Shapley-based method for estimating the impact of containment measures on the number of COVID-19 cases in the state of Rio Grande do Sul, Brazil

Paulo Henrique Simões
Departament of Industrial Engineering
Pontifical Catholic University of Rio de Janeiro (PUC-Rio)
225 Marquês de São Vicente St, Gávea, Rio de Janeiro (RJ), Brazil
phsimoes33@gmail.com

Fernanda Baião
Departament of Industrial Engineering
Pontifical Catholic University of Rio de Janeiro (PUC-Rio)
225 Marquês de São Vicente St, Gávea, Rio de Janeiro (RJ), Brazil
fbaiao@puc-rio.br

Paula Maçaira
Departament of Industrial Engineering
Pontifical Catholic University of Rio de Janeiro (PUC-Rio)
225 Marquês de São Vicente St, Gávea, Rio de Janeiro (RJ), Brazil
paulamacaira@puc-rio.br

ABSTRACT

This work proposes applying Shapley's cooperative game theory to classify and rank the containment measures enforced by the government in response to the COVID-19 pandemic in Brazil, with respective to their impact on the evolution of the number of daily cases of COVID-19 in the state of Rio Grande do Sul. Data about the containement measures was collected by the Oxford COVID-19 Government Response Tracker (OxCGRT) project, and our proposed method applied the XGBoost machine learning algorithm to learn a model using Poisson regression through successive iterations of decision trees. To interpret the model, we used the SHapley Additive exPlanations (SHAP), an algorithm based on Shapley's theory, which is a robust method capable of quantifying the contribution that each containment measure brings to the model. Our experimental resuts evidenced that closing schools was the measure that mostly contributed to reducing the number of cases (SHAP value of –1.01), followed by closing workplaces (SHAP value of –0.77) and restricting agglomerations (SHAP value of –0.73). Public information campaigns presented a SHAP value of zero (0.00), which indicates that they did not contribute to the progress of the number of cases of COVID-19.

KEYWORDS. COVID-19 pandemic. XGBoost. SHAP.

SA. EST&MP. BDA

https://proceedings.science/p/157431?lang=pt-br

1. Introduction

Since the declaration of a public health emergency in Wuhan Province, China in December 2019, several patients with a severe accute respiratory syndrome (SARS) have been diagnosed with COVID-19 disease [Zhu et al., 2020]. The SARS-CoV-2 virus demonstrated a high inter-human transmission capacity, spreading rapidly around the world. On Mar 11, 2020, just over three months after the first case in China, it became a pandemic with thousands of infected people and deaths worldwide. What led the World Health Organization to declare this pandemic was the spreading speed of the new coronavirus, with numerous cases breaking out in Asian countries and quickly moving to Europe, having Italy as its initial epicenter, and later spreading towards other continents [WHO et al., 2020]. Meanwhile, countries in the Americas began to report the first cases of the disease. In Brazil, the first case was confirmed on Feb 26, 2020, in São Paulo, with a travel history coming from Italy. With the accelerated spread of the virus, the country's southeast region (where the state of Rio Grande do sul is located) became the pandemic epicenter in May 2020 [King et al., 2020].

In the absence of vaccines and antiviral drugs to curb the COVID-19 epidemic until November 2020 [Campos, 2020], WHO suggested public authorities worldwide to adopt non- pharmacological public health measures, including isolation, quarantine, social distancing, and the use of masks [Wilder-Smith e Freedman, 2020]. Adopting containment measures significantly impacts the epidemic dynamics of COVID-19, preventing an overload of health systems [Figueiredo et al., 2020]. Researchers from Imperial College London defended that drastic measures of social distancing, extended to the entire population, should be the preferred policy option, albeit this choice depends on the feasibility of its implementation and the social contexts of each country [Ferguson et al., 2020]. Reducing mobility, on the other hand, contributes to slowing the growth of the number of cases within a city and prevents transmission to other locations [Kraemer et al., 2020; Chinazzi et al., 2020; Boldog et al., 2020; Lau et al., 2020].

