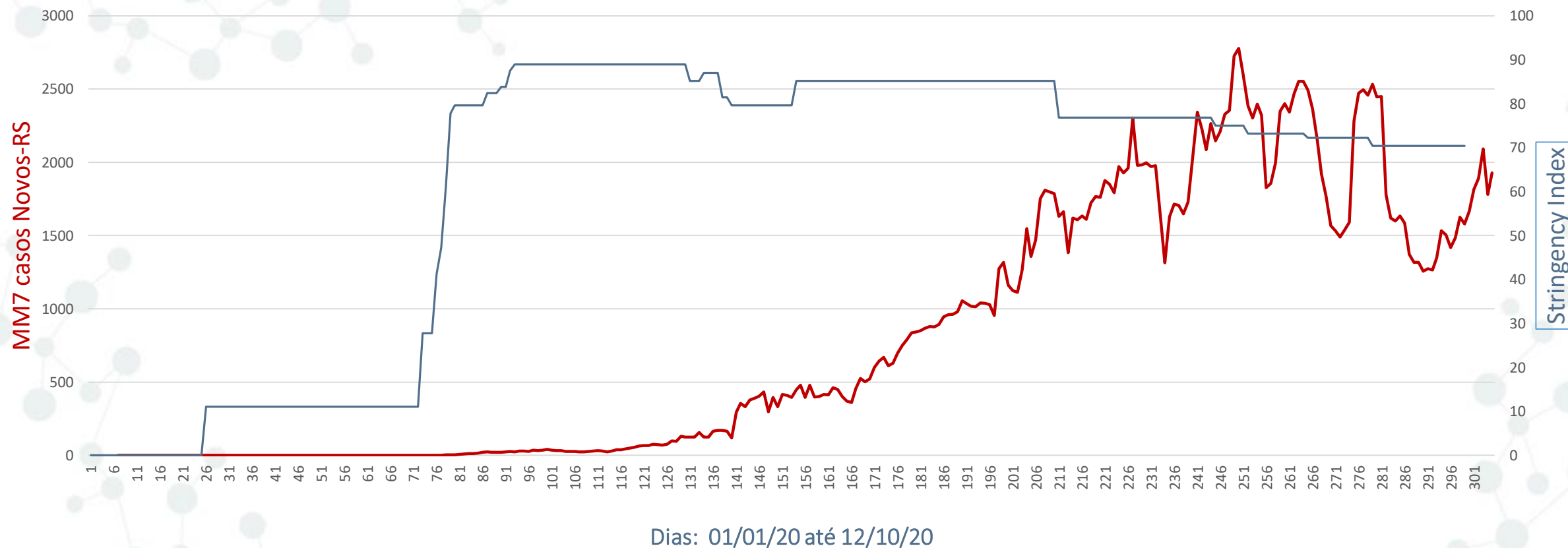
Indicador	C1	C2	C3	C4	C5	C6	C7	C8	H1
Valores	0, 1, 2, 3	0, 1, 2, 3	0, 1, 2	0, 1, 2, 3, 4	0, 1, 2	0, 1, 2, 3	0, 1, 2	0, 1, 2, 3, 4	0, 1, 2

- **Todos os índices são médias simples dos indicadores individuais.**
- **Os valores do índice de restrição estão entre 0 e 100.**

$$\text{Stringency index} = \frac{1}{9} \sum_{j=1}^k 100 \left(C_j \frac{0,71}{N_j} + 0,29 G_j \right)$$

onde: C_j é o indicador , N_j é o valor max do indicador e $G_j = 0$ politica local, 1 politica geral