Portugal was reffered to as an excellent example of practical action, whose authorities acted early in implementing containment measures. The quick and assertive decision of the Portuguese government and the massive adhesion of the population reduced the advance of the first pandemic wave, suggesting that containment measures effectively reduced the contamination rate [Peixoto et al., 2020]. Moreover, Ganem et al. [2020] showed that the non-adoption of social distancing measures could hamper ICUs' capacity for COVID-19 in Brazil, and in particular in the São Paulo Metropolitan Region. The authors also suggested that the set of adopted social distancing measures could have avoided the overload of the health system (maintaining hospital occupation at 76%), saving the death of almost 90 thousand people during the epidemic. In addition, another Brazilian study showed that the maintenance and strengthening of measures such as social distancing, quarantine, and case isolation are necessary to avoid the collapse of health systems in the country [Canabarro et al., 2020]. Thus, there is evidence that the pandemic control strategy is effective when case isolation and quarantine are combined with a set of social distancing measures that cover the entire population [Nussbaumer-Streit et al., 2020].

However, there are no evidences about which measures are more effective, which makes it dificult to assess, for example, whether travel restriction is more effective (or not) than closing schools. In particular in Brazil, where there was no centralization of containment measures by the Federal Government due to political issues, each state adopted a different set of measures, with a different schedule, which intrinsically represented regional differences in actions against the pandemic. Therefore, a detailed analysis of the measures adopted in each state is pertinent in that a general study on Brazil would not capture these local particularities, biasing the study.

https://proceedings.science/p/157431?lang=pt-br

Hence, this study aims to measure the effect of containment measures in reducing the number of new cases of COVID-19 using information from the state of Rio Grande do Sul, where the first case identified was on Feb 29, 2020 and, in the 41st epidemiological week, there were 34,337 confirmed cases. It should be noted that Brazilian states have broad social, economic, and political differences [Silva e de Oliveira, 2020]. This study focuses on investigating the same nine indicators that make up the Stringency Index created by the University of Oxford, an aggregated form of government responses, showing the level of rigor imposed by local governments.

For this purpose, the SHAP (SHapley Additive exPlanations) algorithm will be used, which was inspired by methods of model interpretability. Lundberg et al. [2019] proposed a unified approach to explain the output of machine learning models. The application of SHAP is justified since the XGBoost (Extreme Gradient Boosting) algorithm is not directly interpretable. Although XGBoost works fine, has excellent metrics and few prerequisites, we do not have as much control over how it operates internally concerning its variables.

SHAP is based on the Shapley value that comes from cooperative game theory. It can attach importance to the factors that make up the model. Specifically, Shapley values can be used to find a fair division that defines how the total importance should be distributed among the factors. Thus, it is possible to understand the contribution of each variable to the model [Algaba et al., 2019].

That said, it is then possible to establish a hierarchy of effectiveness among all containment measures, whose importance is necessary to implement public policies to fight the pandemic and reduce the number of cases caused by COVID-19. This process can assist in decision-making by health authorities for a possible flexibilization of social distancing measures. Finally, it enables a partial resumption of society's daily activities and reduces economic impacts arising from stricter restrictions.

2. Methods

2.1. Database

The Oxford COVID-19 Government Response Tracker (OxCGRT) [Hale et al., 2021], is an initiative from Oxford University that is responsible for monitoring and reporting the different government responses to COVID-19 around the globe. The monitored interventions (a.k.a. containment measures) are encoded into 23 indicators, including for example school closings, travel restrictions, and vaccination policy. These containment measures policies are scaled into scores reflecting to what extent each government action was enforced, and the scores are aggregated into a set of policy indexes. In the present work, the dataset from the project report OxCGRT - BSG-WP-2020-036-PT [Petherick et al., 2020] was used, since it deals with the responses to the COVID-19 pandemic in Brazil.

The Brazilian project within the OxCGRT initiative is carried out in partnership with EBAPE-FGV of Rio de Janeiro, USP, and the University of the State of Pará. The data offers a systematic way of tracking the government responses to COVID-19 in the country and state jurisdictions. This database provides information on the 26 states, the federal district, and 52 cities, including 26 capitals and the cities with the highest number of inhabitants in each state that are not capitals. The database version used in the present work included 12 indicators regarding closure and containment policies and health policies followed by the number of new daily cases of COVID-19, from Jan 1 to Oct 14, 2020.

Brazil's state-level database is presented in different formats. In the present work, we use data that captured the total number of policies enforced in a given jurisdiction. This format is identified by the suffix "TOTAL". It includes measures adopted by the highest levels of government

that may supersede local policies, such as the federal government's ban on entry of foreigners, which is also applied to all federal units.

The nine indicators used in this study are the same ones used to construct the OxCGRT database Constraint Index: C1-closing of schools, C2-closing of workplaces, C3-cancellation of public events, C4-restrictions on agglomerations, C5-restrictions on public transport, C6-home confinement, C7-restrictions to internal movement, C8-international travel restrictions, and H1-public information campaign.

2.2. Statistical analysis

To model the impact of the nine containment measures categories in reducing the number of daily cases of COVID-19 in the state of Rio Grande do Sul, a tree model for Poisson regression was built using Extreme Gradient Boosting (XGBoost). Poisson regression is indicated for cases of discrete counts [McCullagh e Nelder, 2019].

XGBoost belongs to the category of Ensemble Learning models, which are based on training several simple models to produce a more robust final model. The algorithm builds decision trees for each iteration, where the models are no longer trained independently but sequentially, based on adjusting the previously trained models [Chen et al., 2022].

In order to optimize the hyperparameters of the XGBoost model, we used a visual implementation of the `caret` [Kuhn, 2021] package of R software [R Core Team, 2021]. This is an optimization process for XGBoost's hyperparameters. It focuses on visually evaluating the steps in the process while applying a logical, simple, and gradual approach to adjust the model. This approach uses five steps to optimize hyperparameters. In the first step, the eta learning rate and the number of iterations are established; in a second step, the maximum depth of a tree and minimum sum of instance weight are defined; in the third step of the process, the subsample ratio of columns and subsample ratio of lines are optimized; in the fourth step, experiments with different gamma values are performed; finally, in the fifth step, the learning rate is optimized. The rationale was to initially use a higher learning rate to adjust the hyperparameters. Those parameters will be used to tune the final model with a larger number of trees and a lower learning rate. Another obligation at this stage of the process is to use cross-validation methodology, since simple random resampling of the time series is not the best way to resampling the data [Hyndman e Athanasopoulos, 2018]. Hence, we applied blocking techniques to split the original dataset into training and testing sets along the time series, using TimeSlices function from the Caret package.

Poisson regression was chosen for this study because it uses counting variables. The XGBoost algorithm fits this Poisson regression through the construction of a random forest (i.e., an ensemble of decision trees). To address the lack of intelligibility of this random forest model, the SHapley Additive ExPlanations (SHAP) was applied to calculate the contribution values of the nine features composing the input dataset (i.e., the types of containment measures). Specifically, SHAP value is characterized by the average marginal contribution among all possible combinations of the factors that make up the model, and represent a fair distribution of the total importance among them [Lundberg et al., 2019].