Rio Grande do Sul - BR

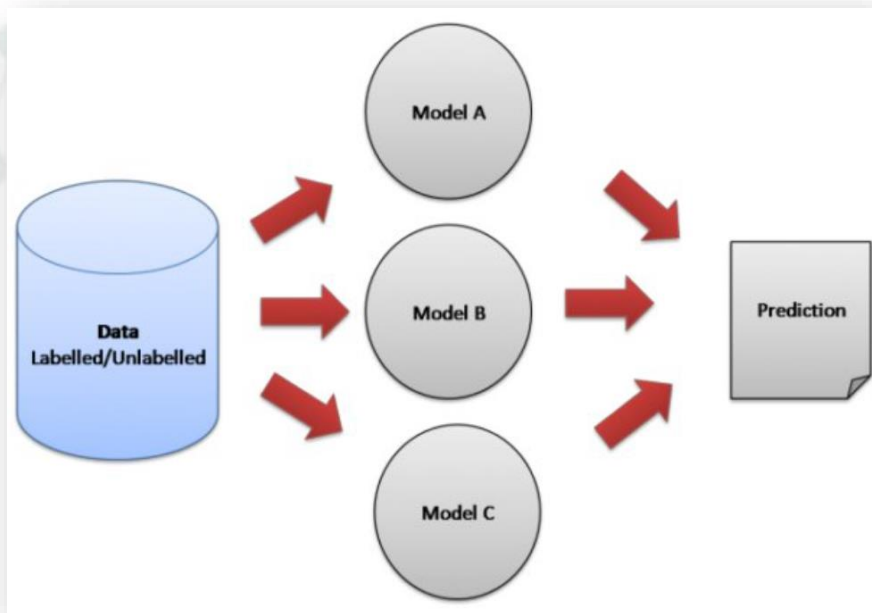


Método Ensemble:

A construção de um modelo utilizando metodologia Ensemble consiste em duas etapas:

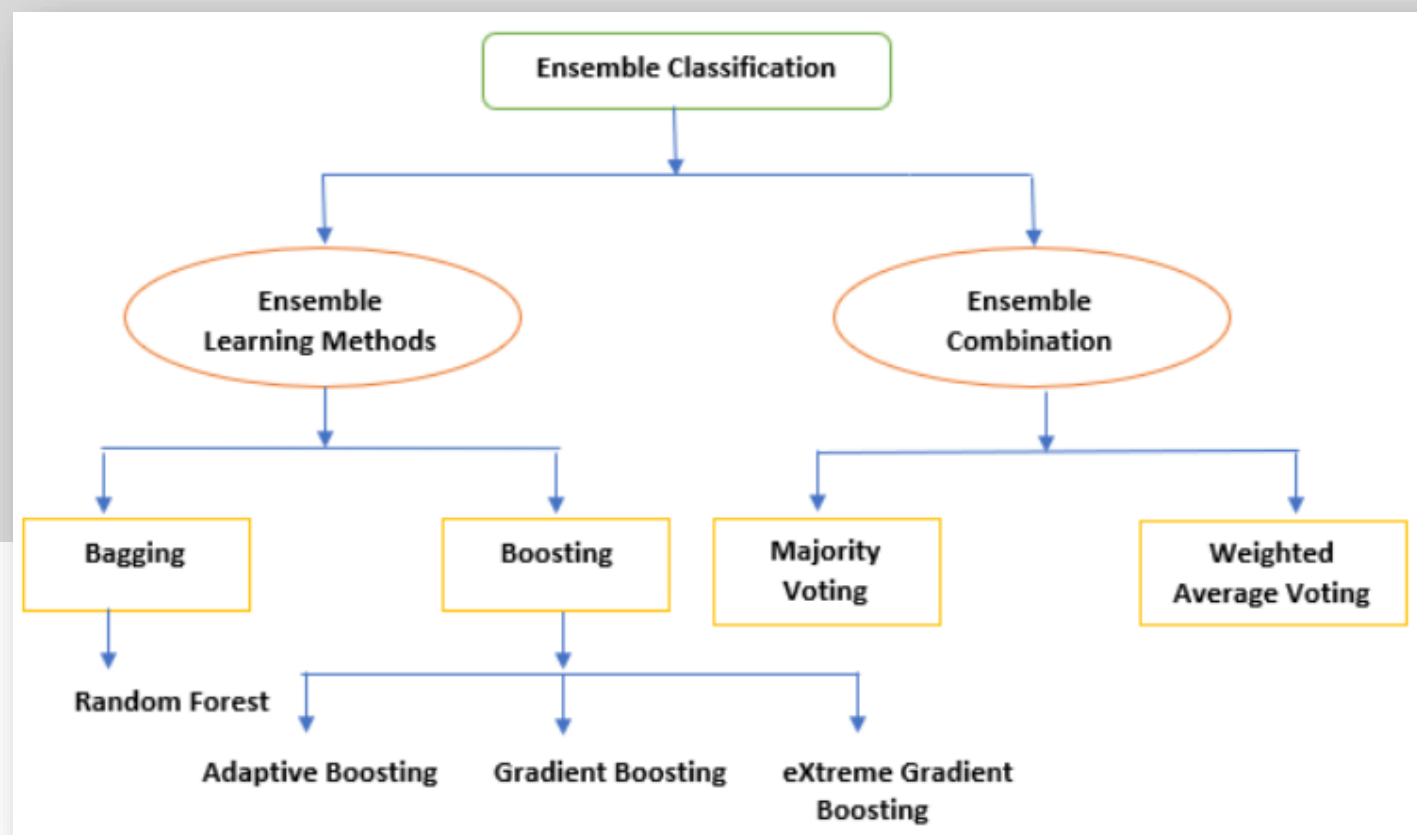
1º -> construção de diversos modelos.

2º -> combinação de suas estimativas.

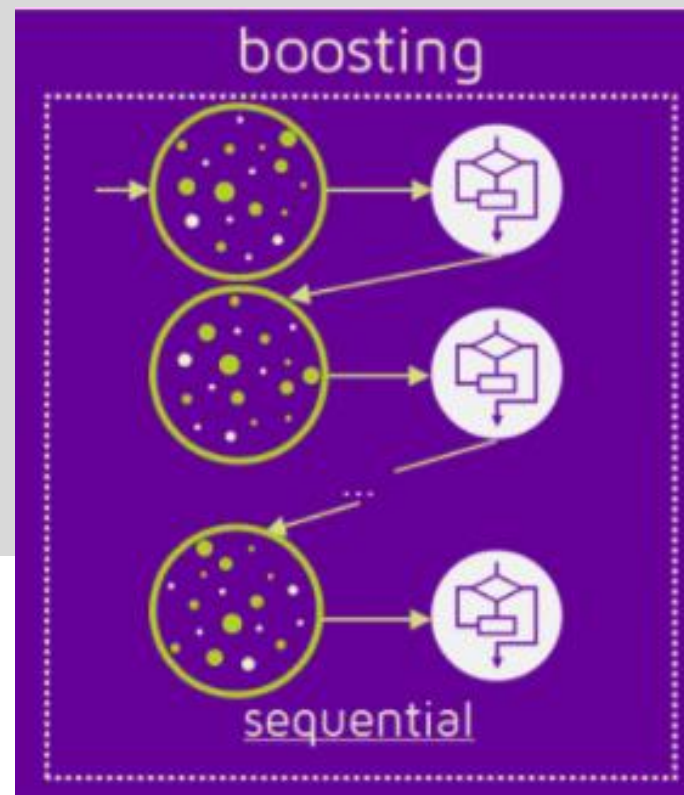


Categorias de Ensemble:

- ✓ Bagging.
- ✓ Boosting.
- ✓ Voting.