3. Results

The time series used corresponds to the number of new daily COVID-19 cases in the state of Rio Grande do Sul from Jan 2020 to Dec 10, 2020 (345 days). We applied 7-day moving average (7MM) to smooth the time-series of COVID-19 cases, to minimize the variability of the report rate within each week, including a high underreporting rate on weekends (PRADO, 2020). Figure 1 shows the restriction index and the 7MA of cases. It is possible to observe a very rapid increase

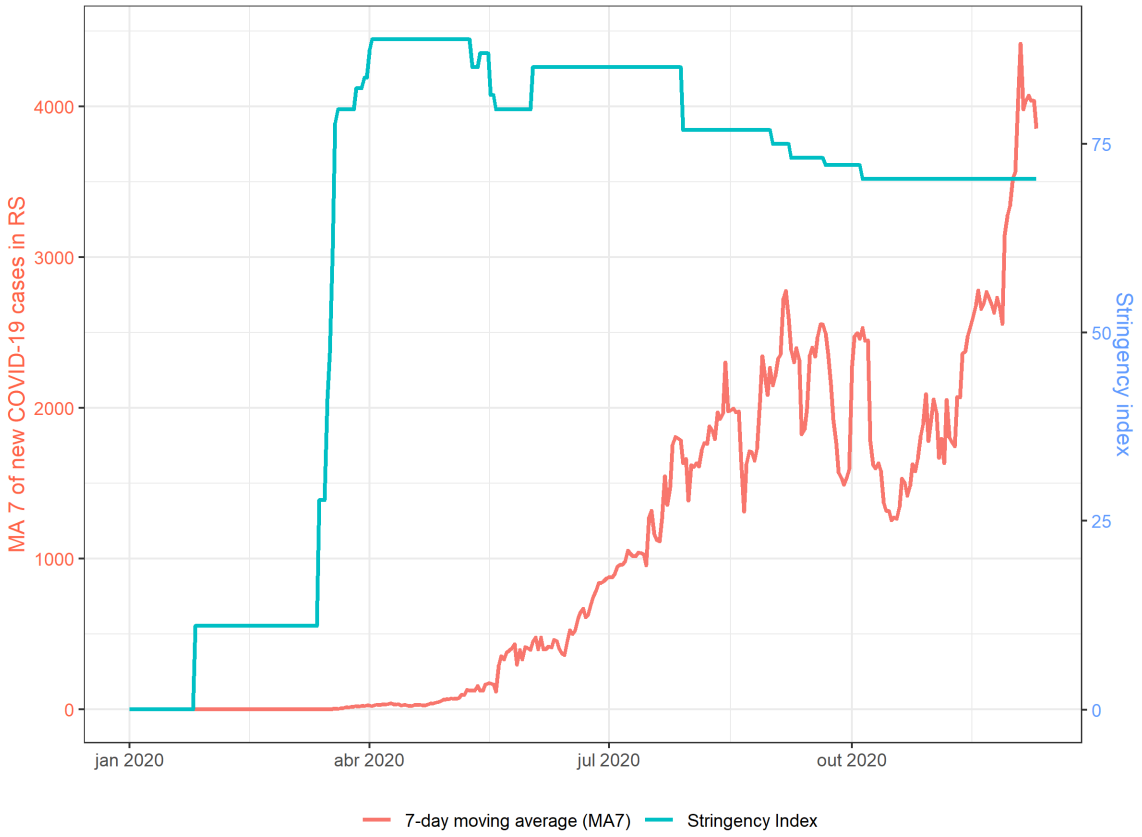


Figura 1: Stringency index & 7-day moving average (MA7) of new cases of COVID-19 in the state of Rio Grande do Sul, Brazil.

in restrictions imposed by the government of RS/Brazil in the initial pandemic period and then a decrease in these restrictions, while the number of cases continued to rise.

Since the containment measures do not take effect immediately but, rather, between 7 to 14 days after being imposed (PEIXOTO et al., 2020), we assessed the fitness of the Poisson Regression model built by the XGBoost algorithm, by applying a time lag ranging between 1 and 21 days. The lag that minimized the Root Mean Square Error (RMSE) was 13 days, as shown in Figure 2. Meaning that, on average, the containment measures are more effective 13 days after their implementation.

After constructing the regression model, SHAP is applied to assess the contribution of containment measures to the trajectory of the number of COVID-19 cases in the state of Rio Grande do Sul. A Shapley value was calculated for each day and each containment measure. The final SHAP value is the average of the 345 days.

In Figure 3, the negative results are directly related to the reduction in the number of cases. The lower the Shapley value, the more significant their contribution to reducing cases of COVID-19. Thus, “closing schools”(C1) was the measure that most contributed to reducing the number of cases with –1.01, followed by “closing workplaces”(C2) with –0.77 and “restrictions on agglomerations”(C4) with –0.73. “Public information campaigns”(H1) had zero SHAP value, meaning that they did not contribute to reducing the cases of COVID-19.

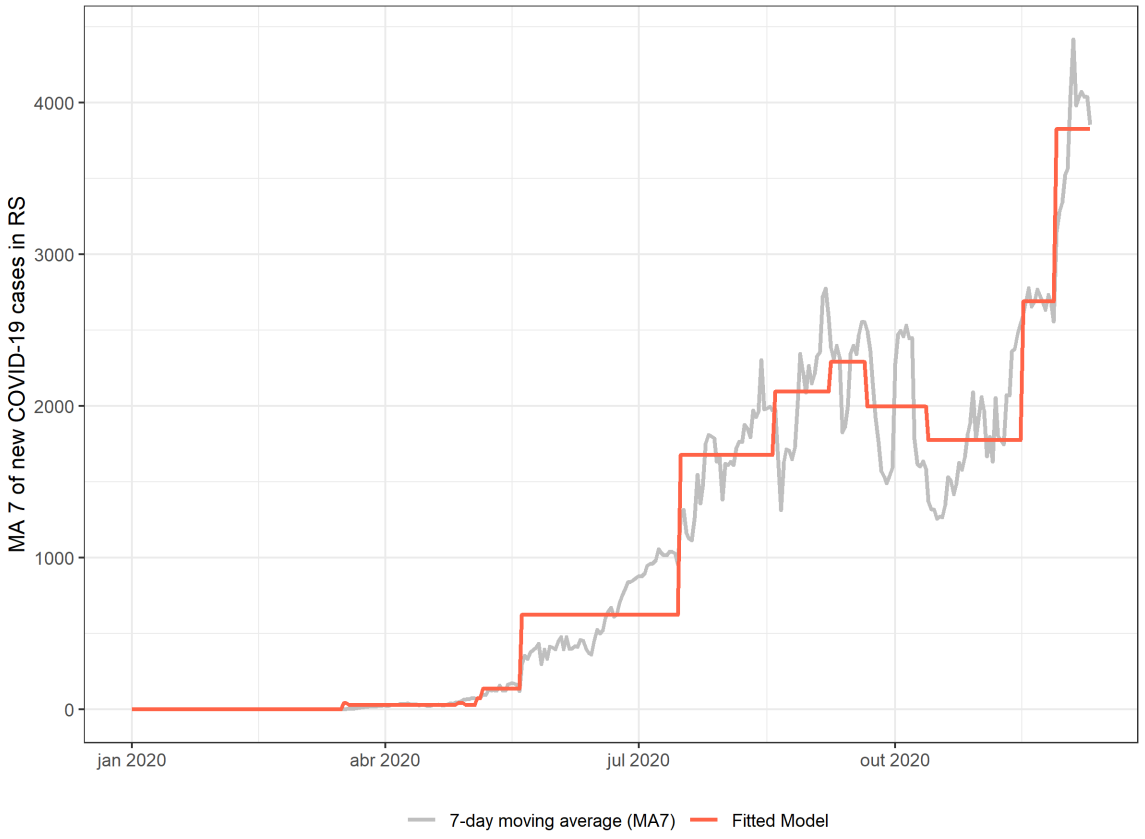


Figura 2: Poisson-fitted regression model by the XGBoost algorithm (red) and 7-day moving average (MA7) (grey) of the number of COVID-19 cases in the state of Rio Grande do Sul, Brazil.