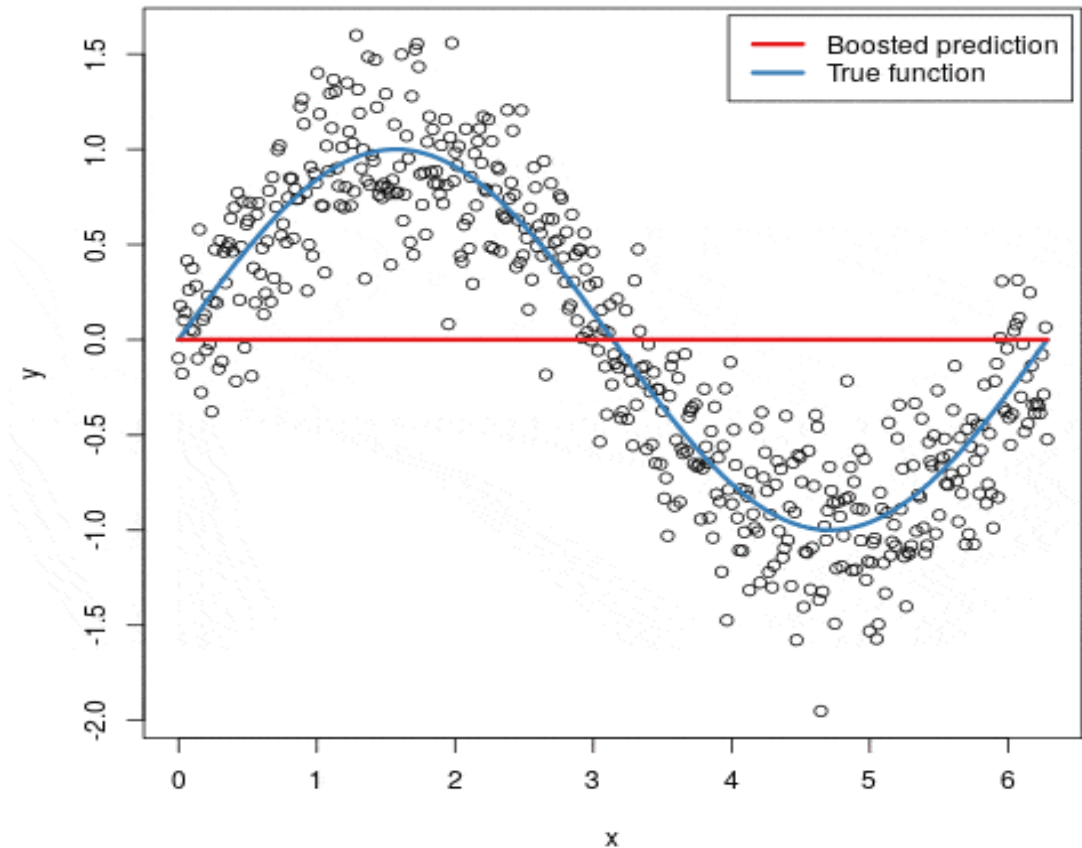


Gradient Boosting -> é usado para a construção de múltiplos modelos do mesmo tipo, onde cada novo modelo aprende a corrigir os erros gerados pelo modelo anterior, dentro da sequência de modelos criados.



Extreme Gradient Boosting

Cada novo modelo tentam prever a partir do erro dos modelos anteriores em vez de tentar prever independentemente o target.



List of the Tunable Hyperparameters

Hyperparameters	default	Descrição
Nrounds	100	Número de árvores
Max depth	6	Máxima profundidade da arvore
eta	0,3	Taxa de aprendizado
gamma	0	Usado para ajuste de regularização
Colsample bytree	1	Amostragem de coluna
Min child weight	1	Peso mínimo da folha
subsample	1	Amostragem de linha

- ❑ Nessa fase da modelagem utilizaremos as 5 etapas descritas a seguir para tunar os hiperparâmetros do modelo.
- ❑ Desenvolvido por Jani Pelkola para competição no kaggle.

Etapas do processo

Step 1: Number of Iterations and the Learning Rate;

Hiperparameters	Nrounds	Max depth	eta	Gamma	Colsample bytree	Min child weight	Subsample
Step 1	1000	2	0,3	0	1	1	1

Step 2: Maximum Depth and Minimum Child Weight;

Hiperparameters	Nrounds	Max depth	eta	Gamma	Colsample bytree	Min child weight	Subsample
Step 2	550	2	0,3	0	1	1	1

Step 3: Column and Row Sampling;

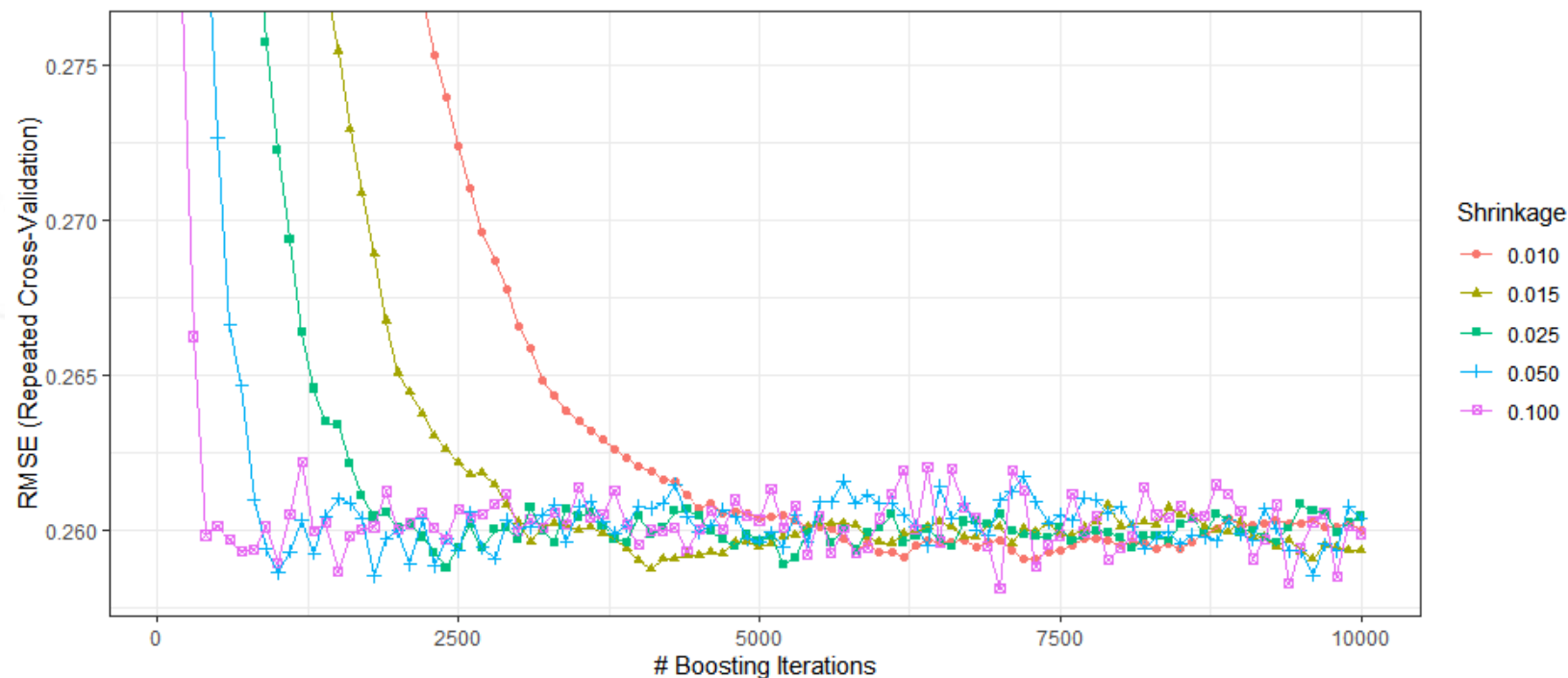
Hiperparameters	Nrounds	Max depth	eta	Gamma	Colsample bytree	Min child weight	Subsample
Step 3	300	2	0,3	0	0,6	1	0,5

Step 4: Gamma;

Hiperparameters	Nrounds	Max depth	eta	Gamma	Colsample bytree	Min child weight	Subsample
Step 4	800	2	0,3	0	0,6	1	0,5

Etapas do processo:

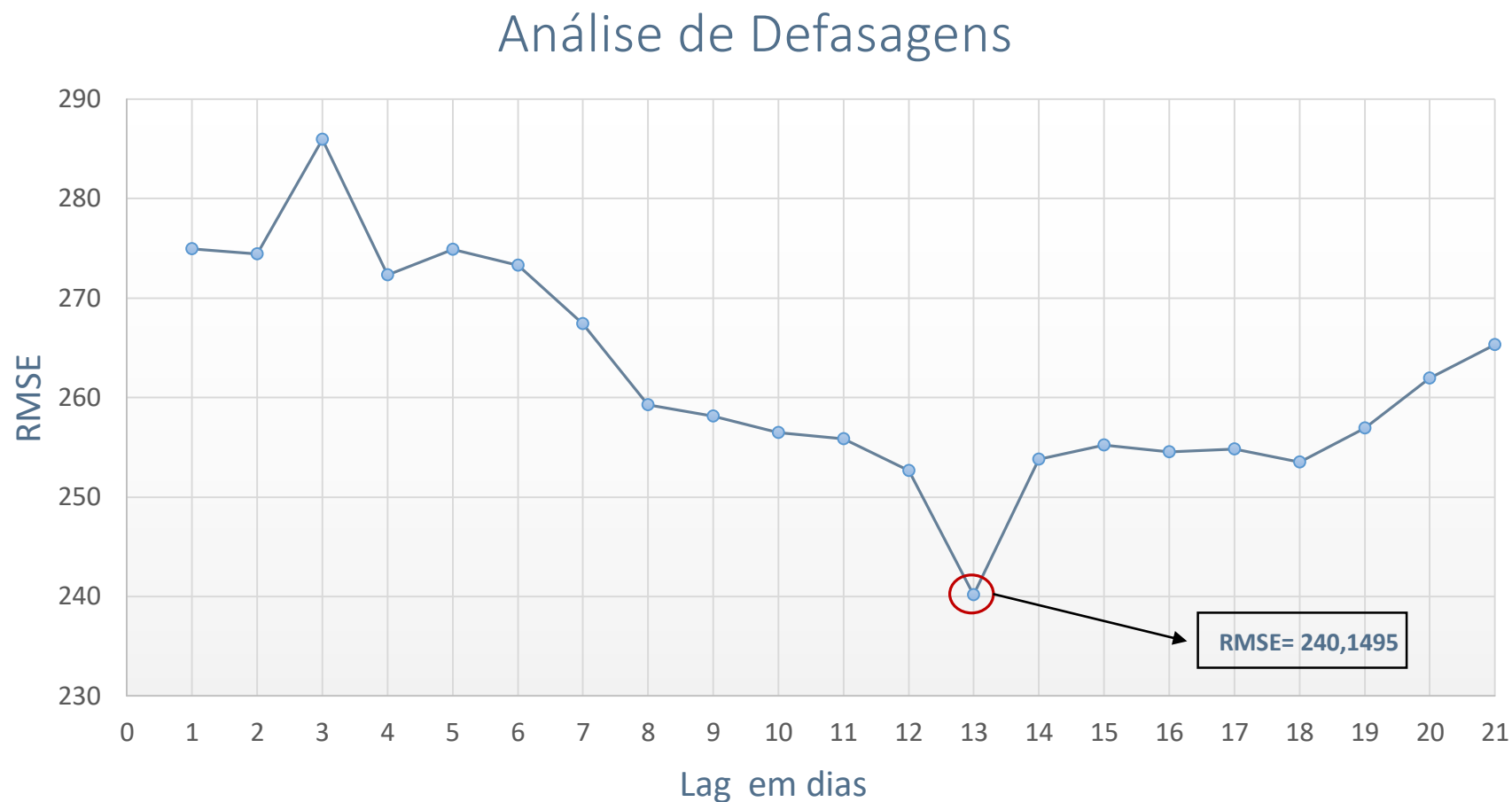
Step 5: Reducing the Learning Rate;