4. Discussion and Conclusion

The experiment conducted in this study assessed the impact of the nine containment measures (C1 to C8) and of public information campaigns (H1) that were enforced in the state of the Rio Grande do Sul, Brazil, on combating COVID-19. For this, the eXtreme Gradient Boosting (XGBoost) algorithm was applied to build a Poisson regression model that fits the COVID-19 daily cases curve, using the containment measures as explanatory variables.

Shapley's method was applied to the final model to assign a fair value to each of the containment measures, thus ranking them with respect to their impact in reducing the number of COVID-19 cases.

Based on the results obtained for the period of the study, we can state that closing schools (C1) had the most significant impact on reducing the number of new cases of COVID-19 in RS, followed by the closure of workplaces (C2) and restrictions on agglomerations (C4), which were also very impactful on this reduction. On the other hand, public information campaigns (H1) did not reduce the number of cases of COVID-19.

Likely, the effect observed in the study is directly linked to the human factor when dealing with restrictions and socioeconomic factors. When applied to low- and middle- income countries (LMIC) like Brazil, these elements make it difficult to implement measures such as home confinement. According to the preliminary results of the [Nadalin et al., 2020] there were 133,021

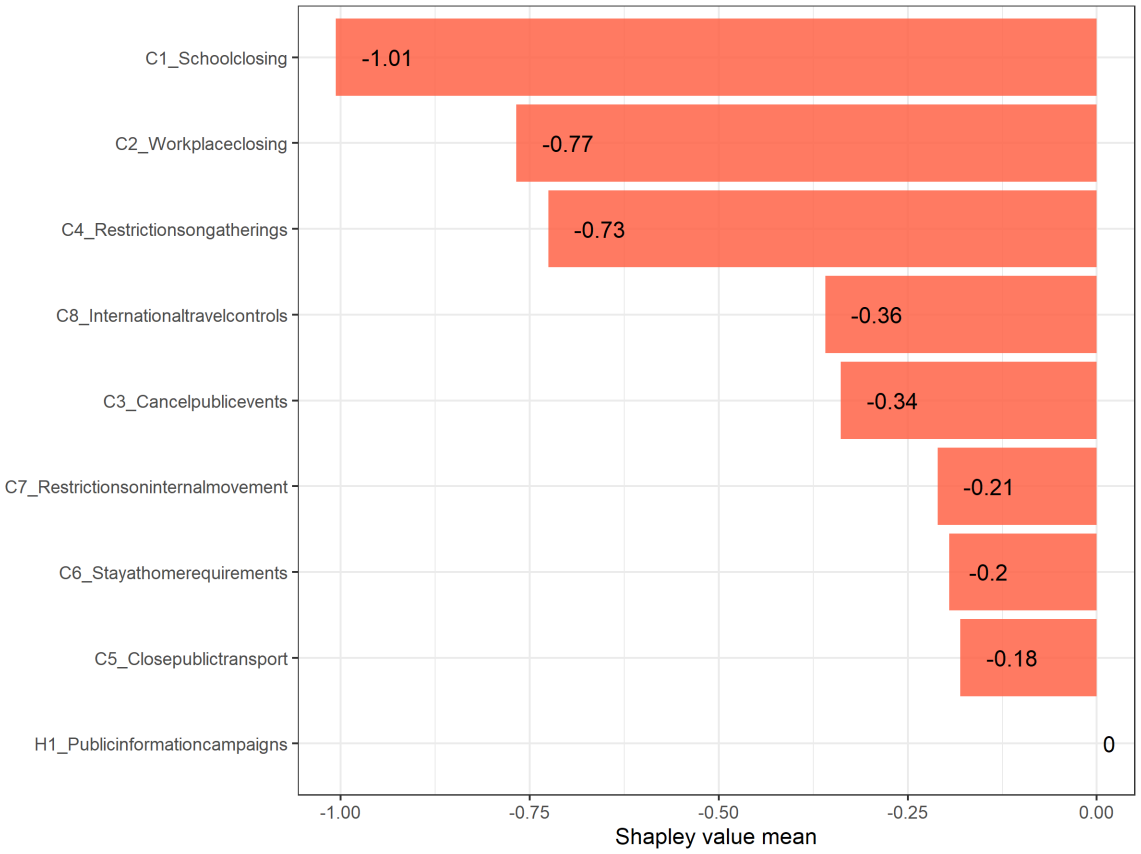


Figura 3: Shapley Values Mean for each type of containment measure in reducing the number of COVID-19 cases in the state of Rio Grande do Sul/Brazil.

households in a subnormal agglomeration situation in the state of Rio Grande do Sul, which represents 3.5% of the houses in the state. This informations corroborate with one of the results in this work, that (C6) stay at home requirements had a little contribution to containment measures.

Another issue addressed in the study was the estimation of the time interval from implementing the containment measure to its actual effectiveness in reducing the number of cases of COVID-19. Given that the lag of 13 days resulted in the smallest error, it can be inferred that this was the time needed for the containment measures to mostly affect the curve of cases of RS. This result corroborates the existing literature. In Portugal, it took about two weeks (14 days) after implementing the containment measures to reduce new cases and the number of hospitals and intensive care units [Peixoto et al., 2020].