Hyperparameters	Step 5
Nrounds	8900
Max depth	2
eta	0.1
gamma	0
Colsample bytree	0,6
Min child weight	1
subsample:	0,5

Foram utilizados 21 dias para a análise de defasagem. A defasagem que minimizou o RMSE foi a de 13 dias como mostra o gráfico a seguir:

Lag	RMSE
1	274,9452
2	274,4291
3	285,9297
4	272,3123
5	274,9075
6	273,3154
7	267,4233
8	259,2779
9	258,1478
10	256,4694
11	255,8655
12	252,6536
13	240,1495
14	253,7656
15	255,1970
16	254,5430
17	254,8144
18	253,5248
19	256,9394
20	261,9282
21	265,2892



Teoria dos jogos

Cooperativos

(Como ser mais justo)

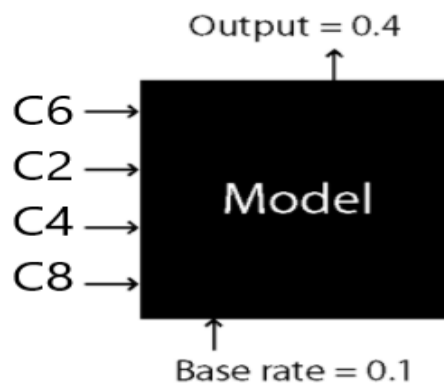
Shapley Value

Competitivos

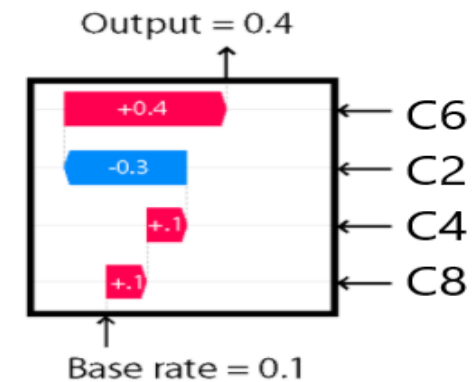
(Como ser mais esperto)

Equilíbrio de Nash

A implementação do algoritmo de SHAP desenvolvido por Scott M. Lundberg é uma técnica recente para interpretar modelos de caixa preta que possibilita a identificação da contribuição de cada variável que compõe o modelo.