As conclusion, these study implemented a model to measure and rank the measures adopted in the state of RS, Brazil, that most contributed to reducing the number of cases of COVID-19. The final model was able to reveal the marginal contribution of each containment measure, in other words, to infer which of them had the most significant impact during the first year of the pandemic. This process can and should improve and optimize public policies to combat the virus spread, especially in the Brazilian scenario of social inequality and lack of a centralized governance.

5. Acknowlodgments

This work was developed as part of the EFFECT-BRAZIL Project [1], which is part of the Grand Challenges ICODA pilot initiative delivered by Health Data Research UK and funded by the Bill & Melinda Gates Foundation and the Minderoo Foundation. This work was supported by the Brazilian Coordination for the Improvement of Higher Level Personnel (CAPES) under grant number [001]; the Brazilian National Council for Scientific and Technological Development (CNPq) under grants numbers [422470/2021-0 and 422810/2021-5], the Carlos Chagas Filho Research Support Foundation of the State of Rio de Janeiro (FAPERJ) under grants numbers [211.308/2019, 211.086/2019, 211.645/2021, and 201.348/2022]. The authors would also like to thank Blavatnik School of Government, FGV EBAPE - Brazilian School of Public and Business Administration, and USP - University of São Paulo, for contributing to the construction and updating of the Global Oxford COVID-19 Government database Response Tracker (OxCGRT).

Referências

Algaba, E., Fragnelli, V., e Sánchez-Soriano, J. (2019). *Handbook of the Shapley value*. CRC Press.

Boldog, P., Tekeli, T., Vizi, Z., Dénes, A., Bartha, F. A., e Röst, G. (2020). Risk assessment of novel coronavirus covid-19 outbreaks outside china. *Journal of clinical medicine*, 9(2):571.

Campos, G. W. d. S. (2020). O pesadelo macabro da covid-19 no brasil: entre negacionismos e desvarios. *Trabalho, Educação e Saúde*, 18.

Canabarro, A., Tenório, E., Martins, R., Martins, L., Brito, S., e Chaves, R. (2020). Data-driven study of the covid-19 pandemic via age-structured modelling and prediction of the health system failure in brazil amid diverse intervention strategies. *Plos one*, 15(7):e0236310.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., e Yuan, J. *xgboost: Extreme Gradient Boosting*, 2022. URL `https://CRAN.R-project.org/package=xgboost`. R package version 1.6.0.1.

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Mu, K., Rossi, L., Sun, K., et al. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science*, 368(6489):395–400.

Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., et al. (2020). Impact of non-pharmaceutical interventions (npis) to reduce covid-19 mortality and healthcare demand. *Imperial College COVID-19 Response Team London*.

Figueiredo, A. M. d., Figueiredo, D. C. M. M. d., Gomes, L. B., Massuda, A., Gil-García, E., Vianna, R. P. d. T., e Daponte, A. (2020). Determinantes sociais da saúde e infecção por covid-19 no brasil: uma análise da epidemia. *Revista Brasileira de Enfermagem*, 73.

Ganem, F., Mendes, F. M., de Oliveira, S. B., Porto, V. B. G., de Araújo, W. N., Nakaya, H. I., Diaz-Quijano, F. A., e Croda, J. (2020). The impact of early social distancing at covid-19 outbreak in the largest metropolitan area of brazil. *MedRxiv*.

[1] https://icoda-research.org/project/dp-effect-brazil/

https://proceedings.science/p/157431?lang=pt-br

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., et al. (2021). A global panel database of pandemic policies (oxford covid-19 government response tracker). *Nature human behaviour*, 5(4):529–538.

Hyndman, R. J. e Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

King, A., Andrus, J. K., e Figueroa, J. P. (2020). Financial crisis at paho in the time of covid-19: a call for action. *The Lancet*, 396(10244):96.

Kraemer, M. U., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., Group†, O. C.-. D. W., Du Plessis, L., Faria, N. R., Li, R., et al. (2020). The effect of human mobility and control measures on the covid-19 epidemic in china. *Science*, 368(6490):493–497.

Kuhn, M. *caret: Classification and Regression Training*, 2021. URL `https://CRAN.R-project.org/package=caret`. R package version 6.0-90.

Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J., e Khosrawipour, T. (2020). The positive impact of lockdown in wuhan on containing the covid-19 outbreak in china. *Journal of travel medicine*.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., e Lee, S.-I. (2019). Explainable ai for trees: From local explanations to global understanding. *arXiv e-prints*, p. arXiv–1905.

McCullagh, P. e Nelder, J. A. (2019). *Generalized linear models*. Routledge.

Nadalin, V. G., Krause, C., e Neto, V. C. L. (2020). Aglomerados subnormais – ibge. `https://www.ibge.gov.br/geociencias/organizacao-do-territorio/tipologias-doterritorio/15788-aglomerados-subnormais.html`. Acessado: 2020-05-04.

Nussbaumer-Streit, B., Mayr, V., Dobrescu, A. I., Chapman, A., Persad, E., Klerings, I., Wagner, G., Siebert, U., Ledinger, D., Zachariah, C., et al. (2020). Quarantine alone or in combination with other public health measures to control covid-19: a rapid review. *Cochrane Database of Systematic Reviews*, 9.

Peixoto, V. R. F. D., Vieira, A. A. M., Aguiar, P., Sousa, P., e Abrantes, A. V. (2020). "timing", adesão e impacto das medidas de contenção da covid-19 em portugal. *Escola Nacional de Saúde Pública, Universidade Nova de Lisboa*.

Petherick, A., Goldszmidt, R., Kira, B., e Barberia, L. (2020). Do brazil's covid-19 government response measures meet the who's criteria for policy easing? *Blatnavik School of Government*.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL `https://www.R-project.org/`.

Silva, D. F. e de Oliveira, M. L. C. (2020). Epidemiologia da covid-19: comparação entre boletins epidemiológicos. *Comunicação Em Ciências Da Saúde*, 31:61–74.

WHO et al. (2020). Who director-general's opening remarks at the media briefing on covid-19-11 march 2020. 2020.

Wilder-Smith, A. e Freedman, D. O. (2020). Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-ncov) outbreak. *Journal of travel medicine*.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al. (2020). A novel coronavirus from patients with pneumonia in china, 2019. *New England journal of medicine*.