Explanation

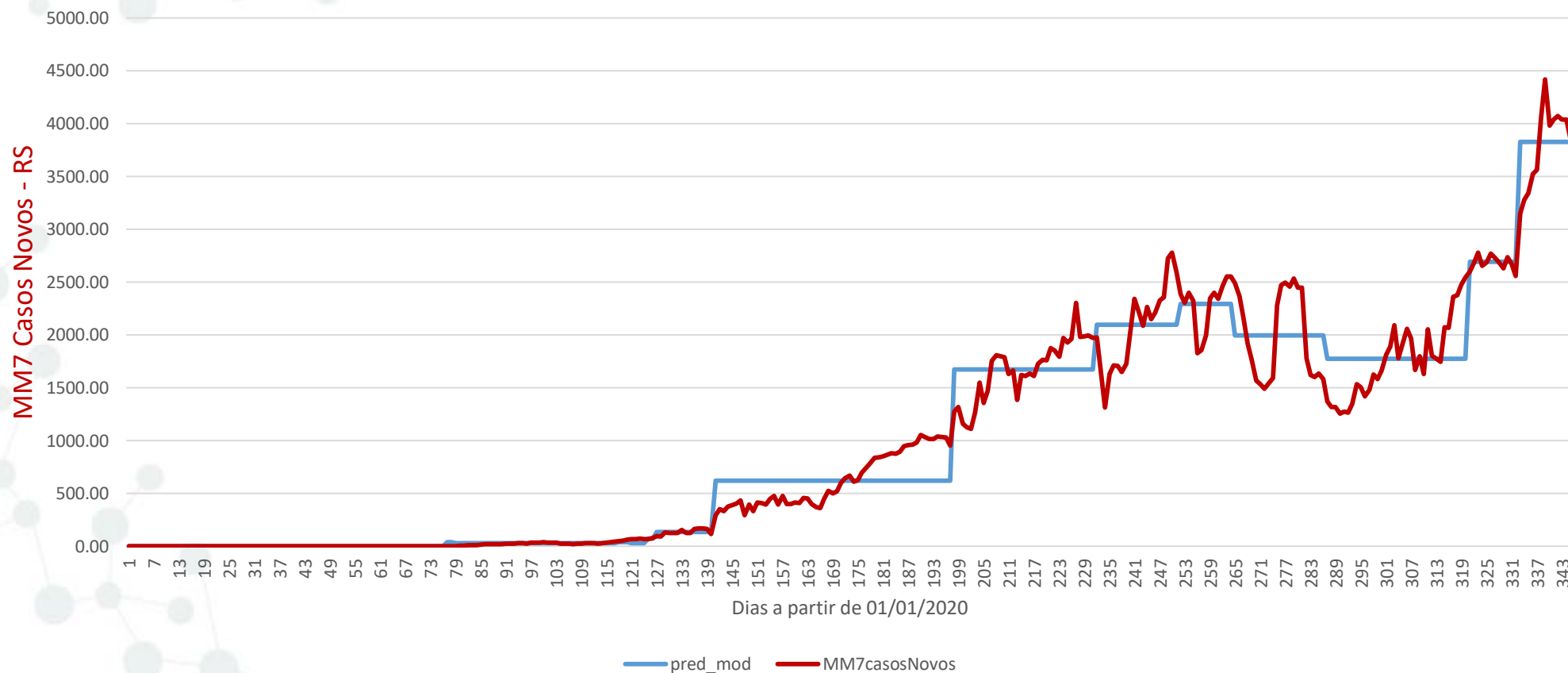


Essa nova abordagem nos permite entender um pouco mais a complexidade dos resultados do modelo preditivo, ao mesmo tempo em que nos permite explorar as relações entre variáveis para casos previstos.



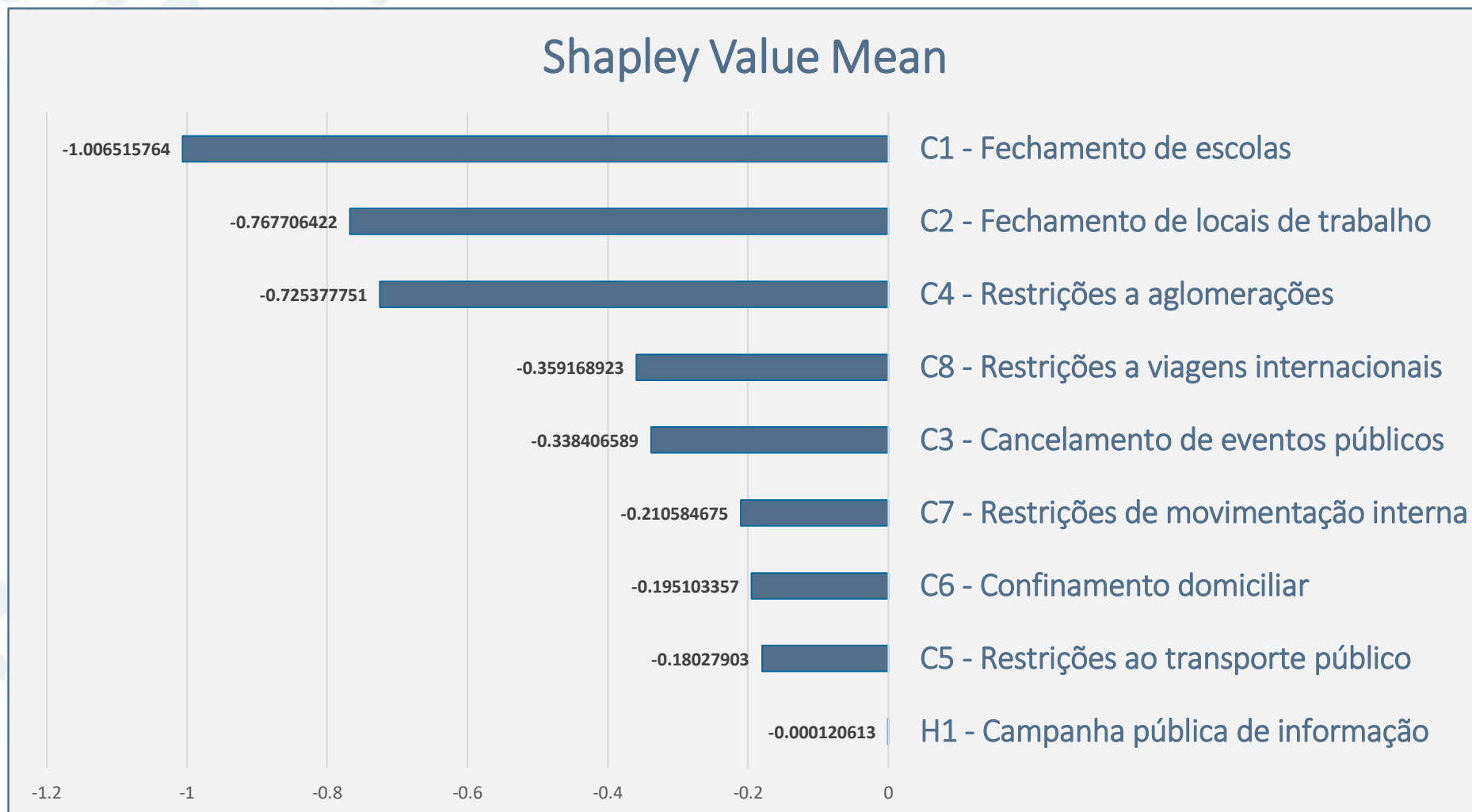
Ajuste do modelo com LAG de 13 dias:

O modelo XGBoost ao número de casos de Covid19 - RS



Obs.: O modelo ajustado foi uma regressão de Poisson.

Shapley value para as Medidas de Contenção aplicadas ao RS:



Resultados :

- ❖ Dado que o Lag de 13 dias foi o que melhor ajustou o modelo, podemos inferir que esse pode ser o tempo necessário desde a Implementação da medida de contenção até sua real efetividade.
- ❖ Podemos também afirmar que **C1 - Fechamento de escolas** foi a medida que teve maior impacto na redução do número de novos casos de covid-19, dado que teve a maior contribuição negativa para o modelo;
- ❖ Seguindo de **C2 - fechamento de locais de trabalho, C4 - Restrições a aglomerações**, que segundo o valor de Shapley respectivamente foram as 3 medidas que mais contribuíram para a redução do número de novos casos de covid-19 no Rio Grande do Sul.

- <https://www.bsg.ox.ac.uk/research/publications/brazils-fight-against-covid-19-risk-policies-and-behaviour>
- A Unified Approach to Interpreting Model Predictions (neurips.cc)
- <https://towardsdatascience.com/interpretable-machine-learning-with-xgboost-9ec80d148d27>
- <https://xgboost.readthedocs.io/en/latest/tutorials/index.html>
- Visual XGBoost Tuning with caret | Kaggle
- A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting (infona.pl)
- SHAP for XGBoost in R: SHAPforxgboost | Welcome to my blog (liuyanguu.github.io)